Tailored fertilizer advice analytical methods

Machine learning approach

Harnessing the power of cutting-edge machine learning models, our first approach builds on the relationship between crop yield and nutrient utilization, primarily nitrogen (N), phosphorus (P), and potassium (K), alongside key biophysical factors influencing crop growth. The biophysical factors include soil properties sourced from available digital sol maps such as soilGrids and iSDA, weather information from Chirps. Once the machine learning model is trained, it can be deployed across scales, accommodating a wide array of NPK combinations to forecast expected yield against the desired NPK matrix. These insights serve as a foundation for optimizing fertilizer advice tailored to specific targets, be it nutrient rates for achieving target yield, maximizing return on fertilizer investment, which can further be aligned with farmers' investment capacity and risk tolerance. This result can also be used to assess nutrient requirements to mitigate the agronomy yield gap and/or estimation of total fertilizer requirement at a national scale in support of decision making in fertilizer purchase, distribution, subsidy etc.

Data Needs: This approach hinges on comprehensive field trials data encompassing a vast spectrum of NPK combinations and corresponding yield measurements, augmented by geo-spatial variables including soil properties, elevation, and climate data. The crop yield data provides an understanding of yield effect due to nutrients added via fertilizers use, and the geo-spatial variables help further fine-tune these effects to locally relevant bio-physical factors.

Complexity: While of moderate complexity, this approach entails aggregating extensive field trials data, preparing geo-spatial layers, and constructing a predictive model. The training and application of machine learning models such as the gradient boosting, random forest, deep learning, and the ensemble models is made straightforward with semi-automated procedures in AgWise. Excellent data exploration skills coupled with agronomy knowledge and understanding of machine learning models are imperative to assess the quality and understand the limitation of the data and make informed decisions in model selection.

The ground truth data is typically aggregated from numerous field experiments conducted across multiple years and institutes. The data origins from a combination of both on-farm and on-station trials from experiments with different protocols such as mono- and inter-cropping, replicated and single treatments, nutrient omission trials to protocols set to test effect of micronutrients, lime and/or organic fertilizer addition, etc. Consequently, the data exhibits the strength and challenges inherent to meta-analysis, necessitating meticulous data cleanup and exploration to mitigate error propagation that could jeopardize model accuracy.

While data quality challenges are ubiquitous across approaches, this method uniquely grapples with model constraints tied to the nutrient levels and range present in the training dataset. Notably, the model's capacity to estimate yield response is confined to the nutrient range covered by the training data. In instance where the NPK rates in the training dataset are not sufficiently large for the target area, the recommendations generated from this approach may fall short of optimal levels particularly in high yield potential areas, potentially leading to missed opportunities for yield or profit maximization. Additionally, the resulting stepwise yield response curve, transitioning from one nutrient level to the next, poses a limitation. When the training data lacks diversity in NPK rates, this issue is exacerbated, resulting in a uniform yield response across a broader range of nutrient rates, contrary to biological norms.

Application: This approach is successfully deployed in developing tailored fertilizer advice in Ethiopia for maize, wheat, sorghum and tef.

