“मशीन लर्निंग और फसल सिमुलेशन मॉडल द्वारा सरसों की उपज का अनुमान”

“MUSTARD YIELD PREDICTION BY MACHINE LEARNING AND CROP SIMULATION MODELS”

Mr. Avinash Goyal

Roll No. 10409

DIVISION OF AGRICULTURAL PHYSICS

ICAR-INDIAN AGRICULTURAL RESEARCH INSTITUTE

NEW DELHI-110012

2023

MUSTARD YIELD PREDICTION BY MACHINE LEARNING AND CROP SIMULATION MODELS

By

AVINASH GOYAL

A Thesis

Submitted to the Faculty of Post-Graduate School,
ICAR-Indian Agricultural Research Institute, New Delhi,
in partial fulfillment of the requirements
for the award of the degree of

DOCTOR OF PHILOSOPHY
IN
AGRICULTURAL PHYSICS
2023

Approved by the advisory committee:

Dr. Ananta Vashisth (Chairperson) ....

Dr. V. K. Sehgal (Co-Chairperson) ....

Dr. D. K. Das (Member) ....

Dr. Joydeep Mukherjee (Member)

Dr. K M Manjaiah (Member) ....

Dr. Shiv Prasad (Member) ....

Division of Agricultural Physics,
ICAR-Indian Agricultural Research Institute
New Delhi-110012, India

Dr. Ananta Vashisth,
Principal Scientist

CERTIFICATE

This is to certify that the thesis entitled **"MUSTARD YIELD PREDICTION BY MACHINE LEARNING AND CROP SIMULATION MODELS"** submitted to the Post-Graduate School, **ICAR-Indian Agricultural Research Institute, New Delhi,** in partial fulfilment of the requirements for the degree of **DOCTOR OF PHILOSOPHY in AGRICULTURAL PHYSICS,** embodies the results of bona fide research work carried out by **Mr. Avinash Goyal**, under my guidance and supervision, and that no part of this thesis has been submitted for any other degree or diploma.

It is further certified that any help or source of information availed during the investigation has been duly acknowledged by him.

Vash.W

Place: New Delhi
Dated: 14/02/2023

(Dr. Ananta Vashisth)
Chairperson
Advisory Committee

ACKNOWLEDGEMENT

It is a genuine pleasure to express my deep sense of thanks and gratitude to all those people whose kind help and guidance made this thesis a reality. I am beginning in the name of God whose blessings gave me the strength and perseverance to complete this study.

I wish to express my deepest sense of gratitude and indebtedness to Dr. Ananta Vashisth, Principal Scientist, Division of Agricultural Physics, as Chairman of my Advisory Committee for her invaluable guidance, constant encouragement, scholarly suggestions, untiring enthusiasm, affectionate behaviour, useful discussion and peerless criticisms during the course of investigation and preparation of manuscript. Her kindness and devotion left an indelible impression in my mind.

I express my esteem and profound gratitude to Dr. (Mrs.) P. Krishnan, Head and professor, Division of Agricultural Physics for providing me the necessary facilities, for her valuable cooperation and help during the course of my studies as well as research work.

I express my gratitude to Dr. V. K. Sehgal, Ex Professor, Division of Agricultural Physics and co-chairman of my advisory committee for providing me the necessary facilities, for her valuable cooperation and help during the course of my studies as well as research work.

It is a great privilege for me to express my esteem and profound gratitude to Dr. Rajkumar Dhakar, Division of Agricultural Physics for his cooperative attitude, constructive and valuable suggestions and help throughout the program of this work.

I convey my respectful sincere thanks to the members of my Advisory Committee, Dr. Joydeep Mukherjee, Division of Agricultural physics, Dr. D. K. Das, Division of Agricultural physics, Dr. K M Manjaiah, Soil Science and Agricultural Chemistry, Dr. Shiv Prasad, Centre for Environment Science and Climate Resilient Agriculture, for his concrete suggestions and commendable help throughout the experimental period.

I express my profuse thanks to all other scientists of Division of Agricultural Physics for their help that they extended to me during my research work.

I am extending my sincere gratitude to technical, administrative and supportive staff members especially to Mrs. Vipin mam, the Division of Agricultural Physics, IARI for their support which has helped me to complete my study in the division.

I sincerely acknowledge the help and support by Dr. D. K. Joshi and Ishwar Chand Sir, in monitoring the field experiments and collecting in-situ observations.

I express my gratitude to my dear seniors Bappa Sir, Amit Sir, Gaurav Sir, Rekha ma'am at ICAR-Indian Agricultural Research Institute, New Delhi for their great company in making my days of Ph.D. memorable and for keeping my spirits high during difficult times.

The unceasing help received from my beloved juniors Vikas, Ramnarayan, Brijesh, Kaushik, Alka, Jai Prakash Bisen, Bhoopesh Punera, Ashish will be affectionately remembered.

I express my heartfelt thanks to my dear friend Mr. Nitin Garg, Mr. Kamlesh Soni, Mr. Mahesh Kumar, Mr. Chirah Maheshwari, Mr. Sushil Kumar, Neeraj Budhlakoti, and Mr. Utkarsh Tiwari for extending their help on numerous occasions.

I am in a dearth of words to express my feelings of indebtedness and gratitude towards my family for all their boundless affection, moral support, eternal love, deep concern, prayers, and personal sacrifices that sustain peace in my life.

I sincerely thank PG School, ICAR-Indian Agricultural Research Institute, New Delhi.

I sincerely acknowledge the support of the entire faculty of the Division of Agricultural Physics & NRL. during my course of study.

Place: New Delhi

Date: 14/02/2023

Avinash Goyal

Avinash Goyal

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

1. Introduction

About 60 percent of people in India are directly or indirectly associated with agriculture. It is a big challenge faced by any government to provide sufficient food supply to the people, especially in areas with the continuing expansion of population and shrinkage in agricultural land. In the last few decades, the focus has been on increasing food grain production to attain self-sufficiency. But unfortunately, oilseed production did not fulfill the demand of the ever-increasing population. So, to achieve self-sufficiency in oilseed production, one should know about the response of the oilseed crops to existing and changing environmental conditions. Oilseed crops represent one of the major groups of crops and play an imperative role in India's agricultural and industrial financial system in India. Rapeseed-mustard is the second most important edible oilseed after groundnut sharing 27.8% of the oilseed economy in India (Mandal *et al.*, 2006).

Crop yield is affected by extreme weather events and climatic variability. It is desirable to develop accurate and dynamic crop yield prediction models to overcome considering the challenge of food security at the domestic and international levels. Accurate information about the history of weather variables and crop yield is useful for making decisions related to agricultural risk management and future predictions. An accurate and timely prediction of crop production with a longer lead time can be very useful, depending on the scale of applications. It allows an agricultural producer to take more informed in-season corrective crop management and financial decision. The government policymaker can take actions such as stocking food supply and strategic resource mobilization in the most insecure areas. Many agricultural industries increasingly rely on crop market outlooks and yield prediction for their decision-making.

Conventionally, the crop yield estimation is based on nationwide crop-cutting experiments (CCE), these results are aggregated at various administrative units, but these estimates come after the harvest. Crop-cutting experiments are laborious, time-consuming, and require more accuracy. Therefore it is important to devise a methodology for crop yield prediction at multi-stage before crop harvest. This will be helpful for taking informed decisions for planners and resource managers. Mainly empirical statistical models and crop simulation models have been used for crop yield prediction (Bocca and

Rodrigues, 2016). Several researchers developed several prediction models based on statistical techniques by time series data. However, time series data is often full of non-linearity and irregularity. Several machine-learning modern techniques has been used, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF), to overcome the problems of predicting non-linearity and non-stationary time series dataset. R statistical software was used to develop prediction models and compare the results of these techniques to determine the best one among them. Modern machine learning techniques are based on climatic variables and crop yield data with limitations in quantifying the absolute changes in crop biophysical parameters and crop yield losses. In contrast, crop simulation models used soil, plant, and climatic variables to monitor crop growth and estimate yields at regional scales. Comparing the theory of crop model and statistical models, crop models require more detailed data on crop growth, soil parameters, and management factors (Lobell and Field, 2007; Xiong *et al.*, 2007; Zhang *et al*., 2008), but in favor statistical models doesn’t need field and management data (Lobell and Burke, 2010). Song et al. (2011) recommended that the optimal combination of different empirical crop yield prediction models is a more reliable approach to overdrawing the limitation of individual approaches at regional scales.

Keeping these research gaps in mind following objectives of the research were formulated:

1. To calibrate and validate crop simulation model for mustard crop at experimental farm and farmer’s fields and evaluate it for mustard yield prediction

2. To develop and evaluate machine learning models for mustard yield prediction.

3. To design and demonstrate a yield prediction methodology by optimal combination of models.

In this study, mustard yield prediction was accomplished using a crop-simulation model in conjunction with various machine-learning techniques. First, the InfoCrop-mustard model was tested on an experimental research farm as well as on farmer fields, and a multi-stage mustard prediction was done for IARI, New Delhi. Second, variable selection and extraction along with ANN, SVM, and RF approaches was used for mustard yield prediction for IARI, New Delhi and five major mustard growing zones of Rajasthan. The R statistical software version 3.1.3 was used to create the best prediction model. To determine the best models for mustard yield prediction for study areas, different model accuracy parameters were used to compare the models. Finally, a variance-based optimum combination approach was used to predict mustard crop yield for study areas.

2. Review of Literature

The oilseed plays an important role in the economy of a nation. Oilseed crops represent one of the major groups of crops and play an imperative role in India's agricultural and industrial financial system. Rapeseed-mustard is the second most important edible oilseed after groundnut sharing 27.8% of the oilseed economy in India (Pradhan *et al*. 2014). The area occupies by rapeseed and mustard is approximately 6.02 million hectares, with a production of 7.98 million tones (DES, 2018). The decision regarding planting *Rabi* season crops depends upon the previous crop harvesting time. The growth and development of mustard crop is highly sensitive to weather variables by changing the microenvironmental conditions (Goyal *et al*., 2018). Weather is an important uncontrollable factor influencing crop growth and development. Mustard is mainly grown in the North-West part of India, semi-arid to sub-tropic regions having well-defined wet and dry winter seasons. Mustard is the most important oilseed crop grown in the *Rabi* season in the North-West part of India, and it is largely grown in Punjab, Haryana, Uttar Pradesh, Bihar, and Rajasthan State. The ideal climate type required for the mustard crop is semi-arid and sub-tropical, with a distinct wet season and a dry winter season. In India, mustard is sown mainly during the first week of October however it is often shifted to the first week of November. Mustard seed yield in different geographical areas are highly related to the spatial variability of weather. Weather is dynamic, continuous, and multi-dimensional. These unfavorable properties make weather prediction a formidable and challenging task for meteorologists. For effective weather prediction modeling, we required computer models, agro-meteorological observatories, and knowledge of patterns and trends of weather variables through historical records. Mainly two kinds of methods are used widely for weather prediction: empirical approach and dynamical approach; empirical models, are generally used for efficient local-scale weather prediction. Many prediction systems combine empirical dynamic approaches for large-scale weather prediction, whereas remote sensing data can effectively be used for large-scale weather prediction. Remote sensing has limitations in quantifying the absolute changes in crop biophysical parameters and crop yield losses.

Statistical and crop models are two important tools for detecting the change in observed yield with predicted yield. Crop models can integrate the knowledge of crop physiology, soil science, agronomy, and meteorology into the models, representing a set of mathematical equations to describe physical, physiological, and chemical processes to simulate crop growth. Comparing the theory of crop model and statistical models, crop models require more detailed data on crop growth, soil parameters, and management factors (Lobell and Field, 2007; Xiong *et al*., 2007; Zhang *et al*., 2008), but in favor statistical models doesn't need field and management data (Lobell and Burke, 2010). The variation in crop yields from location to location and year to year is due to the change in crop growth and development influenced by a change in weather spatially and secularly, making yield estimation a baffling process. The prediction of yield differs spatially and temporally depending on the objectives, such as the yield estimation for a research field, farm, village, or district, and that for a week, month, or longer before harvest. For the prediction of yield on the regional scale, the Directorate of Economics and Statistics of the Ministry of Agriculture follows traditional survey techniques such as crop-cutting experiments.

The results related to different aspects of delay in sowing, crop yield prediction by a weather-based statistical model, crop simulation model, and their optimal combination obtained by earlier workers at the global level for different crops under different treatments were reviewed and presented in the following sub-headings.

1) Effect of dates of sowing

Exact sowing time is one of the most important factors which directly plays a key role in crop yield improvement. Mustard cultivars have variable optimum sowing time; this provides the flexibility of mustard crops for adjusting in multiple cropping systems around the year. Moreover, sowing time in the case of mustard crops does have variable crop response. Thus, optimizing the sowing time for a mustard cultivar will help in exploiting the maximum yield potential of the cultivar. We have further discussed the various parameters get affected by sowing time.

1.1) Growth and development

The thermal environment of the cultivar gets affected by a change in sowing time which ultimately affects the growth and development of the cultivar, leading to a change in life cycle completion (Adak *et al*., 2011a). The plant's early completion of the life cycle due to delayed sowing is because of higher temperature at the latter part of plant growth, which leads to forced maturity. Similarly, compared to early and regular sowing, late-sown crops accrued 9-20 percent lower GDD over the entire crop duration period (Adak *et al*., 2011b). The heat needed to achieve any phenological event in mustard crops differed from cultivar to cultivar, and its accumulation decreased as sowing was postponed (Chand *et al*., 1995). With delayed sowing, Das *et al*. (2009) observed longer vegetative and shorter reproductive stages in mustard. The vegetative stage of a late-sown crop was subjected to lower temperatures and less solar radiation, necessitating a longer period to accumulate a given number of growing degree days. With each delay in sowing, the overall crop length decreased. Temperature increases in the second half of February damaged the productivity of late-planted mustard crops (Ghosh and Chatterjee, 1988). Each crop requires a particular number of growing degree days for the vegetative to reproductive phase. Due to the prevalence of higher temperatures and longer sunshine hours during the post-sowing season, early sowing (1st October) absorbed required growing degree days in relatively less time (Kanth *et al*., 2000).

Weerakoon and Somaratne (2011) used ten mustard accessions (AC 501, 515, 580, 790, 1099, 1814, 2122, 5088, 7788, and 8831) in a study at Nagollagma, Sri Lanka, to determine the relationship between growth and yield during two growing seasons Maha (October to March-receives North eastern showers) and Yala (April to September-receives South western showers). During the Maha season, they obtained the highest yield from three mustard accessions (AC 580, 5088, and 7788), and AC 7788 produced the highest yield in the Yala season, demonstrating adaptability to seasonal variations. In *B. rapa*, delayed sowing increased the maturity phase (Liyong *et al*., 2007; Jun *et al*., 2007). The crop growth rate was found to be higher for normal sowing and lower for delayed sowing for all species. In any given crop growth season, crop growth and development rates are a function of energy receipt and thermal regime (Neogi *et al*., 2005). The higher crop growth rate associated with timely sowing was primarily due to a high leaf area index, which accumulated dry matter at a faster rate per unit leaf area per unit time, resulting in a decrease in dry matter with delay in sowing (Thurling, 1974). Early sown crops in *B. napus* (Robertson *et al*., 2002; Poureisa and Nabipour (2007) and *B. juncea* (Nanda *et al*., 1996). had a longer period of pod filling. Singh *et al*. (2002) stated that a crop's yielding potential is based on a greater proportion of biomass and yield being invested in variation in their edaphic and environmental conditions, which was accomplished by changing sowing dates. According to Pradhan *et al*. (2014), the vegetative period increased with sowing delay (49 days in early sowing, 62 days in normal sowing, and 66 days in late sowing), while the reproductive phase decreased with sowing delay (90 days in early sowing, 80 days in normal sowing and 58 days in late sowing).

1.2) Leaf area index (LAI)

The leaf area index (LAI) is a relevant parameter for leaf growth and development. The LAI is calculated as a ratio of the total one-sided leaf cover area to the ground cover area. LAI is crucial to subsequent crop growth, absorbed fractional PAR, biomass, and seed yield production of crops. Hence, LAI is the most important biophysical parameter used in crop yield prediction models (Krishnan *et al*., 2016), and climatic models (Waring and Landsberg, 2011). LAI is the highly sensitive input parameter in crop simulation models like Info-Crop. Thus, the accurate measurement of LAI is necessary for better crop vigor monitoring and use in modeling purposes and overall crop management practices (Wallach *et al*., 2001; Schirrmann *et al*., 2015). Alteration in the leaf area of a crop canopy may modify crop productivity, photosynthesis rate, and harvest index to a larger extent. The crop environment can be modified by sowing the crop on different dates anticipating reduced seed yield. Pradhan *et al*. (2014) reported a significant interaction between the date of sowing and cultivars concerning LAI and seed yield while working on mustard at ICAR-IARI, New Delhi. Allen and Morgan (1972) used oilseed rape as an experiment crop and found that higher LAI during the flowering period resulted to attain more seed yield, which was positively influenced by several pods/plant and the seeds/pod. Umburanas *et al*. (2019) worked on soybean crops in Parana State, Brazil, and concluded that the value of LAI reduced from 5.78 to 3.88 for late sowing. The mean temperature at the reproductive phase of mustard was increased due to delay in sowing irrespective of cultivars, resulting in reduced LAI (Kumar *et. al*., 2017). Panda *et al*. (2004) experimented on two mustard cultivars with three different dates of sowing and noticed a significant reduction in LAI under delay in sowing for New Delhi conditions. There was a significant reduction in the peak value of LAI from 6.14 to 4.27 with a delay in the sowing of winter maize at Odhisa (Kar and Kumar, 2015). Several researchers studied the mustard crop and found the same results for reduction in LAI under delayed sowing in India (Singh and Singh, 2002; Kumar *et. al*., 2000; Kurmi, 2002).

1.3) Seed yield

The yield of a crop varies with different cultivars and environmental conditions. It is a necessity to analyze the seed yield for different environmental conditions to overcome the losses. Delay in sowing for the *Rabi* season crop was mainly responsible for diminishing the crop's vegetative phase, advanced reproductive phase, and reduced seed yield and above-ground biomass (Thurling and Das, 1980). Tomar (1997) experimented on mustard at Tikamgarh and noted a significant reduction in seed yield for a late sown crop. Dhoble *et al*. (1997) worked in the Prabhani district of Maharashtra emphasized that the 1074 kg/ha seed yield was obtained for 18th October sowing as compared to 436 kg/ha for17th November sowing. An experiment on the mustard crop at Navsari china (Patel *et al.*, 1994) revealed that late sowing is responsible for seed yield reduction. Surekha and Reddy (1996) showed the performance of the mustard cultivar at Rajendranagar, Hyderabad, and reported that the 5th October sown crop had a higher yield than the 5th November sown crop. Pal *et al*. (1996) at Hissar, Gare *et al*. (1996) at Rahuri, Tuteja *et al*. (1996) at Raipur found similar results for mustard crops at different locations. There were nearly 40% losses in seed yield for different mustard cultivars if delayed in sowing time by one month from mid of October (Lallu *et al*., 2010). Different dates of sowing led to changes in the thermal environment of the experimental field resulting from variations in seed and growth parameters (Adak *et. al*., 2011a). Singh *et al*. (2001) revealed that there was about a 25% reduction in the seed yield of Indian mustard with delay in sowing. Pradhan *et al*. (2014) observed the significant interaction between the date of sowing and cultivars concerning seed yield of mustard and concluded that significant reduction in mustard seed yield (about 45%) in late sowing compared to early and normal sowing for all three cultivars at semi-arid environmental condition. Gupta *et al*. (2017) experimented at Jammu and found a significant reduction (about 42%) in mustard seed yield with delayed planting. A similar type of results was obtained by Roy *et al*. (2005) for the mustard crop, Kumar *et al*. (2008) for the soybean crop, and Prakash *et al*. (2010) for the cotton crop. Kar and Kumar (2015) noted 4.6 to 11.7% reduction in seed yield in delayed sowing of winter maize at Bhubaneswar, Odisha.

1.4) Biomass production

Since grain yield is highly dependent on photosynthesis partitioning towards grain filling after anthesis, studies on biomass development and partitioning are of greater importance to crop management. According to Thurling (1974), *Brassica campestris* accumulates about 85 percent of the total dry matter before anthesis, while *Brassica napus* accumulate just 50 percent. In a mustard crop, Rao (1992) found that 12 to 17 percent of dry matter production was accumulated before flowering and the rest after flowering under Delhi conditions. In a study conducted in Varanasi in 1982-83, Krishnamurthy and Bhatnagar (1998) found that dry matter (DM) increased with a positive correlation between DM and LAI at flowering and final DM development and that CGR increased until flower cessation and then dropped. Studies on *Brassica juncea* and *Brassica napus* revealed that when the crop was subjected to moderate temperature stress (28/15^{0}C) for a short period of 6 days to 7 days, dry matter production was unaffected, while when the crop was exposed to high-temperature stress (35/15^{0}C), dry matter production was significantly reduced. In a study conducted by Kar and Chakravarty (2001) on sandy loam soils in Delhi during the years 1993-94 and 1994-95, they found a 6 and 22% reduction in biomass production in Brassica crops sown in the first and third weeks of November, respectively, as compared to crops sown in the third week of October (cv. B.O.54 and Pusa Bold). Giri (2001) found that the October sown crop produced more dry matter (cv. Pusa Jaikisan) than the November sown crop in a field experiment conducted in Delhi. Singh *et al*., (2002) found the highest biomass allocation in leaves (59%), followed by stems and roots at the first flower appearance in semi-arid Haryana. Stems had the highest biomass at the start of seed filling, followed by leaves, roots, and siliquae, while at the end of seed filling and maturity, stems had the highest biomass allocation (43 to 59 percent), followed by siliquae (32 to 63 percent), and leaves (9.5 to 2%). Due to the delay in sowing, Dastidar and Patra (2004) observed a small delay in flowering and maturity. The biological yield could increase as the number of pods grows. Kar and Kumar (2015) revealed that 9.6 to 17.6% reduction in above-ground biomass in the delayed sowing of winter maize at Bhubaneswar, Odisha.

1.5) Radiation use efficiency

In plant growth determination, solar radiation is the key factor. According to Monteith (1994) relation between the solar radiation utilized by the crop and its biomass production rate is called radiation use efficiency. Radiation use efficiency is stated as the quantity of dry biomass above the ground or total dry matter produced per unit intercepted photosynthetically active solar radiation. The quality and intensity of intercepted radiation by the crop, control the growth, development, and yield. Out of the total solar radiation, nearly 50 percent of the solar radiation falls in the infrared region, 41 percent in the visible region, and 9 percent in the ultraviolet region Kahle *et al.* (2003). Solar radiation is an important component of photosynthesis, an energy source utilized for metabolic activities. The capability of interception of incoming radiation and conversion into biomass determined the crop growth in the field conditions (Gifford *et al.* 1984). Long *et al.* (2006) found that photosynthesis and yield potential, could be enhanced by up to 50 percent through effective light utilization. The important impression of double cropping on the resources specifies the variance between storable resources like water and non-storable resources like radiation. Carriglia *et al*., (2004) observed that Water use efficiency (WUE) and radiation use efficiency (RUE) was thoroughly linked. Variation in biomass accumulation was observed due to differences in radiation interception, which primarily depends primarily on LAI (Lindquist *et al*., 2005). Effective conversion of solar radiation interception and efficiency of plant leaf in the conversion of intercepted light into dry matter affects the plant biomass production. Absorption of incoming photosynthetically active radiations is significantly related to cropping geometry and LAI (Plenet *et al.*, 2000). Muchow *et al*. (1993) found that underneath optimum circumstances of water availability, the radiation use efficiency (RUE) rests almost persistent during most of the plant growth cycle and demonstrates no effects due to local atmospheric conditions. Factors such as air temperature, vapor pressure deficit, and water stress are the most common elements which affect the RUE.

However, crop productivity is not entirely dependent on resource capture during the growing season but also on resource use efficiency (Azam-Ali *et al*., 1994; Yang *et al*., 2004).

Pradhan *et al*. (2014) found a significant reduction of about 32 and 26% in radiation use efficiency in delayed sowing mustard crops compared with early and normal sowing, respectively. It may be due to the sharper reduction in above-ground biomass than the corresponding decrease in intercepted PAR. Kar and Kumar (2015) experimented on winter maize and reported that the peak IPAR was higher (89.3%) in the first sown crop, which might be attributed to higher biomass, LAI, and crop height for all cultivars. They also reported a significant reduction in radiation use efficiency with the peak value of 1.97, 1.85, and 1.60 g MJ^{-1} for the first, second and third sowing, respectively.

1.6) Water productivity

Water Productivity (WUE) is defined as a ratio of economic yield (seed or grain) or biomass production concerning crop evapotranspiration. In simple words, it is the amount of crop yield produced per unit of water uptake. Water deficit is one of the most significant abiotic stresses for crops, affecting growth, development, and yield (Micheletto *et al*., 2007). The weather variable during different phenological stages also changed by changing sowing dates for the crop. Since weather variables play an important role in crop evapotranspiration hence, crop evapotranspiration is affected by the delay in sowing. Pradhan *et al.* (2014) experimented on the mustard crop at IARI, a New Delhi research farm, and found that there was about a 15% significant reduction in evapotranspiration. He also concluded that the water use efficiency for delayed sowing was significantly ($p < 0.05$) lower than that of early and normal sowing crops. A significant gain in crop water productivity was found for the fields sown the earliest (maximal WUE around 3.5 kg-m^{-3}) compared to those sown the latest (minimal WUE around 1.5 kg-m^{-3}).

1.7) Oil content

The seed oil content varied under different environmental conditions which can modify by different sowing dates. Kumar *et al*. (2008) reported that a certain ambient temperature should be required for oil accumulation in an oilseed crop. The delay in sowing reduced seed yield and oil content in rapeseed (Pritchard *et al*. 2000; Ozer, 2003). Turhan *et al*. (2011) worked on two different rapeseed cultivars in Turkey and reported a 2.5% reduction in oil content with a delay in sowing. The mustard seed oil content showed a significantly drastic reduction with delayed sowing in the semi-arid condition in India (Adak *et. al*., 2011b; Sharma *et. al*., 2006). Several studies reported that delayed sowing is responsible for rising crop canopy temperature at the time of oil accumulation resulting in decreased oil content in mustard. Robertson *et al*. (2004) did an experiment on canola in Australia and reported that every 1°C increase in temperature above the normal date of sowing during flowering and grain filling reduced the oil percentage by 1.7. Tobe *et al*. (2013) revealed that normally sown canola had a significantly higher percentage of oil content than delay sown crop. Response of rapeseed oil content to changes in sowing date showed significant results; the mustard sowing on 30th September had the highest oil content (40.9%) compared to the 30th October sowing crop (37.7%) (Kumar *et. al*., 2008). Nazeri *et al*. (2018) reported that seed oil content had a drastic variation of about 4% in the late-sown crop for canola cultivars in Iran.

1.8) Harvest index (HI)

Panda *et al.* (2004) suggested that each delay date in sowing after 16th October decreased the harvest index. Lallu *et al.* (2010) observed that November sowing caused a significant reduction in HI (17.9%) as compared to October sowing (22.9 %). Afroz *et al.* (2011) recorded significantly higher HI with 10th November sowing (31.4%) as compared to 20th November (30.4%) and 30th November (29..4%) sowings. Kumari *et al*. (2012) observed that 10th October sowing resulted in significantly higher HI (22.9%) over 20th October sowing (22.2%) and 30th October (21.7%) sowing. Patel (2013), in a field experiment at Tikamgarh, Madhya Pradesh revealed that the 20th October sown crop exhibited a significantly higher harvest index (%) compared to the 4th November and 19th November sown crops.

2) Crop yield prediction

The variation in crop yields from location to location as well as year to year is due to the change in crop growth and development influenced by a change in weather spatially and secularly. It makes yield estimation/prediction a challenging task. The crop yield prediction and estimate are generally two different things. The prediction and final estimates depend on the timing of the release. The crop yield prediction is the art of predicting crop yields and production before the harvest occurs, typically a couple of weeks/months in advance. Prediction can be made before the harvest of the entire crop, whereas estimates are made after the crop harvest (Bouman, 1995). It is more important to consider the effect of the weather and climate in a prediction approach.

From the agricultural point of view, crop yield prediction is an important aspect of giving precise, scientific sound, and independent prediction of crop yield and production to make informed decisions by growers and to make timely import and export decisions by planners and resource managers (Horie *et al.*, 1992). There are various generic prediction methods, and some of them can be applied to crop yield prediction as well (Petr, 1991). Using a proper prediction method and getting reliable results with the same dataset depends on the vision of the predicter. No standard prediction methods available in the literature (Makridadis *et al*., 1998; Armstrong, 2001).

Presently, the crop forecasting activity was operationalized in the Ministry of Agriculture by establishing a center, Mahalanobis National Crop Forecast Centre (MNCFC) in April 2012. MNCFC provides crop production forecasts for 8 major crops through FASAL (Forecasting Agricultural output using Space, Agro-meteorology, and Land-based observations) program at national/state/district level. FASAL project is operational under the Ministry of Agriculture and Farmers Welfare, Govt. of India in collaboration with the Space Application Centre (SAC), Institute of Economic Growth (IEG) and India Meteorological Department (IMD). Under the FASAL scheme, IMD in a joint effort with Agromet Field units (AMFU) situated at distinctive State Agricultural Universities (SAUs), ICAR research institutes, and IITs, develops operational yield estimates for the *Kharif* and *Rabi* seasons utilizing statistical models. FASAL approach integrates inputs from various sources (Singh *et al*., 2017). It involves the econometric models based on time series data at the early stage of crop growth, agromet models during the mid-crop growth, and remote sensing based on forecasts after flowering and the pre-harvest stage of crop growth. An outline of FASAL model had been shown in Figure 2.1.

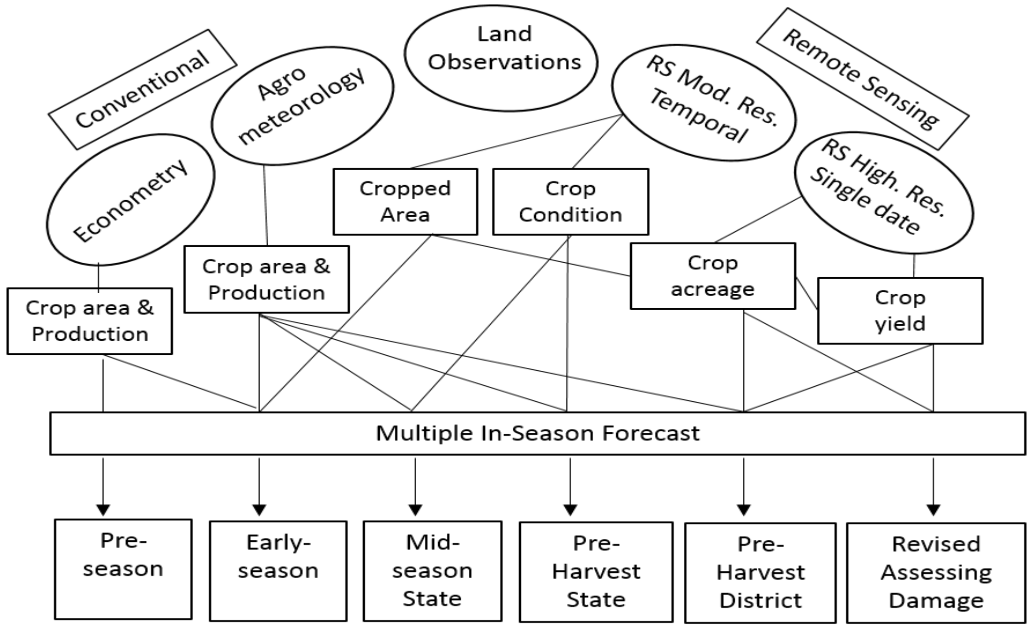

Figure 2.1: Forecasting agricultural output using apace, agro-meteorology, and aand-based observations (FASAL) program

Crop yield prediction methods can be classified into two broad categories: subjective and objective. The subjective approach is used traditionally by sample survey methods such as crop-cutting experiments conducted in India by the Directorate of Economics and Statistics of the Ministry of Agriculture. This method is time-consuming and labor-intensive. In contrast, objective methods rely upon the “models” which describe the plant environment interactions in quantitative terms. In general, there are mainly two approaches, empirical statistical models and crop simulation models to predict crop yield used by researchers considering the variability of weather and climate (Bocca and Rodrigues, 2016).

2.1) Empirical crop models

Traditionally empirical/regression statistical crop models had been developed to predict the nature of crop biometric parameters, economic yield, and above-ground biomass with the help of several factors such as agro meteorological or weather parameters, soil properties, and long-term time or technology trend data. Empirical statistical models are simple and require fewer input data and time than crop simulation models. It is a simple regression technique in which average weighting coefficients are used statistically to make crop weather-based models statistical models (Lobell and

Burke 2010; Shi *et al*. 2013). A long series of historical weather and yield data is required to develop a weighting coefficient to calibrate and validate the model. The empirical statistical model has limited applicability in space and time due to heterogeneity in weather parameters and soil properties of that particular place. The empirical approach doesn't have an explanatory nature or see the effect of a particular input parameter on the resulting output but is a very practical approach to assessing crop yield. Most agro-meteorological models do not respond well to extreme conditions weather. Prediction of crop yield using weather and management inputs into a statistical regression is relatively conventional and used in several research programs of crop yield prediction (NASS, 2006; Lobell *et al*., 2009). Empirical models can give many insights into past crop yields and historical effects and can be utilized to guide the other types of models (Lobell *et al*., 2011).

Fisher (1924) and Hendricks and Scholl (1943) have done pioneer work to develop a crop weather relationship model that requires a small number of weather parameters over a crop season. Fisher hypothesized that weather variables show a specific trend and don’t show abrupt changes within a week. Afterward, Hendricks and Scholl (1943) modified Fisher's method by considering the joint effect of weather variables. This existing model was again modified by Agrawal *et al*. (1986), including weight variables instead of individual weather variables. Crop–weather model was developed for turmeric yield prediction in Coimbatore district, Tamil Nadu, India with a coefficient of determination of 0.89 (Kandiannan *et al*., 2002). Statistical models used for groundnut and ragi in coastal Karnataka based on weather and yield data by Rajegowda *et al*. (2014). Kumari and Kumar (2014) used an ordinal logistic model based on weather data, predicting wheat (Triticum aestivum L.) yield in Kanpur district of Uttar Pradesh. The time-series data of temperature and yield were used to assess the impact of climate change on mustard crop in Haryana. (Shabnam *et al*., 2013). Lobell *et al*. (2013) used statistical models to determine the effects of increases in temperature on maize yield in the USA. Kazmi and Rasul, (2012) did a study on agro-meteorological wheat yield prediction in the rain-fed Potohar region of Pakistan. Verma (2015) did the pre-harvest yield prediction of mustard crop at the zonal level in Haryana through linear mixed models. Garde *et al*. (2015) used stepwise multiple linear regression techniques for pre-harvest wheat crop yield prediction for Varanasi district of eastern Uttar Pradesh. Agrawal *et al*. (2001) developed a weather based crop yield prediction model for wheat and rice in Madhya Pradesh by considering the effect of weather on crop yield. Dutta *et al*. (2001) developed a meteorological model for rice crops with 82% coefficient of correlation for 10 districts of Bihar and predict the crop yield. Pandey *et al*. (2015) developed a statistical model for rice crop in the eastern Uttar Pradesh region and also noticed the effect of single weather variables or a combination of them on crop yield. According to them, sunshine (hour) played a more important role in prediction than wind velocity and rainfall, whereas the joint effect of rainfall and wind velocity was more crucial than other combinations. Singh *et al*. (2014) predicted rice and wheat seed yield through a weather-based yield prediction model for the Uttar Pradesh region with good model accuracy in crop yield prediction. Giri *et al*. (2017) developed a regression equation for 6 districts of Madhya Pradesh for wheat and rice yield prediction with <10% standard deviation. Vashisth *et al*. (2014) used a weather-based statistical model for multistage crop yield prediction of wheat crop at 45 and 25 days before harvesting with 10.7% and 7% percentage deviation, respectively. Gupta *et al*. (2018) developed a statistical crop yield prediction model for mustard for six districts of western Uttar Pradesh with R^2 values ranging between 0.26 to 0.87.

Statistical modeling methods based on machine-learning algorithms can provide alternatives to traditional regression approaches and overcome some of their limitations. These machine-learning techniques belong to the class of algorithmic approaches, which assume the data mechanism as a ‘black box’ or ‘gray box’. Machine-learning techniques have been used increasingly in recent years as niche-based classification modeling tools for predicting species habitat suitability in response to climate change (Lawler *et. al*., 2006; Olden *et. al*., 2008). Researchers are using data-driven models to get accurate predictions by using the available data (Xing *et al*., 2018; Hund *et al*., 2018). In a data-driven model, Machine Learning (ML) algorithms play a major role in achieving better accuracy (Liu *et al*., 2017). This technique suffers an overfitting problem due to correlation in independent variables (Verma *et al*. 2016). So it is most important to overcome this problem by use of variable selection methods (e.g. stepwise multiple linear regression (SMLR), least absolute shrinkage and selection operator (LASSO), or elastic net (ENET) method) or variable extraction method (e.g. principal component analysis) before developing a crop weather model (Das *et al*. 2017).

More recently, machine learning techniques such as artificial neural networks, support vector machines, and random forest have been applied with variable selection and variable extraction techniques for crop yield prediction. The use of the most important variables, which play more than 90% role in model development is called variable selection, whereas variable extraction creates new variables by taking all possible combinations of factors. Variable selection techniques are generally used to remove correlated input parameters in domains where there are many features and comparatively few samples. Whereas variable extraction aims to crate heterogeneity in the number of input variables by creating a new variable from the innovative data set and characterizing it into a lower dimensional space (Kumar *et. al*., 2020).

Feature selection may be done as forwarding selection, backward selection, and a combination of both. Stepwise regression is a combination of the forward and backward selection techniques so that after each step in which a variable was added, all candidate variables in the model are checked to see if their significance has been reduced below the specified tolerance level. Step-wise multiple linear regression is an automated iterative procedure of regression is followed by the elimination of explanatory variables which have the weakest correlation. Zhang (2016) introduced variable selection with the stepwise regression technique as the best subset approach due to various information criteria. The most important technique for variable extraction is Principal component analysis (PCA) which removes homogeneity in variables and creates uncorrelated variables (known as Principal component (PC)). This method entails dimensionality reduction by principal component analysis (PCA) resulting in the most relevant components called 'PCs'. The linear regression is performed between the variable of interest (biophysical parameter) and PCs. The first few PC(s) can explain maximum variability in explanatory variables which are used as explanatory variables in the model and model parameters are easily estimated with reasonable accuracy. Azfar *et al* (2015) developed the model using principal component analysis of weekly data on weather variables for developing a rapeseed and mustard yield prediction model for Faizabad district of UP. Yield prediction were done for three subsequent years 2009-10 to 2011-12.

He reported that the model with six weather variables (maximum and minimum temperature, Morning and evening relative humidity, wind velocity, and sunshine hour) was most appropriate for providing yield prediction one and half months before the harvest. Annu *et al*. (2017) developed a wheat weather-based model in conjunction with MLR as a variable extraction technique.

A salient feature of machine learning models is that they treat the output (crop yield) as an implicit function of the input variables (genotypes and weather components), which could be a highly non-linear and complex mathematical function.

2.2) Yield prediction using artificial neural network (ANN)

In the recent scenario of prediction of yield and weather prediction, artificial neural network (ANN) getting a great deal of attention. Complex problems are solved by this method even if the quality of data is less or lacks precision in data. There are numerous introductory works on ANNs have been done by many workers. Cheng and Titterington (1994) have done a well-detailed study of neural network models vis-a-vis traditional statistical models. They have shown that some statistical methods including regression, principal component analysis, density function, and statistical image analysis can be given neural network expressions. Warner and Mishra (1996) reviewed the relevant literature on neural networks, clarified the learning algorithm, and made a comparison between regression and neural network models in terms of notations, terminologies, and implementation. Zhang *et al*. (1998) gave the overall summary of the work in ANN, providing the procedures for neural network modeling and the general paradigm of the ANNs, especially those used for prediction. The comparative performance of neural networks is studied with traditional statistical methods, and most of the studies concluded with a better performance of ANN over the traditional method. Gaudart *et al*. (2004), for determining the quality in prediction and robustness of the deviation of multilayer perceptron to linear regression, showed relative performance of Multilayer perceptron (MLP) was found better than linear regression. ANNs, data-driven, and self-adaptive methods are working based on prior assumptions of the model. Based on examples, subtle functional relationships among the data are captured even if the underlying relationships are unknown or hard to describe. ANNs can identify and learn correlated patterns between input data sets and corresponding target values through training. After training, ANNs can be used to predict the outcome of new independent input data and have great capacity in predictive modeling, i.e. all the characters relating to the unknown condition are uploaded to the trained ANNs, and then the prediction of yield may be possible.

An artificial neural network (ANN) by its name itself says it is a network of artificial neurons that follows the concept of function as in human brain neurons. The structure of an artificial neural network consists of several layers of processing units /neurons/ nodes (Haykin, 2001). Modeling with ANN involves two important tasks, namely, topology and learning algorithm of a network. The first task topology of networks comprises (a) fixing the number of layers, (b) the number of neurons present for each layer, (c) the node function for each neuron, (d) feedback or feed-forward method, and (e) the pattern of connection between the neurons through layers. All these alterations should be considered for improved performance of the system. The learning phase deals with weight adjustments as well as threshold values (Hagan and Menhaj, 1994). Usually, the data is divided into three non-overlapping sets: the so-called training, validation, and testing set. The training set, consisting larger portion of data, is used to teach the network to get the desired target function. Then the validation set is used to decide when to stop the training process, to avoid overfitting, a situation where the network memorizes the training data rather than learning the law that governs them. The testing data set, which is exposed to the unseen data, is used to measure the performance of the trained network by mean square error (MSE) or root mean square error (RMSE) or normalized root mean square error (nRMSE). Neural Network architectures were developed by using Levenberg Marquardt (LM) Algorithm (Ranganathan, 2004; Hao and Wilamowski 2011) as a training algorithm for weight matrix.

Due to the high variability in yield from year to year, there is a need to provide a reliable yield prediction, which will be helpful in decision-making as well as future planning. Various studies exist in the literature for predicting crop yield with linear and non-linear techniques but the prominent ones among linear are Regression and Autoregressive Integrated Moving Average (ARIMA) Model and for non-linear, Artificial Neural Network (ANN) architecture (Agrawal *et al*., 1986; Zhang, 1998; Sharma *et. al*., 2012; Kumari, *et. al*., 2013 and Kumari, *et. al*., 2014a).

ANN has many distinguishing features that make it attractive to a researcher. It is in contrast to many traditional techniques for time series predictions, such as Regression and ARIMA, which assume that the existing relationships in the problem under study are generated from linear processes and might be inappropriate for most real-world problems that are non-linear. Therefore, there is a need to solve non-linear, time-variant problems, such as in agriculture and other fields, which are uncertain in their behavior and change with time. ANNs are known to provide competitive results to various traditional time series models such as ARIMA model (George *et. al*., 2001; Ho *et. al*., 2002; Mishra and Singh, 2013; Meena *et. al*., 2016). Kumari *et al* (2016) evaluated the performance of Artificial Neural Network (ANN) by comparing it with Multiple Linear Regression (MLR) and Autoregressive Integrated Moving Average (ARIMA) for predicting the yield of pigeon pea for Varanasi region of Uttar Pradesh using 27 years of data (1985-86 to 2011-12). The performance of the model was assessed by root mean squared error (RMSE). As compared to both linear models, ANN was found to be best suitable model having the lowest RMSE with predicted yield during the year 2012-13 for Varanasi region. Emamgholizadeh *et. al*., (2015) used two methods, namely an artificial neural network (ANN) and multiple regression model (MLR) for estimating the seed yield of sesame from readily measurable plant characters (e.g., flowering time of 100% (days), the plant height (cm), the capsule number per plant, the 1000-seed weight (g) and the seed number per capsule). The accuracy of neural network model and multiple linear regression models are validated with field data, seed yield prediction of sesamum indicated that the prediction accuracy of ANN was better with the coefficient of determination, root means the square error of 0.91 and 0.34 t/ha respectively, while MLR model showed R^2 and RMSE value of 0.78 and 0.346 t/ha respectively. Laxmi and Kumar (2011) did yield prediction for rice, wheat, and sugarcane by utilizing different learning algorithms using neural networks by multilayer perceptron architecture. For the development of ANN model, weather indices such as maximum and minimum temperatures, rainfall, and morning relative humidity were considered as input variables, whereas district-level crop yield was taken into consideration as a dependent or output variable. The conjugate gradient descent algorithm used for learning in multiple perceptions was found to be more accurate in most cases for the yield prediction. From these findings, the study of ANN authenticates good potential for accurate prediction of yield. Das *et al*. (2018) performed PCA on 42 weather indices for a developed weather-based rice model using the ANN technique on the west coast of India. He reported that the range of RMSE and nRMSE during validation was superior in PCA-ANN as compared to ANN.

2.3) Yield prediction using support vector machine (SVM)

Support Vector Machine (SVMs) is a kernel-based, nonparametric, supervised machine learning technique used for the prediction and classification of samples in two disjoint clusters (Pal, 2009). The prediction accuracy of SVMs depends on model input variable and kernel types because it is nonparametric (eg: linear, polynomial, radial-based function, and sigmoid) (Ustuner *et al*., 2015). It is an idea for a set of correlated superintended learning techniques that show patterns and rate data being used for regression analysis and classification. It can use two possible class forms making the nonprobabilistic binary linear classifier the prediction for each input (Collobert *et. al*., 2001). Vapnik (1998) first develop a technique that can be used for prediction by support vector machines (SVMs) with good accuracy. SVM generates a hyperplane or more than one hyperplane in a high or infinite dimensional space, which is utilized for regression, classification, or other tasks. The SVM techniques are used for extensive applications in in the prediction of crops and weather along with long-time series data analysis. SVM method was used to overcome the overfitting problem in the input dataset. Different weather variables such as daily maximum and minimum temperature, precipitation, bright sunshine hour, maximum and minimum relative humidity along with long-term seed yield data are required for crop yield prediction (Athani and Tajeshwar, 2017). Gandhi *et al*. (2016) evaluated SVM model and predicted the rice crop yield with 78.76% accuracy by applying SMO classifier using the WEKA tool for 27 districts of Maharashtra state, India. Balakrishnan and Muthukumarasamy (2016) compared SVM and Naive Bayes techniques for the prediction of different crops in Thanjavur district, Tamil Nadu. They reported that the accuracy of SVM and Naive Bayes techniques for rice paddy were 90.5% and 86.3%; 87.6% and 84.9% for cotton; 88.5% and 85.6%for

Sugarcane; 89.3% and 85.4%for groundnut; 86.7% and 82.4% for black gram, respectively. These results clearly showed that the SVMs approach was a better prediction option than the Naive Bayes approach. Su *et al*. (2017) developed a SVM-based Open Crop Model (SBOCM) to predict the growing stages of rice and seed yield at a regional level. Karimi *et al*. (2008) used a fivefold cross-validated support vector technique with aerial hyperspectral data to predict the biophysical parameters and yield of corn. They found more accurate prediction results by SVM technique over the stepwise regression technique. Were *et al*. (2015) did the relative evaluation of support vector regression, artificial neural networks, and random forests for soil organic carbon prediction. They found that the model accuracy assessment parameters favored the prediction by SVM among all other techniques at the unvisited locations. Dey *et al*., (2017) predicted the rice yield (Aus, Aman, and Boro rice) by different models such as; Multiple Linear Regression, SVM, Adaptive Boosting, and Modified Nonlinear Regression. They concluded that the Modified Non-linear Regression model predicted the rice yield in a better way than the other three predefined models, except for predicting Aman Rice, where the support vector machine had more accuracy in prediction. Palanivel and Surianarayanan (2019) reviewed several types of machine learning techniques such as linear regression, artificial neural networks, and Support Vector Machines and found that SVM based prediction models are found to be more suitable for crop yield prediction. SVM theory is a bit intimidating, particularly because the more efficient SVM variants often incorporate difficult-to-understand concepts for a non-expert user. This limits effective cross-disciplinary applications of SVMs. The importance of crop prediction is highly needed for agriculture and the economy. Numbers of studies were done using SVM for classification purposes but very few studies on crop yield prediction used variables selection and extraction by SMLR and PCA, respectively.

2.4) Yield prediction using random forest (RF)

Random forest is a most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks (Breiman, 2001). Random forest is used for classification purposes for categorical variables, but apart from that, it is used as a regression technique for continuous variables. Random Forest operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. More number of decision trees in a forest are needed to develop a robust prediction model. Several homogeneous units of decision trees had been developed to improve the model's performance by splitting each tree variable known as nodes based on the nature of input variables. These variables used to split the data are considered important explanatory variables. The random forest technique avoids overfitting in the training dataset and is based on boots trapping. The predicted value of a continuous response is the mean fitted response from all the individual trees that resulted from each bootstrapped sample. The key advantage of the random forest technique is it can investigate non-linear and hierarchical relationships between the predictors and the response using an ensemble learning approach. In the case of Random Forest, the training parameters that needed specification were: (i) the number of trees to grow the forest (ntree), (ii) the number of randomly selected predictor variables at each node (mtry), and (iii) the minimal number of observations at the terminal nodes of the trees (node size).The default value of ntree =500 and mtry=M/3 for regression problems, where M is the number of features for prediction. Gromping, (2009) reported that RF regression technique provides a more accurate result for highly correlated input variables.

Tulbure *et al*. (2012) used random forest regression to identify important variables for switching grass yields across the USA. They identified nitrogen fertilizer, cultivar, rainfall, stand age, and soil silt levels as the most influential of 22 predictor variables. The variables identified by random forests were then used to build better models of switch grass yield. Fukuda *et al*. (2013) demonstrated Random Forest models to predict the maximum and mean value of mango fruit yield under different irrigated and rainfed conditions. Everingham *et al*. (2016) used the random forest technique for sugarcane yield prediction by using 10 years of weather data. They reported that the accuracy of sugarcane yield prediction varies from 86.4% to 95.5% in September in the year before harvest and January in the year of harvest. Jeong *et al.* (2016) used Random forest and compared it with multiple linear regressions for wheat, maize, and potato tuber. They reported that wheat yield prediction at a global scale had nRMSE values of 16 and 33 for prediction done by Random forest and multiple linear regressions, respectively. Random forest performed better than MLR in all measures of model performance for US maize yield predictions for 30 years with RMSE of 1.13 tons/ha, which is 16.7% of the average observed yield. The result obtained by RF showed more accuracy in the prediction of potato tuber compared with the MLR technique. Random forest have been used as a classification tool rather than a regression tool for predicting ecosystem or crop productivity (Vincenzi *et. al*., 2011). Less literature is available for applications of RF regression in the fields of crop science.

3) Crop simulation models

Crop simulation models (CSM) are an alternative approach to studying the crop ecosystem. Crop simulation models are extensively used to understand the influence of meteorological parameters, soil properties, crop genotype, and crop management practices on various agricultural applications. Dynamic mechanistic crop models are process based and they utilize established physiological processes to mimic the influence of environmental conditions on the growth and yield of crops.

The explosion of information in different agricultural science disciplines necessitated a system modeling approach to understand the interaction between different components of a system and the overall response of the system. System modeling was initially popular in engineering science but originated in the 1960s in agricultural science with the pioneering work of physicist, C. T. de Wit of Wageningen University. The work of modelers at Wageningen University continually evolved and led to the development of many models like MACROS, SUCROS, BACROS, etc. Another pioneer was a chemical engineer, W. G. Duncan, who published on modeling canopy photosynthesis (Duncan *et al*., 1967). Jones *et al*. (2017) reviewed the brief history of agricultural system modeling. Global wheat shortage in the USA during 1972 caused a major boost in studies of agricultural system modeling for predicting regional crop production with the help of remote sensing to understand the behavior of international trade. It led to the development of CERES model for wheat and maize (Ritchie and Otter, 1984; Jones and Kiniry, 1986). Likewise, the pioneer efforts of Duncan, Ritchie, and others have progressed and supported the extensively used DSSAT group of crop models through collaborative efforts among the University of Florida, University of Hawaii, Michigan State University, the International Fertilizer Development Institute, Washington State

University, and others (IBSNAT, 1984; Tsuji *et al*., 1998; Uehara and Tsuji, 1998; Jones *et al*., 2003; Hoogenboom *et al*., 2012). Australia during the 1990s, guided the development of a cropping system model 'APSIM' group model, which is one of the most widely used crop models in the world (McCown *et al*., 1996; Keating *et al*., 2003). The Systems Analysis of Rice Production (SARP) project in 1984 guided the development of the extensively used rice model ORYZA (Penning de Vries *et al*., 1991; Bouman *et al*., 2001). The workers from India also contributed to developing of these models through a collaborative project between IBSNAT and SARP. Apart from this, indigenous efforts of India started with the development of a wheat crop model developed by the IARI and SAC and further developed as WTGROWS (Aggarwal *et al*., 1994). Subsequent efforts by Aggarwal and his colleagues at IARI led to the development of a generic model for annual crops in a tropical environment called InfoCrop (Aggarwal *et al*., 2006b). Many crop models of differing complexity are now available for major food crops; some are listed in Table 2.1. Among them, the most widely used models are DSSAT, APSIM, EPIC, WOFOST, AQUACROP, and STICS.

