

# Agrizero: an O2O adversarial framework and triangulation for agricultural forecasting

Yenming J. Chen\*

Department of Logistics Management, National Kaohsiung University of Science & Technology, Kaohsiung, Taiwan

\*Email: yjjchen@ccms.nkfust.edu.tw

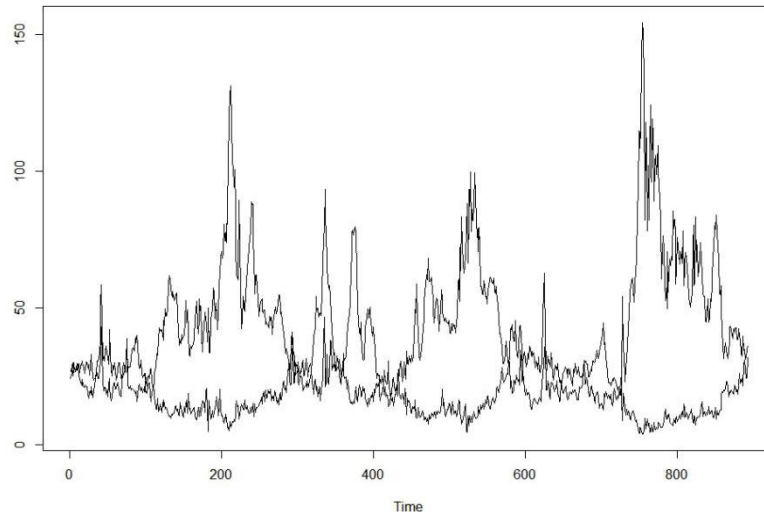
**Abstract.** This study proposes an online to offline (O2O) learning framework, ArgiZero, based on three components: a generative adversarial network, an online auction market, and offline simulated agents (digital twins: buyers, farmers, natures, and markets). The generative time series and digital twins are massively generated in a manner of Monte Carlo but with extremely efficient algorithms. The goal of the generator is to produce time series that are statistically indistinguishable with the records from auction market. The goal of the discriminator is to develop a triangulation method based on semi-modeless assimilation to separate generated from actual time series. Most farmers believe agriculture is impossible to be planned because uncertainty, crowding, crisis, and risk cannot be foreseen and accommodated. This AgriZero framework alleviates the challenges through the techniques of Bayesian deep learning and data assimilation as well as the mega power of GPU computation. Thanks to Bayesian hierarchical estimation, which is akin to deep learning but more sophisticated and longer history. We are able to estimate human behaviour in the agents of buyers and farmers, natural disaster in the agents of natures, and price fluctuations in the agents of markets. The framework has been validated by a large amount of records in vegetable auctions of Taiwan and USA. The hierarchical Bayesian estimation and Monte Carlo Markov Chain particle filters used in hidden Markov model are appreciated during the massive construction of the most probable digital twins. The feature space mapping in wavelet time series, Bayesian deep learning in recurrent neural network, kernel induced Hamiltonian dynamics ABC, and hybrid SDE-kernel based forecasting for time series analysis embodies the particle generator in the GAN structure. We also apply time series clustering, RNN, bagging and boosting, and semi-modeless assimilation to assist the performance of the triangulated discriminator.

## 1. Introduction

Challenges exist in forecasting cultivation yields and crop prices. Market prices may fluctuate without obvious reasons. Unlike the business practices in the industry, no actual demand is specified for a farmer. In other words, no customers place orders for agricultural produce. Price is determined completely in an auction market through a demand-supply-price relation. In the agriculture business, no one promises to buy produce with promised prices. Instead, farmers guess future trends of prices through a history of transactions. Without the help of scientific management, farmers mostly gamble in low bet efficiency. Because all farmers look at the same price history, they may make the same "optimal planting decision" in variety and quantity. This excessive over-supply usually makes produce



price plunge. Whenever a loss is encountered, the government is obligated to allocate tax money to subsidize the losses, and consumers must buy produce with high prices.



