Exploring Global Technologies for Village/Field Level Crop Yield Estimation

Table of Contents

1. Introduction and objectives of the report

In this era of economic liberation and the increasing population, timely and accurate crop yield estimations are essential for government policies, insurance schemes, and other production-based management systems. The production forecasting of major cereal crops is important for national food security, which is composed of ecological, biological, socio-economic, political, and cultural dimensions. As there is an increase in demand for food production, the advanced assessment of crop production methods is becoming important to overcoming challenges, particularly the use of limited resources. On the other side, production risk to agri/crops due to weather anomalies, pest/diseases, and availability of seeds/inputs are exacerbated by climate change/variabilities. One of the options for risk management is crop insurance.

Crop insurance in India, initiated in 1915, is called a rain insurance scheme to insure farmers against drought. After the attainment of independence in 1947, crop insurance gradually started to find mention more often. There are two types of crop insurance policies considered in India: individual and homogeneous areas. In an individual approach, the farmers are compensated to the full extent of the losses, and then the premium is to be paid by him based on his own past yield and loss experience. In this approach, efficient, reliable, and accurate crop yield data is required from the individual farmers for a sufficiently long period to fix premium insurances. In contrast, the homogeneous approach the area comprising homogeneous villages form the point of view of crop production and whose annual variability of crop production would be similar. The homogeneous based approach is efficient in terms of the government policy rules; therefore, various agro-climatically areas are treated as a single unit where individual farmers pay the same rate of premium and receive the same benefits irrespective of their loss.

Different experiments on crop insurance were undertaken in India in early 1970s. The first crop insurance program was introduced in 1972-73 by the "General Insurance" Department of Life Insurance Corporation of India on H-4 cotton in Gujarat. Later, they implemented the experimental scheme to other crops, including groundnut, wheat, and potato, in different states in India. This experiment scheme was based on the "Individual Approach" continued up to 1978-79. Despite these efforts to introduce insurance policies, the individual farm approach was not viable and sustainable in this country. After that, the General Insurance Corporation of India (GIC) introduced a Pilot Crop Insurance Scheme (PCIS) in 1979. The risk was shared by the General Insurance Scheme (GIS) and the respective State Govt. in the ratio of 2:1. This PCIS continued until 1984-85; and the Comprehensive Crop Insurance Scheme (CCIS) was introduced in 1985 with the main objective of the state government's active participation. In this scheme, the Govt. covered crop insurance for notified crops and in notified areas. Further, several schemes were implemented to incur farmers losses including, Experimental Crop Insurance Scheme in 1997, followed by Pilot Scheme on Seed Crop Insurance in 2000, and Farm Income Insurance Scheme in 2003.

Pradhan Mantri Fasal Bima Yojana (PMFBY) was launched in Kharif, 2016 to support agriculture production by providing an affordable crop insurance product to ensure comprehensive risk cover for farmers' against all non-preventable natural risks from pre-sowing to the post-harvest stages of crops. The Scheme has completed eight crop seasons and is being implemented across States/Union Territories (UTs). The PMFBY highlights a few things from the previous ones. One is to cover at the Grama Panchayat level to better capture crop's risks at a lower level. This led to conducting enormous Crop Cutting Experiments (CCEs) at the lower community level. However, performing a number of CCEs is a time-consuming process, cost-effective, requires manual labour to visit the field, and sometimes leads to data uncertainty problems.

To overcome these challenges, the best solution may be the use of technologies to reduce CCEs and estimate yield loss directly. Globally, the countries are using drones, data from satellites, crop models, remote sensing derived indices, mobile platforms, and machine learning techniques, respectively, to estimate direct yields at the lowest community levels. For example, in the US, they use drone technologies to estimate the direct crop loss immediately after extreme weather conditions. This way, they can insure the farmers within a short period after the intensive rainfall event that causes flooding in crops and damages fields. In addition, satellite data can be used to estimate seasonal crop yield variation using derived remote sensing indices. Many spectral indices have been developed with remote sensing data to estimate yields. Mobile platforms offer tailor-made services to the farmers to guide them about weather conditions, nutrient management, and appropriate time for harvesting crops. The efficient use of global technologies for estimating crop yields, managing insurance policies, and continuous monitoring of vegetation conditions are providing valuable services in the agriculture sector.

This study was initiated to review the global best practices in the use of technologies for crop yield estimation and crop insurance to derive best practices and provide recommendations for their use in the Indian crop insurance scheme. The accumulated information should lead to the development of efficient methods for crop insurance policies and yield estimates in India.

Objectives:

1) Review the current methods of crop yield estimation in India for crop insurance purposes.
2) Review the global methodologies of crop yield estimation for crop insurance with their merits and demerits.
3) Analyse different technology-based methods and their potential for use in crop insurance in India.

The ToRs of the project (for rechecking)

1) Review of globally available technologies
2) Their usefulness in Indian condition

3) Comparative analysis of the global technologies vis-a-vis technologies used by the pilot studies conducted by the Department.
4) Recommendation of the best technology for adoption by PMFBY.

The report may be reorganized as per the objectives, if possible.

2. Methods currently used in India for crop yield estimation

2.1. Methods used by Govt. of India

2.1.1. Crop Cutting Experiments (CCE)

Crop Cutting Experiments or CCE refers to an assessment method employed by governments and agricultural bodies to accurately estimate the yield of a crop in a region during a given cultivation cycle. The CCE is a traditional approach where sample locations are selected based on a random stratified sampling of the total area under study. Once the plots are selected, the produce from a section of these plots is collected and analyzed for a number of parameters such as biomass weight, grain weight, moisture, and other indicative factors. The data gathered from this study is extrapolated to the entire region and provides a fairly accurate assessment of the average yield of the region under study.

In addition to conducting these CCEs to estimate total yield, the central government uses this data for its Pradhan Mantri Fasal Bima Yojana (PMFBY), a scheme that assists insurance companies to disburse payment for farmers crop insurance claims seamlessly. The PMFBY requires the data from each state conducting at least four CCEs in every village panchayat level for each crop within a month of the harvest period. The data gathered from CCE is useful to multiple stakeholders in the agricultural value chain. While governments use it for planning agricultural policies and programs for the future, the information helps financial institutions with all the inputs they need before offering loans or insurance coverage if there is a poor harvest or crop failure. The biggest drawback of this traditional method towards CCE is that it is dependent on a number of variables such as administrative setup, type and size of the field staff, farmer cooperation, and harvest conditions. Especially in a scenario where there are nearly **2.5 lakh *gram panchayats* or village councils** in India that are scattered, along with inadequate trained human labour or the time to facilitate these experiments effectively, there needs to be a more efficient way of utilizing the resources and obtaining an accurate yield estimation within the short harvesting window.

The use of technologies in agriculture creates room for precision farming techniques for accurate yield estimates. Compared to the traditional method of CCEs based on random stratified sampling, remote sensing and other technological methods give accurate and timely estimates. This makes the process of sample selection for CCE a lot more scientific and less arbitrary. SmartRisk®, an AI- and ML-powered digital platform by CropIn uses ground-level data and satellite imagery to identify the plots that are apt for these experiments. A dedicated and highly skilled data science team analyses millions of data points and runs them through numerous criteria to zero in on the farm plots to provide the most accurate sample for the region. With the help of this data, authorities can easily identify the right plots to be included in the study, removing all ambiguity from the selection process. The benefits of technologically-aided CCE are not just limited to choosing the right sample for the study. Apart from offering a more optimized method of plot selection, many of these products also help in providing the government and insurance companies with scientific, scalable, and accurate reports for future processing.

Impact on the stakeholders

The use of technology in agriculture has a far-reaching impact on agribusinesses and enables more efficient and accurate decision-making throughout the cultivation cycle. The smart approach towards CCE is beneficial on multiple levels:

- ***Government bodies***: The use of technology helps conducting a large number of CCEs with limited manpower during a short harvesting window. Using a digital platform such as CropIn's reduces the paperwork and subsequent possibilities of human error. Additionally, CropIn provides trained field managers to oversee the process and gather the required data, thus reducing the government's burden. With the help of these scientific methods, the government can improve overall efficiency by utilizing its resources in the best possible way.
- ***Insurance Companies***: The data provides a more accurate report of the crop in question and allows for the timely settlement of claims in a more just manner. The reports derived using CropIn's platform are data-driven and accurate, hence removing the possibilities of fraudulent claims or inaccurate payment disbursals. It also allows insurance companies to customize crop insurance schemes and products based on real-time data gathered from each individual farm.
- ***Farmers***: Digitization of the CCE process enables impartial settlement of claims, which is of immense benefit for farmers. It cuts down the stress on the farmer to provide proof of his/her claims, thereby also reducing the effort and time spent in the process.

2.2. Methods followed by the insurance Agencies /Start-Ups involved with PMFBY

In general, crop insurance models are of two categories: (i) farm-specific *claim-based* insurance products where payouts are based on claimed crop yield losses on *individual farms* in a region, and (ii) area/region-specific (village, district, region) *index-based* insurance products where the rates of payouts are based on a measured independent index that is strongly correlated to farm-level losses in the region. The index can be defined based on output (like average area yield at the end of the season) or input (weather variables correlated to the area yield, such as average rainfall in the area during the growing season). In index insurance, if an individual farmer suffers a yield loss when the area (on average) does not, the farmer's loss will not be indemnified. Thus, in contrast to claim-based insurance that protects individual farmers from losses, index-based insurance protects farmer communities against large-scale shocks.

When numbers of small-holder farmers are large, the claim-based model is not viable because of the high cost of crop loss verification, adverse selection (only farmers most at risk may buy insurance), and moral hazard (individual farmers influence loss assessments in their favor). In Index insurance, *all farmers in the region receive the* *same rate of payout* when the area-index triggers, irrespective of what actually happens on individual farms. Therefore, index insurance has advantages of low monitoring costs (crop losses need not be verified on individual farms), quick uniform payouts, no risk of adverse selection, and (in principle) limits the scope for moral hazard. However, if the area index used in the insurance model is not well-designed, it does not offer risk-reduction value to individual farmers in the area. It can undermine their confidence in investing in crop insurance.

The *PMFBY is an Area-Based Yield Index* insurance model for universal, comprehensive cover to farming communities against crop failure (yield losses) due to multiple non-preventable risks - climate anomalies, prevented sowing, pests, post-harvest losses, and localized natural calamities like hail, floods, drought, fire, inundation, etc. The level of coverage is crop and region-specific. The region covered by the area index is called an *Insurance Unit* (IU). All the insured farmers of a given crop in an IU are covered by the same area index and receive the same rate of payouts when the index triggers. Production conditions for the insured crop are assumed homogeneous in the IU (despite spatial variations among individual farms in soils, water, pests, resources, and management practices). The insurance covers the period from pre-sowing to post-harvest stages of the crop. *It does not cover market risk (price loss) and losses due to inadequate crop management or theft*. PMFBY was operationalized in 2016 and covers all food crops, oilseeds, and commercial and horticultural crops grown in both Kharif and rabi seasons. The scheme is implemented jointly by Central and State Governments and private and public insurance companies.

i. Operational Context (MOAFW, 2020)

For specified crops and seasons (kharif, rabi) notified by respective State Governments:

- **Insurance Unit = Gram Panchayat (GP),** defined by Govt (~ 240000 GPs in India)
- **Insurance Premium** = Lowest bid (L1), determined by respective State Governments after an annual public bidding process among insurance companies. The bids are made separately for individual districts or *for regional clusters of districts* with similar agroclimatic risk profiles for each crop and season. The validity of successful bids has been extended to three years from 2020 to incentivize long-term investments in supporting infrastructure by insurance companies and bring uniformity in premium rates.
- **Sum Insured (per ha)** = States can choose between the two alternatives: Scale of Finance or Average Notional Value of crop.
    - Scale of Finance or the finance required (cost of cultivation + some profit) per unit cultivated area in a district. It is determined at the district level every season by District Level Technical Committees of Banks to give credit to farmers and notify respective State Governments before every season.
    - Notional Average Value of Crop = (Notional Average Yield x MSP). (Farm Gate Price can be used in place of MSP when MSP for the crop is not declared)
- **Average yield** = Historical average crop yield in the IU/GP. It is determined as the average yield of *best 5 years in the immediate past seven years* (notified for each IU/GP by respective State Governments in call for bids by insurance companies)
- **Threshold Yield = (Average yield) x (Indemnity level for the crop in GP)**
- **Indemnity level** = *One of three levels of indemnity* based on crop production risk in the IU: (1) low risk, 90%; (2) moderate risk, 80%; and (3) high risk, 70%. This means that farmers will have to bear 10%, 20%, and 30% of the loss by themselves

in low risk, moderate risk, and high-risk areas, respectively. Indemnity level is notified for each crop and season by respective State Governments at the IU level, based on historical yields of the crop in the district/district cluster and their observed variability.

- **Premium Subsidy**: The PMFBY provides direct subsidies for insurance premiums. Farmers pay only 2% of the premium for kharif crops, 1.5% for rabi crops, and 5% for commercial crops. The balance premium is paid directly to the insurance companies by Central and respective State Governments, 50% by each).
- **Actual Yield** = Average yield of the crop at harvest in the current season for the entire IU/GP. It is estimated from Crop Cutting Experiments (CCE) at randomly sampled locations in the IU/GP, selected by a statistical sampling process. CCE is carried out under Govt. authorities supervision: 4 CCE/IU for major crops; 2 CCE/IU for non-major crops, 8 CCE/IU for commercial crops.
- **Yield Loss** = [(Threshold Yield) – (Actual Yield)]
- **Insurance claim payout/ha**= [(Yield Loss) / (Average Yield)] x Sum Insured
- **Payment of claims to farmers:** Transferred digitally to farmer accounts within three weeks from receipt of CCE crop yield data by the insurance company. If Threshold Yield is lower than the Actual yield in an IU/GP, no farmer in the IU will receive a payout.

The respective State Governments need to include in the call for bids for insurance premium *at the beginning of crop year* the relevant information on Average Yield, Indemnity Level, Threshold Yield, and Sum Insured for each IU/GP in the districts cluster. This information and other relevant data for the IU (farmer data, sown area, coverage, and Claims) are also updated on the PMFBY National Crop Insurance Portal (NCIP) for the purpose of claims calculation and payment to farmers in a transparent way.

In addition to this base insurance cover, the PMFBY also provides for *four add-on* covers:

1. Prevented sowing/failed germination due to prolonged adverse early-season weather (up to 25% of the sum insured),
2. Mid-season adversity (flood or prolonged drought) resulting in yield losses of > 50% of normal yield (up to 25% of sum insured is covered),
3. Crop loss due to localized calamity (hailstorm, landslide, inundation, cloudburst, and natural fire due to lightning affecting part of a notified plot), and
4. Post-harvest loss cover up to two weeks after harvest against unseasonal rain.

Note that for 3 and 4, the IU is the affected field of the insured farmer.

ii. Issues related to Area-Index Design and Measurement

It is important to note that three of the four factors that trigger the area-index for insurance payouts in an IU/GP, namely Average Yield, Threshold Yield, and Sum Insured, are based on historical data. They are explicitly specified apriori in the call for bids for premiums from insurance companies. Only the fourth factor, Actual Yield, is determined for the current season at crop harvest by taking the average of 4 CCE at randomly sampled locations within the IU/GP. Actual Yield is considered to represent the crop yield value for all the insured farmers in the IU for payout purposes. Thus, the single most important component of the PMFBY insurance model is the fair and rapid settlement of crop insurance claims.

<u>In practice,</u> the CCE sampling process in the PMFBY is prone to several delays, biases, errors, and conflicts in claims settlements, and moral hazards, leading to questions about its credibility and reliability. These arise from the following:

1. A large number of diverse IUs/GPs in India (~250000) make it physically impossible and prohibitively expensive to carry out the requisite number of CCEs (4 CCE/IU) per insured crop in the limited time window available before harvest. This has led to delays and even non-settlement of claims. (Note that conventional CCE used for crop production estimation annually is carried out at Block level, and crop information of previous years is used for sampling CCE locations).
2. *Like all index-based insurance models, PMFBY also suffers from two types of intrinsic risks,* ***basis risk, and systemic risk*****.** Basis risk affects *individual* insured farmers (idiosyncratic risk), while systemic risk affects insurance companies and markets**.**
   i. *<u>Basis risk for farmers</u>* refers to the potential mismatch between insurance payouts based on the area yield index and actual losses experienced on individual farms. It *arises from the underlying assumption* of agro-climatic homogeneity over the IU/GP. A uniform Actual Yield of a crop in the current season is estimated from CCE. But in reality, such homogeneity of cropping conditions is rare, as natural resources, management practices, pests, costs, and other factors vary spatially, leading to a wide range of yield variability among individual farms even within the IU/GP. Basis risk, therefore, affects individual farmers differentially. So, farmers in the IU/GP can experience yield losses on their individual farm but not receive a payout (false negatives) as the area-yield index is not triggered by the CCE measured Actual Yield value (Actual Yield in IU from limited CCE >= Threshold Yield for the IU, Type 1 error; Nagendra et al., 2020). Basis risk can therefore make *some farmers worse off than without crop insurance*. Other farmers in the same IU/GP who have not suffered any loss may receive a payout (false positives; Type II error). Such errors can lead to conflicts in claims settlement and encourage some farmers to influence CCE locations and measurements (moral hazard) to lower the estimated yield to become eligible for higher insurance payouts. Insurance companies would have a similar vested interest in raising the estimated Actual Yield to withhold payouts. *Thus, if the positive correlation between individual farm yield and Actual Yield in the IU is high, the lower is the basis risk.*

   ii. *<u>Systemic risk for insurance companies</u>* arises when insurance companies cannot diversify their risk over a large number of farms. Such situations arise after widespread disaster events like droughts, floods, or cyclones when yield losses among farms over wide regions can be highly correlated (the converse of the conditions for basis risk). Large numbers of insurance payouts will need to be made by the companies as large numbers of farms suffer significant yield losses. Given the high start-up costs of insurance companies, such large payouts can lead to a crash of crop insurance markets. This is why private insurers have generally been reluctant to engage in crop insurance without Government support. The PMFBY encourages insurance companies to cover this risk through reinsurance and supports their costs by paying insurance premium subsidies (>80% of premium on average) to them directly. Further, in case of catastrophic losses (claims to premium ratio >1:3.5 or percentage of claims to Sum Insured > 35%, whichever is higher at the national level in a crop season), the Government (Central + State) steps in as the reinsurer, to provide insurance cover to the companies (MOAFW, 2020).