Table 2.1: List of published crop models

Crop	Model Name	Reference
Alfalfa	APSIM	Fick (1981)
Barley	CERES-Barley	Ritchie *et al.* (1989)
Chickpea	CHIKPGRO	Singh and Virmani (1996)
Cotton	ELCOMOD	Marani and Baker (1981)
	GOSSYM	Baker *et al.* (1985)
	KUTIN	Mutsaers (1984)
	COTCROP	Brown *et al.* (1985)
	COTTAM	Jackson *et al.* (1988)
	COTCO2	Wall *et al.* (1994)
Bean	BEANGRO	Hoogenboom *et al.* (1994)

Maize	CORNF	Stapper and Arkin (1980)
	CERES-Maize	Jones and Kiniry (1986)
	GAPS	Buttler (1989)
Peanut	PNUTGRO	Boote *et al.* (1989)
Pearl Millet	CERES-Millet	Ritchie and Alagarswamy (1989)
	RESCAP	Monteith *et al.* (1989)
	BAJRAWAT	Rao *et al.* (1999)
Potato	POTATO	Ng and Loomis (1984)
	SIMPOTATO	Hodges *et al.* (1992)
Rapeseed	BRASSICA	Rao (1992)
Rice	CERES-Rice	Ritchie *et al.* (1986)
	ORYZA-1	Kropff *et al.* (1994)
Sorghum	SORGF	Arkin *et al.* (1976)
	CERES-Sorghum	Alagarswamy and Ritchie (1989)
	RESCAP	Monteith *et al.* (1989)
	SORKAM	Rosenthal *et al.* (1989)
Soybean	SOYMOD	Curry *et al.* (1975)
	GKYCIM	Acock *et al.* (1983)
	SOYGRO	Wilkerson *et al.* (1983)
Sugarcane	AUSCANE	Jones *et al.* (1989)
Wheat	ARCWHEAT1	Weir *et al.* (1984)
	WINTER WHEAT	Baker *et al.* (1985)
	CERES-Wheat	Ritchie *et al.* (1985)
	SWHEAT	Van Keulen and Seligman (1987)

	AFRCWHEAT2	Porter (1993)
	ShootGro 2.0	Wilhelm *et al.* (1993)
	WTGROWS	Aggarwal *et al.* (1994)
	MODWht3	Rickman *et al*. (1996)
	SIRIUS	Jamieson *et al.* (1998)
General	BACROS	De Wit (1978)
	INFOCROP	Aggarwal *et al*.(2006b)
	SUCROS	Van Keulen *et al*. (1982)
	LINTUL	Spitters, (1987)
	MACROS	Penning de Vries *et al.* (1989)
	WOFOST	Van Diepen *et al.* (1989)
	EPIC	Williams *et al.* (1989)
	CropSyst	Stockle *et al.* (1991)
	WOFOST 6.0	Hijmans *et al.* (1994)

A brief timeline history of selected key events and drivers that influenced the development of agricultural system models had been shown in fig 2.2.

The model must be calibrated before its use i.e., model output has to be compared with independent observation datasets. Models are frequently validated with all or some of the data used for model development or calibration (Jones *et al*., 2001), whereas independent data, is not used in model development (McCarl, 1984). While validating the crop models, in most cases comparison of simulated yield with observed yield from short-term field experiments is a standard procedure. Aggarwal *et al*. (1994) validated WTGROWS under potential production environments and found that the model simulated wheat yield accurately at most places with no point outside ±1 standard deviation and an R^2 value of 0.74.

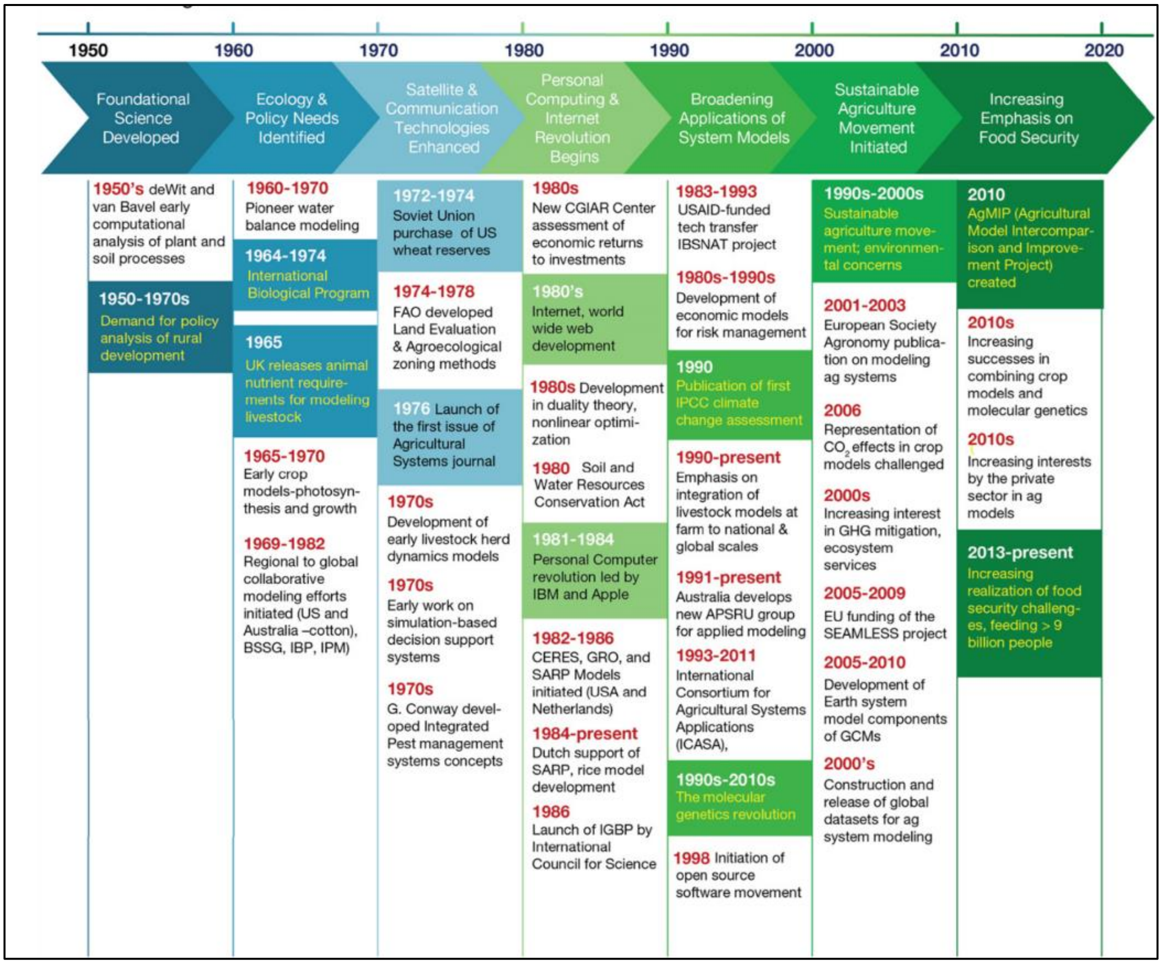

Fig 2.2 Brief timeline history of agricultural system models

Todorovic *et al*. (2009) compared the performance of AquaCrop, WOFOST, and CropSyst models for simulating the growth of sunflowers under different water regimes in Mediterranean environments. The reason for choosing these three models is that they differ in the level of complexity describing crop development, main growth modules driving the biomass simulation, and the number of input parameters. The mechanism of stimulation of biomass in AquaCrop is exclusively water driven while CropSyst is based on both water and radiation-driven module and WOFOST uses the carbon-driven approach for dry matter production. They found that the estimation of the phenological stage was almost similar for all models because of the utilization of the same heat unit concept. During development, CropSyst over-predicted LAI development, whereas; WOFOST provided convincing results before anthesis. Maximum LAI at anthesis is also over-predicted by CropSyst, whereas WOFOST under-predicted LAI by 2%. However, senescence followed the measured LAI trend in all models. All the models underestimated the yield under severe water stress conditions. Simulation of biomass growth by WOFOST was better than the other two models under rain-fed and deficit irrigation because the latter involves the simpler physiological sub-model accounting impact of water stress on biomass growth and its partitioning to yield. Simulation of water use efficiency (WUE) was better with CropSyst than other models under a limited water supply.

Battisti *et al*. (2017) studied the performance of five soybean models (FAO – Agroecological Zone; AQUACROP; DSSAT CSM–CROPGRO–Soybean; APSIM Soybean; and MONICA) under both rainfed and irrigated condition in Southern Brazil. DSSAT and APSIM were found to be better estimators of phenology because they accounted for the effect of water stress and nitrogen limitation on crop development rate besides thermal time and photoperiod. Under the rainfed condition, APSIM and DSSAT simulated a more drastic reduction in LAI due to water deficit during the middle to end of the crop cycle, while MONICA had a less drastic reduction.

Akinseye *et al*. (2017) compared the performance of DSSAT, APSIM, and Samara crop models for West African sorghum cultivars. Simulated phenology and morphology organs during calibration and validation were within the close range of measured values with the evaluation of model error statistics (RMSE and R^2). Except for the highly sensitive variety (IS15401), APSIM and Samara estimates indicate the lowest value of RMSE (< 7 days) against the observed values for phenology events (flowering and maturity) compared to DSSAT model. The grain yield and biomass prediction were less accurate for calibration and validation. The predictions using APSIM were found to be closest to the observed followed by DSSAT and Samara models, respectively. Based on detailed field observations, this study showed that crop models captured well the phenology and leaf development of the photoperiod sensitive (PPS) varieties of West Africa but failed to estimate accurate partitioning of assimilates during grain filling. APSIM and SAMARA are more mechanistic crop models. They have a higher sensitivity to the adjustment of key parameters, notably the specific leaf area for APSIM in low PPS varieties. At the same time, SAMARA shows a higher response to parameter changes for high PPS varieties.

Jin *et al*. (2016) evaluated the algorithms that determine the impacts of heat and drought stress on maize in 16 major maize models by incorporating these algorithms into a standard model, APSIM, and running an ensemble of simulations. Different drought algorithms (*i.e.*, a function of soil water content, soil water supply to demand ratio, and actual to potential transpiration ratio) simulated considerably different patterns of water shortage over the growing season but predicted similar decreases in annual yield. Using the selected combination of algorithms, the simulations showed that maize yield reduction was more sensitive to drought stress than to heat stress for the US Midwest since the 1980s, and this pattern will continue under future scenarios; the influence of excessive heat will become increasingly prominent by the late 21st century.

Eitzinger *et al*. 2004 compared the performance of CERES, WOFOST, and SWAP models in simulating soil water content during the growing season in Austria under different soils (chernozem, sandy chernozem, and fluvisol) with a 2 m profile depth. CERES and SWAP simulated the grain yield of barley and wheat better as compared to WOFOST. All three models simulated soil water content in the profile with similar results. The root means square error (RMSE) range of soil water content was 0.71–4.67% for barley and 2.32–6.77% for wheat, depending on the model and soil type. None of the models simulated total soil water content in the profile significantly better, but there was a general tendency for the models to overestimate soil water depletion. CERES and SWAP mimicked the soil water content dynamics well in the top 0.3 m of the soil. The study showed that the multiple layer approach models (SWAP or CERES) are more sophisticated estimation methods for root growth and soil water extraction.

The results of the WOFOST model did not reflect real conditions, probably because it considers all rooted soil to be homogeneous in terms of soil water content. Inaccuracies caused by applying this assumption to soils with irregular drying and wetting cycles with various amounts of water have also been found by others, for example, Van den Berg and Driessen (2002).

3.1) InfoCrop model

InfoCrop is a web-based crop simulation model developed to understand the impact of various input parameters on specific output parameters. It is designed to simulate the effect of weather variables, soil properties, field management practices, and specific pests and diseases on crop growth and biophysical parameters and their associated environment (Aggarwal *et al*., 2006b). This model is based on developmental stage-dependent crop-specific functions, dry matter available for crop growth is partitioned into roots, leaves, stems, and storage organs. Roots get the priority for allocation and get increased in case the crop experiences water or nitrogen stress. The remaining dry matter is allocated to the above-ground shoot from which a fraction is allocated to leaves and stems. The balanced dry matter is automatically allocated to the storage organ. The number of storage organs and grain filling rate is not directly influenced by water stress, as it influences the effect of water stress on dry matter production. The programming language of the InfoCrop model is Fortran Simulation Translator (FST) (FST/FSE; Graduate School of Production Ecology, Wageningen, The Netherlands; Van Kraalingen, 1995). Another version (user interface) had been written for Info Crop to accelerate its applications in agricultural research and development for people that are not familiar with the programming language. InfoCrop was developed for 13 crops including Mustard. Crop models in InfoCrop are sensitive to weather, soil, varietal parameters, and agronomic management practices**.** A basic representation of InfoCrop has been shown in Fig 2.3.

A detailed study of the InfoCrop model is required for the interactive understanding of input variables on crop plants which plays a critical role in the planning and execution of farm decisions such as the selection of cultivars and soil and water management practices (Krishnan *et al*. 2016). The web-based InfoCrop mustard model is easy to understand. Scientific understanding and policy/decision support were two broad purposes that guided the development of agricultural system models. The sensitivity analysis of the InfoCrop model makes us understand the importance of the input variable's accuracy to develop a robust model. Kumar *et al*. (2017) analyzed the different combinations of weather parameters for the sensitivity analysis of mustard cultivars through the InfoCrop model at Haryana. They reported that variation in maximum temperature and minimum temperature by -1 to 1ºC, rainfall by 10 to 20%, and CO_2 between 415 to 490 ppm respond to the positive impact on biophysical parameters and seed yield. Krishnan and Aggarwal (2018) analyzed the global sensitivity of the InfoCrop model for different soil properties. They used the 10 soil parameters and reported that nitrate content, soil organic carbon, and soil ammonium were the most sensitive variables under the ideal condition (without stress). Apart from that initial soil moisture plays a butterfly effect in water stress conditions.

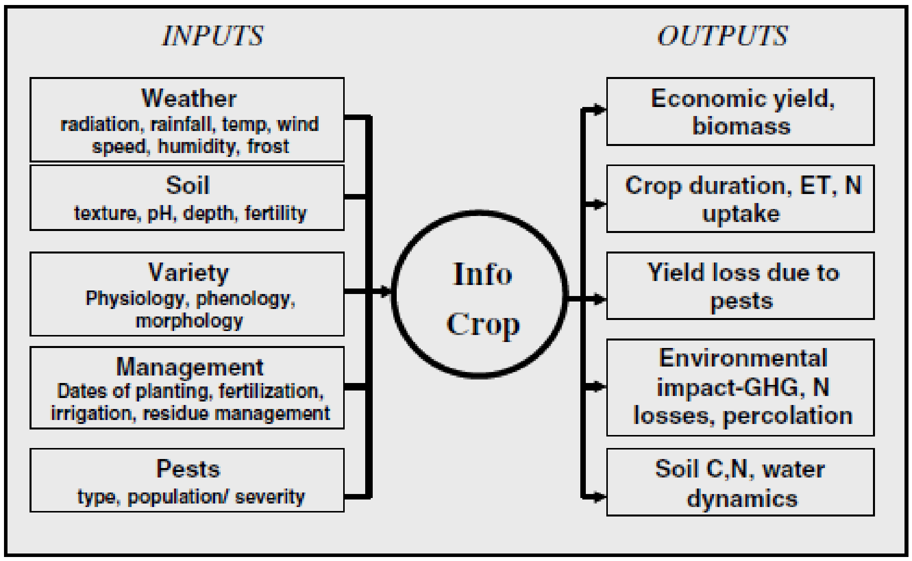

Fig 2.3 Key inputs and outputs of web-based crop simulation model (InfoCrop)

Kumar *et al*. (2015) calibrated and validated InfoCrop for different rice cultivars at Kumarganj, Faizabad (U.P.). The model accuracy parameters (MAE, MBE, and RMSE) were calculated with 10% variations for the simulated value of phonological stages, leaf area index, above-ground biomass, and rice seed yield. In Gujarat, the InfoCrop model was calibrated and validated for maize crops in *Rabi* (Choudhary *et al*., 2014a) and *Kharif* seasons (Choudhary *et al*., 2014b). They reported that InfoCrop underestimates leaf area index whereas, overestimates above-ground biomass and seed yield with a 15% error. Akula (2003) compared the WTGROWS and InfoCrop model for wheat with two years of experimental data at Anand, Gujarat. They reported the underestimated results and noticed that WTGROWS model (4537±874 kg ha^{-1}) performed better than the InfoCrop model (4296±918 kg ha^{-1}) for wheat grain yield, while the actual grain yield was (4608±620 kg ha^{-1}). In the Anand and Panchmahal districts of Gujarat, the simulated grain yield was 2.4 times higher than the observed grain yield by InfoCrop model (Akula *et. al*., 2005b). Several other researchers successfully adapted, calibrated, and validated the InfoCrop simulation model for different crops in different regions rice (Aggarwal *et al*., 2006b), wheat (Aggarwal *et al*., 2006b), potato (Singh *et al*., 2005), cotton (Hebbar *et al*., 2008), and coconut (Kumar *et al*., 2008).

Boomiraj *et al*. (2010) calibrated the InfoCrop model at Delhi and simulated it in different mustard-growing regions of the country such as: (eg; Hissar, Ludhiana, Kanpur, Sriganganagar, Delhi, Gwalior, Pantnagar, Akola, and Varanasi) with the use of field experimental data sets collected from All India Coordinated Research Projects. They revealed that the observed and simulated values in different locations vary from 0.25 to 2.87 t ha^{-1} and 0.26 to 3.14 t ha^{-1} for seed yield, 0.29 to 7.05 t ha^{-1} 0.28 to 7.22 t ha^{-1} for above-ground biomass, respectively. InfoCrop has been successfully adapted, calibrated, and validated in the Delhi region for mustard by Adak *et al.* (2009) and reported that the model overestimated about 45% of the seed yield and biomass with RMSE value of 3.14. Results obtained by simulation of crop growth parameters like leaf area index (LAI), biomass, and seed yield showed that the model needs further refinement. Keerthi *et al.* (2017) calibrate and validate mustard InfoCrop model in an experimental field and reported that the RMSE value 157 and 1364 for simulated seed yield (kg/ha) and biomass (kg/ha), respectively. Gill *et al* (2016) simulated mustard crop phenology, biophysical parameters, and seed yield by InfoCrop model under different temperature scenarios at Punjab Agricultural University Ludhiana, Punjab. They found a linear reduction in seed yield and biomass by raising the maximum and minimum air temperature. Only limited studies have been done to calibrate and validate InfoCrop on the experimental field as well as farmers' field experimented data.

4) Optimal combination of predicted models

Bates and Granger first proposed the optimal combination theory in 1969. They used plug-in weights in the optimal solution based on the estimated variance-covariance matrix with a major emphasis on improving the accuracy of yield prediction. The need for optimal combination is to obtain diversified results because many models developed for yield prediction have similar accuracy, so it is difficult to identify the best prediction model among them. The covariance of two predicted results was used in the combination method to get the better predicted property with the least root mean square error. An optimal combination on a weight basis is used to minimize the expected loss of the combined prediction techniques. The larger weights were responsible for better prediction and reduction in error. This optimized method of weighting is more realistic and further improves prediction accuracy. Generally, the prediction combinations technique has been successfully used in various fields like Gross National Product, currency market, volatility and exchange rates, inflation, interest rates, money supply, stock returns, meteorological data, city populations, etc.

The traditional approach for the optimal combination was a weighted geometric average. But it was modified by Cheng *et al*. (1994) and renamed the induced ordered weighted geometric average (IOWGA) approach. Song *et al*. (2011) combined the predicted energy consumption results by induced ordered weighted geometric average (IOWGA) operator and resulted that the mean average percent error of the combination model was 0.0198, which was smaller than a single model prediction in China. Lajevardy *et al*. (2015) guided the electricity price used for the induced ordered weighted geometric average combination model. Pandey *et al*. (1992) combined the ARIMA model and remote sensing-based technique to improve the accuracy of wheat yield estimated at Hissar, Haryana. They reported that the optimal combination reduced the RMSE and percent deviation compared to both individual approaches. There needs to be more optimal combination literature based on agriculture crop yield prediction.

3. Materials and Methods

In order to achieve the objectives of the present study, an experiment was conducted for mustard crop during *Rabi* 2016-17 and 2017-18 at the research farm of ICAR-Indian Agricultural Research Institute (IARI), New Delhi. The long term weather data was collected by India Meteorological Department (IMD) and National Climate Data Center (NCDC) while mustard yield data was collected from Directorate of Economics & Statistics (DES) and state agricultural department for Delhi and major mustard growing region of Rajasthan. The details of the materials and methods adopted during investigation have been presented in this chapter.

3.1 Experimental farm study

3.1.1 Study area/location

A field experiment was conducted at the experimental farm (Main Block 4C) of the Division of Agricultural Physics of Indian Agricultural Research Institute, New Delhi, located at 28^{o}64’23’’ North latitude and 77^{o}15’27’’ East longitude with an altitude of 228.6 meters above mean sea level.

3.1.1.2 Climate

The climate of New Delhi is sub-tropical and semi-arid with warm and dry summers and cold winters coming under the ‘Trans-Gangetic Plains’ agro-climatic zone. The hottest months (May and June) during summer have daily mean maximum temperature remains in the range of 40-46^{0}C most of the days while temperature decreases from September onwards. The all-time record of the highest daily maximum temperature (46.8^{o}C) was recorded on 29th May, 1998. In January daily mean minimum temperature ranged around 5°C to 7.5°C, it is the coldest month in the winter season. The all-time lowest daily minimum temperature (-1.4^{o}C) was recorded on 9th January, 2006. Annual mean rainfall is around 710 mm and the wettest months are July and August. Maximum rainfall around 80 percent of the annual rainfall is received from July to September during the south-west monsoon period (expected date of arrival of monsoon at Delhi is 29th June). The rest amount is received through “Western Disturbances” during winter months (December to February) and some local convections during pre-monsoon months (March-June). However, air remains dry during most of the year. The monthly mean relative humidity ranged between 35% (June) and 94% (December). Monthly mean wind velocity varies from 3.5 km/hr in October to 4.4 km/hr in April. Air remains dry during most part of a year. Pan evaporation varied between 3.5 to 13.5 mm day^{-1} and reference evapotranspiration varies from 9-15 mm day^{-1}.

3.1.1.3 Soil properties

The sandy loam texture of soil was present at the top layer of the experimental site. Whereas, 15-90 cm soil was sandy clay-loam. Physio-chemical properties of soil at experimental site were determined before initiating the experiment and the values are presented in Table 3.1. It showed that the soil was little alkaline, deficient in organic carbon and available nitrogen and medium in phosphorus and potassium content. The bulk density is 1.57, 1.60, 1.63 and 1.71 Mg m^{-3} for different depths such as 0-15 cm, 15-30 cm, 30-60 cm 60-90cm, respectively. The electrical conductivity (EC) varied from 0.46 dS m^{-1} for the 0-15 cm layer to 0.25 dS m^{-1} for the 60-90 cm layer. The mean value of field capacity (FC) and permanent wilting point (PWP) for 90 cm soil depth was 220 mm and 70 mm, respectively.

Table 3.1: Physiochemical properties of soil at experimental site

Depth (cm)	**P^H**	**EC (dS m^{-1})**	**SOC (g kg^{-1})**	**Particle size distribution**			**Bulk Density (mg m^{-3})**	**Saturated hydraulic conductivity (cm hr^{-1})**	**Soil texture**
				Sandy (%)	**Silt (%)**	**Clay (%)**			
0-15	7.1	0.46	4.2	65.0	16.7	19.1	1.57	1.01	SL
15-30	7.2	0.24	2.2	65.4	10.8	24.9	1.60	0.82	SCL
30-60	7.4	0.24	1.6	63.83	10.0	26.2	1.63	0.71	SCL
60-90	7.5	0.25	1.4	59.53	10.1	30.2	1.71	0.49	SCL

3.1.2 Experimented Crop

Three mustard cultivars (*Brassica juncea* (L.) Czern. & Coss.) such as Pusa Tarak (P. Tarak), RH-406 and Girraj were sown during the *Rabi* season of 2016-17 and 2017-18. The specific characteristics of these cultivars have been described below.

3.1.2.1 Pusa Tarak

This cultivar was developed in 2009 for the National Capital Region of Delhi. It has a short-duration cultivar that requires about 121 days to attain maturity. The average seed yield of Pusa Tarak is 1920 kg ha^{-1}. It is a high test-weighted mustard cultivar (6.00 g/ 1000 seeds) with about 40.0 % oil content. This variety is suitable for multiple cropping systems.

3.1.2.2 RH-406

This cultivar was released in 2013 and notified at a national level for timely sown rainfed conditions of Haryana, Punjab, Delhi and parts of Rajasthan. It is a bold-seeded, high-yielding cultivar with high oil content. It matures in 145 days and the seeds contain 42% oil.

3.1.2.3 Girraj

This cultivar is also known as DRMRIJ-31. This cultivar was released in 2013 and is mainly recommended for Delhi-NCR, Haryana, Punjab and some parts of Rajasthan. It is a bold-seeded, high-yielding cultivar with high oil content that matures between 137-153 days. The seed yield varies between 2225-2750 kg ha^{-1} with 39 to 42.6% oil content.

3.1.3 Experiment details

The outlay of the field experiment site followed split plot design, in which each plot was of 5x4 m^2 size with adequate margins for bunds and irrigation channels (Figure 3.1). The soil of the location is deep to very deep and sandy loam textured throughout the profile. Three Mustard cultivars (cv. Pusa Tarak, RH-406 and Girraj) were raised during the *Rabi* season of 2016-2017 and 2017-18 in split-plot design with a date of sowing as the main plot treatment and cultivars are sub plot treatment. The field was prepared following the usual pre-sowing operations like disking and leveling.

Treatments

All three cultivars were sown at three different dates of sowing: D1: Timely sown (10th Oct 2016), D2: Late sown (25th Oct 2016) and D3: Very late sown (10th Nov 2016) during *Rabi* season 2016-17 and D1: Timely sown (12th Oct 2017), D2: Late sown (26th Oct 2017) and D3: Very late sown (11th Nov 2017) during 2017-18. Manual sowing with the help of hand-held seed drill was done keeping the recommended row-to-row spacing of 45cm and plant-to-plant spacing of 10 cm. The rainfall received during the entire crop growing period was 119.7 mm during 2016-17 and 13.4 mm for the 2017-18 year. A good amount of rainfall (39.1mm) was received at 40th SMW in 2016-17, which met the pre-sowing irrigation requirement. The irrigation treatments were applied at 41 and 80 DAS for the first sown crop, 26, 65 and 112 DAS for the second sown crop and 50 and 90 DAS for the third sown crop during 2016-17. During 2017-18, irrigation treatments were applied at 30, 72 and 113 DAS for a timely sown crop, 58 and 99 DAS for a late sown crop and 43 and 84 DAS for a very late sown crop. NPK nutrients were applied as per recommended dose of fertilizers for mustard crop (80 kg N: 40 kg P_2O_5 : 40 kg K_2O). 50% of N, with 100% of P_2O_5 and K_2O were applied at the time of sowing and remaining 50 % of nitrogen was applied after first irrigation at 30 days after sowing. The recommended cultural practices of weeding and plant protection measures were followed. Laborers were employed to harvest the crop after complete drying. The mustard field experiential during *Rabi* 2016-17 and 2017-18 at the research farm of ICAR-IARI are shown in Plate 3.1 and 3.2.

3.1.4 Field observation and measurements

Periodic observations at 15 days intervals on different parameters such as soil moisture at different depth, leaf area index (LAI), radiation interception, plant height and biomass were measured in each treatment during the crop-growth period. Final biomass, seed-yield and yield attributes were recorded after the harvest. Daily weather data of maximum and minimum temperature, rainfall, morning and evening relative humidity, bright sunshine hours, wind speed and pan evaporation were acquired from the adjacent agromet class-A observatory of IARI during the crop growing period.

Irrigation channel	**D3V3**	**D2V2**	Irrigation channel	**D1V1**
	D3V1	**D2V3**		**D1V2**
	D3V2	**D2V1**		**D1V3**
	D3V3	**D2V2**		**D1V1**
	D3V1	**D2V3**		**D1V2**
	D3V2	**D2V1**		**D1V3**
	D3V3	**D2V2**		**D1V1**
	D3V1	**D2V3**		**D1V2**
	D3V2	**D2V1**		**D1V3**

Fig 3.1 Layout of experimental field ICAR-IARI, New Delhi during *Rabi* 2016-17 and 2017-18

Size of each plot: 4m X 5m

Cultivars: V1: Pusa Tarak, V2: Girraj, V3: RH-406

Sowing time: D1**:** Timely sown, D2: Late sown, D3: Very late sown

Plate 3.1 Experiential field of mustard crop at ICAR-IARI, New Delhi field during Rabi 2016-17

Plate 3.2 Experiential field of mustard crop at ICAR-IARI, New Delhi field during *Rabi* 2017-18

3.1.4.1 Soil moisture

3.1.4.1.1 Initial soil moisture

Initial profile soil moisture on the day of sowing of the crop at various soil profile sections viz., 0-15, 15-30, 30-45, 45-60, 60-75 and 75-90 cm was measured by gravimetric method. In this method, soil sample at different depths was collected in aluminum bins, weighed and placed in the hot air oven and allowed to dry at 105ºC for at least 24 hrs. The weight of dried soil sample was taken and the moisture content on mass basis was calculated using the following formula:

$$SMC\ (\%) = \frac{(M_w - M_d)}{M_d} * 100 \qquad \text{------------------ (1)}$$

Where, *SMC* is soil moisture content in %, *Mw* is weight of non-dried soil samples and *Md* is the sample weight of the dried soil sample.

3.1.4.1.2 Profile soil moisture

Gravimetric method was used to measure soil moisture for upper 15 cm soil. Neutron Moisture Meter (NMM) (CPN-503 DR Hydroprobe, Campbell Pacific Nuclear International Inc. USA) was used to measure profile soil moisture at the different depths of 15-30, 30-45, 45-60, 60-75 and 75-90. Aluminum tubes of 1.5 m length were fitted in the center of each treatment plot to measure profile soil moisture at different depths. In each treatment first standard counts were recorded by keeping the probe in the air, and then counts per minute at every soil depth were noted. Count ratio i.e., ratio of counts per minute at a given depth to standard count, was computed. The soil moisture content was calculated using the calibration equation between gravimetric soil moisture and count ratio. This calibration equation was developed for the same field by Thomas (2013). The calibration equation is given below:

$$SMC\ (\%) = \frac{\left(\frac{C_a}{C_s} - 0.1512\right)}{6.5066} \qquad \text{-------------------- (2)}$$

Where, *SMC* is soil moisture content in percent, *Ca* is NMM count per minute at a specific soil depth and *Cs* is the NMM standard count per minute. The *Ca/Cs* is the count ratio. Profile soil moisture content calculated by addition of soil moisture of each depth up to 90 cm.

3.1.4.2 Leaf area index (LAI)

Now a days, there are numerous direct (destructive) and indirect (non-destructive) techniques available to estimate LAI. The direct methods like computation of the leaf area of plant canopy required subsamples of leaves and LAI is related to dry biomass specific leaf area (SLA). LAI is determined by multiplying SLA with total biomass (Jonckheere *et al.*, 2004). Direct measurement of LAI is more accurate than the indirect method but it is expensive, time consuming, labour-intensive and difficult to apply in large area scale. Thus, the indirect methods of LAI estimation are more popular. The optical instrument plant canopy analyzer is largely used for measuring the LAI non-destructively (Liu, Pattey and Admiral, 2013). LAI were measurements by Plant Canopy Analyzer (LI-COR, USA) at 15 days intervals during crop growing period. The principle of Plant Canopy Analyzer is based on “fish-eye” measurement of diffuse radiation interception by measurement of gap fraction at five zenith angles (0–13, 16-28, 32–43, 47–58, 61–74°) simultaneously. The LAI measurement by plant canopy analyzer based on light intercepted by plant canopy and foliage density. The measured gap fractions are then inverted to get the effective LAI and the instrument was set to take four below and one above-canopy measurements to estimate the LAI. Five LAI readings were recorded in each plot and removed two outfit values. The average of the reaming three values represents each plot LAI.

3.1.4.3 Photosynthetically active radiation (PAR)

Line Quantum Sensor (LI-191, LI-COR) was used with an integrator (LI- 250A, LI-COR) for measuring the incident and intercepted photosynthetically active radiation (PAR) by a mustard canopy. PAR measurements were taken above the canopy with the sensor facing sky to account for incident radiation (I_o) received and the sensor looking downwards for reflected radiation (I_r) from the canopy. Reading was also taken below the canopy keeping the sensor just above the soil but across the rows with the sensor looking upwards for the transmitted radiation (I_t) through the canopy and with the sensor looking downwards for radiation reflected (I_e) from the soil. Three sets of measurements were recorded in each plot and averaged. The above measurements were taken regularly (within a week) on clear days between 11:30 and 12:00 hours IST when disturbances due to leaf shading and solar angle were minimal. These measurements were used to derive fraction intercepted PAR (fIPAR) as given in formula:

$$fIPAR = \frac{(I_0 - I_t)}{I_o} \qquad \text{------------------ (3)}$$

Values for fIPAR for each day after sowing were interpolated between actual measurements by linear interpolation throughout the crop season. Global solar radiation (MJ m^{-2}) was computed from daily bright sunshine hours by using Angstrom coefficients (a=0.32, b=0.46). Computed global solar radiation was multiplied with a factor of 0.48 to calculate daily incident PAR. Daily intercepted PAR was computed by multiplying daily incoming PAR values with corresponding daily *f*IPAR values. The daily IPAR was accumulated corresponding to the crop growth period to calculate total IPAR (TIPAR).

3.1.4.4 Crop phenology

Phenological stages of the crop were identified visually through regular visit to the field and their dates were recorded. The main phenological stages of mustard were: emergence, Seedling emergence, rosette initiation, first flowering, 50 % flowering, pod formation, seed development and physiological maturity.

3.1.4.5 Plant height

Randomly five mustard plants were selected and measured their height with the help of scale at regular intervals (within a week). Afterward, an average was computed to have a representative plant height of that plot.

3.1.4.6 Above-ground biomass

Three plant samples of 50 cm row length were cut just above the soil surface for measuring above-ground biomass. The plant samples are dried in the shade and later dried in an oven at 65°C for 72 hours until constant weight. The weight of dried plant samples was measured using an electric balance. Average above-ground biomass in g m^{-2} was calculated and converted in to kg ha^{-1}.

3.1.4.7 Seed yield, above-ground biomass at harvest and harvest index

The seed yield and biomass of all three mustard cultivars sown on three different dates were measured after the harvest of the crops. The 1 m^2 area has been identified and cut just above the ground surface at the harvesting time. The weight of dried plant samples from the identified area showed the biomass at the harvest time. Crop seed yield was measured after the thrashing of a sampled area with spring balance after drying the crop for 5-7 days under bright sunshine. The seed weight was measured after threshing. Finally, the average seed yield in g m^{-2} was calculated and converted into kg/ha. Seed yield and thousand-seed weight measured after thrashing and winnowing by a small mechanical thrasher. Harvest index (HI) was calculated as:

$$HI(\%) = \frac{Seed\ yield}{Total\ biomass} * 100 \qquad \text{--------------------- (4)}$$

3.1.4.8 Oil content

Indian mustard is an oilseed crop, so it is most needed to estimate oil content. Nuclear Magnetic Resonance (NMR) Method is a non-destructive method for oil estimation. Hence we used low-resolution pulsed H1NMR (model no- PC20 Brukermade, frequency 20MHz) for seed oil estimation of each plot in the Nuclear Research Laboratory, ICAR IARI. For this purpose, 10g of dry and clean seeds from each plot were kept for drying at 105°C in the oven and then kept in a desicator till measurement was taken. About 2-3g desiccated seeds were inserted into the NMR and the signal was recorded. Oil content (percent) was determined using a standard calibration curve. Since the sensitivity of the NMR instrument depends on any change in instrumental components, air temperature and relative humidity, the instrument was calibrated before taking reading each time. The equation of NMR calibration curve with oil content is given as follows:

$$Oil\ content(\%) = \frac{(Signal + intercept)}{(Weight\ of\ seeds \times slope)}\ X\ 100 \qquad \text{------------------ (5)}$$

$$Y=1.3156 \times X-0.0727\ (R^2=0.99) \qquad \text{------------------ (6)}$$

Where, X is the seed weight in grams; and Y is the percent oil content

3.1.5 Computation from field observation

Different thermal indices, radiation use efficiency and water productivity were calculated from temperature, soil moisture and radiation, respectively. The detailed description has been shown below:

3.1.5.1 Thermal indices

Different thermal indices were calculated at physiological maturity under different conservation practices along with conventional practice as given by the equations in table 3.2.

Table 3.2: Calculation for different heat indices

Sl.no.	Indices	Computation	Reference
1	Growing degree days(GDD)	$=\sum\{\{(T_{max}+T_{min})/2\}-T_b\}$	(Iwata,1984)
2	Helio thermal units(HTD)	$=\sum$(DD× SSH)	(Rajput,1980)
3	Photo thermal units(PTU)	$=\sum$(DD× Day length)	(Major *et al.*,1975)
5	Heat use efficiency(HUE)	=Yield/GDD	(Haider *et al.*, 2003)
6	Photo thermal index(PTI)	GDD/Growing day	(Haider *et al.*, 2003)

Where T_{max} is the maximum air temperature during the crop period; T_{min} is the minimum air temperature during the crop period; DD is the degree days and SSH is the bright sunshine hours.

3.1.5.2 Seasonal evapotranspiration (ET) and water productivity (WP)

Consumptive use of water (Evapotranspiration) from each plot was computed by field water balance equation) for mustard crop 2016-17 and 2017-18 (Lenka *et al.*, 2008; Bandyopadhyay *et al.*, 2010 b) as given below:

$$ET = (R + I + C_R) - (r + d + \Delta S) \qquad \text{------------------ (7)}$$

Where, ET is crop evapotranspiration in mm, R is rainfall in mm, I amount of irrigated water applied on each plot in mm, C_R is the upper capillary rise of groundwater in mm, r is the runoff in mm, d is the deep percolation in mm and ΔS is the storage of soil moisture within soil profile in mm.

There are some assumptions to calculate ET by field water balance equation. Upper capillary rise assumed to zero due to low level of ground water table. Runoff water amount is negligible due to presence of bund at each experimental plot. Deep percolation is also assumed to be zero because there is no water saturation condition in the field. The amount of rainfall, irrigation and soil moisture storage are required to measure seasonal evapotranspiration. Thus we can simplify the equation 7 into equation 8.

$$ET = (R + I) - \Delta S \qquad \text{------------------ (8)}$$

Water productivity is calculated with the help of seasonal evapotranspiration. It is the ratio of crop production with respect to evapotranspiration. The crop production may be in terms of seed yield or biomass. The measured unit of water productivity is g/m^2/mm. Water productivity in terms of seed yield (WP_y) and biomass (WP_b) has been shown below:

$$WP_y = \frac{Seed\ Yield}{Crop\ water\ use} \qquad \text{------------------ (9)}$$

$$WP_b = \frac{Final\ above\ ground\ Biomass}{Crop\ water\ use} \qquad \text{------------------ (10)}$$

3.1.5.3 Radiation use efficiency (RUE)

Radiation use efficiency (RUE) is defined as the amount of seed yield and biomass production using each unit of intercepted photosynthetically active radiation (IPAR). The measuring unit of RUE is g MJ^{-1}. RUE in terms of seed yield (RUE_y) and biomass (RUE_b) of the crop were estimated using the following equation:

$$RUE_y = \frac{Seed\ Yield}{TIPAR} \qquad \text{------------------ (11)}$$

$$RUE_b = \frac{Final\ above\ ground\ Biomass}{TIPAR} \qquad \text{------------------ (12)}$$

Where, TIPAR is daily accumulated IPAR during the crop growing period.

3. 2 Crop yield prediction by InfoCrop model at experimental and farmer's fields

3.2.1 Model structure description

InfoCrop is a generic model which simulates the effect of genotype, weather, agronomic management, water, nitrogen, carbon and pests on the growth, development and yield of the crop in tropical agro-environments (Aggarwal *et al*., 1994). The InfoCrop-wheat v2.1 was used in this study. The source code was provided by the model developer from CESCRA, ICAR-IARI. The model was written in Fortran Simulation Translator (FST) language (Van Kraalingen *et al*., 1995). The compiler FST Win 4.2 was used to compile the program. The basic structure of InfoCrop involving growth and development processes follows the structure of MACROS (Penning de Vries *et al*., 1989). The details on the structure of the model and processes accounted by the model can be found in Aggarwal *et al*. (2004). InfoCrop is a production level 4 model, a relational diagram illustrates a flow of information and materials among the state variables of different sub-models given in Figure 3.2. The rates, quantities and auxiliary variables are represented as valves, rectangles and circles, respectively. Solid line show a flow of materials and dashed lines show a flow of information.

In the model, the phenology of the crop is calculated based on thermal time accumulated during three phases viz., sowing to seedling emergence, seedling emergence to 50 % flowering and 50 % flowering to physiological maturity. The photo-period and water stress modify the accumulated thermal time. During sowing to seedling emergence, water stress delays emergence in many crop plants. The thermal time can be increased depending upon the available water fraction in the surface soil layer. During vegetative and reproductive stages, rate of development is modulated by water and nitrogen stress dependent temporary function, MAXSTD (Aggarwal *et al*. 2006a).

In the initial stage of development (when LAI is less than 0.75), leaf growth rate is mainly influenced by temperature and moderated by nitrogen stress and not by water stress. Thereafter, growth rate in LAI (RLAI) is calculated based on initial LAI (LAII), leaf area growth rate (GLAI), death rate of LAI (DLAI) and net loss of LAI due to pests (LALOSS) and transplanting (Aggarwal *et al*. 2004).

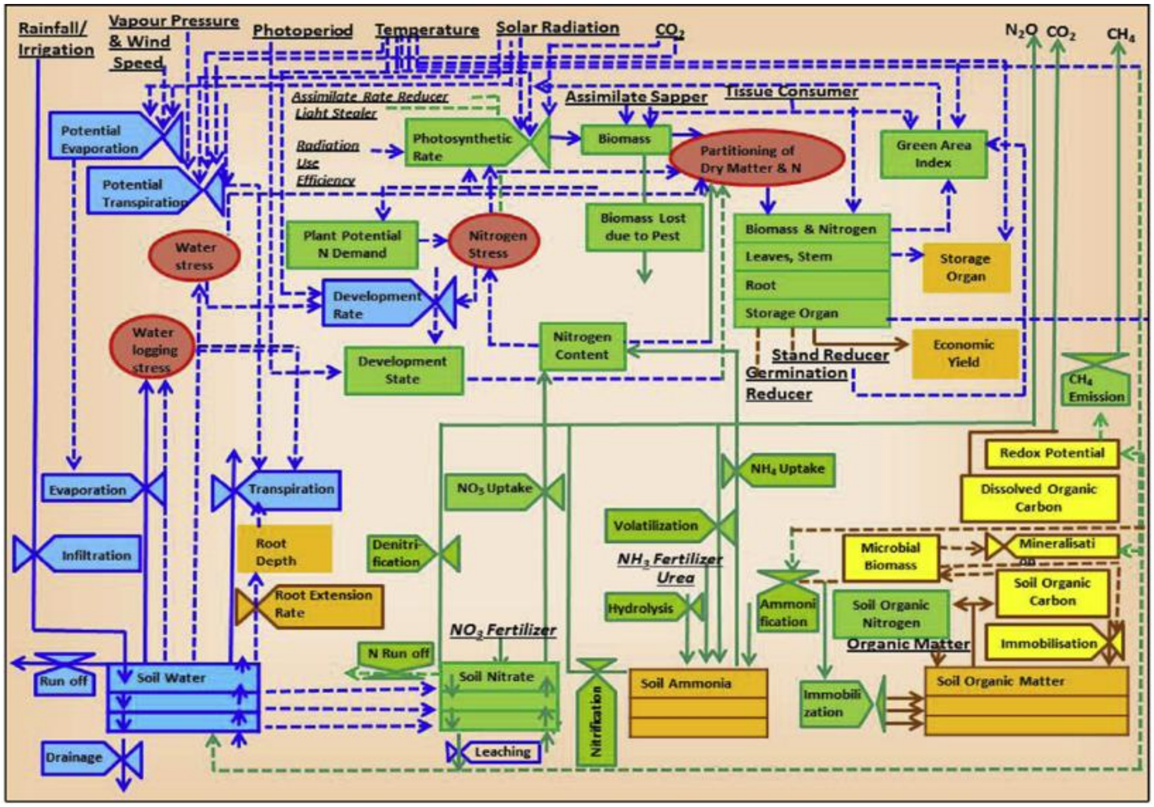

Figure 3.2: The relational (Forrester) diagram of InfoCrop-mustard model describing interrelationship and response among different soil and physiological processes

$$RLAI = LAII + GLAI - DLAI - LAI * (1 - PLTR) - LALOSS \quad \text{------------------ (13)}$$

Simulation of senescence (DLAI) is based on several empirical constants relating to shading, ageing, nitrogen mobilization, temperature, water stress and death due to pests and diseases. The water and nitrogen stress accelerates senescence depending upon its severity.

InfoCrop utilizes the radiation use efficiency (RUE) based approach for dry matter production. Maximum RUE (RUEMAX) is input in the model as a function of crop/cultivar. The RUEMAX of plant is affected by abiotic (temperature, CO_2, nitrogen stress and water stress) and biotic factors. Delay in sowing reduces RUE almost in proportion to severity.

Based on developmental stage-dependent crop-specific functions, dry matter available for crop growth is partitioned into roots, leaves, stems and storage organ. Roots get the priority for allocation and get increased in case the crop experiences water or nitrogen stress. The remaining dry matter is allocated to above-ground shoot from which a fraction is allocated to leaves and stems. The balanced dry matter is automatically allocated to the storage organ. The number of storage organ and grain filling rate is not directly influenced by water stress, as it influences the effect of water stress on dry matter production. Source-sink balance is considered in determining grain yield. The formation of storage organ occurs during the crop-specific period only. The number of storage organ form during this period is calculated by a crop-specific factor that relates the grain number per unit growth and is limited by a maximum number of grains that can be attained by a crop in given environmental conditions. Once the number of grains determined, they are filled up with a rate dependent upon temperature driven potential grain filling rate and level of dry matter available for growth. The growth of the storage organ stops under three conditions: (1) Sink limitation- when weight of storage organ attains the potential grain weight or (2) Source limitation- if there is no dry matter available, or (3) crop attains thermal time driven physiological maturity.

InfoCrop accounts for crop production under water-limited conditions with soil water balance and the effect of water stress factors on crop growth and phenology. It considered the soil depth into three layers of variable thickness. The rate of change in soil water (dW) at any given day for all layers is calculated by the following equation:

$$dW = Irrigation + Rainfall + Capillary\ Rise - Evapotranspiration - Interception - Percolation - Drainage - Runoff \quad \text{-------- (14)}$$

Redistribution of water in the soil is simulated by the tipping bucket approach with a time step of one day. The water stress factor is calculated as:

$$\text{Water stress factor} = \frac{\text{Actual water uptake}}{\text{Potential transpiration}} \quad \text{------------------ (15)}$$

The value of water stress factor ranges from zero (maximum water stress) under depleted soil water conditions to one (no water stress) when actual water uptake is equal to potential transpiration.

The rate of nitrogen uptake by a plant is determined by its developmental stage, nitrogen demand, availability of soil nitrogen, actual transpiration, rooting depth and soil water status. Nitrogen demand for different plant parts viz., root, stem, leaves and storage organs, are calculated separately depending upon rate of growth and maximum N concentration they can accumulate. Distribution of net nitrogen taken from soil to different plant parts are determined by their relative demands. Actual crop N uptake by different plant organs is the minimum of allocation of N and soil N availability. The actual N content of different organs is calculated by integrating their relative rate of change. The latter is the balance of the N partitioned to the organs, N transferred to the SO, and N lost in senescent tissues and due to transplanting and pests.

Nitrogen stress is based on the potential and actual levels of N in different plant parts, analogous to actual/potential transpiration ratio used for determining water stress factor. It also decreases transpiration and raises canopy temperature accelerating phenological development. Between water and nitrogen stresses, the one that is more severe affects the rate of crop development. Radiation use efficiency decreases as nitrogen concentration decrease from an optimal level. N stress also affects the partitioning of dry matter and senescence as with water stress.

The genotype of a crop in InfoCrop is characterized by the thermal time required to attain a phenological stage, factors for leaf/canopy development, factors for light assimilation and source-sink balance of a mustard default variety (Table 3.3).

3.2.2 Model initialization, parameterization and evaluation

Models differ in level of complexity describing crop development, main growth modules driving the biomass simulation, yield formation and in the number of input parameters. Inputs consist of weather data, crop and soil characteristics, and management practices that define the environment in which the crop will develop. The list of input parameters required by the InfoCrop-mustard model is given in annexure-I.

Table 3.3: Varietal characteristics required by InfoCrop-mustard model

Parameter	Unit	Typical value for mustard variety
Base temperature for sowing to germination	°C	5
Thermal time for sowing to germination	°C days	110
Base temperature for germination to 50% flowering	°C	5
Thermal time for germination to 50% flowering	°C days	840
Base temperature for 50% flowering to physiological maturity	°C	5
Thermal time for 50% flowering to maturity	°C days	1150
Optimal temperature for development	°C	24
Maximum temperature for development	°C	40
Sensitivity to photoperiod (between 0.5 and 1.5)	Scale	1
Relative growth rate of leaf area (an indicator of early vigour)	$(°C\ d)^{-1}$	0.008
Specific leaf area	Ha leaf kg^{-1} leaf	0.0022
Extinction coefficient of leaves at flowering		0.6
Radiation use efficiency	g Mj^{-1}	3.62
Root extension growth rate	mm day^{-1}	35
Index of greenness of leaves (between 0.8 and 1.2)	Scale	1
Sensitivity to flooding (between 1.0 and 1.2)	Scale	1
Index of storage organs formation: Slope of the relationship between storage organs number m^{-2} and dry matter accumulated during their formation stage	Number kg dry $matter^{-1}$	3000000
Potential weight of the storage organs	mg storage $organ^{-1}$	8
Nitrogen content of storage organs	%	0.039
Sensitivity of storage organs setting to low temperatures	Scale	1
Sensitivity of storage organs setting to high temperatures	Scale	1

Parameterization of model has to be done before running the model for setting up the growing environment of a crop/variety. The parameters are calibrated to suit the model for local conditions. A variety grown in a simulation model has to be calibrated through the specification of cultivar specific genetic coefficients. InfoCrop-mustard model was parameterized and calibrated for the three different mustard cultivars in sandy loam soil under semi-arid conditions. The details of the varietal genetic coefficients (TTGERM, TTVG, TTGF, SLAVAR, RGRPOT, KDFMAX, POTGWT, RUE) used for calibrating the model are discussed in the next chapter. Default value for other genetic coefficients was used as given in InfoCrop model. The model was calibrated for the first sown cultivar during 2016-17. Model was calibrated for days to emergence, days to 50 % flowering, days to physiological maturity, growth profile of LAI, maximum LAI, biomass and seed yield.

The field experimental data taken under different thermal environments (created through crop sown under different date of sowing) for three different mustard cultivars during the *Rabi* season 2016-17 and 2017-18 were used to validate the model outputs.

3.2.3 Farmer's fields study

3.2.3.1 Study area

The study was conducted for the farmers' fields situated in the Sitara and Mukundpura village of Bharatpur district, Rajasthan, India. The GPS position (latitude and longitude) was recorded at the center of every selected field and their pictorial position are shown in plate 3.3 with the help of google earth pro. According to ICAR classification, Bharatpur district comes under "Flood prone eastern plain" agroclimatic zone. Mainly rainfall received from south-west monsoon in the summer, but western disturbances also bring some rainfall during winter. Mustard is a major crop of *Rabi* season for this area.

3.2.3.2 Soil

The soil of the farmer's field area is non-calcareous, alkaline and yellow to dark brown in color. Soil texture varied from sandy-loam to clay loam. The clay content is more than 50 percent in these areas. The detailed description of soil physical properties at Kumher, Bharatpur district, is shown in table 3.4.

