Soliman, Dr. Marwah, Nathaniel K. Newlands, Dr. Vyacheslav Lyubchich, and Yulia R. Gel. 2023. "Multivariate Copula Modeling for Improving Agricultural Risk Assessment under Climate Variability." *Variance* 16 (1).

Financial and Statistical Methods

Multivariate Copula Modeling for Improving Agricultural Risk Assessment under Climate Variability

Dr. Marwah Soliman[1a], Nathaniel K. Newlands[2b], Dr. Vyacheslav Lyubchich[3c], Yulia R. Gel[4d]

[1] Department of statistics, London School of Economics, [2] Government of Canada (Agriculture and Agri-Food Canada), Summerland Research and Development Centre, [3] Statistics, University of Maryland Center for Environmental Science, [4] Statistics, University of Texas - Dallas

Keywords: Agricultural insurance, Basis risk, Crop yield, Multivariate copula, Normalized difference vegetation index

Variance

Vol. 16, Issue 1, 2023

Agricultural production is highly vulnerable to both short-term extreme weather events and long-term climate variability and change. These impacts propagate further and result in socioeconomic changes affecting farmers, insurers, and other stakeholders across agricultural supply chains. As a result, the most recent challenges in addressing resiliency and sustainability at the face of climate change require development of innovative multivariate methods for quantifying crop yield risk driven by factors that are strongly spatially and temporally dependent. Copulas offer a systematic solution to tackle this spatio-temporal uncertainty quantification problem. However, utility of copulas in agricultural risk assessment and insurance remains largely under-explored. We introduce multivariate copula modeling (MCM) for capturing yield-climate dependence and evaluate its utility by benchmarking its performance on a multi-scale yield-climate dataset against state-of-the-art competing model-based approaches. MCM is found to outperform traditional statistical approaches and better explain complex dependence structure over time and space between crop yield and climate. Our findings highlight the benefits of MCM for reducing basis risk and improving the robustness of insurance premium rate-making.

Address for Correspondence: Marwah.Soliman@utdallas.edu

a Dr. Marwah Soliman is an LSE Fellow at Department of statistics, London school of Economics, London, UK. She received her PhD from Department of statistics at University of Texas at Dallas in 2020, She did a two year postdoc at Department of Statistics in Clemson University, SC, USA before Joining LSE.

b Nathaniel K. Newlands is a Research Scientist with the Government of Canada (Agriculture and Agri-Food Canada) based at the Summerland Research and Development Centre in British Columbia, Canada, and an Adjunct Professor in Geography with the University of Victoria, Canada. His research work addresses public-good food-water-energy nexus issues and tackles broad, integrated, complex global problems to help support and advance global sustainable development. Nathaniel is a co-chair of the Group on Earth Observations (GEO)'s Disaster Risk Reduction Working Group and Subgroup 2 Lead on UNDRR Sendai Framework Monitoring and Global Assessments. He is Vice-Chair of the American Statistical Association (ASA)'s International Relations Committee, and International Board Member (North American Region) of the International Environmetrics Society (TIES).

c Dr. Vyacheslav Lyubchich is an Associate Research Professor of Statistics at the University of Maryland Center for Environmental Science. He received his Ph.D. in Statistics from the Orenburg State University, Russia in 2011. In the same year, he was awarded the Government of Canada postdoctoral fellowship to continue his research in time series methodology at the Department of Statistics and Actuarial Science of the University of Waterloo, Canada. Since 2015, V. Lyubchich is a research faculty and a founding member of the Environmental Statistical Collaborative at the University of Maryland Center for Environmental Science.

d Yulia R. Gel is Professor of Statistics at UT Dallas. Her research interests include statistical foundation of data science, spatio-temporal modeling and predictive analytics. She holds PhD in Mathematics from Saint Petersburg State University, Russia, followed by a postdoctoral position at University of Washington, USA. Prior to UT Dallas, she was a tenured faculty member at University of Waterloo, Canada. She held visiting positions at Johns Hopkins University, University of California, Berkeley, and Newton Institute, Cambridge University, UK. She is served as President of the International Environmetrics Society and as Vice President of the International Society on Business and Industrial Statistics. She is Fellow of the American Statistical Association.

1. INTRODUCTION

CLIMATE-SMART AGRICULTURE

Global food systems, food security, and human health are highly vulnerable to extreme weather and climate events such as windstorms, heatwaves, floods, and droughts (Myers et al. 2017; Tack and Ubilava 2015; Gobin, Tarquis, and Dalezios 2013). The frequency and intensity of extreme events continue to increase as global warming accelerates. Many agricultural production areas are experiencing longer growing seasons and warmer winters due to the climate changes now taking place. This increases variability and losses in yield quantity and quality due to waterlogging, crop heat stress, invasive pests, and overwintering diseases (Bush and Lemmen 2019; Newlands 2016). Reliably attributing specific events in space and time to broader climate variability and climate change is a challenging problem. Accurately predicting the frequency, duration, and intensity of future extreme events is also a major challenge (NAS 2016; Sillmann et al. 2013; Zhang et al. 2011; Easterling et al. 2000). Climate-smart adaptation strategies for the Canadian Prairies, one of the major agricultural growing areas for the global supply of wheat and canola, will very likely be required to mitigate yield risks posed by climate change, such as late seeding decisions due to increased spring precipitation (Renard and Tilman 2019; Carew et al. 2018; Martens, Entz, and Wonneck 2015). Historically, past adaptation strategies explain observed trends in yield (across years 1976–1998), whereby farmers extended and diversified their cropping systems, allocating larger areas to oilseed and pulse crops for extended periods in more humid regions with less risk of yield depression and drought, instead of leaving the areas fallow (i.e., summer fallow) (Campbell et al. 2002). Projected climate change is expected to increase yield variability in both the medium term (2034–2050) and long term (2079–2095), under both low- and high-carbon emission scenarios with production variance effects differing in space and time, as evidenced across crop districts in the province of Manitoba (Carew et al. 2018).

WEATHER INDEX-BASED INSURANCE, BASIS RISK, AND COPULAS

Index-based insurance refers to insurance policies with payoffs contingent on the value (e.g., mean or threshold) of an underlying index such as temperature, rainfall, satellite-derived vegetation, or crop stress measurement indices. Index models are typically based on a single variable (i.e., univariate), not multiple variables (multivariate). More complex insurance products use multiple indices at once or apply a statistical model to compute insurance basis risk, premiums, and payoffs. For example, Erhardt and Engler (2018) developed and validated a multivariate model for spatially-dependent predictions of temperature—an autoregressive moving average model with generalized autoregressive conditional heteroscedasticity of residuals (ARMA(3,0)-EGARCH(3,1))—that used bootstrapping to incorporate parameter uncertainty into the simulations. This model demonstrates how actuarial risk measures of the portfolio can be estimated from bootstrap distributions. To be a reliable indicator, an index must be representative of crop loss across many locations having diverse climate regimes and agricultural practices. We refer readers to recent reviews by Lyubchich et al. (2019), Vyas et al. (2021) and Abdi et al. (2021) on copulas, agricultural insurance, and risk management.

The problem of spatial basis risk is fundamental to index insurance and yet is still largely unexplored in the literature (Yu et al. 2019; Jensen, Barrett, and Mude 2016; Norton, Boucher, and Verteramo Chiu 2015; Norton, Turvey, and Osgood 2012). One reason it may have been underexplored is a practical knowledge gap that exists between the application of more advanced multivariate statistical methods and satellite-derived indices in insurance, despite advances in such methods for joint multivariate modeling of risk in agriculture (Newlands 2016). Recent findings of Salgueiro (2019) indicate that weather index–based insurance policies based on copulas that cover both extreme and non-extreme adverse weather can effectively reduce risk exposure, measured by grape yield variance and expected shortfall. Liu et al. (2020) confirm that MODIS vegetation indices, such as normalized difference vegetation index (NDVI), are better predictors of yield at crop peak growing stages, with multiple linear regression model root-mean-squared error (using leave-one-out cross-validation, hereafter LOOCV) in the range of 18%–28% at the municipality scale (using data for the province of Saskatchewan for years 2000–2016). Deep learning has been more recently shown to better explain spatial dependence at a higher spatial resolution (regional or municipal scale) and to further improve the accuracy of the CAR regional yield forecasts (Tanguy et al. 2015; Newlands 2016). Findings demonstrate that, while deep learning outperforms classical methods at higher spatial resolution, it requires hyperparameter tuning and a sufficient amount of yield data for model training.

The copula is a "mathematical tool for modeling the joint distribution of simultaneous events. From the perspective of tail and systemic risk, the copula allows the decoupling of the marginal distribution (that which is associated with tail risk) from the dependence structure (that which is associated with systemic risk)," and enables modeling them separately with a greater precision (Staudt 2010). Copula methods have been widely applied in energy, finance, hydrology, and insurance because they have proven to be both flexible and accurate when accounting for deviations from the normal distribution—particularly, heavy tails. However, copulas have only recently been applied in agricultural risk assessment and insurance for capturing yield–climate dependence and quantifying the occurrence of extreme events (tail dependencies) (Lyubchich et al. 2019; Nguyen-Huy et al. 2018; Sadri and Burn 2014; Woodard et al. 2011). Staudt (2010) highlights the selection of copulas to appropriately capture distribution tail dependence and account for skewness and kurtosis of underlying data, but also the importance of a natural interpretation of the copula.

EARTH OBSERVATION–BASED INDICES

Indices derived from direct observation of crop growth from satellites (i.e., remotely-sensed or Earth Observational (EO) monitoring) may better track the variability in crop response due to weather and climate. EO monitoring provides data for model training and validation, and typically offers data that are free and publicly available. From an operational monitoring and forecasting standpoint, models based on satellite-derived indices can detect and track changes in crop condition (e.g., growth and stress) more quickly than those that rely on ground-based yield surveys (Zhao et al. 2020; Petersen 2018; Newlands et al. 2014; Lobell 2013). NDVI is a satellite-based vegetation index that measures plant greenness or plant chlorophyll content and photosynthetic activity as a proxy for plant health and biomass, or yield. Healthy plants absorb visible light for use in photosynthesis and reflect near-infrared radiation (NIR) (700–1300 nm wavelengths). When a plant is stressed (e.g., due to heat waves, nitrogen deficiency, or being in senescence or waterlogged from flooding), it reflects less NIR and absorbs less visible light (red portion of the spectrum, 600–700 nm wavelengths) as photosynthetic activity is less efficient due to damaged or poorly developed cell structure. The NDVI is (NIR – Red)/(NIR + Red), ranging within the interval [0, 1], with higher values representing more "green" biomass, cover type, or quantity and vigor of green biomass associated with healthier plants. Vegetation indices such as NDVI are influenced by soil background reflectance, crop type, physiographic region, crop and land management practices, and climate. The use of such vegetation indices alongside historical climate variability (i.e., temperature and precipitation) has provided valuable insights into changing growing conditions across agricultural landscapes in Canada (Kouadio et al. 2014; He et al. 2012; Wall, Larocque, and Léger 2008; Guo 2003), China (Pei et al. 2019; Yuan et al. 2019), the United States (Wang, Rich, and Price 2003), and Africa (Petersen 2018). Turvey and Mclaurin (2012) emphasize the need to calibrate NDVI based on its spatial dependency on temperature and precipitation before it can be widely applied as a useful tool in index-based insurance product design. They highlight the importance of determining the optimal period and spatial extent best suited for maximizing the strength of relationship between NDVI, climate variables, and crop yields.

1.0.1. OBJECTIVE

Our goal is to investigate the utility of multivariate copula modeling (MCM) for capturing the spatiotemporal dependence between NDVI, yields, and climate variables, and its potential for reducing basis risk. That is, we are interested in exploring whether MCM is a viable new alternative to improve insurance ratemaking involving model-based indices for setting payoff level and risk loaded premiums, based on more robust estimation and prediction of the probability of yield loss. We illustrate the utility of the MCM using a multi-scale dataset comprising validated and interpolated climate (daily, 1 km), Earth observational NDVI (weekly, 250 m), and crop yield (annual, municipal-scale) data. The MCM approach is benchmarked against the most popular competing algorithms—namely, multivariate regression with ARMA errors (MR with ARMA errors), generalized least squares (GLS), the Bayesian network (BN) approach, and the ARMA EGARCH model.