Thus at present, in the PMFBY, systemic risk is largely covered by the State through direct payment of premium subsidies to companies and providing for their reinsurance during catastrophic losses to stabilize insurance markets. On the other hand, PMFBY *has no specific provisions to protect farmers from basis risk*. It is just assumed that, by prescribing the IU at GP/Village level, the area of the IU would be sufficiently small for homogeneity to hold among individual farms so that basis risk is minimized. However, even at the village level, agroclimatic conditions and management practices can vary greatly among individual farms. As a result, individual farmers can suffer negative impacts on incomes and livelihoods despite insuring their crops. Such losses can lead to a loss of confidence and acceptance among farmers in crop insurance. This can be disastrous not only for farmers but also for the insurance industry and the government, which invested heavily in the scheme. For successful implementation of PMFBY, the challenge therefore is to keep basis risk at non-significant levels while protecting against large-scale community-wide losses and retaining the operational advantages of the area index.

2.3. Methods used by IASRI for crop yield forecast (Traditional GCES and Technology-based interventions for crop yield estimation)

The subjective methods of estimating crop yield include farmers' assessments, expert opinions, and crop cards, while the objective methods include whole-plot harvesting. These subjective methods suffer from certain limitations in terms of the reliability of the data on crop yield. Although objective methods of measuring yield such as the whole-plot harvesting method are expected to provide reliable estimates, but it is costly and time-consuming.

Crop Cutting Experiment (CCE) technique was developed in India as a method for estimating crop yield based on a sampling of small subplots within cultivated fields by pioneers in sampling and survey design, especially Dr. P.V. Sukhatme of ICAR-Indian Agricultural Statistics Research Institute, New Delhi, and Prof. P.C. Mahalanobis of Indian Statistical Institute (ISI), Kolkata independently. Within a decade, this technique was quickly adopted and globally accepted by the Food and Agriculture Organization of the United Nations (FAO) as a Gold Standard for crop yield estimation in place of the Whole Field Harvest technique. Therefore, now any technique of crop yield estimation is to be compared with the Crop Cutting Experiment Technique in place of the Whole Field Harvest technique as it is regarded as a reliable and objective method for estimating crop yield.

The Directorate of Economics and Statistics (DES), Ministry of Agriculture and Farmers Welfare, Govt. of India adopted the crop-cut method to estimate crop yield of major crops under the General Crop Estimation Surveys (GCES) Scheme. This method is widely adopted by many African and Latin American countries also.

In India, the agricultural statistics system generates reliable yield estimates at the district level, state level, and national level through the General Crop Estimation Surveys (GCES) Scheme. The sampling design adopted for the GCES is stratified multi-stage random sampling with tehsils/taluks/revenue inspector circles/blocks/mandals, etc. as strata and revenue village within a stratum as first stage unit of sampling, survey number/field within each selected village as sampling unit at the second stage and experimental plot of a specified shape and size as the ultimate unit of sampling. In each selected primary unit, two survey numbers/fields growing the experimental crop are selected to conduct crop-cutting experiments. However, in Dadra and Nagar Haveli, three fields are selected instead of two.

Generally, 80-120 experiments are conducted for a crop in a major district where a district is considered as major for a given crop if the area under the crop in the district exceeds 80,000 hectares or lies between 40,000 and 80,000 hectares but exceeds the average area per district in the State. Otherwise, the district is considered a minor for a given crop. Experiments in minor districts are so adjusted that the precision of the estimates is fairly high, and the workload on the field staff is manageable. On average, about 44 or 46 experiments are planned in a minor district. The number of experiments allotted to a district is distributed among the strata within the district roughly in proportion to the area under the crop in the stratum. Generally, the crop cutting is done in a plot of size 5m x 5m size for most of the states' crops. However, in Uttar Pradesh, the shape of the plot is of an equilateral triangle of size 10 meters, and in West Bengal, a circular plot of radius 1.745 meters is taken for crop cut.

The average yield is obtained after harvesting, threshing, weighing, and recording the weight of the produce from the selected plots. In a sub-sample of experiments, further processing of the harvested produce is done to determine the percentage recovery of dried grains or the marketable grain of the produce, depending on the nature of the crop.

In case of three non-land record states i.e., Kerala, Orissa, and West Bengal, for yield estimation, the crop cutting experiments are planned in a sub-sample of the primary units selected for the purpose of area enumeration using a sample survey approach. The general procedure of selecting sampling units remains the same at different stages as in that of other states.

In India, the Indian Council of Agricultural Research (ICAR) and the Indian Space Research Organization (ISRO) jointly conducted the first multi-spectral air-born study for the identification of root-wilt disease in coconut in 1969. The country-level studies related to applications of remote sensing technologies were initiated after the launch of the IRS-IA satellite. Crop Acreage and Production Estimation (CAPE) was one of the important projects in this direction for estimating crop area under wheat, rice, cotton, groundnut, sorghum & mustered. Apart from these national-level projects, a number of small studies have been carried out to develop methodologies for the application of satellite data in various fields of agricultural and rural development by the Department of Space (Dadhwal et al. 1985,1991).

Several methodological studies related to crop area and production estimation have been carried out at Indian Agricultural Statistics Research Institute (IASRI), New Delhi. Singh et al. (1992) used satellite data to stratified crop area for the general crop estimation surveys and obtained a more precise crop yield. Under this study, stratification was done based on Normalized Difference Vegetation Index (NDVI) and Ratio Vegetation Index (RVI). Random sampling was done within each stratum, which is now named smart sampling estimation crop yield in the case of crop insurance.

Singh et al. (1999) also developed a small area estimator of crop yield at the block level. Singh et al. (2002) used satellite data and the farmer's eye estimate for developing a reliable crop yield model. The application of remote sensing and GIS technology for the estimation of land use statistics using spatial models has been explored by Rai et al. (2004). A project entitled “Forecasting Agricultural Output Using Space, Agro-metrology and Land-based Observations (FASAL) was undertaken under the National Crop Forecasting Center (NCFC) of the then Ministry of Agriculture, Govt. of India to meet the requirements of timely nation-wide and multi-crop reliable forecast. A project was also taken up jointly by IASRI, New Delhi, Space Application Center (SAC), Ahmedabad and North-Eastern Space Application Centre (NESAC), Shillong, with the support of the Directorate of Economic and Statistic of Meghalaya State to explore the possibility of estimation of area and production of field crops by integration of remote sensing technology, GIS and field survey. The result of all these studies was very encouraging and indicated that in the future, remote sensing and GIS have a great potential tool to improve the quality of area and production statistics of the country.

With the introduction of Pradhan Mantri Fasal Bima Yojana (PMFBY) on 18 February 2016, which is a yield-based Scheme, the number of CCEs has increased many folds i.e., about one crore CCEs at the national level for crop yield estimation from the existing about 13.00 lakh CCEs for crop yield estimation under GCES scheme which has not only become unmanageable at states level with the diminishing manpower but also results in a significant increase in non-sampling errors.

ICAR-IASRI conducted a study in 2018-19 funded by MoAFW, Govt. of India to reduce the number of CCEs significantly using advanced technologies i.e., remote sensing, GIS, sample survey techniques, spatial interpolation technique, etc. Under this study, an integrated sampling methodology for crop yield Estimation was developed using Remote Sensing, Field Surveys, and Geostatistics for Crop Insurance. An attempt was made to estimate the average yield at the Gram Panchayat (GP) level under this study. The study showed that the number of Crop Cutting Experiments (CCE) could be reduced significantly using the developed methodology.

The methodology developed by IASRI has to be validated at the field level before it can be implemented. In addition to tools like remote sensing, geostatistics, and crop yield modeling using satellite data, weather data, and CCE yield data, etc., it would require further R&D efforts to include the use of drones, advanced sample survey techniques mentioned below:

- Deep stratification
- Geographically Weighted Regression (GWR) in survey sampling
- Calibration approach in survey sampling
- Machine Learning approaches in survey sampling etc.
- Small area estimation

Such a study proposal has been submitted to MoAFW, GoI for funding for non-cereal crops for Kharif 2021 and Rabi 2021-22.

2.4. Constraints of the methods and need for improvements

The Govt. of India using traditional based methods to estimate crop yields. Previously they were used subjective methods; however, these methods suffer from certain limitations in terms of the reliability of the data on crop yield. Although the objective methods of measuring yield, such as the whole-plot harvesting method, are expected to provide reliable estimates, but it is significantly time-consuming and cost-effective. For that, the CCEs plan was proposed all over India at district/ Gram panchayat levels, particularly estimating cereal crops. They have successfully launched in several districts of India to estimate yields to a small scale. However, they have certain limitations including:

- Representativeness of the sampling plan.
- Time-consuming, laborious, and cost-intensive.
- Synchronize the crop maturity/harvesting within conducted CCEs.
- Measurement errors in terms of area, weighing of products, and variations in moisture content.

The critical challenges of agriculture insurance policies faced in India are looking for an optimized solution for estimating yields. The advent of new technologies available to improve yield estimates is discussed in the following chapters. There are specific challenges, and possible suggestions are formulated below.

- More representative sampling plan to reduce the sampling error.
- Reduce the number of CCEs without compromising the accuracy.
- Estimate crop conditions and yield directly.

In recent studies, statistical or processed-based models have become widely used to estimate yields from small-scale to large-scale areas. The data required for these models increases exponentially with an increased crop area size, limiting the scalability of crop yield estimation. The model that includes more processing steps produces uncertainty and the possibility of errors. Combining the crop models with technology-based data is one possible solution for daily monitoring, observing, and simulate yields accurately. The reduced number of CCEs is a challenging task, mainly when the target is to reach accuracy. This study is for exploring other alternative options that estimate crop conditions and yields directly at village-level regions.

3. Advanced global methodologies for crop yield estimation

3.1. Review of prominent global technologies for yield estimation

In the past decade, advances in high-resolution satellite imagery and technology, drone imagery, cloud computing, and digital connectivity through mobiles, as well as advances in crop models, statistical analysis, machine learning, and AI-driven analytics, have been used increasingly in high-resolution crop health monitoring and yield assessment, and also agricultural insurance. However, it is important to evaluate the capabilities of these technologies in mirroring the real-time crop yield risk distribution among individual farms at the IU/GP level to understand and minimize basis risk in PMFBY.

Keeping this in view, the Ministry of Agriculture and Farmers' Welfare commissioned a number of pilot studies in 2019, 2020, and also 2021 involving government agencies, Agritech Start-ups, and Insurance Companies as partners to explore the feasibility of adapting emerging imaging technologies, modeling, and data analytics for expediting crop yield assessment by rationally reducing the numbers of CCE to manageable levels, and also by directly estimating area-yield at IU/GP level.

This is to be done by: (i) combining smart sampling[1] approaches with complementary data and information from alternate sources of crop health indicators that capture multiple adverse risks like dry spells, drought, pest attacks, etc., without impacting the quality of field sampling for Actual Yield estimation, and also to (ii) directly estimate Actual Yield at IU/GP level using new technologies. Towards this end, all contract agencies were mandated to adopt a consistent two-step yield estimation approach:

1. Derive a scientifically designed objective *smart sampling strategy* to optimize numbers and field locations of CCE *to better reflect the actual crop yield distribution in the IU/GP than random sampling*, and conduct CCE in a transparent manner (using digital photo/video recording on mobiles) to eliminate human bias and moral hazard.
2. Develop a scientific protocol to reduce the number of CCE to manageable levels by (i) developing high-resolution pixel-level yield proxy maps (weather indices, vegetation indices) at *district level*; (ii) classifying the District level maps into *four classes of equal frequency at the Block level, categorized as normal, mild, moderate and severe*, each with relatively homogeneous yield/yield proxy index distribution, (iii) identifying CCE locations at block level among blocks classified as 'normal' or 'mild', and (iv) identifying CCE location in all IU/GP in Blocks classified as 'moderate' or 'severe' as per (1) above.

By this combined approach, CCE locations can be identified in IU/GP clusters within districts (instead of in individual GPs) with more CCE locations in 'moderate' and 'severe' clusters, and fewer CCE locations in 'normal' and 'mild' clusters. Such an approach is expected to better reflect the crop yield distribution among farms at the GP level. In the short run, attention will be more on (1) above at GP level to support PMFBY in current and near seasons. But, in the long run, the combined two-step approach of (1) & (2) above will be standardized and used to limit the number of CSE.

The MNCFC developed the following protocol in 2019 for 1 and 2 above:

1. Develop a **crop map** at high spatial resolution (10m-30m) at Block/District level using high resolution (10m-30m) optical/SAR data from multiple satellites and ground truth data. Mapping accuracy must be > 80% to capture the crop distribution at IU/GP level.
2. Develop a map of **crop yield proxy for the selected** using the crop map as a mask (based on multiple satellite data derived time series of indices, up to 20 days before harvest) to capture the spatial variability of yield at IU/GP level.
3. Select CCE fields by (i) stratifying the yield proxy map into 4 quantiles per Block (6 for the district); (ii) Overlay the IU/GP map on the yield proxy map and select 2 or 3 major classes for (low, moderate or high yield), and (iv) Select CCE locations randomly in each stratum (4 CCE locations per IU/GP cluster and 3 points (2 as reserve) for each CCE), (v) overlay the CCE points on cadastral maps to obtain corresponding field locations for CCE and (vi) communicate to State agriculture departments, in real-time, for conducting CCEs.

The protocol is being implemented since Kharif 2019 and is continuing on different crops and States-Districts-Blocks-IU/GP. By this two-step method-based protocol, the

[1] Smart sampling is a stratified random sampling method where a crop yield proxy variable is used to stratify the block or village into different strata of high to low yield. CCE locations are then selected randomly and proportionately from each stratum.

number of selected CCE in a GP ranges between 2 to 4, depending on the number of major clusters in the GP. Further, the GP level yield is not a simple average of CCE yield but an area-weighted average of the CCEs. The following are some of the observations on the protocol.

1. None of the pilot studies used Indian satellite (RESOURCESAT) data.
2. The pilot studies were essentially exploratory, as organizations chose the crops, regions, and technologies according to their expertise and convenience. This is reflected in the arbitrary choice of remote sensing data sources (satellite data, UAV, IoT, digital photographs), and analytics tools (statistical models, crop models, machine learning algorithms), as well as measures of yield estimation accuracy.
3. One important outcome of the pilot studies is the yield data from over 20,000 CCE supervised by MNCFC, across 15 States, 64 districts, and 9 crops, in one season. More such data would become available with the new set of pilot studies during 2020-21.
4. Information is not available on whether results of CCE conducted under the pilot studies were favorably received and used by the Insurance Companies in the GPs where they were available to decide insurance payouts to farmers. There is a need to take the pilot studies to their logical conclusion by evaluating their uptake by the Insurance Companies and conducting field surveys to examine if they were effective in minimizing basis risk for the farmers (that is, if they reduced the number of false negatives). This can be done post factor for 2019, 2020 and included in surveys for proposed pilot studies in 2021.
5. The use of high-resolution multi-temporal and multi-source satellite data requires adherence to rigorous standards of pre-processing to ensure consistency over time and spatial alignment. Several organizations involved in the pilot studies are private agritech start-ups, for which all data, models, and ML processing algorithms are proprietary and subject to rigorous IP protection. The various stages of RS data pre-processing and the assumptions and codes of the models would be difficult to verify their authenticity independently to ensure that necessary standards were observed in data management and interpreting the results.
6. Among the relatively more preferred technologies used in the pilot studies are MODIS, Landsat, and Sentinel for satellite data; ANN and Regression among machine learning algorithms, and gross primary productivity estimation from satellite data for crop models for yield estimation (because of direct availability of relevant data as satellite data products and a limited number of model parameters).
7. In general, the range of commonly reported measures of accuracy of estimated yields at GP level in the pilot studies is R^2 (0.1 to 0.8) and MAE & RMSE (10 to 70%). At this level of error, even with smartly sampled CCE, the requirements yield accuracy and consistency for crop insurance will not be satisfied.
8. Though UAVs were used in some studies, the extent of their significance in yield assessment in the pilot studies is difficult to ascertain.
9. A point of significance in the pilot studies on innovative technologies and in the PMFBY insurance model, in general, is the exclusive attention on the accuracy of actual yield in the current season. Additional key input for the index insurance model to decide payouts is the threshold yield (based on average historical yield in GP). *Threshold yield, too, is subject to the same limitations of accuracy as the actual yield from CCE*. An equally rigorous assessment is required for threshold yield. This can be possible with long-term time series of high-resolution satellite data and crop data.

In general:

- For crop insurance payouts, yield loss estimations are needed more accurately under adverse conditions. Crop models tend to predict yields well in normal weather conditions compared to extreme weather conditions. Only the more complex portfolios of crop models like DSSAT accommodate multiple crop stresses, and that too for a few crops. They also need reliable estimates of many additional model parameters for estimating stress effects on yields. But present crop models can play a significant role in insurance in the preliminary stages of insurance planning for risk assessment and index insurance model parameters, like threshold yield, premium, and sum assured at GP level.
- Applying machine learning techniques to remote sensing data can capture about 70–80% of the yield variability ($R^2 \approx 0.7$–0.8) with 10–15% errors in the reported yield. This level of accuracy is feasible in data-rich homogeneous areas with consistent crop management. Prediction errors can be more for small-holder areas with diverse farming conditions[2]. Some of the predictions in the pilot studies listed above, particularly for paddy, are in the above range of error. Paddy fields tend to be relatively more homogeneous, but fields of other crops are more heterogeneous, and the prediction errors can be larger.
- Application of machine learning and AI, aided by scientific insights from crop models, is the likely way forward in index-based insurance, as RS data and analytics tools progressively improve in crop condition assessment and yield predictions, eventually minimizing the need for CCE. But machine learning prediction algorithms tend to become more effective as more training data is used (than the limited current season CCE crop yields used in the pilot studies). This is because machine learning algorithms 'learn' and get better with more data and insights from additional data. A large and diverse pool of carefully conducted CCEs will be required in the initial several years across the entire range of crops to make adequate authentic training data sets available for the machine learning and AI models. *A highly systematic approach* will be required to *archive and share long time series of both satellite data and crop yields data from CCE and other sources*.

On the whole, the pilot studies provide valuable insights into national capacities and limitations in testing and scaling high-resolution satellite technologies, crop models, and machine learning and AI for crop yield estimation at the GP level. For supporting field-level decisions on insurance in complex landscapes, deeper, more systematic, and rigorous approaches subject to verifiable standards of rigor in methods used are needed to gain confidence that they can accurately capture yield losses under adverse conditions when farmers suffer the most. Strategically improving the collection of reliable ground reference data on crop types and yields with CCEs and other means, use of time series high-resolution satellite data, evaluating prediction accuracies with machine learning algorithms and models, and archiving relevant data systematically in a freely accessible form would facilitate this task.

