Climate Smart Farming – Deployment of Digital Twin Concepts in Agricultural Seed Value Chain

Sundar Raja Vadlamani
President – Supply Chian
SeedWorks International Pvt. Ltd
Hyderabad, India
raja.vadlamani@seedworks.com

Chevala VVS Narayana
Manager – Digital Transformation & Data Scientist
SeedWorks International Pvt Ltd
Hyderabad, India
vvsnarayana.chevala@seedworks.com

Pulkit Mittal
Digital Business Leader,
Tractor Manufacturing and Farming Equipment

Abstract

To ensure food security, seed production systems should innovate in the direction of increasing production while optimizing utilized resources amidst changing climate. Rice (Oryza sativa L.) is one of the pivotal staple cereal crops feeding more than half of the world population. In recent years, hybrid seeds have proven to be an effective solution to increasing yield. Seed production is a complex matrix of exogenous factors such as sowing practices, fertigation, biotic/abiotic stresses, and endogenous factors such as seed setting, floret opening, etc. Based on our field experience and past research studies, a large collection of factors have been identified, which can potentially impact the yield. Several factors, which were represented in this study have been found to impact the listed attributes of global hybrid seed productions such as a number of panicles per given area, spikelet grains per panicle, filled grains per panicle, and grain weight. Our study includes the use of Digital Twins and agent-based simulations to enable interaction between the real world and digital environments for "Acre-by-Acre Prescriptions (ABA)" on agronomical practices as farming recommendations, including the measurement of water consumption, conservation, and benchmarking the right optimum. By using these methods, the yield forecasting is improved to near accuracy of 89% as against current industry benchmarking of 70-75% accuracy, and with ABA prescription models the yield gain improved by 10% and water conserved by 30%.

Keywords
Rice Seed Production, Sustainable Farming, Climate-Smart Agriculture, Digital Twin, Modelling and Simulations

1. Introduction
Rice (*Oryza sativa* L.) is one of the pivotal staple cereal crops feeding more than half of the world population. It is grown in more than 100 countries with China and India contributing to more than 50% of the global production. Annually 700 million tons of paddy (470 million tons of milled rice) are being produced on a total harvested area of approximately 158 million hectares worldwide. With the increase in urbanization, the land available for rice production is steadily going down, which has led to an increased demand for a higher crop yield There are several changes faced by the farmers due to various factors such as water shortage, weather uncertainties, agronomy challenges and inaccurate crop prediction as a consequences of climate change. Seed with superior genetics that are climate resilient and less harming to climate is the differentiator for ensuring the objective. Next the Right Agronomy adopting CSA will be the key to achieve the desire results. Production of Right seeds is a challenge being faced globally especially in Hybrid rice seed production. In recent years, hybrid seeds have proven to be an effective solution

to increase yield. However, the hybrid seed production is becoming a challenge due to climate change and various associated factors which hamper the growth of hybridization in India (Current hybridization is 7%; Raja V, 2016; Gangaiah B, 2019). It is observed that the large variation in the seed yield is due to a multitude of factors including the weather, location, and seed production process. While researchers have conducted localized controlled experiments to study how individual attributes affect yield, there is a lack of accurate data-driven approaches, which integrate large collections of data from multiple sources to predict the yield and provide recommendations of production planning to a seed producer that addresses sustainable farming solutions (climate smart agriculture), resource optimization and risk mitigation. A higher yield in rice seed production increases the profitability of the farmers and ensures seed security to food security. The hybrid paddy seeds are produced by crossing the parent varieties and are found to outperform the traditional varieties (such as OPV (Open Pollinated Varieties) - Open-Pollinated Varieties) in terms of yield and quality of the produce (Ferguson et al. 1991). Hence, to meet the increasing global demand for rice, several research institutes, government agencies, and industry players are developing new and improved rice hybrids that are climate smart with high resource-use efficiency (RUE), with reduced GHG emissions.

Rice seed production precedes the commercial paddy production step and the seeds produced are input to the latter. Increasing yield for seed production directly impacts commercial paddy production due to the availability of high yielding hybrids. Besides, commercial paddy production has received more attention from researchers while the seed production technology for hybrid seeds is relatively underserved due to its niche area. The seed production technology for hybrid seeds is also significantly different and highly complex, with many steps, compared to commercial paddy production. Hybrid seeds are not only required to be produced year-on-year for the subsequent commercial crop cycles but also required to be produced with high quality to achieve a high yield in the fields. Seed production is a complex matrix of exogenous factors such as sowing practices, fertilization methods, etc., and endogenous factors such as seed setting, floret opening, etc. (Biswas, P. K. & Salokhe, V. M., 2001, Kameswara Rao, N. & Jackson, M. T., 1996). Understanding the suitable agronomical practices, early prediction of abiotic/biotic stress, best input quantification is key to achieve better yield and helpful to the seed producers and growers as there is a large variability in the year-to-year hybrid seed production (Figure 2). Hence, forecasting trends of the hybrid seeds supply is challenging, leading to a compounding effect on the availability of commercial paddy production. As seeds are a perishable good, despite their extended development cycle, poor forecasting can have huge impacts on the business of seed producers and the food security of the country. This means that seed producers have a limited period between production and sale of the product to adjust to variations in supply. Players in the seed industry are generally unorganized and focused on small geographic areas and they need huge investments in product development and inventory management. Hence, spatial-temporal variations in yield have a significant impact on these businesses.