Plate 3.3 Study area of the farmers' fields in the Mukundpura and Sitara village near KVK Kumher, Bharatpur, Rajasthan

Table 3.4: Physiochemical properties of soil at Sitara and Mukundpura village

Parameters	**Unit**	**Layer 1 (0-30 cm)**	**Layer 2 (30-60 cm)**	**Layer 3 (60-90 cm)**
Sand content	(%)	24.5	25.0	23.0
Silt content	(%)	20.5	25.0	27.0.
Clay content	(%)	55.0	50.0	52.0
Organic Carbon	(%)	0.40	0.33	0.30
Saturated Hydraulic conductivity	mm/day	89.3	84.6	75.7
Bulk Density	mg/m3	1.30	1.42	1.61
PH of Soil		8.1	8.3	8.5

3.2.3.3 Field observations and measurements

There were 20 farmers selected to validate InfoCrop-mustard model from Mukundpura and Sitara village, Bharatpur. Farmers have varietal, sowing date and management practices variation during mustard growing season among the area**.** The information on crop variety and management practices were collected by interaction with the farmers. Soil moisture at 0-15, 15-30, 30-45, 45-60, and 60-90 cm depth in the farmers' field was recorded by gravimetric method at 15 days interval. Multiple LAI measurements were taken randomly in each of the selected field to account for with-in-field variability. The LAI was measured non-destructively by using plant canopy analyzer (LAI-2000) instrument (Welles and Norman, 1991). The average field mustard LAI on a given date was computed by averaging multiple LAI observations of that field after excluding outliers. Two samples of mature mustard crop were harvested from 1x1 m^2 area in each plot and allowed to dry in air. The weight of total biomass in each plot was measured using a spring balance. After thrashing and winnowing by a small mechanical thrasher, the seed weight was taken to estimate seed yield. All the observations on farmer's field were taken at 15 days interval with the help of KVK, Khumer, Bharatpur.

3.3 Crop yield prediction by empirical models

3.3.1 Data collection

The long-term weather data was collected by India Meteorological Department (IMD) and National Climate Data Center (NCDC), while mustard yield data was collected from Directorate of Economics & Statistics (DES) and state agricultural department for Delhi and major mustard growing region of the Rajasthan. The National Agricultural Research Project (NARP) of ICAR defined the agro-climatic zones on the basis of soil type, temperature, rainfall and geological constraints and divided the country into 127 agro-climatic zones.

According to NARP report, Rajasthan state has ten agro-climatic zones. Five major mustard growing zones are selected for yield predictions (Fig. 3.3.). The climatic condition of the Delhi region has been discussed in the earlier part of this chapter. A detailed overview of studied zone of Rajasthan is presented in table 3.5. Mustard crop is mainly grown during the *Rabi* season due to favourable weather variables.

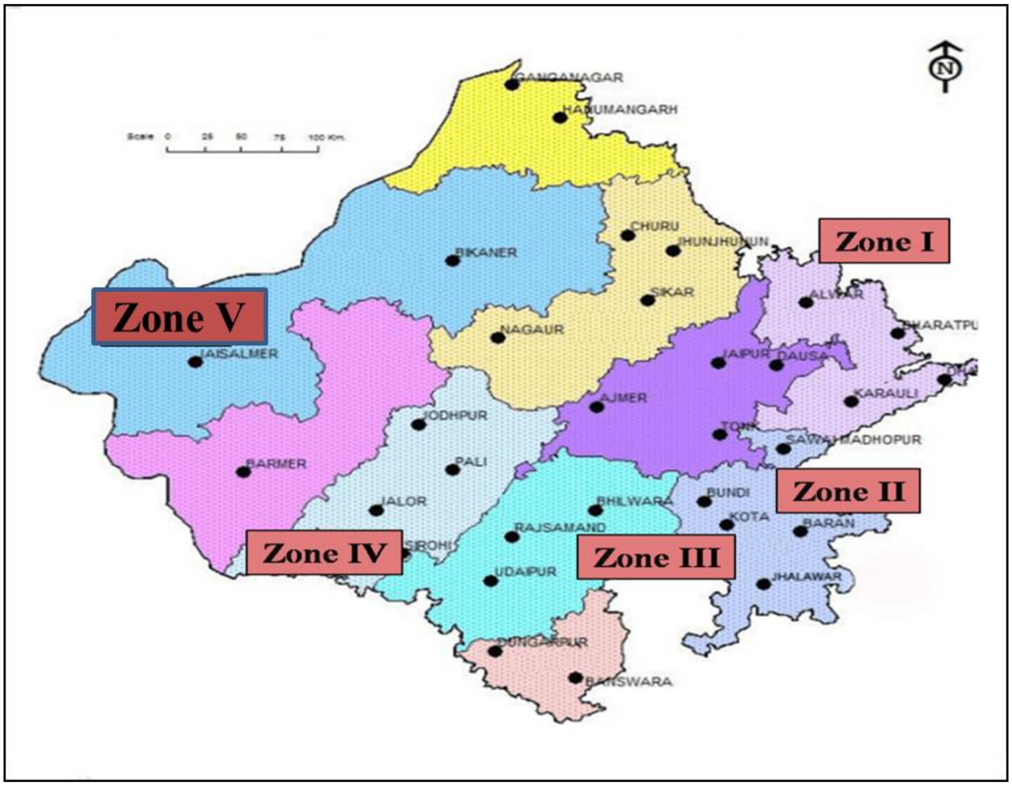

Fig 3.3 Selected zones of Rajasthan for crop yield prediction

Table 3.5: Overview of studied zone of Rajasthan

Parameters		Zone I	Zone II	Zone III	Zone IV	Zone V
Districts		Alwar, Bharatpur	Kota, Sawaimadhopur	Udaipur, Jhalawar	Jodhpur, Pali	Bikaner
Total area (mha)		2.77	2.70	3.36	3	7.7
Avg. Rainfall (mm)		500-700	650-1000	500-900	300-500	100-350
Temp (°C)	**Maximum**	40.0	42.6	38.6	38.0	48.0
	Minimum	8.2	10.6	8.1	4.9	3.0
Soil texture		Clay loam	Clay-loam	Loam	Sandy-loam	Loamy-sand
Climate		Semi-arid	Humid	Sub-humid	Sub-humid	Hyper-arid

Source: http://www.agriculture.rajasthan.gov.in

3.3.2 Preprocessing of data

Missing value in data is the first obstacle to predicting crop yield by long-term weather and yield data. Data gaps were filled using maximum likelihood estimation, which operates by estimating a set of parameters that maximize the probability of getting the estimates from the sample data that is analyzed. It provides a deterministic result (Collins *et al*. 2001). Gap filling of maximum and minimum temperature has been done by Hmisc package while miss-forest package was used for gap filling of relative humidity in R statistical software. Hmisc has two types of functions first is "impute" and other is "aregImpute". Among the both, "aregImpute" works better than impute function because it is based on bootstrapping and predictive mean matching while impute function is based on median. Missforest is a nonparametric imputation method, which is based on random forest mechanism.

Mustard seed yield shows an increasing trend over a long time series data (Fig 3.4). The increasing trend is generally due to improvements in crop production technology over time, such as the introduction of high-yielding/stress tolerant cultivars, higher applications of input resources, and better technology for intercultural operations (Aggarwal *et al*., 2000; Nain *et al*., 2004). To understand the behavior of weather variables on mustard yield and overcome the technological effects, a modification has been proposed here. The modified mustard yield is named as scaled normalized yield. The formula of scaled normalized yield is shown below:

$$Scaled\ normalized\ yield = \frac{Normalized\ yield_i - Normalized\ yield_{min}}{Normalized\ yield_{max} - Normalized\ yield_{min}} \text{ ------------------ (16)}$$

Where, Normalized yield$_i$, Normalized yield$_{max}$, and Normalized yield$_{max}$ are the normalized yield deviations from yield trend for current period, maximum normalized yield and minimum normalized yield among the whole data set, respectively. The normalized yield is calculated as:

$$Normalized\ yield = \frac{Yield_i - Yield_{trend}}{Yield_{trend}} \text{ ------------------ (17)}$$

Where, Yield$_i$ is crop yield of the current period; Yield$_{trend}$ is the trend predicted yield for each year. Yield $_{trend}$ has been calculated by developing a linear regression between yield and time series, such as:

$$Yield\ _{trend} = a + b * Time \text{ ------------------ (18)}$$

Where, a is the intercept and b is the slope of a linear regression between yield and time.

The resultant scaled normalized yield does not show any increasing trend and the mean value of that is nearly 50 percent with zero value of the coefficient of determination (Fig 3.4). It removed the impact of developed technology on crop yield and gave a better representation of weather variables' effect on crop yield. If there is no time trend is present in crop yield data than scaled normalized yield will be equal to observed yield.

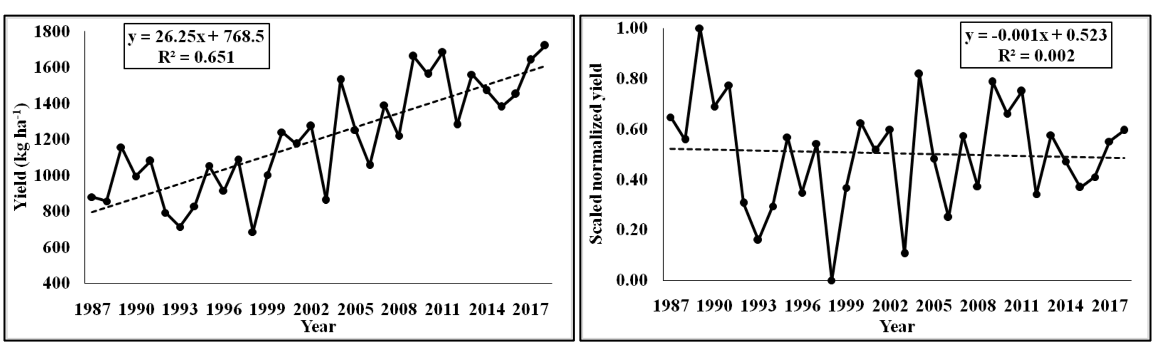

Fig 3.4 Temporal plots of (a) crop yield and (b) scaled normalized crop yield

3.3.3 Weather indices calculation

After calculating scaled normalized yield we developed the Z variables, which were used as predicted variables. There were six weather variables used to find out z variables such as maximum and minimum temperature (T_{max} and T_{min}, °C), morning and evening relative humidity (RH_{max} and RH_{min}, %), rainfall (mm) and bright sunshine hour (SSH, hr) during 40th to 13th standard meteorological week. There are two Z variables: simple Z variables and weighted Z variables. Simple Z variables were developed by summing each weather variable or weather variables interactions between 40th to 13th SMW of each year. Weighted Z variables were developed by the sum-product of weather variable or weather variable interactions to their correlation with adjusted yield. The simple and weighted Z variables were computed by following equations.

$$Z_{ij} = \sum_{w=1}^{m} X_{iw} \qquad \text{------------------ (19)}$$

$$Z_{ij} = \sum_{w=1}^{m} X_{iw} X_{ii'w} \qquad \text{------------------ (20)}$$

Where, Z_{ij} is the simple Z variable; X_{iw} and $X_{ii'w}$ is the value of i^{th} weather variable and their interaction with i' variable for w standard meteorological week; m is the standard meteorological weeks used for model development.

$$Z_{ij'} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw} \qquad \text{------------------ (21)}$$

$$Z_{ij'} = \sum_{w=1}^{m} r_{ii'w}^{j} X_{iw} X_{ii'w} \qquad \text{------------------ (22)}$$

Where, $Z_{ij'}$ is the weighted Z variable; r^{j}_{iw} and $r^{j}_{ii'w}$ are the correlation coefficients of yield with i^{th} and weather variables interaction with i' variable for w standard meteorological week. The details of simple and weighted Z variables are shown in table 3.6.

Table 3.6: Simple and weighted weather indices used for developing mustard prediction model by different techniques

	Simple weather indices						Weighted weather indices					
	T_{max}	T_{min}	Rain	RH_{max}	RHmin	SSH	T_{max}	T_{min}	Rain	RH_{max}	RH_{min}	SSH
T_{max}	Z10						Z11					
T_{min}	Z120	Z20					Z121	Z21				
Rain	Z130	Z230	Z30				Z131	Z231	Z31			
RH_{max}	Z140	Z240	Z340	Z40			Z141	Z241	Z341	Z41		
RH_{min}	Z150	Z250	Z350	Z450	Z50		Z151	Z251	Z351	Z451	Z51	
SSH	Z160	Z260	Z360	Z460	Z560	Z60	Z161	Z261	Z361	Z461	Z561	Z61

Where, Z10 is summation of maximum temperature of each SMW for crop growing period. Like that, we use Z20 for minimum temperature, Z30 for rainfall, Z40 for morning relative humidity, Z50 for evening relative humidity and Z60 for bright sunshine hour. Z120 means the summation of interaction of maximum and minimum temperatures for each SMW. Other Z variables were computed by interaction of Tmax*Rain (Z130), Tmax*RHmax (Z140), Tmax*RHmin (Z150), Tmax*SSH (Z160), Tmin*Rain (Z230), Tmin*RHmax (Z240), Tmin*Rhmin(Z250), Tmin*SSH (Z260), Rain*RHmax (Z340), Rain*RHmin (Z350), Rain*SSH (Z360), RHmax*RHmin (Z450), RHmax*SSH (Z460) and RHmin*SSH (Z560).

Z11 is weighted Z variable for maximum temperature of each SMW for crop growing period. Like that, we use Z21 for minimum temperature, Z31 for rainfall, Z41 for morning relative humidity, Z51 for evening relative humidity and Z61 for bright sunshine hour. Z121 means the summation of interaction of maximum and minimum temperature for each SMW. Other Z variables are computed by interaction of Tmax*Rain (Z131), Tmax*RHmax (Z141), Tmax*RHmin (Z151), Tmax*SSH (Z161), Tmin*Rain (Z231), Tmin*RHmax (Z241), Tmin*RHmin(Z251), Tmin*SSH (Z261), Rain*RHmax (Z341), Rain*RHmin (Z351), Rain*SSH (Z361), RHmax*RHmin (Z451), RHmax*SSH (Z461) and RHmin*SSH (Z561).

3.3.4 Selection and extraction of variables

These Z variables are very closely correlated to each other. Sometimes the irrelevant variables developed a good agreement and increased complexity in the model. So, it is important to reduce the correlation to avoid the over-fitting problem in model development. Multiple linear regression technique is the most widely used for yield prediction. But, it can not be employed for long-term time series data because number of computed Z variables are more than the number of yield data. Therefore variable selection and extraction is necessary prior to reduce the dimensionality of the data in crop yield prediction modelling. Stepwise multiple linear regression (SMLR) models ran in SPSS for selection of highest important variables. It is a combination of forward and backward regression method. A most effective variable is considered for the addition and the subtraction of least important variable based on R^2 and F-test. The principal component analysis model ran in SPSS to extract the variables. The principal components (PCs) selected on the basis of eigen values (>1) were able to describe more than 90 percent variability of the input data set. A new set of variables was computed by all input variables in the variable extraction technique, whereas in the variable selection technique the most significant input variables were used.

3.3.5 Machine learning techniques

There are several techniques used in crop yield prediction. Three different types of machine learning techniques (artificial neural network, support vector machine and random forest) were used to develop good crop yield prediction models for mustard crop at Delhi and major mustard growing zones of Rajasthan. All the machine learning techniques were run in the R statistical software version 3.1.3. using two-thirds of the data for training and one-third of the data for testing. A detailed discussion of the machine learning techniques is given as follows:

3.3.5.1 Artificial neural network (ANN)

Artificial neural network consists of many artificial neurons connected to a network architecture. Neural network has various architectures to approximate any linear function, such as feed-forward network, feed-back network, lateral network, etc. ANN is composed of three layers, namely input layer, hidden layer and output layer. Multilayer perceptron (MLP) technique is one of the more popular neural network types than other neural network types. The neurons are arranged in a successive pattern, through which information will flow uni-directionally from the input layer to the output layer through the hidden layer. This network interpreted as input-output model, with weights and threshold (biases) as free parameters of the model. Artificial neural network work through the optimized weighted value of variables, the method by which the optimized values are attained is called learning. In the learning process, it tries to teach to produce the output based on the corresponding input provided. Learning will complete when the trained neural network can update the optimal weights and produce the output within the desired accuracy corresponding to the input pattern. The main objective of the neural network is to produce its output having reduced discrepancies with the target output value, which will help to transform the input into meaningful output.

The main problem for implementing the ANN is to find the parameters, which are used for cross-validation, such as a number of units in the hidden layers and nodes.

We use “caret” package for cross-validation, “ggplot2” package for data visualization and “nnet” package in R statistical software to perform ANN exercises. Ten fold cross-validation has been used for prediction by ANN method using R version 3.1.3 (Kuhn 2008). Size and decay is the regularization parameter to avoid over-fitting in “nnet” package that represent the number of units in the hidden layer and parameters of weight decay in nodes, respectively. A schematically representation of the ANN model for prediction has been shown in fig 3.5.

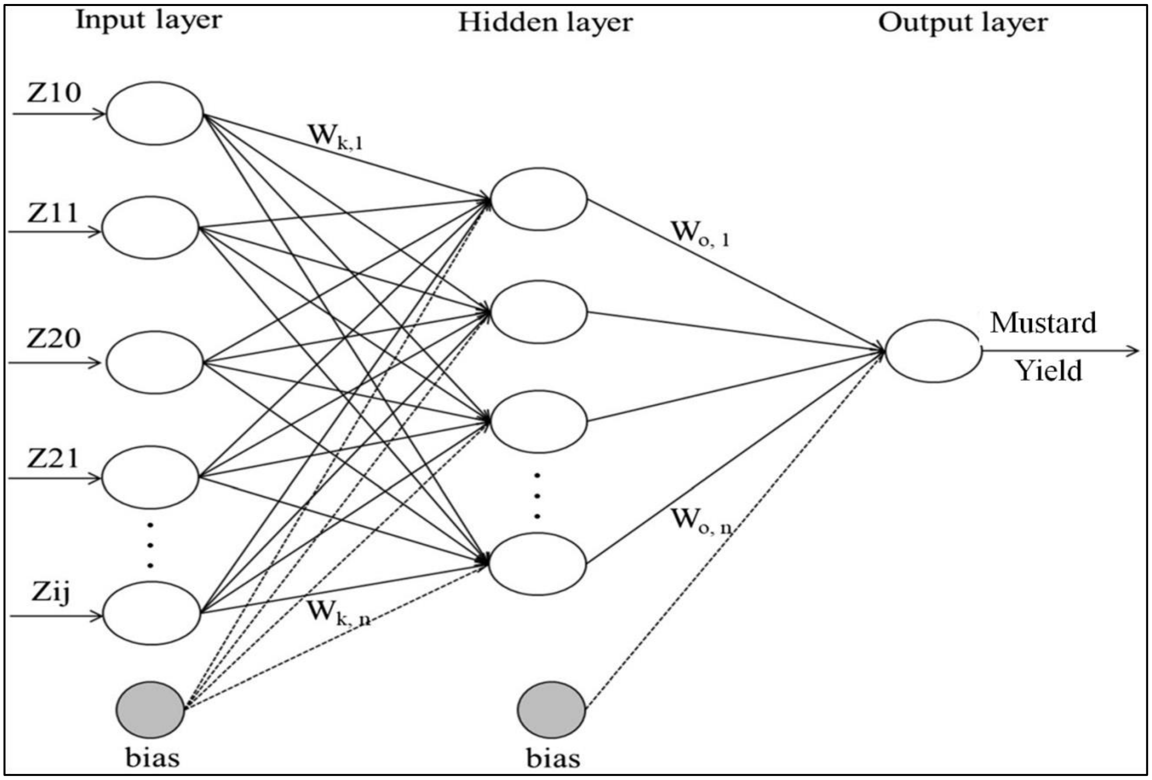

Fig 3.5 Schematically representation of the basic ANN model

3.3.5.2 Support vector machine (SVM)

Support Vector Machines (SVMs) is a kernel-based, nonparametric, supervised machine-learning technique used for to predict and classify samples in two disjoint clusters (Pal, 2009). Nonparametric methods estimate the relationship between the variable of interest and seed yield based on the regression. It is most important to know several terminologies for a better understanding of SVM techniques, such as Kernel, Hyper-plane, Boundary Line and Support vectors. Kernel is the function to convert any linear model to a nonlinear model and lower dimensional data into high dimensional data. Hyper-plane is the separation line between two input data sets and it will help to predict the continuous output. The boundary line creates the two side margins from hyper-plane. A hyper-plane separation is good if more distance covered by data variables, which indicates less prediction error. SVMs are useful tool with high accuracy for prediction and classification due to their capability to handle small training data sets. The schematically representation of SVM for crop yield prediction has been shown in Fig 3.6.

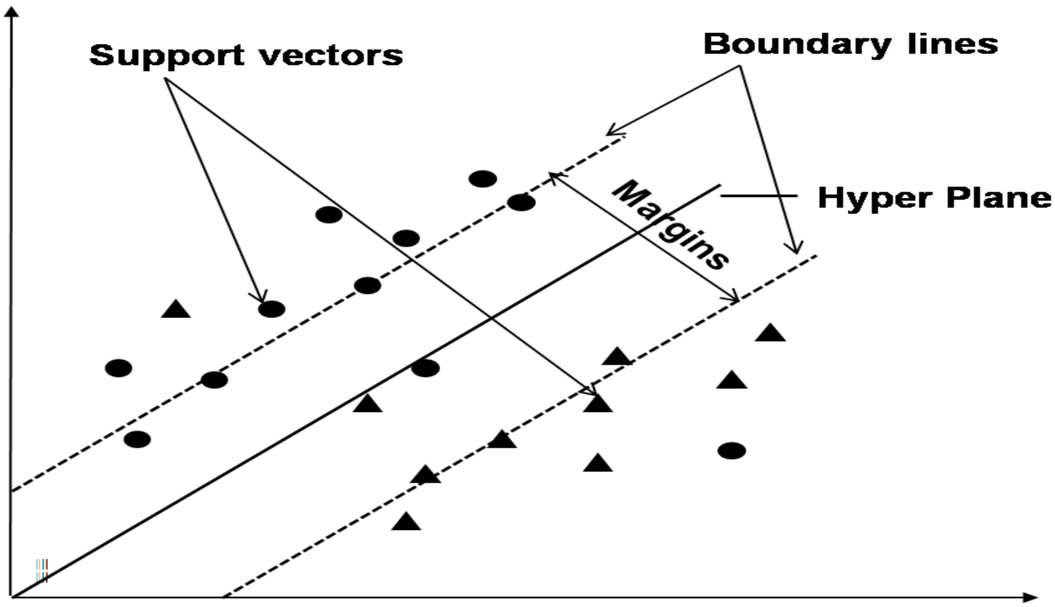

Fig 3.6 Schematically representation of the basic SVM model

In this study, The SVM approach is used to create functions from a set of labeled training data. This SVM approach used a regression function for crop yield prediction. Support vector regression (SVR) is the implementation of the SVM method for regression and function approximation (Smola and Scholkopf, 2004). We used "e1071" package for SVM analysis, "caret" for cross-validation and "ggplot2" for data visualization in R statistical software version 3.1.3. Gamma and cost parameters are used for a cross-validation. Gamma defines the distance of a single data point from the hyper-plane, whereas a value of cost decided the smoothness of hyper-plane (large C smooth boundary). A low value of gamma and a large value of cost represents more accurate model for prediction.

3.3.5.3 Random forest (RF)

Random forest is an ensemble machine-learning technique. It creates a forest to enhance the performance of a single decision tree by bootstrapping. The combination of all the trees improves th prediction. Each tree is different from another one due to the presence of nodes and branches. The schematically representation of the random forest has been shown in fig 3.7.

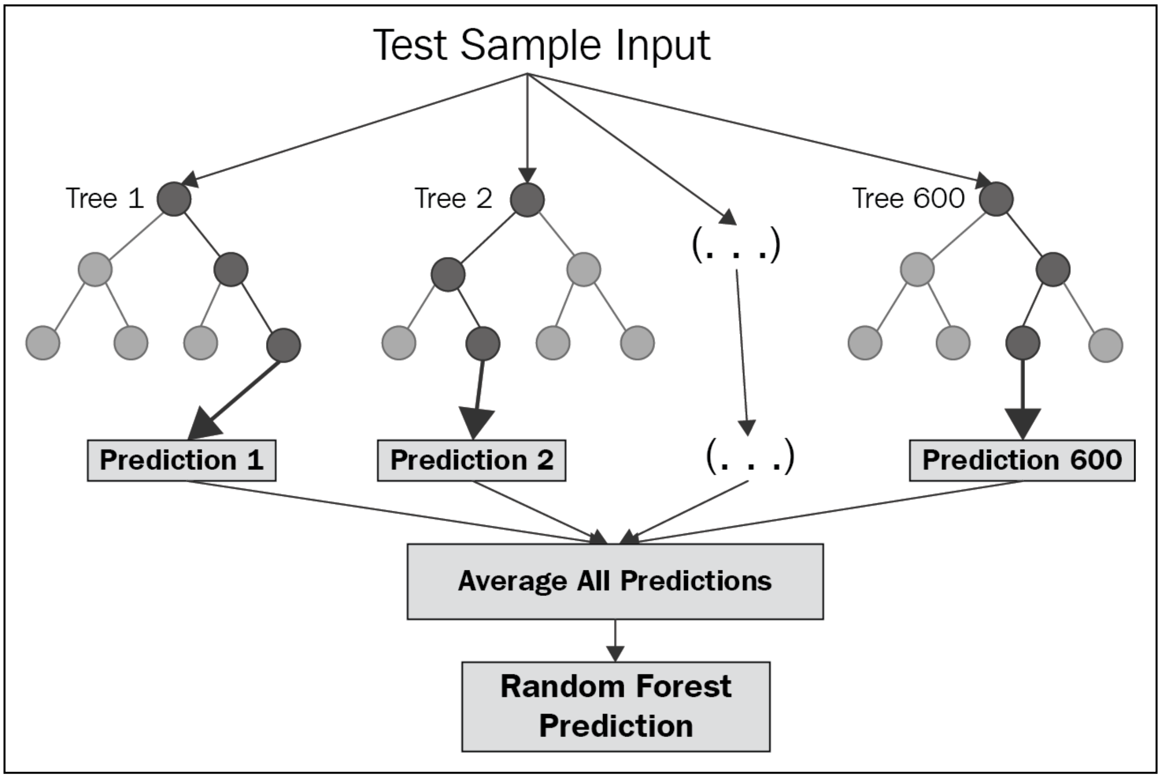

Fig 3.7 Schematically representation of the basic random forest model

Crop yield prediction by RF is done by use of "Random Forest" package. We used "caret" for cross-validation and "ggplot2" for data visualization in R statistical software version 3.1.3 (Breiman, 2001). It is most important to decide the value of ntree and mtry for RF regression. The value of ntree ensures that, every input row gets at least a few times, so that ntree should not have less value. The default value of ntree is 5. The value of mtry decides the number of variables randomly sampled at each split. The default value of mtry for classification is a square root of a number of variables and for regression the number of a variable is divided by three. More mtry showed a good agreement between trees. A flowchart of crop yield prediction by different techniques is presented in fig 3.8.

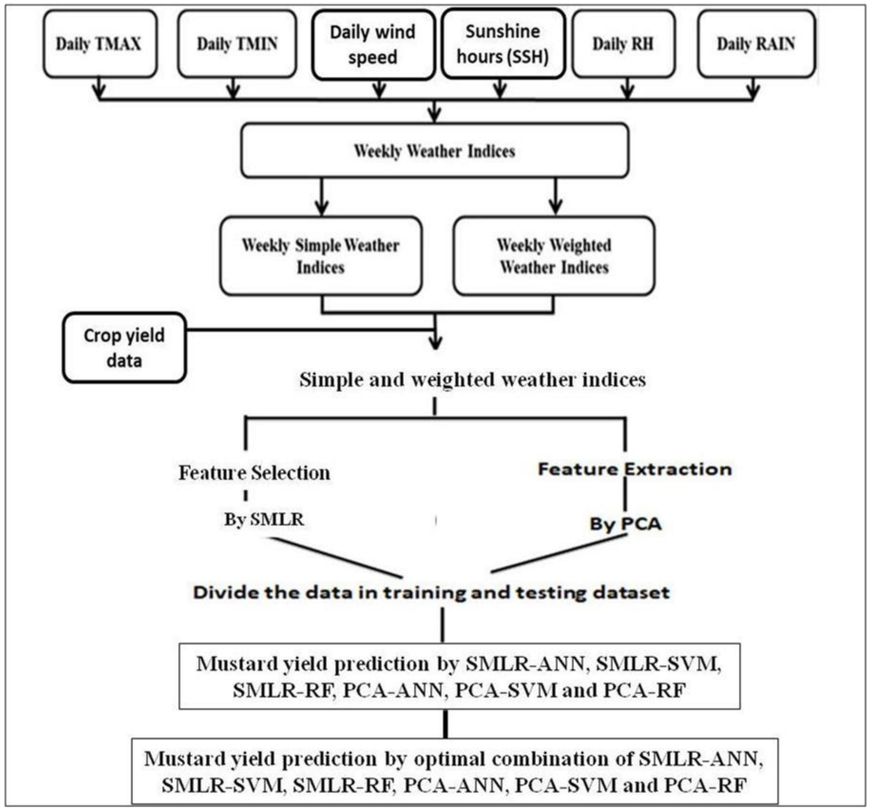

Fig. 3.8 Flowchart for developing crop yield prediction model

3.4 Optimal combination of different models

The need for optimal combinations to obtain diversified results arises because many prediction models have similar accuracy, so it is difficult to identify the best prediction model among them. The predicted results having the least root mean square error were used for optimum combination to get the better accuracy of the crop yield prediction. Optimal combination on the basis of weights to minimize the error in the crop yield prediction. The larger weights are responsible for better prediction and reduction in error.

We used the variance of a validated dataset to obtain optimal yield. The data sets were processed for analysis of variance and to develop an optimal combination model for all possible combinations of ANN, SVM and RF by using MS-Excel office 2013. The equation 23 is used for optimal combination based on variance.

$$X = \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}\right]^{-1} \left[\frac{X_1}{\sigma_1^2} + \frac{X_2}{\sigma_2^2} + \frac{X_3}{\sigma_3^2}\right] \quad \text{------------------ (23)}$$

Where, X_1, X_2 and X_3 are independent measurements (Testing data set) and σ_1^2, σ_2^2 and σ_3^2 are the variance of independent measurements (Testing data set**)** by ANN, SVM and RF, respectively.

3.5 Statistical test

The data were analyzed using SPSS (version 16.0), Excel package (version 13.0), R statistical software version 3.1.3. Analysis of variance as applicable for split-plot design was used to test least significant differences among the various treatment means and their interactions using statistical analysis. MS Excel version 13.0 was used to draw required graphs. Root mean square error, normalized root mean square error and ratio of performance to deviation were used for comparing the accuracy of the models.

3.5.1 Root mean square error (RMSE)

It measures the difference between predicted values and observed values. By this test, model performance during the calibration as well as validation period can be determined. It is also helpful in comparing individual model performance with other predictive models.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi - Oi)^2} \quad \text{------------------ (24)}$$

Where, RMSE is a root mean square error, Pi is the predicted value, O_i is the observed value and N is the number of observations

3.5.2 Normalized root mean square error (nRMSE)

If Pi, O_i, N and M are notated as predicted value, observed value, number of observations and mean of observed value, nRMSE can be written as the formula given below. Normalized mean square error expressed in percentage, values close to zero indicate better model performance. nRMSE is a measure (%) of the relative difference of estimated versus observed data. The prediction is considered excellent with the nRMSE <10 %, good if 10–20 %, fair if 20–30 %, and poor if >30 % (Jamieson *et al.*, 1991)

$$nRMSE = \frac{100}{M} * \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi - Oi)^2}$$ ------------------ (25)

3.5.3 Ratio of performance to deviation (RPD)

Viscarra Rossel *et al*. (2006) used RPD as an accuracy parameter to evaluate the prediction accuracy of the developed models. The Performance of the model based on the ratio of performance to deviation (RPD) values are shown in table 3.7 which shows that model should be recommended at most 1.4. RPD was calculated using the following formula:

RPD=Sd/SEP ----------------------------- (26)

Where SEP= standard error of estimation, which is calculated as root mean square error

$$SEP = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Oi - Pi)^2}$$

Sd= Standard deviation of the sample

$$Sd = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Oi - M)^2}$$

Where Pi: estimated values, Oi: observed value; N: number of observation, M: mean of observed value respectively.

Table 3.7: Performance of model based on ratio of performance to deviation (RPD) value

RPD value	Category	Recommendation
Less than 1.0	very poor model	not recommended for use
1.0 and 1.4	poor model	not recommended for use
1.4 and 1.8	fair model	may be used for prediction
1.8 and 2.0	good model	quantitative predictions
2.0 and 2.5	very good model	quantitative prediction
More than 2.5	excellent model	quantitative prediction

3.5.4 Percent deviation

It is the difference between predicted and observed yield with reference to observed yield. The positive value of percent deviation shows overestimation and the negative value shows underestimation of a model. Percent deviation is calculated using the following formula:

$$Percent\ deviation = \frac{Pi-Oi}{Oi}*100 \qquad \text{------------------ (27)}$$

Where, P_i is the predicted value and O_i is the observed value.

3.7.4 Mean absolute error (MAE)

Mean absolute error (MAE) of an estimator measures the average magnitude of deviation predicted data set that is, the average difference between the estimated values and what is estimated. MAE is calculated by using the following formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Pi - Oi| \qquad \text{------------------ (28)}$$

3.7.5 Coefficient of determination(R^2)

The coefficient of determination or R-squared represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one.

$$R^2 = 1 - \frac{\sum(P_i - O_i)^2}{\sum(P_i - O_{mean})^2} \qquad \text{------------------ (28)}$$

Where, P_i = predicted value, O_i = observed value, and O_{mean} = mean of observed value

4. Results

To estimate "Mustard yield prediction by machine learning and crop simulation models" experiments were conducted for mustard crop during *Rabi* season 2016-17 and 2017-18 at research farm of IARI, New Delhi. Results obtained from the field experiments and the objectives on different aspects are presented in this chapter.

4.1.1 Weather during the crop growing period

The daily weather data recorded at the IARI agro-met observatory near the experiment area during the crop growing season 2016-17 and 2017-18 were taken for analysis. The weekly average values were computed from daily observed values of weather variables such as Weekly mean maximum and minimum temperature, total weekly rainfall, mean relative humidity and bright sunshine hours.

The maximum temperature varied from 18 to 35°C, whereas minimum temperatures varied from 3 to 23 °C in both the years. The trend of the maximum and minimum temperatures is shown in Fig. 4.1. The figure clearly showed the decreasing value of the maximum and minimum temperature till 2nd standard meteorological week (SMW); afterward they followed the increasing trend till crop physiological maturity during both the growing years 2016-17 and 2017-18. It was important to notice that the maximum temperature for 2016-17 was higher than 2017-18 during the early crop growing period, but at the time of maturity, the maximum temperature for 2017-18 was slightly higher than 2016-17. On the contrary, the minimum temperature for 2017-18 was higher than 2016-17 till the reproductive stage of mustard and after that there was a drastic increment in minimum temperature for 2017-18 sown crop.

There is an inverse relationship between the temperature and relative humidity. Fig. 4.2 represent the changing pattern of RHmax and RHmin for the entire *Rabi* crop growth period during 2016-17 and 2017-18. The maximum and minimum relative humidity was completely different until the 45th SMW for 2016-17 and 2017-18. Maximum relative humidity was nearly constant throughout the crop growing season for both years. The minimum relative humidity peaks were obtained on 50th, 1st, 4th, 7th, and 9th SMW in both years. A smoother temperature and relative humidity curve represented less variation during 2017-18, whereas weather variables of 2016-17 showed more variability.

The amount of rainfall and its distribution are shown in Fig. 4.3. The rainfall received during the entire crop growing period was 119.7 mm during 2016-17 and 13.4 mm for the 2017-18. There were five rainy days in 2016-17 and two rainy days in 2017-18. A good amount of rainfall (39.1mm) received at 40th SMW in 2016-17 met the pre-sowing irrigation requirement. The year 2016-17 was relatively wet in terms of the amount and distribution of rainfall compared to 2017-18. The bright sunshine hour is important for photosynthetic activity. Weekly mean bright sunshine hours ranged between 0.2 hours at the 45th SMW and 9.0 hours at the 13th SMW (Fig. 4.4). The major drops in bright sunshine hours were recorded in 42nd, 43rd and 44th SMW due to fog/cloudiness in both years.

4.1.2 Crop phenology

The days required to achieve different phenological stages for three mustard cultivars at three different dates of sowing were recorded during 2016-17 and 2017-18 (Table 4.1). The sowing dates for mustard were 10th October, 25th October and 10th Nov ember, 2016 for the first year experiment and 12th October, 26th October and 11th November, 2017 for the second year experiment. The results clearly showed that there were more days required to attain 50 percent flowering with delay in sowing for all cultivars in both years. Delay sowing took more days for emergence and flowering compared to timely sowing of mustard crop during both years. Apart from that, there was a reduction in crop growth period with a delay in sowing for all cultivars in both years.

The 50 percent flowering occurred at 64, 51 and 63 days after sowing in timely sown crop, 67, 53 and 66 days after sowing in late sown crop and 70, 55 and 70 days after sowing in very late sown crop for RH-406, Pusa Tarak and Girraj in the timely and very late sowing during the 2016-17 crop growing year, respectively. In 2017-18 crop growing year, the corresponding days for 50 percent flowering were 65, 50 and 65 in timely sown crop, 68, 53 and 67 in late sown crop and 70, 56 and 71 in very late sown crop for RH-406, Pusa Tarak and Girraj, respectively. The physiological maturity was reached at 137, 117, 138 days after sowing and 125, 108, 128 days after sowing for RH-406, Pusa Tarak and Girraj for timely sown crop and very late sown crop in 2016-17 crop

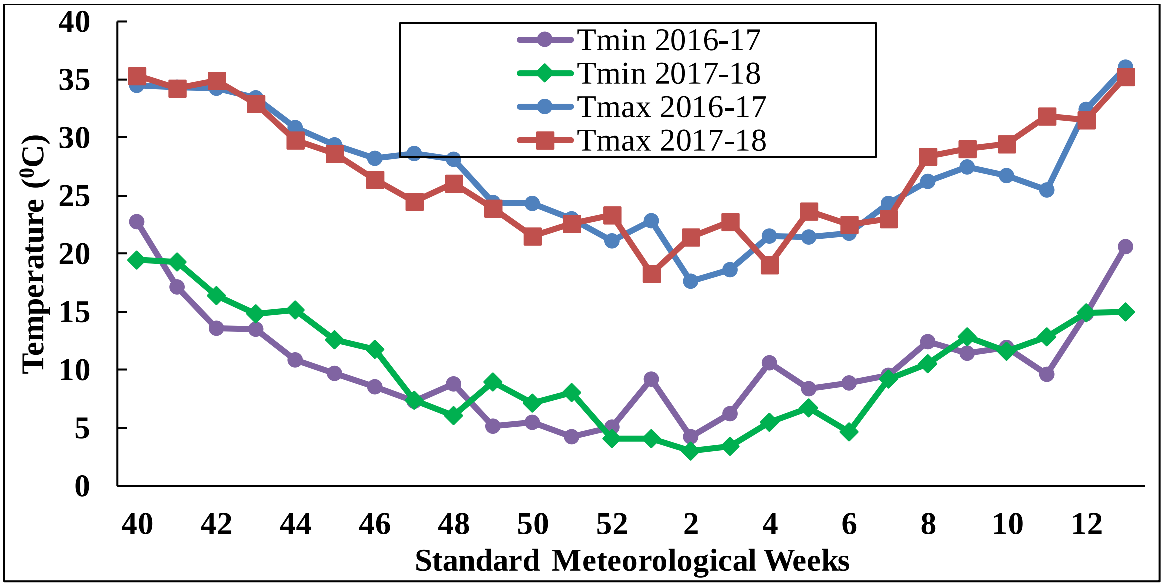

Fig 4.1 Weekly maximum and minimum temperature at IARI, New Delhi observatory during *Rabi* 2016-17 and 2017-18 crop seasons

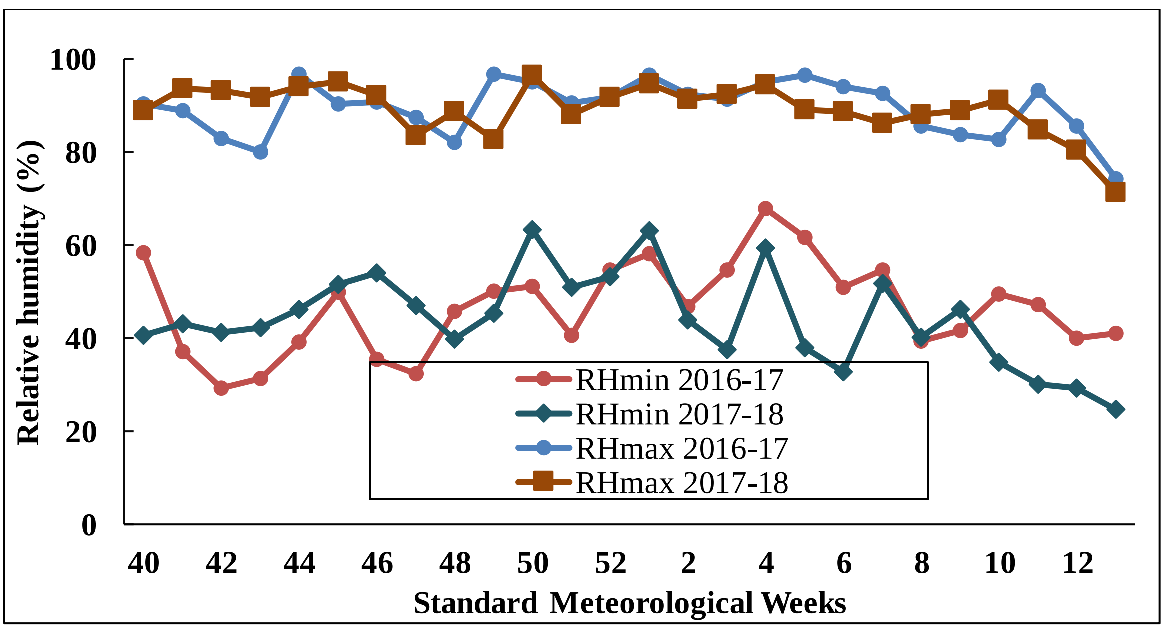

Fig 4.2 Weekly maximum and minimum relative humidity at IARI, New Delhi observatory during *Rabi* 2016-17 and 2017-18 crop seasons

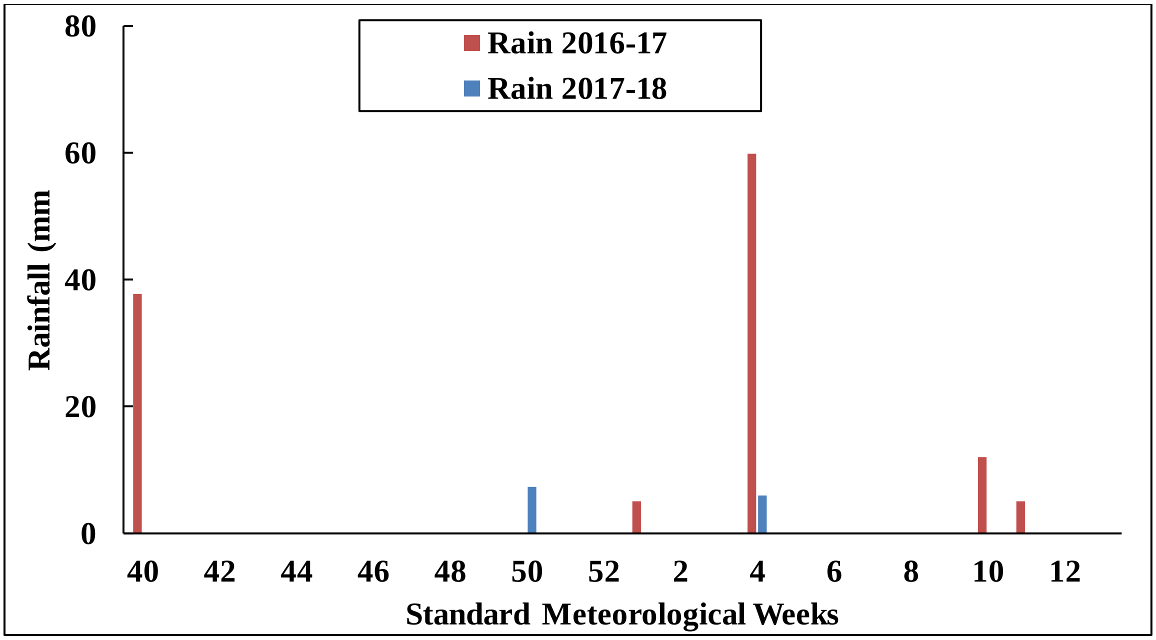

Fig 4.3 Weekly rainfall at IARI, New Delhi observatory during *Rabi* 2016-17 and 2017-18 crop seasons

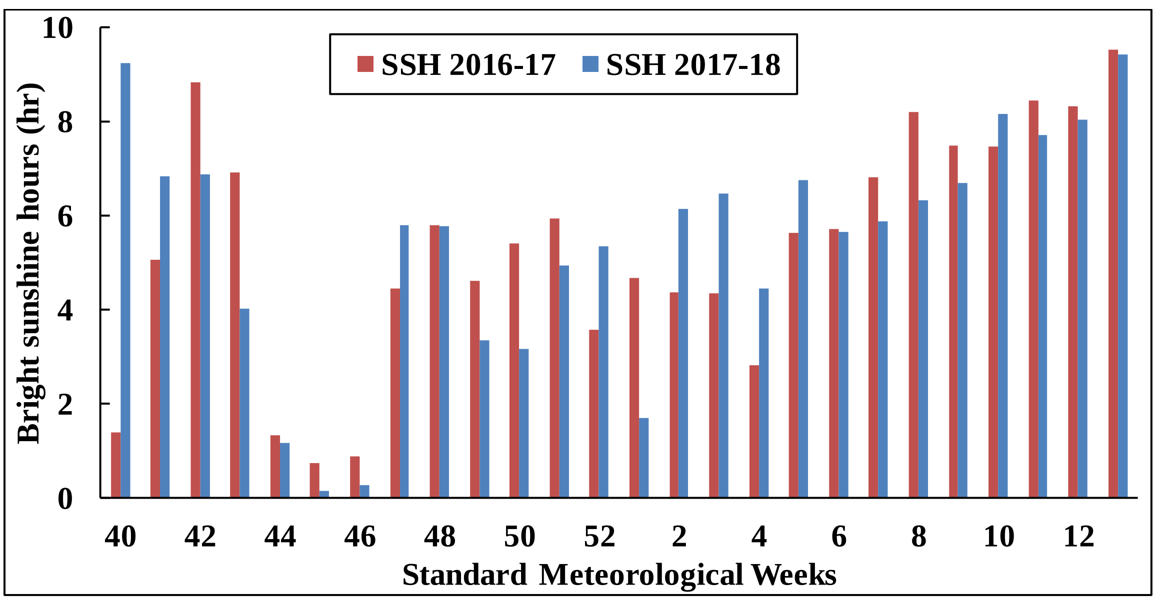

Fig 4.4 Weekly bright sunshine hour at IARI observatory during *Rabi* 2016-17 and 2017-18 crop seasons

year. Similar results were obtained for 2017-18 crop year with the addition of 2 to 4 more days to achieved physiological maturity. PusaTarak had the shortest crop growth duration among all cultivars.

Table 4.1 Crop Phenology of different mustard cultivars under different dates of sowing during *Rabi* 2016-17 and 2017-18.

2016-17									
Phenological stages (days after sowing)	**RH 406**			**Pusa Tarak**			**Girraj**		
	D1	**D2**	**D3**	**D1**	**D2**	**D3**	**D1**	**D2**	**D3**
Seedling emergence	6	7	9	4	5	7	7	8	9
Early Veg. phase	14	16	19	10	13	16	15	16	18
First flower	55	57	61	40	43	44	54	57	60
50 per cent flowering	64	67	70	51	53	55	63	66	70
Pod formation	97	95	93	76	75	74	96	94	92
Seed development	116	113	105	100	97	93	116	112	106
Physiological maturity	137	132	125	117	114	108	138	133	128
Crop harvest	146	140	131	127	124	116	146	140	131
2017-18									
Phenological stages (days after sowing)	**RH 406**			**Pusa Tarak**			**Girraj**		
	D1	**D2**	**D3**	**D1**	**D2**	**D3**	**D1**	**D2**	**D3**
Seedling emergence	6	8	9	5	6	8	6	8	9
Early Veg. phase	13	16	18	11	13	17	14	15	18
First flower	56	59	62	40	41	44	55	58	61
50 percent flowering	65	68	70	50	53	55	65	67	71
pod formation	99	98	95	78	77	75	99	98	95
Seed development	118	115	109	102	99	95	119	114	108
Physiological maturity	140	137	129	118	115	110	140	134	128
Crop harvest	149	144	134	130	125	118	149	144	134

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

4.1.3 Radiation Interception and extinction coefficient

The seasonal profiles of fIPAR for different mustard cultivars at different dates of sowing during both the years are shown in Fig. 4.5. In all the cases, fIPAR showed an increasing trend during vegetative growth, then plateauing and later decreasing with the progress of the season in the reproductive stage. The peak fIPAR value ranged between 0.89 to 0.98 in all the treatments. The peak fIPAR value occurred at 77, 75, 73 days after sowing for Pusa Tarak, 93, 89, 85 days after sowing for RH-406 and 93, 89, 85 days after sowing for Girraj at timely, late and very late sown conditions, respectively during 2016-17. Similar results were obtained during the period 2017-18. Due to the delay in sowing for all cultivars during 2016-17 and 2017-18, fIPAR had a lesser value and the peak of fIPAR occurred earlier as compared to the timely sowing. There was no significant difference in fIPAR during the early growing period, whereas the values of fIPAR were significant after the vegetative phase.

The extinction coefficient (k) was determined separately for each treatment in both years with least-square regression by calculating the slope of the relationship between ln(1- fIPAR) and LAI with an intercept set to zero. The extinction coefficient varied between 0.45 and 0.62 among the treatments in both years. No significant change was observed in extinction coefficients across cultivars and sowing dates treatments.

4.1.4 Leaf area index (LAI)

The leaf area index is an important parameter for crop growth studies since it is useful in interpreting the capacity of a crop for producing dry matter in terms of the utilization of intercepted radiation and the amount of photosynthesis synthesized. The seasonal profile of the mustard leaf area index (LAI) for all three cultivars at different dates of sowing during 2016-17 and 2017-18 are depicted in Fig. 4.6. LAI showed a similar pattern as fractional intercepted photosynthetically active radiation (fIPAR). At first, LAI increased at a slow rate during the initial growth phase and then increased with increasing rate till flowering and pod formation stages afterward, it started to follow decreasing pattern till physiological maturity due to the senescence of crop leaves. The peak value of LAI was attained at 85-90, 78-83 and 72-79 days after sowing in the timely sown, late sown and very late sown crop, respectively for RH-406 and Girraj in both the year. In case of Pusa Tarak, the peak LAI value varied from 80 to 65 days for timely and very late sown crop in both the years.