[2] Benami, E., Jin, Z., Carter, M.R. *et al.* Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nat Rev Earth Environ* **2,** 140–159 (2021)

3.2. Identifying the technologies potentially suitable for India

Agricultural monitoring on a regional and national level based on remote sensing, weather, and other indicators has been in place for decades. There are currently eight main global and regional scale agricultural monitoring systems in operation. These global agricultural monitoring systems provide near real-time information on crop conditions for food security, agricultural markets, and early warnings on disasters. The major such systems include: (i) FAO’s Global Information and Early Warning System (GIEWS), (ii) Famine Early Warning Systems Network (FEWSNET) of USAID; (iii) Crop Watch of China; (iv) Monitoring Agriculture with Remote Sensing (MARS) system of European Commission (EC); (v) United States Department of Agriculture's Foreign Agricultural Service (USDA-FAS) monitoring system; (vi) GEOGLAM (Group on Earth Observations (GEO) Global Agricultural Monitoring), system; (vii) World Food Program Seasonal Monitor; and (viii) the Anomaly Hot Spots of Agricultural Production (ASAP) system of European Commission. USDA-FAS was the first system to provide globally comprehensive information on crop production and crop condition. The Group on Earth Observations (GEO) recognized that national and regional monitoring systems cannot effectively monitor agriculture at all scales and launched GEOGLAM to coordinate information sharing across all the agricultural monitoring systems. These systems are publicly accessible and provide information on crop conditions through cloud-based platforms and services. All the systems use meteorological data and remote sensing information, and many of the system's meteorological data are also derived from remote sensing.

Table 1
Data and model inputs used by each monitoring system obtained from the questionnaire. Check marks or text indicate that the systems use the data or model inputs while a dash indicates non-usage or non-applicability. The following acronyms are used to indicate meteorological sources: AM, gridded data from Atmospheric Model; RS, gridded data estimated by a RS-based model; I, gridded data interpolated from meteorological ground station data.

Data and model inputs			GIEWS	FEWS NET	MCYFS	CropWatch	USDA-FAS	GEOGLAM	Seasonal Monitor	ASAP
Meteorological data source used	Precipitation		RS	RS	I	RS	RS	RS	RS	AM
	Temperature		RS	RS	I	RS	RS	RS	–	AM
	Evapotranspiration		–	RS	I	RS	RS	RS	–	AM
	Solar radiation		–	RS	I	RS	RS	–	–	–
	Relative humidity		–	RS	I	–	–	–	–	–
	Wind speed		–	RS	I	–	–	–	–	–
	Snow coverage		RS	–	I	–	RS	–	–	–
	Total cloud cover		–	–	I	–	–	–	–	–
	Water vapor pressure		–	–	I	–	–	–	–	–
	Atmospheric pressure		–	RS	–	–	–	–	–	–
Remote sensing	Products	Vegetation indices (e.g. NDVI, VHI, fAPAR)	✓	✓	✓	✓	✓	✓	✓	✓
		Soil moisture, FAO-ASIS	✓	✓	–	–	✓	✓	–	–
	Sensors	Passive Radar (10–50 km) (e.g. SMOS, SMAP, SSM/I, TMI)	–	✓	–	✓	✓	✓	–	–
		Active radar (20 m – 50 km) (e.g.ASCAT, JASON, Sentinel 1)	–	–	–	✓	✓	–	–	✓
		Geostationary (5 km, e.g. FY2)	–	–	–	✓	–	–	–	–
		Coarse resolution optical (250 m – 1 km) (e.g. AVHRR, MODIS, Proba-V, FengYun-3)	–	✓	✓	✓	✓	✓	✓	✓
		Very high to high resolution optical (80 cm – 30 m) (e.g. Landsat, Sentinel 2, Gaofen1&2, ZY3)	–	✓	–	✓	✓	–	✓	✓
Crop models used	Water balance models (e.g. GWSI, WRSI)		✓	✓	–	–	✓	✓	–	✓
	Biophysical/simulation models (e.g.WOFOST, Wheat Ritchie (CERES), Sorghum Vanderlip and Reeves, GDD, Corn Hanway, FAO-ASIS, WARM for rice, etc.)		✓	–	✓	–	✓	–	–	–
	NDVI models		✓	–	✓	✓	✓	–	–	✓
	Statistical and bespoke models (e.g. FAO-ASIS)		✓	–	✓	✓	✓	–	–	–
Auxiliary data used	Cropland maps		✓	✓	✓	✓	✓	–	✓	✓
	Crop type		✓	–	✓	✓	✓	✓	–	✓
	Crop calendar		✓	✓	✓	✓	✓	✓	✓	✓
	Soil information		✓	–	✓	✓	✓	–	–	✓
	European Media Monitor outputs		–	–	✓	–	–	–	–	✓
	Agricultural Census		✓	–	✓	–	✓	✓	–	–
	Agricultural Surveys		–	✓	–	✓	✓	–	–	–
	Small area crop statistics		✓	–	✓	✓	✓	–	–	–
	Agricultural contacts		–	–	–	✓	✓	✓	–	–
	DEM		–	–	✓	✓	✓	✓	–	–
	Climate or agroecological zones		✓	✓	–	✓	–	–	–	–
	Surface water availability		✓	–	–	–	✓	–	–	–
	Commodity prices		✓	–	–	–	✓	–	–	–
	Soil map		✓	–	✓	✓	✓	–	–	✓
	Admin borders		✓	–	✓	✓	✓	✓	✓	✓
	Livelihood zones		✓	✓	–	–	–	–	–	–
	Windshield survey (field observation)		✓	✓	✓	✓	✓	✓	✓	✓
Forecasts from the system	Crop conditions		✓	–	✓	✓	✓	✓	–	✓
	Agro-climate/potential Biomass		–	–	–	✓	–	–	–	–
	Cropland utilization		✓	–	✓	✓	✓	–	–	–
	Cropping intensity		✓	✓	–	✓	✓	–	–	–
	Crop area		✓	–	–	✓	✓	–	–	–
	Crop yield		✓	✓	–	✓	✓	–	–	–
	Crop production anomalies		–	–	–	✓	✓	–	✓	✓
	Crop area affected by critical anomalies		✓	–	–	✓	✓	✓	–	✓
	Crop stage		✓	–	–	–	✓	–	–	–
	Season start		✓	–	–	–	✓	–	–	–
	Number of dry days		–	–	–	–	✓	–	–	–
	NDVI values		✓	–	–	✓	✓	–	–	–

AET = actual evapotranspiration; PET = potential evapotranspiration; ET = evapotranspiration; RH = relative humidity; DEM = digital elevation model; NDVI = normalized difference vegetation index; ASI = Agricultural Stress Index; GDD = growing degree days, VCI = vegetation condition index, TCI = temperature condition index, VHI = vegetation health index, RFE = rainfall estimate, SPAM = spatial allocation model, WRSI = water requirement satisfaction index.

3.3. Comparative analysis of the global technologies vis-à-vis technologies used by the pilot studies conducted by DAC&FW

The pilot studies on using innovative technologies to estimate actual crop yield at (GP) level at the end of the crop season were conducted by the Department of Agriculture, Cooperation and Farmers Welfare (DAC&FW) through public and private technology organizations (including several Agritech start-ups). The technologies experimented with included high Spatio-temporal remote sensing data, UAVs, crop simulation models, machine learning, artificial intelligence, etc. Each organization is leading a pilot study covered at least five districts, spread over three states in different agro-climatic zones, and three crops (including at least one multi-picking crop and one non-cereal crop). At least 10 GPs were covered in each block of every selected district. Yield data from a minimum of 5 scientifically sampled CCEs in each GP were used for developing the crop yield model. The results of the pilot studies are summarized in Table 1.

Table 1. Summary of PMFBY pilot studies at GP level on innovative use of technologies for a smart sampling of CCE points and crop yield estimation[3].

No	Organization	Technologies used		Yield estimation Accuracy (for GP) RMSE, NRMS, MAE, etc
		Remote sensing	Analytics (yield estimation, crop mapping; smart sampling)	
1.	**AMNEX**, Ahmedabad, (Agritech Start-up) Crops: paddy, cotton Crop yield estimated at GP level using physically-based regression and CNN with inputs from optical and microwave RS data	1. Optical data: Sentinel 2A optical bands (to derive FAPAR, LSWI) 2. Microwave data: Sentinel 1A (SAR) – microwave backscattering coefficient	1. Modified Monteith model with FAPAR and LSWI) 2. Linear regression models with SAR data 3. ML- Deep CNN with sentinel 2A data (cotton only)	Paddy: Model agreement (R^2): 74-85% NRMSE: 15-25% GP level accuracy range: 40-100% Cotton Model agreement (R^2): 75-80% NRMSE: 20-25% GP level accuracy range: 57-100% MAPE: 18%
2.	**Cropin**, Bangalore (Agritech Start-up) Crops: paddy, cotton, soybean Crop yield estimated at GP level based on RS data and identify Smart sampling locations for CCE	1. Sentinel 2A/B 2. Sentinel-1 (SAR) 3. Landsat 8 OLI/TIRS 4. Weather data 1 &2 for crop detection, smart sampling, and sowing calendar 3&4 for yield model	ML algorithms for crop classification Yield estimation using FAO yield model, based on growing period ET deficit (Doorenbos & Kassam)	Paddy: NRMSE: 12-16% Cotton NRMSE: 10-27% Soybean: NRMSE: 14-18%
3.	**ICRISAT** Hyderabad Crops Paddy, maize, chickpea, groundnut Crop yield estimated at GP level from crop models with inputs from RS data, weather data, and soil data	1. Sentinel 2A data for crop classification 2. Landsat 30m data for LAI estimation for models	1. Cluster algorithm (ISODATA)for crop classification and mapping 2. Crop models CERES for Rice and Maize DSSAT for Chickpea and groundnut	Rice: R^2: 0.32; 0.61 RMSE: 989; 583 kg/ha RRMSE:0.28;0.10 MAE: 722; 437 Maize: R^2: 0.57 RMSE: 755 (?) RRMSE:0.33 MAE: 568 Groundnut: R^2: 0.387 RMSE: 218 (?) RRMSE:0.18

[3] Source: MNCFC (2020) Gram Panchayat Level yield estimation using Technology under PMFBY: A summary of the Pilot Studies conducted during Kharif (2019-20).

				MAE: 178 Chickpea: R^2: 0.527 RMSE: 193 (?) RRMSE:0.19 MAE: 150
4.	IFPRI, Delhi Crops Paddy Crop yield estimated at GP level using crop models with inputs from smartphones RS data, and weather data, validated with CCE data	1. Smartphone: Multiple digital, georeferenced photographs for phenology and damage identification 2. Sentinel 2 (for LAI ?) 3. Weather- IMD 4. Weather-Gridded products	1. Crop models (APSIM) 2. Non-Linear Statistical models derived from crop model outputs 3. Statistical models adjusted for phenology and crop damage	Accuracy of statistical models evaluated against the simulated model, not observed yields
5.	LeadsConnect NOIDA Crops Paddy, Cotton Bajra Crop yield estimated at the **district level** using crop models with inputs from satellite data, weather data, and soil data and validated with CCE data	1. Satellite data for NDVI, LAI, NMDI,	1. Crop Models CASA DSSAT Statistical models (AHP-LASSO)	Yield predictions at District level (not GP) Model accuracy determined by correlation: CASA: 0.05 – 0.7 AHP-LASSO: 0.77-0.85
6.	NCML, Gurgaon Crops Paddy, cotton maize, millet, soybean, guar Crop yield estimated at the **district level** using ANN, with input data from satellites; weather variables and soil moisture, CCE	MODIS-500m, 8-day composites for LAI and FPAR MODIS- 250m NDVI 16-day composite AMSR2-10km surface soil moisture Weather data –AWS Weather data- IBM 4x4 km grid Farmer surveys CCE data	ANN	MAPE (%): Paddy: 13.7-15.5 Cotton: 12.6 – 21.5 Maize: 87.6 Millet: 19.1 Soybean:42.2 Guar: 39.1
7.	RMSI Crops paddy, cotton pigeon pea Evaluated three ML algorithms for Crop yield estimation at GP	MODIS 250m: NDVI, EVI, LAI; Landsat 30m: NDVI, EVI, LAI, CWSI Sentinel 2: CCCI	ANN, PLSR, Random Forest Multi layer perceptron (MLP),	$R^{2:}$ Paddy: 0.7-0.9; UAV: 0.11 Cotton: 0.75 Pigeon pea: 0.67

	level using ANN, with input data from satellites at two resolutions (250m and 30m) at three combinations of crop stages, weather and soil moisture; and with weather and soil moisture data only (no optical data)	Sentinel 1 SAR: soil moisture UAV: for NDVI, CCCI, CWSI IMD: Rainfall, Temperature		
8.	Skymet <u>Crops</u> paddy, cotton pigeon pea Estimated pixel level (30m) crop yield using ANN models with input data from satellites, UAV, weather data grid, and smart sampled CEE	Seninel2 (in 3 districts) and Sentinel 1(in 2 districts for NDVI, NDWI, LAI, fAPAR UAV in one area IMD/Skymet weather grid for GDD	ANN (analysis with all data resampled to 30m resolution for pixel yield) Area weighted pixel yield computed for GP	R^2 at district level (not GP level) Paddy:0.84-0.98 Cotton0.89-0.99 Ppigeon pea:0.92-0.96
9.	TRINITY <u>Crops</u> paddy, cotton maize Estimated crop yield Using ANN and linear regression with input data from satellites, UAV, soil data, weather, and CEE	Sentinel, Landsat and UAV for: NDVI, EVI, ARVI, NDWI, SAVI and NDSI	ANN for 3 scenarios (optimistic, normal, pessimistic) Linear regression	R^2 and MAE for ANN and regression models <u>($R^{2)}$</u> Paddy: 0.63-0.85 Cotton: 0.29 Maize: 0.68 <u>ANN (MAE %)</u> Paddy: 4.9-10 Cotton: 20.0 Maize: 13.0 <u>Regression(MAE %)</u> Paddy: 9-16.9 Cotton: 22.4 Maize: 11.7
10.	WRMS <u>Crops</u> paddy, cotton potato Estimated crop yield using multi-parameter regression, **own crop model**, and RNN input data from satellites, UAV, IOTs (soil moisture), crop condition photographs, farm management surveys, and AWS	Sentinel 1, 2 for NDVI, LAI, NDRE UAV at some locations	Multi-Parameter Regression analysis **In-house developed Crop model** (hybrid model integrating biophysical models with remote sensing and IoT data, and using ANN to estimate some parameters) Recurrent Neural Networks	RMSE: <u>Regression</u>: Paddy: 8% Cotton: 6% - 27% Potato: 23% <u>RNN</u>: Paddy: 10% Cotton: 9% - 30% Potato: NA <u>Crop model</u>: Paddy: 8% Cotton: 5% - 29% Potato: 17%

				With farm pictures and farming information Paddy: R^2 : 0.6; RMSE:15% Cotton: R^2 - 0.7; RMSE:15% Without farm pictures and farming information Paddy: R^2 0.4; RMSE: 20% Cotton: R^2: 0.43; RMSE :25%
11.	Niruti, Hyderabad Crop: Rice Crop yield prediction using TOPS NASA's flexible modeling system that integrates ecosystem models with satellites and surface weather data to produce ecosystem nowcasts and forecasts	MODIS and Sentinel to estimate LAI, FPAR	Yield based on gross primary productivity and HI from MODIS data Yields downscaled from MODIS to Sentinel to generate CCE sampling scheme	Paddy: R^2: 0.4 ; MAE: 20%

3.4. Prioritization of suitable technologies for India

In Europe, the European Commission's (EC) Joint Research Centre (JRC) utilizes the **MARS** (Monitoring Agricultural Resources) Crop Yield Forecasting System (MCYFS) to provide timely forecasts of crop yields for the European Union. For the sustainable management of the agricultural market of the EU, the JRC has been making in-season forecasts of expected crop yield for the major crops in the European Union (EU) Member States (MS) since 1993. The JRC has developed and runs a crop yield forecasting system since 1992 that provides timely forecasts of crop production, including biofuel crops, for Europe and other strategic areas of the world. They provide accurate in-season yield estimates performed on a monthly basis during the growing season. In contrast, to forecast crop production, forecasts of the end-of-season crop yield are multiplied by sown areas. Several interconnected software is used to perform statistical analysis and crop yield forecasts, including Crop Growth Monitoring System (CGMS) models and CoBo (the statistical Control Board) tools. The main important component of CGMS is World Food Studies (WOFOST) model. The WOFOST is an open-source model that is biophysically based, dynamic, and explanatory model that performs across a range of meteorological, soil, and agro-climatic conditions. The model potentially simulates every stage of crop growth determined by the defining factors carbon dioxide (CO_2), temperature, solar radiation, and crop characteristics. Despite its usage in simulation potential yield estimates, observed meteorological data is interpolated on a regular 25 km grid based on the distance, altitude, and climatic region. The WOFOST model runs on the intersection between the 25 km meteorological grid and the European soil map units. Therefore, model outputs are aggregated to the national level and used as decadal predictors in the statistical analysis. In order to assess the model performance and results, Velde & Nisini (2019) examined the MARS-crop yield forecasting system for the European Union from 1993 to 2015. They addressed three questions related to the MARS performance: **First**, how good has the system performed? This was investigated by calculating several accuracy indicators (i.e., the mean absolute percentage error, MAPE). **Second**, do forecasts improve during the season? This was evaluated by comparing the accuracy of the first, the pre-harvest, and the end-of-campaign forecasts. **Third**, have forecasts systematically improved year-to-year? This was quantified by testing whether linear models fitted to the median of each crop's national absolute relative forecast errors. Overall the analysis indicated that the accuracy differs from crop and MS, with the media MAPE ranging from 6.31 to 10.15% for respectively for potato and wheat. The crop yield forecasts tend to improve during the growing season, and no structural improvements were found for other crops.

GEOGLAM is the Group on Earth Observations Global Agricultural Monitoring Initiative. It is an international G20-endorsed program geared toward enhancing the use of Earth observations (EO) to strengthen decision-making, action-making, and policy in terms of food security and sustainable agriculture. The main objective of this initiative is to leverage and build upon existing models, programs using Earth observation technologies, capacity development, monitoring, and development activities. The EO data is effectively helpful for monitoring agricultural conditions globally. The datasets, which are included NDVI, precipitation, temperature, Evaporation Stress Index, and soil moisture.