The paper is organized as follows. Section 2 describes the multivariate data in detail. In Section 3, we detail the competing models (including the MCM approach) and outline how crop yield–NDVI bias can be estimated. In Section 4, we describe training and validation of the models. In Section 5, we demonstrate and evaluate the use of the MCM model for pricing crop insurance premium, benchmarking its performance against a classic burn analysis approach. In Section 6, we present and discuss the modeling results, and Section 7 concludes the paper.

2. DATA DESCRIPTION

CROP YIELD

Historical panel data on crop yield (tonnes per acre) (2000–2018) were obtained from the Manitoba Agricultural Services Corporation (MASC) via their publicly available online MMPP Variety Yield Data Browser.[1] Spring wheat (i.e., hard red spring wheat (*Triticum aestivum L.*)) is selected as a representative crop having high spatial coverage as it is historically one of the major crops grown in Manitoba (alongside canola and barley) (Campbell et al. 2002). A subset of 37 municipalities distributed across the province were selected based on the previously assembled yield and climate data (Tanguy et al. 2015). The municipal regions vary substantially by their harvested area, from 2 to 3,572 km.[2] The municipalities of Stuartburn, West St. Paul, La Broquerie, Dauphin, and Kelsey were removed because they had insufficient annual-yield data for spring wheat, leaving a total of 32 regions.

CLIMATE

Climate data for Manitoba were acquired from Daymet,[2] providing spatially-interpolated estimates of daily maximum and minimum temperatures (°C) and precipitation (mm) at 1 km spatial grid resolution (Thornton et al. 2017). Daily maximum and minimum temperatures are aggregated into weekly means, and daily precipitation is aggregated into weekly total precipitation, to correspond to the weekly timescale of the NDVI. A boundary file (LCSD00a16a ESRI

1 MMPP Variety Yield Data Browser, https://www.masc.mb.ca/masc.nsf/mmpp_browser_variety.html

2 Daymet, https://daymet.ornl.gov

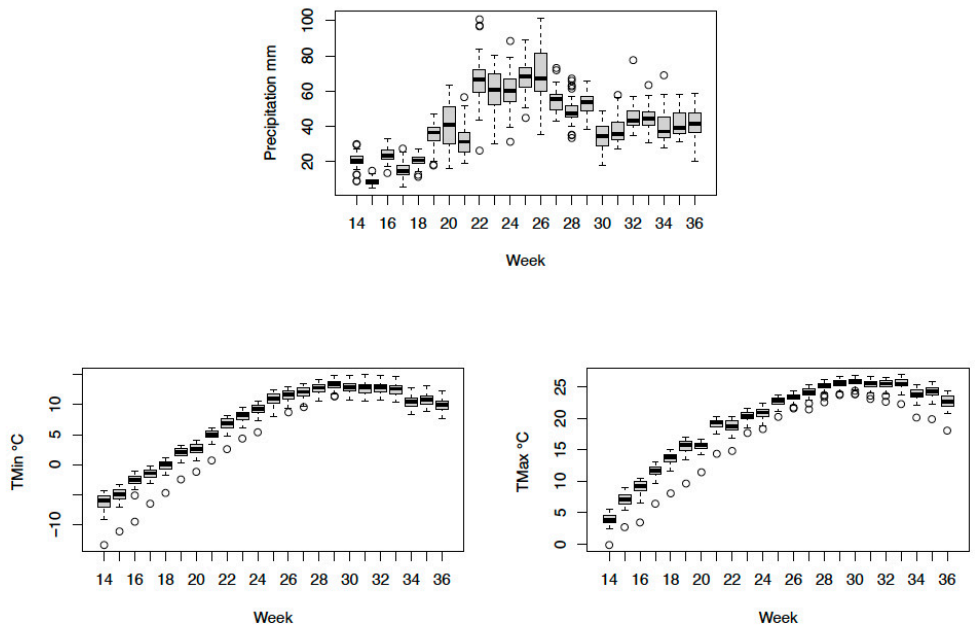


Figure 1. Historical variability of observed total precipitation (top), minimum and maximum temperatures (bottom), averaged over municipalities (2000{2018).

shapefile) available from Statistics Canada is used to delineate the municipalities as census subdivisions.[3] Centroids of each municipality (i.e., shapefile polygon) are references to obtain values of the climate variables for each region. Some municipalities have multiple centroids that consist of spatially disconnected areas. To reduce potential error in referencing centroid values, the shapefile is processed to ensure only one centroid point corresponds to each municipality, using the "Feature to point" (inside) tool in ArcGIS (ESRI Inc.). See Figure 1 showing variability of the aggregated data.

NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI)

Statistics Canada's Crop Condition Assessment Program (CCAP) provides weekly-composite NDVI data at a 250 m spatial resolution from the AVHRR (Advanced Very High-Resolution Radiometer) and the MODIS (Moderate Resolution Imaging Spectroradiometer; instrument of the Terra satellite). The CCAP NDVI product correlates well with other value-added products, such as the Canadian Drought Monitor (DroughtWatch, AAFC), and is used by Statistics Canada in national crop production reporting and by Agriculture and Agri-Food Canada (AAFC) in operational forecasting of crop yield of the major crops (wheat, canola, soybean, corn) across Canada through the growing season (Canadian Crop Yield Forecaster, CCYF) (Chipanshi et al. 2015; Kouadio et al. 2014; Newlands et al. 2014). Data derived from the MODIS satellite imagery at the 250 m resolution[4] are used in this modeling study. The MODIS sensor's red and NIR channel ranges are 630–670 nm and 841–876 nm, respectively, and lie within the red and NIR spectrum bands of 600–700 nm and 700–1300 nm indicated earlier. The NDVI dataset consists of quality-controlled weekly data (i.e., 7-day composites, Monday to Sunday) during 2000–2018 for the growing season months of April to mid-October or Julian weeks 14 (April 9–15) through week 41 (October 8–14). The aggregation into weekly composites reduces error due to cloudy days among other error corrections required for the satellite imagery. A subset of the data for the 2000–2018 reference period for the months April to September (i.e., weeks 14 to week 36) was processed for input to models coded using the R statistical language and open-source R libraries. NDVI values are extracted from the data files using the ArcGIS "Extract values to table" tool.

NDVI typically increases at the beginning of the growing season, then peaks and levels off in mid-season, then decreases to the end of season (Wall, Larocque, and Léger 2008). Basnyat et al. (2004) have determined that the optimal time when NDVI is most correlated with spring wheat yield of spring-seeded crops maturing in August in the Canadian Prairies (southern Saskatchewan semi-arid field sites) is between July 10–30. During late July to early August, crops have generally completed vegetative growth and

3 Boundary file, https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm

4 MODIS NDVI (Statcan), https://open.canada.ca/data/en/dataset/dc700f75-19d8-4913-9846-78615ca93784

are in the grain-filling stage, when crop ripening differences within a farmer's field can enhance NDVI, pointing to its use for delineating in-field management zones. For the data used in this study, the NDVI response shows a consistent growing season peak each year between Julian weeks 25–27, (i.e., June 18–July 8), which is slightly earlier than the reported optimal time of July 10–30 estimated by (Basnyat et al. 2004). The maximum value of NDVI can be computed over an entire growing season and used to predict crop yield, as an alternative metric to mean NDVI that is computed through the weeks of the growing season. For this reason, we construct and test models for both mean and maximum NDVI. Maximum NDVI varies spatially and symmetrically with most values clustered within 0.95–1.2 t/acre of the yield and with yield estimates corresponding to high NDVI (0.87–1.00), with some regions having consistently higher/lower yields. In Canada, the average farm size in 2011 was 778 acres (i.e., 315 ha or 3.15 km^2), whereby NDVI at the 250 m resolution can enable vegetation monitoring of conditions within a single field. If a field is cropped during one week, remaining neighboring fields in the pixel can influence the resulting NDVI so that the index does not fluctuate substantially. Such an effect has been calibrated and error-corrected by Statistics Canada, along with other corrections to minimize effects of atmospheric residual contaminants and cloud cover.

3. METHODS

Exploratory analysis of the multivariate data included quality checks, standard variable importance, pair-wise correlation, spatial correlation analysis, and distribution normality tests. These tests are either detailed in this section or alongside test results presented in the Results and Discussion (Section 6).

YIELD–NDVI CALIBRATION

Crop yield data bear inherent spatial and temporal uncertainty. Field survey–based estimates, for example, may have coefficients of variation of up to 1%–5% for major crops like wheat, and between 6% and 25% for specialty crops and small areas[5] (Paliwal and Jain 2020). Crop yields also tend to increase through time due to technological change, improvements in seed biotechnology, and better management practices. Such trends are typically removed (yield data are detrended) using various detrending techniques. However, assumptions behind detrending typically do not consider spatial correlation and temporal correlation between yield and climate, and are typically unrealistic and oversimplified (e.g., trends are assumed to be linear in time and spatially homogeneous). Such assumptions often lack independent, evidence-based support and are linked with many confounding variables that vary substantially across complex agricultural landscapes, such as crop genetics/cultivar type, soil, water management, fertilizer and soil nutrient availability, herbicide use for pest control and disease-related yield loss, and energy and other historical technical improvements in farming equipment. Detrending yield and modeling yield anomalies do not necessarily provide greater model explanatory power and reliability, if the data are detrended under unrealistic assumptions unsupported by data evidence (e.g., linearity, normality, spatial independence).

Crop yield and its variability in time and space are driven by both climate and non-climate variables that include harvesting technology, genetics, weather patterns, soil, and agronomic management practices (e.g., farmer decisions on fertilizer application, tillage and crop hybrid selection, irrigation management, row spacing, planting date and depth) (Kukal and Irmak 2018). To account for non-climate drivers, we specify a link function that relates crop yield to the remotely-sensed index (NDVI_t) as

$$Y_t = g(\mathrm{NDVI}_t),$$

where $g(\cdot)$ is a specified functional relationship between crop yield and NDVI. We utilize a first-order, linear function, $Y_t = \kappa_o \max(\mathrm{NDVI}_t) + \kappa_1$, where κ_o is the slope parameter and κ_1 is the intercept for estimating bias in the yield and NDVI relationship that takes into account non-climate variables. The κ_o and κ_1 coefficients are assumed to be spatially independent (i.e., constant for all municipalities). Analysis performed in the prior field studies shows that NDVI is near-linearly related to photosynthetically active radiation absorbed by a plant canopy, and hence in light-dependent physiological processes, such as photosynthesis, occurring in the upper canopy (Panek and Gozdowski 2021; Liu et al. 2019; Glenn et al. 2008).

In some modeling studies, crop yield Y_t is assumed to depend on NDVI and climate variables independently (separability assumption) (Deines et al. 2021), while others assume NDVI, as a measure of vegetative growth, itself depends on climate variables (Pei et al. 2019; Wang, Rich, and Price 2003) (non-separability assumption) (Yuan et al. 2019). Here, we assume the general case of non-separability of the NDVI–climate relationship. The same crop yield–NDVI calibration relationship is assumed. For a given yield–NDVI calibration, different models are then used to model the relationship of NDVI to climate and to benchmark our MCM approach against traditional multivariate regression approaches.