**Figure 1.** Transaction quantities and price for celery in three consecutive years.

By inspecting figure 1, the motivation for our agriculture research is well understood. The graph shows the transaction quantity and price for celery in three consecutive years. The supply quantities oscillate yearly and prices response reversely. Apparently, the most profitable timing cannot be at the high and low peaks of production. An experienced farmer usually uses their intuition to guess a right timing and quantity for their cultivation plan. Traditionally, because of multiple constraints imposed by the natural characteristics of crops, such as weather, temperature, and scale of economies, people mistakenly believe that they have limited freedom in making the optimal decisions for cultivation. Without the possibility of control, allowing for the contribution of forecast becomes unconventional. This mis-understanding is probably the reason why prior studies in cultivation planning are few. However, almost all farmers, regardless of whether they are doing in a right way or not, implicitly apply newsvendor theory to their planting quantity decisions. According to their intuition, farmers usually plant an increased number of crops if they envision an increase in the probability of high demand (demand  $\gg$  supply) and higher profit (price  $\gg$  cost). If this intuition is toward the right direction, scientific approaches can definitely help farmers avoid mistakes and improve their managerial yields.

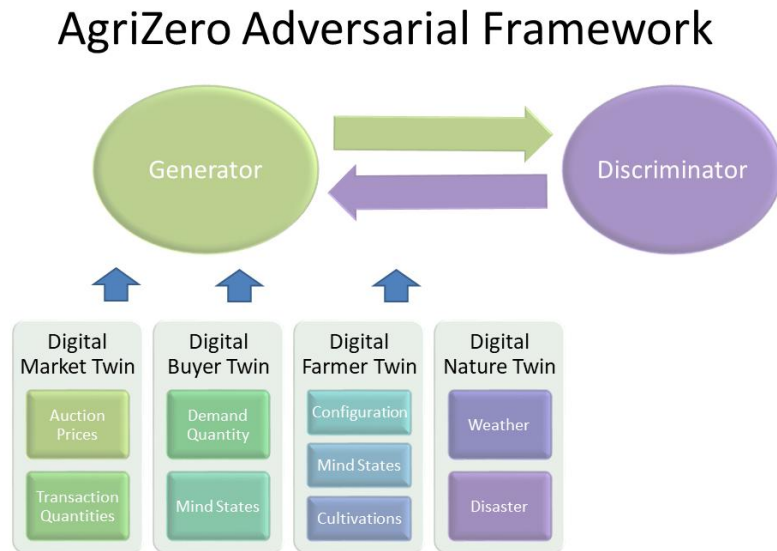
We propose an online to offline (O2O) learning framework, ArgiZero to tackle the challenges in agriculture business. Basically, the generative time series and digital twins are massively generated in a manner of Monte Carlo but with extremely efficient algorithms.

## 2. Literature

This framework is related to three streams of literature. We are mostly concerned with business and practical aspects of the agriculture supply chain as well as with the mechanism to make the system sustainable. To make the research practical, big data analytics and deep learning in generative models are investigated.

This framework exploits machine learning in stochastic modelling. Deep generative models recently received tremendous attention because of their superior performance. However, parameter estimation for latent variables within the models could be challenged due to highly non-linearity and chaotic state trajectories. [1] proposed a generative adversarial net in which two models are trained iteratively. The two models form a minimax two-player game such that a generative model is trained to mimic given data and the other discriminative model is trained to evaluate the performance of the target generative model. This approach effectively replaces the ordinary mean square error metric used

traditionally. [2] developed deep convolutional to replace the inner product operator in the middle layers for application of anomaly detection, classification, and clustering. [3] summarized the algorithms of unsupervised deep learning for time-series feature extraction. [4] reviewed the deep learning evolution in the form of literature chronicles. On the other hand, the deep generative models can help the MCMC sampling more efficient. [5] facilitated neural dynamics of recurrent networks for the sampling of stochastic computation and suggested that the stochastic computation is closer to the mechanism of biological brain.