4. Use of yield estimating technologies for crop insurance

4.1. Smart Sampling (Satellite-based sampling) for conducting CCEs

The increasing number of CCEs for conducting a crop yield estimation which results in time-consuming, redundancy, and rise data uncertainty issues. Since the harvest time period is very short, carrying out a number of CCEs is extremely difficult with limited manpower and time. Using advanced technologies to bring optimize the solution for minimizing CCEs of the insurance unit level (village, village panchayat, block, revenue circle, Mandal, or taluk) is essential for overall crop management practices for yield prediction. Between 2018-19, the Indian agriculture ministry conducted pilot studies to develop satellite and other advanced technologies for CCE optimization. However, the results indicated that there is a chance to reduce CCEs by 30-70% by using satellite technologies and other models while maintaining good accuracy. Consequently, the Indian agriculture ministry is intended to adopt new methodologies for covering a large number of districts, whereas pilot studies are underway for technology-based direct yield estimation.

The CCEs based on traditional approaches should immediately replace technology-based platforms, thus improving precision agriculture. The Government of India emphasizing the potential use of technologies like drones, mobile cameras with GPS, satellite-based acquisition images. The data inferred from these platforms is processed through photogrammetric software to analyze the field for optimizing CCEs to a certain extent. The use of drones in agriculture for reliability, accuracy assessment, and speed of conducting CCEs. At the same time, mobile devices for real-time transmission of CCE data with GPS, date, and time stamping. Satellite technologies are helpful for day-to-day observations of vegetation conditions, crop status, and yield estimates. Tailor-made services offer the best solutions to the farmers about soil moisture levels, adequate fertilizer application to crops, and crop identification depends on the predicted weather conditions. The importance of these technologies in precision agriculture is detailed in the following chapters.

4.2. Two-step yield estimation (conducting large CCEs only when the crop situation is moderate or severe)

1) One of the major requirements of the PMFBY is carrying out a large number of Crop Cutting Experiments (at least 4 per GP, for a major crop) for yield estimation, which is the basis for claim computation. However, conducting such a larger number of Crop Cutting Experiments (CCEs) has become a very cumbersome task, considering the short harvest period within which all CCEs have to be completed.
2) This may also result in compromise of the quality, rigor, and accuracy of the CCEs. In order to overcome this, the Government of India is trying to implement various approaches with the help of technology, such as, i) smart sampling under which CCE sites are being selected based on satellite remote sensing-based proxy yield parameters. The government is also carrying out, through 12+1 agencies, a large number of pilot studies for direct yield estimation using technology.
3) Another significant approach which the Government is proposing to adopt is 2- step yield estimation. The basic idea behind the 2-step yield estimation is to assess/categorize the crop loss incurred due to adverse climatic conditions, pest infestation, etc. based on technical parameters (remote sensing, weather, field survey, etc.) and carrying out the large number (as defined in PMFBY) of CCEs wherever the situation is ‘severe’ or ‘moderate.’ Wherever the situation is ‘mild’ or ‘normal’ based on the technical parameters, the reduced number of CCEs (as defined for a higher administrative unit) can be conducted. In other words, the required number (4 per GP) of CCEs will be carried out only in the places where there is a situation of crop loss, and a limited number of CCEs will be carried out where the crops are ‘normal.’
4) A typical example of a similar approach is followed in drought assessment where weather, satellite, and ground parameters are used to assess the drought situation, and then ground truth for loss assessment is carried out wherever the situation is ‘Moderate’ or ‘Severe.’ All States are following this procedure for drought declaration.
5) In order to implement a similar approach in the 2-step yield estimation, there is a need to identify the technology parameters separately for different climate induced disasters, pest infestation, etc, which have been notified in the PMFBY. Also there is a need to generate thresholds values of different parameters to categorize a situation into ‘severe,’ ‘moderate’, ‘mild’ or ‘normal’. The Government proposes to

form an Expert Committee to identify the categorizing parameters for different disasters and also identifying the thresholds. After these parameters and thresholds are decided, pilot studies would be carried out to verify these.

4.3. Technology-based direct yield estimation

4.3.1. Drone-based Technologies

With the advent of new geospatial technologies, Unmanned Aerial Systems (UAS), referred to as Unmanned Aerial Vehicles (UAV) that are known to be drones, play a valid role in the agriculture sector over the last two decades. Many research studies demonstrate the importance of drones in agriculture to enhance crop growth, precise monitoring, nutrient conditions, and surveillance of crop yields of large agriculture farms. They are becoming very popular and gaining more attention in the agriculture area fields to help farmers, insurance policymakers, and government agencies to make any sort of decisions towards the sustainable management of agricultural practices. Drones can cover entire fields from smaller to larger farms, which are essentially useful for inaccessible and high-altitude regions. They are associated with sophisticated technologies fitted with GPS and very high-resolution cameras, which the pilot can operate to track and control drones over the larger distances used by onboard GPS-enabled remote systems. Modern drones are integrated with Wi-Fi technologies to provide real-time video of a flight over smartphones or tablets.

Drones are manufactured for a wide range of applications, facilitated by a variety of functions. However, among the most promising areas, agriculture is regarded as the most important and challenging sector to meet farmer's required challenges to get maximum crop yield. The following are the various applications and advantages of drones in agriculture and their functions (Puri et al., 2017):

1. *Agriculture farm analysis*: Drones are packed with high-end technologies, flying over the agriculture areas to check the farm condition at the commencing of the crop of any season. They generate 3-D maps for soil analysis that are significantly useful for farmers to decide the specific time for seed plowing. The information provided by the drones is also useful for irrigation, analyze soil conditions, fertility rates, and nutrient management of fields for better crop growth.
2. *Time-saving*: Filed observations are always time-consuming, and it is more difficult during hot climates to inspect the fields from time to time to observe crop conditions, particularly with large distances of fields. Drones can cover the areas with the lesser time within regular, frequent intervals to know the status of the crop.
3. *Higher crop yield:* With the help of drones, the more care can be taken, the more yield can be estimated based on the continuous crop monitoring techniques, including the accurate application of pesticides, and water turns the higher yield.
4. *Better monitoring:* The images obtained from the drones, processed through appropriate software to derived vegetation indices such as NDVI maps to better track vegetation health, wet & dry conditions, and plant transpiration rates.

Compared with satellite technologies, the smart use of drones in agriculture is an effective approach that provides a birds-eye view of fields to the farmers (Tripicchio et al., 2015), thus provides precise information about the crop conditions.

The very high-resolution images obtained from the drones are helpful for the estimation of above-ground biomass (AGB), yield assessment, crop management, and optimization. Consequently, drones are using for spraying fertilizers in farm fields to target pests to save time and cost, thus enhancing yield production. The raw images captured by the drones are processed through the most advanced software for georeferencing and the ortho-mosaicing process. For example, the Pix4Dmapper Desktop software (https://www.pix4d.com/)

Natural disasters like storms or floods can decimate a season's profits. In order to measure the crop damage levels, the insurance loss adjuster must ensure the field scale assessment before finalizing the insurance policies. Therefore, crop damage assessments based on traditional approaches have been complex procedures where the policymaker has to visit the fields and inspect the damage areas unless it is possible for small-scale fields. It is relatively difficult for an insurance loss adjuster to identify all areas in a damaged field. Particularly larger scales and often inaccurate to measure the entire plot using field measurement techniques (i.e., manual tools) require a lot of time to estimate the area impacted. Obtaining manual measurements can be a major challenge, if not impossible. These struggles to obtain an overall assessment of the farm and ground-based measurements in poor conditions or after an extreme weather event often leads to guesswork or inaccurate reporting. Because of this, it is difficult to adjust the insurance policymakers to decide to prepare loss schemes, and sometimes it's obvious that can lead to over-or underpayment, valuation disagreements, hold-ups, and other complications.

In the US, the insurance agencies play a vital role in approving policies backed by the US government and the USDA RMA agency. These policies insure farmer's crops to a certain value of production. After an extreme weather event, the farmers should identify the crop damage areas and file a written complaint to the insurance agencies within 72 hours of problem discovery. The insurance loss adjuster makes an appointment for an on-site visit. Usually, the loss adjuster flies a drone over the flooded areas to collect the data. Further processing of the data to assess the crop loss, the insurance agencies make use of the software that is approved by the USDA RMA is "Pix4Dfields. The measurement methods of the Pix4Dfields fall under the USDA RMA methods. The loss adjuster digitizes the damaged areas in processing software according to the visuals based on the soil moisture conditions and water standing areas. In the Pix4Dfields software, the custom index tool provides the level of damage and area. The tool associated with the custom index formula: (nir-red)/(nir+red) + 0/min(0; (nir-red)/(nir+red) - 0.25), that calculates all the vegetation pixels except the bare soil (i.e., damaged areas). The loss adjuster can quantity the results into acres and severity of the damage. This is how the USDA RMA government issue the insurance policies to the farmers against extreme weather events.

Consequently, the Mexican company located in the northern state of Coahuila has realized the crucial importance of Pix4Dfields software to verify the quality and compare different treatments of crops for livestock. They are using Agriculture 4.0 methods to offer the grower solutions to implement precision agriculture in this area. The Agriculture 4.0 term is associated with the next big trends facing the agriculture industry. The main objective of this goal is to focus on precision agriculture using internet technologies, big data solutions for making efficient decisions. They performed experiments with the same variety of oat crops in two different soil types with two different treatments to crosscheck which method gives the best results. The results of this study are helpful for the livestock industry to increase crop yield efficiency.

Similarly, a DroneBee is a precision farming company in Italy offering highly technological services in precision farming and agronomic consultancy to farms, agricultural consortiums, research institutions, and insurance agencies. Recently, they have conducted two projects using RGB/multispectral/thermal sensors mounted on drones to analyze crop health and identify stressed or ill areas in vineyards. They chose two important areas, including Chianti and Bolgheri in Italy are producing high-quality wines. The Chianti project's main goal was to establish a precision irrigation system over 15 ha of vineyards to recover from grape yield losses. They used drones for the entire area to collect RGB images for analysis. The NDVI (Normalized Difference Vegetation Index) and CWSI (Crop Water Stress Index) indices were analyzed, respectively, to identify the stressed areas that will receive more water than healthy areas. The study demonstrated that the efficient use of drones at an appropriate time for agriculture water management is necessary for crop productivity. Whereas in Bolgheri project, they aimed to analyze 3 ha of vineyards for preparing the selective harvesting areas that produce high-quality wines. This has been done manually using a smartphone by following the map imported in Google Maps. Further, they have tested an innovative and specific algorithm to assess vines canopy geometry used to evaluate optimal pesticide dose. During these two projects, different data analytics were used for precision agriculture goals emphasized the potential use of drones in agriculture.

Orly et al. (2020) estimated the yield maps from apple orchards in Randwijk, near Wageningen, using UAV Imagery and a Regional Convolution Neural Network algorithm. They have accurately predicted the number of apples in each tree using a linear regression model were compared with apple count made in-situ by agrotechnician. This study demonstrated that in the R^2 agreement obtained from the regression model, which is 0.80, it is possible to detect the number of fruits in individual trees from UAV images. In addition to their findings, the Deep Learning algorithms give promising values and, therefore, a great potential method of estimating yields of other fruits.

In Australia (Stein et al., 2016), a multi-sensor framework was used to estimate every piece of fruit in a commercial mango orchard. They have extended their study to use the LiDAR component that automatically generates image masks for each canopy, allowing each fruit to be associated with the corresponding tree. A total of 522 trees and 71,609 mangoes were scanned, and where 16 trees were counted by hand for validation. The results indicated that single, dual, and multi-view methods could provide accurate yield estimates, but among them, only the multi-view approach can provide better results without calibration.

Weed production is another state of worry that reduces crop yield and quality. Rumex obtusifolius is a dangerous and fastest-growing weed, and it is one of the most common weeds in production grasslands in the Netherlands. A study (Valente et al., 2019) showed that using high-resolution RGB images obtained from a UAV; weeds can be detected up to 90% from a 6mm/pixel ortho mosaic generated from the aerial survey and deep learning. This method can be suitable for detecting weeds in the grasslands over the larger areas.

The most common software for processing raw images to generate yield maps includes Pix4D (www.pix4d.com/), AgisSoft PhotoScan (www.agisoft.com), and Photomodeler (www.photomodeler.com/).

Valente et al. (2020) proposed an automated method for counting plants from very high-resolution UAV imagery. The study examined to count 10 weeks old spinach plants in an experimental field with a surface area of 3.2 ha. They have examined the Excess Green Index and Otsu's method – and transfer learning using convolution neural networks to identify and count plants, respectively. Additionally, the Pix4D software was used to determine the aerial surveying flying path over the spinach field. The validation of data was prepared for 1/8 that of the surface area. The results demonstrated that the proposed methodology could count plants with an accuracy of 95% for a spatial resolution of 8 mm/pixel in an area up to 172 m^2. Further, they decreased spatial resolution by 50%, the maximum additional counting error achieved is 0.7%. Therefore, this method can be helpful for the estimation of crop yield for small-scale regions.

Matt Wade (2015) identified that using a drone's surface model to estimate crop yields & assess plant health gives better results and visualize biomass. Using Pix4Dmapper software to transform drone imagery into a digital surface model (DSM) for visualizing plant biomass. The method includes that by taking the difference between the drone-sourced DSM at the full canopy and topography of the field, it is easy to estimate the height of the crop. This provides valuable information about the biomass and furthers the data useful for the yield estimation. Drone data is best adopted for providing quick tips on fertilization. Apart from drone-based images, high spatial resolution satellite data, biophysical models, smart sampling, CropSnap, IoT, etc., are also being used in this large-scale pilot study.

4.3.2. Satellite-Based Yield Estimation

Nowadays, technology-based remote sensing methods provide more accurate results and are widely used in different applications. Compared to the field or direct estimated methods, satellite-based remote sensing models have become popular and gain more attention in the agriculture sector. Most of the satellite datasets are freely available from low scale to very high resolutions that facilitate the data from a smaller area to different constellations. The satellite captures the land information as an RGB image format associated with different spectral bands. However, each band should be processed through the image processing software for analyzing the different features of the Earth's surface. The first multispectral earth observation satellite was launched by NASA in 1972 to collect information about the Earth from space. Further, successive Landsat series have successfully emerged with an appropriate optimal ground resolution to efficiently track land use and document land changes due to urbanization, vegetation change, drought, wildfire, and biomass changes. However, optical remote sensing data limitation is that the images are not clear to distinguish the objectives at the time of cloudy nature, and especially problematic for collecting data during night time. Later on, the European Space Agency has launched the Synthetic Aperture Radar (SAR) to overcome many of these obstacles. SAR can monitor the Earth's surface during day and night, through most weather conditions, and the signal can penetrate the vegetation canopy. Meanwhile, the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the NASA EOS Terra satellite has visible, near-infrared, and shortwave infrared bands useful for calculating vegetation indices (Ziao et al., 2005). A number of satellite datasets are available to monitor everyday earth surfaces that provide more useful information to the policymakers, authorities, government, and public sectors. Table 2 outlines the major satellite datasets available for the agriculture field to monitor crop growth and yield estimation.

Over the last five decades, government and private space agencies have launched numerous earth observations (EO), enabling progressively increasing access to data at higher spatial, temporal, and spectral resolutions (Table 2). Among the most important are the Landsat missions since the 1970s and MODIS missions since the 1980s, which provided continuous time series data over the past five decades, which is largely responsible for the development of remote sensing science and enabled new applications in agricultural and economic development, including agricultural insurance.

Table 2. Indicative list of selected satellites/sensors and gridded data products of relevant indicators available in public and private domains for crop monitoring and insurance.

No.	Satellite/Sensor	Launch date	Ongoing / end date	Spatial resolution	Temperature sensor spatial resolution	Temporal resolution days	Category	Imagery Products available for crop monitoring *	Agency	Public/ private
1.	AVHRR	08/24/1981	ongoing	1.1 km		1.0,0.5	Biophysical	Vegetation Indices	NOAA	public
2.	MODIS -Terra	12/18/1999	ongoing	250m,500m,1km	1km	1.0	Biophysical	Temperature, LST	NASA	public
3.	MODIS-Terra	12/18/1999	ongoing	250m,500m,1km	1km	1.0	Biophysical	Vegetation Indices	NASA	public
4.	VIIRS	11/28/2011	ongoing	500m		1.0	Biophysical	Vegetation Indices	NASA/NOAA	public
5.	VIIRS_T	11/28/2011	ongoing	750m		1.0	Weather	Temperature, LST	NASA/NOAA	public
6.	ECOSTRESS	06/29/2018	ongoing	30m, 60m	60m	1-7 days	Biophysical	Plant Temperature, Evaporative Stress Index	NASA	public
7.	Landsat-1 MSS	07/23/1972	01/06/1978	30m, 60m		16	Biophysical	Vegetation Indices	NASA	public
8.	Landsat-2 MSS	01/22/1975	02/25/1982	30m, 60m		16	Biophysical	Vegetation Indices	NASA	public
9.	Landsat-3 MSS	03/05/1978	03/31/1983	30m, 60m		16	Biophysical	Vegetation Indices	NASA	public
10.	Landsat-4 MSS	07/16/1982	12/14/1993	30m, 60m		16	Biophysical	Vegetation Indices	NASA	public
11.	Landsat-5 TM	03/01/1984	01/15/2013	30m, 60m	120m	16	Biophysical	Vegetation Indices	NASA	public
12.	Landsat-7 ETM+	04/15/1999	ongoing	30m, 60m	60m	16	Biophysical	Vegetation Indices	NASA	public
13.	Landsat-8 OLI/TIRS	02/11/2013	ongoing	30m, 60m	100m	16	Biophysical	Vegetation Indices	NASA	public
14.	ASTER	Dec 1999	ongoing	15m, 30m, 90m	15m	1	Biophysical	Elevation, Surface reflectance, Emissivity, LST	NASA, METI(Japan)	public
15.	Sentinel-1A SAR	04/03/2014	ongoing	10m		5	Biophysical	Soil Moisture, ET	ESA	public
16.	Sentinel-1B SAR	04/25/2016	ongoing	10m		5	Biophysical	Soil Moisture, ET	ESA	public
17.	Sentinel-2A MSI	06/23/2015	ongoing	10m,20m,60m		5	Biophysical	Vegetation Indices, FAPAR, LAI	ESA	public
18.	Sentinel-2B MSI	03/07/2017	ongoing	10m,20m,60m		5	Biophysical	Vegetation Indices	ESA	public