During 2016-17, the peak values of LAI were 3.9, 4.8 and 4.6 in the timely sown crop, 3.5, 4.5 and 4.25 for the late sown crop, 2.39, 3.06 and 2.94 for the very late sown crop in Pusa Tarak, RH-406 and Girraj, respectively. The peak values of LAI were 4.01,

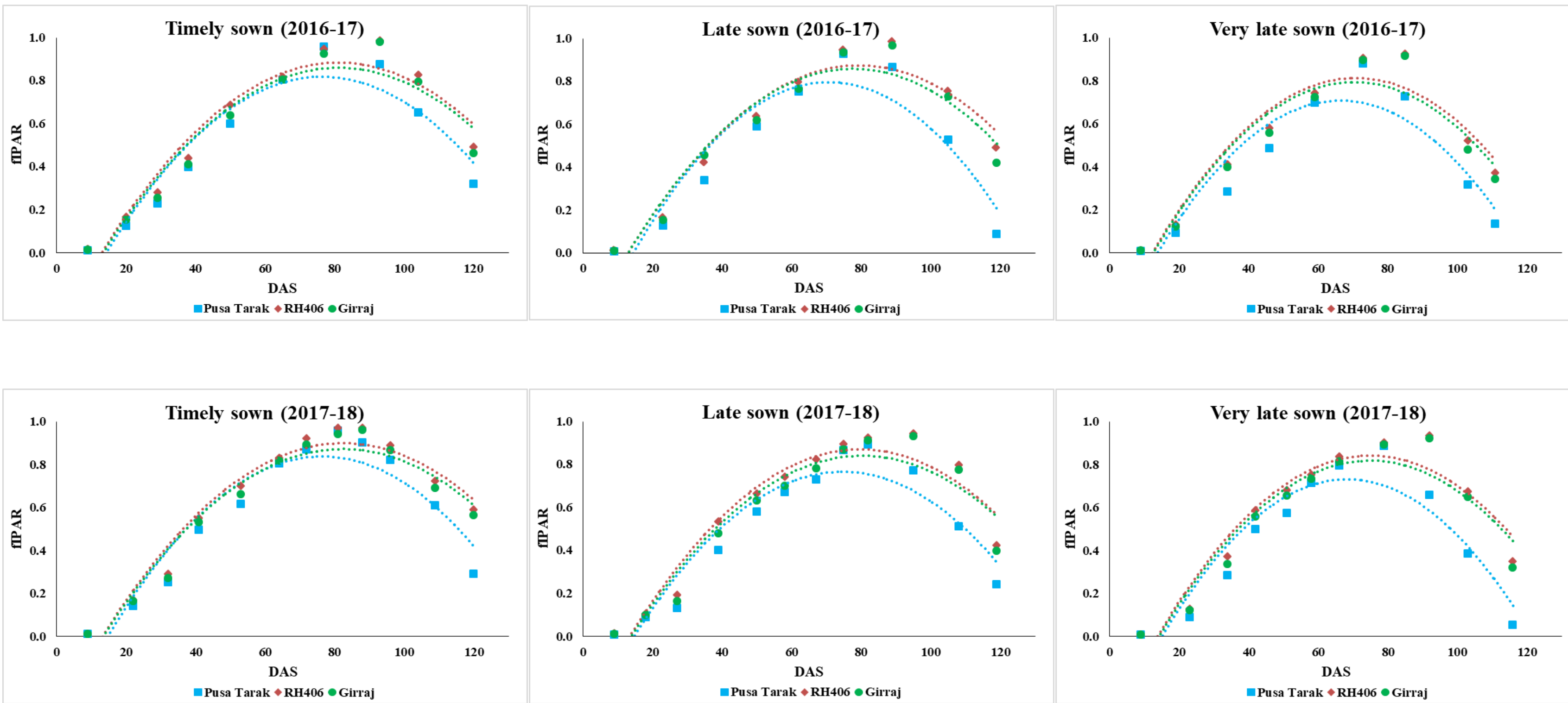

Fig 4.5 fIPAR at different days after sowing for all three cultivars during *Rabi* 2016-17 and 2017-18 under different sowing conditions

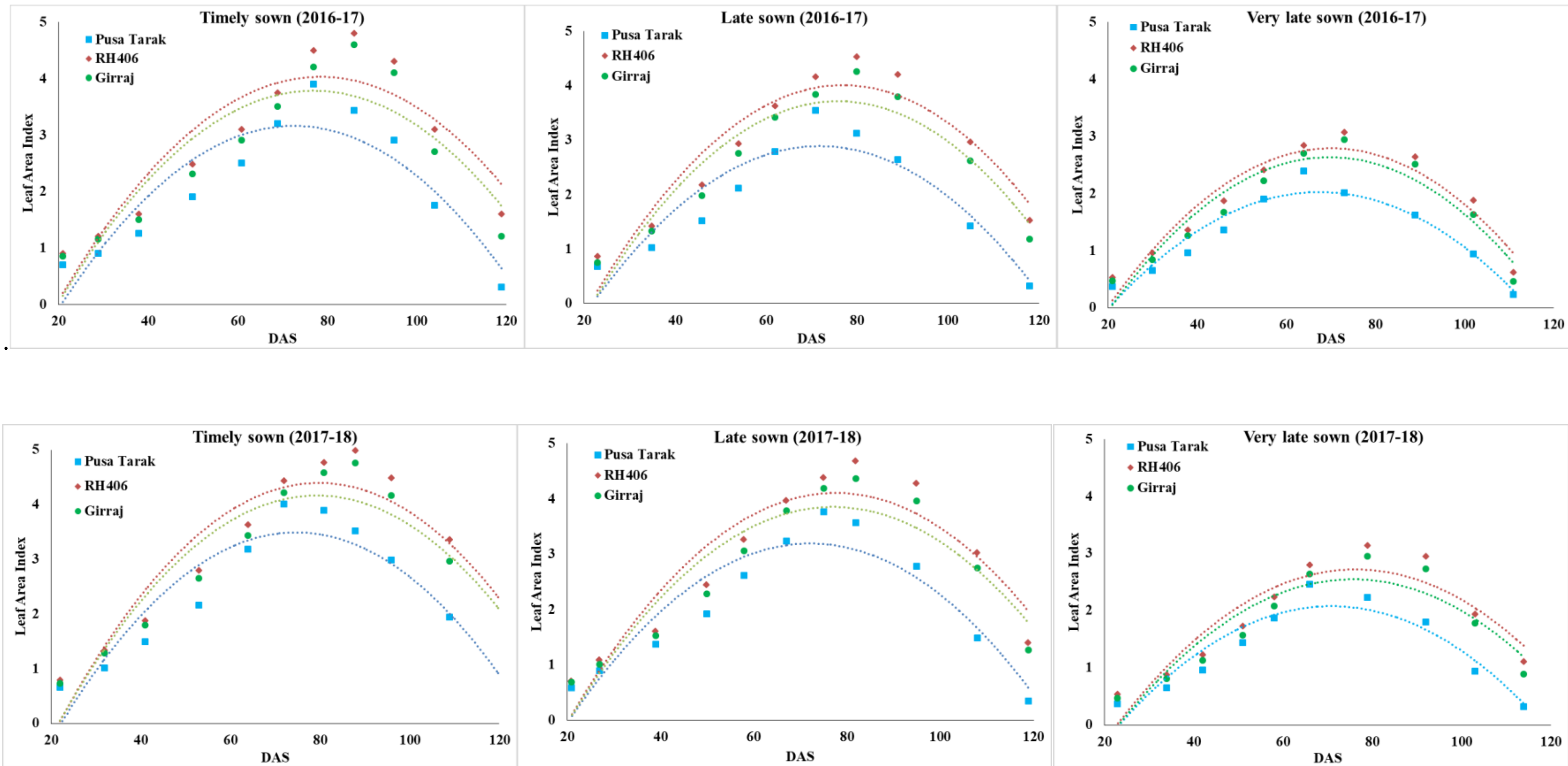

Fig 4.6 Leaf area index (LAI) at different days after sowing for all three cultivars during *Rabi* 2016-17 and 2017-18 under different sowing conditions

4.98 and 4.76 under timely sown crop, which reduced to 2.46, 3.13 and 2.95 under very late sown mustard crop for Pusa Tarak, RH-406 and Girraj, respectively during 2017-18. RH-406 has the highest value of LAI followed by Girraj and Pusa Tarak throughout the *Rabi* season in 2016-17 and 2017-18. The climatic condition of 2017-18 was more favorable to accomplish more LAI than 2016-17 for all dates of sowing. The date and peak of attaining maximum LAI shifts towards a lower value under delay in sowing, which indicates delay in sowing resulted in significant reduction in crop growth and a shortening of crop growing period.

4.1.5 Above-ground biomass

The above-ground biomass obtained on different dates for different treatments and cultivars during *Rabi* 2016-17 are shown in Table 4.2. There was a significant difference observed in above-ground biomass accumulation, for different cultivars, for different sowing dates and their interaction after 30 DAS. The values shown for the effect of sowing dates are averaged over the values of cultivars. Similarly, the values shown under the effect of cultivars are averaged over the effect of sowing dates. The significance of difference in any two pairs of values was least significant difference (LSD) at 95 % confidence level (p=0.05).

The final above-ground biomass was significantly higher for RH-406 (8867 kg ha^{-1}) followed by Girraj (8577 kg ha^{-1}) and Pusa Tarak (7553 kg ha^{-1}). The average above-ground biomass was 9506 kg ha^{-1} for the timely sowing and reduced to 8543 kg ha^{-1}, 6946 kg ha^{-1} for the late and very late sowing, respectively. In the sowing dates and cultivars interaction study, RH-406 in the timely sowing showed significantly highest value of above-ground biomass (10200 kg ha^{-1}), and Pusa Tarak in the very late sowing showed the lowest value (6460 kg ha^{-1}).

A higher accumulation of biomass was observed during 2017-18 than that of 2016-17 (Table 4.3). There was significant difference among cultivars, sowing dates and their interaction for above-ground biomass measured at different days after sowing. The above-ground biomass was found to be highest for RH-406 as also observed during 2016-17. The highest values of above-ground biomass were 10820, 9990 and 8900 kg ha^{-1} for the timely sowing, 9540, 9220 and 7830 kg ha^{-1}for the late sowing, 8200, 7760 and 6600 kg ha^{-1} for the very late sowing in RH-406, Girraj and Pusa Tarak, respectively.

Table 4.2 Above-ground biomass (kg ha^{-1}) for different mustard cultivars sown on different dates during *Rabi* 2016-17.

Treatments	Days after sowing (DAS)						
	30	**45**	**60**	**75**	**90**	**105**	**At Harvest**
Effect of sowing date							
D1	228	586	1333	2943	4905	6934	9506
D2	207	491	1090	2503	4487	6547	8543
D3	144	433	912	2366	4248	6227	6946
$LSD_{0.05}$	NS	13.39*	27.21*	87.20*	115.21*	66.60*	126.09*
Effect of cultivars							
Pusa Tarak	152	386	860	2000	3752	5685	7553
RH 406	217	588	1313	3129	5203	7237	8867
Girraj	209	535	1162	2651	4685	6785	8577
$LSD_{0.05}$	NS	7.74*	32.14*	50.20*	80.53*	52.08*	103.09*
Interaction effect							
Pusa Tarak (D1)	185	454.5	1054	2296	4014	5962	8640
RH 406 (D1)	244	678	1580	3518	5640	7710	10200
Girraj (D1)	256	627	1366	3014	5062	7130	9680
Pusa Tarak (D2)	182	385.45	813	1897	3662	5630	7560
RH 406 (D2)	223	566	1261	2995	5160	7200	9100
Girraj (D2)	216	522.9	1194.8	2618	4640	6810	8970
Pusa Tarak (D3)	91	319	714	1809	3580	5463	6460
RH 406 (D3)	185	521	1098	2876	4809	6801	7300
Girraj (D3)	155	458	924	2413	4354	6416	7080
$LSD_{0.05}$	NS	17.17*	52.71*	87.50*	160.84*	98.61*	191.45*

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

* Significant at 95 % confidence level (p=0.05).

Table 4.3 Above-ground biomass (kg ha^{-1}) for different mustard cultivars sown on different dates during *Rabi* 2017-18.

Treatments	Days after sowing (DAS)						
	30	**45**	**60**	**75**	**90**	**105**	**At Harvest**
			Effect of sowing date				
D1	249	726	1585	3590	5539	7474	9903
D2	241	563	1168	3087	5036	6748	8863
D3	162	470	1071	2748	4838	6968	7520
$LSD_{0.05}$	13.97	55.72*	26.32*	145.65*	76.34*	183.96*	164.97*
			Effect of cultivars				
Pusa Tarak	173	447	975	2509	4269	6072	7776
RH 406	248	710	1548	3705	5855	7819	8200
Girraj	230	602	1302	3211	5288	7299	7760
$LSD_{0.05}$	14.44	41.06*	38.42*	51.15*	119.45*	100.29*	151.15*
			Interaction effect				
Pusa Tarak (D1)	212	575	1210	2893	4520	6462	8900
RH 406 (D1)	276	857	1932	4254	6392	8268	10820
Girraj (D1)	258	743	1614	3622	5703	7692	9990
Pusa Tarak (D2)	206	462	886	2540	4432	5909	7830
RH 406 (D2)	264	675	1417	3579	5605	7414	9540
Girraj (D2)	253	553	1201	3142	5071	6922	9220
Pusa Tarak (D3)	102	304	828	2093	3854	5845	6600
RH 406 (D3)	205	597	1294	3282	5568	7775	8200
Girraj (D3)	180	510	1090	2870	5092	7283	7760
$LSD_{0.05}$	NS	NS	60.15*	161.76*	184.73*	230.63*	202.12*

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

* Significant at 95 % confidence level (p=0.05).

The result showed that the date of sowing significantly affected mustard crop biomass in both years. The timely sown crop had a significantly higher yield than a late sown crop in the semi-arid environment of the New Delhi, indicating that delay in sowing to mid-November adversely affected the crop growth and above-ground biomass. The mustard crop growth observed during the 2017-18 was more healthy and productive than the 2016-17 crop seasons.

4.1.6 Seed yield, oil content and harvest index

Seed yield, oil content and harvest index (HI) of mustard for different sowing dates and cultivars during 2016-17 and 2017-18 are shown in Table 4.4.

Table 4.4 Seed yield, oil content and harvest index of different mustard cultivars sown on different dates during *Rabi* 2016-17 and 2017-18.

Treatments	Seed yield (kg ha^{-1})		Oil content (%)		Harvest index (%)	
	2016-17	2017-18	2016-17	2017-18	2016-17	2017-18
Effect of sowing date						
D1	2083	2156	38.14	41.95	21.88	21.74
D2	1638	1715	36.76	40.46	19.13	19.31
D3	1112	1188	34.48	38.37	15.96	15.81
$LSD_{0.05}$	51.19*	32.15*	2.17*	1.32*	1.54*	0.52*
Effect of cultivars						
Pusa Tarak	1389	1468	35.51	38.44	18.07	18.61
RH 406	1786	1863	36.28	40.28	19.86	19.30
Girraj	1658	1728	37.59	42.05	19.04	18.86
$LSD_{0.05}$	66.65*	52.56*	0.78*	1.18*	0.83*	0.76*
Interaction effect						
Pusa Tarak (D1)	1840	1900	37.43	40.88	21.29	21.34
RH 406 (D1)	2280	2374	38.20	42.96	22.35	21.94
Girraj (D1)	2128	2192	38.80	43.00	21.98	21.94
Pusa Tarak (D2)	1380	1453	35.77	38.13	18.26	18.57
RH 406 (D2)	1820	1904	36.63	41.11	20.00	19.96
Girraj (D2)	1715	1789	37.87	42.16	19.13	19.41
Pusa Tarak (D3)	946	1051	33.33	36.32	14.65	15.92
RH 406 (D3)	1257	1310	34.00	37.78	17.22	15.98
Girraj (D3)	1132	1204	36.10	41.00	16.00	15.52
$LSD_{0.05}$	89.51**	80.72**	0.78*	0.64*	NS	NS

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

* Significant at 95 % confidence level (p=0.05).

The final seed yield production at harvest was significantly reduced in the very late sown crop (1112 kg ha^{-1}) as compared to the timely sown (2083kg ha^{-1}) and late sown crop (1638 kg ha^{-1}) during 2016-17. Similar results was obtain for seed yield during 2017-18, seed yield were observed 2156, 1715, and 1188 kg ha^{-1} for the timely sown, late sown and very late sown crop, respectively. Significantly higher seed yield 1786 and 1863 kg ha^{-1} was observed for RH-406 followed by 1658 and 1728 kg ha^{-1} for Girraj and 1389 and 1468 kg ha^{-1} for Pusa Tarak during 2016-17 and 2017-18 crop seasons, respectively. Among all the cultivars, RH-406 recorded the significantly highest seed yield of 2280, 1820 and 1257 kg ha^{-1} followed by Girraj 2128, 1715 and 1132 kg ha^{-1} and Pusa Tarak 1840, 1380 and 946 kg ha^{-1} for timely, late and very late sown crop during 2016-17. In the interaction effect during 2017-18, RH-406 maintained the highest production among all cultivars for different treatments.

The seed oil content is an inherent characteristic of a cultivar but showed variation under different environmental conditions, which can modify by different sowing dates. Oil content of an oilseed crop requires a certain ambient temperature during the oil accumulation period. The significant result in mustard oil content was 37.43, 38.20, 38.80 percent for timely sown crop, 35.8, 36.6, 37.9 percent for late sown crop and 33.3, 34.0, 36.1 percent for very late sown crop in Pusa Tarak, RH-406 and Girraj, respectively during 2016-17 (Table 4.4). The percentage oil content during 2017–2018 was 42.9, 41.1 and 37.8 for RH-406, 43.0, 42.2 and 41.0 for Girraj and 40.9, 38.1 and 36.3 for Pusa Tarak in timely, late and very late sown crop, respectively. About 4 percent significant reduction was observed in mustard oil due to delay in sowing for both the years. The oil content for different cultivars Pusa Tarak, RH-406 and Girraj were 35.5, 36.3, 37.6 percent and 38.4, 40.3, 42.0 percent during 2016-17 and 2017-18 crop season. Girraj had the highest oil content among all cultivars for all treatments.

The harvest index significantly varied from 21.9, 19.1 and 15.9 during 2016-17 and 21.7, 19.3 and 15.8 during 2017-18 crop growing year for timely, late and very late sown crop, respectively. RH-406 had significantly highest HI 19.9 percent followed by Girraj 19.0 percent and Pusa Tarak 18.1 during 2016-17. RH-406 had observed similar result during 2017-18. The interaction effect of cultivars with different dates of sowing was non-significant during 2016-17 and 2017-18. There was a significant reduction in harvest index with delay in sowing and it was mostly due to more sharper reduction in seed yield compared with reduction in biomass.

4.1.7 Radiation use efficiency (RUE)

It is envisaged that the radiation use efficiency is a specific varietal character of any crop but it can be altered under different microclimatic conditions during crop growing season. Radiation use efficiency (RUE) based on seed yield and above-ground biomass of different mustard cultivars sown on different dates during 2016-17 and 2017-18 are shown in Table 4.5.

Table 4.5 TIPAR and RUE of different mustard cultivars under different dates of sowing during *Rabi* 2016-17 and 2017-18.

Treatments	TIPAR ($MJ\ m^{-2}$)		RUEy (g/MJ)		RUE_b (g/MJ)	
	2016-17	2017-18	2016-17	2017-18	2016-17	2017-18
Effect of Sowing date						
D1	363.9	372.6	0.58	0.58	2.63	2.67
D2	341.6	353.9	0.48	0.49	2.51	2.52
D3	313.1	327.9	0.36	0.36	2.25	2.31
$LSD_{0.05}$	5.03*	5.39*	0.03*	0.01*	0.02*	0.02*
Effect of cultivars						
Pusa Tarak	278.7	290.4	0.49	0.5	2.7	2.67
RH 406	377.1	389.9	0.47	0.47	2.34	2.43
Girraj	362.9	374.1	0.45	0.46	2.35	2.4
$LSD_{0.05}$	5.00*	4.33*	0.02*	0.02*	0.04*	0.05*
Interaction effect						
Pusa Tarak (D1)	299.8	313.8	0.61	0.61	2.88	2.84
RH 406 (D1)	403.2	410.2	0.57	0.58	2.53	2.64
Girraj (D1)	388.8	393.9	0.55	0.56	2.49	2.54
Pusa Tarak (D2)	285.9	288.2	0.48	0.50	2.64	2.72
RH 406 (D2)	377.9	396.1	0.48	0.48	2.41	2.41
Girraj (D2)	361.1	377.6	0.47	0.47	2.48	2.44
Pusa Tarak (D3)	250.3	269.4	0.38	0.39	2.58	2.45
RH 406 (D3)	350.3	363.3	0.36	0.36	2.08	2.26
Girraj (D3)	338.8	351.1	0.33	0.34	2.09	2.21
$LSD_{0.05}$	8.54*	8.01*	NS	NS	0.06*	0.05*

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

* Significant at 95 % confidence level (p=0.05).

RH-406 showed significantly higher values 377.1 and 389.9 MJ m^{-2} of total intercepted photosynthetically active radiation (TIPAR over the total growth period), followed by Girraj 362.9 and 374.1 MJ m^{-2} and Pusa Tarak 278.7 and 290.4 MJ m^{-2} during 2016-17 and 2017-18, respectively (Table 4.5). The value of TIPAR was higher 372.6 MJ m^{-2} in timely sown crop compared to late sown crop 353.9 MJ m^{-2} and very late sown crop 327.9 MJ m^{-2} during 2017-18 as compared to corresponding value during 2016-17. The interaction effect between cultivars and date of sowing was also significant for TIPAR during both the years.

The RUE based on seed yield (RUE_y) were 0.58, 0.48 and 0.36 g MJ^{-1} and RUE based on biomass (RUE_b) were 2.63, 2.51 and 2.25 g MJ^{-1} for timely, late and very late sown crop, respectively during 2016-17. The highest values of RUE was 0.5 and 2.7 g MJ^{-1} for Pusa Tarak, followed by 0.47 and 2.34 g MJ^{-1} for RH-406 and 0.45 and 2.35 g MJ^{-1} for Girraj on seed yield and biomass basis, respectively. The values of RUE_y and RUE_b remain the same during 2017-18 for all the cultivars. The RUE_y was non-significant for cultivars and date of sowing interaction effect Whereas, RUE_b was significant for cultivars and date of sowing interaction effect during both the years. Decrease in RUE (based on seed yield and above-ground biomass) was mostly due to more reduction in seed yield and above-ground biomass as compared to the corresponding reduction in TIPAR. The result showed that Pusa Tarak had highest RUE on seed yield and biomass basis among all cultivars. Hence, Pusa Tarak may be a good option to cultivate in very late sowing conditions.

4.1.8 Soil moisture content

Soil moisture content during mustard crop for different sowing dates was measured upto a soil depth of 90 cm at 15 days intervals by thermo-gravimetric method for 0-15 cm depth and neutron moisture meter for 15-90 cm depth during both the years and presented in Fig. 4.7. The straight horizontal lines represent mean value of field capacity (FC) and permanent wilting point (PWP) for 90 cm soil depth. The soil moisture content remained well within the FC (220mm) and PWP (70mm) for all the treatments. The peak values in the soil moisture was obtained at 41 and 80 days after sowing (DAS) for the timely sown crop, 26, 65 and 112 DAS for the late sown crop and 50 and 90 DAS for the very late sown crop, correspond to either irrigation or rainfall events during 2016-

17. During 2017-18, the peak value of soil moisture content was found at 30, 72 and 113 DAS for the timely sown crop, 58 and 99 DAS for the late sown crop and 43 and 84 DAS for the very late sown crop. Pusa Tarak had highest soil moisture content compared to the RH406 and Girraj in all dates of sowing for the both years. It may be attributed to higher soil moisture content related to less evapo-transpiration and less soil moisture extraction in producing biomass and seed yield.

4.1.9 Water productivity (WP)

The crop evapotranspiration and water productivity (WP) based on seed yield and above-ground biomass of different mustard cultivars for different sowing dates during 2016-17 and 2017-18 are presented in Table 4.6. The seasonal evapotranspiration (ET) was 219, 242 and 240 mm in 2016-17 and 222, 260 and 249 during 2017-18 for Pusa Tarak, RH-406 and Girraj, respectively. Timely sown crop had 11.3 and 4.06 percent higher ET than very late and late sown crop, respectively during 2016-17. During 2017-18, ET reduction was 5.11 and 9.84 percent compared to the late and very late sown crop. Interaction between cultivars and sowing dates during both the year had non-significant value.

The water productivity was highest 0.85 and 3.86 g/m^2/mm in timely sown crop followed by 0.69 and 3.61 g/m^2/mm in late sown crop and 0.51 and 3.18 g/m^2/mm in very late sown crop on the basis of seed yield (WP_y) and biomass (WP_b), respectively during 2016-17 crop season. The values of WP_y and WP_b remain same value during 2017-18 crop seasons as that during 2016-17 for all the cultivar. Significantly higher water productivity was observed in RH-406 compared to the Girraj and Pusa Tarak on seed yield and biomass basis for both the years. The interaction effect between crop cultivars and sowing dates was non-significant for both the years.

The reduction in water productivity is mainly due to more quick reduction in seed yield and biomass without less sharp reduction in seasonal evapotranspiration, which clarifies the positive correlation of water productivity with seed yield and above-ground biomass.

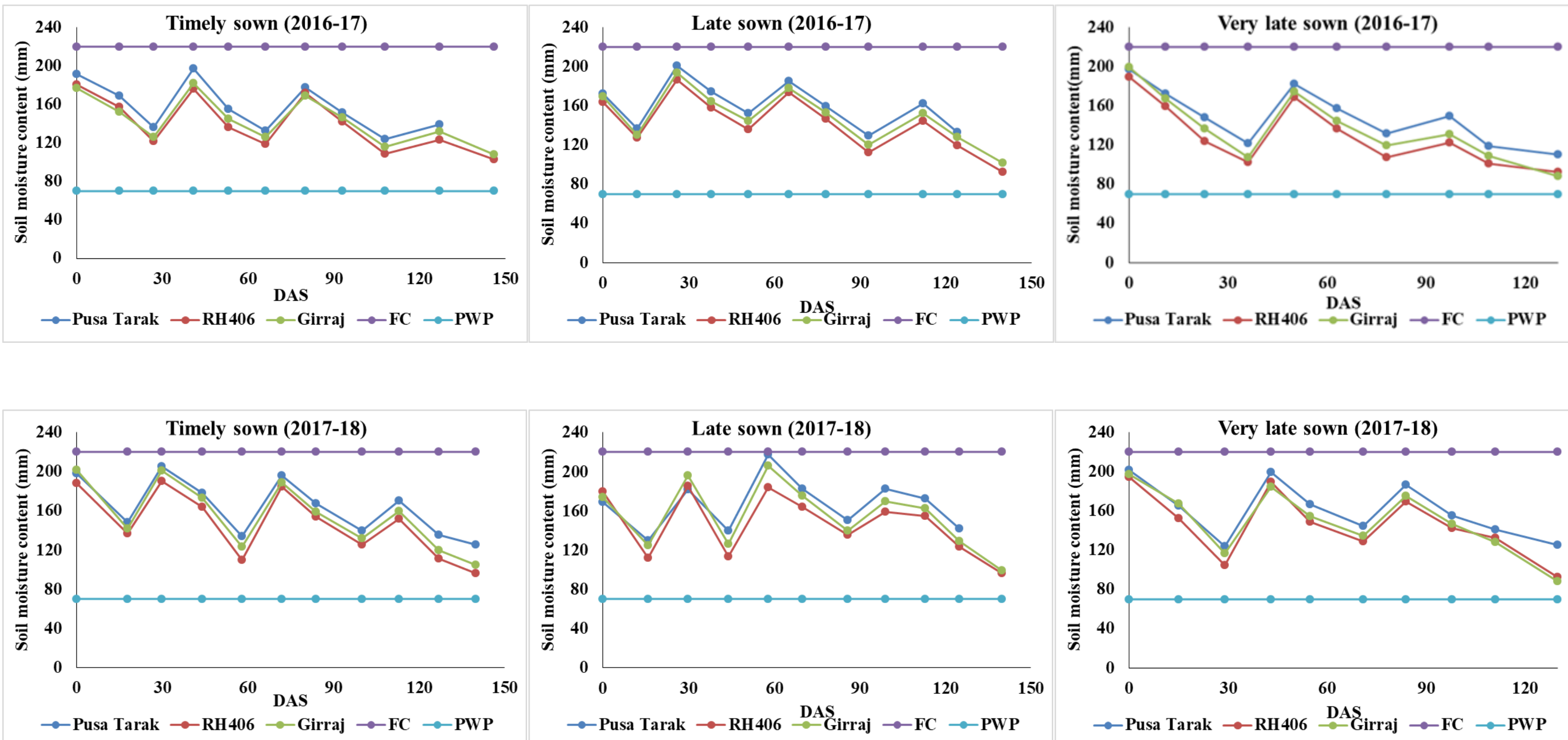

Fig 4.7 Soil moisture content at different days after sowing for all three cultivars during *Rabi* 2016-17 and 2017-18 under different sowing conditions

Table 4.6 Evapotranspiration and water productivity of different mustard cultivars sown on different dates during *Rabi* 2016-17 and 2017-18

Treatments	**ET (mm)**		**WP_y ($g/m^2/mm$)**		**WP_b (g/m2/mm)**	
	2016-17	**2017-18**	**2016-17**	**2017-18**	**2016-17**	**2017-18**
			Effect of Sowing date			
D1	246	254	0.85	0.85	3.86	3.89
D2	236	241	0.69	0.71	3.61	3.67
D3	218	229	0.51	0.52	3.18	3.28
$LSD_{0.05}$	6.55*	5.72*	0.05*	0.03*	0.15*	0.11*
			Effect of cultivars			
Pusa Tarak	219	222	0.63	0.66	3.43	3.49
RH 406	242	260	0.73	0.73	3.64	3.76
Girraj	240	249	0.69	0.69	3.57	3.58
$LSD_{0.05}$	5.45*	4.10*	0.03*	0.02*	0.09*	0.08*
			Interaction effect			
Pusa Tarak (D1)	232	236	0.79	0.81	3.73	3.77
RH 406 (D1)	257	270	0.89	0.88	3.96	4.01
Girraj (D1)	249	257	0.86	0.85	3.89	3.89
Pusa Tarak (D2)	220	220	0.63	0.66	3.45	3.55
RH 406 (D2)	243	250	0.75	0.76	3.75	3.82
Girraj (D2)	248	253	0.69	0.71	3.62	3.64
Pusa Tarak (D3)	207	209	0.46	0.50	3.13	3.15
RH 406 (D3)	227	237	0.55	0.55	3.22	3.45
Girraj (D3)	222	241	0.51	0.50	3.19	3.22
$LSD_{0.05}$	NS	NS	NS	NS	NS	NS

D1= Timely sown crop, D2 = Late sown crop, D3= Very late sown crop

* Significant at 95 % confidence level (p=0.05).

4.2 Crop yield prediction by InfoCrop-mustard model

4.2.1 Calibration of InfoCrop-mustard model at experimental field

The model was initialized each time prior to mustard sowing during *Rabi* 2016-17 and 2017-18. The model calibration involved initially determining the values of phenology coefficients and then the coefficients describing growth and seed development. The model should be calibrated for crop growing under ideal or non-stress conditions. The measured genetic coefficients of each mustard cultivar used for calibrating the InfoCrop-mustard model are listed in Table 4.7. The details of soil parameters are given in materials and method section in Table 3.4. InfoCrop- mustard model was calibrated for the emergence period, 50 % flowering period, physiological maturity period, LAI peaks and profiles, above-ground biomass and seed yield. For calibrating InfoCrop-mustard model, the parameters were adjusted for timely sown mustard during *Rabi* 2016-17. Each mustard cultivar's genetic coefficients were obtained using the best fit method, i.e. by iteratively changing coefficient values to found the relationship between simulated and measured values (within 10 percent range). We ran the mustard exe file on cmd prompt for calibration and validation of InfoCrop- mustard model.

Table 4.7 Varietal coefficients specified in InfoCrop-mustard model for different mustard cultivars

Genetic constants	Acronym	Pusa Tarak	RH406	Girraj
Thermal time – germination period (°C days)	TTGERM	80	115	105
Thermal time- reproductive period (°C days)	TTVG	665	820	800
Thermal time – grain filling period (°C days)	TTGF	730	1000	960
Specific leaf area of cultivar (ha leaf kg $^{-1}$ leaf)	SLAVAR	0.00404	0.00423	0.00420
Potential rate of growth (mm day^{-1})	RGRPOT	0.0216	0.022	0.022
Light extinction coefficient	KDFMAX	0.59	0.65	0.68
Potential weight of a grain (mg)	POTGWT	8.42	8.9	8.78
Radiation use efficiency (g MJ^{-1})	RUE	2.88	2.53	2.49

4.2.2 Validation of InfoCrop-mustard model at experimental field

After the calibration of the model for the timely sowing date, the InfoCrop-mustard model was simulated for staggered sown field experiments during *Rabi* 2016-17 and 2017-18. The model performance was evaluated by comparing model simulations with independent experimental datasets which were not used in model calibration. The phenological developments, profile and peak value of LAI, above-ground biomass and seed yield were used in this study for model validation as described in the following sub-sections.

4.2.2.1 Phenological development

In the InfoCrop model, phenology of the crop was simulated for three different mustard cultivars (Pusa Tarak, RH-406 and Girraj), which are based on an accumulation of degree days instead of calendar days during three phases viz., sowing to seedling emergence, seedling emergence to flowering and flowering to physiological maturity. The accumulated degree days are modified by the maximum temperature, minimum temperature and photo-period during crop growing period. The InfoCrop-mustard model was validated for three developmental stages, i.e. germination, 50% flowering and physiological maturity for all cultivars with different sowing dates. The simulation of phenological development is most important for model.

Simulation of days required from germination along with 1:1 line as scatter plot for Pusa Tarak, RH-406 and Girraj during both the years are depicted in Fig. 4.8. There was hardly one day difference between observed and simulated values. InfoCrop-mustard model overestimated the days for germination to 50% flowering. The RMSE value for germination days was <1, for all cultivars from the InfoCrop-mustard model developed.

The results showed that observed and simulated duration for 50% flowering occurred between 50 to 55 days for Pusa Tarak, 64 to 70 days for RH-406 and 63 to 71 days for Girraj under timely, late and very late sowing Fig. 4.9**.** Simulated duration for 50% flowering by InfoCrop-mustard model was underestimated in timely sown and overestimated in late and very late sown crop for all the cultivars.

Days simulated for physiological maturity was underestimated for Pusa Tarak and overestimated for RH-406 (Fig. 4.10). RMSE values for simulation days for physiological maturity were 0.89 for Pusa Tarak, 1.67 for RH-406 and 1.34 for Girraj. Better precision in phenology simulation may be attributed to model accounting for the effect of date of sowing on thermal time accumulation.

4.2.2.2 Leaf area index

In InfoCrop model, during initial stage of development (when LAI is less than 0.75), leaf growth rate is mainly influenced by temperature and moderated by nitrogen stress and not by water stress. After that, the growth rate in LAI (RLAI) is calculated based on initial LAI (LAII), leaf area growth rate (GLAI), death rate of LAI (DLAI) and a net loss of LAI due to pests (LALOSS) (Aggarwal *et al.* 2004).

The InfoCrop-mustard simulation model was reasonably good for simulating LAI of all the cultivars during both the years (Fig. 4.11). The simulated peak LAI value was higher than the observed value in different cultivars during both the years. The more deviation was observed in LAI value during late sown crop along with the 1:1 scattered line, and this may be due to the temperature stress condition at the later stage of development. LAI was overestimated compared to the observed value by InfoCrop-mustard simulation model because premature leaf senescence due to the reduction in the crop duration was not fully diverted to leaf area by the model. RMSE and nRMSE values for simulation of LAI by InfoCrop-mustard model were 0.59, 0.64 and 0.59 and 18.4, 14.8 and 18.5 for Pusa Tarak, RH-406 and Girraj during 2016-17 and 2017-18 crop seasons, respectively.

4.2.2.3 Above-ground biomass and seed yield

Above-ground biomass was simulated by the InfoCrop-mustard model for Pusa Tarak, RH-406 and Girraj during crop growth period of *Rabi* 2016-17 and 2017-18. A good agreement (based on nRMSE value) was found between simulated and observed values of accumulation in above-ground biomass (Fig. 4.12) and seed yield (Fig. 4.13). InfoCrop utilizes the radiation use efficiency (RUE) based approach for dry matter production. Maximum RUE (RUEMAX) is input in the model as a function of crop/cultivar. The RUEMAX of a plant is affected by abiotic (temperature, CO_2, nitrogen and water stress) and biotic (pest and disease) factors.

RMSE values during simulation of biomass for both years were 1186, 920 and 1265 kg ha^{-1} and nRMSE values were 15.87, 10.74 and 14.7 for Pusa Tarak, RH-406 and Girraj, respectively. For simulation of the seed yield during both the years, RMSE values were 189 kg ha^{-1}, 201 kg ha^{-1} and 200 kg ha^{-1} and nRMSE values were 14.04, 11.64 and 12.47 for Pusa Tarak, RH-406 and Girraj, respectively. Values of nRMSE were less than 15 for model simulation of above-ground biomass and seed yield for different treatments. The deviation from the observed biomass and the seed yield was highest in delayed

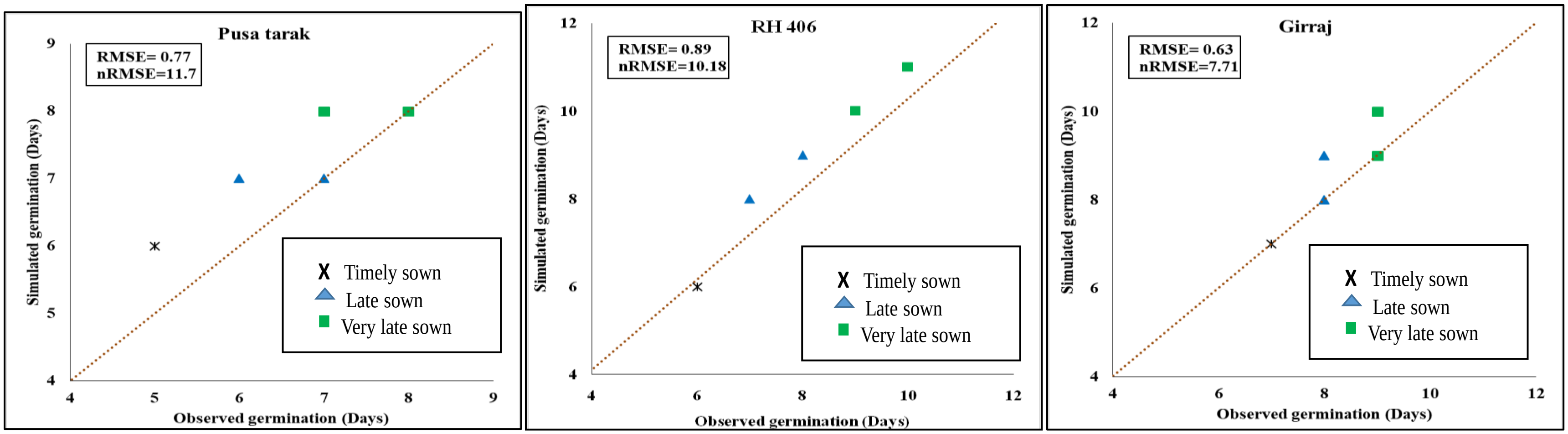

Fig 4.8 Simulation of days for germination for all three mustard cultivars during both experimental years

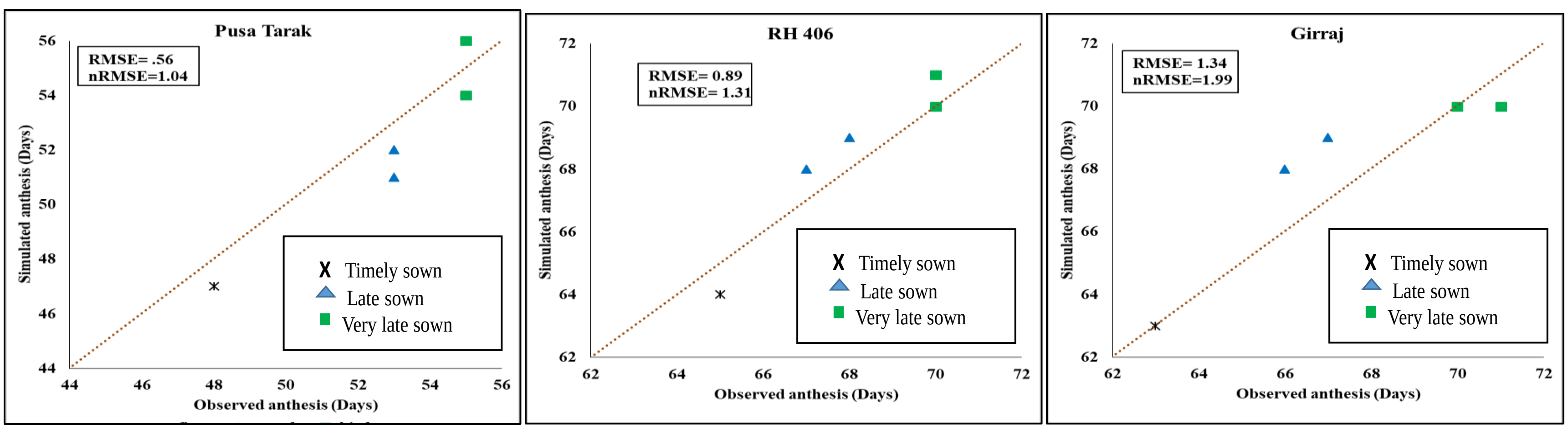

Fig 4.9 Simulation of days for 50 percent flowering for all three mustard cultivars during both experimental years

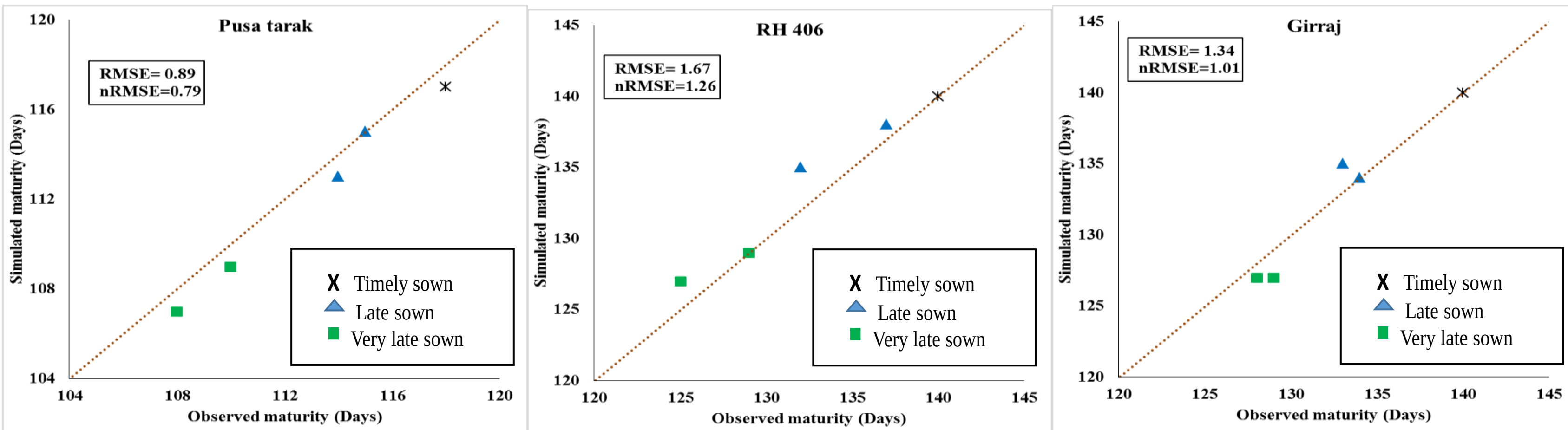

Fig 4.10 Simulation of days for maturity for all three mustard cultivars during both experimental years

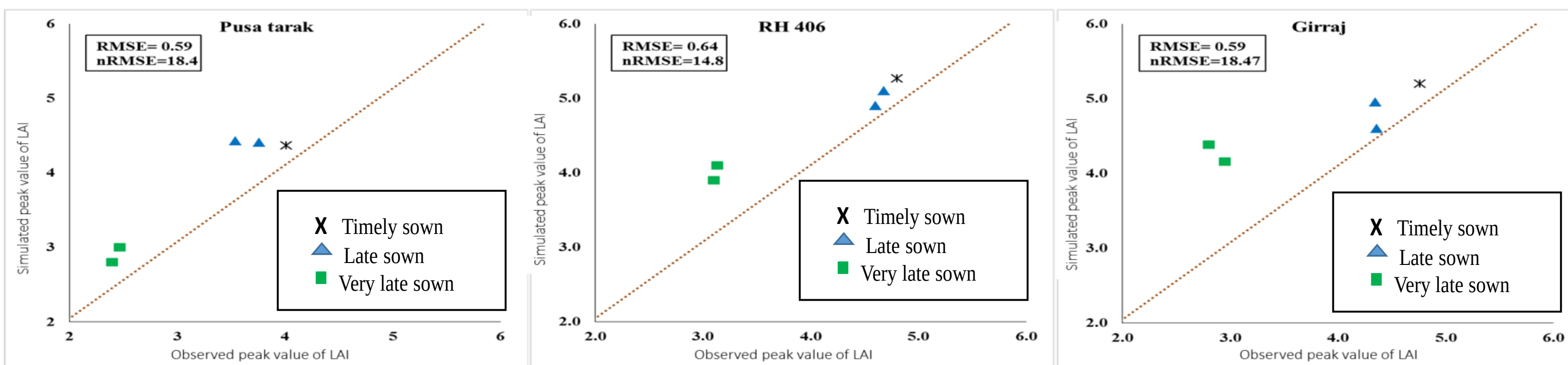

Fig 4.11 Simulation of peak value for LAI of all three mustard cultivars during both experimental years

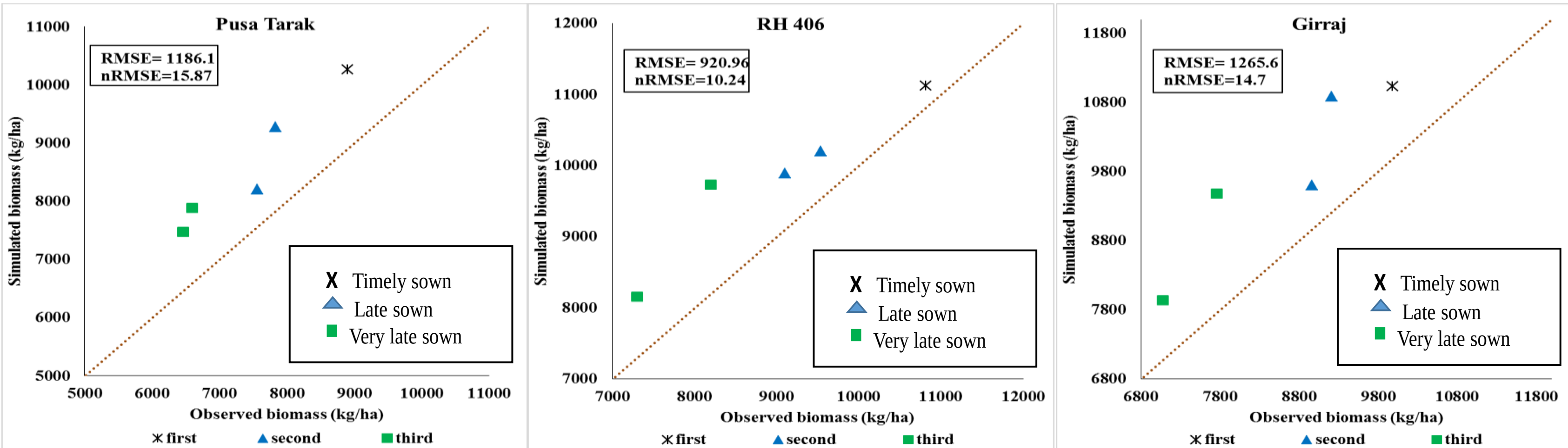

Fig 4.12 Simulation of above ground biomass for all three mustard cultivars during both experimental years

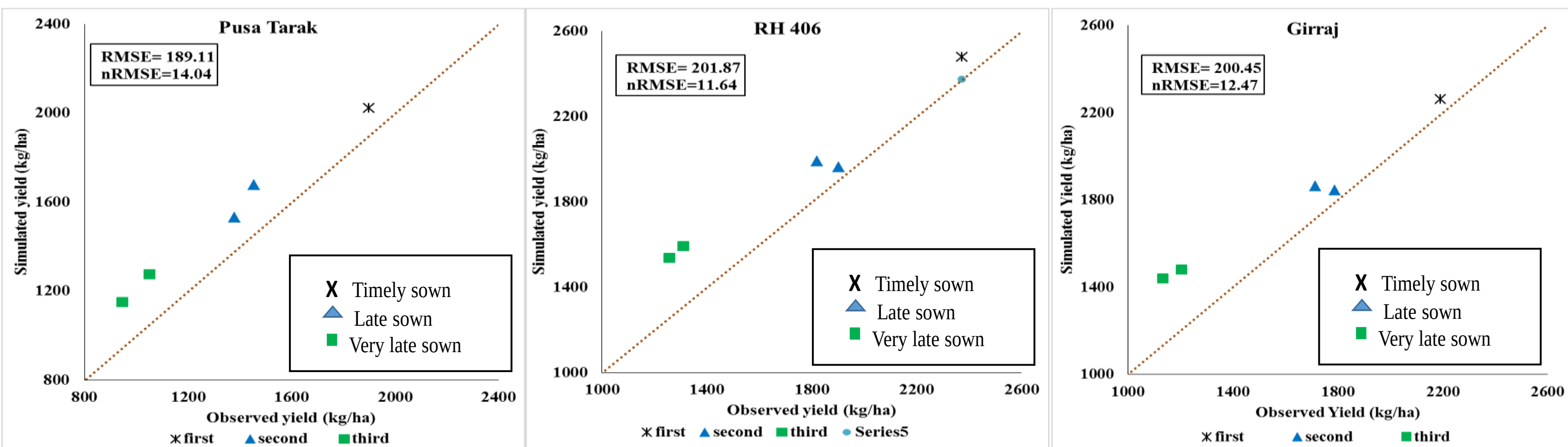

Fig 4.13 Simulation of seed yield for all three mustard cultivars during both experimental years

sowing during both the years. It implies that model accuracy was good for simulating the above-ground biomass and seed yield of mustard for timely sown crop.

4.2.3 Weather conditions at farmer's field during *Rabi* 2017-18

The daily weather data was collected from KVK, Kumher, Bharatpur, Agro-met observatory during *Rabi* 2017-18 crop seasons. The weekly average values were computed from daily observed values of weather variables such as Weekly mean maximum and minimum temperature, total weekly rainfall, mean relative humidity and bright sunshine hours (Fig. 4.14). The maximum temperature varied from 12.5 to 40°C, whereas minimum temperatures varied from 1.9 to 21.5°C. There is an inverse relationship between the temperature and relative humidity. Maximum relative humidity ranged between 54 to 97 % and minimum relative humidity ranged between 15 to 88 %. The rainfall received during the entire crop growing period was 3.8 mm. The weekly mean bright sunshine hours was ranged between 0.2 hours and 9.2 hours. The average wind speed varied from 0.67 to 8.67 km/hour during the crop grown period.

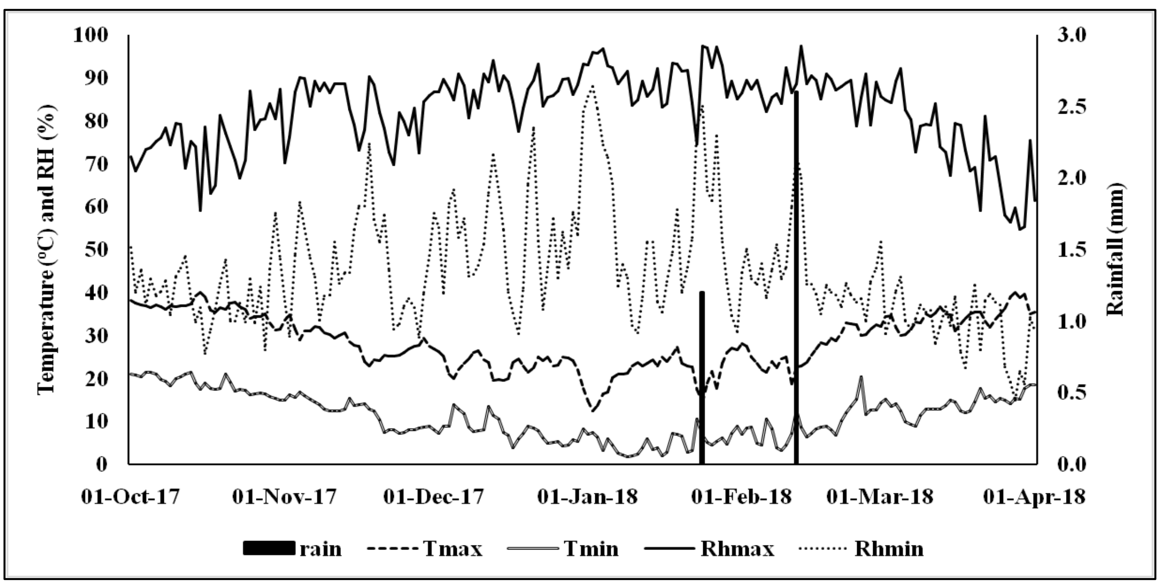

Fig .4.14 Daily weather conditions at KVK, Kumher during *Rabi* 2017-18.

4.2.4 Validation of InfoCrop-mustard model at farmers' field

The major mustard-growing state in India is Rajasthan. Bharatpur district plays a dominant role in the production and productivity of mustard. To validate InfoCrop-mustard model at farmer's field 20 farmers were selected from Mukundpura and Sitara village of Bharatpur district. The location (latitude and longitude) of farmer's field selected for the observation has been shown in material and method (plate 3). There were variations in cultivars, sowing dates and management practices at the farmer's field table 4.8.