The main challenge encountered by scientists and corporates involved in rice research and production in the world is to find appropriate solutions for major issues such as understanding the optimal agronomy, high/low temperature in the temperate areas, problems of water use efficiency, availability and pollution, land constraints, major biotic stresses, improvement of rice yield, with emphasis on bridging yield gaps, raising yield ceiling and reversing yield decline, rice quality, decline in investment for increased rice production, high costs of production and the flow of rice information (Ismael Fofana, 2011). In agricultural research, efforts are made to know the hidden regularities of some aspects of soils, crop agronomy and other related biological phenomena. It may be yield potentials of a crop, effects of pest incidence on crop yields, effects of changes in climate on pests and crop, effect of fertilizer application on crop yields, effect of cross-fertilization of different crop varieties, and the like. The present knowledge concerning these aspects may be inadequate with respective hybrid seed production. For a seed producer, there are two main challenges – understanding the hidden regulators, predicting the yield for every farm-level, and deciding a good production plan, which includes deciding where, when, and how much should the seed producer invest in sowing the parent seeds to maximize the yield of hybrid seeds. Also, several studies have investigated the methane flux in rice field which were correlated to the soil water content during crop development. Effective tracking to substantially enhance water use efficiency (WUE) through benchmarking the water resource consumption are potential to reduce the carbon emissions. This study will address both these challenges

1.1 Objectives

Our simulations use data for creating virtual models for Hybrid rice seed production cycle from seedlings to harvesting. Considering the above inferences, the modelling and simulations need to assess the impact of attributes on yield, identify the crop recommendations under different crop phases to address the duration of vegetative and reproductive phases, synchronization and seed setting, along with optimal resource use. Based on these needs, this work focused on the following objectives:

1. Optimize the supply chain seed production planning against the target intake volumes using yield estimates, objectives and constraints to recommend end-to-end plan while maximizing yield and minimizing cost.
2. To demonstrate advanced machine learning techniques such as Naive Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees (XGBoost) for crop yield estimation
3. To analyze the adapted machine learning techniques using evaluation metrics such as R2, RMSE (Root Mean Square Error), MAE, MSE (Mean Squared Error) (Mean Squared Error), MAPE, CV, and NSME.
4. Simulate the best accurate prediction technique to evaluate the relationship between crop attributes to understand the impact on yield
5. Bring insights to understanding agronomy for "Acre-by-Acre Prescriptions (ABA)" as farming recommendations

With this we would like to sustainably increase rice hybrid seed productivity through farm recommendations, adapting and building resilience to climate-change with enhanced resource use efficiency while reducing and/or removing greenhouse gas emissions with optimal water consumption, wherever possible.

2. Literature Review

Intensive field studies were published to review the status of regional rice production, focusing on the gaps between potential and actual yield in the Asia-Pacific region countries (Papademetriou et al. 1999) while taking account of the factors responsible for it. E.A.Siddiq (Papademetriou et al. 1999) in the report highlighted the yield gaps in India to the consolidation of yield by correction of yield destabilizing factors is, however, considered as the more promising short-term strategy. Moreover, this research by R.C Chaudary (Papademetriou et al. 1999) is knowledge shared on farmer-front rice cultivation addressing strategies for bridging the gap in rice by means of vertical (increase of yield) and horizontal (rice area) expansion with reduction in yield losses, by means of models to bridge the gap/narrowing the gap to aim not only to increase rice yield and production but also to improve efficiency of land and labour use, to reduce the cost of production, and to increase sustainability. Exploitable yield gaps of rice are often caused by several factors including physical, biological, socio-economic, and institutional constraints, which can be effectively improved through participatory and an integrated approach. India's irrigation sector is one of the largest in the world (A. Narayanamoorthy, 2022). Most studies on contrast to high rice yield indicate water as main factor for yield gaps and yield variability from experiments stations to farms (Papademetriou et al. 1999). However, not many studies have analysed the water consumption in rice-based irrigation systems. Hundertmark et. Al (2003) developed a framework for water use in rice-based irrigation systems and a strategy for improving the system performance, especially its water productivity. More recently, deep learning has also been used in many fields of natural and agricultural science. Much of the success of deep learning is based on the ability of neural networks to recognize patterns in high-dimensional space. In advance to modelling, there were concepts to enable models to create digital representations of connected environments with an open modelling language. Model buildings, factories, farms, energy, networks, railways, stadiums, even entire cities. By tracking the past and helping predict the future of any connected environment it is known as – **Digital Twins** (Michael Grieves, 2002).

Rice field water uses (to account for evapotranspiration plus seepage and percolation) an average, about 2,500 Liters of water need to be supplied (by rainfall and/or irrigation) to a rice field to produce 1 kg of rough rice grain. This average number is derived from many experimental data at the individual field level across Asia (Bouman, Bas. 2009) - So, on an average the productivity of rough Rice in Telangana/ CG/Odisha/KS is 2500 Kgs per acre and the average consumption of water per acre is 6250000 litres. (6.25 million litres)/1.6 million gallons. We highlight the seed yield per acre in the cited states approximately 600 kilograms (kgs) per acre and the bench marked water consumption per Kg of seed production is 10000 to 11000 Litres. (Table 1). Using the inch levels in paddy fields at weekly intervals, the approximate water consumption is arrived based on the following mathematical equation. One inch is $^{1}/_{12}$ feet so one acre of water 1 inch deep is 43560 ($^{1}/_{12}$) = 3630 cubic feet of water. There are 231 cubic inches in an American gallon. A cubic foot is 12 inches by 12 inches by 12 inches and thus one cubic foot is 12*12*12 = 1728 cubic inches. So, 3630 cubic feet is 3630*1728 = 6272640 cubic inches. Thus, the amount of water needed to cover one acre to a depth of 1 inch is $^{6272640}/_{231}$ = 27154.3 gallons (1 Gallon = 4.54 liters)

3. Methods

3.1 Feature extraction, selection and construction

Since the dataset has many input variables, feature construction involves transforming a given set of input features to generate a new set of more powerful features which are then used for prediction and are interpretable.