**Figure 2.** AgriZero adversarial framework.

### 3. Proposed framework

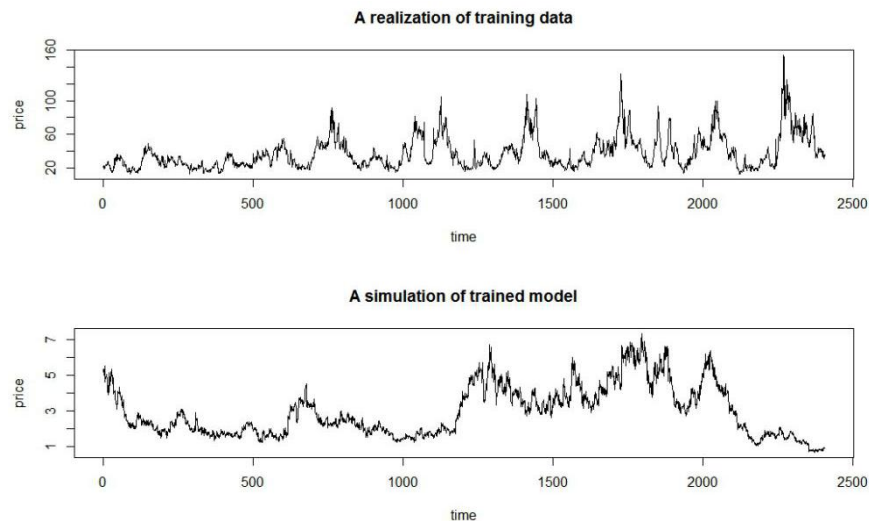
Our AgriZero framework incorporates big data estimation techniques and price dependent demand model into an integrated decision model, as shown in figure 2. Therefore, our model is different from others in three perspectives: 1) the performance of training is not fixed by a metric like  $mse$  but it will be discriminated by another learning agent, which devote to challenge the former's decisions, 2) the long consequence of a decision will firstly reflect to a virtual market which can embed all kinds of factors, and 3) the reinforcement of decision logic can be immediately tested by the online transaction results.

This framework uses partially observed Markov process (POMP) models as a fundamental computation platform for the generative agents. The models are also known as hidden Markov models (HMM) or state space models, which are powerful for time series analysis. We focus on Monte Carlo statistics, where the process and noises are nonlinear and non-Gaussian. The underlying algorithms include sequential Monte Carlo MLP, Hamiltonian MLP, nonlinear iterated filtering, particle MCMC, and approximate Bayesian computation (ABC). As shown in figure 3, the prediction of celery price and quantity through the learning of POMP in the three years are promising.