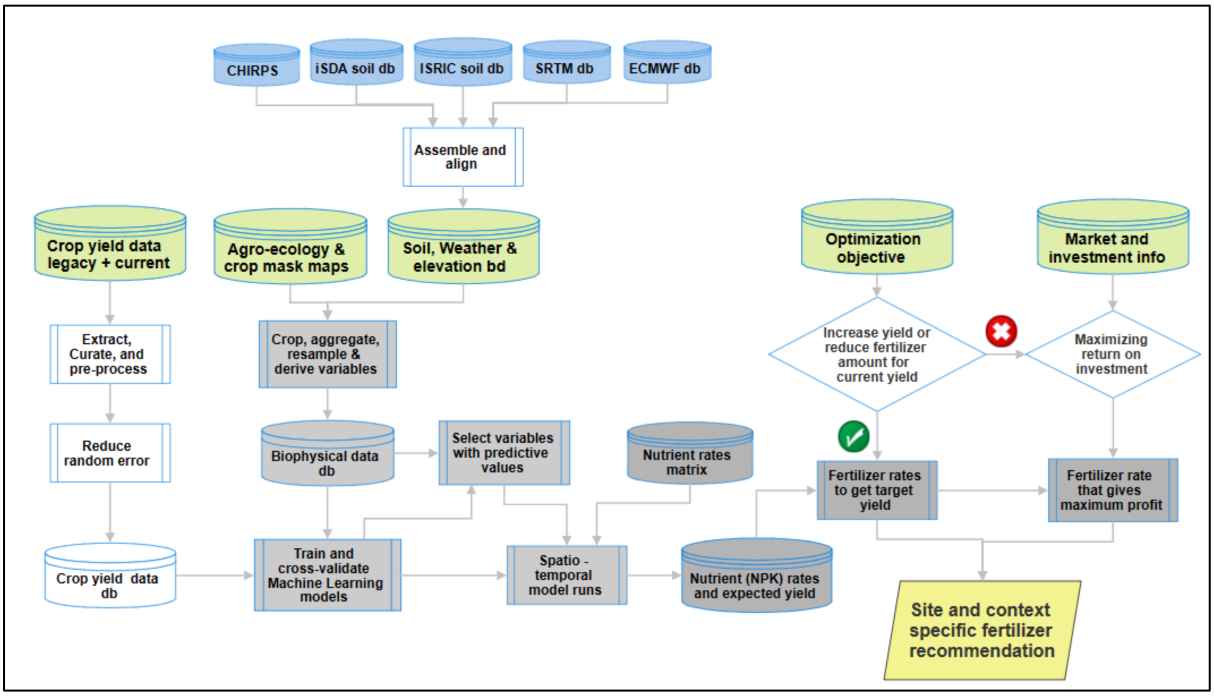

Figure 1. The machine learning approach

QUEFTS-based approach

As an alternative to the machine learning approach, this approach leverages the QUEFTS (Quantitative Evaluation of Fertility of Tropical Soils) crop model to produce tailored fertilizer advice. QUEFTS's strength lies in its ability to model the dynamic relationship between soil nutrients supply and yield, accounting for crop specific nutrient uptake capacity and the maximum attainable yield at a given location. The estimation of apparent soil nutrient supply (ANS), representing the soil's NPK available to a crop considering indigenous soil nutrients, water availability, and field management effects, is central to this method. Achieving this estimation relies on an optimization algorithm rooted in reverse QUEFTS methodology, iteratively adjusting ANS values to best fit observed yield responses while minimizing the residual sum of squared errors of the estimation. This iterative and location-specific approach ensured that ANS estimations are tailored to the unique conditions present at each trial location. Subsequently, ANS values can be aggregated based on similar recommendation domains, such as agroecological zones, to define the NPK availability for crops in specific areas. Forward QUEFTS is then employed with large sets of combinations of NPK to predict yields based on ANS. Like the machine learning approach, the results can be further refined to provide fertilizer advice aimed at meeting yield or profit increases.

Data Needs: The QUEFTS approach requires quality field trials data with at least three treatments being tested at a location with yields, along with information on NPK combinations. While data from nutrient omission trials, including full NPK, a control with zero fertilizer input and the three omission treatments, are ideal, they are not mandatory.

Complexity: Moderate in complexity, this approach necessitates meticulous data curation and error reduction, coupled with QUEFTS methodology application. It is less data intensive than the machine learning approach as it can account for crop specific parameters to model biologically realistic yield response curves from fewer nutrient rate levels. As a result, it can predict yields for NPK levels that are not available in training dataset safeguarding users in high yield potential areas from conservative fertilizer advice. This approach however has limited application when there is limited field trials data in a highly variable environment basically because of the inability to explain the soil nutrient supply of the target area using few ANS obtained from the experimental sites. This limitation lays the groundwork for the subsequent approach integrating QUEFTS and the ML methods to estimate ANS at scale.

Application: The QUEFTS approach is successfully deployed in developing tailored fertilizer for maize and rice in Rwanda.