3.2 Model description

Statistical analysis is adapted primarily, namely multiple linear regression (MLR), to determine the effect of some independent variables on dependent variables to compute the linear dependence of the variables. It defines an association between known (x) and unknown variables (y) based on the random noise and its parameters, and it is expressed as below:

$y_i = \beta X_i + \varepsilon_i$ (1)

where y_i denotes a predicted rate; $X_i = (1, x_1, x_2, x_3, \ldots x_n)$ are the terms for the explanatory vector variables; $\beta = (\beta_0, \beta_1, \beta_3, \ldots \beta_k)^T$ represents a vector coefficient; ε_i denotes a random error for ith observation. The input data comprises approximately 200 crop-related attributes covering the entire crop duration period which were collected using the field inspection app (figure 2). Further, normalization is adapted to prepare data reduction and remove the data redundancy for machine learning applications.

Regression analysis focuses on a dependent variable and series of other changing variables – making it particularly useful for prediction and forecasting. Figure 2 depicts the conceptual framework of this paper. Machine learning model development involves data acquisition from multiple sources, data processing, feature engineering to make it suitable for building the model, choosing algorithm to build the model, build model, compute performance metrics and choosing the best performing model as explained in an end-to-end system (figure 3). Comprehensive analysis of various machine learning models was assessed for crop yield estimations, it includes Naive Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees (XGBoost).

3.2.1 Naive Bayes

Naive Bayes is a high-bias, low-variance classifier, and it can build a good model even with a small data set. It is simple to use and computationally inexpensive. Typical use cases involve text categorization, including spam detection, sentiment analysis, and recommender systems. The fundamental assumption of Naive Bayes is that, given the value of the label (the class), the value of any Attribute is independent of the value of any other Attribute.

3.2.2 Deep Learning

Deep Learning is based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using backpropagation. The network can have many hidden layers consisting of neurons with tanh, rectifier and maxout activation functions. Advanced features such as adaptive learning rate, rate annealing, momentum training, dropout and L1 or L2 regularization enable high predictive accuracy. Each compute node trains a copy of the global model parameters on its local data with multi-threading (asynchronously) and contributes periodically to the global model via model averaging across the network.
The operator starts a 1-node local H2O cluster and runs the algorithm on it.

3.2.3 Decision Tree

A decision tree is a tree like collection of nodes intended to create a decision on values affiliation to a class or an estimate of a numerical target value. Each node represents a splitting rule for one specific Attribute. For classification this rule separates values belonging to different classes, for regression it separates them in order to reduce the error in an optimal way for the selected parameter criterion. The building of new nodes is repeated until the stopping criteria are met. A prediction for the class label Attribute is determined depending on the majority of Examples which reached this leaf during generation, while an estimation for a numerical value is obtained by averaging the values in a leaf. This Operator can process containing both nominal and numerical Attributes. The label Attribute must be nominal for classification and numerical for regression.

3.2.4 Random Forest

A random forest is an ensemble of a certain number of random trees, specified by the number of trees parameter. These trees are created/trained on bootstrapped sub-sets of the dataset provided as an input. Each node of a tree represents a splitting rule for one specific Attribute. Only a sub-set of Attributes, specified with the subset ratio criterion, is considered for the splitting rule selection. This rule separates values in an optimal way for the selected parameter criterion. For classification the rule is separating values belonging to different classes, while for regression it separates them in order to reduce the error made by the estimation. The building of new nodes is repeated until the stopping criteria are met.

3.2.5 Gradient Boosted Model

A gradient boosted model is an ensemble of either regression or classification tree models. Both are forward-learning ensemble methods that obtain predictive results through gradually improved estimations. Boosting is a flexible nonlinear regression procedure that helps improving the accuracy of trees. By sequentially applying weak classification algorithms to the incrementally changed data, a series of decision trees are created that produce an ensemble of weak prediction models. While boosting trees increases their accuracy, it also decreases speed and human interpretability. The gradient boosting method generalizes tree boosting to minimize these issues.

3.3 Model selection

Tuning hyperparameters of machine learning models and selecting best models with optimal hyperparameter values is necessary to achieve high prediction accuracies. Cross-validation is commonly used to evaluate the predictive performance of fitted models by dividing the training set to train and validation subsets. An Automated hyperparameter tuning finds the best parameters for each model. Figure 2 depicts the modelling and simulation workflow, describes the model deployment and real-time integration of operation data from ERP system, automated feature engineering, input simulator to understand the crop attributes which can be taken as insights to improve the on-field execution.

3.4 Mathematical optimization approaches

Mathematical optimization is a proven, powerful AI (Artificial Intelligence) technology that has delivered tremendous benefits in terms of efficiency and profitability for companies around world and across many industries. As mathematical optimization is being used in an expanding array of mission-critical applications by business today and has delivered such significant business value, with no surprise that this AI technology is gaining traction among decision makers within these companies (Kruk Serge, 2018). The applications were widely used in the areas of planning, logistics, operational research, resource scheduling, and other business processes. Here we highlight the importance of these techniques to optimize the seed production planning while maximizing yield and minimizing production cost.

3.5 Approaches

To attain the designated objectives, the following steps need to demonstrate the application of the selected machine learning algorithms.

- Collect the data using available sources
- Distribute the data into two segments: training data (70%) and testing data (30%).
- Develop the machine learning model to assess the crop yield.
- Determine the evaluation metrics for each model.
- Simulate the crop yield using adapted techniques to turn predictive models into Acre-by-Acre Prescriptive Actions
- Optimize the supply chain production planning using yield estimates, objectives and constraints
- Recommend the optimal plan for seed production planning using observed outcomes of model, to maximize yield and minimize cost.

4. Data Collection

4.1 Filed site application

The historical data of paddy seed production of SeedWorks has been used for the study. We have well-established data collection mechanisms to collect comprehensive data across geographies. For data collection, the seed production team employed dedicated staff who visit farms regularly and enter data. The staff uses an internal mobile application, named Siddhi, for capturing key agronomy data of seed production up to 200 attributes for each field from sowing to harvest which is integrated to ERP System in real-time. The user interfaces for the Siddhi app are shown in Figure 3. The raw data collected from this app will be processed through the model architecture shown in Figure 4. In the past few years, we have collected millions of data points and were fed to various models for this analysis. Hence, multivariate time series data for each location will be directly used as inputs to the prediction models.