MULTIVARIATE REGRESSION MODEL WITH ARMA ERRORS

We consider a multivariate regression model with autoregressive and moving average (ARMA) errors (Bianco et al. 2001; Gomez and Maravall 1996) given by

5 Statistics Canada, https://www150.statcan.gc.ca/n1/pub/22-002-x/2011008/technote-notetech2-eng.htm

$$NDVI_t = \sum_{k=1}^{m} \beta_k x_{tk} + \epsilon_t,$$
$$\epsilon_t = \phi_1 \epsilon_{t-1} + \cdots + \phi_p \epsilon_{t-p} + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + a_t,$$

where $\epsilon_t \sim \text{ARMA}(p, q)$; ϕ_i $(i = 1, \ldots, p)$ and θ_j $(j = 1, \ldots, q)$ are autoregressive and moving average coefficients, respectively; p is the autoregressive order, q is the moving-average order (p and q are selected using the lowest value of Akaike information criterion), and a_t is white noise (i.e., $a_t \sim \text{WN}(0, \sigma^2)$). NDVI_t is the response, $\sum_{k=1}^{m} \beta_k x_{tk}$ are m regressors with m regression coefficients β_k. The regressors are: minimum temperature (°C), maximum temperature (°C), total precipitation amount (mm), municipality, and year.

GENERALIZED LEAST SQUARES (GLS) MODEL

We employ the method of generalized least squares (Nelder and Wedderburn 1972) to estimate parameters of a linear regression model when the residuals are correlated with possibly non-constant variance:

$$NDVI = X\beta + \epsilon, \qquad \text{E}[\epsilon] = 0, \ \text{Var}[\epsilon] = \sigma^2 \Sigma.$$

NDVI is a $n \times 1$ vector of observations of the dependent variable (either mean NDVI or maximum NDVI across the growing season); X is a $n \times k$ matrix of predictors (i.e., minimum temperature (°C), maximum temperature (°C), total precipitation amount (mm), municipality, and year); β is a $k \times 1$ vector of the unknown parameters; and ϵ is a $n \times 1$ vector of correlated errors with zero mean and heteroscedastic variance, where σ^2 is a positive constant and Σ is a $n \times n$ matrix. We used the R package *nlme* to fit the GLS model (Pinheiro et al. 2014).

BAYESIAN NETWORK (BN) MODEL

A Bayesian network is a probabilistic graphical model that uses Bayesian inference to compute probabilities (Gao et al. 2019; Huang, Yang, and Jiang 2019; Silander and Myllymaki 2012; Tsamardinos, Brown, and Aliferis 2006; Friedman, Geiger, and Goldszmidt 1997). BN is a directed acyclic graph (DAG) that models conditional dependence where nodes represent the variables and edges represent the conditional dependencies:

$$p(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{N} p(X_i | S_{x_i}),$$

where $S_{x_i}$is the set of edges that are X_i parent node in a graph. We use the *bnlearn* package in R to fit the BN model to data using the Bayes conditional probability strategy for parameter learning (Scutari and Ness 2012). The naive structure (or unstructured) is typically assumed based on all climate variables as predictors (i.e., parent nodes) of crop yield or its proxy NDVI (i.e., child node). The maximum temperature (i.e., $Temp.Max$, °C), minimum temperature ($Temp.Min$, °C), total precipitation (*pcp*, mm), and municipality are parent nodes. In the case of max NDVI, the conditional probability is given by

$$P(Max.NDVI \mid Temp.Max, Temp.Min, Municipality, pcp).$$

A structured network is learned from the observed climate and NDVI data and includes the temporal and spatial effects for each municipality based on NDVI and climate variability as parent nodes for both mean and maximum NDVI, alongside other conditional dependencies between the climate variables.

ARMA-EGARCH MODEL

To model a time series using an EGARCH process, let Z_t be a white noise process with mean 0 and variance 1. ϵ_t denotes error (return residuals) and σ_t is the square root of the conditional variance with non-negative process. The EGARCH(p, q) model is represented as follows:

$$\epsilon_t = \sigma_t Z_t,$$
$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $\alpha_0 > 0$, $\alpha_i > 0$ for $i = 1, \ldots, q$, and β_j for $j = 1, \ldots, p$.

The ARMA model is the model with p autoregressive terms and q moving-average terms. This model contains the AR(p) and MA(q) models,

$$y_t = \mu + \sum_{i=1}^{m} \phi_i y_{t-i} + \sum_{j=1}^{n} \theta_j e_{t-j}.$$

ARCH models assume the variance of the current error term to be a function of the previous errors: often the variance is related to the squares of the previous innovations.

The GARCH processes are generalized ARCH processes in the sense that the squared volatility σ_t^2 is allowed to depend on previous squared volatilities, as well as previous squared values of the process.

In the real world, the return processes may be non-stationary, so we combine the ARMA model and the EGARCH model, where we use ARMA to fit the mean and EGARCH to fit the log variance.

We utilized an ARMA(3,0)-EGARCH(1,1) model (Lama et al. 2015; Chen and Chen 2016; Erhardt and Engler 2018), specified by:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t,$$
$$\epsilon_t = \sigma_t Z_t,$$
$$\ln(\sigma_t^2) = \alpha_o + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where y_t is the response which is the NDVI, and y_{t-i} is the lagged NDVI with lag = 3.

MULTIVARIATE COPULA MODEL (MCM)

A copula is a function that links univariate marginal distributions to the joint multivariate distribution of those variables, with the aim of capturing as much of the entire dependence structure in the multivariate distribution as possible. If a joint distribution is multivariate normal, then a Gaussian copula captures the entire dependence structure. Sklar's theorem provides the theoretical foundation for the application of copulas: where $x_1, x_2, \ldots, x_n$ are random variables with continuous distribution functions $F_1, F_2, \ldots, F_n$ and joint distribution function F, then there exists a unique copula C a distribution function on $[0, 1]^n$ with uniform marginals such that for all $x \in R^n$,

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \quad (1)$$

We define this joint distribution based on mean NDVI or maximum NDVI, together with maximum temperature, minimum temperature, and precipitation. A multivariate copula of dimension 4 was specified. The two main elliptical copulas, namely Gaussian and t-copula, are selected. These copulas do not have a closed form and are restricted to have radial symmetry, whereas Archimedean copulas such as Frank, Joe, Clayton, and Gumbel have a closed form and allow for a number of different dependent structures. Five different copula models for estimating yield-weather dependence are evaluated by Woodard et al. (2011), who report that the Gaussian and t-copulas (as well as kernel, bootstrap, and Clayton copulas) tend to perform well in terms of efficiency, although the Clayton copula performs poorly at low coverage levels.

For the marginal distributions, we select normal distributions for NDVI, maximum temperature, and minimum temperature, and gamma distribution for precipitation, based on χ^2 goodness-of-fit tests. After fitting the multivariate distribution, we select from it a random sample of specific size. The size of the random sample was either 608 for yearly aggregated data for max NGVI, or 13,984 for yearly data for mean NDVI. The dependence structure is captured and specified in the copula function, for example normal and t-copula, by specifying the correlation matrix and the degrees of freedom (for t-copula). We use the *copula* package in R to fit the copula function (Hofert et al. 2014), and the function *fitCopula* in R with the BFGS quasi-Newton variable metric optimization method with 1000 maximum iterations (set the *maxit* parameter in the *optim.control* function) to find the value of the correlation matrix, and appears in the copula function like t-copula has also degrees of freedom to be specified and normal copula. The correlation matrix is estimated using the maximum likelihood method using simulated pseudo-observations as realizations of the random variables to the copula models. The *mvfc* R function is next used to generate the multivariate distribution based on the marginals and copula (refer to Equation 1), and then a random sample is drawn from the multivariate distribution using the *rmvdc* R function.

4. MODEL TRAINING AND VALIDATION

We train the models using data from year 2000 to year 2017.[6] An out-of-sample forecast validation for the year 2018 is performed that compares the MCM performance in forecasting the NDVI to the other benchmark models. This approach aims to predict maximum NDVI as a proxy of yield for 2018 as a validation or test year in which both yield and climate data are available. The approach uses current and historical values of climate-related predictors. To generate future predictions (i.e., forecasts), the approach would require future, near real-time (NRT) weather forecasts (and/or longer-term climate model projections). The prediction mean square error (PMSE) and prediction mean absolute error (PMAE) are used as two popular validation metrics of dispersion and deviation between model-based forecasts and observational data (Willmott and Matsuura 2005). Let y_i be the observed value and $\hat{y}_i$ be the corresponding forecasted value, then for a validation set of size n,

$$PMSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

$$PMAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|.$$

In addition to *forecast* validation for the year 2018, leave-one-out cross-validation (LOOCV) is performed as a *hindcast* validation using all the training data (2000–2017). The LOOCV leaves one year of data out of the training data, i.e., for the 18 years of data in the original sample, 17 years are used to train the model and one year is used for validation. This procedure is iterated so each year is used as the validation year once, then model errors are averaged for all trials, to quantify overall effectiveness.

The Wasserstein distance metric is the distance between two probability distributions on a given metric space (Piccoli and Rossi 2016; Li 2000; Bernton et al. 2019; Bhatia, Jain, and Lim 2019; Olkin and Pukelsheim 1982). Let $(\mathcal{X}, d)$ be a Polish metric space, and let $p \in [1, \infty)$. For any two probability measures μ, υ on $\mathcal{X}$ the Wasserstein distance of order p between μ and υ is defined as (Villani 2009)

$$W_p(\mu, \upsilon) = \Big(\inf_{\pi \in \Pi(\mu,\upsilon)} \int_{\mathcal{X}} d(x,y)^p d\pi(x,y)\Big)^{1/p}.$$

In contrast to point-based measures of forecasting performance, Wasserstein distance allows to account for matching between the true and predicted spatiotemporal distributions rather than only for their location. Model performance improves as PMSE, PMAE, and Wasserstein distance decrease.

SPATIAL DEPENDENCE

Spatial correlation (i.e., the tendency for regions that are close together to have similar NDVI) is assessed based on a *global* Moran I test conducted on maximum NDVI values for each municipality (Moran 1950), because insufficient spatial-referenced yield information is available within each region to enable a local Moran's test within each municipality. The moran.test function in the *spdep* R package is used (Bivand et al. 2019). This test has the null hypothesis of spatial independence (H_0: $I = 0$) versus the alternative hypothesis of spatial autocorrelation (H_a: $I \neq 0$). Values of I usually range from -1 to $+1$, with values significantly below its expected value $-1/(n-1)$ indicating negative spatial autocorrelation, and values significantly above this value indicating positive spatial autocorrelation (n is the sample size).

6 Data and R code are available by request from the authors.

5. MODEL-BASED CROP INSURANCE PRICING

When an index does not match a farmer's actual insured losses, basis risk is present. Either the insurance product is poorly designed, or the index itself does not robustly capture spatial and temporal dynamics of the insured variable, e.g., weather variability or extreme events that affect yield. The advantage of the MCM is that it can capture more representative dynamics of yield change through a growing season in relation to weather variables, and can be used to derive statistical information on claims that is not represented in observed values (e.g., historical data) such as extreme events.

The pricing of weather-based crop-yield insurance contracts based on the MCM for NDVI is demonstrated and benchmarked to the classical burn analysis approach (Gyamerah, Ngare, and Ikpe 2019; López Cabrera, Odening, and Ritter 2013; Taib and Benth 2012). We next describe this benchmarking approach in greater detail.

