Crop-yielding Prediction using Neural Network for Stochastic Differential Equation Parameters Estimation

Dr. Ayman Elgharabawy
Biological Data Science Institute (BDSI)
The Australian National University
Canberra ACT 2600 Australia
ayman.gh@anu.edu.au
+61 466054162

Vito Rihaldijiran
Biological Data Science Institute (BDSI)
The Australian National University
Canberra ACT 2600 Australia
vito.rihaldijiran@anu.edu.au
+61 493 944 883

Abstract

Crop-yielding prediction is a complex, multivariate, random, dynamic, and computationally intensive problem. This project employs a collection of stochastic differential equations (SDEs) in Itô form, using an artificial neural network (ANN) to predict the parameters of the SDE for the chaotic time series of wheat yield on a daily, monthly, and yearly basis. Our goal is to develop a novel and unique stochastic wheat model that provides the most accurate predictions based on SDE parameter estimation. The project will focus on defining new SDE parameters that can be accurately estimated using a neural network under specific noise level regimes to predict crop yield. The dataset used for this project is sourced from the European Commission's MARS Crop Yield Forecasting System (MCYFS), which includes data on spring barley, potato, grain maize, and sunflower. The historical data is available in both daily and yearly scales.

Research Objective

The objective of this research is to develop a high-performance wheat-yield prediction model for large-scale datasets using a stochastic approach combined with machine learning for parameter estimation. This method aims to significantly reduce computational time compared to traditional and deep neural network models, which often face overfitting issues due to unbalanced data.

Research Output

In conducting this study, we will accomplish the following specific deliverables:

- Develop a novel stochastic differential equation (SDE) model tailored for crop yield prediction. This model will accurately forecast crop yields by incorporating innovative drift and diffusion functions.

- Identify and estimate a specific set of parameters within the SDE model that are crucial for accurate yield predictions. These parameters will be determined using advanced machine learning techniques.

- Design and implement new drift functions that represent the deterministic trends in crop growth and yield over time.

- Develop diffusion functions that effectively capture the stochastic nature and inherent variability in crop yield data due to environmental factors and other uncertainties.

- Validate the performance of the new SDE model through rigorous testing on extensive historical crop yield data, ensuring its robustness and reliability in various scenarios.

- Demonstrate the computational efficiency of the proposed model, showcasing its ability to handle large-scale datasets while mitigating common issues like overfitting, which are prevalent in traditional and deep neural network models.

Introduction

Crop yield prediction is a highly complex and intricate problem due to its nature as a multivariate random dynamic system. This complexity stems from the multitude of factors influencing crop yields, which encompass weather conditions, soil types, and variations in both genotype and phenotype. These variations alone introduce nearly thousands of individual factors that need to be considered.

The factors influencing crop yield are numerous and varied. Weather-related factors include temperature, precipitation, humidity, and sunlight, all of which can fluctuate widely and have significant impacts on crop growth. Soil type adds another layer of complexity, as different soils have varying capacities for water retention, nutrient availability, and root support. Genotype and phenotype variations introduce additional complexity, as different crop varieties and their specific traits interact differently with environmental conditions.

When scaled to large datasets, the computational demands of analyzing and predicting crop yields become immense. Traditional and deep machine learning models face significant challenges in handling the high dimensionality of data, which includes thousands of genotype, phenotype, and environmental variable features. Processing such large volumes of data is computationally expensive, requiring substantial resources and time.

Moreover, these factors are not only numerous but also interdependent and often hidden, adding to the difficulty of making accurate predictions. This complexity is akin to predicting stock exchange rates, where numerous invariant and hidden factors, such as market trends, economic indicators, and investor behaviors, must be accounted for to make reliable predictions. In both cases, the challenge lies in accurately modeling a system influenced by a wide array of dynamic and interrelated variables.

In summary, crop yield prediction requires advanced computational approaches to manage and analyze the vast and complex datasets involved. Innovative methods that can efficiently handle the high dimensionality and dynamic nature of the data are essential to improving prediction accuracy and making effective use of the available information.

Research Method

Parameter Estimation

Given a one-dimensional time-homogeneous stochastic differential equation (SDE):

$$dX = \mu(X;\theta)dt + g(X;\theta)dW \tag{1}$$

the task is to estimate the parameter θ from a sample of $(N+1)$ observations $X_0, X_1, ..., X_n$ of the process at known times $t_0, t_1, ..., t_n$. In the statement of equation (1), dW is the differential of the Wiener process (Brownian motion), $\mu(X;\theta)$ is the instantaneous drift, and $g(X;\theta)$ is the instantaneous diffusion.