19.	Sentinel-3 SLSTR	02/16/2016	ongoing	1km		5	Weather	Temperature, LST	ESA	public
20.	Sentinel-3A	02/16/2016	ongoing	300m		1, 2 days	Biophysical	Vegetation Indices	ESA	public
21.	Sentinel-3B OLCI	04/25/2018	ongoing	300m		1,2 days	Biophysical	Vegetation Indices	ESA	public
22.	Sentinel-3B SLSTR	04/25/2018	ongoing	500m	1km	1,2 days	Weather	Temperature, LST	ESA	public
23.	Sentinel 3B SAR	04/25/2018	ongoing	300m		1,2 days				
24.	Sentinel-5P TROPOMI	10/13/2017	ongoing	7km	17		Biophysical (Atmosphere)	Air quality, Ozone, GHGs	ESA	public
25.	SPOT-1	02/22/1986	12/31/1990	20m		1	Biophysical	Vegetation Indices	CNES	public
26.	SPOT-2	01/22/1990	07/31/2009	20m		1	Biophysical	Vegetation Indices	CNES	public
27.	SPOT-3	09/26/1993	11/14/1997	20m		1	Biophysical	Vegetation Indices	CNES	public
28.	SPOT-4	03/24/1998	07/31/2013	20m		1	Biophysical	Vegetation Indices	CNES	public
29.	SPOT-5	05/04/2002	03/31/2015	10m,20m		1	Biophysical	Vegetation Indices	CNES	public
30.	SPOT-6	09/11/2012	ongoing	6m		1	Biophysical	Vegetation Indices	CNES	public
31.	SPOT-7	06/30/2014	ongoing	6m		1	Biophysical	Vegetation Indices	CNES	public
32.	Pleiades 1A HiRI	12/17/2011	ongoing	2.8 m		1	Biophysical	Vegetation Indices	CNES	public
33.	Pleiades 1B HiRI	12/02/2012	ongoing	2.8m		1	Biophysical	Vegetation Indices	CNES	public
34.	RADARSAT-1	11/04/1995	12/14/2007	10 -100m		7	Biophysical	Vegetation structure, Soil Moisture, ET	CSA	Public
35.	RADARSAT-2	12/14/2007	ongoing	3-100m		7	Biophysical	Vegetation structure, soil moisture, ET	CSA	public
36.	RADARSAT-Constellation	06/12/2019	ongoing	3-100		4		Vegetation structure, Soil Moisture, ET	CSA	public
37.	ALOS 1,2,3-AVNIR	2006, 2014, 2019	ongoing	10m		5	Biophysical	Vegetation indices	JAXA (Japan)	public
38.	ALOS 1,2,3-PALSAR	2006, 2014,2019	ongoing	10m		3	Biophysical	Soil moisture		
39.	ALOS 1,2,3-PRISM	2006,2014, 2019	ongoing	2.5m; ALOS 3, 0.8m)		5	Biophysical	Elevation		

40.	Resourcesat-1 LISS-3	10/17/2003	ongoing	23.5m			Biophysical	Vegetation Indices	ISRO	public
41.	Resourcesat-1 LISS-4	10/18/2003	ongoing	5.8m			Biophysical	Vegetation Indices	ISRO	public
42.	Resourcesat-1 AWIFS	10/19/2003	ongoing	56m			Biophysical	Vegetation Indices	ISRO	public
43.	Resourcesat-2 LISS-3	4/20/2011	ongoing	23.5m			Biophysical	Vegetation Indices	ISRO	public
44.	Resourcesat-2 LISS-4	4/21/2011	ongoing	5.8m			Biophysical	Vegetation Indices	ISRO	public
45.	Resourcesat-2 AWIFS	4/22/2011	ongoing	56m			Biophysical	Vegetation Indices	ISRO	public
46.	OCO-2	07/02/2014	ongoing	3km	16		Biophysical	Solar induced fluorescence	NASA	public
47.	OCO-3	05/04/2019	ongoing	3km			Biophysical	Solar induced fluorescence	NASA	public
48.	CBERS 1	10/14/1999	8/1/2003	20m, 80m	160m	3, 26	Biophysical	Vegetation indices	CNSA / INPE	public
49.	CBERS 2	10/21/2003	12/31/2009	20m, 80m,260m	160m	3,5,5	Biophysical	Vegetation indices	CNSA / INPE	public
50.	CBERS 3	12/1/2013	12/1/2013	10m, 20m, 40m	80m	3,5,26	Biophysical	Vegetation indices	CNSA / INPE	public
51.	CBERS 4	12/1/2014	ongoing	10m, 20m, 40m, 64m	80	3,5,26	Biophysical	Vegetation indices	CNSA / INPE	public
52.	CBERS 4A	12/20/2019	ongoing	8m, 16m, 55m		5,31	Biophysical	Vegetation indices	CNSA / INPE	public
53.	GOSAT	2009	ongoing	10.5 km		3	Biophysical	Green house gases (GHG)	JAXA	public
				0.5, 1.5 km		3	Biophysical	Clouds and aerosols		
54.	AMSRE									
55.	SMAP	01/31/2015	ongoing	36km		2 to 3 days	Biophysical	Soil Moisture	NASA	
56.	SMOS	11/02/2009	ongoing	40km		2 days	Biophysical	Soil Moisture	ESA	public
57.	TRMM	11/27/997	2014	5km		1 day	weather	Precipitation	JAXA/NASA	public
58.	GMI (GPM)	2/27/2014	ongoing	11km			Weather	Precipitation	NASA/JAXA	public

59.	GOES-1-17	10/16/1975	ongoing	1km		1day	Weather	Soil Moisture, ET	NOAA/NASA	public
60.	Meteosat-1-7	11/23/1977	1984	1km		30 min	Weather	Precipitation, Soil Moisture, ET	EUMETSAT/E	public
61.	Meteosat-8-11	08/28/2002	ongoing	1km		5, 15 min	Weather	Precipitation, Soil Moisture, ET	EUMETSAT/E	public
62.	CHIRPS (Gridded Data products)	01/01/1981	present	5km		1.0	Weather	Precipitation	UCSB	public
63.	CHIRTS(Gridded Data products)	01/01/1983	present	5km		1.0	Weather	Temperature	UCSB	public
64.	GRIDMET (Gridded Data products)	1979	present	4km				Rainfall, Relative Humidity, T_{min}, T_{max}, wind speed, vapour pressure deficit, reference evapotranspiration	UC Merced	Publ;ic
65.										
66.	VanderSat	06/01/2002	ongoing	100m			Biophysical	Soil Moisture, ET	Vandersat	private
67.	IKONOS	1999	2015	4m		3	Biophysical		Digital Globe	private
68.	WorldView-1	09/18/2007	ongoing	0.5m-2m		1.7	Biophysical	Vegetation Indices	Maxar (Digital Globe)	private
69.	WorldView-2	10/08/2009	ongoing	0.5m-2m		1.1	Biophysical	Vegetation Indices	Maxar	private
70.	WorldView-3	08/23/2014	ongoing	0.3m-1.24m			Biophysical	Vegetation Indices	Maxar	private
71.	WorldView-4	11/11/2016	ongoing	0.3m-1.24m			Biophysical	Vegetation Indices	Maxar	private
72.	RapidEye -Planet	08/29/2008	03/31/2020	5m			Biophysical	Vegetation Indices	Planet	private
73.	PlanetScope	06/22/2016	ongoing	3m,5m			Biophysical	Vegetation Indices	Planet	private
74.	SkySat (Planet)	11/21/2013	ongoing	0.72m			Biophysical	Vegetation Indices	Planet	private
75.	aWhere	01/01/2006	ongoing	9km			Weather	Precipitation, Temperature	aWhere	private
76.	ICEYE-X1 to X10; SAR	01/01/2018	ongoing	0.25 to 5m, 5 to 20m		hourly to daily	Weather	Soil Moisture, flood mapping	ICEYE	private

*Algorithms are available in respective toolkits of several satellites/sensors (AVHRR, MODIS, VIIRS, LANDSAT, SENTINEL) for derived vegetation status indicators of biophysical importance in representing important vegetation properties of crops to assess crop condition and yield estimation, like LAI (leaf area index), FAPAR (Fraction of absorbed photosynthetically active radiation, FVC (Fractional Vegetative Cover), and others.

Satellite remote sensing offers a promising data tool to assess crop yields. Many studies examined a high yield correlation between satellite-based estimates and traditional sources [Ref]. Nevertheless, the satellite data available since the 1970s but shows limited applications in the agriculture field. Several recent developments have enabled extensive use of remote sensing data towards crop mapping, vegetation monitoring, disease detection, yield prediction, and many other agricultural functions. Firstly, access to satellite data has significantly become easier these days because of the additional satellites launched into an orbit, including Landsat-8, Sentinel series, and private satellites. Second, the archives of all the imageries are freely available to the public, and it becomes more easier to access them via Google cloud platforms without downloading data being analyzed. Finally, new algorithms have been developed for modeling crop yields from satellite measurements. In contrast, traditional methods are standardized methods to assign for a specific crop or region, unlike new algorithms that are more scalable for every crop prediction.

For example, Azzari et al. (2017) have introduced a Scalable satellite-based Crop Yield Mapper (SCYM) based on a coarser temporal resolution but for fine spatial resolution of Landsat data to estimate yield. It combines crop model simulations with imagery and weather data to generate 30 m resolution yield estimates, demonstrated that there is no further need for ground calibration. SCYM model employs simulations from crop models followed by regression analysis for final crop yield imagery uses vegetation indices that are available during the entire growing season. In contrast, they modeled the second method to compare the yield estimates for larger scalable areas. In this case, they used fine temporal resolution and coarse spatial resolution such as the MODIS data to track variation in vegetation indices. This method, referred to as PEAKVI, was developed by the Group on Earth Observations Global Agricultural Monitoring Program (GEOGLAM; http://geoglam-crop-monitor.org/). The PEAKVI method relates yield to the peak value of spatial composition of all the vegetation indices. Both the methods, SCYM and PEAKVI, were used to generate a final crop yield map utilizing the archives of all the available satellite imageries for more than a decade in three countries and validated against ground-based estimates at the administrative levels. The three different study regions were chosen to test the model with varying crops, field sizes, and landscape heterogeneity, such as maize crop in the US corn belt, wheat in India (i.e., Indo-Gangetic Plains), and maize in southern Zambia. The results indicated that the performance of SCYM and MODIS PEAKVI in capturing spatial variability, the results in India showed better correlation with R^2 values above 0.45 in most years as compared with the US. On the other hand, Landsat PEAKVI performed poorly in India's maize, with R^2 below 0.25. Overall, the best average performance for capturing spatial yield variation is MODIS-based SCYM, with an average RMSE of 0.56 ton/ha. The factors affecting the weak correlation showed in India because the yield data was not available in some of the districts and could be the low temporal variability in this irrigated system. Therefore, this method shows the best results when the complete ground data is available in all Indian districts.

Remote sensing applications in small-holder agriculture significantly face considerable challenges because of small field size, heterogeneity in management practices, landscape fluctuations, and the widespread presence of trees within the fields (Vancutsem et al., 2012). Despite these complications, some studies achieved land use classification accuracies from small-holder farming systems using high-resolution satellite images (Lebourgeois et al., 2017; Debats et al., 2016). On the other hand, quantitative estimates of small-holder yields from remote sensing data also offer the potential way to improve scalability and objectivity compared to estimates from field survey data (Lambert et al., 2018). The method used to estimate crop yield from satellite data is referred to as the crop growth model, where they simulate the relationship between plants, vegetation indices, and the environment to predict crop yield. The model requires a large amount of input data, including weather, land use, soil, plant, and management practices. In theory, crop models perform better yield estimation achieve with high accuracy at a fine-scale representation of the land, soil, and weather data. But, in practice, they are constrained by data limitations, biased towards temperature specifications, variations in fertilizer applications, and high-yielding cultivars (Burke & Lobell, 2017; Srivastava et al., 2016). A set of recent studies focused on yield estimation from crop models using earth observation data (Kasampalis et al., 2018; Wit & Diepen, 2008). Numerous research studies have been developed and tested crop growth models for yield estimation. The most common reliable crop models are outlined in Table 3. Since the crop growth models have certain limitations because of the lack of spatial information on the actual conditions of the field. Remote sensing data can fill the gap of the missing spatial information needed by the crop models for improved yield prediction. In the 21st century, geospatial technologies have taken advantage of crop models; thus, the yield can be predicted at any given field resulting from improved models, advanced technologies, spatial datasets, and ground-based measurements. Crop growth models are mainly classified into two different types: Empirical or dynamic models depending on the model's design. The main advantage of crop growth models is explained in the following chapters.

Table 3. List of most common crop models.

Crop Model	Website Link
1. APSIM	http://www.apsim.info/
2. AgrometShell	http://www.hoefsloot.com/agrometshell.htm
3. Aquacrop	http://www.fao.org/aquacrop
4. CERES-wheat	http://nowlin.css.msu.edu/wheat_book/
5. CERES	http://dssat.net/
6. CROPGRO-Soybean	http://ecobas.org/www-server/rem/mdb/cropgro.html
7. Cropsyst	http://modeling.bsyse.wsu.edu/CS_Suite_4/CropSyst/index.html
8. DAISY	https://soil-modeling.org/resources-links/model-portal/daisy
9. DSSAT	http://dssat.net/
10. EPIC	https://epicapex.tamu.edu/
11. FarmSim	http://models.pps.wur.nl/node/961
12. Fasset	http://www.fasset.dk/
13. GLAM	https://www.see.leeds.ac.uk/research/icas/research-themes/climate-change-and-impacts/climate-impacts/glam/
14. HERMES	http://www.zalf.de/de/forschung_lehre/software_downloads/Seiten/default.aspx

15.	InfoCrop	https://www.quantitative-plant.org/model/InfoCrop
16.	LINTUL	http://www.csa.wur.nl/UK/Downloads/LINTUL/%22%3Ehttp://www.csa.wur.nl/UK/Downloads/LINTUL/%3C/a%3E
17.	MONICA	http://monica.agrosystem-models.com/
18.	ORYZAv3	https://sites.google.com/a/irri.org/oryza2000/about-oryza-version-3
19.	STICS	http://www6.paca.inra.fr/stics_eng
20.	SUCROS	https://models.pps.wur.nl/simple-and-universal-crop-growth-simulator-sucros
21.	SWAP	http://www.swap.alterra.nl/
22.	WOFOST	http://www.wageningenur.nl/en/Expertise-Services/Research-Institutes/alterra/Facilities-Products/Software-and-models/WOFOST.htm

(i) Statistical modelling for crop yield estimation with remote sensing data

Empirical models are also known as correlative or statistical models. These models are expressed as regression equations that establish a correlation of using one or a few parameters of the observed data. They provide limited information about the mechanisms that give rise to the response (Kasampalis et al., 2018). Therefore, they are not suitable for the uncertainty conditions than dynamic models. Statistical models are categorized into three types, time-series, panel, and cross-sectional methods (Lobell & Burke, 2010). These models involved in yield prediction mainly depend on empirical relationships between measured or observed historical data against climate change scenarios (Schlenker & Lobell, 2010). This model's main advantage over the process-based approach is that it includes weather-induced pathogens and air pollution. However, only a few inputs are required and relatively high predictive power if sufficient training attributes are available (Li et al., 2019). Nevertheless, it has certain limitations because of the co-linearity between the predictors, moderate data extrapolation issues, and data dependency of prediction performance. Since the statistical model has remained the best for yield prediction, the model has to address the above challenges in the near future (i.e., up to 2050) to predict the effects of climate change on crop yield. Therefore, several research studies have conducted empirical models based on regressions between yield measurements and spectral vegetation indices (Kern et al., 2018).

(ii) Process-based crop models

Dynamic models are also known as process-based models. They simulate crop progression through time using differential equations to describe crop development (Rauff & Bello, 2015). They require more input data parameters to simulate the model than statistical methods. Especially, they need physiological data from field experiments to run the model for final yield estimation. However, the importance of a process-based model at a regional scale level is limited due to the lack of spatial information on model parameters (Moulin et al., 2001). This method is highly advantageous to policymakers in that they can explore future climate trends using the existing data and estimate yield responses to changes in spatial and temporal patterns of climate and management practices (Jones et al., 2017). These models applied to the small field to the continental scale areas run as the basis over an hour, days, or seasons. The process-based models or dynamic models are classified into field-scale models and regional scale models. Field-scale models comprehensively simulate plant functions; ORYZA2000 is the most common method in this area (Li et al., 2017). The regional-scale model often incorporates a mixture of canopy processes and broader scale crop-climate relationships. General Large Area Model (GLAM) is the popular yield estimation method in a regional-scale model (Challinor et al., 2004).

4.3.3. Spectral Indices

In recent decades, remote sensing acts as a major promising data source for crop monitoring and final yield prediction in large spatial extents. The derived indices from different spectral bands are significantly used in vegetation monitoring, crop growth, and check available soil moisture conditions across the fields. The most common indices are used for monitoring crop development like the Normalized Difference Vegetation Index (NDVI), biophysical parameters such as the Leaf Area Index (LAI), or Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) are widely used in relation to plant status, productivity, therefore, they are essential to estimate crop yield (Huang et al., 2015; Ferencz et al., 2004). In addition, other remote sensing indices were also proposed to support the crop yield estimations. For example, the Normalized Difference Water Index (NDWI) is related to soil moisture conditions, which can assess plant status and production, relatively enhancing crop yield estimation (Bolton & Friedl, 2013). Since the 2000s, the Moderate Resolution Imaging Spectroradiometer (MODIS) became one of the popular data sources, thus provides data every two days gives high radiometric and spatial accuracy.

(i) NDVI

Normalized Difference Vegetation Index (NDVI) quantifies vegetation with the difference between near-infrared and red light. NDVI is the most common vegetation index used in agriculture to observe plant health growth conditions. NDVI value ranges from -1 to 1, where negative values represent water bodies or dry areas, and positive values close to dense vegetation, respectively.

$$NDVI = (NIR - Red) / (NIR + Red)$$

(ii) SAVI

The Soil Adjusted Vegetation Index (SAVI) was designed to minimize soil brightness influences. The general equation for SAVI is similar to the NDVI; however, a soil adjustment factor (L) was added to the NDVI to correct soil noise effects (soil color, soil moisture, soil variability across the region, etc.).

$$SAVI = ((NIR-Red)/(NIR+Red+L))*(1+L)$$

L is a variable ranging from -1 to 1, depending on the amount of green vegetation present in the area. L is set to zero during high green vegetation analysis, and low green vegetation analysis regions require L is equal to 1. The SAVI index is useful for analyzing young crops (i.e., arid regions with sparse vegetation, dry soil moisture areas).

(iii) ARVI

The Atmospherically Resistant Vegetation Index (ARVI) is the first vegetation index that is not prone to atmospheric factors (i.e., aerosol). Kaufman and Tanre invented the ARVI formula is NDVI corrected for atmospheric scattering effects in the red reflectance spectrum using the measurements in blue wavelengths.

$$ARVI = (NIR – (2 * Red) + Blue)/ (NIR + (2 * Red) + Blue)$$

Compared with other indices, the ARVI agriculture index is more robust to topographic effects, which is highly an effective monitoring tool for tropical mountain regions often polluted by soot coming from slash-and-burn agriculture. This index is mainly used for high atmospheric aerosol content (i.e., rain, dust, fog, smoke, air pollution).