Let an index-based yield insurance claim size or payoff, X, be given by

$$X(\tau_1, \tau_2) = k \sum_{s=\tau_1}^{\tau_2} \max\left(I(s) - c, 0\right), \tag{2}$$

where I is the index; s is an index over the measurement period, τ_1 to τ_2; c is the coverage level ($0 \leqslant c \leqslant 1$) for the yield insurance contract to be attractive to a farmer; and k is a conversion factor between an index to payoff in units of CAD per acre. We set c as 0.87 in NDVI units, which in yield units is 0.70 or 70% of actual yield (i.e., 0.87 multiplied by the 0.807 yield–NDVI conversion factor). We compute k as the product of a current (2021) crop dollar value for spring wheat of $230 per t of yield ($6.26 lbs/ac) and the yield–NDVI bias calibration factor obtained for the MCM of 0.807 t/index unit to obtain a fixed crop-specific value for k of $185.61 per acre. Average farm size in Manitoba was 1,192 acres (2016).[7]

We adopt a standard approach for pricing a yield contract based on the expected value of claim size, $X(\tau_1, \tau_2)$, that is discounted by an annual interest rate, r, where $r > 0$ and assumed constant (López Cabrera, Odening, and Ritter 2013; Taib and Benth 2012). An insured farmer, who pays a yield insurance premium at time $t \leqslant \tau_1$, receives the amount of $X(\tau_1, \tau_2)$ at the end of the measurement period of the index, i.e., time τ_2, such that if one purchases insurance today and pays the premium then $t = 0$. The times τ_1 and τ_2, for example, might correspond to the start and end of the growing season, and $t = \tau_1$ would mean that a farmer purchases insurance at the start of the growing season, assuming an index is available for pricing over the entire period (τ_1,τ_2). To obtain the present value with a risk-loaded premium level, $\gamma_{\alpha\%}$ for a given probability of exceedance, $\alpha\%$, expressed for contract price, P (CAD per acre), as

$$P(t, \tau_1, \tau_2) = \exp(-r(\tau_2 - t)\left[X(\tau_1, \tau_2)|F_t\right]), \tag{3}$$

where F_t denotes the filtration or all available information (e.g., historical climate or yield records) in an insurance market up to time t, which the insurance company considers in contract pricing. Continuously compounding discount rates are assumed along with an *annual* interest rate of 5% (corresponding to a daily discounting factor of $r = 0.00014$, or weekly factor of $r = 0.00096$).

Burn analysis is used to determine the insurance premium rate benchmark and assumes that the expected loss rate in the future and the past loss distribution are the same. Burn analysis uses historical data to evaluate the value of derivative contracts based on the basic hypotheses that: (1) the time series of temperature is stable; (2) the data of each year are independent and subject to the same distribution (Jewson, Brix, and Ziehmann 2005). In our burn analysis, the empirical distribution of the payoff $X(\tau_1, \tau_2)$ based on max NDVI values associated with each municipality and years 2000–2018 (i.e., $n = 608$) is used to compute contract price. The contract price changes yearly in the burn analysis benchmark, as the historical payoff is measured only once a year, so this approach can only be used at the yearly scale. In contrast, the MCM index can be simulated at the annual or weekly scale. For comparison purposes, an equal number ($n = 608$) of predicted values of *annual* max NDVI from the MCM for each year and municipality were used. Recall that the MCM was trained on observed weekly data ($n = 13{,}984$) across the 32 municipalities and 19 years.

The profit Π^{LG} (i.e., total loss or gain) for a farmer is the difference between total claim sizes $X(\tau_1, \tau_2)$ and the price $P(t, \tau_1, \tau_2)$ of the contract entered into at time t,

$$\Pi^{LG} = X(\tau_1, \tau_2) - P(t, \tau_1, \tau_2). \tag{4}$$

The value-at-risk (VaR) at a statistical significance level $\alpha = 0.05$ (i.e., 5%) of the loss distribution, Π^{LG}, under a fair premium, is used to determine a fair pricing level, γ_{fair} as $P(X \geqslant \gamma_{fair}) = \alpha$, where X is the claim size. This level γ_{fair} is termed the exceedance probability. The risk-loaded premium is assumed to be a certain fraction of this pricing level (i.e., also 5%) termed the premium risk-loading factor, β. The risk-loaded contract price, $P_{risk-loaded}$, above the fair contract price, P_{fair}, based on the risk-loaded premium, $\gamma_{risk-loaded}$, compensates an insurance company for bearing the yield risk, and is given by

$$P_{risk-loaded} = P_{fair} + \gamma_{risk-loaded} = P_{fair} + \beta\gamma_{fair}. \tag{5}$$

6. RESULTS AND DISCUSSION

Correlation analysis of the data shows that *mean* NDVI (weekly scale) is more strongly (and positively) correlated with the climate variables than *maximum* NDVI (annual scale) (Figure 2). These results apply standard Spearman and Kendall pairwise rank correlation analysis that is nonlinear and does not carry any assumptions about the data distribution. Correlation coefficients between 0.10 and 0.29

7 Manitoba Crop Highlights – 2016 Census of Agriculture, https://www.gov.mb.ca/agriculture/markets-and-statistics/crop-statistics/pubs/crop-highlights-census.pdf

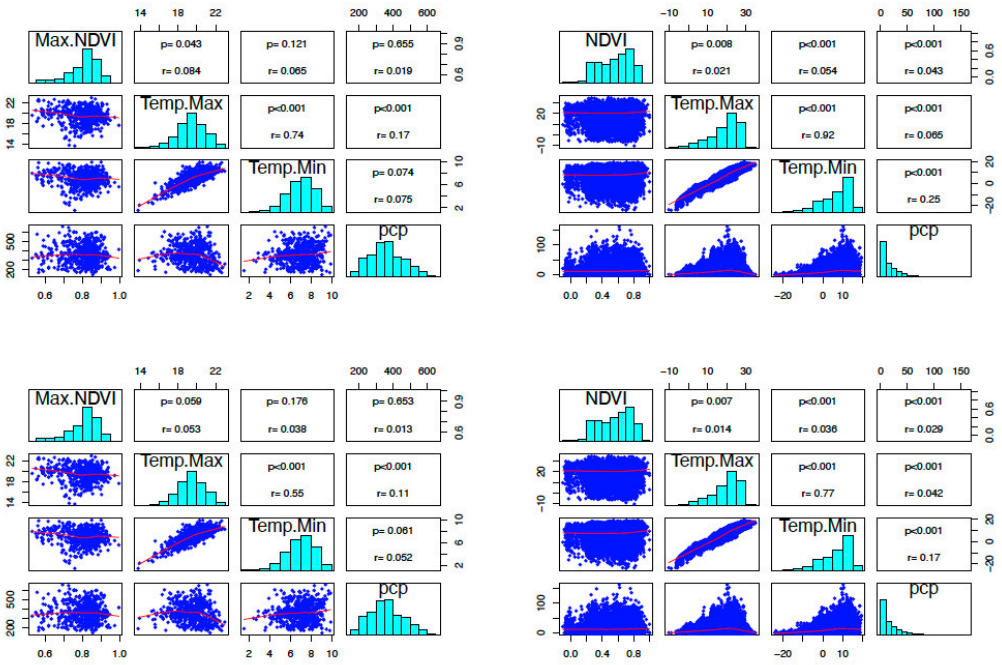


Figure 2. Estimates of pairwise correlations of the observed variables.

Distributions of annually-aggregated maximum NDVI (Left, n = 608) and weekly-aggregated mean NDVI (Right; n = 13;984) (April{September growing season, 2000{2018, 32 municipalities) are shown. (Top) Spearman rank correlation, and (Bottom) Kendall rank correlation.

represent a weak association, coefficients between 0.30 and 0.49 represent a medium association, and coefficients of 0.50 and above represent a strong association or relationship. The association between NDVI (mean or maximum) and any *individual* climate variable alone is weak. This points to the need for a multivariate model that combines variables in an additive way, or a multivariate model with nonlinear assumptions being needed to explain NDVI (i.e., crop vegetation variability) in response to climate variability. Observed distributions for NDVI (unitless), minimum and maximum temperatures (°C), and total precipitation (mm) are also shown accompanying the correlation results. The climate data were determined to be normally-distributed for both minimum ($p = 0.18$) and maximum ($p = 0.10$) temperature, and gamma-distributed for precipitation amount ($p = 0.03$) at 1% significance level.

Testing for spatial correlation in NDVI across municipalities in the study region resulted in a global Moran's I critical z-score ($z = 0.1862$) with a p-value of 0.0116 (for $n = 32$ locations). This z-score is less than the critical value 2.58 at the 99% confidence level, hence the null hypothesis cannot be rejected, meaning that there is no significant evidence of spatial correlation.

Based on the highly statistically significant results of the Durbin-Watson test and the score test for non-constant error variance, we conclude that the multivariate model with ARMA errors and the GLS model, which allow for correlated errors and unequal variances, are justified to be used as modeling benchmarks. (The tests are performed using the R package *lmtest* (Hothorn et al. 2019) and the R package *car* (Fox et al. 2012).)

Variable importance ranking based on the portion of total observed variance explained by each variable, from most to least important, was determined to be: municipality, minimum temperature, maximum temperature, precipitation, and year. In the case of unsupervised, structural learning, a notably different DAG of variable interdependence and spatial dependence (i.e., via an introduced municipality node) resulted, as compared to naive or unstructured learning (Figure 3). The multivariate copula function is able to represent the overall range of inter-annual observed variability reasonably well, but has difficulty in representing higher variability observed in individual years (e.g., 2002, 2004, 2009, 2014), based on 25% and 75% distribution quantiles for both the normal and t-copula function (Figure 4).

The performance of MCM, along with the benchmark models, in predicting mean and maximum NDVI across all municipalities is summarized for the out-of-sample validation test year (i.e., 2018) (Table 1), and for LOOCV (2000–2017) (Table 3). The multivariate regression with ARMA errors (MR with ARMA) model had $p = 2$, $q = 2$ for mean NDVI and maximum NDVI, as this had the lowest Akaike information criterion, an estimator of prediction error, among different orders of p autoregressive terms and q moving-average terms.

For the test year of 2018, the ARMA-EGARCH model yields the lowest PMAE and PMSE, that is, point-based measures of forecasting utility. In turn, the MCM copula-based approach delivers the best performance in terms of Wasserstein distance, that is, the metric that accounts for deviations among the true and forecasted *spatiotemporal distributions* rather than only their means (as in the case of

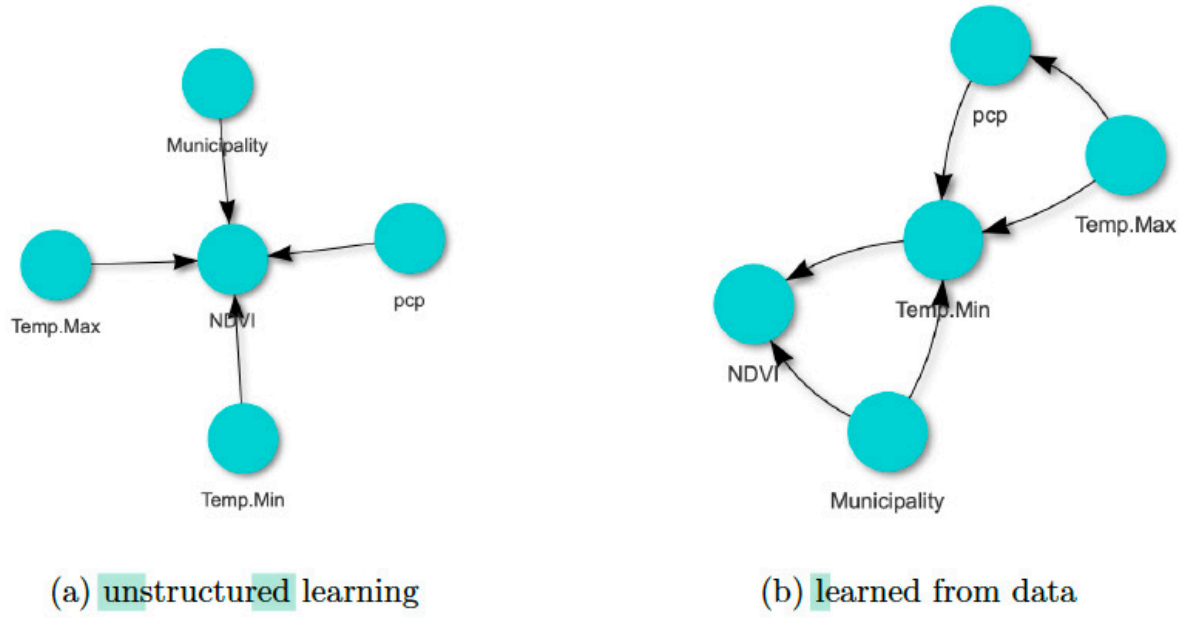


Figure 3. BN model: a) unstructured, and b) learned structure.