4.2 Data

Field measurements of crop data include a diverse collection of factors that are identified to have potential impact on the crop yield as summarized below:

- Soil attributes, Time of soaking, Date of transplanting, Age of transplanting, Tillering, fertigation, Primordial initiation, the reproductive growth stages, Panicle exertion, synchronization status, 50 % flowering, pollen availability, sigma exertions, no of days of supplementary pollination, Number of panicles per given area, water source etc.
- Number of spikelets per panicle, Duration of reproductive phase, average daily temperatures during booting (21-days period beginning from one week after primordial initiation to heading), Average night temperatures, Average diurnal variations, Average relative humidity
- Number of low temperature (<15 C) & cloudy days during pollen formation and seed setting days, Number of rainy days during flowering, Number of high temperature (> 40 C) days during flowering and dough stage and abortion percentage.
- Test weight, Temperature during the dough phase, seed moisture
- Other: Spikelet Sterility (%), plant height, flowering date

Furthermore, the data collection includes the water levels measured and recorded as inches at weekly intervals.

5. Results and Discussion

In this paper, efforts were made to build a system that can make predictions on crop yield and to generate an optimized large-scale production planning. We predicted the crop yield based on data from multiple sources and using seven different machine learning models such as Linear Regression, Deep Learning and Gradient Boosting Trees out of which the GBT's proved to be even more efficient. The optimized planning, model accuracy and recommendations have increased productivity by 10%. Simulations were performed for the accurate prediction of crop yield and to assess the system's recommendations to bring more insights at farm-level. We also found that the prediction accuracy will rely on various factors such as regional difference, type of algorithm used and the agricultural zone.

5.1 Numerical Results

Table 1: Water consumption per acre

Depth of Water	**Per week consumption (Galloons)**	**Required for weeks**	**Total consumption (Gallons)**
1/4 Inch (Thin layer)	6750	16	108000
1 Inch (2.54 cms)	27000	16	432000
2 Inch (5 cms)	54000	16	864000
3 Inch (7.5 cms)	81000	16	1296000