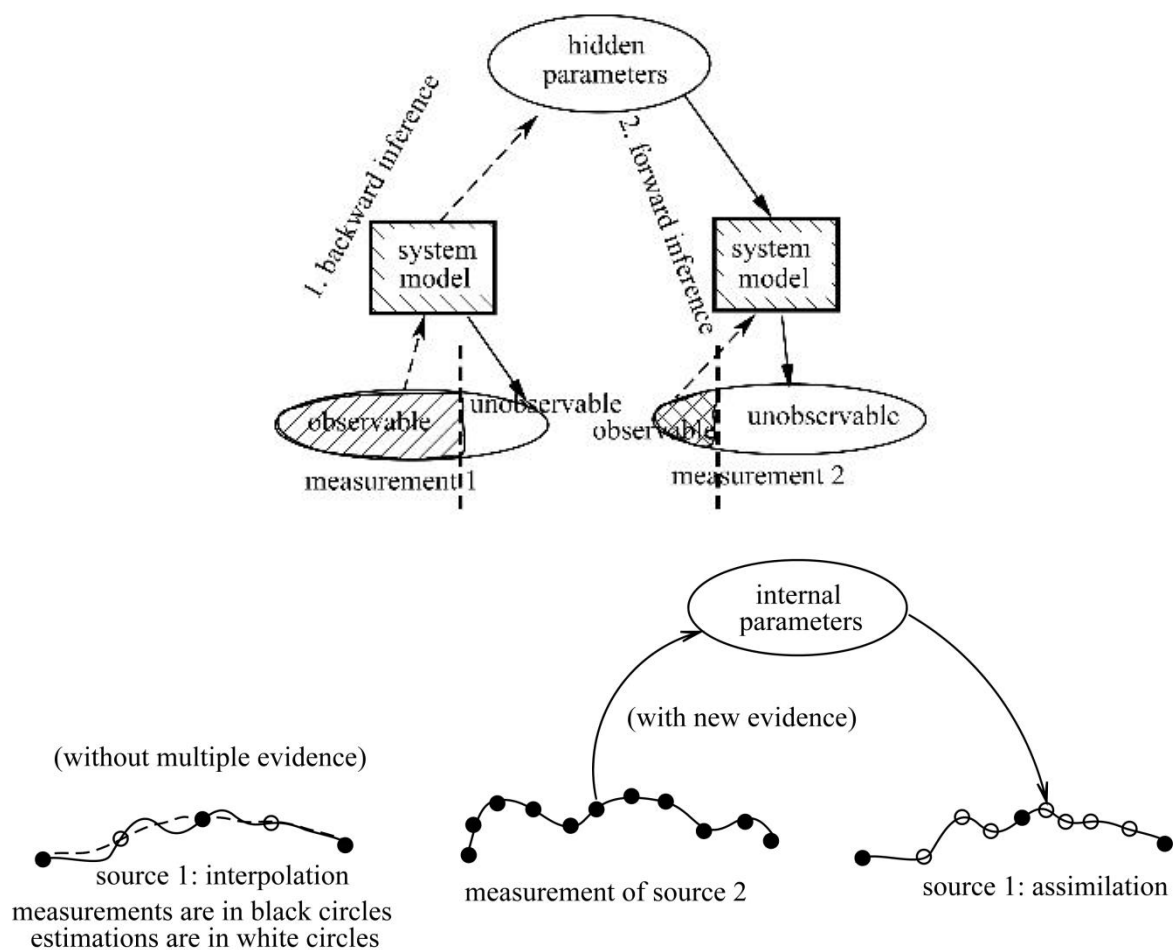
Triangular data assimilation used in the AgriZero framework refers to the quantitative methods by which the observations of variables are combined with system behaviours to provide estimates of internal states and key parameters. The discriminator can easily to use the framework of triangular data assimilation to boost the performance of generative agents. General data assimilation applies sparse observed information to estimate system internal states by making assumptions based on the fact that the system behaviours are typically governed by a set of pre-defined equations or statistical models. A suggested treatment can therefore be derived once all internal states are clear.

The estimation of the unmeasured demands depicted in the left panel of figure 4 is not accurate with a single measuring source. True demand should be presented in the solid line, but the estimation falls to the dashed lines due to the lack of sufficient information. However, through other sources, as

shown in the right panel of figure 4, the estimation becomes more accurate. New evidences can be derived from observations that result from the common causes.