Temp.Min and Temp.Max is minimum and maximum daily temperature (°C); pcp is daily precipitation (mm).

PMAE and PMSE). The results are not surprising as MCM particularly targets better forecasting of spatiotemporal distributions.

The LOOCV validation testing shows the BN model predicts well the mean and maximum NDVI, based on the PMSE and PMAE, but the MCM predicts the best, based on the Wasserstein distance.

Overall, the best-fitting MCM prediction error for the 2018 test year and LOOCV testing was smallest for maximum NDVI, compared to mean NDVI. As noted previously by (Turvey and Mclaurin 2012), integrated NDVI and maximum NDVI (i.e., not mean NDVI) have been shown repeatedly to be the best predictors of crop yield. Overall, substantial reductions in prediction error were achieved by the MCM compared to the standard regression models, such as multivariate regression with ARMA and GLS.

Maps of PMSE and PMAE are shown for each model (Figure 5), where darker regions have lower prediction errors. The white regions are municipalities that are not in our study. The BN and MCM and ARMA-EGRACH are able to reduce prediction error across most municipalities, compared to the MR with ARMA errors model and the GLS benchmark models. Stepanov et al. (2020), in a recent study using NDVI and climate data (2010–2018, 9 years) to predict soybean yield also at a municipality or regional scale, obtained RMSE values as high as 16% (see Table 11, page 17). With a longer historical record of data, data for more municipalities, and further modeling work, current prediction error may be further reduced for the MCM and other evaluated models.

The crop yield–NDVI calibration for the BN and MCM best-performing models indicates that bias increases slightly when using the BN model to predict crop yield (spring wheat) but is far more clustered within the 1σ confidence ellipse than the MCM (Figure 6). For the MCM, the estimated coefficients (κ_0 intercept, κ_1 slope) were 0.807348 and -0.006623, respectively, compared to the BN model, which had fitted values of 0.84058 and -0.004879; for ARMA-EGARCH they were 1.28652 and -0.07623. Most of the predicted values for these models lie within the 90% confidence ellipse for the BN model and the 99% (3σ) ellipse for the MCM. Note there is error in the crop yield estimates that is not quantified for each year and municipality and may contribute to bias.

Observed maximum NDVI are more strongly centered between 0.8–0.85 compared with the model-predicted values, which are more broadly distributed between 0.7–1.0 (Figure 7). The associated claim size distribution for the burn analysis is more peaked and clustered around claim sizes between 5–10, but the MCM provides a more uniform distribution of smaller claims and a reduced number of higher claims. Insurance pricing based on the MCM with normal copula shows a wider distribution than the burn analysis benchmark that has most profit values less than 5 $/acre (Figure 8) based on an equal sample size comparison between observed and model predictions ($n = 608$).

The MCM generated a lower exceedance level (risk loaded at 5%) and a slightly higher premium price, but a much lower loss probability and high gain probability for the farmer, as compared with the burn analysis benchmark based on observed data only (Table 3). For the burn analysis, the probability of loss is only slightly higher than the probability of gain. The resultant premium price of the MCM, 12.5524 $/acre, compares to estimates of actual premiums of 8.59–9.76 $/acre that can be above 10 $/acre in some years.

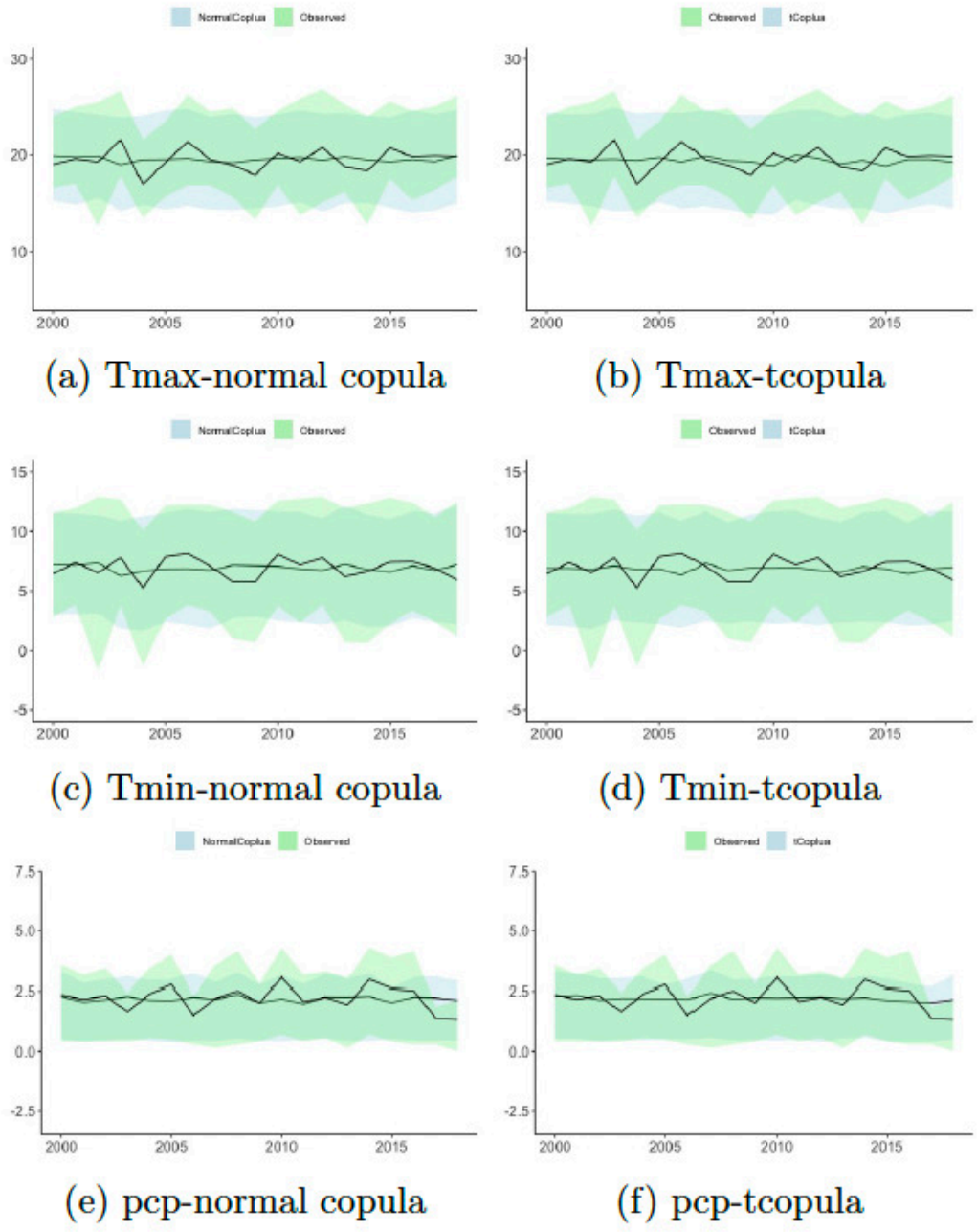


Figure 4. Multivariate copula function (blue) (simulated) compared to observational data (green).

Left insets are for nC, and right insets are for tC. Top to bottom: observed mean of daily maximum temperature (Tmax) (°C), mean of daily minimum temperature (Tmin) (°C), mean of daily precipitation amount (pcp) (mm). The smooth (black) line is the model predicted mean, shown superimposed over the variability in the observed mean (black line) over years. The shaded areas show the 25% and 75% quantiles of observed variability for observed (green shading) versus copula model (blue shading).

7. CONCLUSIONS

We have evaluated the utility and performance of an MCM approach for forecasting crop yield based on satellite NDVI data and incorporating spatial dependence with climate (temperature and precipitation). Our results illustrate the importance of using validation metrics that measure deviation at distribution extremes, such as the Wasserstein distance. Furthermore, we have found that models incorporating spatial dependence (i.e., BN and MCM) outperform standard or classical regression approaches. Our findings demonstrate that the use of MCM to incorporate complex spatial and temporal dependencies in crop yield is closely linked with the dynamics of confounding variables, like temperature and precipitation. Our results demonstrate that MCM has significant benefits and potential as an objective, consistent, and robust statistical methodology for integrating remote-sensing (such as NDVI), climate, and weather data in the design of novel, improved index-based insurance products. In the future, such insurance products may increasingly utilize satellite-derived indices (i.e., coverage, accuracy, reliability).

Table 1. Validation test year (2018): Performance of the models for a) mean NDVI, and b) maximum NDVI, based on the historical data/training dataset (2000{2017) (n = 13;248) and out-of-sample validation for year 2018 (*n* = 736). Prediction mean square error (PMSE) (t^2/ha^2) and prediction mean absolute error (PMAE) (t/ha) averaged across all municipalities for each model are provided: Multivariate regression with ARMA errors, b) Generalized Least Squares (GLS), c) Bayesian network (BN), d) Normal copula (nC), and, tcopula (tC), e) ARMA-EGARCH. Wasserstein metric shows the distance between the observed and predicted NDVI distributions. Best-fitting models for mean NDVI and maximum NDVI models are indicated in boldface.

a) Model prediction error for year 2018 (*mean* NDVI)			
Model	PMSE	PMAE	Wasserstein metric
MR ARMA	0.046	0.182	0.152
GLS	0.043	0.178	0.204
MCM with Ncopula	0.084	0.237	**0.043**
MCM with t-copula	0.083	0.236	**0.041**
BN	0.037	0.157	0.087
ARMA-EGARCH	**0.025**	**0.123**	0.107

b) Model prediction error for year 2018 (*maximum* NDVI)			
Model	PMSE	PMAE	Wasserstein metric
MR ARMA	0.0661	0.2430	0.1522
GLS	0.0799	0.2697	0.2043
MCM with Ncopula	0.0118	0.0828	**0.0226**
MCM with t-copula	0.0103	0.0792	**0.0216**
BN	0.0129	0.0825	0.0334
ARMA-EGARCH	**0.0074**	**0.0257**	0.0862

Table 2. Leave-one-out cross-validation (LOOCV): Performance of the models for a) mean NDVI, and b) maximum NDVI based on LOOCV (2000{2018) (averaged over the 32 municipalities and 19 years) (*n* = 13;984).

a) Model prediction error (*mean* NDVI).			
Model	PMSE	PMAE	Wasserstein metric
MR ARMA	0.078	0.236	0.205
GLS	0.043	0.177	0.201
MCM with Ncopula	0.082	0.234	**0.039**
MCM with t-copula	0.083	0.230	**0.040**
BN	**0.030**	**0.138**	0.088
ARMA-EGARCH	0.0424	0.175	0.204

b) Model prediction error (*maximum* NDVI).			
Model	PMSE	PMAE	Wasserstein metric
MR ARMA	0.0088	0.0659	0.0852
GLS	0.0069	0.0635	0.0838
MCM with Ncopula	0.0147	0.0988	**0.0221**
MCM with t-copula	0.0146	0.0883	**0.0237**
BN	**0.0019**	**0.0326**	0.0367
ARMA-EGARCH	0.0062	0.0608	0.0817

Using remote-sensing indices to track basis risk (yield and yield loss) is particularly attractive because they reduce the need for costly and laborious crop yield field surveys, providing better, more timely coverage over large agricultural areas. The use of such data (e.g., vegetation growth/stress indices) can also help to enhance model predictive performance and provide greater temporal and/or spatial coverage and sampling over a region of interest. Rather than just relying on limited historical records of annual yield estimates for model training, the use of satellite data