(iv) EVI

Enhanced Vegetation Index (EVI) was invented by Liu & Huete to simultaneously correct NDVI results for atmospheric influences and soil background signals, mainly in the dense canopy area. The EVI value ranges from -1 to 1, and for healthy vegetation, it varies between 0.2 and 0.8.

$$EVI = 2.5 * ((NIR-Red)/((NIR)+(C1*Red)-(C2*Blue)+L))$$

Where C1 & C2 coefficients correct for aerosol scattering present in the atmosphere, and the L factor is to adjust soil and canopy background. The MODIS standard coefficients values of EVI for C1=6, C2=7.5, L=1, respectively. The EVI is mostly used for analyzing larger areas with a huge amount of chlorophyll with minimum topographic effects.

(v) GCI

Green Chlorophyll Index (GCI) is used to estimate leaf chlorophyll content in various species of plants. It reflects the chlorophyll content in plants useful for identifying the physiological state of plant growth conditions. GCI decreases in stressed plants; therefore, it is a tool to measure plant health status.

$$GCI = (NIR)/(Green)-1$$

It is helpful in monitoring the impact of seasonality, environmental stress, applied pesticides on plant health.

(vi) SIPI

Structure Insensitive Pigment Index (SIPI) is suitable for analyzing vegetation with the variable canopy structure. It estimates the ratio of carotenoid to chlorophyll: the increased value signals of stressed vegetation.

$$SIPI = (NIR – Blue) / (NIR – Red)$$

High SIPI values (increased carotenoids and decreased chlorophyll) are often indicators of plant disease associated with the loss of chlorophyll in plants. It is used for monitoring plant health in regions with high variability in canopy structure or LAI for early detection of plant disease or other causes of stress.

(vii) NBR

Normalized Burn Ratio (NBR) is used to highlight burned areas following the fire. The NBR index is calculated by both NIR and SWIR wavelengths: healthy vegetation shows high reflectance in the NIR spectrum, whereas the recently burned areas of vegetation reflect highly in the SWIR spectrum.

$$NBR = (NIR – SWIR) / (NIR + SWIR)$$

The NBR index is highly recommended for agriculture and forestry to detect active wildfires, analyze burn severity, and monitor vegetation survival after the burn.

The important vegetations indices that can be used in agriculture are listed in Table 4.

Table 4. Vegetation indices for use in agriculture.

S.No.	Index Name	Formula
1	Aerosol free vegetation index 1600	$\left(NIR - 0.66\,\frac{1600nm}{NIR + 0.661600nm}\right)$
2	Aerosol free vegetation index 2100	$\left(NIR - 0.5\,\frac{2100nm}{NIR + 0.562100nm}\right)$
3	Ashburn Vegetation Index	$2.0[800{:}1100] - [600{:}700]$
4	Atmospherically Resistant Vegetation Index	$\frac{NIR - RED - y(RED - BLUE)}{NIR + RED - y(RED - BLUE)}$
5	Atmospherically Resistant Vegetation Index 2	$-0.18 + 1.17\left(\frac{NIR - RED}{NIR + RED}\right)$
6	Blue-wide dynamic range vegetation index	$\frac{0.1NIR - BLUE}{0.1NIR + BLUE}$
7	Chlorophyll vegetation index	$NIR\,\frac{RED}{GREEN^2}$
8	Corrected Transformed Vegetation Index	$\frac{NDVI + 0.5}{\lvert NDVI + 0.5\rvert} * \sqrt{\lvert(NDVI) + 0.5\rvert}$
9	Difference NIR/Green Green Difference Vegetation Index	$NIR - G$
10	Difference Vegetation Index MSS	$2.4[800{:}1100] - [600{:}700]$
11	Enhanced Vegetation Index	$2.5\frac{NIR - RED}{(NIR + 6RED - 7.5BLUE) + 1}$
12	Enhanced Vegetation Index 2 -2	$2.5\frac{NIR - RED}{NIR + 2.4RED + 1}$
13	Enhanced Vegetation Index 2	$2.4\frac{NIR - RED}{NIR + RED + 1}$
14	Global Vegetation Moisture Index	$\frac{(NIR + 0.1) - (SWIR + 0.02)}{(NIR + 0.1) + (SWIR + 0.02)}$
15	Green atmospherically resistant vegetation index	$\frac{NIR - (GREEN - (BLUE - RED))}{NIR - (GREEN + (BLUE - RED))}$
16	Green Normalized Difference Vegetation Index	$\frac{NIR - [540{:}570]}{NIR + [540{:}570]}$
17	Green Optimized Soil Adjusted Vegetation Index	$\frac{NIR - G}{NIR + G + Y}$
18	Green Soil Adjusted Vegetation Index	$\frac{NIR - G}{NIR + G + L}(1 + L)$
19	Ideal Vegetation index	$\frac{NIR - b}{a * RED}$
20	Infrared percentage vegetation index	$\frac{\frac{NIR}{NIR + RED}}{2}(NDVI + 1)$
21	Mid-infrared vegetation index	$\frac{[700{:}1300]}{[1570{:}1780]}$
22	Misra Green Vegetation Index	$-0.386[500{:}600] - 0.53[600{:}700] + 0.535[700{:}800] + 0.532[800{:}1100]$
23	Misra Yellow Vegetation Index	$0.723[500{:}600] - 0.597[600{:}700] + 0.206[700{:}800] - 0.278[800{:}1100]$
24	Modified Normalized Difference Vegetation Index RVI	$\frac{RVI - 1}{RVI + 1}$
25	Modified Soil Adjusted Vegetation Index	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - RED)}}{2}$

26	Modified Soil Adjusted Vegetation Index hyper	$(0.5)\left((2800nm+1) - \sqrt{(2800nm+1)^2 - 8(800nm-670nm)}\right)$
27	Modified Triangular Vegetation Index 1	$1.2(1.2(800nm-550nm) - 2.5(670nm-550nm))$
28	Modified Triangular Vegetation Index 2	$\left(1.5\frac{1.2(800nm-550nm) - 2.5(670nm-550nm)}{\sqrt{(2800nm+1)^2 - (6800nm - 5\sqrt{670nm}) - 0.5}}\right)$
29	Nonlinear vegetation index	$\frac{[780:1400]^2 - RED}{[780:1400]^2 + RED}$
30	Normalized Difference MIR/NIR Normalized Difference Vegetation Index (in case of strong atmospheric disturbances)	$\frac{MIR - NIR}{MIR + NIR}$
31	Normalized Difference NIR/Blue Blue-normalized difference vegetation index	$\frac{NIR - BLUE}{NIR + BLUE}$
32	Normalized Difference NIR/MIR Modified Normalized Difference Vegetation Index	$\frac{NIR - MIR}{NIR + MIR}$
33	Normalized Difference NIR/Red Normalized DifferenceVegetation Index, Calibrated NDVI - CDVI	$\frac{NIR - RED}{NIR + RED}$
34	Normalized Difference Vegetation Index 690-710	$\frac{NIR - [690:710]}{NIR + [690:710]}$
35	Normalized Difference Vegetation Index C	$\frac{NIR - RED}{NIR + RED}\left(1 - \frac{SWIR - SWIR_{min}}{SWIR_{max} - SWIR_{min}}\right)$
36	Optimized Soil Adjusted Vegetation Index	$(1+Y)\frac{800nm - 670nm}{800nm + 670nm + Y}$
37	Optimized Soil Adjusted Vegetation Index 1510	$\frac{(1+L)(800nm - 1510nm)}{800nm + 1510nm + L}$
38	Optimized Soil Adjusted Vegetation Index 2	$(1+0.16)\frac{750nm - 705nm}{750nm + 705nm + 0.16}$
40	Perpendicular Vegetation Index	$\left(\frac{1}{\sqrt{a^2+1}}\right)(NIR - ar - b)$
41	Red-Edge Stress Vegetation Index	$\frac{718nm + 748nm}{2} - 733nm$
42	Renormalized Difference Vegetation Index	$\frac{800nm - 670nm}{\sqrt{800nm + 670nm}}$
43	Simple Ratio 800/670 Ratio Vegetation Index	$\frac{800nm}{670nm}$
44	Simle Ratio 860/1240	$\frac{860nm}{1240nm}$
45	Simple Ratio NIR/G Green Ratio Vegetation Index	$\frac{NIR}{G}$
46	Simple Ratio NIR/Red Difference Vegetation Index, Vegetation Index Number (VIN)	$\frac{NIR}{RED}$
47	Simple Ratio Red/NIR Ratio Vegetation Index	$\frac{RED}{NIR}$
48	Single Band 705	$705nm$
49	Single Band 735	$735nm$
50	Soil Adjusted Vegetation Index	$\frac{800nm - 670nm}{800nm + 670nm + L}(1+L)$

51	Soil and Atmospherically Resistant Vegetation Index	$(1+L)\frac{800nm-(Rr-y(RB-Rr))}{800nm+(Rr-y(RB-Rr))+L}$
52	Soil and Atmospherically Resistant Vegetation Index 2	$2,5\frac{NIR-RED}{1+NIR+6RED-7,5BLUE}$
53	Soil and Atmospherically Resistant Vegetation Index 3	$(1+0,5)\frac{833nm-658nm}{833nm+658nm+0,5}$
54	Soil-adjusted vegetation index	$\frac{NIR}{RED+\frac{b}{a}}$
55	Specific Leaf Area Vegetation Index	$\frac{NIR}{RED+SWIR}$
56	Spectral Polygon Vegetation Index	$0.4(3.7(800nm-670nm)-1.2\|530nm-670nm\|)$
57	Transformed Soil Adjusted Vegetation Index	$\frac{B(NIR-B*R-A)}{RED+B(NIR-A)+X(1+B^2)}$
58	Transformed Soil Adjusted Vegetation Index 2	$\frac{a.NIR-a.RED-b}{RED+a.NIR-a.b}$
59	Transformed Vegetation Index	$\sqrt{(NDVI)+0,5}$
60	Triangular Vegetation Index	$0.5(120(750nm-550nm)-200(670nm-550nm))$
61	Vegetation Condition Index	$\frac{NDVIj-NDVImin}{NDVImax-NDVImin}*100$
62	Vegetation Index 700	$\frac{700nm-[660:680]}{700nm+[660:680]}$
63	Weighted Difference Vegetation Index	$NDVI-a*RED$
64	Wide Dynamic Range Vegetation Index	$\frac{0.1NIR-RED}{0.1NIR+RED}$

Source: https://www.indexdatabase.de/search/i-search.php?s=vegetation&offset=2

4.3.4. Remote Sensing Data Fusion Model

Every satellite sensor has its own procs and cons, depending on its data availability between different spatial, spectral, and temporal resolutions. Particularly, the data uncertainty issues have been rising when the data is being used for agriculture precision analysis. Satellite data plays a key role in providing data in every stage of crop growth, from the time of the sowing period to the harvesting time. However, there is a prominent trade-off between the spatial and temporal resolution of a single sensor (Shen et al., 2021). High resolution and high revisiting frequency cannot be achieved by the same sensor. However, agricultural analysis requires continuous time-series data at fine resolution. Many satellite sensors have a very high temporal resolution of one day or several days, but their spatial resolution ranges between 250 m to 1 km. For example, MODIS datasets are available every one or two days, but due to their coarse spatial resolution, they cannot meet the requirements of precision agriculture monitoring. On the contrary, the sensors like optical (i.e., Landsat TM/ETM+/OLI) and radar (Sentinel 2 MSI) datasets have very high spatial resolutions of 30 m and 10 m. Nevertheless, the revisit cycle of a single sensor lasts ten or more days, which is a drawback for agriculture analysis. Optical data has mostly been covered with clouds that further degrade the temporal resolution and the quality of NDVI data. To overcome the above challenges for obtaining the best quality of NDVI data from a small scale to large scale agriculture monitoring, the data fusion models are the most appropriate methods in remote sensing techniques. Therefore, merging the high spatial resolution and high temporal resolution of different sensors is an effective solution to get high-quality NDVI data at a low cost (Liao et al., 2017; Zhu et al., 2010).

Many studies have developed spatiotemporal fusion models with certain assumptions for different application purposes. The first proposed data fusion method was the spatial and temporal adaptive reflectance fusion model (STARFM), widely used in data fusion techniques. The STARFM achieves accurate results in homogeneous regions during prediction time in stable land cover regions. Whereas in non-homogeneous areas, the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) was proposed by Zhu et al. (2010).

High Spatial Resolution	CubeSAT, Planet Lab Planetary scope	3.125 m
	Sentinel 2, ESA European Space Agency	10 m
	Landsat, USGS US Geological Survey	30 m
	MODIS, NASA	500 m

High Temporal Resolution	MODIS, NASA	daily
	CubeSAT, Planet Lab Planetary scope	Every ~3 days
	Sentinel 2, ESA European Space Agency	~ weekly
	Landsat, USGS US Geological Survey	every ~2 weeks

Fig. 1. Spatial and temporal resolutions of satellite imagery.

Figure 1 depicts the increasing high spatial and temporal resolution of satellite imagery. CubeSat data (3.125 m) is available at the very fine spatial resolution, followed by the Sentinel (10 m), Landsat (30 m), and MODIS (500 m), respectively. In contrast, MODIS datasets are available at the highest temporal resolution compared with other datasets. CubeSAT has a high temporal resolution every three days, whereas the Sentinel series provides weekly data, and Landsat gives data every two weeks. With the differences in different spatial and temporal resolutions, researchers have developed data fusions models to aggregate them into fine spatial and high temporal resolutions. However, in order to collect timely high-resolution data, these models are useful for monitoring everyday crop status with MODIS and combining other satellite datasets.

4.3.5. Remote Sensing Data Assimilation Model

Several factors are influencing the crop models when monitoring crop growth conditions and yield at a regional scale (Zhao et al., 2013). The crop parameters, farming management practices, and planting patterns are challenging to obtain from the different sources that lead to data uncertainty issues at the regional scale level. Therefore, it is difficult to validate the reliability of modeling results. In recent decades, remote sensing information has mainly been used by statistical or semi-empirical models to estimate crop yield. The assimilation of remote sensing information with crop models to improve yield estimation is a new trend in agriculture. The coupling of remote sensing spectral indices and crop models has allowed scientists to invert the crop parameters needed for model fitting (Wang, 2008). Therefore, this tool is used for them monitoring crop growth and estimating crop yield over a region. Remote sensing technologies can provide crop growth conditions at a field scale level against the weather factors on crop growth and development. It is essential to combine crop models and remote sensing data in order to inspect the crop status at every stage of crop growth (Revill et al., 2013).

Wit & Diepen (2007) predicted a regional yield by using assimilation of the crop data with the Ensemble Kalman filter (EnKF) for improving modeling results. They used KnKF to assimilate the microwave sensor coarser resolution for correcting errors in the WOFOST crop model. The study area was chosen in different countries, including Spain, France, Italy, and Germany, for the simulation period between 1992-2000. The results were evaluated with the statistical field yield at regional and national scales. Overall they achieved accurate results in estimating winter wheat in the study regions. In another study (Zhao et al., 2013), they have extended the WOFOSAT model to check whether to improve yield results or not. They evaluated a potential use of crop model (PyWOFOST) with the combination of remote sensing indices (MODIS LAI) and WOFOST model to simulate maize growth and yield in Northeastern China. The assimilation of remote sensing information and crop model for analyzing the impact of data uncertainties reflects the prediction yield model. The regional maize yield was estimated by using the PyWOFOST model and validated with the statistical yield values. The results demonstrated that the estimation of maize yield with remote sensing and crop model had shown better simulation than the one without assimilation. Fang et al. (2008) proposed a method to estimate corn yield in the state of Indiana by integrating Crop-System Model-CERES-Maize with the MODIS leaf area index (LAI). The estimated crop yield compared reasonably well with the US Department of NASS statistics for most regions. Using the seasonal LAI in the model enhances the results compared with the highest LAI values. Planting, seasonal growth variation, fertilizer application rates, and management practices were also estimated at a regional level. The integration of remotely sensed data with crop models through data-assimilation methods is an effective approach to estimate crop yield at the regional scale level.

4.3.6. Machine Learning-based Yield Estimation

Machine learning has emerged with big data technologies and high-performance cloud computing programs for enabling numerous opportunities for data-driven science in the multidisciplinary agri-technologies domain. There are eight different types of machine learning models are available in this field applicable to data classification and decision-making systems. Remote sensing-based methods require the processing of a large amount of data from different platforms. Therefore, ML models alleviate data complexity issues due to their ability to handle enormous input data and process non-linear tasks (Chlingaryan et al., 2018). At the beginning of the 21st century, improving crop production, quality, and accurate yield estimation while reducing operational costs and environmental effects is crucial in precision agriculture.

Table 6. Abbreviations used in machine learning tools.

ANNs	artificial neural networks

BM	bayesian models
DL	deep learning
DR	dimensionality reduction
DT	decision trees
EL	ensemble learning
IBM	instance-based models
SVMs	support vector machines

Machine learning methodologies involve a learning process with the objective to learn from a set of attributes (i.e., training data) to perform a task. The data in ML consists of a group of attributes, usually termed as features or variables. The feature data type can be in the form of nominal (enumeration), binary (0 or 1), ordinal (A+ or B-), or numeric (integer, real number, etc.). The statistical or mathematical models measure the performance of the ML models, and algorithms are used. However, for accuracy assessment analysis, the training data was further divided into testing and validating input samples. ML tasks are mainly classified into two categories; supervised and unsupervised learning. The supervised learning method defines as that maps an input to an output based on example input-output pairs. It derives a function from labeled training data consisting of a set of training examples. The reviewed articles have been predominantly focused on four major categories: crop management, livestock management, water management, and soil management. ML methodologies in the crop section were divided into several categories, including yield prediction, weed detection, crop mapping, vegetation growth assessment, and species recognition.