Brownian Motion and SDEs

Brownian motion, also known as Wiener process, is a fundamental concept in stochastic calculus and plays a crucial role in modeling the random behavior in various systems. In the context of SDEs, Brownian motion represents the stochastic component of the system. The term dW in the SDE denotes the infinitesimal increments of the Wiener process, capturing the random fluctuations.

The SDE $dX = \mu(X;\theta)dt + g(X;\theta)dW$ comprises two parts:

- The drift term $\mu(X;\theta)dt$ represents the deterministic trend of the process, describing how the system evolves over time on average.
- The diffusion term $g(X;\theta)dW$ captures the random noise or variability in the system, which is essential for modeling real-world phenomena where uncertainty and variability are inherent.

Estimating the parameter θ involves determining the values that best describe the observed data within this stochastic framework. This requires advanced techniques to handle the complexity and randomness introduced by the Brownian motion.

Crop SDE Model

The proposed SDE model will introduce new functions for the drift and diffusion components to accurately reflect the dynamics of crop yields. The model will be designed to capture the intricate dependencies and stochastic nature of the agricultural processes.

Evaluation Methods: Euler and Stratonovich

The output model will be evaluated using the Euler and Stratonovich methods, two prominent numerical techniques for solving SDEs.

Euler Method

The Euler-Maruyama method is a straightforward extension of the Euler method for ordinary differential equations to the stochastic setting. It provides a simple yet effective way to approximate the solution of an SDE. Given the SDE:

$$dX = \mu(X;\theta)dt + g(X;\theta)dW,$$

the Euler method discretizes the time interval $[0, T]$ into small steps of size Δt. The approximate solution is obtained iteratively as:

$$X_{n+1} = X_n + \mu(X_n;\theta)\Delta t + g(X_n;\theta)\Delta W_n,$$

where ΔW_n represents the increments of the Wiener process. This method is computationally efficient and easy to implement, making it suitable for large-scale simulations.

Stratonovich Method

The Stratonovich method, on the other hand, is another approach to solving SDEs that is particularly useful when dealing with systems driven by physical processes. Unlike the Euler method, which uses the Itô interpretation of stochastic integrals, the Stratonovich method interprets the integral in a manner more consistent with classical calculus.

The Stratonovich approximation for the SDE is given by:

$$X_{n+1} = X_n + \mu(X_n;\theta)\Delta t + \frac{1}{2}\left[g(X_{n+1};\theta) + g(X_n;\theta)\right]\Delta W_n.$$

This method can provide more accurate results in certain cases, especially when the noise terms are multiplicative or when the system's physical properties dictate the use of Stratonovich calculus.

Both the Euler and Stratonovich methods will be employed to evaluate the proposed crop SDE model, ensuring robust and reliable predictions of crop yields. These evaluations will involve extensive testing and validation using historical crop yield data to assess the model's accuracy and computational efficiency.

Research Activities

Activity	Timeline	Date	Outcome
Implementing the SDE model based on small testing data using Python	1-3 months	TBA	Initial implementation and testing of the SDE model on small datasets
Tuning the model by adding parameters and functions	1-3 months	TBA	Improved accuracy and functionality of the model
Implementing NN that changes the parameters based on the output performance	2-3 months	TBA	Adaptive model that updates parameters dynamically
Optimize the ANN hyperparameters, the SDE functions, and number of yielding on real datasets	1-2 months	TBA	Optimized model ready for large-scale data application

Table 1: Activity and timeline

Skills Required

Stochastic Differential Equations.

- Advanced math
- Stochastic Differential Equations
- Machine Learning and Deep Learning
- Programming (Python)
- Agricultural Science and Crop Modelling
- Research and Technical Writing

References

- Zhi Xie, Don Kulasiri, Sandhya Samarasinghe, and Channa Rajanayaka. The estimation of parameters for stochastic differential equations using neural networks. Inverse Problems in Science and Engineering, 15:629–641, 2007.
- Geng Yuan, Hiroaki Mukaidani, and Tadashi Shima. Uncertainty quantification in neural networks using stochastic differential equations. 2023 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE), pages 1543–1548, 2023.

- Aldi Eka Wahyu Widianto, Endah R.M. Putri, Imam Mukhlash, and Mohammad Iqbal. European high-dimensional option pricing using backward stochastic differential equation-based convolutional neural network. Proceedings of the 2023 6th International Conference on Mathematics and Statistics, 2023.
- Dilli Paudel, Hendrik Boogaard, Allard de Wit, Sander Janssen, Sjoukje A. Osinga, Christos Pylianidis, and Ioannis N. Athanasiadis. Machine learning for large-scale crop yield forecasting. Agricultural Systems, 2020.