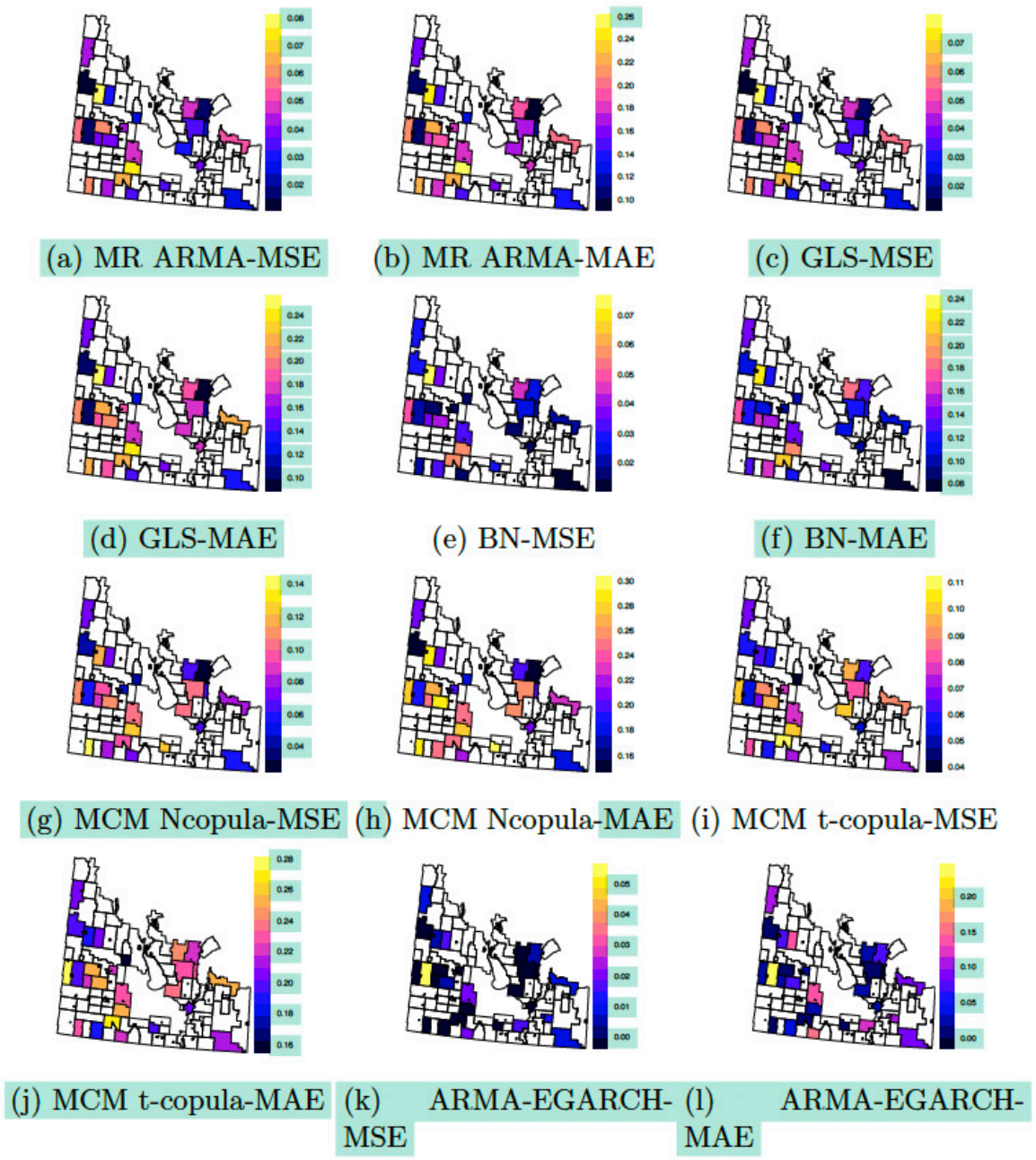


Figure 5. LOOCV: Spatial variability in PMSE (left) and PMAE (right) across the selected regional municipalities in Manitoba, Canada based on: 1) Multivariate regression with ARMA errors, 2) Generalized least squares (GLS), 3) Bayesian network (BN), 4) Normal copula (nC), and 5) t-copula (tC) models, 6) ARMA-EGARCH.

such as NDVI enables developing weekly models for forecasting crop yield and basis risk through the growing season in relation to the timing of extreme events or trends. In tandem with new and improved techniques for data fusion, using data from many satellite image providers, there are anticipated coverage and monitoring capabilities, newly developed remote-sensing indices, and new satellites and constellations of satellites that have been launched with more frequent revisit times than previous missions, having sensors with much higher spatial resolution (e.g., Europe's Sentinel-2 and Canada's Radarsat Constellation Mission or RCM). In addition, satellite data could be supplemented by finer-scale imagery collected by drones (i.e., low-cost remotely piloted aircraft systems or RPAS)—a rapidly emerging technology that can also provide high-resolution and timely geospatial data to support model-based agricultural risk decision making.

This study illustrates available R libraries for implementing, fitting, and benchmarking MCMs, for the scientific community and other practitioners who are interested in employing the models. There is increasing interest in the MCM method, as evidenced by a new toolbox called the Multivariate Copula Analysis Toolbox (MvCAT), recently developed for MCM use and application by the statistics

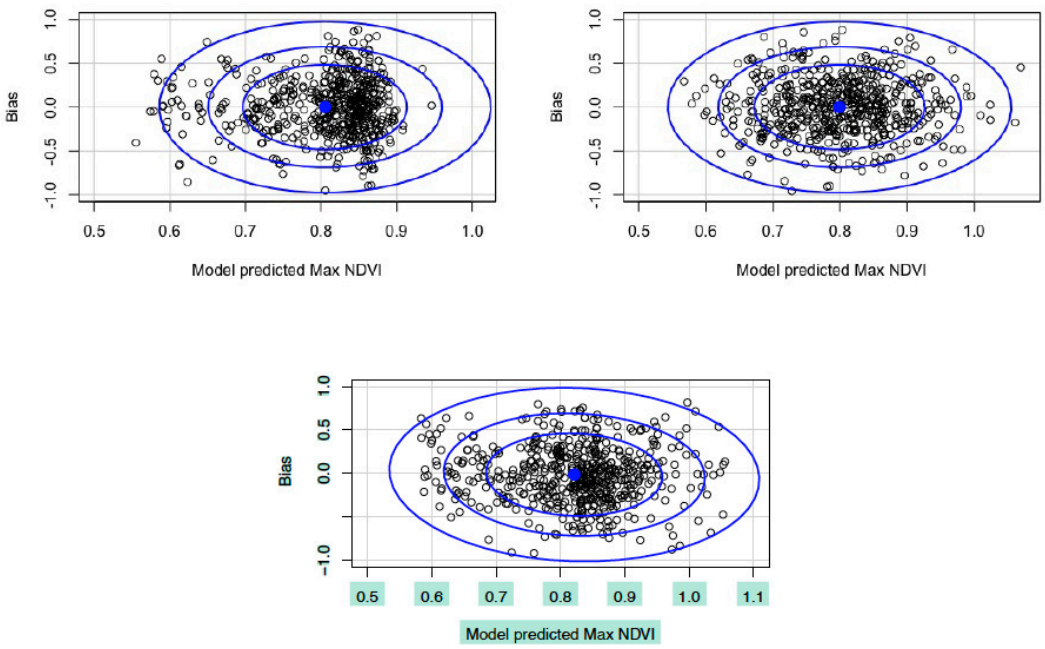


Figure 6. Estimated bias between observed crop yield (t/acre) (spring wheat) and model predictions of maximum NDVI.

Results are shown for the two best- tting models, namely: BN (top left) and MCM with Normal copula (top right) (2000{2018), and ARMA-EGARCH (bottom) $n = 608$. Con dence ellipses containing 65% (1 σ), 90%, and 99% (3 σ) of model predictions are shown.

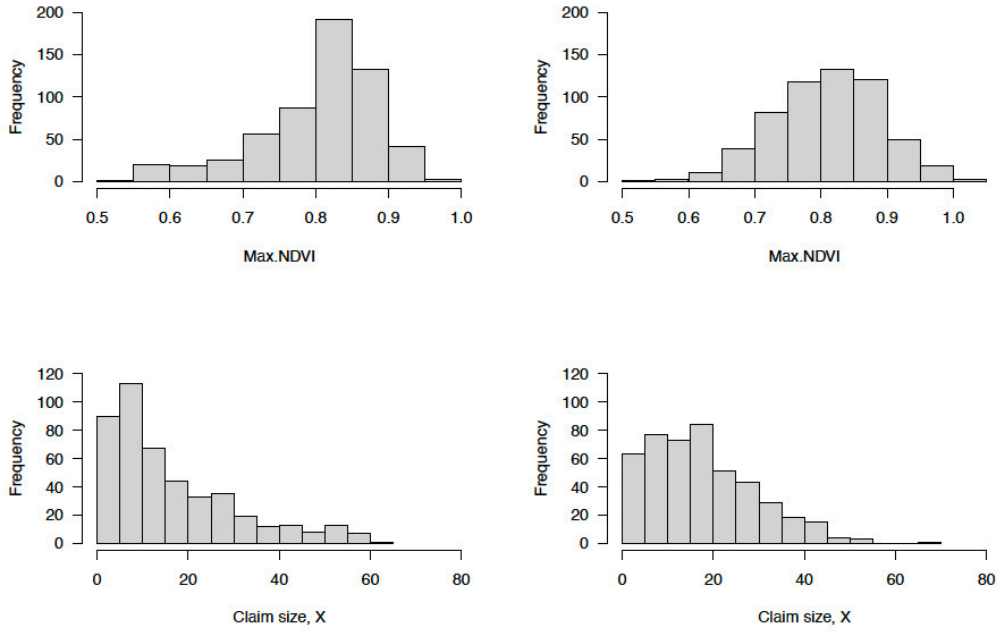


Figure 7. Histogram of maximum NDVI from observed data (top left) and best-fitting MCM (top right) and resulting claim size distribution from the empirical distribution from the burn analysis (bottom left) and MCM predictions (bottom right).

and actuarial science community (Sadegh, Ragno, and AghaKouchak 2017). This toolbox employs a Bayesian framework with a Gaussian likelihood function for residuals that infers copula parameters for a wide range of copula families. The findings reported here on MCM performance and utility to multivariate problems support much broader use of MCMs in risk modeling.

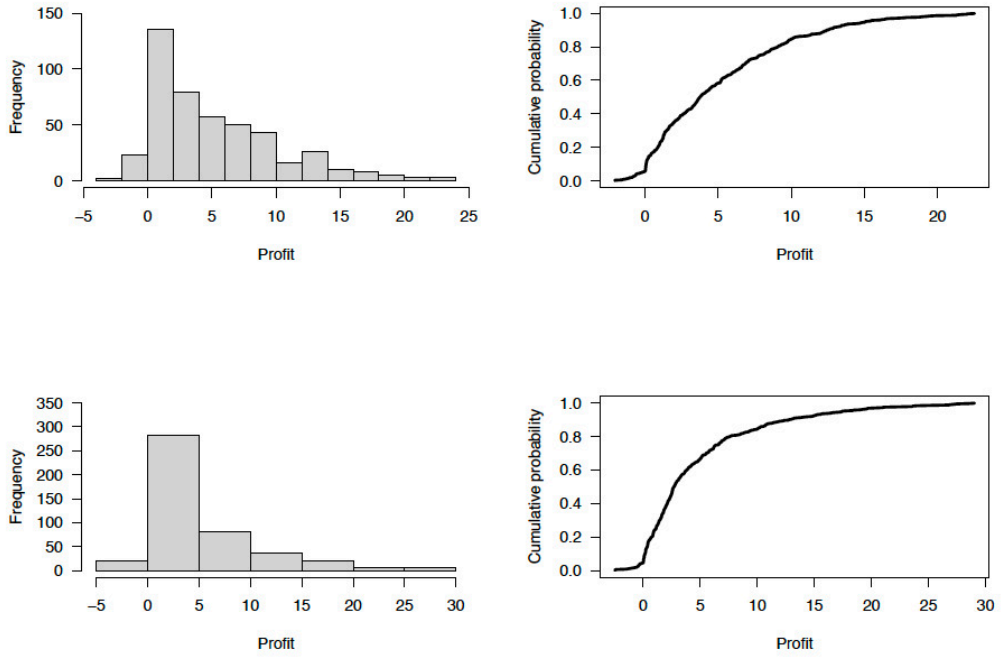


Figure 8. Insurance pricing calculation: histogram of profit and associated cumulative probability distributions for the best-fitting MCM (top) and burn analysis (bottom) for annual max NDVI for the 32 municipalities, 2000-2018 (*n* = 608).