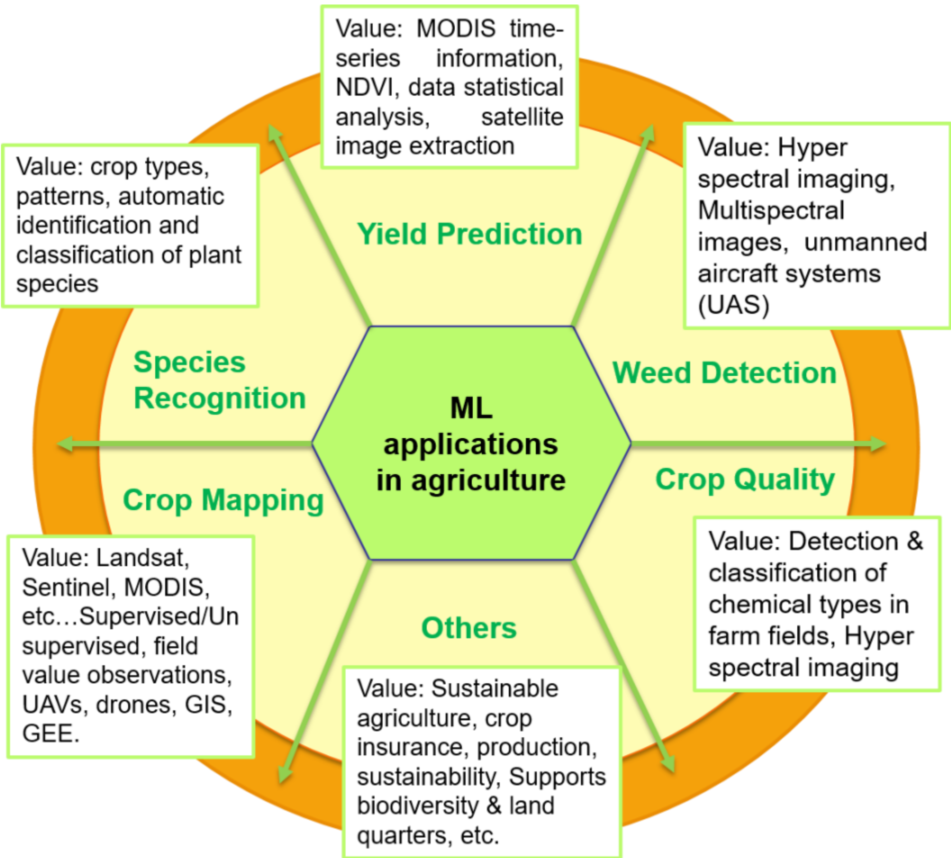

Fig. 2. Applications of machine learning in agriculture.

Yield Prediction

Achieving maximum crop yield at minimum cost is one of the main goals of sustainable agricultural production. Therefore, early detection and management of problems that are related to crop yield can assist farmers in increasing crop yield, making profits, and developing subsequent crop management goals.

Many studies have assimilated the results of yield predictions on the basis of a field survey (data) by exploiting remote sensing and GIS techniques. Field-based yield predictions are subjective, laborious, cost-effective, time-consuming, and are only feasible for small-scale assessment and monitoring of croplands. Most of the time raises data uncertainty and data redundant issues. ML method is an effective approach in yield prediction analysis, whereas an efficient, low-cost, and non-destructive method paved the way for successful agriculture yield prediction analysis. Several studies reported the essential use of ML techniques to achieve accurate yield prediction for the different crops in recent years (Pantazi et al., 2016; Kung et al., 2016; Subhadra et al., 2016). The most successful ML techniques have been Artificial Neural Networks (Craninx et al., 2008; Fortin et al., 2011; Gandhi et al., 2016; Khosla et al., 2020), Random Forest (RF) (Ramos et al., 2020; Kang et al., 2020), Support Vector Regression (SVR), and Convolution Neural Networks.

Recently, Kang et al. (2020) reported a comparative study of least absolute shrinkage and selection operator (LASSO) regression, SVR, RF, eXtreme Gradient Boost (XGBoost), as well as two deep neural networks—LSTM and CNN for assessment of county-level maize yield prediction in the US Midwest. They validated the models using the simple Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), correlation coefficient (r), and bias. In comparison, the XGBoost algorithm outperforms other algorithms both in accuracy and stability than RF model. It achieved the lowest errors across the produced crop yield models. In contrast, deep neural networks such as LSTM and CNN are not advantageous. However, the results were contributed to asses different machine learning algorithms for the estimation of seasonal yield forecasts.

Another study compared two ML algorithms between Boosted Regression Trees (BRT) and Support Vector Machines (SVM) to predict regional winter wheat yields (Stas et al., 2016). They constructed the model based on NDVI derived indices from low-resolution SPOT VEGETATION imagery. After the evaluation, BRT showed the highest accuracy than the SVM model by overcoming the overfitting problem. In contrast, Nosratabadi et al. (2020) proposed a new crop yield prediction model based on an artificial neural networks-imperialist competitive algorithm (ANN-ICA) and artificial neural networks-gray wolf optimizer (ANN-GWO) algorithms. The results demonstrated that the ANN-GWO model proved a better performance in the crop yield estimation than ANN-ICA.

Similarly, Drummond et al. (2003) performed a study of estimation grain yield using a dataset that was composed of soil properties and topographic characteristics inferred from stepwise multilinear regression (SMLR), projection pursuit regression (PPR), and several types of neural networks for 10 "site-years". The results indicated that the neural networks performed well after evaluating the model while compared with the

SMLR and PPR in the every-site year. Another study disseminated the results using a hybrid approach (crop modeling + ML) to predict corn yield (Shahhosseini et al., 2021). The five ML models (linear regression, LASSO, LightGBM, random forest, and XGBoost) and six ensemble models were used for prediction. Overall, they achieved the best accuracy rate after adding the simulation crop model variables to the ML model. Further concluded that adding weather information alone is not sufficient for model prediction to the ML, hydrological input parameters are needed for enhancing yield prediction rates. Whereas in non-homogenous regions, Linear discriminant analysis (LDA) performed better than other trained models, including logistic regression, K-nearest neighbor, decision tree, naïve Bayes, and SVM (Mupangwa et al., 2020). Adding as many climate input parameters to the crop models improves the accuracy of yield estimation.

Table 7. List of various models used for crop yield prediction.

S. No.	Literature	Year	Crop Type	Model	Proposed model
1	Kang et al.,	2020	Maize	LASSO, SVR, RF, XGBoost, LSTM, and CNN	XGBoost
2	Stas et al.,	2016	Crop yield	BRT, SVM	BRT
3	Nosratabadi et al.,	2020	Crop yield	ANN-ICA, ANN-GWO	ANN-GWO
4	Drummond et al.,	2003	Crop yield	SMLR, PPR, Neural Networks	Neural Networks
5	Shahhosseini et al.,	2021	Corn	(crop modeling + ML); [linear regression, LASSO, LightGBM, random forest, and XGBoost]	crop modeling + ML
6	Mupangwa et al.,	2020	Crop yield	LDA, logistic regression, K-nearest neighbor, decision tree, naïve Bayes, and SVM	LDA
7	Khaki & Wang	2019	Maize	DNN, Lasso, shallow neural networks (SNN), and regression tree (RT)	DNN
8	Marko et al.,	2016	Soyabean	Weighted histograms regression, Conventional regression algorithms	Weighted histograms regression
9	Russello et al.,	2018	Crop yield	CNN	CNN

4.3.7. Emerging Paradigm: Earth Observation Big Data and Analysis Cloud Platforms

Access to free, time series, global moderate to high-resolution satellite remote sensing data has progressively increased over the last four decades, starting with AVHRR (1981 to present) and MODIS (1999 to present), followed by the publication by USGS of global Landsat (30m resolution) data Archive (1970 to present) in 2008/2009 under US open data initiative, and by the European Space Agency of Sentinel satellites high-resolution data (10m, 20m, and microwave; 2015 onwards). A host of other public and private satellites also now provide access to high resolution (up to sub-meter) time series Earth Observation (EO) data (Table 8), making the current EO data pool vastly different from a decade ago. These developments enabled scientists, businesses, and policymakers in various domains, including agriculture, to visualize the enormous potential of time series EO data to address a wide range of important environmental, economic, and social problems at local, regional, and global scales. But, the growing variety and volume (several petabytes) of EO data far exceeds the memory, storage, and processing capabilities of traditional remote sensing data storage, distribution, and processing locally on personal computers. It is also not technically feasible to carry out mandatory pre-processing of long-time series of raw EO data in the traditional paradigm. These factors have limited EO data used for application development in the traditional paradigm to only very small portions of available data.

The major hurdle in working with time series global scale EO data from diverse sources is in providing the proper connections between data, applications, and users. Overcoming this hurdle necessitates a paradigm shift in EO data retrieval, storage, and analysis from local processing on personal computers towards (i) adoption of next-generation infrastructures based on cloud platforms and big data technologies for data storage and processing, and (ii) for automating the pre-processing stages of raw EO data into Analysis Ready Data (ARD). Analysis Ready Data are *time-series stacks of satellite imagery* that are ready for a user to analyze with minimal or no additional pre-processing of the imagery. They are a packaged product created after pre-processing raw EO data through multiple standard stages that include: **searching and downloading** data from various providers, image fusing, clipping the data to cover only the area of interest, correcting for geometry, sensors, radiometry, and atmosphere, identifying pixels shadowed by clouds or with poor quality data, and lining up the images pixel for pixel, by geospatially co-registering and resampling the data. Developing long-term continuous ARD sets that are consistent (over time from the same sensor and across multiple sensors) is a work in progress that is evolving with new developments in both sensor technologies and algorithms.

The idea of ARD has also shifted the burden of pre-processing EO data from individual users to data providers and lowered technical barriers for users to fully utilize EO data. Usually, ARD is provided as tiled interoperable, registered stacks of both Top of Atmosphere (TOA) reflectance and atmospherically corrected surface reflectance products, and with explicit quality assessment information and appropriate metadata for traceability. Users then work directly with pre-processed and arranged data in a coherent time series stack for their area of interest instead of a bunch of randomly placed overlapping images. In the new paradigm, users also can significantly increase the scope and limits of their analysis by working with powerful cloud-computing platforms (instead of personal computers) and advanced analyses with time series, ML, AI, or other models to address complex problems.

Several approaches to EO big data infrastructures for storage and processing on cloud platforms have evolved to enable and accelerate applications in different domains in recent years. Their common goals include helping users to achieve: (i) easier access to EO data, (ii) easier use (storage and processing) of EO data in the cloud, (iii) easier EO big data analytics in the cloud, and (iv) better usability through tailored imagery web applications. Google was among the first to enable the shift towards using EO big data cloud platforms when it introduced the Google Earth Engine (GEE) in 2010 to enhance the use of satellite imagery for large-scale and time series applications. GEE set a benchmark in enabling universal access to its high power cloud computing resources for fast retrieval and processing of time series ARD from diverse sensors (nearly all sensors and gridded products in Table 8). In addition to ARD of multiple satellite data, the platform provides: (i) a large repository of other geospatial data, including environmental variables, weather, and climate forecasts, land cover, topography, and socioeconomic data, and (ii) a portfolio statistical, ML and AI tools, for a wide array of applications (Gorelick et al., 2017). Its library comprises 600 + EO analysis-ready datasets and 1000+ analytical tools. Each data source available on GEE has its own time series of EO/ARD data organized into a stack called Image Collection. Users can access Google Earth Engine with only an active internet connection and a Google account (250 GB free quota). With regular updates of EO data, tools, and features, the platform adapts to a wide range of user requirements and expertise. Users can also analyze their private data on the GEE platform with the help of backend data and analytical tools. A summary of features of GEE and some other currently available EO data cloud platforms for potential use in crop monitoring and insurance is given in Table 8.

Table 8. Cloud-based platforms for EO data access and analysis.

Platform	Satellite Datasets	Access	Special features
Google Earth Engine (GEE), 2010 (operated by Google)	Near real-time and archived Global ARD Datasets with corresponding cloud masks; derived time series products of major biophysical variables (NDVI, EVI, etc.); Gridded weather data products ; (nearly all sensors and gridded data sets of Table 5 are included on the platform)	Free and unlimited access to all public sensor time-series data and storage for research, education, and non-profit use; Limited access to some private satellite datasets (Planet; MAXAR)	Users can leverage storage and computing resources of the GEE platform; allows scaling to large regional and global analysis of time series data on the platform; open-source code for extracting data and processing with a range of statistical and ML/AI algorithms; analysis results can be displayed on the fly in GEE browser and can also be extracted to user systems for integration with other data; users can contribute own datasets and algorithms and develop mobile apps with GEE platform data and analysis tools; sufficiently user-friendly for non-specialists in RS or ML/AI tools; Closed source software, so cannot guarantee reproducibility as source code can change
Geospatial Big Data Platform GBDX (Maxar), 2018	ARD of MAXAR (WorldView) data; and open data of Sentinel and Landsat	Access based on the purchase of subscription; Also offers a free Community Edition called GBDX Notebooks that gives free access to	Leverages Amazon Web Services (AWS) to deliver scalable storage and computing resources that can be used for geospatial analytics and AI machine learning applications; Does not allow export of derivatives or the open images except some limited extraction in the Community Edition (

		open data (Landsat/ Sentinel) and some MAXAR data	6 GB instance and 20 GB of drive space).
Radiant Earth (Open source Platform of Radiant Earth Foundation, 2016)	Library of ARD of Sentinel-2 and Landsat data, ML algorithms, and Training data sets	Free access to data, algorithms, and training data sets for applications of ML/AI to support decisions in critical areas like agriculture, forests, disaster management for sustainable global development; Free access to available cropland data to develop crop masks	Target audience Global Development Community (NGOs, Academics, entrepreneurs); Not designed to scale to large regional analyses; Focus on localized studies for ML applications; Repository of Training Data sets for ML algorithms; Allows maintaining own personal projects and bringing in additional EO and secondary data; Allows sharing of data, training data sets, and algorithms with the community.
Sentinel Hub Playground (Sinergise)	ARD of Sentinel, Landsat- MODIS and DEM	Flexible pricing from free to basic and enterprise-level uses; Free access to explore and download satellite imagery for non-commercial/ research use; Paid access through specific protocols and API, data processing, mobile application data access, higher access limits	Between GBDX and GEE in function; Limited variety of EO data; Closer to GEE in terms of free access to data, analysis tools, and direct display of results in the browser; Allows customized analysis scripts but not sharing of scripts among users; Closed source code, so cannot guarantee Reproducibility; presents the best balance between the analyzed capacities. The drawback of the ODC solution is mainly the lack of support for reproducibility of science, which is not found in the others either. On the other hand, the other capacities evaluated are at least partially met.
SEPAL (System for Earth Observation Data Access, Processing, and Analysis for Land Monitoring), 2018 FAO –a platform for forest and land monitoring	Open source ARD from GEE and directly from other sources (including Sentinel, Pleiades, WorldView, etc.)	Free access, largely meant for developing countries with limited access to satellite data resources.	More focused on infrastructure management and provision of tools for EO data analyses; So. big data challenges are not directly addressed; Combination of GEE (for EO data) and open-source software ORFEO Toolbox (open source remote sensing data processing software for multiple sources), R and others for analysis; GEE is used for data retrieval and the Amazon Web Services Cloud (AWS) is used for data storage and infrastructure for computing analyses.
Open Data Cube - ODC First Developed as Australian Geoscience Data Cube (AGDC, 2017); Modified to allow diverse	Data cubes (time series stacks) of ARD of Landsat-5/7/8, Sentinel-1/2, MODIS ALOS-1/2, ASTER DEM	Available under Apache 2.0 license as a suite of applications These repositories include .	Generic framework composed of a series of data structures and tools for organization and analysis of massive EO data sets; Open source code distributed through Github repositories which include web interface modules for data visualization, data statistics extraction tools as well as Jupiter

users, datasets, and national or regional use options	and others		notebooks with examples of access and use of indexed data in ODC; It does not allow sharing of applications and data Different National ODC implementations are operational in Australia, Switzerland, Kenya, UK, Taiwan, and many other countries

High-resolution, high-frequency, consistent, and more detailed time series satellite data-based crop monitoring is needed over extended periods for effective implementation of crop insurance schemes nationwide, including PMFBY. The data management and analysis challenges arising from the huge time series data volumes can be overcome only with new cloud computing infrastructures, technologies, and data architectures, such as those listed in Table 2. Among these, GEE is currently the most developed with access to more data and analysis resources. However, since it is a closed platform, it cannot guarantee reproducibility source code can change. Sentinel Hub has relatively limited EO data resources and also has a similar drawback as GEE for reproducibility. SEPAL is more focused on infrastructure management and the provision of tools for retrieving and analyzing EO data. It does not directly address the big data challenges of EO data storage and processing. Radiant Earth is focused more on small area studies and machine learning tools. ODC provides a toolkit to facilitate application development by the user. The ODC, SEPAL, and Radian Earth are open source platforms, and their code is available in open repositories. ODC is the only platform that gives a user direct access to data and its infrastructure and data processing capabilities. It provides public documents of the platform governance process and of how to create or incorporate new features into the platform, including new software tools. It allows high replicability, provides for high scalability for storage and processing, and the maximum opportunity for data access interoperability. ODC also uses distributed data storage to minimize data movement during processing so that processing occurs where data is stored. While GEE is the most useful ready to use platform or users with its multiple source ARD, the library of processing tools, but the transient nature of its code, and uncertainty of its availability for the long term raise questions about reliability for use in the nationally important long term public schemes like PMFBY. For such schemes, ODC may be the preferred platform for national schemes despite the considerable effort involved initially in building the platform because of its generic open-source, scalable framework, distributed data storage, data and storage scalability, and interoperability of data sources.

In India, most research and application development with Indian satellite data (Resourcesat series) has so far been in the traditional paradigm, processing only a small portion of available data for relatively few periods on personal computers. Some early attempts at developing ARD time-series stacks for public access and building a national ODC have only recently been initiated by ISRO.

4.3.8. Mobile Camera Platforms

Technology innovation is a significantly important aspect of agricultural development and productivity. The use of mobile ICT (information and communication technology) provides a more efficient and cost-effective method in agriculture to exchange knowledge and information with the farmers. Mobile technologies enable farmers to share information and receive tailor-made advice on farming practices. Therefore, farmers can benefit through mobile technology to access key information such as pest and disease reports, weather conditions, and market prices. In recent years, digital maps have been available to assimilate the land use and soil conditions, which are helpful for farmers to understand where important differences in soil nutrient stocks are found. According to Beza et al. (2017), mobile data technology can be used to improve services to farmers allowing them to enhance nutrient management. This can be achieved in three ways: **First**, mobile platforms are used to collect the basic information related to the field data and past organic and inorganic fertilizer information, and crop yields. This information can be used to understand the time-series responses of crop yield production to the fertilizer application. **Second**, mobile devices are used to geo-reference the field under consideration. Therefore, field location is then linked to soil maps, spatial and temporal weather patterns, and NDVI data to give location-specific fertiliser recommendations to the farmers. Finally, information from farmers and fields can effectively use to estimate the crop yields at smallholder scales. It can also improve communication between farmers and extension workers who are unable to visit the field every time.