**Figure 3.** The learning results of POMP and prediction of celery price and quantity in three years.

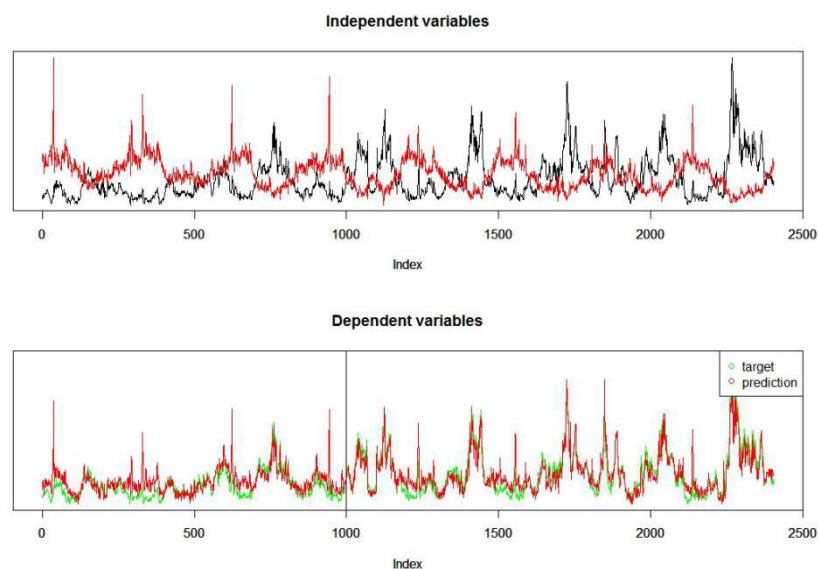


**Figure 4.** Triangular Assimilation.

#### 4. Results

Solving a parameter estimation problem is highly complicated and mostly nonlinear and combinatorial. Combining recent advances in machine learning and stochastic achievements offer an excellent means to achieve the goal. Bagging and boosting has been proven effective in machine learning. The algorithm of adaBoost classification was reported to have excellent capability. In the classification work, the adaBoost classification algorithm successfully combines the merits of multiple algorithms to tackle the challenges. In this framework, the machine learning work is supported mainly by the high-efficiency MCMC and deep learning.

To estimate the joint probability densities, the hierarchical Bayesian learning and MCMC methods were used in the algorithm. The MCMC methods represent a broad range of integration techniques and can approximate a multidimensional integral by generating samples through an ergodic Markov chain, thereby ensuring that the generated samples converge to the desired distribution based on the ergodic theorem. We apply the hierarchical Bayesian estimation and Monte Carlo Markov Chain particle filters during the massive construction of the most probable digital twins. The feature space mapping in wavelet time series, Bayesian deep learning in recurrent neural network, kernel induced Hamiltonian dynamics ABC, and hybrid SDE-kernel based forecasting for time series analysis embodies the particle generator in the GAN structure. figure 5 shows the result of forecasting by applying time series clustering, RNN, bagging and boosting, and semi-modelless assimilation.



**Figure 5.** The forecasting result of the AgriZero.

To conduct high-performance computation algorithms, we apply contemporary GPU, in particular NVIDIA GPU. Because NVIDIA is among the most dominant in the graphics hardware market, their proprietary CUDA programming interface has become the industry standard to distributing thread tasks to parallel programming cores with the scalable ability.

To further overcome misclassification errors, the AdaBoost algorithm combines a collection of weak classifiers (around 500 to 1000) to approach the Bayes classifier iteratively. To improve the classification rate, training data with high misclassification rates are emphasized and boosted in the next iteration. The final classifier is generally obtained by collecting votes from all the classifiers, although it could also be a simple linear combination of the classifiers.

#### 5. Conclusions

Our AgriZero framework is novel in helping investigate the effects of business and managerial insights in terms of the optimal control theory, O2O adversarial framework, and triangular assimilation. We established a modeling work for cultivation planning and will focus on model identification as

well as Markov time series prediction. This study contributes to the field of operations management in a holistic view of time series analysis in government data.

With our AgriZero framework, farmers may increase their profits by using modern management technologies. For whatever reasons that agriculture management research has been ignored for a long time, we wish to raise the possibility of increasing the efficiency of crop production and reducing the waste through the help of modern AI technologies.

## References

- [1] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems*. pp 2672-80
- [2] Yuan, Chao, Amit Chakraborty. 2016. Deep convolutional factor analyser for multivariate time series modeling. *Data Mining (ICDM), 2016 IEEE 16th International Conference on*. IEEE, pp 1323-28
- [3] Långkvist, Martin, Lars Karlsson, Amy Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters* 42 pp 11-24
- [4] Schmidhuber, Jürgen. 2015. Deep learning in neural networks: An overview. *Neural networks* pp 61 85-117
- [5] Buesing, Lars, Johannes Bill, Bernhard Nessler, Wolfgang Maass. 2011. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS computational biology* 7(11) e1002211