5.2 Graphical Results

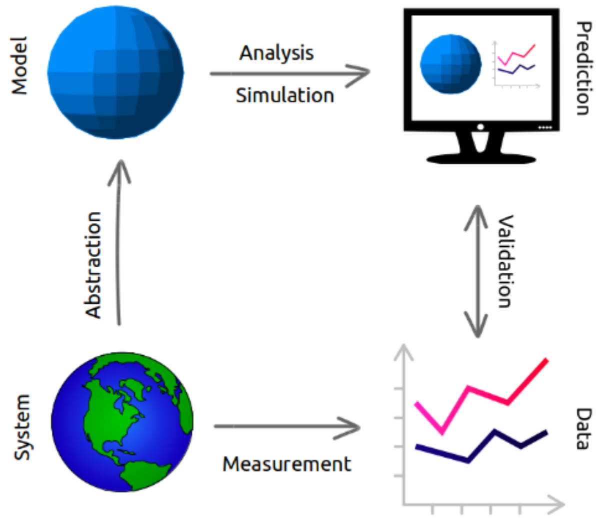

Figure 1: Digital Twin Concepts – Data, System, Modelling and Validations

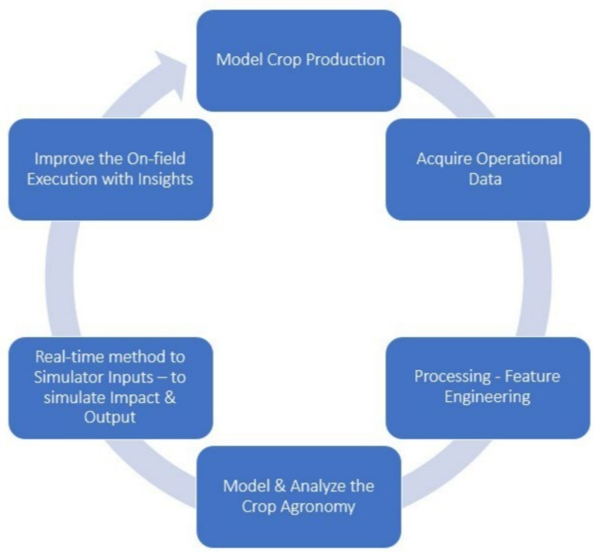

Figure 2: Modelling and Simulation – Workflow

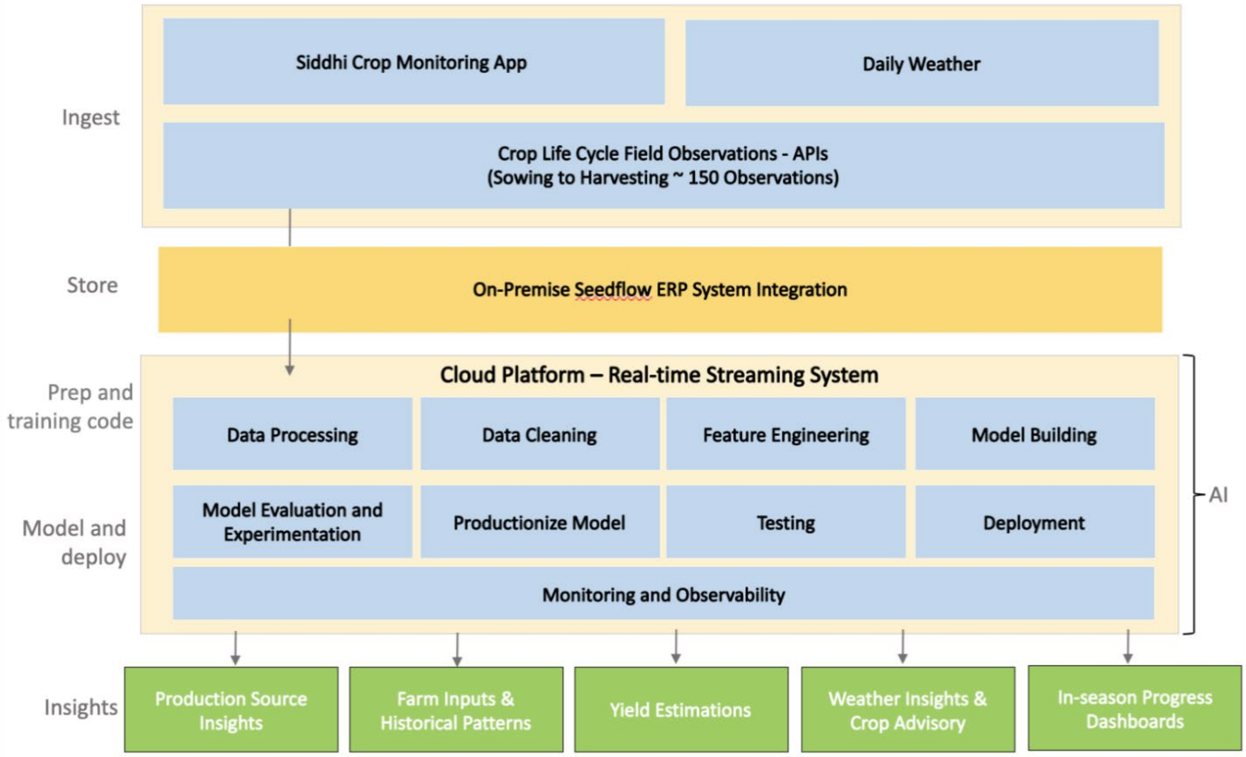

Figure 3: Data Collection, Modelling and Monitoring – System Outputs

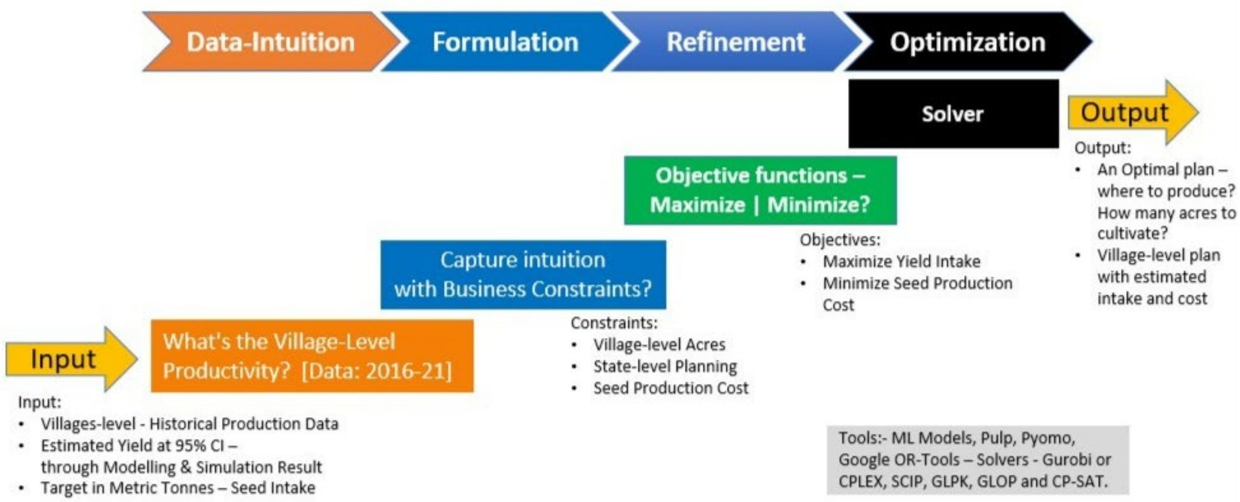

Figure 4: Mathematical Optimization – Objectives, Constrains, Refinement for Seed Production Planning