The importance of internet technologies in precision agriculture has created a new concept called e-Agriculture, to develop various applications for farming surveillance. Nowadays, farmers have a great need for information from crop planning to final yield products. E-Agriculture is broadly integrated with GIS, remote sensing, and data from automated weather stations (AWS) or from systems and sensors used in location monitoring. These technologies allow farmers to access data and use it as specific information for better decision-making. There is a number of methods developed using mobile technologies for crop protection and diagnosis, plant disease, insufficient nutrient levels. For example, Plantix is a diagnosis application developed by PEAT GmbH (Berlin, Germany) that detects diseases, pests, and nutritional deficiencies using an image of the plant to be analyzed. The application uses image recognition and deep learning (DL) algorithms to detect damage in crops. BioLeaf (Foliar Analysis), developed at the Federal University of Mato Grosso do Sul (Campo Grande, Brazil), is an application that using image captured by mobile phones or uploaded from a photo gallery allows to identify automatically and in situ the regions of leaves with lesions caused by insects. Similarly, Jaywant College of Engineering & Management (Maharashtra, India) developed an E-agree application that provides a variety of services to farmers. These include detecting plant diseases, a marketplace that helps farmers to sell/buy products online, a weather reporting system that plays a crucial role in decision making, as well as soil types information to understand which crop is suitable for their fields. In addition, ADAMA Bullseye was developed by ADAMA Agricultural Solutions (Tel Aviv, Israel), has a database to help farmers identifying pests and diseases in rice, almonds, tomatoes, apple, watermelon, and cotton crops. For example, if farmers are not able to identify the disease in crops, they are allowed to send a picture of their crop for evaluation by quantified technicians. In order to identify disease detection, the Plant Disease app was developed by the Technical Educational Institue of Thessaly (Larissa, Greece) to identify vines disease through leaf photos with an accuracy greater than 90%.

5. Recommendation of the best technology for adoption by PMFBY

It would be appropriate to suggest a scalable mechanism that includes immediate short-term measures to continue the PMFBY implementation without interruption, with progressive improvements to its governance, and mid to long term seamless transition to institutional re-arrangements for technology-driven solutions for more effective implementation of PMFBY:

5.1. In the short run

1) Continue implementing the existing area-yield-index model based on 4 CCE (though these may not be adequate for GP level small area estimates), but with improved governance by following the Karnataka model of web platform based transparent planning and monitoring of four CCE/GP 12. The Karnataka model has shown the feasibility of 4 timely, transparent, and verifiable CCEs in all GPs of the State for all crops. This model is based on: (a) advance planning and display on the public web portal of randomly sampled CCE locations in each GP for all insured crops, for the entire State, (b) advance display of CCE schedules and state govt personnel allocated for each GP, and, (c) concurrent video capture and uploading of the entire CCE process on a web portal for the permanent record for verification by insurance companies, State Govt, farmers, and others. The above steps can be implemented on the present PMFBY portal or individual State Govt portals.
2) Carry out concurrent audits of individual farm yields and actual insurance payouts to farmers using farmer surveys at randomly sampled GPs, to assess the basis risk in the GPs, and ascertain the efficacy of claims settlement process (timely CCE and claim settlement; farmer satisfaction). An appropriate sampling design for assessing GP level basis risk for such audit will need to be identified.
3) For the GPs selected in (2) above, evaluate alternate proxies for GP averaged actual crop yield in the area-index insurance model (e.g., the weighted average of pixel-based vegetation indices composites) derived from high resolution (5-30 m) satellite data (e.g., Landsat, Sentinel, Resourcesat), and also assess basis risk relative to the area-yield- index.
4) For the GPs selected in (2) above, evaluate the use of high resolution (5-30 m) alternate yield proxies for estimating current season average crop yield using pixel-level satellite vegetation indices composites, information on other cofactors like weather and soil moisture, and analytics tools including ML and AI, as well as assess for reduction of basis risk compared to the present PMFBY area-yield-index.
5) Repeat (2) to (4) for all years of implementation of PMFBY (2016-cont.) using archived data.

Note that:

(1) and (2) above essentially refer to the present CCE process and area-yield-index, but with the wider adaptation of improved governance and accountability systems (already in use in some states), and (3) to (5) help to prepare the base in the medium term for the transition to innovative use of new technologies, analytics, cloud computing infrastructure, as well as new institutional systems and processes in the longer term.

5.2. For the mid to long-term

Upgrade MNCFC to an independent and autonomous NICFI with the required infrastructure and staffing to operationalize a countrywide, open-source, high-resolution crop monitoring platform. The institution must have the requisite infrastructure and multidisciplinary expertise in agricultural remote sensing; big data cloud infrastructures (preferably Open Data Cube (ODC) platform), technologies and standards; and generating multi-source analysis-ready high resolution key agricultural remote sensing time series data sets for crop monitoring, small area crop yield and yield loss assessments, insurance, statistics, and other relevant areas.

Essentially the NICFI Platform must generate the following products that insurance implementing agencies can use:

- Time series pre-processed ARD of multisource surface reflectance cloud-free composites at 10 to 30 m resolution or higher
- Monthly dynamic crop masks
- Weekly/biweekly/monthly vegetation status maps

The platform must be free and open-source to allow users to generate near real-time products tailored to their specific needs.

6. Resources and policy needs for implementation of the best technology

We employ a scoping review methodology in this study to explore the global technologies available for village-level crop yield estimation. The main objective of this research study is to inform this scoping review is: What types of government policies and schemes facilitate changes in agricultural yield production at the Grama Panchayat level? Our analysis provides an overview of published literature to identify key trends, deficits & gaps in insurance policies, modern approaches in yield estimates, scalability, and spatial and temporal comparison of crop strategies that are particularly confined to small-scale regions. To meet the requirements of this review, we categorize the types of:

- ❖ Technology-based intervention methods are inherently essential in agriculture precision technologies as they have been providing accurate analysis without compromising the cost and time.
- ❖ Best policies and programs used to encourage farmers for developing sustainable yield production of growing a particular crop
- ❖ Rapid evaluation of research methodologies are used to assess these policies
- ❖ Outcomes of these schemes must compare with already existing methods in order to draw a conclusions
- ❖ Time frame assessment should be sufficient to capture policies and programs implemented in insurance policies and eliminate quantitative restrictions

In a particular context, policies and programs are established in relation to political, social, and economic concerns. Understanding context can contribute to policy implementation, uptake and impact by identifying policy levers. Efficient agricultural policies are essential to meet the increasing demand for food production sustainably; therefore, policy monitoring and evaluation provide needed evidence for governments to ensure that their agri-food policies address these challenges well. Agricultural policy packages need to be coherent and efficient to enable the sector to develop its full potential and achieve key public policy objectives. But today's farmers, advisors, and policymakers have difficulty in choices due to a wide range of technologies are available or under development. On the other hand, it is also challenging to comprehend or compare the regional assessment studies with other countries based on crop type, weather patterns/climate, socio-economic factors, and population. *Technology adoption* is a broad concept, and it is greatly influenced by the implementation, development, and application at the field level compared with the existing traditional methods. It can also be affected by education, training, and imparting information to the farmers, which form on the basis of farmer's knowledge. Most of these technologies originated from outside the farm sector; their effective use may not be substantial in the whole-agri-food-sector at the farm level. The adoption of technologies for sustainable management of farming systems is challenging for farmers, insurance agencies, policymakers, and government sectors. The agricultural sector needs to employ various emerging technologies and farming practices across different farming systems and structures in order to meet the demand from consumers and the public for food.

6.1. Policy aspects

There are certain areas of the policy aspects that need to be considered:

- The overall policy framework should be relatively consistent and coherent: This requires a more integrated approach in terms of formulating objectives, research design, plan, implementation, development priorities, and targeting policy measures at the appropriate level
- Marketing challenges and consumer requirements: The policies need to be verified with the international trading environments, world market conditions, global competitiveness, infrastructure, and ultimately based on users tastes
- Tracking technology adoption system for sustainable agriculture: There is a need for greater follow up to track adopting technology could help ensure that corrections are made before too much is invested in the wrong way
- A more extensive participatory methodology will work on adapting technologies for sustainable cultivating frameworks, including the scope of stakeholders.

References:

M. Stas, J. Van Orshoven, Q. Dong, S. Heremans and B. Zhang, "A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series SPOT-VGT," 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2016, pp. 1-5.

Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J. & Kitchen, N. R. Statistical and neural methods for site-specific yield prediction. Trans. ASAE 46, 5 (2003).

S. Nosratabadi, K. Szell, B. Beszedes, F. Imre, S. Ardabili and A. Mosavi, "Comparative Analysis of ANN-ICA and ANN-GWO for Crop Yield Prediction," 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), 2020, pp. 1-5.

Ramos, A.P.M., Osco, L.P., Furuya, D.E.G., Goncalves, W.N., Santana, D.C., Teodoro, L.P.R., Junior, C.A.S., Capristo-Silva, G.F., Li, J., Baio, F.H.R., Junior, M., Teodoro, P.E., Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. Computers and Electronics in Agriculture, 178, 105791.

Chlingaryan, A.; Sukkarieh, S.; Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151, pp: 61-69.

Pantazi, X.-E.; Moshou, D.; Alexandridis, T.K.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 2016, 121, 57–65.

Kung, H.-Y.; Kuo, T.-H.; Chen, C.-H.; Tsai, P.-Y. Accuracy Analysis Mechanism for Agriculture Data Using the Ensemble Neural Network Method. Sustainability 2016, 8, 735.

Subhadra, M., Debahuti, M., Gour Hari, S., 2016. Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper. Indian J. Sci. Technol. 9.

Craninx, M.; Fievez, V.; Vlaeminck, B.; De Baets, B. Artificial neural network models of the rumen fermentation pattern in dairy cattle. Comput. Electron. Agric. 2008, 60, 226–238.

Fortin, J.G., Anctil, F., Parent, L., Bolinder, M.A., 2011. Site-specific early season potato yield forecast by neural network in Eastern Canada. Prec. Agric. 12, 905–923.

N. Gandhi, O. Petkar and L. J. Armstrong, (2016). "Rice crop yield prediction using artificial neural networks," 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2016, pp. 105-110.

Khosla, E.; Dharavath, R.; Priya, R. (2020). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environment, Development and Sustainability (2020) 22:5687–5708.

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environ. Res. Lett. 15 064005.

Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S.V. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci Rep 11, 1606 (2021).

Mupangwa, W., Chipindu, L., Nyagumbo, I., Mkuhlani, S. & Sisito, G. Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. SN Appl. Sci. 2, 952 (2020).

Khaki, S. & Wang, L. (2019). Crop yield prediction using deep neural networks. Front. Plant Sci, 22.

Marko, O., Brdar, S., Panic, M., Lugonja, P., and Crnojevic, V. (2016). Soybean varieties portfolio optimisation based on yield prediction. Comput. Electron. Agric. 127, 467–474.

Russello, H. (2018). Convolutional neural networks for crop yield prediction using satellite images. IBM Center for Advanced Studies.

Azzari, G.; Jain, M.; Lobell, D.B. (2017). Towards fine-resolution global maps of crop yields: Testing multiple methods and satellites in three countries. Remote Sensing of Environment. http://dx.doi.org/10.1016/j.rse.2017.04.014.

Vancutsem, C.; Marinho, E., Kayitakire, F.; See, L.; Fritz, S. (2012). Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. Remote Sens., 5(1):19-41.

Lebourgeois, V.; Dupuy, S.; Vintrou, E.; Ameline, M.; Butler, S.; Begue, A. (2017). A combined random forest and obia classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, VHRS, and DEM). Remote Sens., 9(3), p. 259.

Debats, S.R.; Luo, D.; Estes, L.D.; Fuchs, T.J.; Caylor, K.K. (2016). A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens. Environ., 179, pp:210-221.

Lambert, M.J.; Sibiry Traore, P.C.; Blaes, X.; Baret, P.; Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sensing of Environment, 216, pp:647-657.

Burke, M. and Lobell, D.B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder african systems. Proc. Natl. Acad. Sci., 114 (9), pp: 2189-2194.

Srivastava, A.K.; Mboh, C.M.; Gaiser, T.; Webber, H.; Ewert, F. (2016). Effect of sowing date distributions on simulation of maize yields at regional scale a case study in central Ghana, West Africa. Agric. Syst., 147, pp:10-23.

Wit, AJW & Diepen, C.A. (2008). Crop growth modeling and crop yield forecasting using satellite-derived meteorological inputs. International Journal of Applied Earth Observation and Geoinformation, 10, pp:414-425.

Kasampalis, D.A.; Alexandridis, T.K.; Deva, C.; Challinor, A.; Moshou, D.; Zalidis, G. (2018). Contribution of Remote Sensing on Crop Models: A Review. J. Imaging, 4, 52. https://doi.org/10.3390/jimaging4040052.

Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolking, S.; Li, C.; Salas, W.; Moore III, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sensing of Environment, vol (95), 480-492.

Vikram Puri, Anand Nayyar & Linesh Raja. (2017). Agriculture drones: A modern breakthrough in precision agriculture, Journal of Statistics and Management Systems, 20:4, 507-518, DOI: 10.1080/09720510.2017.1395171.

Tripicchio, P.; Massimo, S.; Dabisias, G.; Ruffaldi, E.; Avizzano, C. A. (2015). Towards smart farming and sustainable agriculture with drones. International Conference on Intelligent Environments. DOI 10.1109/IE.2015.29.

Schlenker, W., Lobell, D.B. (2010). Robust negative impacts of climate change on African agriculture. Environ. Res. Lett. 5, 014010. https://doi.org/10.1088/1748-9326/5/1/014010.

Li, Y.; Guan, K.; Yu, A., Peng, B.; Zhao, L., Li, B.; Peng, J. (2019). Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US Field Crops Research, 234 (2019), 55-65.

Kern, A.; Barcza, Z.; Marjanović, H.; Árendás, T.; Fodor, N.; Bónis, P.; Bognár, P.; Lichtenberger, J. (2018). Statistical modeling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agricultural and Forest Meteorology, 300-320.

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., Wu, W. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. Agric. For. Meteorol. 204, 106–121.

Ferencz, Cs., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, Gy., Timár, G., Molnár, G., Pásztor, Sz., Steinbach, P., Székely, B., Ferencz, O.E., Ferencz-Árkos, I. (2004). Crop yield estimation by satellite remote sensing. Int. J. Rem. Sens. 25, 4113–4149.

Bolton, D.K., Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. Forest Meteorol. 173, 74–84.

Shen, Y., Shen, G.; Zhai, H., Yang, C., Qi. K. (2021). "A Gaussian Kernel-Based Spatiotemporal Fusion Model for Agricultural Remote Sensing Monitoring," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3533-3545.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G. (2010). "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610-2623.

Liao, C., Wang, J., Pritchard, I., Liu, J., Shang, J. (2017). "A Spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions," *Remote Sens.*, vol. 9, no. 11, pp. 1125.

Rauff, K.O.; Bello, R. (2015). A review of crop growth simulation models as tools for agricultural meteorology. Agric. Sci, 6, 8.

Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, HCJ; Herrero, M.; Howitt, R.E.; Janssen, S.; Keating, B.A.; Munoz-Carpena, R.; Porter, C.H.; Rosenzweig, C.; Wheeler, T.R. (2017). Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agric. Syst, 155, 269–288.

Challinor, AJ; Wheeler, T.R.; Craufurd, P.Q.; Slingo, JM; Grimes, D.I.F. Design and optimization of a large-area process-based model for annual crops. Agric. For. Meteorol. 2004, 124, 99–120.

Li, T.; Angeles, O.; Marcaida, M.; Manalo, E.; Manalili, M.P.; Radanielson, A.; Mohanty, S.(2017). From oryza2000 to oryza (v3): An improved simulation model for rice in drought and nitrogen-deficient environments. Agric. For. Meteorol, 237–238, 246–256.

Wang, D.W. (2008). Methods and Application of Assimilating Remote Sensing Data and Crop Growth Model. Beijing Normal University, Beijing, China, pp. 1–141.

Lobell, D.B. & Burke, M.B. (2010). On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol, 150, 1443–1452.

Kern, A.; Barcza, Z.; Marjanovic, H.; Arendas, T.; Fodor, N.; Bonis, P.; Bognar, P.; Lichtenberger, J. (2018). Statistical modeling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agricultural and Forest Meteorology, 260-261, pp:300-320.

Mishra, A.K.; Ines, A.V.M.; Das, N.N.; Khedun, P.; Singh, V.P.; Siva Kumar, B.; Hansen, J.W. (2015). Anatomy of a local-scale drought: Application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study. Journal of Hydrology, 526, pp:15-29.

Zhao, Y.; Chen, S.; Shen, S. (2013). Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. Ecological Modelling, 270, pp:30-42.

Revill, A., Sus, O., Barrett, B., Williams, M., 2013. Carbon cycling of European croplands: a framework for the assimilation of optical and microwave Earth observation data. Remote Sens. Environ. 137, 84–93.

Wit, AJW & Diepen, C.A. (2007).Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Agricultural and Forest Meteorology, 146, pp:38-56.

Moulin, S., Launay, M., Guerif, M. (2001). The crop growth monitoring at a regional combination of remote sensing and process-based models. In: International Symposium Crop monitoring and Prediction at Regional Scales, Tsukuba, Japan, February 2001.

Fang, H.; Liang, S.; Hoogenboom, G.; Teasdale, J.; Cavigelli, M. (2008). Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. Int. J. Remote Sens. 32(4), 1039-1065.

Orly, E.A.; Manuel, P.R.; Jorge, M.G.; Joao, V.(2020). A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a deep learning technique. Front. Plant Sci., 15, 2020.

Stein, M., Bargoti, S., Underwood, J. (2016). Image-Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. *Sensors* 16, 1915. doi: 10.3390/s16111915.

Valente, J.; Doldersum, M.; Roers, C.; Kooistra, L. (2019). Detecting rumex obtusifolius weed plants in grasslands from UAV RGB imagery using deep learning. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W5.

Valente, J.; Sari, B.; Kooistra, L.; Kramer, H.; Mucher, S. (2020). Automated crop plant counting from very high-resolution aerial imagery. Precision Agriculture, 21:1366-1384.

Matt Wade. (2015). https://www.sensefly.com/blog/drone-surface-model-estimate-crop-yield-plant-health/

	Drones	**Satellite**	**Lidar**
Small scale	✓	✓	✓
Continental-scale	×	✓	✓
Resolution	✓	✓	✓
Cost	high	free/expensive	
availability			
Accuracy			

Main objectives of the study

Minimize the number of CCEs

Explore the technology-based yield methods

TOPS (Terrestrial Observation and Prediction System)

Terrestrial Observation and Prediction System (TOPS) is a new innovative approach to exploit the latest technologies to estimate crop yield.

1. Climate Variability Impact Index, defined as the monthly contribution to overall anomalies in growth during a given year, is derived from the 1 km MODIS Leaf Area Index.

3. spectral crop growth profile and biomass production models