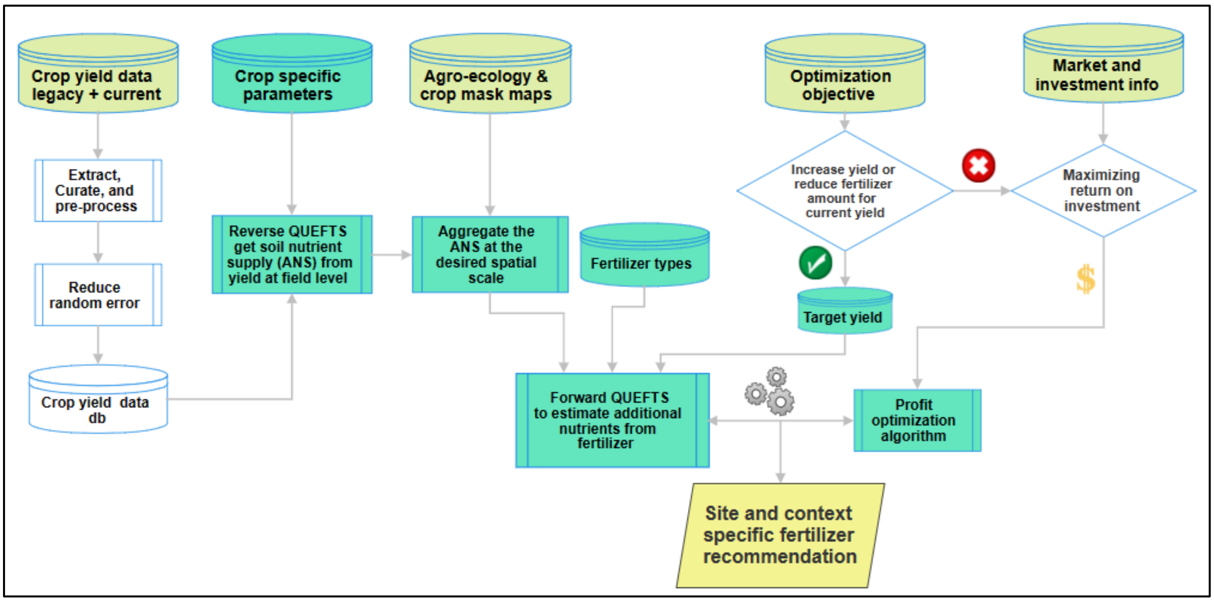

Figure 2. QUEFTS-based nutrient rate estimation approach

QUEFTS- based and Machine Learning approach

Advancing upon the methodologies of the previous approaches, this integrated approach combines QUEFTS principles with machine learning techniques to extrapolate the estimation of apparent soil nutrient supply (ANS) to a more variable environment by accounting the variation in the geo-spatial factors into account. Machine learning models are trained to establish intricate relationships between ANS derived from reverse QUEFTS and geo-spatial variables at experimental sites. This trained model is then used to estimate ANS at scale, at desired resolutions, across various locations. While it is naturally logical to attribute soil nutrient supply bio-physical factors such as sol properties, elevation, and weather variables, the current geo-spatial data available at scale, can account for approximately 30% - 40% of the structured spatial variation of soil nutrient supply. To enhance the model's predictive capabilities substantially, the inclusion of local soil fertility indicators, particularly current yield class information, has been proven instrumental. The current yield class can be established from survey or field trials data and curated by agronomists working in the target area.

Data Needs: This approach requires field trials data to estimate the current yield class and the ANS using reverse QUEFTS. The field trials data should meet the quality standards outlined in the QUEFTSalone approach, with the addition of data or indications of current yield class being essential. Extensive geo-spatial data is also needed to apply the machine learning models.

Complexity: Of higher complexity, this approach integrates QUEFTS and machine learning techniques, requiring extensive data assessment to establish current yield levels. Ideally deployed within a system enabling users to input historical yield data from their fields and facilitating real-time computation for customized advice adjustments based on user-specific parameters such as farm size, fertilizer types and prices, and investment capacity. In such setup, it is possible to deliver hyper-contextualized advisory based on user input data on farm size, fertilizer types and prices, investments capacity, etc.

Application: This integrated approach is successfully deployed in developing tailored fertilizer advice for maize in Kenya and potato for Rwanda.

