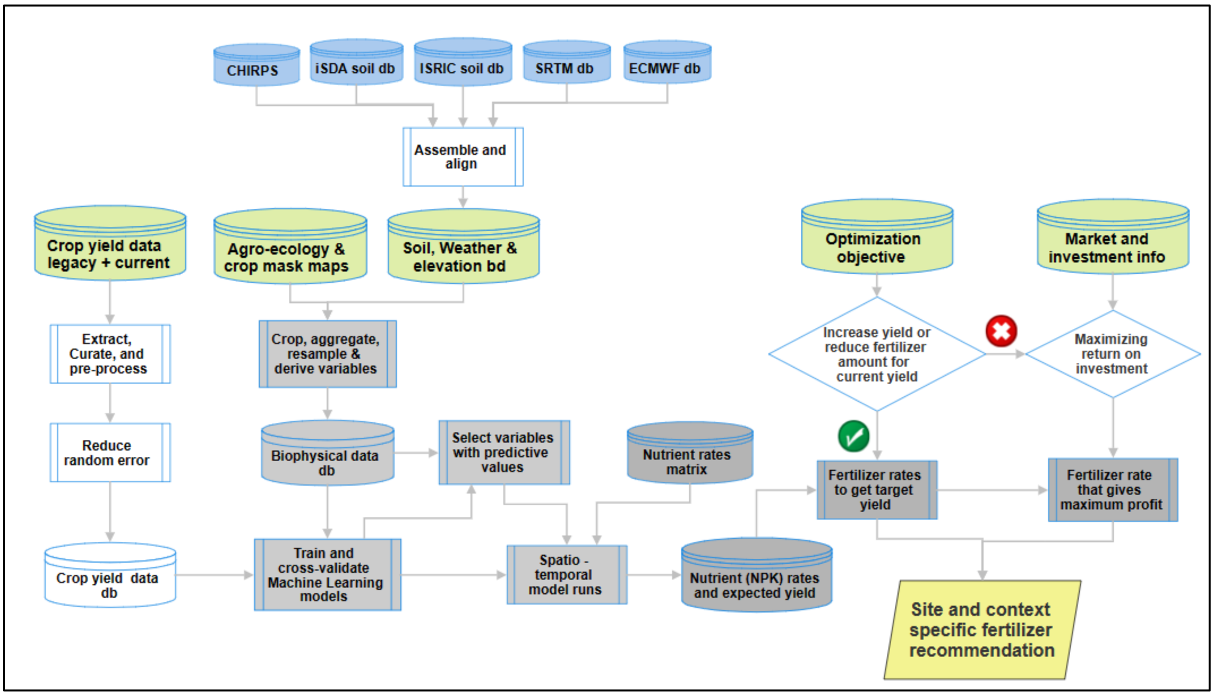
Tailored fertilizer advice analytical methods

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.