Table 3. Exceedance level (γ), average premium price ($/acre), and probability of loss and gain for 5% of risk loading for the MCM and burn analysis benchmark.

Pricing method	Exceedance level, γ ($\alpha = 5\%$)	Premium, $P(\tau_1, \tau_2)$ ($/ac)	Probability of loss,$p(X < P(\tau_1, \tau_2))$	Probability of gain,$p(X > P(\tau_1, \tau_2))$
Burn analysis	49.9465	11.6834	0.5032	0.4967
MCM	41.8361	12.5524	0.3600	0.6399

Note: *P* denotes price, and *p* denotes probability.

In future research, we aim to expand the current MCM to one that has a widely-applicable and flexible network structure (i.e., as a Copula Bayesian Network, CBN). This form could then be compared to Deep Bayesian Networks (DBNs) and recent results obtained for these models (Tanguy et al. 2015). Future research investigations will seek to incorporate data on other agricultural crops and explore the inclusion of multiple satellite remote-sensing indices and indices of weather extremes. Finally, inspired by the recent applications of topological data analysis in actuarial sciences (Jiang et al. 2022; Yuvaraj et al. 2021), we will investigate the utility and limitations of integrating the copula methods with topological summaries of crop yields and atmospheric data.

ACKNOWLEDGMENTS

This work is supported, in part, by the Canadian Agricultural Partnership (CAP) of Agriculture and Agri-Food Canada (AAFC) and the Casualty Actuarial Society (CAS). The authors thank Dr. T. A. Porcelli for providing helpful comments and editing of the final manuscript, and Molina Tek (GIS Analyst, AAFC) for assistance in the satellite data manipulation and processing.

Submitted: May 04, 2021 EDT. Accepted: February 16, 2022 EDT. Published: May 03, 2023 EDT.

REFERENCES

Abdi, Mukhtar Jibril, Nurfarhana Raffar, Zed Zulkafli, Khairudin Nurulhuda, Balqis Mohamed Rehan, Farrah Melissa Muharam, Nor Ain Khosim, and Fredolin Tangang. 2021. "Index-Based Insurance and Hydroclimatic Risk Management in Agriculture: A Systematic Review of Index Selection and Yield-Index Modelling Methods." *International Journal of Disaster Risk Reduction* 67:102653.

Basnyat, P., B. McConkey, G. P. Lafond, A. Moulin, and Y. Pelcat. 2004. "Optimal Time for Remote Sensing to Relate to Crop Grain Yield on the Canadian Prairies." *Canadian Journal of Plant Science* 84 (1): 97–103.

Bernton, Espen, Pierre E Jacob, Mathieu Gerber, and Christian P Robert. 2019. "On Parameter Estimation with the Wasserstein Distance." *Information and Inference: A Journal of the IMA* 8 (4): 657–76. https://doi.org/10.1093/imaiai/iaz003.

Bhatia, Rajendra, Tanvi Jain, and Yongdo Lim. 2019. "On the Bures–Wasserstein Distance between Positive Definite Matrices." *Expositiones Mathematicae* 37 (2): 165–91. https://doi.org/10.1016/j.exmath.2018.01.002.

Bianco, A. M., M. García Ben, E. J. Martínez, and V. J. Yohai. 2001. "Outlier Detection in Regression Models with Arima Errors Using Robust Estimates." *Journal of Forecasting* 20 (8): 565–79. https://doi.org/10.1002/for.768.

Bivand, R, M Altman, L Anselin, R Assunção, O Berke, A Bernat, and G Blanchet. 2019. "Spdep: Spatial Dependence: Weighting Schemes, Statistics and Models. R Package Version 1.1-2."

Bush, E, and D S Lemmen. 2019. "Canada's Changing Climate Report." Natural Resources Canada/CMSS/Information Management. https://doi.org/10.4095/314614.

Campbell, C. A., R. P. Zentner, S. Gameda, B. Blomert, and D. D. Wall. 2002. "Production of Annual Crops on the Canadian Prairies: Trends during 1976–1998." *Canadian Journal of Soil Science* 82 (1): 45–57. https://doi.org/10.4141/s01-046.

Carew, R., T. Meng, W.J. Florkowski, R. Smith, and D. Blair. 2018. "Climate Change Impacts on Hard Red Spring Wheat Yield and Production Risk: Evidence from Manitoba, Canada." Edited by Christian Willenborg. *Canadian Journal of Plant Science* 98 (3): 782–95. https://doi.org/10.1139/cjps-2017-0135.

Chen, Kuan-Ju, and Kuan-Heng Chen. 2016. "Analysis of Energy and Agricultural Commodity Markets with the Policy Mandated: A Vine Copula-Based ARMA-EGARCH Model."

Chipanshi, Aston, Yinsuo Zhang, Louis Kouadio, Nathaniel Newlands, Andrew Davidson, Harvey Hill, Richard Warren, et al. 2015. "Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) Model for in-Season Prediction of Crop Yield across the Canadian Agricultural Landscape." *Agricultural and Forest Meteorology* 206 (June):137–50. https://doi.org/10.1016/j.agrformet.2015.03.007.

Deines, Jillian M., Rinkal Patel, Sang-Zi Liang, Walter Dado, and David B. Lobell. 2021. "A Million Kernels of Truth: Insights into Scalable Satellite Maize Yield Mapping and Yield Gap Analysis from an Extensive Ground Dataset in the US Corn Belt." *Remote Sensing of Environment* 253 (February):112174. https://doi.org/10.1016/j.rse.2020.112174.

Easterling, D. R., J. L. Evans, P. Ya Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje. 2000. "Observed Variability and Trends in Extreme Climate Events: A Brief Review." *Bulletin of the American Meteorological Society* 81 (3): 417–25. https://doi.org/10.1175/1520-0477(2000)081.

Erhardt, Robert, and David Engler. 2018. "An Extension of Spatial Dependence Models for Estimating Short-Term Temperature Portfolio Risk." *North American Actuarial Journal* 22 (3): 473–90. https://doi.org/10.1080/10920277.2018.1444496.

Fox, John, Sanford Weisberg, Daniel Adler, Douglas Bates, Gabriel Baud-Bovy, Steve Ellison, David Firth, et al. 2012. "Package 'Car.'" *Vienna: R Foundation for Statistical Computing*.

Friedman, Nir, Dan Geiger, and Moises Goldszmidt. 1997. "Bayesian Network Classifiers." *Machine Learning* 29 (2–3): 131–63.

Gao, Xiao-guang, Zhi-gao Guo, Hao Ren, Yu Yang, Da-qing Chen, and Chu-chao He. 2019. "Learning Bayesian Network Parameters via Minimax Algorithm." *International Journal of Approximate Reasoning* 108 (May):62–75. https://doi.org/10.1016/j.ijar.2019.03.001.

Glenn, Edward, Alfredo Huete, Pamela Nagler, and Stephen Nelson. 2008. "Relationship Between Remotely-Sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape." *Sensors* 8 (4): 2136–60. https://doi.org/10.3390/s8042136.

Gobin, A., A. M. Tarquis, and N. R. Dalezios. 2013. "Weather-Related Hazards and Risks in Agriculture." *Natural Hazards and Earth System Sciences* 13 (10): 2599–2603. https://doi.org/10.5194/nhess-13-2599-2013.

Gomez, Victor, and Agustin Maravall. 1996. "Programs TRAMO (Time Series Regression with ARIMA Noise, Missing Observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series). Instructions for the User." *Documento de Trabajo* 9628.

Guo, X. 2003. "Relationships between NDVI and Climatological Variability in the Prairie Ecozone of Canada." *Prairie Perspective, Geographical Essays* 6:32–46. https://pcag.uwinnipeg.ca/Prairie-Perspectives/PP-Vol06/Guo.pdf.

Gyamerah, Samuel Asante, Philip Ngare, and Dennis Ikpe. 2019. "Hedging Crop Yields Against Weather Uncertainties—A Weather Derivative Perspective." *Mathematical and Computational Applications* 24 (3): 71. https://doi.org/10.3390/mca24030071.

He, Yuhong, Xulin Guo, Paul Dixon, and John F. Wilmshurst. 2012. "NDVI Variation and Its Relation to Climate in Canadian Ecozones." *The Canadian Geographer* 56 (4): 492–507. https://doi.org/10.1111/j.1541-0064.2012.00441.x.

Hofert, Marius, Ivan Kojadinovic, Martin Maechler, and Jun Yan. 2014. "Copula: Multivariate Dependence with Copulas." *R Package Version 0.999-9*, C225. http://CRANR-project.org/package=copula.

Hothorn, Torsten, Achim Zeileis, Richard W Farebrother, Clint Cummins, Giovanni Millo, David Mitchell, and Maintainer Achim Zeileis. 2019. "Package 'Lmtest.'"

Huang, Zhiming, Lin Yang, and Wen Jiang. 2019. "Uncertainty Measurement with Belief Entropy on the Interference Effect in the Quantum-like Bayesian Networks." *Applied Mathematics and Computation* 347 (April):417–28. https://doi.org/10.1016/j.amc.2018.11.036.

Jensen, Nathaniel D., Christopher B. Barrett, and Andrew G. Mude. 2016. "Index Insurance Quality and Basis Risk: Evidence from Northern Kenya." *American Journal of Agricultural Economics* 98 (5): 1450–69. https://doi.org/10.1093/ajae/aaw046.

Jewson, Stephen, Anders Brix, and Christine Ziehmann. 2005. *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*. Cambridge University Press. https://doi.org/10.1017/cbo9780511493348.

Jiang, Tian, Meichen Huang, Ignacio Segovia-Dominguez, Nathaniel Newlands, and Yulia R. Gel. 2022. "Learning Space-Time Crop Yield Patterns with Zigzag Persistence-Based LSTM: Toward More Reliable Digital Agriculture Insurance." *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (11): 12538–44. https://doi.org/10.1609/aaai.v36i11.21524.

Kouadio, Louis, Nathaniel Newlands, Andrew Davidson, Yinsuo Zhang, and Aston Chipanshi. 2014. "Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale." *Remote Sensing* 6 (10): 10193–214. https://doi.org/10.3390/rs61010193.

Kukal, Meetpal S., and Suat Irmak. 2018. "Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production." *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-21848-2.

Lama, Achal, Girish K. Jha, Ranjit K. Paul, and Bishal Gurung. 2015. "Modelling and Forecasting of Price Volatility: An Application of GARCH and EGARCH Models." *Agricultural Economics Research Review* 28 (1): 73. https://doi.org/10.5958/0974-0279.2015.00005.1.

Li, David X. 2000. "On Default Correlation: A Copula Function Approach." *The Journal of Fixed Income* 9 (4): 43–54. https://doi.org/10.3905/jfi.2000.319253.

Liu, Jiangui, Ted Huffman, Budong Qian, Jiali Shang, Qingmou Li, Taifeng Dong, Andrew Davidson, and Qi Jing. 2020. "Crop Yield Estimation in the Canadian Prairies Using Terra/MODIS-Derived Crop Metrics." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:2685–97. https://doi.org/10.1109/jstars.2020.2984158.

Liu, Jiangui, Jiali Shang, Budong Qian, Ted Huffman, Yinsuo Zhang, Taifeng Dong, Qi Jing, and Tim Martin. 2019. "Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada." *Remote Sensing* 11 (20): 2419. https://doi.org/10.3390/rs11202419.