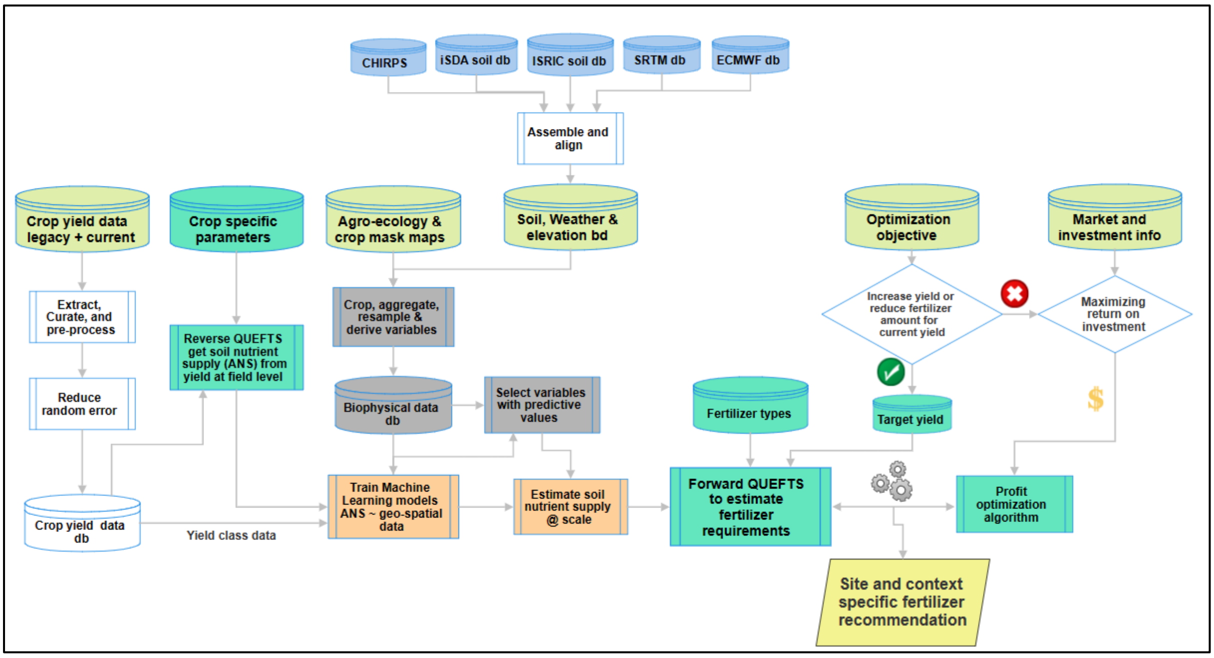

Figure 3. The integration of QUEFTS and Machine Learning approach

Integrated approach: QUEFTS, Machine Learning and Crop models

This is the most advanced approach of AgWise whereby use of crop simulation models such as DSSAT, APSIM, LINTUL and Oryza is integrated with the QUEFTS plus machine learning approach. One the limitations of the previous two approaches where QUEFTS is used is how the attainable yield is defined. The key limitation of the previous QUEFTS-based approaches is how the attainable yield which is a requirement of QUEFTS is set. So far, the attainable yield was set at the yield obtained from the highest NPK treatment plus 20%. The additional 20% was a deliberate choice to mitigate the influence of management practices on the observed yield at the highest fertilizer rates, and secondly, to acknowledge that the nutrient rates utilized in most of the field rials data are not sufficiently high enough to mitigate all nutrient limitations. This approach however can still lead to an underestimation of the attainable yield at a location especially when the highest rate is still in the linear zone of the yield response curve and /or when the yield from such plots is low due to biotic and/or abiotic factors.

This limitation was addressed by using the water limited yield estimated using crop simulation models as attainable yield. In addition, this integration enables AgWise to assess and account for the effect of weather on yield gap both in terms of climate scenarios and effect of panting dates and cultivar selection and provide climate smart tailored fertilizer advisory.

Data Needs: In addition to the data requirement outlined in the third approach, this method needs daily weather data spanning several years and crop management information for running crop simulation models. The sourcing of geo-spatial data and the procedures for running crop models are semi-automated within AgWise streamlining processes for efficient and timely response.