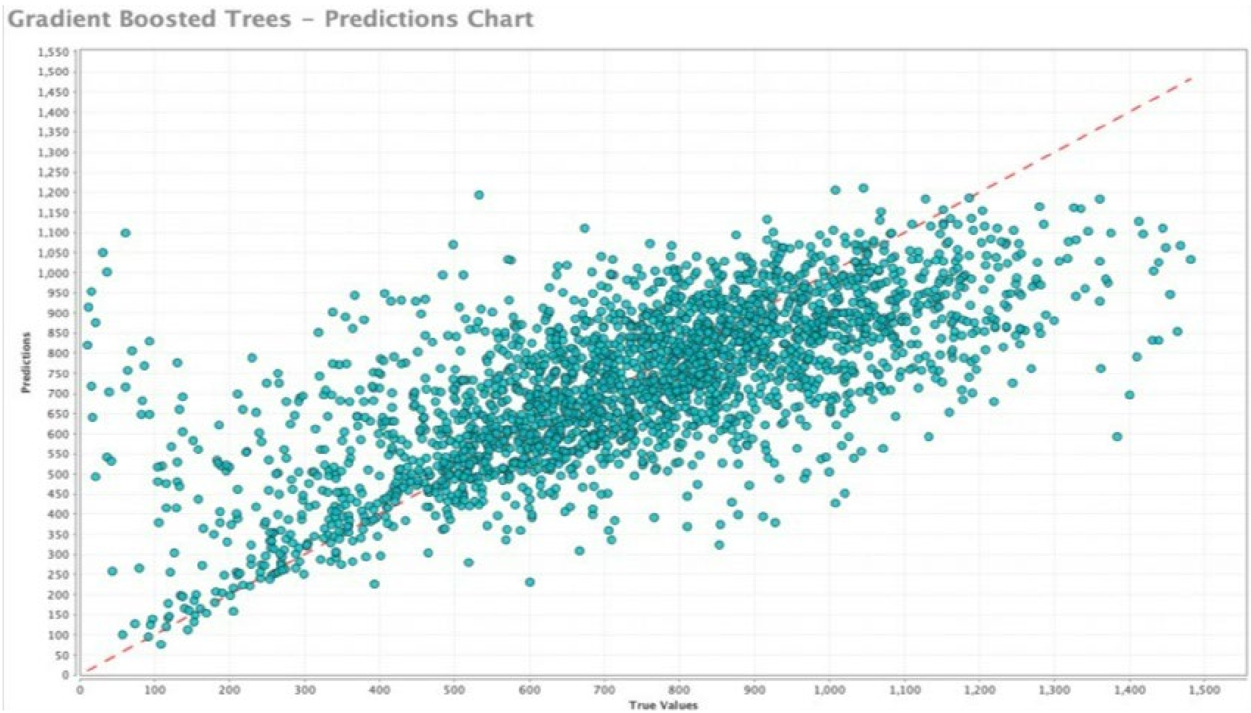

Figure 5: Model Performance (Cross-plot) for the Gradient Boosting Regressor algorithm in the seed production yield estimation

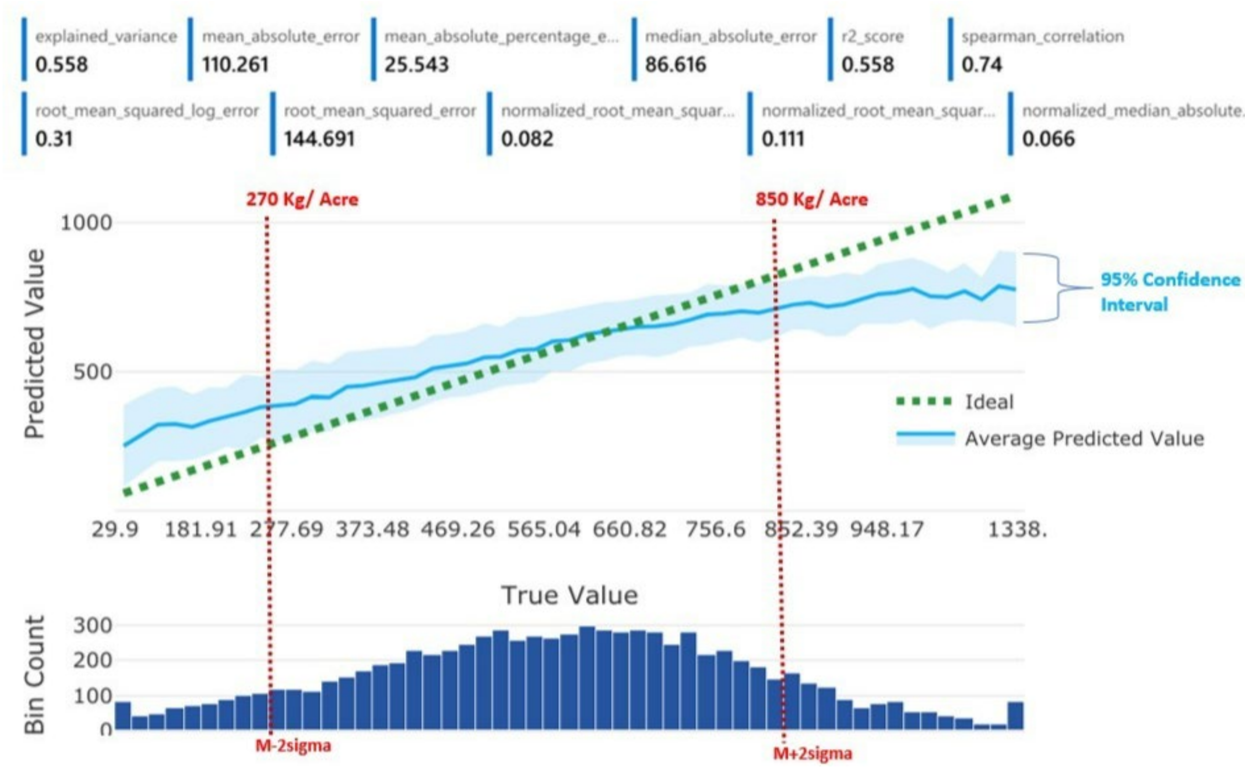

Figure 6: Model Validation – Gradient Boosted Trees performance against validation dataset