Table 4.8 Location of farmer's field along with cultivars, sowing dates, mean of peak LAI, biomass and seed yield of mustard

Farmer's name	Cultivar	Sowing Date	Mean of peak LAI	Biomass (kg ha^{-1})	Seed yield (kg ha^{-1})
Mohan lal	RH-406	10-Oct	4.3	8500	2000
Nawab singh	RH-406	13-Oct	4.1	8200	1930
Hukum Singh	RH-406	15-Oct	4.1	8100	2160
Dharmpal	RH-406	16-Oct	3.9	7100	1800
Atibhan	RH-406	18-Oct	3.5	6800	1680
Jagnath	RH-406	12-Oct	4.4	9000	2200
Hari Om	RH-406	19-Oct	3.8	7900	1750
Bhadur Singh	RH-406	26-Oct	3.2	7200	1650
Panna Lal	RH-406	12-Oct	4.1	8200	2350
Jagtap	RH-406	16-Oct	4.3	8000	2000
Suraj Singh	Girraj	10-Oct	4.5	8180	1980
Devi Singh	Girraj	15-Oct	4.2	8800	2240
Vijay singh	Girraj	15-Oct	4.3	6800	1800
Bhoodev	Girraj	25-Oct	3.8	7600	1960
KVK, Kumher	Girraj	12-Oct	4.4	6700	1620
Hari Singh	Girraj	11-Oct	4.1	8250	1800
Man singh	Girraj	18-Oct	3.7	6900	1640
Uddam Singh	Girraj	25-Oct	3.1	7800	2200
RamNarayan	Girraj	13-Oct	4.5	7200	1990
Suresh	Girraj	17-Oct	4.2	8180	1970

The location (latitude and longitude) of farmer's field selected for the observation has been shown in materials and method section plate 3. The date of sowing, cultivar sown, mean of peak value of LAI over farmer's field, seed yield and biomass for different farmers are shown in table 4.8. Crop yield is a complex characteristic influenced by many factors such as weather, soil, management practices, etc. In order to study, the effect of management practices on wheat yield at farmers' field, four management factors, viz. sowing date, N application rate, P_2O_5 application rate and number of irrigation applied to the crop were taken. The dominating cultivars in the study area were RH-406 and Girraj. There was about 20 days of variation in the date of sowing from 10th to 26th of Oct. But apart from that, there was less variation in fertilizer application and irrigation scheduling. At the farmer's field, observed peak values of LAI were varied from 3.3 to 4.4, biomass was ranged between 6800 to 9000 kg ha^{-1} and seed yield was ranged from 1650 to 2350 kg ha^{-1} for cultivar RH-406. Cultivar Girraj showed LAI variation between 3.1 to 4.5, above-ground biomass variation between 6800 to 8800 kg ha^{-1} and seed yield variation between 1620 to 2240 kg ha^{-1} at farmers' field.

4.2.4.1 Leaf area index (LAI)

Performance of simulation of InfoCrop-mustard model with observed values for peak values of LAI for RH-406 and Girraj are shown in Fig. 4.15 (a, d). At farmer's field, InfoCrop-mustard model overestimates the peak value of LAI. At farmer's field LAI simulation by InfoCrop-mustard simulated model had RMSE values for 0.88 and 0.83 kg ha^{-1} and nRMSE values 22.1 and 21.0 for RH-406 and Girraj, respectively. Since the nRMSE value is more than 20% and less than 25 %, hence a fair agreement was found between the simulated and the observed peak value of LAI for Mukundpura and Sitara village during 2017-18. Model predictions for peak LAI simulation at farmer's field level were fair with nRMSE < 25 %.

4.2.4.2 Above-ground biomass

The simulation of above-ground biomass was validated for farmer's field. The observed value showed a good agreement well along the 1:1 scattered line in Fig. 4.15 (b, e). The observed above-ground biomass at harvest is greatly influenced by management practices and varied from about 6800 to 9000 kg ha^{-1} in farmers' fields during 2017-18. The simulated above-ground biomass ranged between 7200 to 10350 kg ha^{-1}. The values of RMSE for simulation of above biomass by InfoCrop-mustard model were 1422.9 and

1478.7 kg ha^{-1} and Values of nRMSE were 18.0 and 19.3 for RH-406 and Girraj, respectively at Mukundpura and Sitara village of Bharatpur district in farmer’s field. InfoCrop-mustard model had overestimations in the simulation of above-ground biomass. The model predictions were good with nRMSE < 20 % for simulation of above-ground biomass at farmer’s field level.

4.2.4.3 Seed yield

Simulation of grain yield for RH-406 and Girraj on farmer’s field by InfoCrop-mustard model are shown in Fig. 4.15 (c, f). In InfoCrop-mustard model, source-sink balance is considered in determining seed yield. Mustard seed yield is influenced by the date of sowing and weather variables during crop growing seasons. The observed seed yield was varied between 1650 to 2350 kg ha^{-1} for RH-406 and 1620 to 2240 kg ha^{-1} for Girraj at Mukundpura and Sitara village of Bharatpur district at farmer’s field. The RMSE values for seed yield during validation were 332.9 kg ha^{-1} for RH-406 and 350 kg ha^{-1} for Girraj for 2017-18 crop. nRMSE values of model simulation for seed yield were 17.1 for RH-406 and 18.1 for Girraj, respectively. Results showed that simulated LAI, above-ground biomass and seed yield were overestimated by InfoCrop-mustard model at farmer’s field. Because model calibration was done by actual dataset generated at ICAR-IARI, New Delhi research farm during Rabi 2016-17 for the same cultivars. The InfoCrop-mustard model performs better for simulating seed yield than above-ground biomass and LAI.

4.2.5 Multistage mustard crop prediction at experimental field:

The InfoCrop-mustard model predict the crop seed yield and biomass at 50 percent flowering stage and pod formation stage by using weather parameters till 50 percent flowering and pod formation stage for experimental field of IARI, New Delhi, during *Rabi* 2017-18 crop season.

4.2.5.1. Above-ground biomass:

Mustard above-ground biomass prediction through InfoCrop-mustard model at two different stages 50 percent flowering stage and pod formation stage are shown in table 4.9.

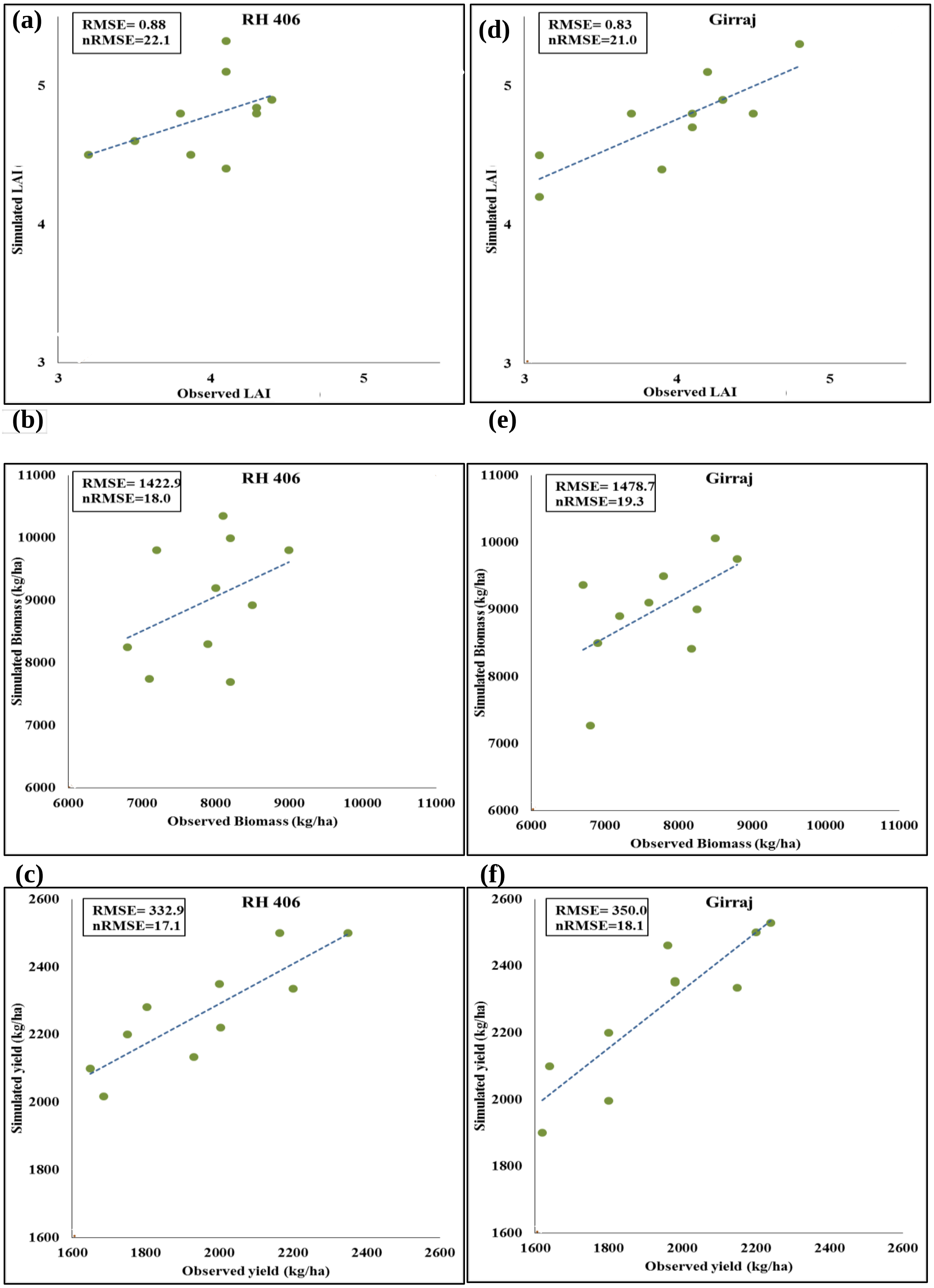

Fig 4.15 Simulation of peak value of LAI (a, d), above ground biomass (b, e) and seed yield (c, f) for RH-406 and Girraj at farmer's field

Table 4.9 Multistage mustard above-ground biomass prediction during *Rabi* 2017-18

Cultivars	Sowing Time	Days after sowing	Biomass (kg ha^{-1}) Observed	Biomass (kg ha^{-1}) Predicted	Percentage deviation	RMSE (kg ha^{-1})	nRMSE (%)
Prediction at 50 percent flowering stage							
Pusa Tarak	Timely	50	8900	10551	18.6	2144.9	27.6
	Late	53	7830	9839	25.6		
	Very late	55	6600	9253	40.2		
RH-406	Timely	65	10820	12587	16.3	2392.4	25.1
	Late	68	9540	11964	25.4		
	Very late	70	8200	11058	34.9		
Girraj	Timely	65	9990	11657	16.7	2077.3	23.1
	Late	67	9220	11392	23.6		
	Very late	71	7760	10094	30.1		
Prediction at pod formation stage							
Pusa Tarak	Timely	78	8900	9706	9.1	1443.2	18.6
	Late	77	7830	9392	19.9		
	Very late	75	6600	8376	26.9		
RH-406	Timely	99	10820	11969	10.6	1668.1	17.5
	Late	98	9540	11083	16.2		
	Very late	95	8200	10355	26.3		
Girraj	Timely	99	9990	10975	9.8	1412.0	15.7
	Late	98	9220	10598	15.0		
	Very late	95	7760	9524	22.8		

The percent deviation of simulated from observed above-ground biomass done at 50 percent flowering stage was lowest for timely sown crop (18.6, 16.3 and 16.7%) followed by late (25.6, 25.4 and 23.6%) and very late sown crop (40.2, 34.9 and 30.1%) for Pusa Tarak, RH-406 and Girraj, respectively. Percent deviation of above-ground biomass prediction done at the pod formation stage was lower than the percent deviation of above-ground biomass prediction done at the flowering stage. Percent deviation were 9.1, 19.9 and 26.9 for Pusa Tarak., 10.6, 16.2 and 26.3 for RH-406 and 9.8, 15.0 and 22.8 for Girraj in timely sown, late sown and very late sown crop, respectively.

Multistage prediction for mustard crop biomass done at 50 percent flowering and pod formation stages by InfoCrop-mustard had RMSE values 2144.9 and 1443.2 kg ha^{-1} for Pusa Tarak, 2392.4 and 1668.1 kg ha^{-1} for RH-406 and, 2077.3 and 1412.0 kg ha^{-1}, for Girraj respectively. nRMSE was calculated to find out the accuracy of the model. For above-ground biomass prediction of mustard done at pod formation stage, nRMSE values were less than 20%, and prediction done at 50 percent flowering stage nRMSE values were more than 20%. It indicates that model performed good for above-ground biomass prediction of mustard done at pod formation stage and fair for prediction done at50 percent flowering stage. Prediction accuracy was found to be better for timely sown crop compared to late and very late sown. Positive values of percent deviation indicate over-estimation in biomass prediction.

4.2.5.2 Seed yield

The seed yield prediction done by the InfoCrop-mustard model at two different stages 50 percent flowering and at pod formation are shown in table 4.10. The percent deviation of seed yield prediction done by the InfoCrop-mustard model at 50 percent flowering stage from observed yield was 15.6, 11.5 and 17.7 in timely sown crop, 24.5, 15.8 and 22.0 in late sown crop and 59.4, 64.2 and 56.2 in very late sown crop for Pusa Tarak, RH-406 and Girraj, respectively. Percentage deviation of predicted yield done at pod formation stage from observed yield was 8.0, 15.9 and 40.3 for Pusa Tarak, 7.8, 13.7 and 37.1 for RH-406 and 4.6. 17.5 and 36.8 kg ha^{-1} for Girraj in timely, late and very late sown crop. The nRMSE values were less than 20% for seed yield prediction done at pod formation stage, whereas nRMSE vaues were between 25 to 30% for seed yield prediction done at 50 percent flowering stage for all the cultivars.

Table 4.10 Multistage mustard seed yield prediction during *Rabi* 2017-18

Cultivars	Sowing Time	Days after sowing	Biomass (kg ha^{-1}) Observed	Biomass (kg ha^{-1}) Predicted	Percentage deviation	RMSE (kg ha^{-1})	nRMSE (%)
Prediction at 50 percent flowering stage							
Pusa Tarak	Timely	50	1900	2196	15.6	449.1	30.6
	Late	53	1453	1810	24.5		
	Very late	55	1051	1675	59.4		
RH-406	Timely	65	2374	2648	11.5	539.3	29.0
	Late	68	1904	2203	15.8		
	Very late	70	1310	2152	64.2		
Girraj	Timely	65	2192	2580	17.7	504.2	29.2
	Late	67	1789	2183	22.0		
	Very late	71	1204	1880	56.1		
Prediction at pod formation stage							
Pusa Tarak	Timely	78	1900	2052	8.0	292.0	19.9
	Late	77	1453	1685	15.9		
	Very late	75	1051	1474	40.3		
RH-406	Timely	99	2374	2560	7.8	336.6	18.1
	Late	98	1904	2164	13.7		
	Very late	95	1310	1798	37.1		
Girraj	Timely	99	2192	2292	4.6	318.56	18.4
	Late	98	1789	2102	17.5		
	Very late	95	1204	1647	36.8		

This indicates that the InfoCrop-mustard model perform good for seed yield prediction done at pod formation stage and fair for seed yield prediction done at 50 % pod formation stage. The multistage prediction by the InfoCrop-mustard model for biomass and seed yield done at pod formation stage was better than prediction done at 50 percent flowering stage. The multistage prediction done by the InfoCrop-mustard model for biomass and seed yield was better for timely sown crop followed by late and very late sown crop.

4.3 Mustard crop yield prediction by empirical models

There are various machine learning techniques used for crop yield prediction. To fulfill the objective of the research work, to develop and evaluate weather based empirical models for mustard yield prediction, few techniques such as artificial neural network (ANN), support vector machine (SVM) and random forest (RF) were selected. The Z variables developed by weather parameters were taken as input parameters for developing a crop yield prediction model using ANN, SVM and random forest techniques. The z variables were selected by variable selection using the stepwise multiple linear regression (SMLR) technique and variable extraction using the principal component analysis (PCA) technique. The principal components (PCs) was selected on the basis of eigenvalues (>1) were able to describe more than 90 percent variability of input data set for the region. The prediction graph (Fig 4.16 to 4.27) show three lines: the outer blue one is the prediction interval (95%), the inner blue one is the confidence interval (95%), and the red one is the regression line. A 95 % confidence interval means there is a 95 % probability that the true best fit line for the population lies within the confidence interval. A 95 % prediction interval means there is 95 % y value to be found for a certain x value will be within the interval range.

4.3.1 Mustard crop yield prediction for IARI, New Delhi

Weather and yield data were collected for the last 35 years for developing the crop yield prediction model and predicting the yield of mustard crop for the IARI, New Delhi. Calibration and validation of the model developed using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques in R statistical software version 3.1.3.for IARI, New Delhi are shown in Fig. 4.16. Crop yield prediction for the IARI, New Delhi was done by these developed models.

Model performances during calibration and validation are shown in Table 4.11. Results showed that the model developed by SMLR-ANN, SMLR-SVM and SMLR-RF performed good with RMSE value 151.2, 150.1 and 223.8 kg ha^{-1} during calibration, respectively. The RMSE value during validation was highest for model developed using SMLR-RF techniques (250.8 kg ha^{-1}) followed by SMLR-ANN (246.7 kg ha^{-1}) and SMLR-SVM (236.2 kg ha^{-1}), respectively. Mean absolute error (MAE) values during calibration were 108.3, 91.3 and 183.7 kg ha^{-1}, respectively and during validation were 179.9, 196.2 and 229.1 kg ha^{-1}, respectively, for models developed by SMLR-ANN, SMLR-SVM and SMLR-RF. The values of nMAE during calibration were 5.64, 4.75 and 9.55 and during validation were 9.02, 9.84 and 11.48 for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively.

Table 4.11 Performance of the mustard yield prediction models developed by different techniques for IARI, New Delhi

Accuracy	**SMLR-ANN**		**SMLR-SVM**		**SMLR-RF**	
Parameters	**Calibration**	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	108.36	179.86	91.28	196.21	183.70	229.11
nMAE (%)	5.64	9.02	4.75	9.84	9.55	11.48
RMSE (kg ha^{-1})	151.25	246.71	150.14	236.21	223.84	250.83
nRMSE (%)	7.87	12.37	7.81	11.84	11.64	12.57
RPD	2.79	1.70	2.81	1.78	1.88	1.68
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	65.24	139.74	38.31	155.96	186.05	244.70
nMAE (%)	3.44	6.80	2.02	7.59	9.82	11.91
RMSE (kg ha^{-1})	101.18	214.24	39.83	187.43	217.41	271.52
nRMSE (%)	5.34	10.43	2.10	9.12	11.48	13.22
RPD	4.15	1.90	10.55	2.17	1.93	1.50

The nRMSE values calculated during calibration were 7.87, 7.81 and 11.64 and during validation, they were 12.37, 11.84 and 12.57 for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The model predictions for the IARI, New Delhi were good having nRMSE value < 15 % for all three developed models. The ratio of performance to deviation (RPD) for the model developed by SMLR-ANN, SMLR-SVM and SMLR-RF techniques were 2.79, 2.81 and 1.88 during calibration, and 1.70, 1.78 and 1.68 during validation, respectively. Based on value of model accuracy parameters (RMSE, nRMSE, and RPD), SMLR-SVM model performed better, followed by SMLR-ANN and SMLR-RF for mustard yield prediction of IARI, New Delhi.

Performance during calibration and validation of the models developed using variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM), and variable extraction by PCA and RF (PCA-RF) techniques in R statistical software vesion 3.1.3. for IARI, New Delhi are shown in Fig. 4.17. Crop yield prediction was done by these developed models, PCA-ANN, PCA-SVM and PCA-RF for the Delhi region. Performances during calibration and validation period for developed model are shown in Table 4.11. The result showed that the RMSE values during calibration were lowest for PCA-SVM (39.8 kg ha^{-1}) followed by PCA-ANN (101.2 kg ha^{-1}) and PCA-RF (217.4 kg ha^{-1}), respectively. During validation, PCA-RF showed the highest RMSE value 271.5 kg ha^{-1} followed by 214.2 kg ha^{-1} for PCA-ANN and 187.4 kg ha^{-1} for PCA-SVM. Mean absolute error (MAE) during calibration, was lowest for PCA-SVM (38.3 kg ha^{-1}) followed by PCA-ANN (65.2 kg ha^{-1}) and PCA-RF (186.0 kg ha^{-1}). The values of MAE during validation were the lowest for PCA-ANN (139.7 kg ha^{-1}) followed by PCA-SVM (155.5 kg ha^{-1}) and PCA-RF (244.7 kg ha^{-1}) for IARI, New Delhi.

During calibration, nMAE values were 3.44, 2.02 and 9.82 and during validation, they were 6.80, 7.59 and 11.91 for PCA-ANN, PCA-SVM and PCA-RF, respectively. During calibration, values for nRMSE were 5.34, 2.10 and 11.48 and during validation, they were 10.43, 9.12 and 13.22 for PCA-ANN, PCA-SVM and PCA-RF, respectively. The model predictions for the IARI, New Delhi were excellent having nRMSE values < 10 % for models developed using PCA-SVM and good having nRMSE values 10.43 and 13.22 for the model developed using PCA-ANN and PCA-RF techniques. The ratio of

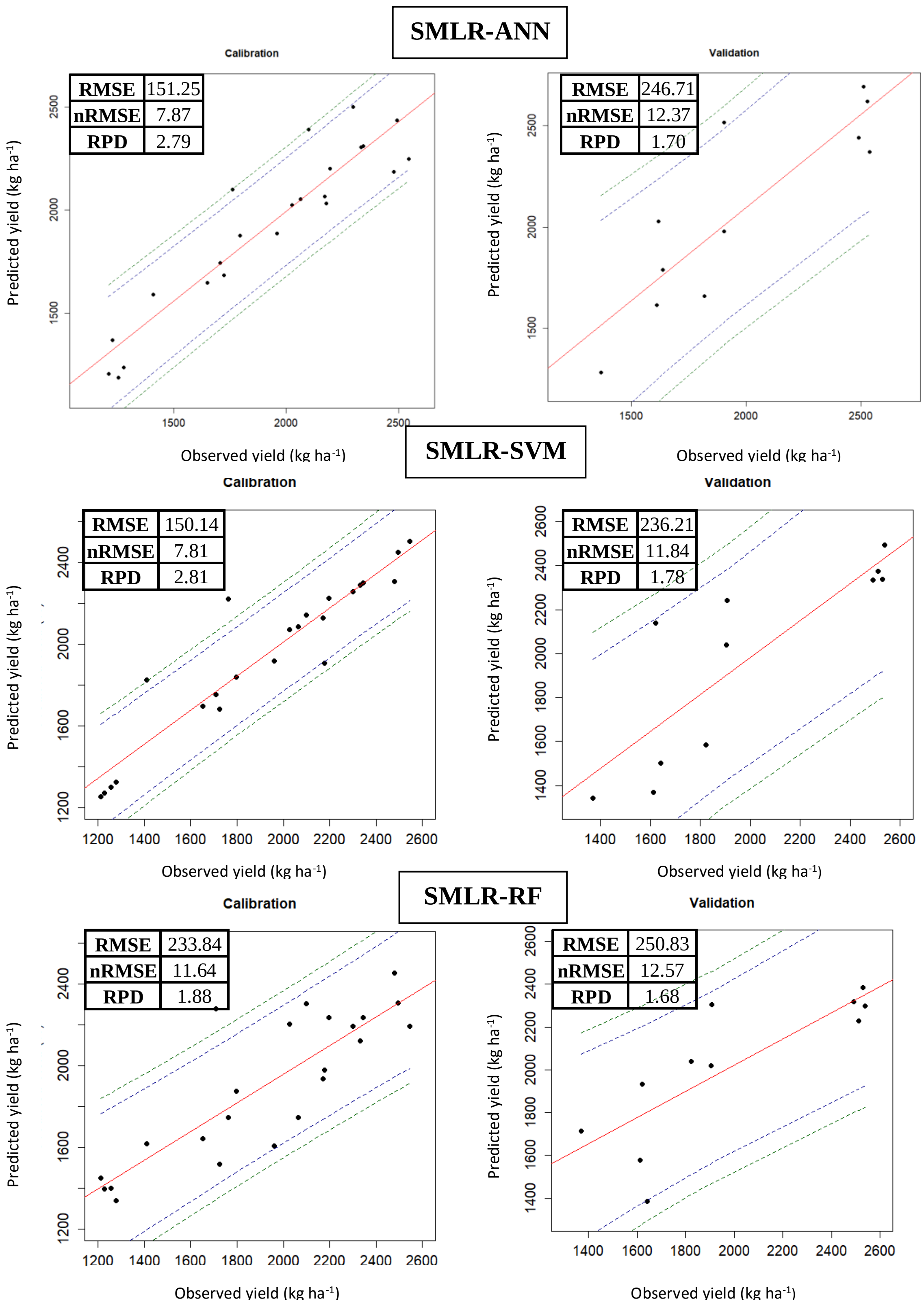

Fig 4.16 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for IARI, New Delhi

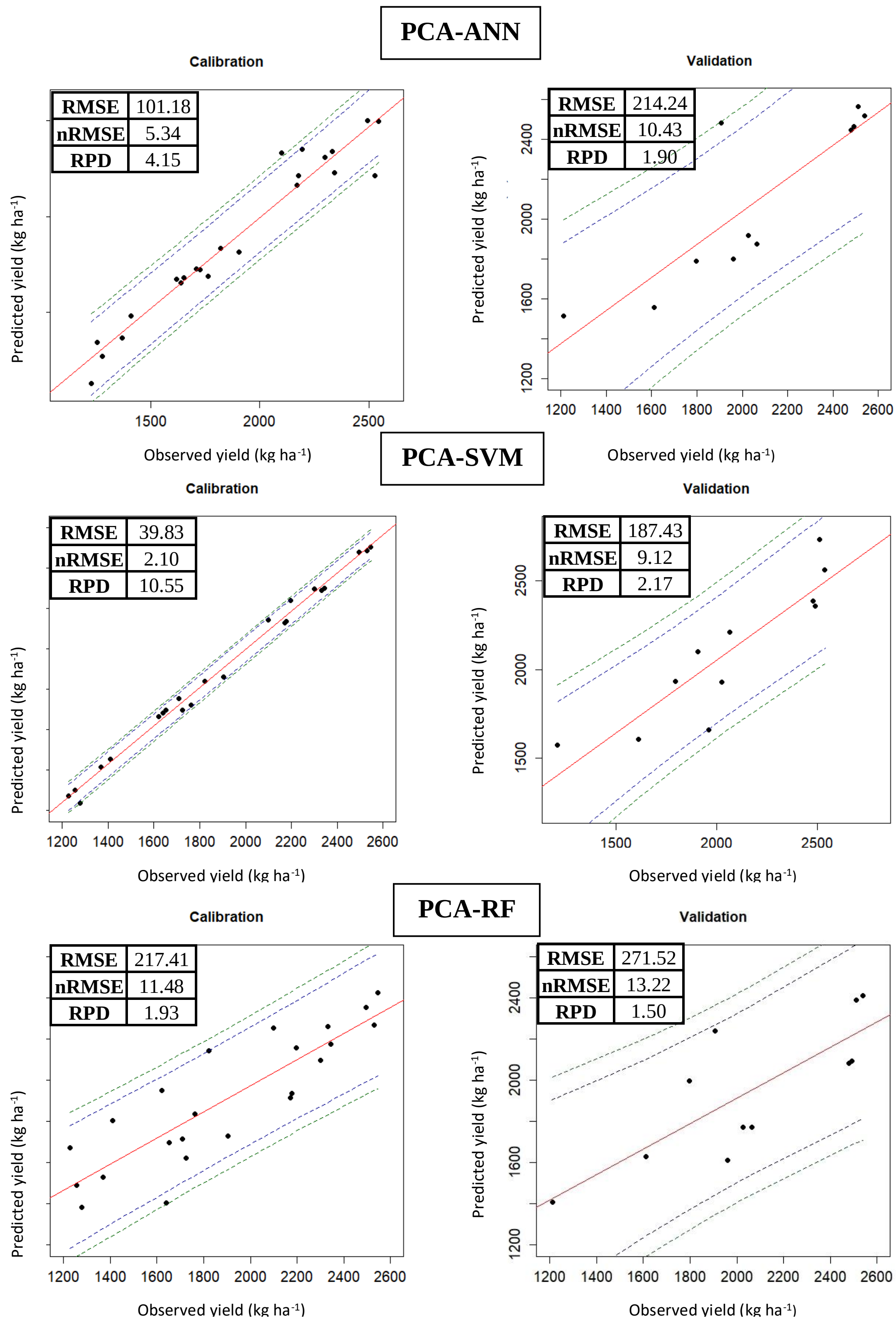

Fig 4.17 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for IARI, New Delhi

performance to deviation (RPD) values for the model developed by PCA-ANN, PCA-SVM and PCA-RF were 4.15, 10.55 and 1.93 during calibration, and 1.90, 2.17 and 1.50 during validation, respectively. Based on value of model accuracy parameters (RMSE, nRMSE, and RPD), among all the three models developed using principal component analysis extraction techniques, PCA-SVM was performing best followed by PCA-ANN and PCA-RF respectively for mustard yield prediction of IARI, New Delhi..

By comparing the models developed either using variable selection by SMLR or variable extraction by PCA on the basis of RMSE, nRMSE, and RPD, SVM is performing best, followed by ANN and RF.

4.3.2 Mustard crop yield prediction for different zones of Rajasthan:

The National Agricultural Research Project (NARP) classified the agro-climatic zones on the basis of soil type, temperature, rainfall and geological constraints. ICAR had divided the country into 127 agro-climatic zones and decided their appropriate boundaries. According to the NARP report, Rajasthan state has ten agro-climatic zones (Fig. 3.3). Five major mustard growing zones selected for yield predictions are illustrated in the following sub-sections.

4.3.2.1 Zone 1 (Alwar-Bharatpur):

The climatic conditions of the Alwar and the Bharatpur are similar, as reported in NARP. Weather data along with the yield of the mustard crop were collected for the last 32 years for each district in Zone 1. The model was developed for Zone 1 using variable selection either by SMLR or variable extraction by PCA and ANN, SVM and RF techniques.

Calibration and validation of the model developed using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques in R statistical software version 3.1.3. for Zone -1 are shown in Fig. 4.18. Crop yield prediction was done for Zone 1 by using the developed models, SMLR-ANN, SMLR-SVM and SMLR-RF. During calibration and validation, performances of different models developed are shown in Table 4.12. The results showed values of RMSE during calibration were 171.63, 86.81 and 193.57 kg ha^{-1}, respectively, for the models developed using SMLR-ANN, SMLR-

SVM and SMLR-RF techniques. The values of RMSE during validation, were maximum for the model developed using SMLR-ANN (220.35 kg ha^{-1}) followed by SMLR-SVM (198.74 kg ha^{-1}) and SMLR-RF (197.08 kg ha^{-1}), respectively. The Mean absolute error (MAE) values were 136.02, 52.75 and 145.55 kg ha^{-1} during calibration, and 167.88, 159.08 and 134.16 kg ha^{-1} during validation for the model developed by SMLR-ANN, SMLR-SVM and SMLR-RF, techniques, respectively. The nRMSE values during calibration were lowest for SMLR-SVM (6.97 %) followed by SMLR-ANN (13.77 %) and SMLR-RF (15.82 %); however during validation, the nRMSE value was lowest for SMLR-RF (16.30 %) followed by SMLR-SVM (16.73 %) and SMLR-ANN (18.55 %). Mustard yield prediction for Zone 1 was good for all the developed models having nRMSE value < 20 %. During calibration, the values of nMAE were 10.91, 4.23 and 11.30, and during validation, they were 14.14, 13.40 and 11.68 for SMLR-ANN, SMLR-SVM and SMLR- RF, respectively. The RPD values were 1.80, 3.63 and 2.12 during calibration, and 1.60, 1.56 and 1.40 during validation by SMLR-RF, SMLR-SVM and SMLR-ANN, respectively. The results on the basis of model accuracy parameters, RMSE, nRMSE, and RPD showed that SMLR-SVM model performed better followed by SMLR-ANN and SMLR-RF during calibration. However, during validation, SMLR-RF model performed better followed by SMLR-SVM and SMLR-ANN. This may be due to over fitting in calibration. Over fitting occurs when model tries to cover all the data points or more than the required data points present in the given dataset. SVM and ANN tries to cover more similar data point than RF.

Performance of the models during calibration and validation developed for mustard yield prediction for Zone-1, using variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM), and variable extraction by PCA and RF (PCA-RF) techniques in R statistical software version 3.1.3. are shown in Fig. 4.19. Mustard crop yield prediction were done by these developed models, PCA-ANN, PCA-SVM and PCA-RF for Zone -1. Model performance was assessed using MAE, nMAE, RMSE, nRMSE and RPD values. Performances during calibration and validation period for developed model are shown in Table 4.12. The values of MAE during calibration of different models PCA-ANN, PCA-SVM and PCA-RF were 111.21, 46.32 and 151.12 kg ha^{-1}, respectively, and during validation, were 134.51, 128.97 and

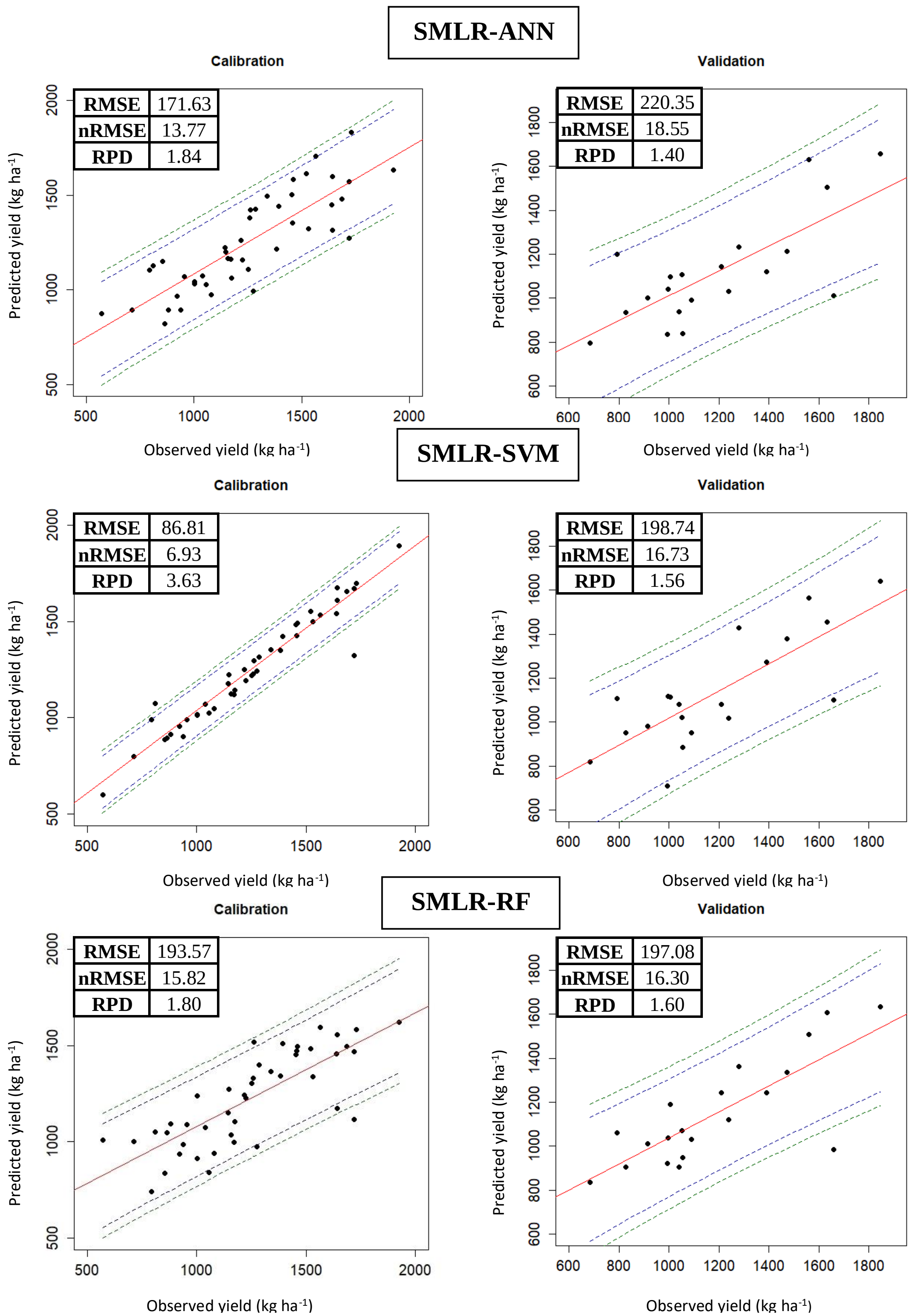

Fig 4.18 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for Zone I of Rajasthan

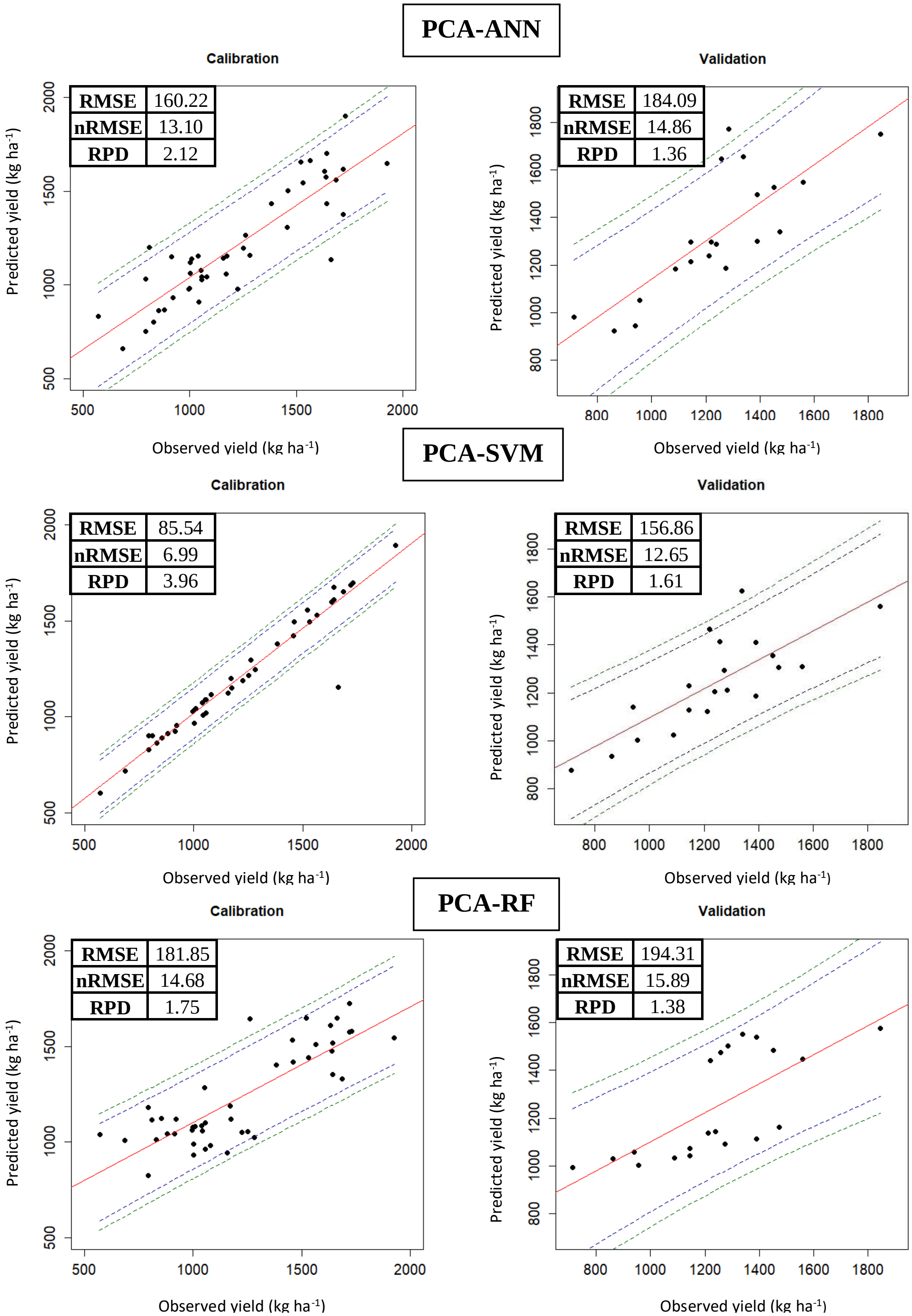

Fig 4.19 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for Zone I of Rajasthan

160.80 kg ha^{-1}. The values of nRMSE during calibration were 9.10, 3.79 and 12.36 and during validation, were 10.86, 10.41 and 12.98, respectively, for model developed by PCA-ANN, PCA-SVM and PCA-RF. The RMSE value during calibration was lowest for PCA-SVM (85.54 kg ha^{-1}) followed by PCA-ANN (160.22 kg ha^{-1}) and PCA-RF (181.85 kg ha^{-1}) and during validation, was 156.86 kg ha^{-1} for PCA-SVM, 184.09 kg ha^{-1} for PCA-ANN and 194.29 kg ha^{-1} for PCA-RF, respectively.

Table 4.12 Performance of the mustard yield prediction models developed by different techniques for Zone I of Rajasthan

Accuracy parameters	**SMLR-ANN**		**SMLR-SVM**		**SMLR-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	136.02	167.88	52.75	159.08	145.55	134.16
nMAE (%)	10.91	14.14	4.23	13.40	11.30	11.68
RMSE (kg ha^{-1})	171.63	220.35	86.81	198.74	193.57	197.08
nRMSE (%)	13.77	18.55	6.97	16.73	15.82	16.30
RPD	1.84	1.40	3.63	1.56	1.80	1.60
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	111.21	134.51	46.328	128.979	151.12	160.80
nMAE (%)	9.10	10.86	3.789	10.409	12.36	12.98
RMSE (kg ha^{-1})	160.22	184.09	85.549	156.861	181.85	194.29
nRMSE (%)	13.10	14.86	6.996	12.659	14.68	15.89
RPD	2.12	1.36	3.96	1.61	1.75	1.38

The RPD values were 2.12, 3.96 and 1.75 during calibration, and 1.36, 1.61 and 1.38 for validation, by PCA-ANN, PCA-SVM and PCA-RF, respectively. The values of nRMSE during calibration were lowest for PCA-SVM (6.99 %) followed by PCA-ANN (13.10 %) and PC-RF (14.68 %). During validation the value of nRMSE were lowest for PCA-SVM (12.66 %) followed by PCA-ANN (14.86 %) and PCA-RF (15.89 %). Mustard yield prediction for Zone 1 were good having nRMSE value < 20 % for all the developed model PCA- ANN, PCA-SVM and PCA-RF techniques. Based on the model accuracy parameters MAE. nMAE, RMSE, nRMSe and RPD, among all the six models developed for mustard yield prediction for Zone -1, PCA-SVM performed best followed by PCA-ANN, PCA-RF, SMLR-RF, SMLR-SVM and SMLR-ANN.

4.3.2.2 Zone II (Sawai madhopur-Kota)

According to the NARP report, the Sawai-madhopur and the Kota districts of Rajasthan comes under the same agro-climatic zone II. Weather and yield data of the mustard crop were collected from Sawai-madhopur for 30 years, and Kota district for 28 years of Rajasthan. The model was developed using variable selection by stepwise regression and principal component analysis using ANN, SVM, RF techniques in R statistical software.

Calibration and validation of the model developed using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques for Zone II are shown in Fig. 4.20. Mustard crop yield prediction for Zone II were done by these developed models SMLR-ANN, SMLR-SVM and SMLR-RF.

The performance of the prediction models were analyzed using RMSE values. The result showed RMSE values 138.16 kg ha^{-1}, 175.34 kg ha^{-1} and 199.03 kg ha^{-1} during calibration, 174.69 kg ha^{-1}, 175.34 kg ha^{-1} and 199.03 kg ha^{-1} during validation for the model developed by SMLR-SVM, SMLR-ANN, and SMLR-RF techniques, respectively (Table 4.13). The nRMSE values during calibration were lowest for SMLR-SVM (13.11 %) followed by SMLR-ANN (14.23 %) and SMLR-RF (16.12 %). During validation, nRMSE value was 16.00 % for SMLR-SVM, 16.06 % for SMLR-ANN and 18.23 % for SMLR-RF. All three developed models performed good for mustard yield prediction for zone II having nRMSE less than 20 %. Ratio of performance to deviation (RPD) values for the model developed by SMLR-ANN, SMLR-SVM and SMLR-RF techniques were 2.08, 2.26, 1.83 during calibration and 2.03, 2.04, 1.79 during validation, respectively. Mean absolute error (MAE) during calibration was 114.1, 93.9 and 135.5 kg ha^{-1}, and during validation were 131.2, 142.6 and 149.8 kg ha^{-1}, respectively for the model developed by SMLR-ANN, SMLR-SVM and SMLR-RF techniques. The value of nMAE during calibration were 10.83, 9.20 and 12.85 and during validation they were12.01,

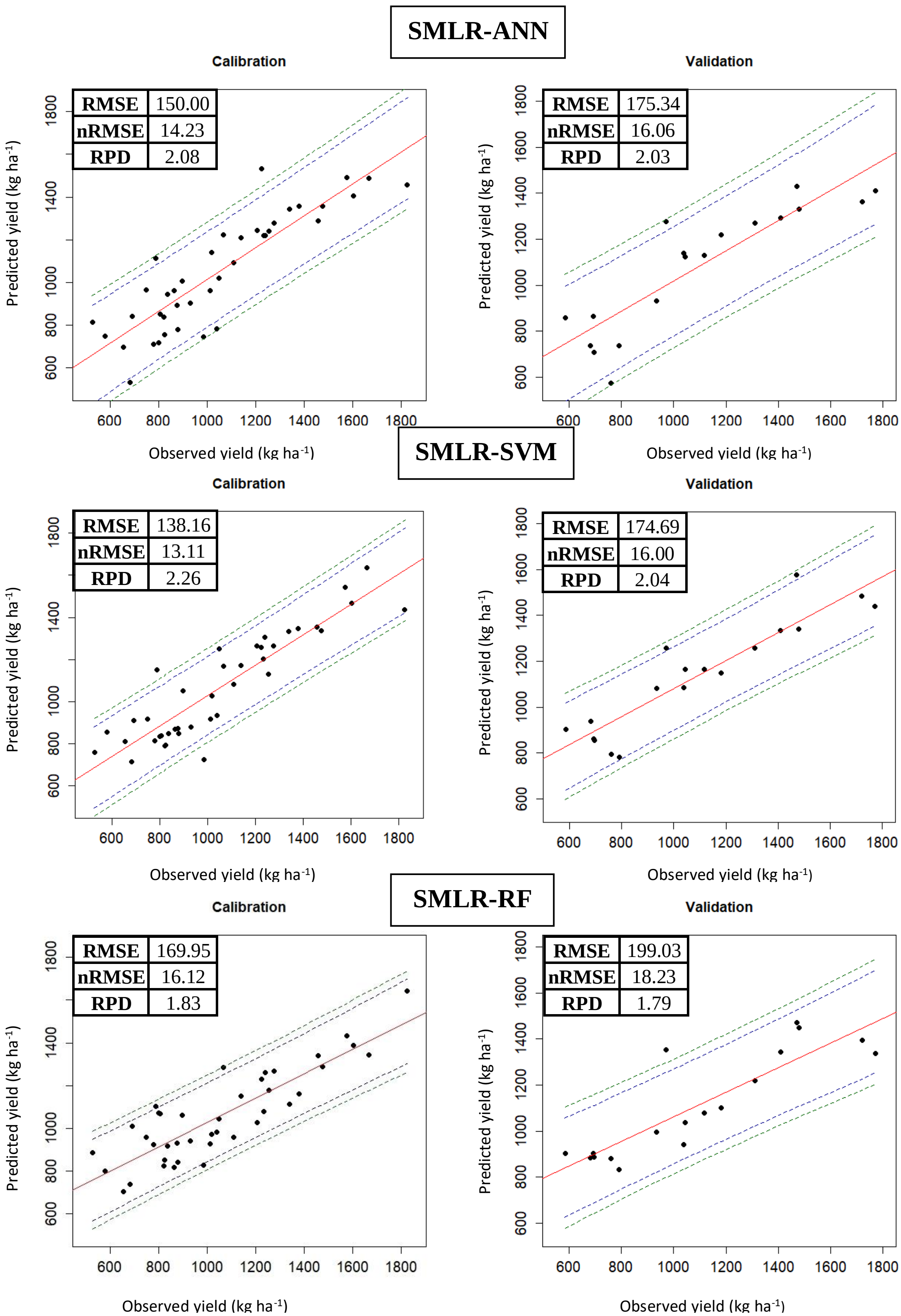

Fig 4.20 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for Zone II of Rajasthan

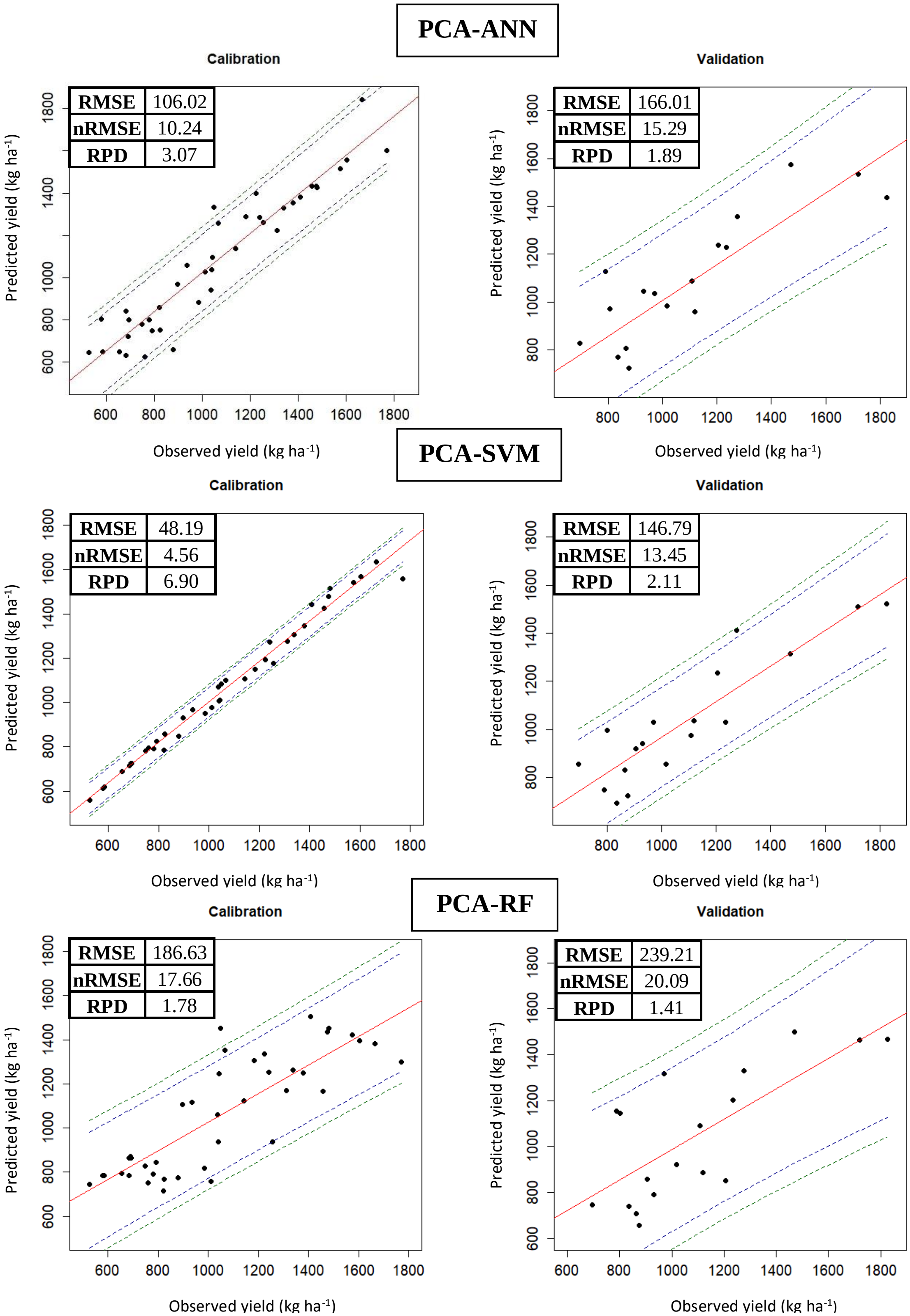

Fig 4.21 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for Zone II of Rajasthan

13.06 and 13.72 for SMLR-ANN, SMLR-SVM and SMLR-RF respectively. Results showed that on the basis of model accuracy parameters RMSE, nRMSE and RPD, among all three models developed, the SMLR-SVM model performed better followed by SMLR-ANN and SMLR-RF for mustard yield prediction of Zone II of Rajasthan.