Complexity: With the highest complexity among the four approaches, this method demands advanced skills in utilizing crop simulation models and conducting extensive data analytics. However, the integration of these advanced techniques enables AgWise to offer unparalleled precision and insight in delivering tailored fertilizer advisory, thereby maximizing agricultural productivity and sustainability.

Application: This integrated approach is successfully deployed in developing tailored fertilizer for cassava in Nigeria, Tanzania, Rwanda and Ghana

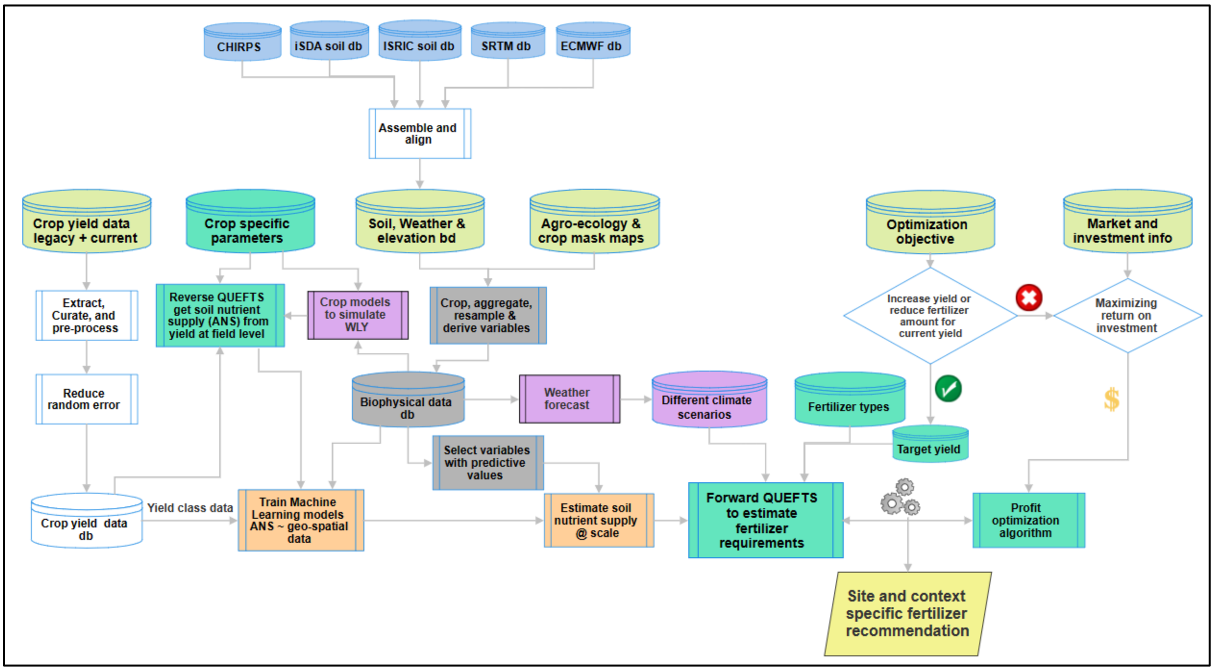

Figure 4. The advanced AgWise approach for climate smart tailored fertilizer advice

How to choose which approach to use:

Data Requirements: The complexity and accuracy of the models depend on the availability and quality of data. While the first two approaches rely mainly on field trials data, the third and fourth approaches incorporate extensive geo-spatial and climate data.

Complexity: The complexity increases from the first approach to the fourth, with the incorporation of advanced modeling techniques and data requirements.

Accuracy: Generally, as the complexity increases, the accuracy and precision of fertilizer advice tend to improve, especially in estimating soil nutrient supply and yield potential under varying conditions.

Applicability: The choice of approach depends on the availability of data, computational resources, and the specific requirements of the target audience. While simpler models may suffice for initial fertilizer advice, more complex models offer better precision and adaptability to diverse conditions.

In summary, each approach has its strengths and limitations, and the choice depends on the specific context and objectives of the fertilizer advisory system.