Lobell, David B. 2013. "The Use of Satellite Data for Crop Yield Gap Analysis." *Field Crops Research* 143 (March):56–64. https://doi.org/10.1016/j.fcr.2012.08.008.

López Cabrera, Brenda, Martin Odening, and Matthias Ritter. 2013. "Pricing Rainfall Futures at the CME." *Journal of Banking & Finance* 37 (11): 4286–98. https://doi.org/10.1016/j.jbankfin.2013.07.042.

Lyubchich, Vyacheslav, Nathaniel K. Newlands, Azar Ghahari, Tahir Mahdi, and Yulia R. Gel. 2019. "Insurance Risk Assessment in the Face of Climate Change: Integrating Data Science and Statistics." *WIREs Computational Statistics* 11 (4): e1462. https://doi.org/10.1002/wics.1462.

Martens, Joanne R. Thiessen, Martin H. Entz, and Mark D. Wonneck. 2015. "Review: Redesigning Canadian Prairie Cropping Systems for Profitability, Sustainability, and Resilience." *Canadian Journal of Plant Science* 95 (6): 1049–72. https://doi.org/10.4141/cjps-2014-173.

Moran, P. A. P. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1/2): 17. https://doi.org/10.2307/2332142.

Myers, Samuel S., Matthew R. Smith, Sarah Guth, Christopher D. Golden, Bapu Vaitla, Nathaniel D. Mueller, Alan D. Dangour, and Peter Huybers. 2017. "Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition." *Annual Review of Public Health* 38 (1): 259–77. https://doi.org/10.1146/annurev-publhealth-031816-044356.

NAS. 2016. *Committee on Extreme Weather Events and Climate Change Attribution of the National Academies of Sciences, Engineering, and Medicine. Attribution of Extreme Weather Events in the Context of Climate Change*. Washington, DC: National Academies Press.

Nelder, John Ashworth, and R. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society: Series A (General)* 135 (3): 370–84.

Newlands, Nathaniel K. 2016. *Future Sustainable Ecosystems: Complexity, Risk, and Uncertainty*. Boca Raton, FL: Taylor and Francis, CRC Press.

Newlands, Nathaniel K., David S. Zamar, Louis A. Kouadio, Yinsuo Zhang, Aston Chipanshi, Andries Potgieter, Souleymane Toure, and Harvey S. J. Hill. 2014. "An Integrated, Probabilistic Model for Improved Seasonal Forecasting of Agricultural Crop Yield under Environmental Uncertainty." *Frontiers in Environmental Science* 2 (June):17. https://doi.org/10.3389/fenvs.2014.00017.

Nguyen-Huy, Thong, Ravinesh C Deo, Shahbaz Mushtaq, Duc-Anh An-Vo, and Shahjahan Khan. 2018. "Modeling the Joint Influence of Multiple Synoptic-Scale, Climate Mode Indices on Australian Wheat Yield Using a Vine Copula-Based Approach." *European Journal of Agronomy* 98 (August):65–81. https://doi.org/10.1016/j.eja.2018.05.006.

Norton, Michael T., Stephen Boucher, and Leslie Verteramo Chiu. 2015. "Geostatistics, Basis Risk, and Index Insurance." Agricultural and Applied Economics Association (AAEA).

Norton, Michael T., Calum Turvey, and Daniel Osgood. 2012. "Quantifying Spatial Basis Risk for Weather Index Insurance." *The Journal of Risk Finance* 14 (1): 20–34. https://doi.org/10.1108/15265941311288086.

Olkin, I., and F. Pukelsheim. 1982. "The Distance between Two Random Vectors with given Dispersion Matrices." *Linear Algebra and Its Applications* 48 (December):257–63. https://doi.org/10.1016/0024-3795(82)90112-4.

Paliwal, Ambica, and Meha Jain. 2020. "The Accuracy of Self-Reported Crop Yield Estimates and Their Ability to Train Remote Sensing Algorithms." *Frontiers in Sustainable Food Systems* 4 (25): 1–10. https://doi.org/10.3389/fsufs.2020.00025.

Panek, Ewa, and Dariusz Gozdowski. 2021. "Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries." *Agronomy* 11 (2): 340. https://doi.org/10.3390/agronomy11020340.

Pei, Zhifang, Shibo Fang, Wunian Yang, Lei Wang, Mingyan Wu, Qifei Zhang, Wei Han, and Dao Nguyen Khoi. 2019. "The Relationship between NDVI and Climate Factors at Different Monthly Time Scales: A Case Study of Grasslands in Inner Mongolia, China (1982–2015)." *Sustainability* 11 (24): 7243. https://doi.org/10.3390/su11247243.

Petersen, Lillian. 2018. "Real-Time Prediction of Crop Yields from MODIS Relative Vegetation Health: A Continent-Wide Analysis of Africa." *Remote Sensing* 10 (11): 1726. https://doi.org/10.3390/rs10111726.

Piccoli, Benedetto, and Francesco Rossi. 2016. "On Properties of the Generalized Wasserstein Distance." *Archive for Rational Mechanics and Analysis* 222 (3): 1339–65. https://doi.org/10.1007/s00205-016-1026-7.

Pinheiro, José, Douglas Bates, Sikat DebRoy, and Deepayan Sarkar. 2014. "R Core Team (2014) Nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-117." *Available At*. http://CRANR-744project.org/package=nlme.

Renard, Delphine, and David Tilman. 2019. "National Food Production Stabilized by Crop Diversity." *Nature* 571 (7764): 257–60. https://doi.org/10.1038/s41586-019-1316-y.

Sadegh, Mojtaba, Elisa Ragno, and Amir AghaKouchak. 2017. "Multivariate Copula Analysis Toolbox (MvCAT): Describing Dependence and Underlying Uncertainty Using a Bayesian Framework." *Water Resources Research* 53 (6): 5166–83. https://doi.org/10.1002/2016wr020242.

Sadri, Sara, and Donald H. Burn. 2014. "Copula-Based Pooled Frequency Analysis of Droughts in the Canadian Prairies." *Journal of Hydrologic Engineering* 19 (2): 277–89. https://doi.org/10.1061/(asce)he.1943-5584.0000603.

Salgueiro, Andrea Martínez. 2019. "Weather Index-Based Insurance as a Meteorological Risk Management Alternative in Viticulture." *Wine Economics and Policy* 8 (2): 114–26.

Scutari, Marco, and Robert Ness. 2012. "Bnlearn: Bayesian Network Structure Learning, Parameter Learning and Inference." *R Package Version* 3.

Silander, Tomi, and Petri Myllymaki. 2012. "A Simple Approach for Finding the Globally Optimal Bayesian Network Structure." *arXiv Preprint arXiv:1206.6875*.

Sillmann, Jana, Viatcheslav V Kharin, FW Zwiers, Xuebin Zhang, and D Bronaugh. 2013. "Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 2. Future Climate Projections." *Journal of Geophysical Research: Atmospheres* 118 (6): 2473–93.

Staudt, A. 2010. "Tail Risk, Systemic Risk and Copulas." *Casualty Actuarial Society E-Forum* 2.

Stepanov, Alexey, Konstantin Dubrovin, Aleksei Sorokin, and Tatiana Aseeva. 2020. "Predicting Soybean Yield at the Regional Scale Using Remote Sensing and Climatic Data." *Remote Sensing* 12 (12): 1936. https://doi.org/10.3390/rs12121936.

Tack, Jesse B., and David Ubilava. 2015. "Climate and Agricultural Risk: Measuring the Effect of ENSO on U.S. Crop Insurance." *Agricultural Economics* 46 (2): 245–57. https://doi.org/10.1111/agec.12154.

Taib, Che Mohd Imran Che, and Fred Espen Benth. 2012. "Pricing of Temperature Index Insurance." *Review of Development Finance* 2 (1): 22–31. https://doi.org/10.1016/j.rdf.2012.01.004.

Tanguy, M., J. Hannaford, L. Barker, C. Svensson, F. Kral, and M. Fry. 2015. "Gridded Standardized Precipitation Index (SPI) Using Gamma Distribution with Standard Period 1961-2010 for Great Britain [SPIgamma61-10]." NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/94C9EAA3-A178-4DE4-8905-DBFAB03B69A0.

Thornton, PE, MM Thornton, BW Mayer, Y Wei, RS Devarakonda, RS Vose, and RB Cook. 2017. *Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 3 ORNL DAAC Oak Ridge Tenn.* USA.

Tsamardinos, Ioannis, Laura E. Brown, and Constantin F. Aliferis. 2006. "The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm." *Machine Learning* 65 (1): 31–78. https://doi.org/10.1007/s10994-006-6889-7.

Turvey, Calum G., and Megan K. Mclaurin. 2012. "Applicability of the Normalized Difference Vegetation Index (NDVI) in Index-Based Crop Insurance Design." *Weather, Climate, and Society* 4 (4): 271–84. https://doi.org/10.1175/wcas-d-11-00059.1.

Villani, Cédric. 2009. "The Wasserstein Distances." *Optimal Transport*, 93–111. https://doi.org/10.1007/978-3-540-71050-9_6.

Vyas, Shalika, Tobias Dalhaus, Martin Kropff, Pramod Aggarwal, and Miranda P M Meuwissen. 2021. "Mapping Global Research on Agricultural Insurance." *Environmental Research Letters* 16 (10): 103003. https://doi.org/10.1088/1748-9326/ac263d.

Wall, Lenny, Denis Larocque, and Pierre-Majorique Léger. 2008. "The Early Explanatory Power of NDVI in Crop Yield Modelling." *International Journal of Remote Sensing* 29 (8): 2211–25. https://doi.org/10.1080/01431160701395252.

Wang, J., P. M. Rich, and K. P. Price. 2003. "Temporal Responses of NDVI to Precipitation and Temperature in the Central Great Plains, USA." *International Journal of Remote Sensing* 24 (11): 2345–64. https://doi.org/10.1080/01431160210154812.

Willmott, CJ, and K Matsuura. 2005. "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30 (1): 79–82. https://doi.org/10.3354/cr030079.

Woodard, Joshua D., Nicholas D. Paulson, Dmitry Vedenov, and Gabriel J. Power. 2011. "Impact of Copula Choice on the Modeling of Crop Yield Basis Risk." *Agricultural Economics* 42 (November):101–12. https://doi.org/10.1111/j.1574-0862.2011.00555.x.

Yu, Jisang, Monte Vandeveer, Jerry D Volesky, and Keith Harmoney. 2019. "Estimating the Basis Risk of Rainfall Index Insurance for Pasture, Rangeland, and Forage." *Journal of Agricultural and Resource Economics* 44 (1835-2019–069): 179–93.

Yuan, Wei, Shuang-Ye Wu, Shugui Hou, Zhiwei Xu, and Huayu Lu. 2019. "Normalized Difference Vegetation Index-Based Assessment of Climate Change Impact on Vegetation Growth in the Humid-Arid Transition Zone in Northern China during 1982–2013." *International Journal of Climatology* 39 (15): 5583–98. https://doi.org/10.1002/joc.6172.

Yuvaraj, Monisha, Asim K. Dey, Vyacheslav Lyubchich, Yulia R. Gel, and H. Vincent Poor. 2021. "Topological Clustering of Multilayer Networks." *Proceedings of the National Academy of Sciences* 118 (21): e2019994118. https://doi.org/10.1073/pnas.2019994118.

Zhang, Xuebin, Lisa Alexander, Gabriele C. Hegerl, Philip Jones, Albert Klein Tank, Thomas C. Peterson, Blair Trewin, and Francis W. Zwiers. 2011. "Indices for Monitoring Changes in Extremes Based on Daily Temperature and Precipitation Data." *WIREs Climate Change* 2 (6): 851–70. https://doi.org/10.1002/wcc.147.

Zhao, Yan, Andries B Potgieter, Miao Zhang, Bingfang Wu, and Graeme L Hammer. 2020. "Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling." *Remote Sensing* 12 (6): 1024. https://doi.org/10.3390/rs12061024.