Table 4.13 Performance of the mustard yield prediction models developed by different techniques for Zone II of Rajasthan

Accuracy parameters	**SMLR-ANN**		**SMLR-SVM**		**SMLR-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	114.12	131.17	96.97	142.60	135.46	149.81
nMAE (%)	10.83	12.01	9.20	13.06	12.85	13.72
RMSE (kg ha^{-1})	150.00	175.34	138.16	174.69	169.95	199.03
nRMSE (%)	14.23	16.06	13.11	16.00	16.12	18.23
RPD	2.08	2.03	2.26	2.04	1.83	1.79
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	82.68	130.47	38.05	123.95	153.84	177.50
nMAE (%)	7.82	12.02	3.60	11.36	14.56	16.26
RMSE (kg ha^{-1})	108.20	166.01	48.19	146.79	186.63	219.31
nRMSE (%)	10.24	15.29	4.56	13.45	17.66	20.09
RPD	3.07	1.89	6.90	2.11	1.78	1.41

Calibration and validation of model developed for Zone-II, using variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM) and, variable extraction by PCA and RF (PCA-RF) techniques in R statistical software 3.1.3. are shown in Fig. 4.21. Mustard yield prediction were done by these developed models, such as PCA-ANN, PCA-SVM and PCA-RF for Zone II of Rajasthan state. Model performance was assessed using MAE, nMAE, RMSE, nRMSE and RPD values. Performances during calibration and validation for developed models are shown in Table 4.13. The RMSE values during calibration were lowest for PCA-SVM (48.19 kg ha^{-1}) followed by PCA-ANN (108.24 kg ha^{-1}) and PCA-RF (186.63 kg ha^{-1}), and during validation, were 146.79 kg ha^{-1}, 166.01 kg ha^{-1} and 219.31 kg ha^{-1} by PCA-SVM, PCA-ANN and PCA-RF, respectively. The RPD values were 3.07, 6.90 and 1.78 during calibration, and 1.89, 2.11 and 1.41 during validation, by PCA-ANN, PCA-SVM and PCA-RF, respectively. The value of nRMSE during calibration was lowest for PCA-SVM (4.56 %) followed by PCA-ANN (10.24 %) and PC-RF (17.66). During validation, nRMSE value was lowest for PCA-SVM (13.45 %) followed by PCA-ANN (15.29 %) and PCA-RF (20.09 %). The value of nRMSE was less than 20 %, indicates that all three models developed performed good for mustard yield prediction for Zone II of Rajasthan. The value of MAE during model calibration was highest in PCA-RF 153.84 kg ha^{-1} followed by 82.68 kg ha^{-1} for PCA-ANN and 38.05 kg ha^{-1} for PCA-SVM. During validation MAE values were 123.9, 130.5 and 177.5 kg/$ha^{-1}$for models developed by PCA-SVM, PCA-ANN and PCA-RF techniques, respectively.

The RPD values were more than 2 for PCA-SVM and PCA-ANN during calibration, whereas during validation, only PC-SVM model had RPD value more than 2. The high value of RPD and low value of RMSE and nRMSE indicates good agreement between model outputs and observed values. Among all the three models developed, PCA-SVM has a low nRMSE 13.45 % and a high 2.11 RPD value as compared to other two models. Hence PCA-SVM model performed better, followed by PCA-ANN and PCA-RF. On the basis of model accuracy parameters, RMSE, nRMSe and RPD, among all six models developed for mustard yield prediction for Zone II of Rajasthan, PCA-SVM performed bests followed by PCA-ANN, SLMR-SVM, SLMR-ANN. SLMR-RF and PCA-RF.

4.3.2.3 Zone III (Udaipur-Jhalawar)

Udaipur and Jhalawar have the same soil type and climatic normal; therefore, according to the NARP report, they were put together in Zone III of Rajasthan. A long-term weather and yield data of 25 years were collected for developing and predicting mustard crop yield for Zone III of Rajasthan. Calibration and validation of the model developed using variable selection by SMLR, and ANN (SMLR-ANN), variable selection by SMLR, and SVM (SMLR-SVM), and variable selection by SMLR, and RF

(SMLR-RF) techniques in R software version 3.1.3. for Zone III of Rajasthan are shown in Fig. 4.22. Mustard yield prediction for Zone III was done using developed models SMLR-ANN, SMLR-PC and SMLR-RF and model performance are given in Table 4.14. During calibration, MAE value was found to be highest for SMLR-ANN (120.03 kg ha^{-1}), followed by SMLR- RF (114.22 kg ha^{-1}) and SMLR-SVM (75.59 kg ha^{-1}). During validation, MAE values were highest for SMLR-ANN (126.34 kg ha^{-1}), followed by SMLR- RF (124.22 kg ha^{-1}) and SMLR- SVM (107.76 kg ha^{-1}). The values of nMAE were 12.37, 7.79 and 11.77 during calibration, and 13.96, 11.91 and 13.73 during validation, for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. RMSE values during calibration were 156.82, 108.69 and 147.20 kg ha^{-1}, and during validation, they were 171.78, 13.52 and 159.96 kg ha^{-1}, for SMLR-ANN, SMLR-SVM and SMLR-RF.

Table 4.14 Performance of the mustard yield prediction models developed by different techniques for Zone III of Rajasthan

Accuracy parameters	**SMLR-ANN**		**SMLR-SVM**		**SMLR-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	120.03	126.34	75.59	107.76	114.22	124.22
nMAE (%)	12.37	13.96	7.79	11.91	11.77	13.73
RMSE (kg ha^{-1})	156.82	171.78	108.69	130.52	147.20	159.96
nRMSE (%)	16.16	18.98	11.20	14.42	15.16	17.68
RPD	1.67	1.36	2.41	1.80	1.80	1.78
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	77.71	122.64	31.36	103.65	109.34	141.20
nMAE (%)	8.13	13.05	3.28	11.16	11.44	15.02
RMSE (kg ha^{-1})	107.88	158.91	55.77	153.46	134.18	189.85
nRMSE (%)	11.29	16.91	5.84	11.50	14.04	20.20
RPD	2.28	1.79	3.46	2.36	1.99	1.40

During calibration, nRMSE values were 16.16, 11.20 and 15.16 %, and during validation, they were 18.98, 14.42 and 17.68 % for SMLR-ANN, SMLR-SVM and SMLR-RF. Based on nRMSE values results showed that all three models for mustard yield prediction for Zone III of Rajasthan performed good having nRMSE value less than 20 %. RPD values during calibration were 1.67, 2.41 and 1.80, and during validation, were 1.36, 1.80 and 1.78, for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The RPD value was more than 2 for SMLR-SVM and SMLR-ANN, but for SMLR-RF, it varied between 1.5 to 2 during calibration and validation. Based on model accuracy parameters RMSE, nRMSE and RPD values, among all three models, SMLR-SVM performed better, followed by SMLR-ANN and SMLR-RF.

Model calibration and validation developed by variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM), variable extraction by PCA and RF (PCA-RF) techniques in R statistical software version 3.1.3. for Zone III of Rajasthan are presented in Fig. 4.23. Mustard yield prediction was done for Zone III of Rajasthan state using developed model PCA-ANN, PCA-SVM and PCA-RF, and model performance are shown in table 4.14. The Value of RMSE during calibration and validation was lowest 55.17 and 153.46 kg ha^{-1}, for PCA-SVM; 107.88 and 158.91 kg ha^{-1}, for PCA-ANN; 134.18 and 189.85 kg ha^{-1}, for PCA-RF; for Zone III of Rajasthan (Table 4.14). Mustard yield prediction model developed using PCA-SVM techniques had the lowest mean absolute error (31.3, 103.7 kg ha^{-1}), followed by PCA-ANN (77.71, 122.6 kg ha^{-1}) and PCA-RF (109.3, 142.2 kg ha^{-1}) during calibration and validation, respectively. The nMAE values during calibration were 8.13, 3.28 and 11.44 %, and during validation, were 13.05, 11.16 and 15.02 % for PCA-ANN, PCA-SVM and PCA-RF, respectively. The values of nRMSE were 11.29, 5.84 and 14.04 % during calibration, and 16.91, 11.50 and 20.20 % during validation for model developed by PCA-ANN, PCA-SVM and PCA-RF techniques, respectively. Model performance was good for PCA-SVM and PCA-ANN, having nRMSE value less than 20 % and fair for PCA-RF having nRMSE value 20.20 %. The RPD values during calibration, were 2.28, 3.46 and 1.99, and during validation, were 1.79, 2.36 and 1.40, respectively. The model developed by PCA- SVM showed better agreement between observed and predicted RPD values.

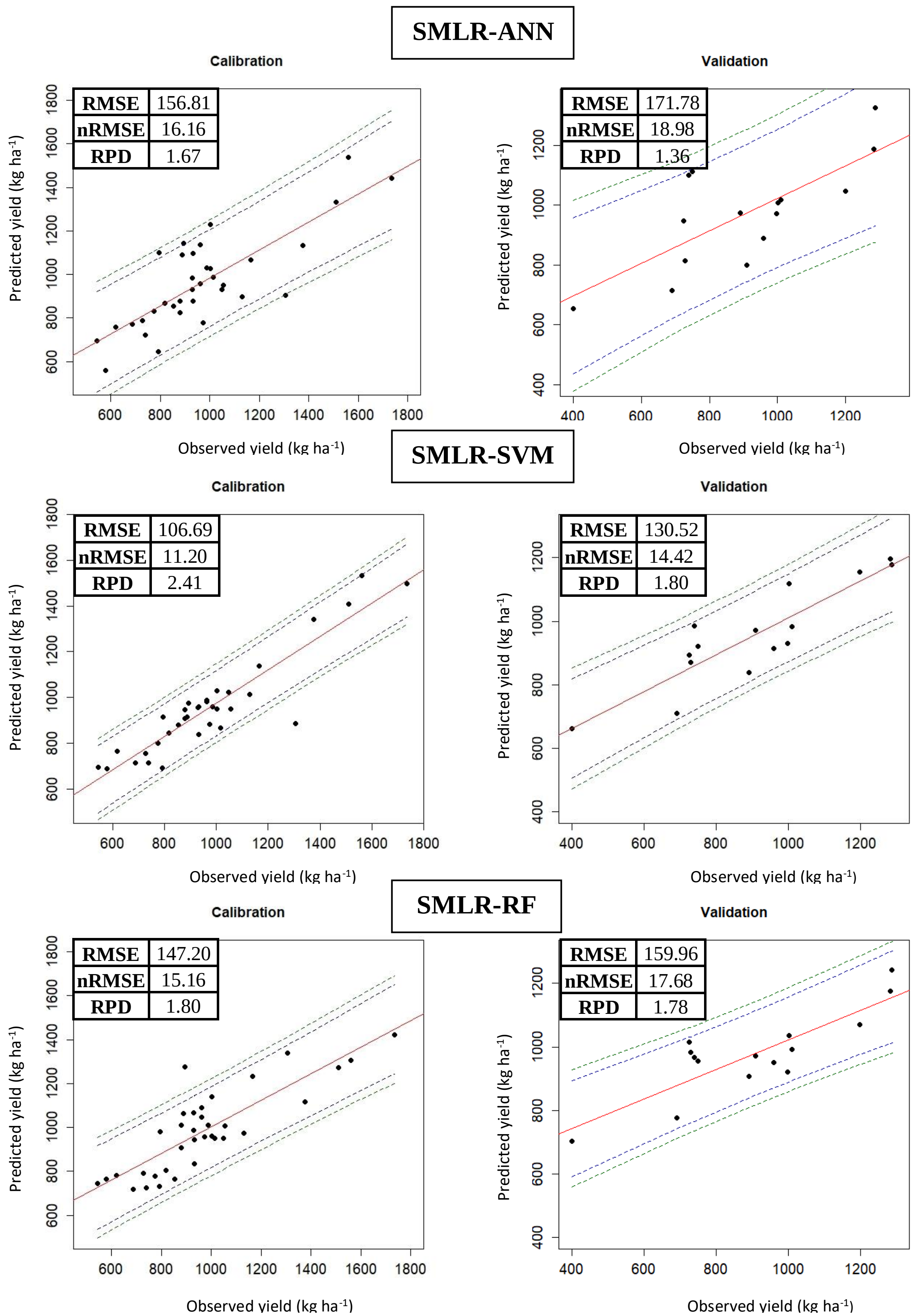

Fig 4.22 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for Zone III of Rajasthan

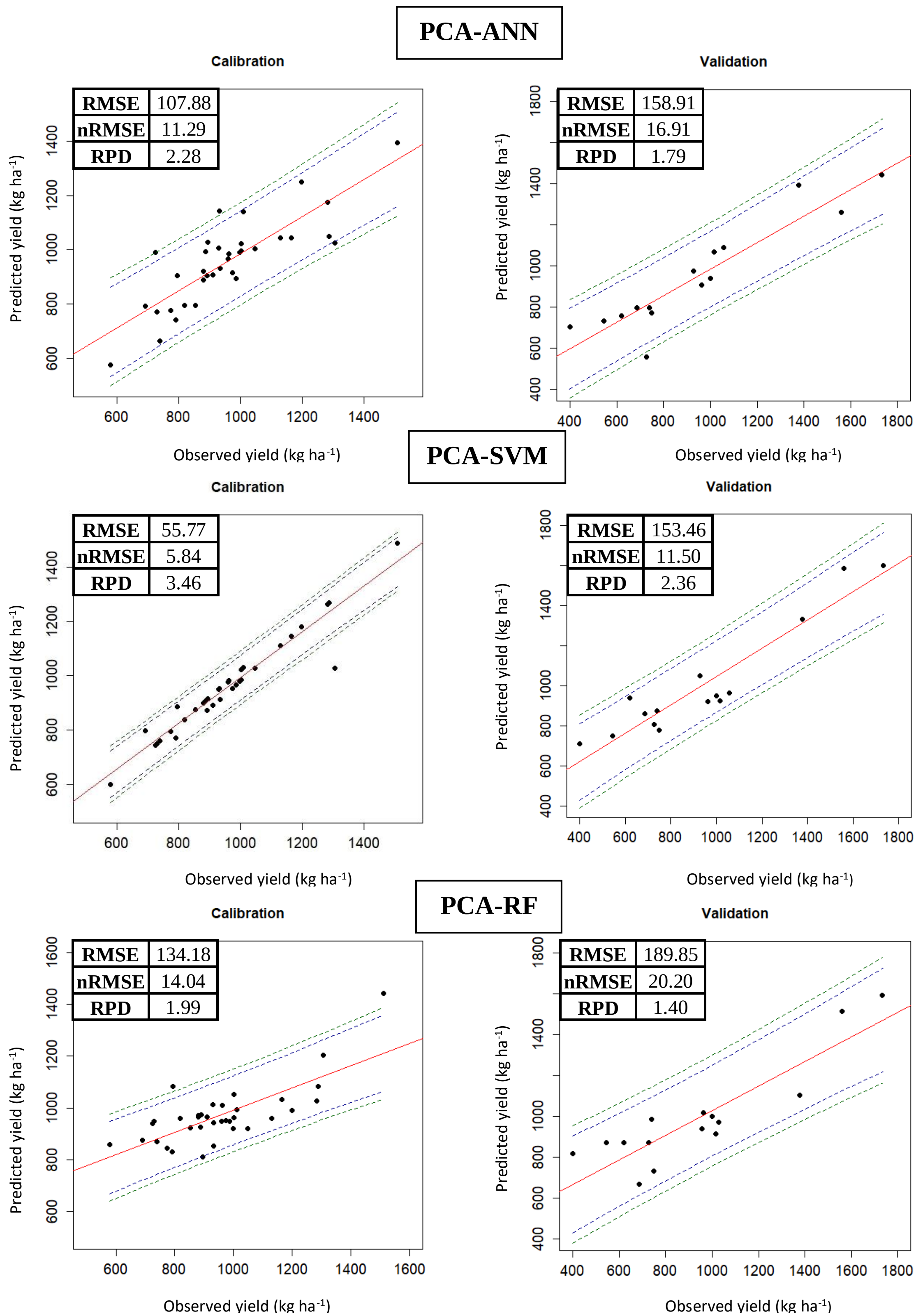

Fig 4.23 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for Zone III of Rajasthan

On the basis of model accuracy parameters, RMSE, nRMSE and RPD, among all developed models, PCA-SVM performed better, followed by SMLR-SVM, PCA-ANN, SMLR-RF, SMLR-ANN and PCA-RF respectively for mustard yield prediction for Zone III of Rajasthan.

4.3.2.4 Zone IV (Pali-Jodhpur)

The climatic conditions of Pali and Jodhpur districts are similar as reported by the NARP; therefore, both distract were put together in Zone IV of Rajasthan. Weather data along with the yield of mustard crop for the last 31 years were collected for each distracts. Collected data was used to develop and predict mustard yield by variable selection by stepwise regression model and variable extraction by principal component analysis, using ANN, SVM and RF techniques in R software version 3.1.3.

Calibration and validation of the model developed using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques in R software version 3.1.3. for Zone IV of Rajasthan are shown in Fig. 4.24. Performance of mustard yield prediction done for Zone IV of Rajasthan using developed model SMLR-ANN, SMLR-SVM and SMLR-RF are given in Table 4.15. The values of RMSE during calibration were 160.39, 104.81 and 163.47 kg ha^{-1}, respectively for SMLR-ANN, SMLR-SVM and SMLR-RF. During validation, RMSE value was found highest for SMLR-ANN (181.25 kg ha^{-1}), followed by SMLR-RF (165.92 kg ha^{-1}) and SMlR-SVM (162.41 kg ha^{-1}). The mean absolute error (MAE) values were 133.0, 72.3 and 134.9 kg ha^{-1} during calibration, and 139.9, 125.1 and 133.4 kg ha^{-1} during validation, for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The nMAE values were 15.08, 8.19 and 15.30 % during calibration, and 16.72, 14.94 and 15.93 % during validation for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The nRMSE value was lowest for SMLR-SVM (11.88 and 19.40 %), followed by SMLR-ANN (18.53 and 19.52 %) and SMLR-RF (18.19 and 21.23 %) during calibration, and validation, respectively. The RPD values were 2.30, 1.47 and 1.50 during calibration, and 1.33, 1.30 and 1.19 during validation, for SMLR-RF, SMLR-SVM and SMLR-ANN, respectively. Based on model accuracy parameters, RMSE, nRMSE and RPD values, among all three developed models, SMLR-

SVM performed better, followed by SMLR-RF and SMLR-ANN for mustard yield prediction for Zone IV of Rajasthan.

Table 4.15 Performance of the mustard yield prediction model developed by different techniques for Zone IV of Rajasthan.

Accuracy parameters	**SMLR-ANN**		**SMLR-SVM**		**SMLR-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	133.01	139.96	72.27	125.08	134.93	133.35
nMAE (%)	15.08	16.72	8.19	14.94	15.30	15.93
RMSE (kg ha^{-1})	160.39	181.85	104.81	162.41	163.47	165.92
nRMSE (%)	18.19	21.73	11.88	19.40	18.53	19.82
RPD	1.50	1.19	2.30	1.33	1.47	1.30
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	87.63	147.69	30.27	106.27	139.17	157.26
nMAE (%)	10.13	16.87	3.50	12.14	16.09	17.96
RMSE (kg ha^{-1})	119.86	177.81	43.64	162.10	178.12	200.00
nRMSE (%)	13.86	20.31	5.05	17.52	20.59	22.85
RPD	1.95	1.32	5.36	1.45	1.31	1.18

Model calibration and validation developed by variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM), and variable extraction by PCA and RF (PCA-RF) techniques in R statistical software version 3.1.3. for Zone IV of Rajasthan are presented in Fig. 4.25. Mustard yield prediction was done for Zone IV of Rajasthan using developed model PCA-ANN, PCA-SVM and PCA-RF and model performance was assessed using MAE, nMAE, RMSE, nRMSE and RPD values (table 4.15). During calibration, MAE values were 87.63, 30.37 and 139.17 kg ha^{-1}, respectively. During validation, MAE value was found to be highest for PCA-RF (157.26 kg ha^{-1}), followed by PCA-ANN (147.69 kg ha^{-1}), and PCA-SVM (106.27 kg ha^{-1}). The nMAE values were 10.13, 3.50 and 16.09 % during calibration, and 16.87, 12.14

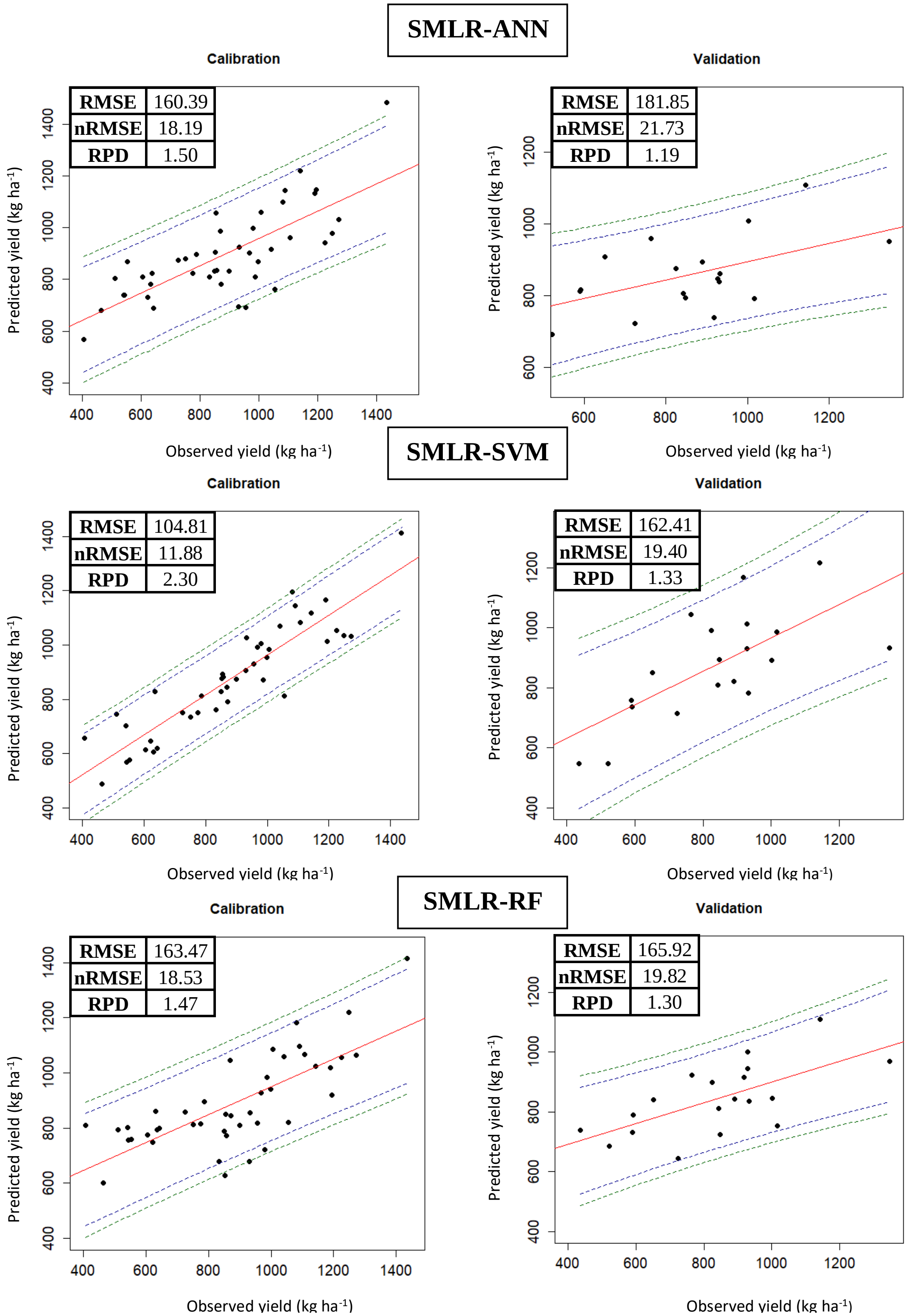

Fig 4.24 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for Zone IV of Rajasthan

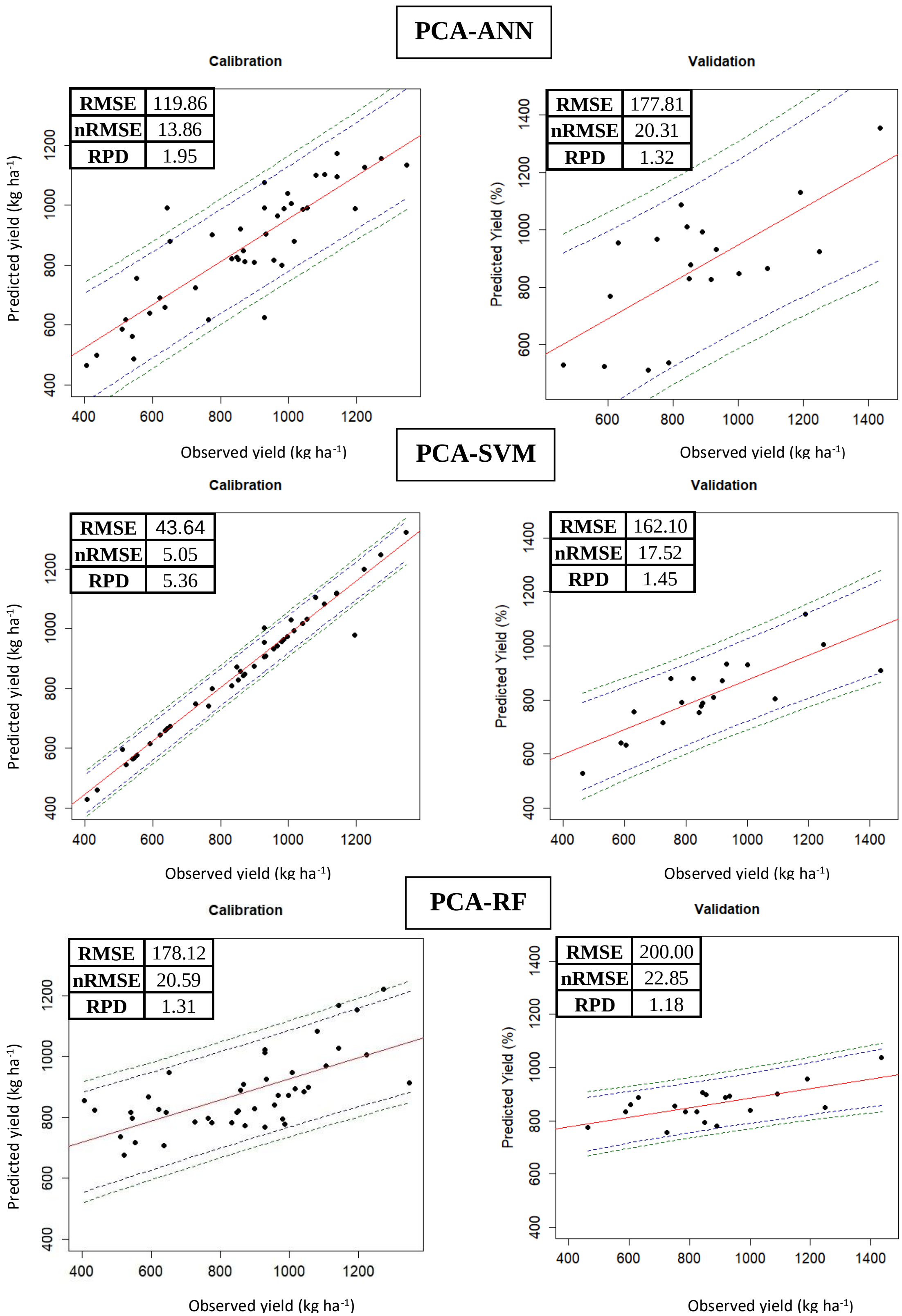

Fig 4.25 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for Zone IV of Rajasthan

and 17.96 % during validation, by PCA-ANN, PCA-SVM and PCA-RF, respectively. During calibration and validation, RMSE value was lowest for PCA-SVM (43.64 and 162.10 kg ha^{-1}), followed by PCA-ANN (119.86 and 177.81 kg ha^{-1}), and PCA-RF (178.12 and 200.00 kg ha^{-1}). The RPD values were 1.95, 5.36 and 1.31 during calibration, and 1.32, 1.45 and 1.18 during validation, for PCA-ANN, PCA-SVM and PCA-RF, respectively. The value of nRMSE was lowest during calibration and validation for PCA-SVM (5.05 and 17.52 %), followed by PCA-ANN (13.86 and 20.31%), and PCA-RF (20.59 and 22.85 %). Results showed on the basis model accuracy parameters, RMSE, nRMSE and RPD values, three model developed by variable extraction methods, PCA-SVM performed best followed by PCA-ANN and PCA-RF and among all six developed models, PCA-SVM performed better, followed by SMLR-SVM, SMLR-RF, PCA-ANN, SMLR-ANN and PCA-RF, respectively, for mustard yield prediction of Zone IV of Rajasthan .

4.3.2.5 Zone V (Bikaner)

The weather and yield data of mustard crop was collected for the last 22 years from the Bikaner district of Rajasthan. Data were processed and a model was developed for predicting mustard yield using variable selection by stepwise regression and principal component analysis, using ANN, SVM, RF techniques in R statistical software version 3.1.3..

Calibration and validation of the model developed using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques for Zone V are shown in Fig. 4.26. Mustard crop yield prediction for Zone V of Rajasthan was done by these developed models, SMLR-ANN, SMLR-SVM and SMLR-RF. Model performance for mustard yield prediction of the Zone V of Rajasthan done using developed models, SMLR-ANN, SMLR-SVM and SMLR-RF are given in Table 4.16. The performance of the models was analyzed using MAE, nMAE, RMSE, nRMSE and RPD values. The RMSE values during calibration were 73.29, 94.96 and 132.50 kg ha^{-1}, and during validation, were 138.80, 155.92 and 143.42 kg ha^{-1} for SMLR-SVM, SMLR-ANN, and SMLR-RF, respectively. The nRMSE value during calibration and validation was lowest for SMLR-SVM (7.96 and 13.99 %), followed by SMLR-ANN (10.32 and 15.72 %), and

SMLR-RF (14.20 and 14.42 %), respectively. The values of nRMSE during validation showed that all three models developed for mustard yield prediction for the Zone V of Rajasthan performed good having nRMSE<20 %. RPD values during calibration were 2.43, 3.14 and 1.74, and during validation, were 1.27, 1.52 and 1.38 for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The RPD values, which signify the model accuracy was high for SMLR-SVM compared with SMLR-ANN and SMLR-RF. The MAE value was lowest during calibration for SMLR-SVM (57.9 kg/ha^{-1}), followed by SMLR-ANN (78.44 kg/ha^{-1}), and SMLR-RF (108.3 kg/ha^{-1}).

Table 4.16 Performance of the mustard yield prediction model developed by different techniques for Zone V of Rajasthan.

Accuracy parameters	SMLR-ANN		SMLR-SVM		SMLR-RF	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	78.44	143.61	57.91	123.07	108.34	121.39
nMAE (%)	8.52	14.48	6.29	12.40	11.77	12.23
RMSE (kg ha^{-1})	94.96	155.92	73.29	138.80	132.50	143.02
nRMSE (%)	10.32	15.72	7.96	13.99	14.20	14.42
RPD	2.43	1.27	3.14	1.52	1.74	1.38
	PCA-ANN		**PCA-SVM**		**PCA-RF**	
	Calibration	**Validation**	**Calibration**	**Validation**	**Calibration**	**Validation**
MAE (kg ha^{-1})	32.76	112.17	31.75	87.24	114.62	172.99
nMAE (%)	3.58	11.15	3.48	8.67	12.54	17.19
RMSE (kg ha^{-1})	41.12	144.84	48.76	118.26	138.99	189.22
nRMSE (%)	4.50	14.39	5.34	11.75	15.21	18.80
RPD	5.26	1.54	4.44	1.89	1.56	1.18

During validation, MAE value was lowest for SMLR-SVM (123.1 kg/ha^{-1}), followed by SMLR-RF (121.4 kg/ha^{-1}), and SMLR-ANN (143.6 kg/ha^{-1}). The nMAE values during calibration were 8.52, 6.2 9 and 11.77 %, and during validation 14.48, 12.40 and 12.23 % for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. On the

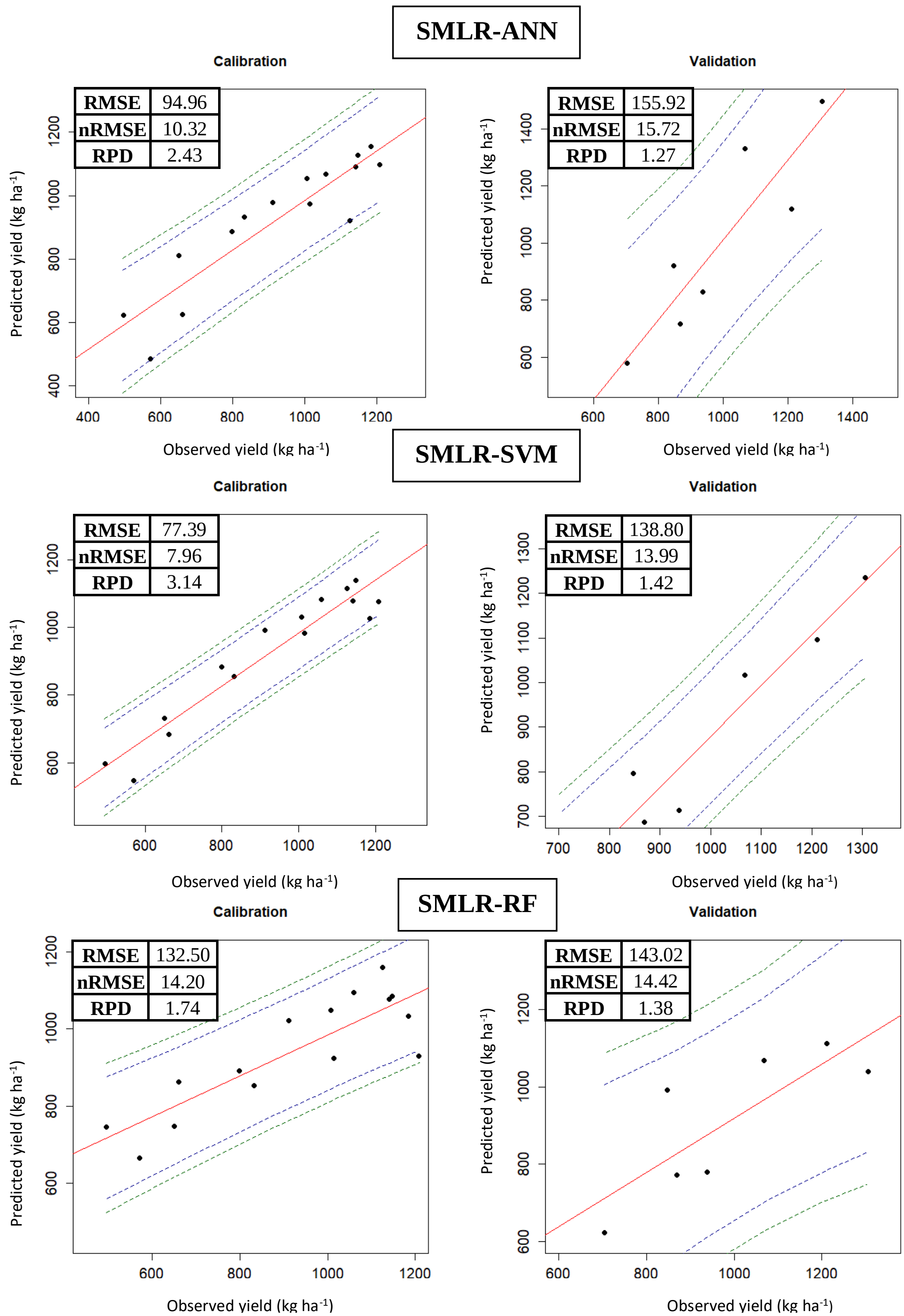

Fig 4.26 Mustard yield prediction by SMLR-ANN, SMLR-SVM and SMLR-RF for Zone V of Rajasthan

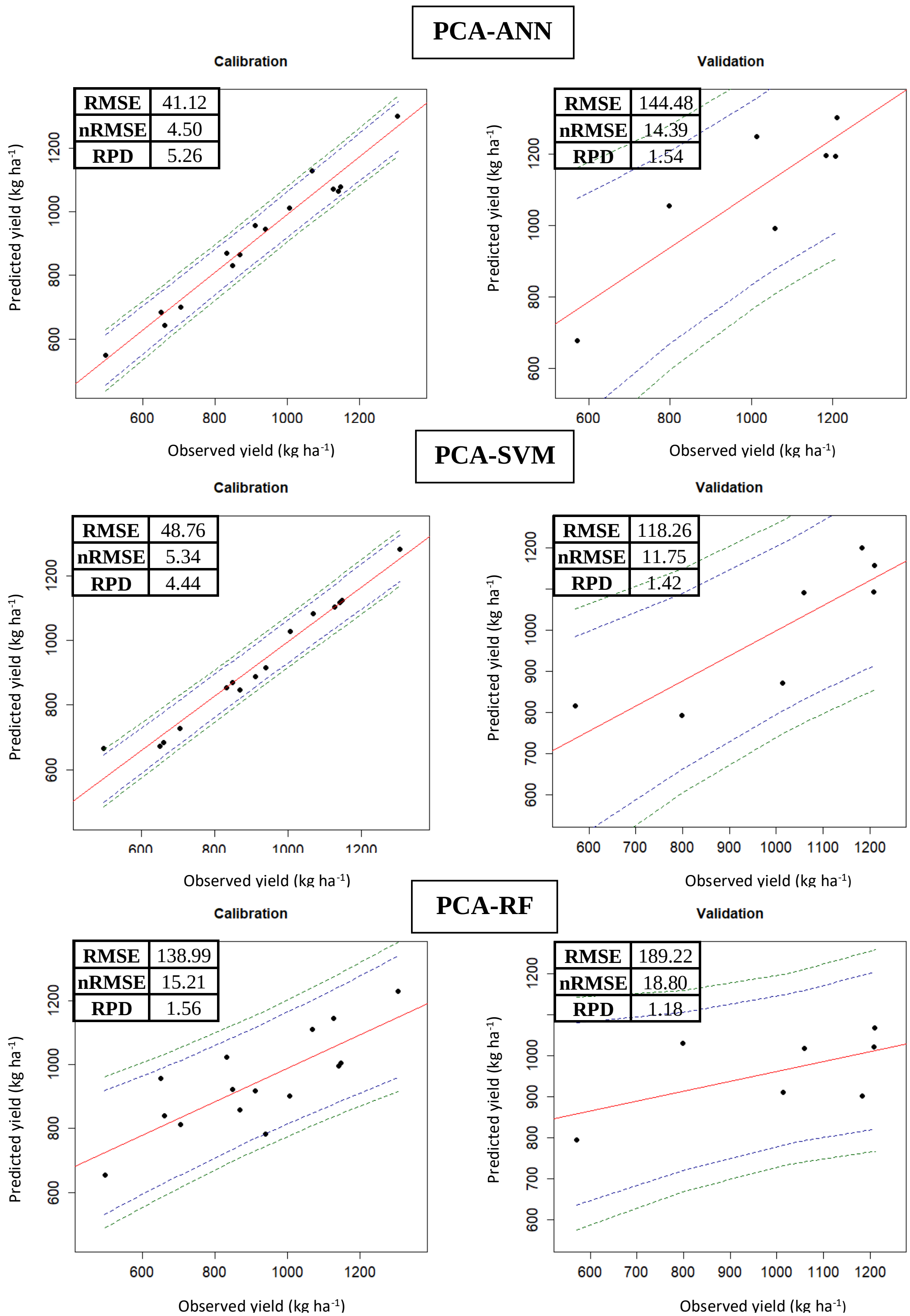

Fig 4.27 Mustard crop yield prediction by PCA-ANN, PCA-SVM and PCA-RF for Zone V of Rajasthan

basis of model accuracy parameters RMSE, nRMSE and RPD values, among the three models developed by the variable selection by SMLR, SMLR-SVM performed better, followed by SMlR-RF and SMlR-ANN for mustard yield prediction for Zone V of Rajasthan state.

Calibration and validation of model developed by variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM) and variable extraction by PCA and RF (PCA-RF) techniques in R statistical software version 3.1.3. for the Zone V of Rajasthan are presented in Fig. 4.27. Mustard yield prediction was done for the Zone V of Rajasthan using developed model PCA-ANN, PCA-SVM and PCA-RF. Model performance assessed using MAE, nMAE, RMSE, nRMSE and RPD values are shown in table 4.16. The MAE value was highest during calibration for PCA-RF (114.62 kg ha^{-1}), followed by PCA-ANN (32.76 kg ha^{-1}), and PCA-SVM (31.75 kg ha^{-1}). During validation MAE value was highest for PCA-RF (121.39 kg ha^{-1}), followed by PCA-ANN (112.17 kg ha^{-1}), and PCA-SVM (87.24 kg ha^{-1}). The nRMSE values during calibration were 3.58, 3.48 and 12.54 %, and 11.15, 8.67 and 17.19 % for SMLR-ANN, SMLR-SVM and SMLR-RF, respectively. The RMSE values during calibration were 48.76, 41.12 and 138.99 kg ha^{-1}, and during validation, were 118.26, 144.84 and 189.22 kg ha^{-1} for PCA-SVM, PCA-ANN and PCA-RF, respectively. The nRMSE had the lowest value during calibration and validation for PCA-SVM (5.34 and 11.75 %), followed by PCA-ANN (4.50 and 14.39 %) and PCA-RF (15.21 and 18.80 %), respectively. The RPD values were 5.26, 4.44 and 1.56 during calibration, and 1.54, 1.89, and 1.18 during validation, for model developed by PCA-ANN, PCA-SVM and PCA-RF techniques. A model is reliable and robust for low value of MAE, RMSE, nRMSE and a high value of RPD. Based on model accuracy parameters, RMSE, nRMSE and RPD values, among all the six models developed for mustard yield prediction for the Zone V, PCA-SVM performed best, followed by SMLR-SVM, PCA-ANN, SMLR-RF, SMLR-ANN and PCA-RF. Results illustrate that models developed either by variable selection or variable extraction, SVM techniques performed best among all developed models.

4.4 Mustard yield prediction by an optimal combination of developed models

An optimal combination of models was used for predicting mustard yield. It chooses weights to minimize the expected errors of the combined prediction. There were four combinations, ANN+SVM+RF, ANN+SVM, ANN+RF and SVM+RF used to combine the predicted results for Delhi and the different Zones of Rajasthan. The accuracy of the prediction was improved by combining the results of different methods. The model accuracy parameters such as coefficient of determination (R^2), MAE, RMSE, and nRMSE were used in this study for model validation, as described in the following sub-sections.

4.4.1 Mustard yield prediction by an optimal combination for IARI, New Delhi

The results obtained from an optimal combination techniques used for the model developed by PCA-ANN, PCA-SVM and PCA-RF techniques are presented in Fig. 4.28. The RMSE value for mustard yield prediction by different optimal combination models for the IARI, New Delhi was lowest 179.9 kg ha^{-1} for PCA(ANN+SVM), followed by 203.1 kg ha^{-1} for PCA(ANN+SVM+RF), and 216.6 kg ha^{-1} for PCA(SVM+RF) and 230.72 kg ha^{-1}for PCA(ANN+RF). The nRMSE had lowest value 8.75 % for PCA (ANN+SVM), followed by 9.88 % for PCA(ANN+SVM+RF), 10.55 % for PCA(SVM+R) and 11.23 % for PCA(ANN+RF). Optimal combination techniques used for model developed by PCA(ANN+SVM) and PCA(ANN+SVM+RF) performed excellent with nRMSE less than 10 % and model developed by PCA(SVM+RF) and PCA(ANN+RF) performed good with nRMSE value 10.55 and 12 % for mustard yield prediction of IARI, New Delhi. The values of MAE were 131.8, 171.9, 185.5 and 200.9 kg ha^{-1} for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF), respectively. The coefficient of determination was highest for PCA(ANN+SVM) with R^2 value 0.81, followed by PCA(ANN+SVM+RF) with R^2 value 0.75 and PCA(SVM+RF) with R^2 value 0.74 and PCA(ANN+RF) with R^2 value 0.70.

Optimal combination techniques were used for model developed by SMLR-ANN, SMLR-SVM and SMLR-RF techniques are presented in Fig. 4.29. Optimal combination techniques were used to predict the mustard yield for IARI, New Delhi by combination of SMLR-ANN, SMLR-SVM and SMLR-RF. Mustard yield prediction done by all four combinations had MAE value 151.9 kg ha^{-1}lowest for SMLR(SVM+RF),

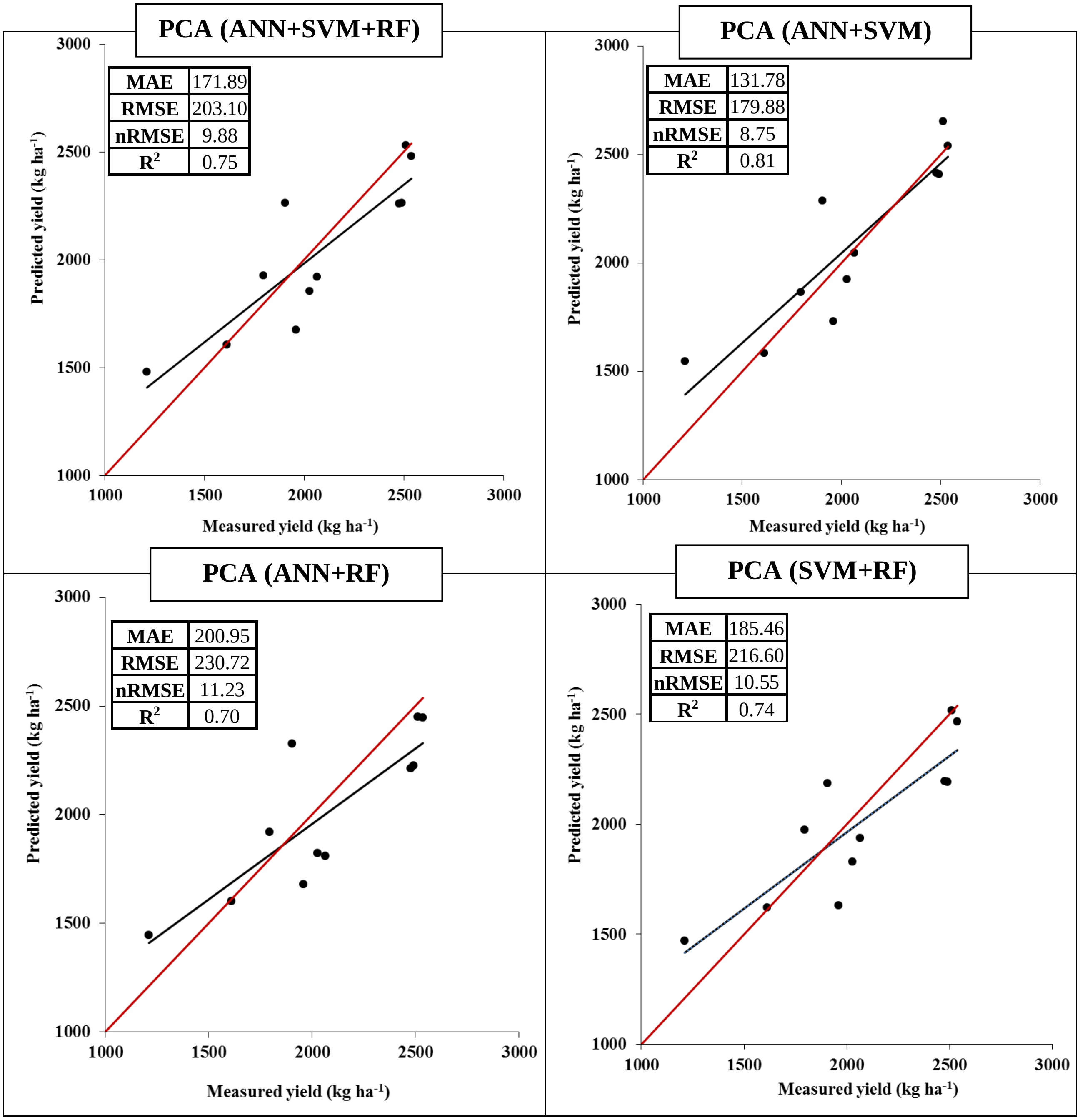

Fig 4.28 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for IARI, New Delhi

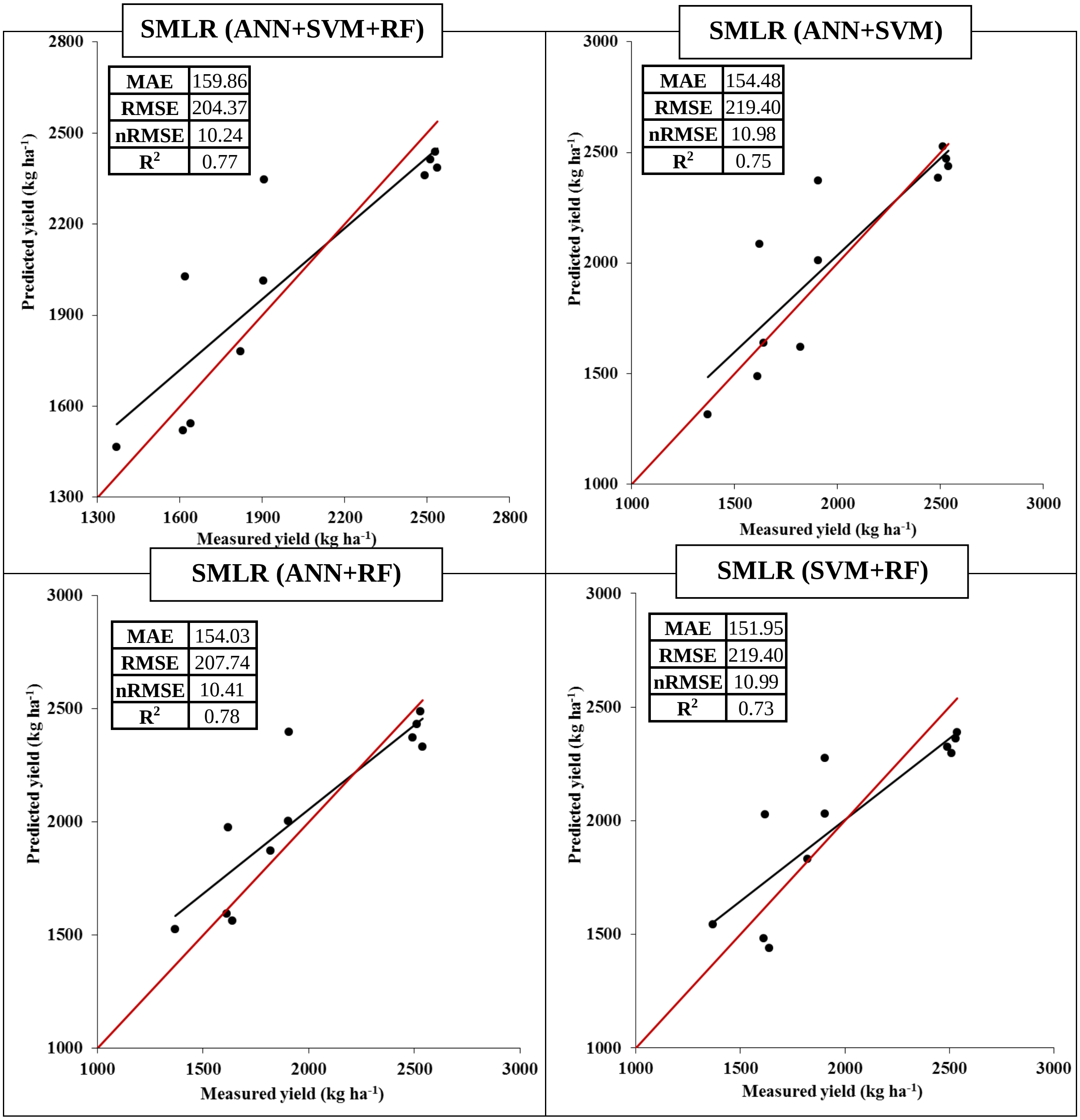

Fig 4.29 Mustard yield prediction using optimal combination of SMLR-ANN, SMLRSVM and SMLR-RF for IARI, New Delhi

followed by 154.0 kg ha^{-1} for SMLR(ANN+RF), 154.8 kg ha^{-1} for SMLR(ANN+SVM), and 159.86 kg ha^{-1} for SMLR(ANN+SVM+RF). The value of RMSE and nRMSE was 204.4 kg ha^{-1} and 10.2 % lowest for SMLR(ANN+SVM+RF), followed by 207.7 kg ha^{-1} and 10.41% for SMLR(ANN+RF), 219.0 kg ha^{-1} and 10.94 % for SMLR(ANN+SVM) and 219.4 kg ha^{-1} and 11.0 % for SMLR(SVM+RF), respectively. Model developed by all four combinations performed good with nRMSE value less than 11 %. Coefficient of determinant was highest for SMLR(ANN+RF) with R^2 value 0.78, followed by SMLR(ANN+SVM+RF) with R^2 value 0.77, SMLR(ANN+SVM) with R^2 value 0.75 and SMLR(SVM+RF) with R^2 value 0.73.