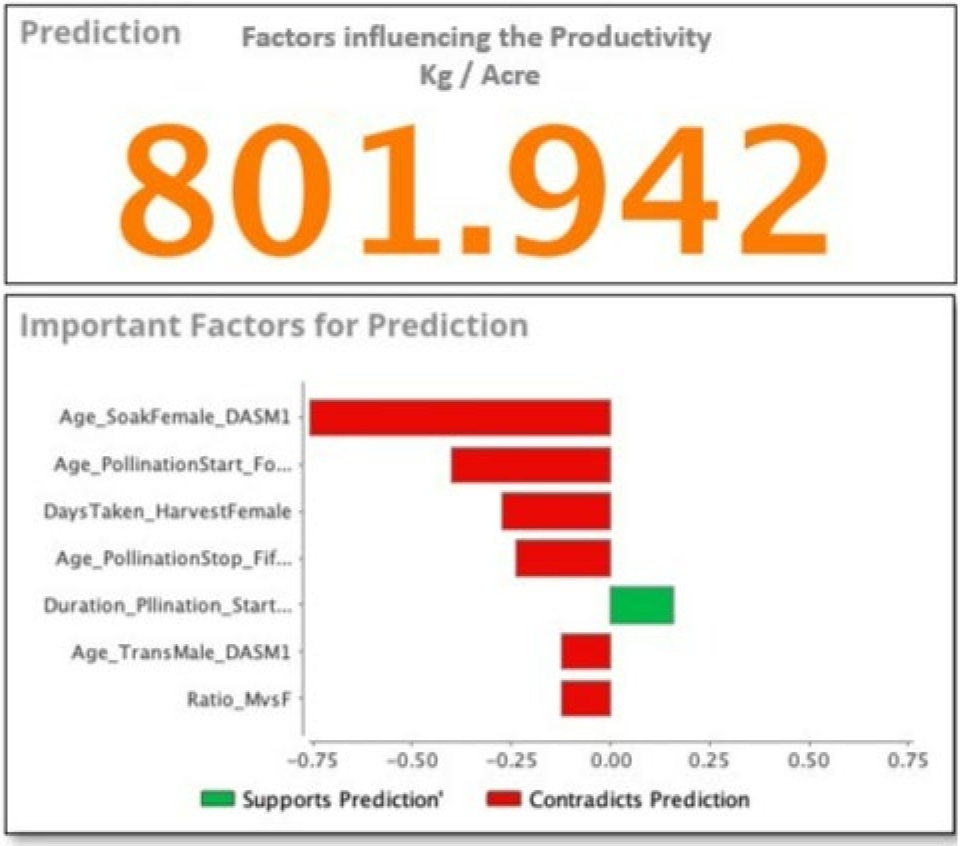

Figure 7: Simulation - Factors Influencing the Yield Prediction and Order of Importance

5.3 Proposed Improvements
Machine learning approaches are increasingly used to extract patterns and insights to extract patterns and insights from ever-increasing streams of geospatial data, but current approaches may not be optimal when system behavior is dominated by spatial and temporal context. Here, rather than amending classical machine learning approaches, we propose the use of ever-increasing streams of geospatial data in contextual cues as part of deep learning (an approach that is able to extract spatio-temporal features automatically) to gain further process understanding of Earth system science problems, improving the predictive ability of forecasting and modelling of spatial connections across multiple timescales. For the satellite image data, we can use the Convolutional Neural Networks (CNN) based Variational Autoencoder (VAE) to find latent embeddings of images of farmland patches. While researchers have used GTWR models for real-estate market prediction, the same can be used for yield prediction in modeling temporal dynamic behavior.

5.4 Validation
Quantification and validation of the simulation results were based mainly on the relative mean bias error (MBE: relative difference between simulated and observed values), the coefficient of determination (r^2: square of the Pearson correlation coefficient), root mean square error (RMSE), and the coefficient of variation (CV: ratio of standard deviation to the mean value) as highlighted in Figure 6.

6. Conclusion
An integrated crop management approach (water, soil fertility/nutrients, agronomy, weeds/pest/diseases, and climate) is vital to maximize the productivity and profitability of rice seed production farmers. All technologies and practices should be used synergistically to help farmers increase and/or maintain gain yields at the same or reduced cost. Improving the seed quality and increasing the yield will enhance seed availability and farmers' profitability.

In this paper a benchmark data-driven model concept for seed production – planning and execution is presented. It focuses on mathematical optimization tools for large-scale automated planning on villages and acreages with given on-field constraints while maximizing yield and reducing cost. This study can serve as a benchmark and foundation for accelerated research in this area. Further, the digital-twins concepts with respect to rice seed production can be improved by incorporating the geospatial dataset which enables spatial-temporal knowledge. There is also a need for understanding the right optimum water for rice seed production. Based on our studies rice seed production currently consumes 10 thousand liters (about the volume of a storage unit) of water per kg of seed produced which can ideally be reduced to 7 thousand liters (about the volume of a storage unit) of water per kg with AWD methods and practicing thin layer water level with much compromise on seed yield.

References

Ismael Fofana., Simulating the impact of climate change and adaptation strategies on farm productivity and income, *FPRI discussion paper*, pages: 17, 2011

DRR., Package and practices for cultivation of hybrid rice, Selected Agricultural Technologies, *A Compendium: Directorate of Rice Research (ICAR)*, pp.1-5, 2004.

Babu, V. Ravindra, K. Shreya, Kuldeep Singh Dangi, G. Usharani, and A. Siva Shankar., Correlation and path analysis studies in popular rice hybrids of India, *International Journal of Scientific and Research Publications* 2, no. 3, 1-5, 2012

Chlingaryan, Anna, Salah Sukkarieh, and Brett Whelan., Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture, *A review: Computers and electronics in agriculture*, 151, 61-69, 2018

Lijie Zhou, Shanshan Liang, Kimberley Ponce, Severino Marundon, Guoyou Ye, Xiangqian Zhao., Factors affecting head rice yield and chalkiness in indica rice, *Field Crops Research*, vol. 172, Pages 1-10, 2015.

Biswas, P. K. & Salokhe, V. M., Effects of planting date, Intensity of tiller separation and plant density on the yield of transplanted rice, Journal of Agricultural Science, Cambridge 137, 279–287, 2001.

Kameswara Rao, N. & Jackson, M. T., Effect of sowing date and harvest time on longevity of rice seeds, *Seed Science Research,* 7, 13–20, 1996.

Krishnan, P., & Surya Rao, A., Effects of genotype and environment on seed yield and quality of rice, *The Journal of Agricultural Science*, 143(4), 283-292, 2005.

John E. Sheehy, P.L. Mitchell, Anaida B. Ferrer., Decline in rice grain yields with temperature: Models and correlations can give different estimates, *Field Crops Research*, Volume 98, Issues 2–3, 2006, Pages 151-156

Ferguson, J. M., R. D. Keys, F. W. McLaughlin, and J. M. Warren., Seed and Seed Quality, Available: https://content.ces.ncsu.edu/seed-and-seed-quality#section_heading_5235 *NC State Extension Publications*, January 1991.

Performance Of Private Sector Rice Hybrids Under Different Agro-Climatic Zones Of Uttar Pradesh. Available: https://upcaronline.org/upload/5527c2a8139beKharif%20Rice%20Report%20(2014).pdf, *U.P. Council of Agricultural Research*, 2020.

Huang, Bo, Bo Wu, and Michael Barry., Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices, *International Journal of Geographical Information Science,* 24, no. 3, 383-401, 2010.