On the basis of model accuracy parameters R^2 MAE, RMSE, nRMSE values, among all the eight optimal combinations used for mustard yield prediction for the IARI, New Delhi, PCA(ANN+SVM) performed best followed by PCA(ANN+SVM+RF), SMLR(ANN+SVM+RF), SMLR(ANN+RF), PCA(SVM+RF), SMLR(ANN+SVM), SMLR(ANN+RF), PCA(ANN+RF). All the combinations of different techniques showed overestimation for low yield and underestimation for high yield, whether based on variable selection by stepwise regression or variable extraction by principal component analysis.

4.4.2 Yield prediction by an optimal combination for Zone I of Rajasthan

Alwar and Bharatpur districts of Rajasthan come under the same agro-climatic Zone I. The results obtained from optimal combination techniques used for model developed by PCA-ANN, PCA-SVM and PCA-RF techniques are presented in Fig. 4.30. There were four combinations. Mustard yield prediction for Zone I of Rajasthan was done using these optimal combinations. The performance of optimal combination prediction models were analyzed using RMSE, nRMSE, MAE and R^2 value. The RMSE value for yield prediction were 145.7 kg ha^{-1}for PCA(ANN+SVM), 154.9 kg ha^{-1} for PCA(ANN+SVM+RF), 162.5 kg ha^{-1} for PCA(SVM+RF) and 168.6 kg ha^{-1} for PCA(ANN+RF) combination. The nRMSE value was 11.75 % lowest for PCA(ANN+SVM), followed by 12.05 % for PCA(ANN+SVM+RF), 13.11 % for PCA(SVM+RF) and 13.60 % for PCA(ANN+RF). The values of coefficient of determinant R^2 were 0.73, 0.63, 0.59 and 0.58 for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF) combinations, respectively. The MAE value was 116.57 kgha^{-1} lowest for ANN+SVM, followed by 128.61 kgha^{-1} for PCA(ANN+SVM+RF), 136.48 kgha^{-1} for PCA(ANN+RF) and 139.40 kgha^{-1} for PCA(SVM+RF) combination, respectively. On the basis of model accuracy parameters R^2, RMSE, nRMSE values, among the four combination, PCA(ANN+SVM) performed best followed by PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF).

The results obtained from optimal combination techniques used for model developed by SMLR-ANN, SMLR-SVM and SMLR-RF techniques are presented in Fig. 4.31. The RMSE values of mustard yield prediction done for Zone I of Rajasthan by the optimal combinations SMLR(ANN+SVM), SMLR(ANN+SVM+RF), SMLR(SVM+RF) and SMLR(ANN+RF) were 205.9, 145.8, 139.9 and 143.4 kg ha^{-1}, respectively. nRMSE value was 15.97 % lowest for SMLR(SVM+RF), followed by 16.6 % for SMLR(ANN+SVM+RF), 17.0 % for SMLR (ANN+RF0) and 17.33 % for SMLR(ANN+SVM), respectively. The MAE values was 139.9 kg ha^{-1} for SMLR(SVM+RF), 143.4 kg ha^{-1} for SMLR(ANN+RF), 145.8 kg ha^{-1} for SMLR(ANN+SVM+RF) and 158.9 kg ha^{-1} for SMLR(ANN+SVM), respectively. R^2 was 0.65 for SMLR(SVM+RF), 0.62 for SMLR(ANN+SVM) and 0.60 for SMLR(ANN+ SVM) and SMLR(ANN+RF). A high value of R^2 and the low value of MAE, RMSE and nRMSE value indicates good agreement between observed and predicted values. The combination model based on variable extraction by PCA showed more overestimation for lower range of yield and less underestimation for higher range of yield. On the contrary, the combination model based on variable selection by SMLR showed less overestimation at a lower range of yield and more underestimation at a higher range of yield. Among the four combinations based on variable selection, SMLR(ANN+SVM) performed best followed by SMLR(ANN+SVM+RF), SMLR(SVM+RF) and SMLR(ANN+RF).

On the basis of model accuracy parameters RMSE, nRMSE, and R^2, among all the models developed either by variable extraction or variable selection for mustard yield prediction of Zone I of Rajasthan, the combination of PCA(ANN+SVM) performed best, followed by PCA(ANN+SVM+RF), PCA(SVM+RF), PCA(ANN+RF), SMLR(SVM+RF), SMLR(ANN+SVM+RF), SMLR(ANN+RF) and SMLR(ANN+SVM).

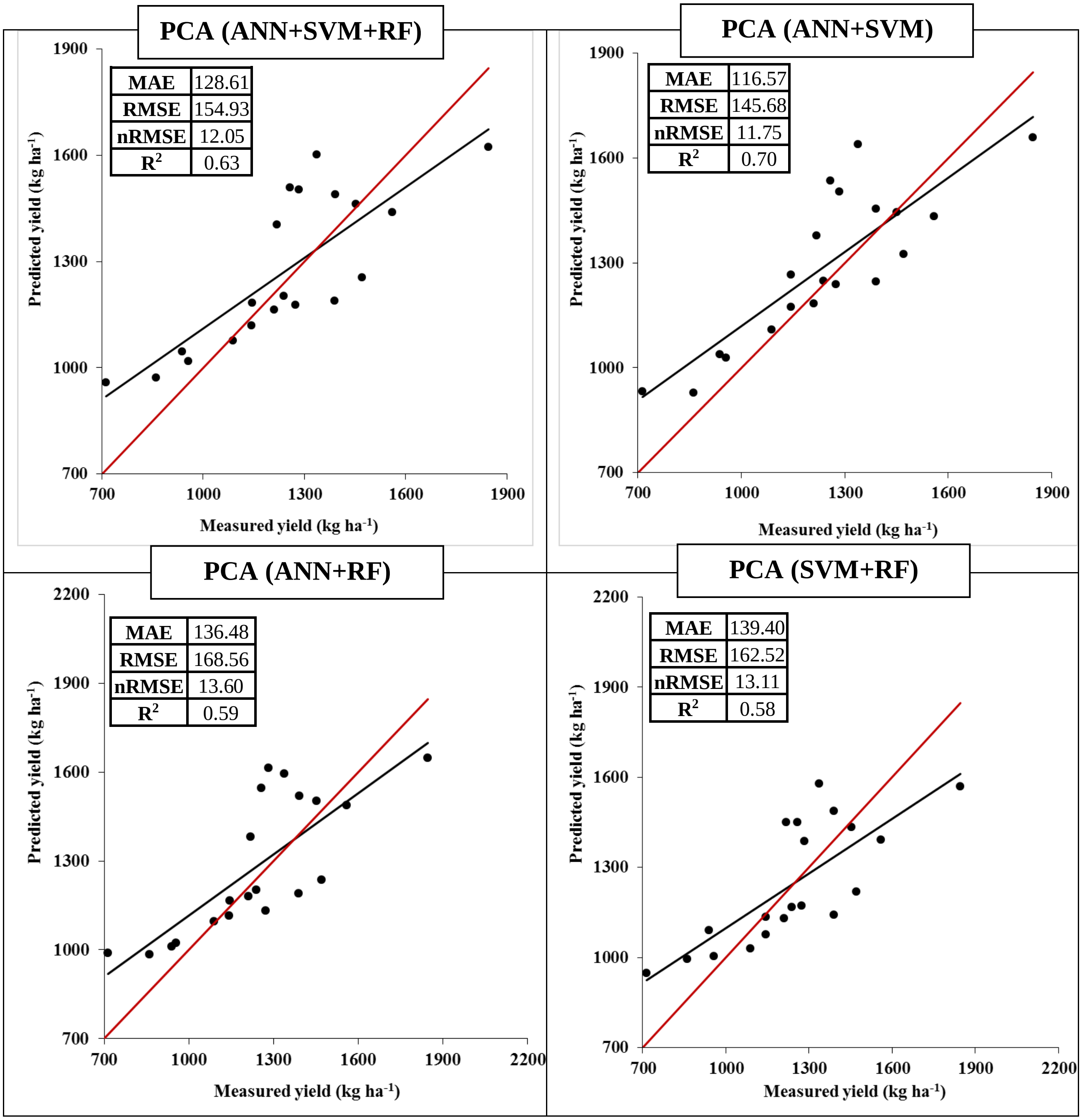

Fig 4.30 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for Zone I of Rajasthan

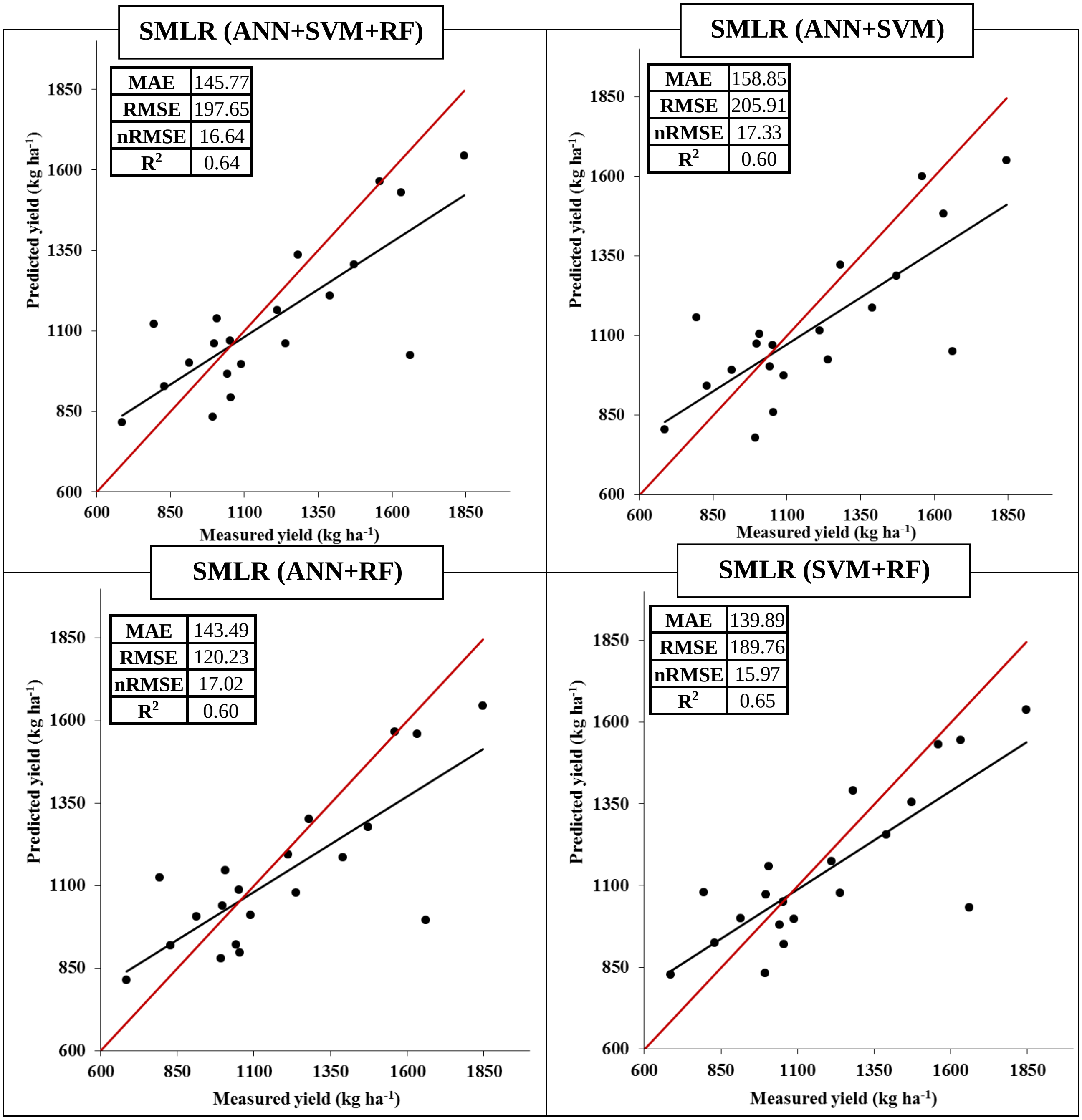

Fig 4.31 Mustard yield prediction using optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF for Zone I of Rajasthan

4.4.3 Mustard yield prediction by an optimal combination for Zone II of Rajasthan

According to the NARP report, Sawai madhopur and Kota districts of Rajasthan comes under the same agroclimatic zone (Zone II). The results obtained from optimal combination techniques used for the model developed by PCA-ANN, PCA-SVM and PCA-RF techniques are presented in Fig. 4.32. Mustard yield prediction for Zone II of Rajasthan was done by the optimal combination of PCA-ANN, PCA-SVM and PCA-RF model. There were four optimum combination. MAE values were 125.8, 105.0, 18.6 and 170.1 kg ha^{-1}, respectively for optimal combination the PCA(ANN+SVM+RF), PCA(ANN+SVM), PCA(ANN+RF) and PCA(SVM+RF) respectively. The RMSE value was 138.4 kg ha^{-1}lowest for PCA(ANN+SVM), followed by 151.9 kg ha^{-1} for PCA(ANN+SVM+RF), 170.1 kg ha^{-1} for PCA(SVM+RF) and for 159.86 kg ha^{-1} PCA(ANN+RF), respectively. nRMSE value ranged between 11.8 to 20.4 %. The value of nRMSE was 11.85 % lowest for the model developed using PCA(ANN+SVM) combination, followed by 13.32 % for PCA(ANN+SVM+RF), and 15.58 % for PCA(SVM+RF), and 20.41 % for PCA(ANN+RF). Performace of mustard yield prediction for Zone II of Rajasthan done by optimal combinations were good for PCA(ANN+SVM), PCA(ANN+SVM+RF) and PCA(SVM+RF) having nRMSE value less than 20% and fair for PCA(ANN+RF) having nRMSE value 20.41 %. The R^2 values were 0.83, 0.82, 0.81 and 0.80 for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF), respectively. On the basis of model accuracy parameters RMSE, nRMSE and R^2 values, among the four combinations based on variable extraction, PCA(ANN+SVM) performed better followed by PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF) for mustard yield prediction of the Zone II of Rajasthan.

The performance of model obtained from optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF are presented in Fig. 4.33. Mustard yield prediction for Zone II of Rajasthan was done by the optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF model. The MAE values for optimal combinations SMLR(ANN+RF), SMLR(ANN+SVM+RF), SMLR(ANN+SVM) and SMLR(SVM+RF) were 125.6, 126.1, 130.9 and 141.6 kg ha^{-1}, respectively. RMSE value was 167.5 kg ha^{-1} lowest for SMLR(ANN+SVM), followed by 176.5 kg ha^{-1} for SMLR(ANN+SVM+RF), 181.2 kg ha^{-1} for SMLR(ANN+RF) and 185.1 kg ha^{-1} for SMLR(SVM+RF), respectively. The model developed by optimal combination performed good for all the combinations having nRMSE value 15.33, 16.16, 16.59 and 16.95 % for SMLR(ANN+SVM), SMLR(ANN+SVM+RF), SMLR(ANN+RF) and SMLR (SVM+RF) respectively. R^2 value was 0.83, 0.82, 0.81 and 0.80 for SMLR(ANN+SVM), SMLR(ANN+SVM +RF), SMLR(SVM+RF) and SMLR(ANN+RF) respectively. On the basis of model accuracy parameters RMSE, nRMSE and R^2 values, among all the combinations, PCA (ANN+SVM) performed best for mustard yield prediction for Zone II of Rajasthan, followed by PCA(ANN+SVM+RF), SMLR(SVM+ANN), PCA(SVM+RF), SMLR(ANN+SVM+RF), SMLR(ANN+RF), SMLR(SVM+RF) and PCA(ANN+RF).

4.4.4 Mustard yield prediction by an optimal combination for Zone III of Rajasthan

Udaipur and Jhalawar have the same soil type and climatic normal, according to the NARP report, put together in Zone III of Rajasthan state. The results obtained for mustard yield prediction done for Zone III of Rajasthan state from optimal combination of PCA- ANN, PCA-SVM and PCA-RF are presented in Fig. 4.34. The MAE values were 103.1, 105.3, 111.4 and 118.2 kg ha^{-1} for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF), respectively. The value of RMSE was 138.9 kg ha^{-1} lowest for PCA(ANN+SVM), followed by 141.1 kg ha^{-1} for PCA(ANN+SVM+RF), 150.9 kg ha^{-1}for PCA(ANN+RF) and 15.7 kg ha^{-1} for PCA(SVM+RF) combination. The R^2 values were 0.93, 0.92, 0.91 and 0.90 for PCA(ANN+SVM+RF), PCA(ANN+SVM), PCA(ANN+RF) and PCA(SVM+RF), respectively. The nRMSE values were 14.8, 15.0, 16.1 and 16.24 % for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF), respectively. The nRMSE values were less than 20 % for all the four combinations indicates that performance of mustard yield prediction for Zone III of the Rajasthan were good by all the combinations.

The performance of the mustard yield prediction done by the optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF are presented in Fig. 4.35. The value of MAE was 106.3 kg ha^{-1} lowest for SMLR(ANN+SVM), followed by 111.1 kg ha^{-1} for SMLR(ANN+SVM+RF), 115.5 kg ha^{-1} for SMLR(SVM+RF) and 117.9 kg ha^{-1} for SMLR(ANN+RF). The value of RMSE was 143.5 kg ha^{-1} lowest for SMLR(SVM+R),

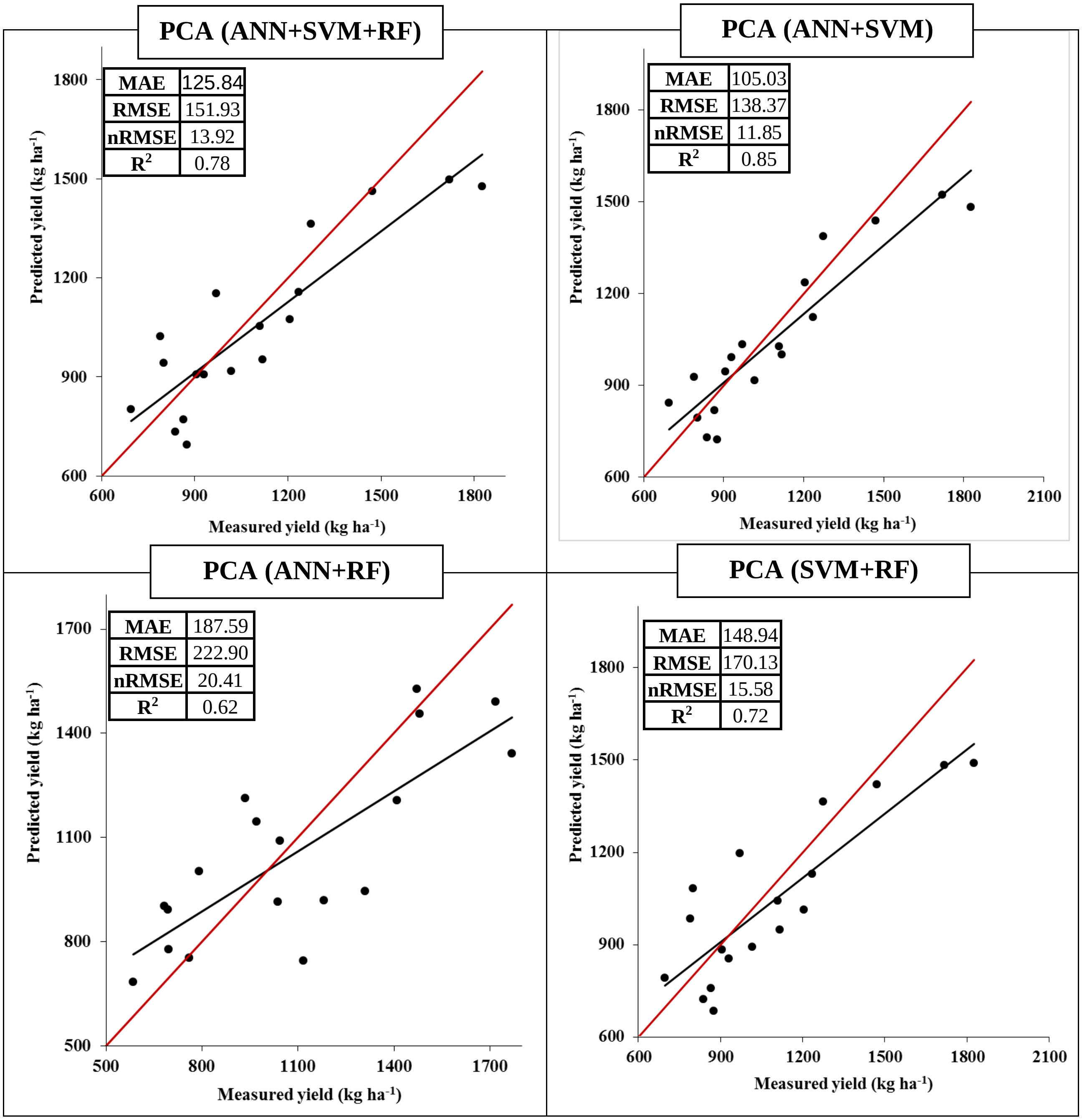

Fig 4.32 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for Zone II of Rajasthan

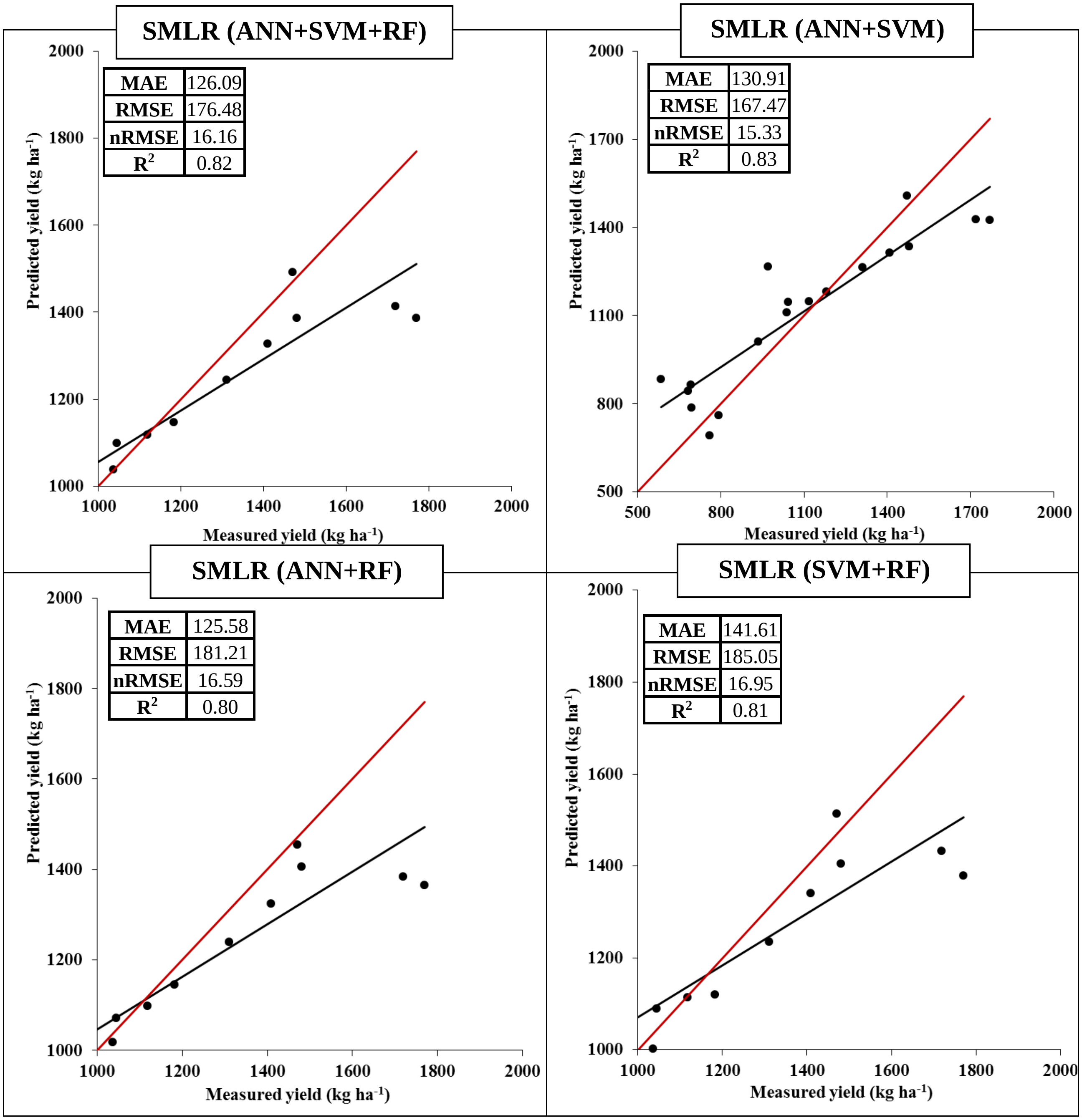

Fig 4.33 Mustard yield prediction using optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF for Zone II of Rajasthan

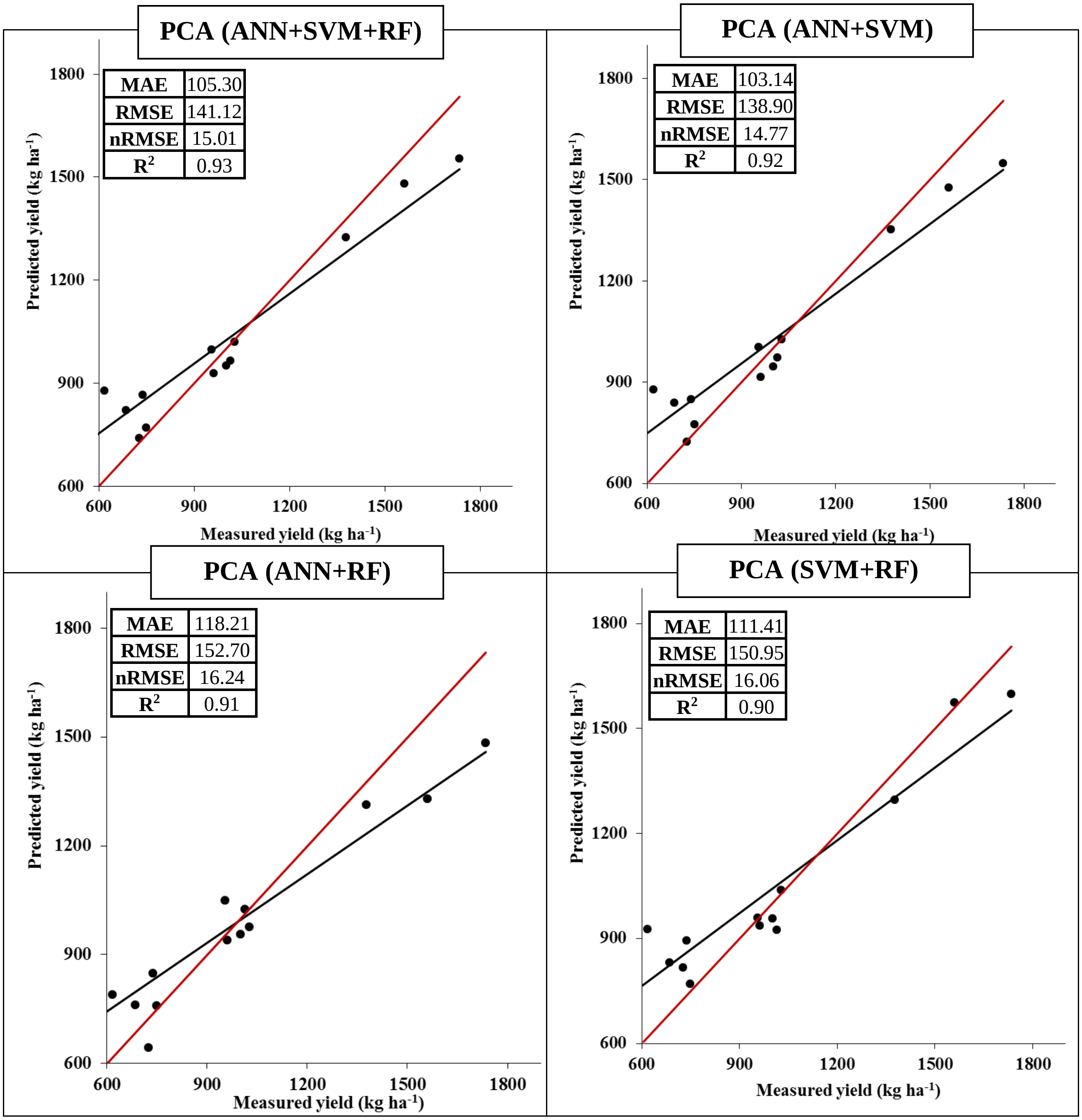

Fig 4.34 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for Zone III of Rajasthan

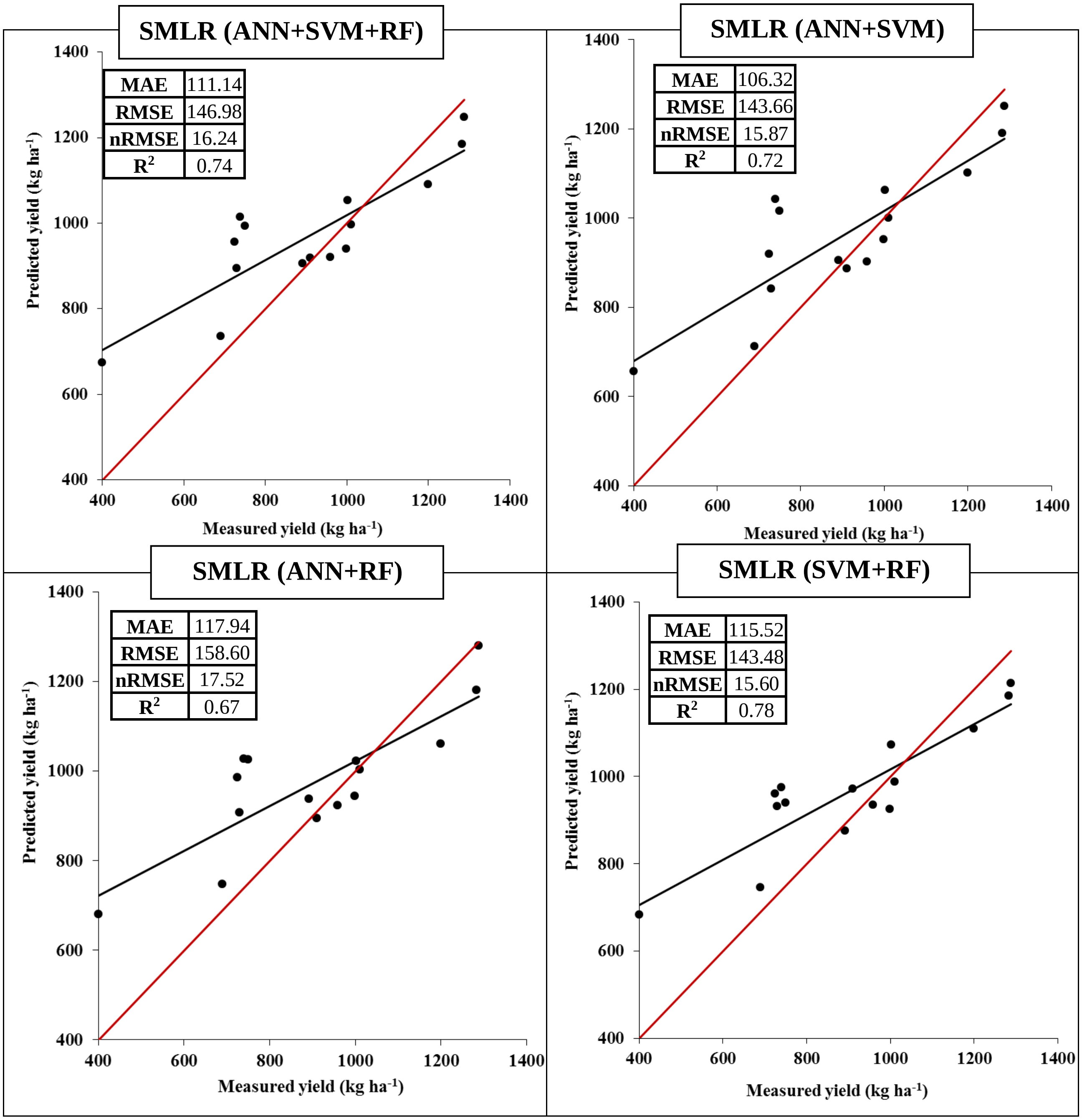

Fig 4.35 Mustard yield prediction using optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF for Zone III of Rajasthan

followed by 143.7 kg ha^{-1} for SMLR(ANN+SVM), for 147.0 kg ha^{-1} SMLR(ANN+SVM+RF) and 158.6 kg ha^{-1} for SMLR(ANN+RF), respectively. R^2 values were 0.78, 0.74, 0.72 and 0.67 for SMLR(SVM+RF), SMLR(ANN+SVM+RF), SMLR(ANN+SVM) and SMLR(ANN+RF), respectively. The nRMSE had lowest value 15.6 % for SMLR(SVM+RF), followed by 15.87 % for SMLR(ANN+SVM), 16.24 % for SMLR(ANN+SVM+RF) and 17.5 % for SMLR(ANN+RF) respectively. The mustard yield prediction done by all the four combination performed good having nRMSE value less than 10 %. Based on model accuracy parameter RMSE, nRMSE and R^2 values among all three combinations SMLR(SVM+RF) performed better followed by SMLR(ANN+SVM), SMLR(ANN+SVM+RF) and SMLR(ANN+RF).

Among all the models developed either by variable extraction or variable selection, on the basis of model accuracy parameter RMSE, nRMSE, R^2 and MSE, the combination of PCA(ANN+SVM) performed best for mustard yield prediction for the Zone III of Rajasthan followed by PCA(ANN+SVM+RF), SMLR(SVM+RF), SMLR(ANN +SVM), PCA(SVM+RF), PCA(ANN+RF) and SMLR(ANN+SVM+RF), SMLR(ANN+RF). All the combinations overestimated for low yield and underestimated for high yield.

4.4.5 Mustard yield prediction by an optimal combination for Zone IV of Rajasthan

The climatic conditions of Pali and Jodhpur districts are similar as reported by the NARP, and they are put together in Zone IV of Rajasthan state. The results obtained from optimal combinations of PCA-ANN, PCA-SVM and PCA-RF are presented in Fig. 4.36. There were four combinations and mustard yield prediction for Zone IV of Rajasthan was done by these four optimal combinations. RMSE values was lowest 147.7 kg ha^{-1} for PCA(ANN+SVM), followed by 168.2 kg ha^{-1} for PCA(ANN+SVM+RF), 177.0 kg ha^{-1} for PCA(ANN+RF) and 183.5 kg ha^{-1} for PCA(SVM+RF) combination. The nRMSE value had lowest value 16.9 % for PCA(ANN+SVM), followed by 19.2 % for PCA(ANN+SVM+RF), 20.2 % for PCA(SVM+RF) and 21.0 % for PCA(ANN+RF) respectively. The values of MAE were 116.4, 133.2, 14.3 and 143.6 kg ha^{-1} for PCA(ANN+SVM), PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF), respectively. The mustard yield prediction done by optimal combination performed good for PCA(ANN+SVM) and PCA(ANN+SVM+RF) having nRMSE value less than 20 %

and performed fair for PCA(SVM+RF) and PCA(ANN+RF) having nRMSE value 20.2 and 21.0 %. R^2 values were 0.67, 0.65, 0.63 and 0.53 for PCA(SVM+RF), PCA(ANN+SVM+RF), PCA(ANN+SVM) and PCA(ANN+RF), respectively. Based on model accuracy parameters RMSE, nRMSE, R^2 and MAE values, among all four optimal combinations for the mustard yield prediction done for Zone IV of Rajasthan, PCA(ANN+SVM) performed better followed by PCA(ANN+SVM+RF), PCA(SVM+RF) and PCA(ANN+RF) combination.

Mustard yield prediction for Zone IV of Rajasthan was done by optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF. The performance of the mustard yield prediction done by all the four combinations for Zone IV of Rajasthan are presented in Fig 4.37. The MAE values were 114.1, 116.9, 125.5 and 131.2 kg ha^{-1} for SMLR(ANN+SVM), SMLR (ANN+SVM+RF), SMLR(SVM+RF) and SMLR(ANN+RF) combination, respectively. The RMSE values were 153.88 kg ha^{-1} for SMLR(SVM+RF), 157.4 kg ha^{-1} for SMLR(ANN+SVM+RF), 157.5 kg ha^{-1} for SMLR(ANN+SVM), and 169.1 kg ha^{-1} for SMLR(ANN+RF), respectively. R^2 values were 0.52 for SMLR(ANN+SVM), 0.51 for SMLR(ANN+SVM+RF) and SMLR(SVM+RF), 0.42 for SMLR(ANN+RF), respectively. The mustard yield prediction done for Zone IV of Rajasthan by optimal combination performed good for SMLR(SVM+RF), SMLR(ANN+SVM+RF) and SMLR(ANN+SVM) having nRMSE value 18.38 % for SMLR(SVM+RF) and 18.8 % for for SMLR(ANN+SVM+RF) and SMLR(ANN+SVM). Performance of SMLR(ANN+RF) was fair with nRMSE value 20.2 Based on model accuracy parameter RMSE, nRMSE, R^2 and MSE values, among all four combinations SMLR(SVM+RF) was better followed by SMLR (ANN+SVM+RF), SMLR(ANN+SVM) and SMLR(ANN+RF) combination.

Among all the model developed either by variable extraction or variable selection on the basis of model accuracy parameter RMSE, nRMSE and R^2 the combination of PCA(ANN+SVM) performed best for mustard yield prediction of the Zone IV of Rajasthan, followed by SMLR(SVM+RF), SMLR(ANN+SVM+RF), SMLR(ANN +SVM), PCA(ANN+SVM+RF), PCA(SVM+RF), SMLR(ANN+RF) and PCA(ANN+RF).

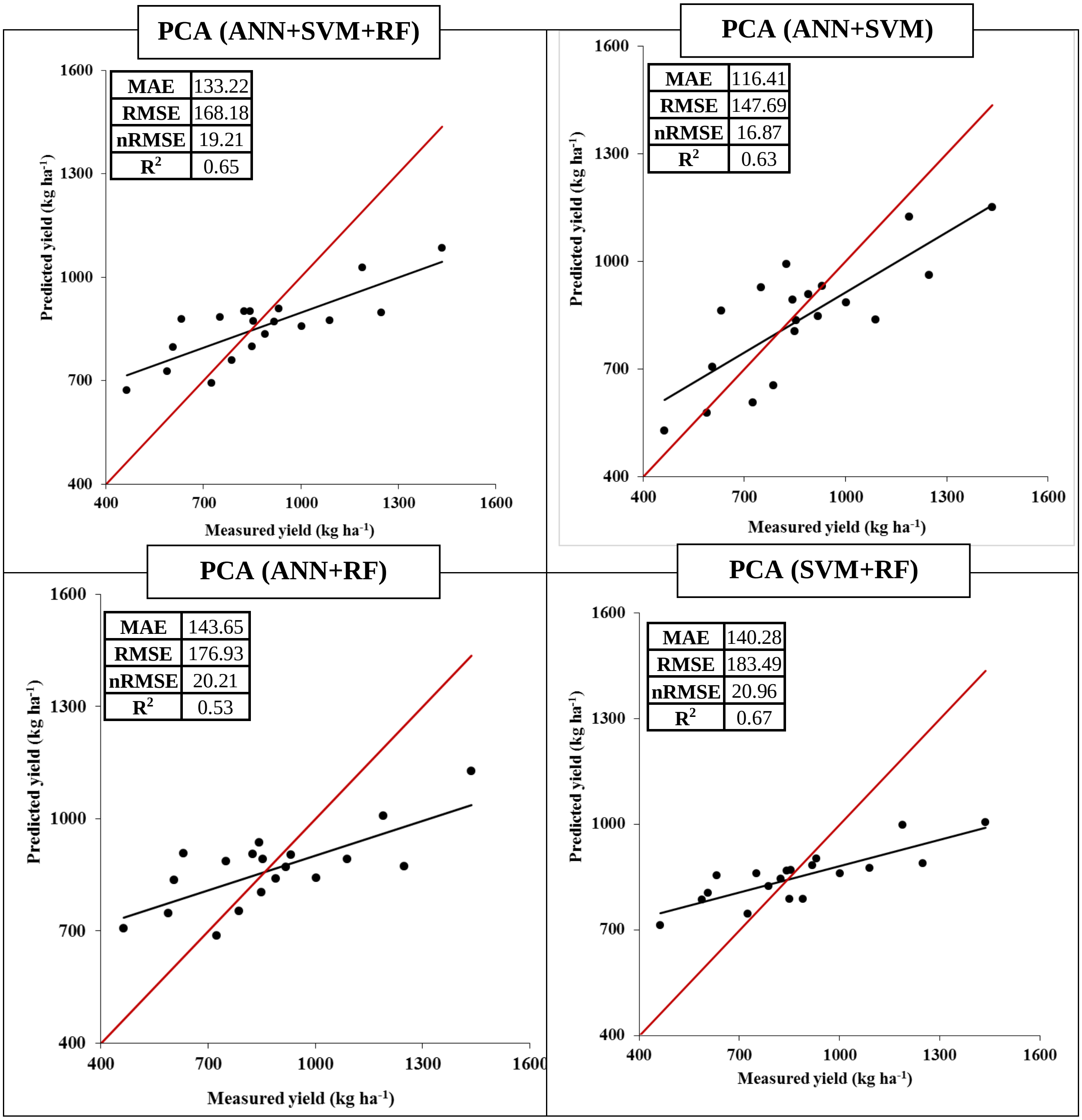

Fig 4.36 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for Zone IV of Rajasthan

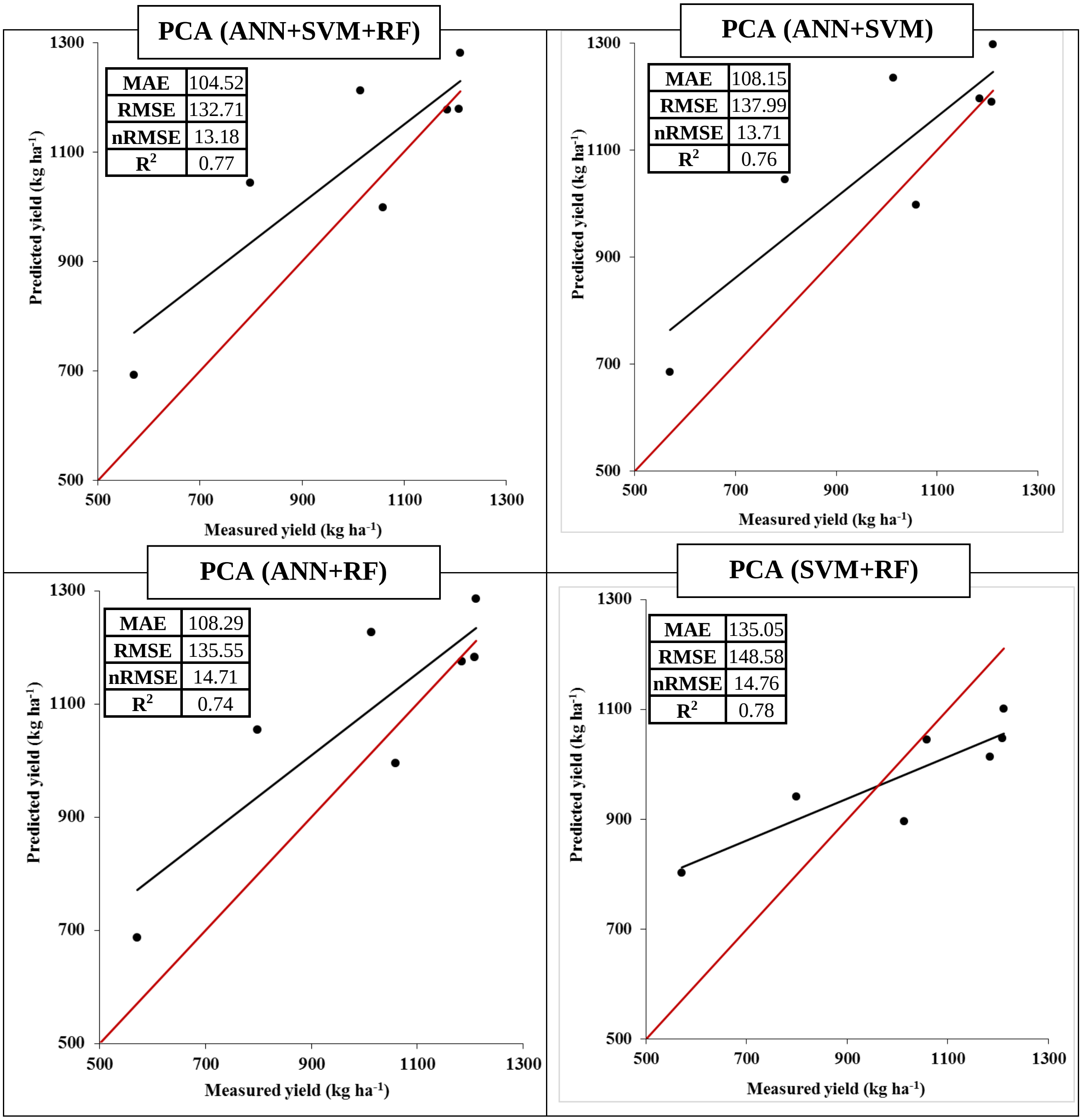

Fig 4.38 Mustard yield prediction using optimal combination of PCA-ANN, PCA-SVM and PCA-RF for Zone V of Rajasthan

4.4.6 Mustard yield prediction by an optimal combination for Zone V of Rajasthan:

According to the NARP report, the climatic condition of Bikaner district is located in Zone V of Rajasthan state. The results obtained from optimal combinations of PCA-ANN, PCA-SVM and PCA-RF are presented in Fig. 4.38. Mustard yield prediction for Zone V of Rajasthan was done by optimal combination of PCA-ANN, PCA-SVM and PCA-RF. The MAE value was lowest 104.5 kg ha^{-1} for PCA(ANN+SVM+RF), followed by 143.4 kg ha^{-1} for PCA(ANN+SVM), 145.8 kg ha^{-1} for PCA(ANN+RF) and 158.9 kg ha^{-1} for PCA(SVM+RF). The values of RMSE were 132.7, 135.6, 138.0 and 148.6 kg ha^{-1} for PCA(ANN+SVM+RF), PCA(ANN+RF), PCA(ANN+SVM) and PCA(SVM+RF), respectively. The value of nRMSE was lowest 13.2 % for PCA(ANN+SVM+RF) followed by 13.7 % for PCA(ANN+SVM), 14.8 % for PCA(ANN+RF) and PCA(SVM+RF). All the four combinations performed good with nRMSE value less than 15 %. Based on model accuracy parameter RMSE, nRMSE and R^2, among all four combinations, mustard yield prediction done for the Zone V of Rajasthan state was best for PCA(ANN+SVM+RF) combination, followed by PCA(ANN+SVM), PCA(ANN+RF) and PCA(SVM+RF) .

Mustard yield prediction for Zone V of Rajasthan was done by optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF. The performance of the mustard yield prediction done by all the four combinations for Zone V of Rajasthan are presented in Fig. 4.38.

The performance of the mustard yield prediction for Zone V of Rajasthan done by the four combinations developed by optimal combinations of SMLR-ANN, SMLR-SVM and SMLR-RF are presented in Fig. 4.39. Mustard yield prediction for Zone V of Rajasthan was done by optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF. The MAE values for SMLR(ANN+SVM+RF), SMLR(ANN+SVM), SMLR(ANN+RF) and SMLR(SVM+RF) was 105.9, 107.0, 110.3 and 116.4 kg ha^{-1}, respectively. The RMSE value was 111.3 kg ha^{-1}lowest for SMLR(ANN+SVM+RF), followed by 111.5 kg ha^{-1}for SMLR(ANN+RF), 120.8 kg ha^{-1} for SMLR(ANN+SVM) and 110.3 kg ha^{-1} for SMLR(SVM+RF), respectively. The value of nRMSE was 11.22 % lowest for SMLR(ANN+SVM+RF), followed by 11.24 % for SMLR(ANN+RF), 12.17 % for SMLR(ANN+SVM) and 13.05 % for SMLR(SVM+RF). Value of nRMSE was less than 15 % for all the four combination indicates that, all the four combinations performed good for mustard yield prediction of Zone V of Rajasthan. Based on model accuracy parameter RMSE, nRMSE and R^2, among the four combinations, mustard yield prediction done by SMLR(ANN+SVM+RF) was best followed by SMLR(ANN+RF), SMLR(ANN+SVM) and SMLR(SVM+RF).

Among all the models developed either by variable extraction or variable selection on the basis of model accuracy parameter RMSE, nRMSE, R^2 and MSE, the combination of SMLR(ANN+SVM+RF) performed best for mustard yield prediction of Zone V of Rajasthan followed by SMLR(ANN+RF), SMLR(ANN+SVM), SMLR(SVM+RF), PCA(ANN+SVM+RF), PCA(ANN+SVM), PCA(ANN+RF) and PCA(SVM+RF). All the combinations overestimated mustard yield for Zone V of Rajasthan.

4.5 Performance of different models developed for mustard yield prediction

The performance of different models and optimal combinations for mustard yield prediction was compared and based on the nRMSE value best model for mustard yield prediction for IARI, New Delhi and the five zones of Rajasthan were selected and presented in table 4.17. We have developed three models by variable selection using SMLR (SMLR-SVM, SMLR-ANN and SMLR-RF) and three models by variable extraction using PCA (PCA-SVM, PCA-ANN and PCA-RF). There were total eight optimal combinations, four combinations (SVM+ANN+RF, SVM+ANN, SVM+RF and ANN+RF) each for models developed by variable selection and models developed by variable extraction. Based on the nMRSE values best model developed individually or in optimal combination were selected for different study area. Results showed among the three models SMLR-ANN, SMLR-SVM, SMLR-RF developed by variable selection, SMLR-SVM performed best for mustard yield prediction done for all the six study areas. Similarly model developed by variable extraction by PCA, PCA-SVM performed best for mustard yield prediction done for all the six study areas. Among all six models developed either variable selection by SMLR or variable extraction by PCA for mustard yield prediction for study area, PCA-SVM performed best for all the study areas.