Gamboa, John Cristian Borges., Deep learning for time-series analysis, *arXiv preprint arXiv*:1701.01887 2017.

J. Masison, J. Beezley, Y. Mei, Hal Ribeiro, A. C. Knapp, L. Sordo Vieira, B. Adhikari., A modular computational framework for medical digital twins, *PNAS*, 118 (20) e2024287118, May 2021.

S. Van Der Walt, S. C. Colbert, G. Varoquaux., The NumPy array: A structure for efficient numerical computation, *Comput. Sci. Eng,* 13, 22–30, 2011.

F. Tao, Q. Qi., Make more digital twins, *Nature* 573, 490–491, 2019.

Maged N. Kamel Boulos and Peng Zhang., Digital Twins: From Personalised Medicine to Precision Public Health, *J Pers Med*, 11(8): 745, Aug 2021.

Breiman. L., Random forests, *Mach. Learn*, 2001.

Friedman JH., Greedy Function Approximation: *A Gradient Boosting Machine. Ann. Stat*, 29:1189–1232, doi: 10.1214/aos/1013203451, 2001.

Guyon I, Elisseeff A., An Introduction to Variable and Feature Selection, *J. Mach. Learn. Res,* **3:**1157–1182, 2003

Papademetriou, M.K. (ed.), Dent, F.J. (ed.), Herath, E.M. (ed.)., Bridging the rice yield gap in the Asia-Pacific region. 20, Bangkok, Thailand, October 5-7, 1999. 1020-6221|1014-2789|18194591|0428-9560

Gauch. H.G., Statistical Analysis of Regional Yield Trials: AMMI Analysis of factorial Designs, *Elsevier Publ. Co*, 278 pp, 1992.
Gleick, P.H. (ed.)., Water in Crisis: A Guide to the World's Fresh water Resources, *Oxford University Press*, New York. 1993.
Gomez, K.A., On-farm assessment of yield constraints: methodological problems. In: Constraints to high yields on Asian rice farms: an interim report. *IRRI*, Los Baños, Philippines. 1-16 pp. 1977.
A. Narayanamoorthy., Temporal Trends and Regional Patterns in Development of Irrigation in India, *The Irrigation Future of India*, 01 January 2022, pp 49-72.
Hundertmark, W. Abdourahmane, A.T., A diagnostic model framework for water use in rice-based irrigation systems, *International Water Management Institute*, Research Report 74, Colombo, Sri Lanka, 2003
Brenowitz, N. D. and Bretherton, C. S., Prognostic Validation of a Neural Network Unified Physics Parameterization, *Geophysical Research Letters*, 45, 6289–6298, Available: https://doi.org/10.1029/2018GL078510, http://doi.wiley.com/10.1029/2018GL078510, 2018.
Chevallier, F., Chéruy, F., Scott, N. A., and Chédin, A., A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget, *Journal of Applied Meteorology*, 37, 1385–1397, 1998.
Yuval, J. and O'Gorman, P. A., Use of machine learning to improve simulations of climate, Available: http://arxiv.org/abs/2001.03151, 2020.
Rasp, S., Pritchard, M. S. and Gentine, P., Deep learning to represent subgrid processes in climate models, *Proceedings of the National Academy of Sciences of the United States of America*, 115, 9684–9689, Available: https://doi.org/10.1073/pnas.1810286115, 2018.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat., Deep learning and process understanding for data-driven Earth system science, *Nature*, 566, 195–204, Available: https://doi.org/10.1038/s41586-019-0912-1, http://www.nature.com/articles/s41586-019-0912-1, 2019.
Bouman, Bas. (2009). How much water does rice use? Rice Today. 8. 28-29.
Raja V, Hybrid rice India 2016, Available: https://www.linkedin. com/pulse/hybrid-rice-india-2016status-raja-vadlamani/, 2016
Gangaiah B, Rice Hybrids and High Yielding Varieties for Augmenting Public Distribution System Demand in Andaman and Nicobar Islands, India. *International Journal of Current Microbiology and Applied Sciences*, 10.20546/ijcmas.2019.806.382, 2019

Acknowledgements
We would like to acknowledge SeedWorks International Pvt Ltd for extending the required support to design, execute and evaluate the project.

Biography / Biographies
Sunder Raja Vadlamani is currently as a President of Supply Chain at SeedWorks International Pvt Ltd. He is a graduate in Agriculture by training & alumni of Institute of Agriculture Science, Banarus Hindu University and Asian Institute of Management, Manila, Philippines. His association with seed industry is 34+ years with various multinational corporates before joining SeedWorks International Pvt Ltd. His primary contributions and focus areas are supply chain, agricultural value chain, seed production, innovation, and sustainable farming.

Chevala VVS Narayana, is currently a Manager – Digital Transformation and Data Scientist at SeedWorks International Pvt Ltd. Before, he served as Manager – IT and Senior Analyst between 2015-2020 with different corporate organizations, and as a senior research fellow between 2012-2015 at the National Institute of Plant Genome Research (NIPGR), India. He holds an undergraduate degree in Biotechnology from Department of Biotechnology, PR Govt College, Andhra University, a Master's in Bioinformatics from Department of Bioinformatics, Pondicherry Central University (PCU), India. Before joining to SeedWorks International Pvt Ltd, he worked in academics as Senior Research Fellow in Department of Biotechnology, India funded projects at NIPGR for almost 3 years. In addition to his academic experience, he has national industrial experience such as; Data Engineer in Godrej & Boyce Pvt Ltd. India, Senior Analyst in Bluedart-DHL Pvt Ltd, India, and Senior Manager - IT in UnitedHealth Group. He has published two articles in leading scientific journals indexed by Web of Science, PubMed, PubMed Central, Scopus, Google Scholar, such as Nature Scientific Reports and Journal of Experimental Botany. His primary research areas are genomics, agricultural innovation, statistical modeling, optimization, digitalization and manufacturing strategies, and project management.