Results showed that performance of the model developed by variable selection by SMLR (SMLR-ANN, SMLR-SVM, and SMLR-RF) and optimal combination of SMLR-

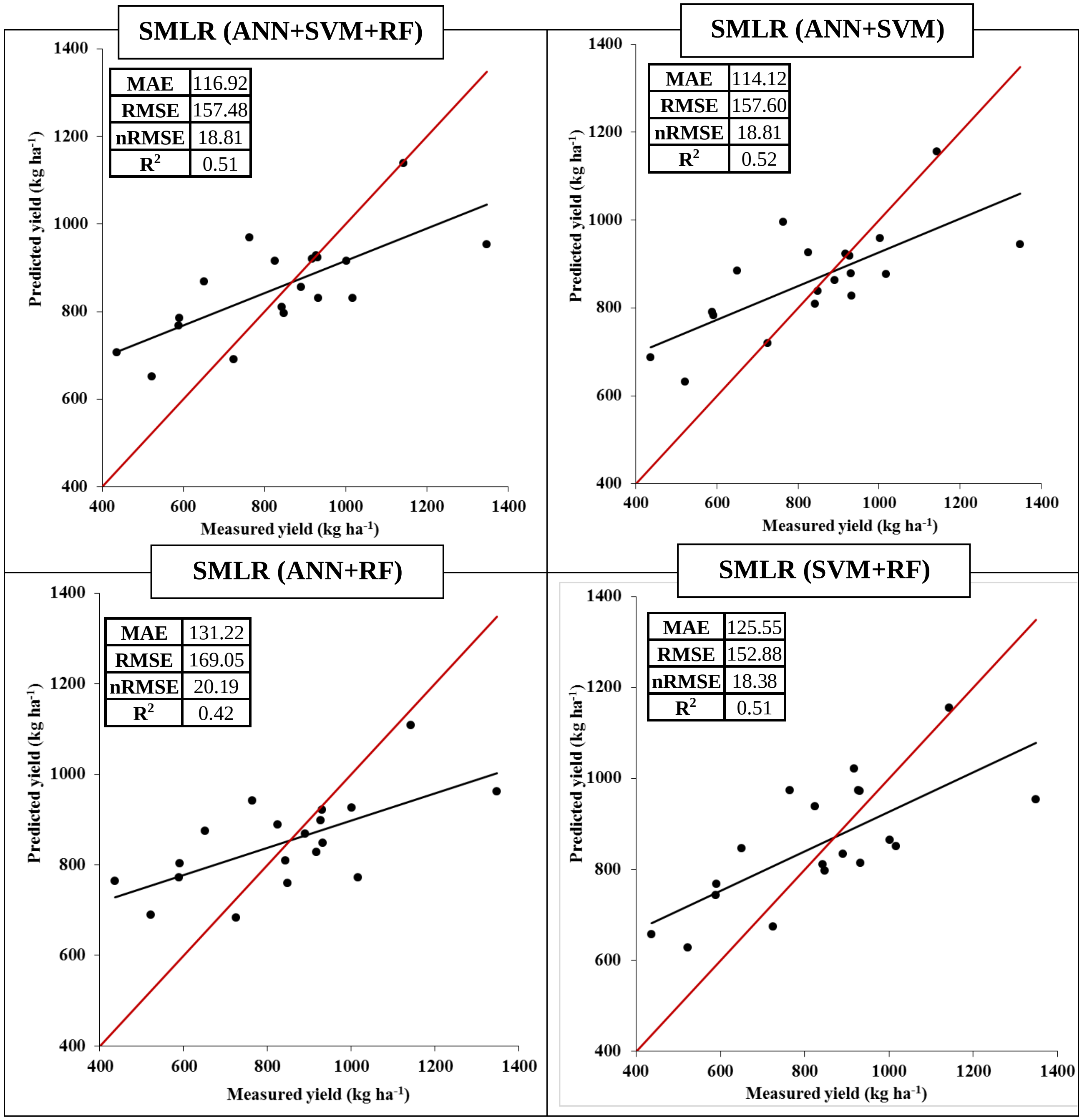

Fig 4.37 Mustard yield prediction using optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF for Zone IV of Rajasthan

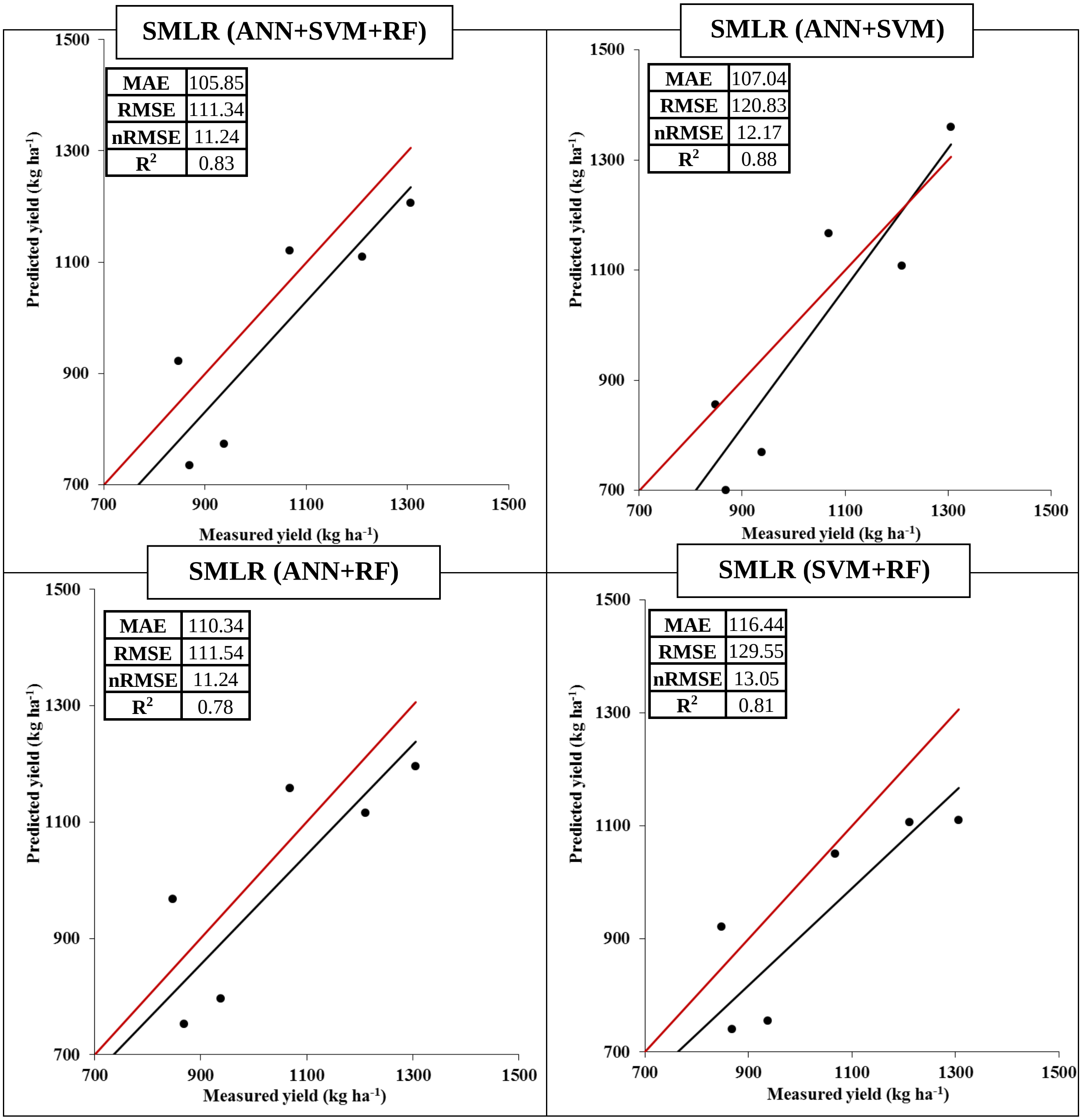

Fig 4.39 Mustard yield prediction using optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF for Zone V of Rajasthan

ANN, SMLR-SVM, and SMLR-RF, optimal combination performed better than the individual. Optimal combination of SMLR(ANN+SVM+RF) performed best for IARI, New Delhi and Zone V, optimal combination of SMLR(ANN +SVM) performed best for Zone II and optimal combination of SMLR(SVM +RF) performed best for Zone I, Zone III and Zone IV of study areas.

By comparing the performance of the model developed by variable extraction by PCA (PCA-ANN, PCA-SVM, and PCA-RF) and optimal combination of PCA-ANN, PCA-SVM, and PCA-RF, optimal combination performed better than the individual for all the study areas except for Zone III and Zone V individual model PCA-SVM performed better than optimal combination. Optimal combination of PCA(ANN+SVM) performed best for all the study areas except for Zone V optimal combination of PCA(ANN+SVM+RF) performed best for mustard yield prediction.

By comparing the performance of the model developed either by variable selection by SMLR individual and optimal combination or variable extraction by PCA individual and optimal combination, variable extraction by PCA performed better than variable selection by SMLR.

Table 4.17: Performance of models developed for mustard yield prediction for study areas

Study area	**Variable selection by SMLR**		**Variable extraction by PCA**	
	Individual	**Optimal Combination**	**Individual**	**Optimal Combination**
IARI, New Delhi	SVM (11.84)	ANN+SVM+RF (10.24)	SVM (9.12)	ANN+SVM (8.75)
Zone I	SVM (16.73)	SVM+RF (15.9)	SVM (12.65)	ANN+SVM (11.7)
Zone II	SVM (16.00)	ANN+SVM (15.3)	SVM (13.45)	ANN+SVM (11.8)
Zone III	SVM (14.42)	SVM+RF (15.6)	SVM (11.50)	ANN+SVM (14.8)
Zone IV	SVM (19.40)	SVM+RF (18.38)	SVM (17.52)	ANN+SVM (16.9)
Zone V	SVM (13.99)	ANN+SVM+RF (11.22)	SVM (11.75)	ANN+SVM+RF (13.2)

*Figures in the brackets showed the nRMSE value

5.Discussion

Mustard yield in different geographical areas are highly influece by the spatial weather variability. Weather is dynamic, continuous, and multi-dimensional, these weather properties make challenging for developing the crop yield prediction model. The response of crop yields to temperature variations may depend on the relative warming of minimum and maximum temperatures (Stone, 2001; Peng *et al*., 2004). It is a big challenge faced by any government to provide sufficient food supply to the people, especially in that areas with the continuing expansion of population and shrinkage in agricultural land. In India growth rate of mustard production and productivity was less during the past few decades. The productivity of mustard crops is highly vulnerable to weather variability during crop growing period. Due to weather variability during crop growing period in different years, there were variation in crop yield during different years. Developing a fine-scale crop yield predicting system under limited field data availability for a large area is the need of the hour. This study aimed to develop a prediction models for mustard crops. It is also obvious that the output from each component of the system is not free from their independent error. It is highly uncertain how these errors will propagate into the system output. Therefore, evaluating the output from such a system is very important and before that, the appropriate selection of each component is also imperative. It is an important task to do reliable crop yield prediction on farm, regional and national levels at a particular time (Bouman *et al*., 1997).

Keeping these above facts in view, the study was conducted, (a) to observe the impact of temperature stress on biophysical parameters, seed yield, radiation use efficiency and water productivity, (b) to calibrate and validate 'InfoCrop mustard model at the experimental field and farmer's field (c) multi-stage mustard yield prediction using InfoCrop at 50 % flowering and at pod formation stage, (d) to develop mustard yield prediction models by different techniques (ANN, SVM, and RF) for IARI, New Delhi and major mustard growing zones of Rajasthan, (e) optimal combination of different developed models to improve the accuracy of mustard yield prediction.

Results showed that timely sown crop had taken more time for total growing periods, followed by late and very late sown crops. There were more days required to attain 50 percent flowering with a delay in sowing. The prevailing high temperatures during the pod filling stage in very late sown crops probably hastened the maturity. The duration of the growing period under different environmental conditions is supported by the works done on mustard, wheat, soybean, and maize crops (Roy *et al*., 2005; Pradhan *et al*., 2014; Kar and Kumar, 2015; Umburanas *et. al*., 2019), under different agro-ecological regions.

The leaf area index is a meaningful parameter widely used for crop growth and development studies. It gives a better understanding of interpreting a crop's dry matter production with respect to the proper utilization of incoming solar radiation for photosynthesis. The LAI and fIPAR showed similar growth patterns during the crop-growing seasons. Goyal *et al*. (2018) confirm the logarithmic relationship between LAI and fIPAR. There are about 0.09 differences in the peak value of fIPAR and about 36 to 38 percent reduction in LAI under delay in sowing. The total interception PAR was higher in timely sown crop compared to late sown. It may be due to longer crop duration and higher LAI . Similar observation was found by Pradhan *et al*., 2018. The days required for attaining peak LAI occured early with delay in sowing, which may be due to the presence of low temperature, less amount of solar radiation throughout the vegetative stage and terminal heat stress during the reproductive stage, which causes faster leaf area development with a lower value. Thermal stress is responsible for retarding crop growth (Jones *et al.*, 1980), underdevelopment of cells and tissues (Mccrec and Davis, 1974), and less photosynthesis (Oppenheimer, 1960). These results are comparable with Sani *et al.* (2008), Cossani *et al*. (2012), and Umburanas *et al*. (2019) for maize, wheat, and soybean, respectively. The light extinction coefficient varied by 0.13 with delayed in sowing of mustard. The increase in light extinction coefficient may be due to the modification in canopy toward planophiles, resulting in less radiation interception.

Above-ground biomass varied for mustard cultivars with dates of sowing. The decrease in above-ground biomass of mustard under very late sown conditions may be due to the shorter duration of reproductive crop growth. The shorter duration of the crop growth period provides an insufficient period for transporting the photosynthate to sink. The reduced biomass production in late and very late sown crops may be due to the unfavorable environmental conditions that prevailed at the reproductive and maturity stages. There is about a 25 to 30 percent reduction in above-ground biomass due to delay in sowing. Above-ground biomass has a linear relationship with the crop LAI (Vashisth *et al.* 2011). The reduction in above-ground biomass with delayed sowing from recommended time was reported for mustard (Pradhan *et al*., 2018), for wheat (Rane *et al*., 2007 and Sehgal *et al*., 2018) for maize (Kar and Kumar, 2015), for soybean (Umburanas *et. al*., 2019) and for cotton (Prakash *et al*.,2010).

The results reported that seed yield is reduced by 44 to 50 percent, oil content is reduced by 3 to 4 percent and harvest index (HI) is reduced by 5 to 6 percent with the delayed of sowing date in mustard. All varieties of timely sown crops had higher seed yield than late sown and very late sown crops. High temperature during the flowering and pod formation stage for late sown crops reduced the dry matter accumulation into the seed, shorten the pod formation period and reduced the seed yield. Pradhan *et al*. (2014) observed the significant interaction between the date of sowing and cultivars with respect to seed yield of mustard and found about 45 % reduction in mustard seed yield for late sown crop. A similar type of results were obtained by Kumar *et al*. (2008) for soybean crop, Prakash *et al*. (2010) for cotton crop, and Roy *et al*. (2005) for mustard crop.

The seed oil content is an inherent characteristic of a cultivar but showed variation under different environmental conditions, which can be modify by sowing on different dates. Kumar *et al*. (2017) reported that a certain ambient temperature should be required for oil accumulation in an oilseed crop. This result is supported by (Pritchard *et. al*., 2000; Ozer, 2003; Turhan *et. al*., 2011; Adak *et. al*., 2011a). Harvest index reduced with delay in sowing due to more reduction in seed yield than the above-ground biomass. Reduction in harvest index for very late sown crop is supported by (Panda *et. al,*.2004; Lallu *et. al.*,2010; Afroz *et. al*., 2011).

Radiation use efficiency enhanced, with increased production of biomass and seed yield, and decreased in total intercepted PAR. The RUE was significantly reduced in late sown mustard crops. There was about 15 percent reduction in RUE based on biomass and 38 percent reduction in RUE based on seed yield for the very late sown crop as compared to timely sown crop. This might be due to the reduction in seed yield and above-ground biomass with delay in sowing. Also delay in sowing reduced leaf area index, which affects the interception of PAR and reduction in photosynthesis. Giunta and Motzo, 2004 revealed that the staggered date for crop sown generally affects the biophysical parameters of the crop, canopy coverage, and radiation utilization. These findings are in conformity with the results reported by (Li *et al*., 2008a; Han *et al*., 2008 and Pradhan *et al*.,2014).

Water productivity depend on biomass production or seed yield and seasonal evapotranspiration (ET). There was 12 percent reduction in ET in the late sown crop because of shorter crop duration and less LAI. The reduction in water productivity is mainly due to reduction in seed yield and biomass without reduction in seasonal evapotranspiration, resulted the positive correlation of water productivity with seed yield and above-ground biomass. Field studies of water productivity and radiation use efficiency are challenging due to the lack of simple measurement techniques and the complexity of the traits (Narayanan *et al*., 2013; Hall *et al*., 1990; Sinclair and Muchow, 1999). The results were supported by Pradhan *et al*., 2014 for the mustard crop.

The LAI, above-ground biomass, seed yield, oil content, RUE and WP during *Rabi* 2017-18, were higher compared to *Rabi* 2016-17. This might be due to more favourable weather conditions during *Rabi* 2017-18 than during *Rabi* 2016-17 experimental years. The minimum temperature during the reproductive stage was high in *Rabi* 2016-17 as compared to *Rabi* 2017-18. This indicates that minimum temperature during the reproductive stage affect the crop seed yield. Higher temperature during the vegetative phase and lower temperature during the reproductive phase plays a significant role in better crop growth and development (Hatfield, 2008). Nishad, (2017) reported that the reduction in biomass, seed yield, and harvest index was 115.5 g, 447 kg ha^{-1}and 3.25 % as per one degree rise in minimum temperate during reproductive phase. Pusa Tarak showed the lowest seed yield, biomass, and LAI due to its short duration growing property among all cultivars.

The InfoCrop mustard model is calibrated and validated to study the impact of high temperature on crop growth, biophysical parameters, and seed yield in the experimental field as well as in the farmer's field . The temperature stress was generated by the delayed sowing of mustard crops in different areas. This study showed excellent simulation for germination, 50 percent anthesis, and physiological maturity with InfoCrop mustard model having nRMSE value less than 10 %. A good agreement was found for LAI, seed yield, and biomass at the experimental field of IARI, New Delhi. The simulation accuracy was more for timely sown crop than delayed sown crop. This might be due to the InfoCrop mustard model being calibrated for timely sown crop, which leads to a high degree of uncertainty about weather variability on biophysical parameters and their variability. This has been supported by (Adak *et al.* 2009; Boomiraj *et al*. 2010; Keerthi *et al.* 2017; Gill *et al*. 2016). Several other researchers successfully adapted, calibrated, and validated the InfoCrop simulation model for different crops in different regions, rice (Aggarwal *et al*., 2006b), wheat (Aggarwal *et al*., 2006b), potato (Singh *et al*., 2005), cotton (Hebbar *et al*., 2008), and coconut (Kumar *et al*., 2008). Krishnan et al. (2016) evaluated the web-based InfoCrop wheat model for wheat growth and development, and its performance was found to be satisfactory.

The InfoCrop-mustard model is also used for multi-stage prediction of biomass and seed yield. The InfoCrop-mustard model showed that results of multi-stage in-season yield prediction are closer to the observed yield for prediction done at pod filling stage than done at 50 percent flowering stage. The mean percent deviation of mustard biomass and seed yield prediction by observed yield done at 50 percent flowering stage was 25.77 and 31 percent , whereas at the pod formation stage was 17.4 and 20.0, respectively. This indicates better performance of the model at the pod formation stage. This may be due to the more data used for prediction done at the pod formation stage as compared to prediction done at 50 percent flowering stage. The InfoCrop-mustard model overestimated the mustard prediction. Several researchers confirmed our findings on different crops such as CERES-wheat model by Bannayan *et al*.(2003); CERES-maize by Quiring *et al*. (2008); AquaCrop-maize model by Kipkorir *et al*. (2007); InfoCrop-maize Maize by Vashisth *et al* (2018); InfoCrop-wheat Wheat by Vashisth *et al* (2019).

The value of nRMSE increased by 5-7 % for seed yield, 3-5 % for biomass, and 6-8 % for LAI when we simulate the InfoCrop mustard model in farmer’s fields located at Sitara and Mukundpura villages of Bharatpur District, Rajasthan. In comparison to experimental fields, the situation of farmer’s fields is more challenging owing to large-scale variability in sowing conditions, diversity in management practices, and unavailability of precise measurements. Dhakar *et al*. (2019) successfully validated InfoCrop-wheat model under varied conditions in farmer’s fields located in Mumtajpur and Lokra villages of Haryana. All previous studies on evaluating the InfoCrop model were in the experimental research field and very few at the farmer's field. This study clearly shows that the InfoCrop-mustard model is very suitable for simulating mustard yield at farmer's field of the study areas.

Mustard yield prediction is done using variable selection by SMLR or variable extraction by PCA and ANN, SVM and RF (SMLR-ANN, SMLR-SVM, SMLR-RF, PCA-ANN, PCA-SVM and PCA-RF) techniques for all six study areas. The PCA-SVM and PCA-ANN is highly effective with good nRMSE values rather than the SMLR- SVM and SMLR-ANN. SMLR-RF performed better than PCA-RF. The variable selection and variable extraction techniques were used to prevent overfitting and reduce model complexity. There is no elimination of input variables in the feature extraction technique, whereas feature selection tries to eliminate less relevant input variables, thereby decreasing the complexity of the model. Variable extraction techniques empower us to determine if a combination of the original features can generate new features, which are more discriminating in the outcome than the original features. Feature selection keeps a subset of the original features, while feature extraction creates new ones. The new variables developed by PCA are orthogonal, which means that they are uncorrelated to each other power to explain the model output. So it is beneficial to use PCA than SMLR to find out appropriate input variables in any machine learning technique. Feature extraction is useful for improving the performance of regression models, improving the stability against noise, avoiding over-fitting, reducing the training and testing time, and reducing the measurement and storage requirements. Several researchers used variable extraction techniques for crop yield prediction (Azfar *et. al.*, 2015; Annu *et. al.*, 2017, Suzuki *et. al.*, 2020). However Jolliffe (1982) predicted good results by feature selection-based machine learning techniques.

Based on model accuracy parameters, MAE, nMAE, RMSE, nRMSE, and RPD, the SVM performed better than ANN and RF either based on variable selection or variable extraction for mustard yield prediction of study areas. It may be due to the use of the radial basis function or polynomial kernels function in the SVM technique, so it has more flexibility over ANN and RF techniques. Additionally, ANN is more easily influenced by the minimum value of output than SVM. Random Forests needs a larger number of instances to work its randomization concept well and generalize to the novel data. More decision trees in a forest are needed to develop a robust prediction model. The value ntree and mtry decide the number of variables randomly sampled at each split. PCA is the linear combination of input variables. PCA explain about 99 percent variability, which are used as explanatory variables in the model and model parameters. Apart from that, we use more input variables selected from variable selection techniques to explain variability explanatory variables for crop yield prediction. The least performance of PCA-RF as compared to others is might be due to less value of ntree and mtry for building a model. Thus it is recommended that if you have a small amount of input variables compared to possible variations of the instances then you should avoid the RF technique. Azfar *et al*., (2014) found prediction accuracy in terms of RMSE and R^2 values by SVM techniques was better than ANN and RF. Ahmad (2017) noticed that ANN performed marginally better than RF to predict the energy consumption of construction in terms of RMSE of 4.97 and 6.10, respectively. Palanivel and Surianarayanan (2019) reviewed several types of machine learning techniques, linear regression, artificial neural networks, and support vector machine and found that crop prediction done by SVM model was better as compared to other models. Song *et al.* (2011) recommended that the optimal combination of different crop yield prediction models is a more reliable approach to overdrawing the limitation of individual approaches at regional scales. The study showed that among all possible combinations done by PCA-ANN, PCA-SVM and PCA-RF for predicting mustard yield, PCA(SVM +ANN) is more accurate than others for IARI, New Delhi and all five zones of Rajasthan. Whereas, SMLR(SVM +RF) predict better mustard yield for IARI, New Delhi and study zones of Rajasthan as compared to all other possible combinations done by SMLR-ANN, SMLR-SVM and SMLR-RF. This may be due to the more variables available by feature selection for decision tree making. Hsiao and Wan (2013) concluded that a combination of different prediction models can develop a more reliable prediction model to overdrawn the limitation of each model.

6.Summary and Conclusions

Accurate and timely crop yields prediction is required for crop management, food supply, strategic resource mobilization, trading, import, export, insurance, financial decision, etc. Conventionally, the crop yield estimation is based on nationwide crop-cutting experiments, the results are aggregated at various administrative units, but these estimates come after the harvest. Several researchers developed prediction models based on statistical techniques by time series data. However, time series data are often non-linear and irregular. Machine-learning techniques have been used to overcome the problems of prediction by non-linear and non-stationary time series datasets. Another tool for crop yield predictions is crop simulation models used to monitor crop growth and estimate yields using soil, plant and climatic variables.

For mustard yield prediction by machine learning and crop simulation models, an experiment was conducted during *Rabi* 2016-17 and 2017-18 for a mustard crop at ICAR-IARI, New Delhi research farm. Three varieties of mustard, RH-406, Pusa Tarak and Girraj were sown on three different dates (timely, late and very late) to generate different weather conditions during different growth stages. Periodic observations of crop phenology, leaf area index, biomass, IPAR and soil moisture content were done at fifteen days interval. Seed yield, oil content and harvest index were measured after harvest. The InfoCrop model was calibrated from the observation taken from *Rabi* 2016-17 sown mustard crop and model validation was done from the observation taken during *Rabi* 2017-18 for the same variety sown under the same treatment. Simulation of phenology, LAI, above-ground biomass and seed yield for RH-406, Pusa Tarak and Girraj cultivars of mustard sown at the IARI research farm were done by InfoCrop model. Twenty farmers were selected from Mukundpura and Sitara villages of Bharatpur district to validate the InfoCrop-mustard model at the farmer's field. Simulation of LAI, biomass and seed yield for RH-406 and Girraj cultivars of mustard in the farmer's field were done by the InfoCrop model. Multistage mustard yield predictions were done by InfoCrop model during *Rabi* 2017-18 at 50 percent flowering and at pod formation stages. For the development of mustard yield prediction models by machine learning techniques, long term weather and mustard yield data were collected from IARI, New Delhi and five zones of Rajasthan. The models for mustard yield prediction were developed; 1) using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), and variable selection by SMLR and RF (SMLR-RF) techniques, 2) variable extraction using PCA and ANN (PCA-ANN), variable extraction using PCA and SVM (PCA-SVM), and variable extraction using PCA and RF (PCA-RF) techniques, 3) optimal combination of SMLR-ANN, SMLR-SVM and SMLR-RF, and 4) optimal combination of PCA-ANN, PCA-SVM and PCA-RF. The salient findings of our study are given below.

- There were more rainfall and a high minimum relative humidity during *Rabi* 2016-17 than during *Rabi* 2017-18. The temperature pattern shows a decreasing trend from sowing upto 1st SMW and, after that shows an increasing trend. The minimum temperature was more during *Rabi* 2016-17 than during *Rabi* 2017-18.
- Among all the cultivars sown on different dates, crop growing period for Pusa Tarak is shortest.
- More days were required to obtain 50 percent flowering for all the cultivars in delayed sowing because of low values of maximum and minimum temperatures during that period.
- With delay in sowing crop duration of all the cultivars was less due to increase in temperature during pod formation stage.

- The peak value of LAI attained at 85-90, 78-83 and 72-79 days after sowing in the timely, late and very late sown crop, respectively for RH-406 and Girraj in both years. In case of Pusa Tarak, the peak LAI value varied from 80 to 65 days in the timely, late and very late sown crop, respectively in both the years.

- RUE decreased with delay in sowing. Pusa Tarak is a short duration cultivars having lowest value of TIPAR as compared to other two cultivars hence, Pusa Tarak showed the highest RUE among all cultivars, whereas it had the lowest yield and biomass. RUE showed non-significant results based on yield and significant results based on biomass.

- In the timely sown crop during *Rabi* 2016-17, cultivars RH-406 showed the highest biomass, fIPAR and peak LAI. Similar, results were found during *Rabi* 2017-18. The peak value of fIPAR and LAI occurred early in delayed sown crops..

- Change in the growing thermal environment due to delay in sowing reduced LAI, radiation interception, biomass, seed yield, radiation use efficiency and harvest index.

- RH-406 had better crop growth and seed yield among all three cultivars sown on all three dates.

- The weather condition during *Rabi* 2017-18 was more favorable to accomplishing more LAI compared to weather condition during *Rabi* 2016-17 for all dates of sowing. The date and peak value of LAI occurred early in late sown crop, indicating that the terminal heat stress caused faster development of leaf area with low value.

- The mustard sowing dates at farmer's field were between 10th October, 2017 (timely sown) and 25th October, 2017 (late sown) which, affects the crop yield and biomass of Sitara and Mukundpura villages of Bharatpur district, Rajasthan.

- Better accuracy in simulation of phenology by Info-Crop may be attributed to that model accounting for the effect of the date of sowing on thermal time accumulation.

- Simulation of above-ground biomass and seed yield for different treatments by InfoCrop model was good with nRMSE value less than 15%. Percentage deviation of simulated above-ground biomass and grain yield from observed value were highest in very late sown crop during both the years.

- The peak values of LAI at Mukundpura and Sitara villages in the farmer's field during *Rabi* 2017-18 were overestimated by the InfoCrop-mustard model. Model performance for LAI simulation at the farmer's field was fair, with nRMSE value less than 22.1 %.

- Biomass simulation for cultivars RH-406 and Girraj at Mukundpura and Sitara villages in the farmer's field during *Rabi* 2017-18 by the InfoCrop-mustard model were good, with nRMSE value less than 20 %.

- InfoCrop-mustard model performed better for simulating seed yield than above-ground biomass and LAI at farmer's fields. nRMSE value during simulation of seed yield were 17.1 and 18.1 % for RH-406 and Girraj, respectively.

- InfoCrop-mustard model was able to simulate growth, development and yield of mustard crop. The model overestimates within a reasonable error at the experimental farm, IARI, New Delhi and with a larger error at the farmer’s field as compared to experimental farm, IARI, New Delhi.

- The biomass prediction of mustard done by InfoCrop at the pod formation stage was good having nRMSE value between 15.7 to 18.6 % and fair at the 50 % flowering stage having nRMSE value between 23.1 to 27.6 %.

- The percentage deviation of predicted mustard yield by InfoCrop at the pod formation stage from observed yield was lower than the predicted mustard yield done at the 50 percent flowering stage.

- The multi-stage prediction done by the InfoCrop-mustard model for biomass and seed yield was better for timely sown crop at the pod formation stage than the 50 percent flowering stage.

- The PCA-SVM model performed best among all the six models (SMLR-ANN, SMLR-SVM, SMLR-RF, PCA-ANN, PCA-SVM and PCA-RF) developed for mustard yield prediction for all six study areas.

- Based on model accuracy parameters, performance of the models developed by SMLR-ANN, SMLR-SVM, SMLR-RF and optimum combination of these three models, optimum combination performed better than the individual. Optimum combination of SMLR(ANN+SVM+RF) performed best for IARI, New Delhi and Zone V, optimum combination of SMLR(ANN +SVM) performed best for Zone II, and optimum combination of SMLR(SVM +RF) performed best for Zone I, Zone III, Zone IV of study area.

- On the basis of model accuracy parameters, performance of the models developed by PCA-ANN, PCA-SVM, PCA-RF and optimum combination of these three models, optimum combination performed better than the individual for all the study area except for Zone III and Zone V individual model SVM performed better than optimum combination. Optimum combination of PCA(ANN+SVM) performed best for all the study area except for Zone V, where optimum combination of PCA(ANN+SVM+RF) performed best for mustard yield prediction.

- By comparing the performance on the basis of model accuracy parameters, models developed either by variable selection by SMLR or variable extraction by PCA, model developed by variable extraction by PCA performed better than model developed by variable selection by SMLR.

Abstract

Accurate and timely prediction of crop yields is necessary for crop management and planning decisions of the government regarding storage, import, export, etc. For mustard yield prediction by machine learning and simulation models, experiments were conducted during *Rabi* 2016-17 and 2017-18 at ICAR-IARI, New Delhi research farm. Three cultivars of mustard, RH-406, Pusa Tarak and Girraj were sown on three different dates. Periodic observations on crop phenology, leaf area index, above-ground biomass, fIPAR and soil moisture content were done at fifteen days interval. Seed yield, oil content and harvest index were measured after harvest. InfoCrop model was calibrated from the field observations taken during *Rabi* 2016-17 sown mustard crop of the same variety under the same treatment. Model validation was done from the observation taken during *Rabi* 2017-18 sown crops under similar treatments. Simulations of phenology, LAI, biomass and seed yield were done by InfoCrop-mustard model for RH-406, Pusa Tarak and Girraj cultivars sown at ICAR-IARI, New Delhi research farm. InfoCrop-mustard model was validated at farmer’s field selected from Mukundpura and Sitara villages, Bharatpur district, Rajasthan. Simulation of LAI, biomass and seed yield for RH-406 and Girraj cultivars at farmer's field were done by InfoCrop model. Mustard above-ground biomass and seed yield prediction were done at 50 percent flowering and pod formation stages by InfoCrop model during *Rabi* 2017-18. For the development of models by machine learning techniques, long-term weather and mustard yield data were collected from IARI, New Delhi and five zones of Rajasthan. The Models were developed for mustard yield prediction using variable selection by SMLR and ANN (SMLR-ANN), variable selection by SMLR and SVM (SMLR-SVM), variable selection by SMLR and RF (SMLR-RF) techniques, variable extraction by PCA and ANN (PCA-ANN), variable extraction by PCA and SVM (PCA-SVM), and variable extraction by PCA and RF (PCA-RF) techniques. Optimal combinations of developed model were done on the basis of weights to minimize the error in the crop yield prediction.

Results showed that because of favorable weather conditions during different phenological stages, timely sown crops had higher values of LAI, above-ground biomass, RUE and seed yield, followed by late sown and very late sown crops. Change in the crop

growing thermal environment due to delay in sowing reduced LAI, radiation interception, biomass, seed yield, radiation use efficiency and harvest index. Among all three cultivars, RH 406 is giving better crop growth and seed yield with respect to all dates of sowing. Pusa Tarak is a short-duration variety among all cultivars. The weather conditions during *Rabi* 2017-18 was more favorable for better crop growth and yield as compared with during *Rabi* 2016-17 for all three dates of sowing.

The variations in sowing dates were found at farmer's field which affected the crop yield and above-ground biomass of Sitara and Mukundpura village of Bharatpur, Rajasthan. Better accuracy in the simulation of phenology by InfoCrop may be attributed to that model accounting for the effect of the date of sowing. Simulation of above-ground biomass and seed yield by InfoCrop model was good with nRMSE value less than 15 %. The percent deviation from observed biomass and seed yield was highest in late sown crop during both the years. Model predictions at the farmer's field for cultivar RH-406 and Girraj during *Rabi* 2017-18 for LAI simulations were fair with nRMSE value less than 22.1 %, and for biomass, simulations were good with nRMSE value less than 20 %. The nRMSE value of model simulation for seed yield at farmer's field was 17.1 % for RH-406 and 18.1 % for Girraj, respectively. The percentage deviation of mustard yield prediction by observed yield done by InfoCrop at the pod formation stage was lower than that done at 50 percent flowering stage and also lower for timely sown crop, followed by late and very late sown crop.

The models developed either by variable selection by SMLR or variable extraction by PCA and using ANN, SVM, RF techniques; SVM model performed best for mustard yield prediction for all the six study areas. Among all six models, PCA-SVM performed best for all the study areas. Performance of the model developed by SMLR-ANN, SMLR-SVM, SMLR-RF techniques and optimum combination of these three models, an optimum combination performed better than the individual. An optimum combination of SMLR(ANN+SVM+RF) performed best for Delhi and Zone V, SMLR(ANN +SVM) performed best for Zone II and SMLR(SVM +RF) performed best for Zone I, Zone III and Zone IV of study areas.

Performance of the model developed by PCA-ANN, PCA-SVM, PCA-RF techniques and optimum combination of these three models, optimum combination performed better than the individual for all the study area except for Zone III and Zone V individual model SVM performed better than optimum combination. The optimum combination of PCA(ANN+SVM) performed best for all the study areas except for Zone V optimum combination of PCA(ANN+SVM+RF), performed best for mustard yield prediction. By comparing the performance on the basis of model accuracy parameters, models developed either by variable selection by SMLR or variable extraction by PCA, model developed by variable extraction by PCA performed better than model developed by variable selection by SMLR hence, can be used for district level mustard yield prediction.

मशीन लर्निंग और फसल सिमुलेशन मॉडल द्वारा सरसों की उपज का अनुमान

सार

फसल प्रबंधन, भंडारण, आयात, निर्यात आदि के संबंध में सरकार के नियोजन निर्णयों के लिए फसल की पैदावार का सटीक और समय पर पूर्वानुमान आवश्यक है। मशीन लर्निंग और फसल सिमुलेशन मॉडल द्वारा सरसों की उपज का अनुमान के लिए, रबी 2016-17 और 2017-18 के दौरान भा.कृ.अनु.स., नई दिल्ली के अनुसंधान फार्म में एक प्रयोग किया गया । सरसों की तीन किस्में आरएच-406, पूसा तारक और गिरराज को तीन अलग-अलग समय पर बोया गया। फसल चरण, पत्ती क्षेत्रफल सूचकांक, बायोमास, एफआईपीएआर और मिट्टी की नमी पर आवधिक अवलोकन 15 दिन के अंतराल पर किया गया । फसल के बाद बीज उपज, तेल की मात्रा और फसल सूचकांक को मापा गया। इंफोक्रॉप मॉडल को रबी 2016-17 में बोई गई सरसों की किस्मों के अवलोकन से अंशांकित किया गया था। मॉडल सत्यापन रबी 2017-18 के दौरान बोई गई सरसों की किस्मों के अवलोकन से किया गया । भा.कृ.अनु.स., नई दिल्ली के अनुसंधान फार्म में बोए गए आरएच-406, पी. तारक और गिरराज के लिए इन्फोक्रॉप-सरसों मॉडल द्वारा फेनोलॉजी, पत्ती क्षेत्रफल सूचकांक, बायोमास और बीज उपज का अनुकरण किया गया। भरतपुर जिले के मुकुंदपुरा एवं सतारा गांव से चयनित किसान के खेत में इन्फोक्रॉप-सरसों मॉडल का सत्यापन किया गया। पत्ती क्षेत्रफल सूचकांक, बायोमास और बीज उपज का अनुकरण आरएच-406 और गिरराज के लिए इन्फोक्रॉप द्वारा किसान के खेत में किया गया। रबी 2017-18 के दौरान इंफोक्रॉप मॉडल द्वारा 50 प्रतिशत फूल और अनाज भरने के चरण में जमीन के ऊपर बायोमास और बीज उपज का पूर्वानुमान लगाया गया । अनुभवजन्य मॉडल के विकास के लिए, भा.कृ.अनु.स., नई दिल्ली और राजस्थान राज्य के पांच क्षेत्रों से दीर्घकालिक मौसम और सरसों की उपज के आंकड़े एकत्र किए गए । एसएमएलआर द्वारा परिवर्तनीय चयन और एएनएन, एसवीएम, आरएफ तकनीकों, पीसीए द्वारा परिवर्तनीय निष्कर्षण और एएनएन, एसवीएम, आरएफ तकनीकों द्वारा सरसों की उपज के पूर्वानुमान के लिए मॉडल विकसित किए गए | फसल उपज के पूर्वानुमान में त्रुटि को कम करने के लिए वजन के आधार पर विकसित मॉडल का इष्टतम संयोजन किया गया ।

परिणामों से पता चला कि विभिन्न फीनोलॉजिकल चरणों के दौरान अनुकूल मौसम की स्थिति के कारण, समय पर बोई गई फसलों में पत्ती क्षेत्रफल सूचकांक, बायोमास, विकिरण उपयोग

दक्षता और अनाज की उपज अधिक , इसके बाद देर से बोई गई और बहुत देर से बोई गई फसलों में पाया गया। फली बनने की अवस्था के दौरान तापमान और पानी की कमी के कारण सभी किस्मों में देर से बोई गई फसल की अवधि कम पायी गयी । बुवाई में देरी के कारण बढ़ते गर्म वातावरण के कारण पत्ती क्षेत्रफल सूचकांक, विकिरण अवरोधन, बायोमास और बीज उपज में कमी पायी गयी, जिसके परिणामस्वरूप विकिरण उपयोग दक्षता और फसल सूचकांक में कमी पायी गयी । तीनों किस्मों में बुवाई की सभी तिथियों में आरएच 406 में बेहतर फसल वृद्धि और बीज उपज पाया गया। तीनों किस्मों में पूसा तारक एक छोटी अवधि की किस्म है। रबी 2017-18 के दौरान मौसम की स्थिति बुवाई की तीनों तिथियों के लिए रबी 2016-17 की तुलना में बेहतर फसल वृद्धि और उपज के लिए अधिक अनुकूल थी। राजस्थान के भरतपुर जिले के सतारा और मुकुंदपुरा गांव में किसानों के खेत में बुवाई की तारीख में भिन्नता पाई गई, जिसने फसल की उपज और जमीन के ऊपर के बायोमास को प्रभावित किया। इन्फोक्रॉप द्वारा फेनोलॉजी के अनुकरण में बेहतर सटीकता के लिए इस मॉडल को बुवाई की तारीख के प्रभाव का लेखांकन के लिए जिम्मेदार ठहराया जा सकता है। इंफोक्रॉप मॉडल द्वारा जमीन के ऊपर बायोमास और बीज उपज का अनुकरण एनआरएमएसई मूल्य 15% से कम के साथ अच्छा था। दोनों वर्षों के दौरान देर से बोई गई फसल में देखे गए बायोमास और बीज उपज में प्रतिशत विचलन सबसे अधिक था। पत्ती क्षेत्रफल सूचकांक सिमुलेशन के लिए रबी 2017–18 के दौरान किसान के खेत में आरएच -406 और गिरराज के लिए मॉडल पूर्वानुमान 22.1% से कम एनआरएमएसई मूल्य के साथ उचित थी और बायोमास सिमुलेशन के लिए एनआरएमएसई मूल्य 20% से कम के साथ अच्छा था। किसान के खेत में बीज उपज के लिए मॉडल सिमुलेशन का एनआरएमएसई मूल्य क्रमशः आरएच-406 के लिए 17.1 और गिरराज के लिए 18.1 था। इंफोक्रॉप द्वारा अनाज भरने के चरण में प्रेक्षित उपज के आधार पर सरसों की उपज के पूर्वानुमान का प्रतिशत विचलन 50 प्रतिशत फूल चरण में किए गए पूर्वानुमान से कम था और समय से बोई गई फसल में सबसे कम उसके बाद देर से और बहुत देर से बोई गई फसल में पाया गया ।

मॉडल चाहे एसएमएलआर द्वारा परिवर्तनीय चयन और एएनएन, एसवीएम, आरएफ तकनीकों का उपयोग करके विकसित किया गया हो या पीसीए द्वारा परिवर्तनीय निष्कर्षण और एएनएन, एसवीएम, आरएफ तकनीकों का उपयोग करके विकसित किया गया हो, एसवीएम मॉडल

ने सभी छह अध्ययन क्षेत्रों के लिए सरसों की उपज पूर्वानुमान के लिए सर्वश्रेष्ठ पाया गया। सभी छह मॉडलों में से, पीसीए-एसवीएम सभी अध्ययन क्षेत्रों के लिए सर्वश्रेष्ठ पाया गया।

एसएमएलआर–एएनएन, एसएमएलआर–एसवीएम, एसएमएलआर–आरएफ तकनीकों द्वारा विकसित मॉडल और इन तीन मॉडलों का इष्टतम संयोजन में से, इष्टतम संयोजन मॉडल अकेले मॉडल की तुलना में बेहतर पाया गया। एसएमएलआर (एएनएन+एसवीएम+आरएफ) इष्टतम संयोजन दिल्ली और जोन V अध्ययन क्षेत्रों के लिए सर्वश्रेष्ठ पाया गया, एसएमएलआर (एएनएन + एसवीएम) जोन II अध्ययन क्षेत्र के लिए सर्वश्रेष्ठ पाया गया और एसएमएलआर (एसवीएम +आरएफ) जोन I, जोन III और जोन IV अध्ययन क्षेत्रों के लिए सर्वश्रेष्ठ पाया गया ।

पीसीए–एएनएन, पीसीए–एसवीएम, पीसीए–आरएफ तकनीकों और इन तीन मॉडलों का इष्टतम संयोजन द्वारा विकसित मॉडल में से इष्टतम संयोजन मॉडल जोन III और जोन V अध्ययन क्षेत्रों को छोड़कर सभी अध्ययन क्षेत्रों के लिए अकेले मॉडल की तुलना में बेहतर पाया गया। जोन III और जोन V अध्ययन क्षेत्रों के लिए इष्टतम संयोजन की तुलना में एसवीएम मॉडल बेहतर पाया गया। सरसों की उपज के पूर्वानुमान के लिए पीसीए (एएनएन+एसवीएम) इष्टतम संयोजन जोन V अध्ययन क्षेत्र को छोड़कर सभी अध्ययन क्षेत्रों के लिए सर्वश्रेष्ठ पाया गया। जोन V अध्ययन क्षेत्र के लिए पीसीए (एएनएन+एसवीएम+आरएफ) इष्टतम संयोजन सर्वश्रेष्ठ पाया गया । मॉडल सटीकता मापदंडों के आधार पर एसएमएलआर द्वारा परिवर्तनीय चयन या पीसीए द्वारा परिवर्तनीय निष्कर्षण द्वारा विकसित मॉडल के प्रदर्शन की तुलना करने पर पीसीए द्वारा परिवर्तनीय निष्कर्षण द्वारा विकसित मॉडल, एसएमएलआर द्वारा परिवर्तनीय चयन द्वारा विकसित मॉडल की तुलना में बेहतर पाये गये इसलिए इनका उपयोग जिला स्तर पर सरसों की उपज का पूर्वानुमान के लिए किया जा सकता है।

Bibliography

Acock, B., Reddy, V.R., .Whisler, E. D., Baker, D. N., McKinion, J. M., Hodges, H. F. and Boote, K. J. (1983). The soybean crop simulator GL YCIM: Model Documentation 1982, Report No.2, United States Department of Energy, Carbon Dioxide Research Division, Office of Energy Research, Washington D.C., USA.

Adak, T., Chakravarty, N. V. K., and Saxena, R. (2009). Growth and yield prediction in mustard using InfoCrop simulation model. *Journal of Agrometeorology*, **11**(2): 156-161.

Adak, T., Chakravarty, N. V. K., Muthukumar, M., Deshmukh, P. S., Joshi, H. C. and Katiyar, R. K. (2011a). Evaluation of biomass and thermal energy utilization efficiency of oilseed Brassica (*Brassica juncea*) under altered micro-environments. *Biomass Bioenergy*, **35**: 2254-2267.

Adak, T., Narjary, B., Chakravarty, N. V. K., (2011b). Response of Brassica to microenvironment modification under semi-arid agroecosystem. *Indian Journal of Agricultural Sciences*, **81**: 744-750.

Afroz, M. M., Sarkar, M. A. R., and Bhuiya, M. S. U. (2011). Effect of sowing date and seed rate on yield performance of two mustard varieties. *Journal of the Bangladesh Agricultural University*, **9**: 5-9.

Aggarwal, P. K., and Kalra, N. (1994). Analyzing the limitations set by climatic factors, genotype, and water and nitrogen availability on productivity of wheat II. Climatically potential yields and management strategies. *Field Crops Research*, **38**(2): 9 3-103.

Aggarwal, P. K., Banerjee, B., Daryaei, M. G., Bhatia, A., Bala, A., Rani, S., Chander, S., Pathak, H. and Kalra, N. (2006b). InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. I. Model description. *Agricultural Systems*, **89:** 1–25.

Aggarwal, P. K., Kalra, N., Chander, S., and Pathak, H. (2004). Infocrop—A generic simulation model for annual crops in tropical environment (p. 132). New Delhi: Indian Agricultural Research Institute.

Aggarwal, P. K., Kalra, N., Chander, S., and Pathak, H. (2006a). InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. I. Model description. *Agricultural systems*, **89**(1): 1-25.

Aggarwal, P. K., Kalra, N., Singh, A. K., and Sinha, S. K. (1994). Analyzing the limitations set by climatic factors, genotype, water and nitrogen availability on productivity of wheat I. The model description, parametrization and validation. *Field Crops Research*, **38**(2): 73-91.

Aggarwal, P.K., Bandyopadhyay, S.K., Pathak, H., Kalra, N., Chander, S., Kumar, S., (2000). Analysis of yield trends of the rice-wheat system in north-western India. *Outlook on Agriculture* **29** (4): 259-268.

Agrawal, R., Jain, R. C. and Jha, M. P. (1986). Models for studying rice crop weather relationship. *Mausam*, **37** (1): 67-70.

Agrawal, R., Jain, R. C. and Mehta, S. C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis. *Indian Journal of Agricultural Science*, **71**(7): 487-490.

Ahmad, M. W., Mourshed, M., and Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, **147**: 77-89.

Akinseye, F. M., Adam, M., Agele, S. O., Hoffmann, M. P., Traore, P. C. S., and Whitbread, A. M. (2017). Assessing crop model improvements through comparison of sorghum (*sorghum bicolor* L. moench) simulation models: a case study of West African varieties. *Field Crops Research*, **201:** 19-31.

Akula, B. (2003). Estimating wheat yields in Gujarat using WTGROWS and INFOCROP model. Ph.D. (Agri. Meteorology) thesis submitted to GAU, Sardar Krushinagar, pp: 50-68.

Akula, B., Shekh, A. M. and Parmar, R. S. (2005b). Estimation of variability in wheat production levels in middle Gujarat using simulation model. *Journal of Agrometeorology*, **7**(2): 208-213.

Alagarswamy, G. and Ritchie, J. T. (1989). Phasic development in CERES Sorghum model, In: Hodges, T. (ed.), Predicting Crop Phenology, CRC Press, Boca Raton, FL, USA. pp, 144-151.

Allen. E. J. and Morgan, D. G. (1972). A quantitative analysis of the effects of nitrogen on the growth, development and yield of oilseed rape. *The Journal of Agricultural Science,* **78**: 315-324.

Annu, Sisodia B. V. S. and Rai V. N. (2017). An application of principal component analysis for pre- harvest forecast model for wheat crop based on biometrical characters. *International Research Journal of Agricultural Economics and Statistics,* **8**: 83–87.

Arkin, G. F., Vanderlip, R. L. and Ritchie, J. T. (1976). A dynamic grain sorghum growth model, *Transactions of the ASAE*, **19**: 622 - 630.

Armstrong, J. S. (ed.), 2001. Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston, Kluwer Academic Publishers.

Athani, S. S. and Tejeshwar, C. H. (2017). Support Vector Machine-Based Classification Scheme of Maize Crop, *IEEE Conference*, 84-88.

Azam-Ali, S.N., Crout, N.M.J. and Bradley, R.G. (1994). Perspectives in modelling resource capture by crops. In: Monteith, J.L., Scott, R.K., Unsworth, M.H. (Eds.), *Resource Capture by Crops*. Nottingham University Press, Nottingham, pp. 125–148.

Azfar, M., Sisodia, B. V. S., Rai, V. N. and Devi, M. (2015). Pre-harvest forecast models for rapeseed and mustard yield using principal component. *Mausam*, **4**: 761–766.

Baker, D. N., Whisler, F. D., Parton, W. J., Klepper, E. L., Cole, C.V. and Willis, W.O. (1985). The development of winter wheat: A physical physiological process model, In: Wilson, W.O. (ed.), ARS Wheat Yield Project, ARS-38, USDA-ARS, Washington D.C., pp. 176-187.

Balakrishnan, N. and Muthukumarasamy, G. (2016). Crop Production-Ensemble Machine Learning Model for Prediction. *International Journal of Computer Science and Software Engineering*, **5**(7): 148-153.

Bandyopadhyay, K. K., Misra, A. K., Ghosh, P. K., Hati, K. M., Mandal, K. G. and Moahnty, M. (2010b). Effect of irrigation and nitrogen application methods on input use efficiency of wheat under limited water supply in a Vertisol of Central India. *Irrigation Science,* **28**: 285–299.

Bannayan, M., Crout, N. M. J. and Hoogenboom, G.(2003). Application of the CERES-Wheat Model for Within-Season Prediction of Winter Wheat Yield in the United Kingdom. *Agronomy journal*, **95 (**1): 114-125.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, **20**(4): 451-468.

Battisti, R., Sentelhas, P. C. and Boote, K. J. (2017). Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. *Field crops research*, **200:** 28-37.

Bocca, F. F. and Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computer and Electronics in Agriculture,* **128**: 67–76.

Boomiraj, K., Chakrabarti, B., Aggarwal, P.K., Choudhary, R. and Chander, S. (2010). Assessing the vulnerability of Indian mustard to climate change. *Agriculture. Ecosystem Environment,* **138**(5): 265-273.

Boote, K. J., Jones, J. W., Hoogenboom, G., Jagtap, S. S. and Wilkerson, G. G. (1989). PNUTGRO: Peanut crop growth simulation model: User's guide, Florida Agricultural Experiment, University of Florida, Gainsville, Florida, USA.

Bouman, B. A. M. (1995). Crop modelling and remote sensing for yield prediction. *Netherland Journal of Agricultural Sciences,* **43**: 143 - 161.

Bouman, B. A. M., Kropff, M. J., Wopereis, M. C. S., Ten Berge, H. F. M. and VanLaar, H. H. (2001). ORYZA 2000: Modeling lowland rice, IRRI Books, Philipines.

Bouman, B. A. M., Van Diepen, M.C., Vossen, P. and Van der Wal, T. (1997). Simulation and systems analysis tools for crop yield forecasting. *Applications of Systems Approaches at the Farm and Regional Levels,* **1**: 325-340.

Breiman, L. (2001). Random forests. *Kluwer Academic Publishers. Manufactured in The Netherlands,* **45**: 5-32.

Brown, L. G., Jones, J. W., Hesketh, J. D., Hartsog, J. D., Whisler, F. D. and Harris, F. S. (1985). COTCROP: Computer simulation of growth and yield, Mississippi Agricultural and Forestry Experiment Station, Information Bulletin No. 69, Mississippi State, USA.

Buttler, I. W. (1989). Predicting water constraints to productivity of corn using plant-environment simulation models, Ph.D. Thesis, Cornell University, Ithaca, USA.

Carriglia, O. P., Sadras, V. O. and Andrade, F. H. (2004). Intensification of agriculture in south eastern Pampas: Capture and efficiency in the use of water and radiation in double-cropped wheat-soybean. *Field crops research.* **87**: 117-129.

Chand, M., Bangarwa, A. S., Kumar, P. and Pannu, R.K. (1995). Crop weather relationship in *Brassica* species. *Agricultural Science Digest,* **15**(4): 197-200.

Cheng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical science*, **9**(1): 2-30.

Choudhary, D., Patel, H. R. and Pandey, V. (2014b). Calibration and validation of InfoCrop model v.1.0 for yield and yield attributing characters of *kharif* maize in middle Gujarat region. *Trends Bioscience*, 7(13): 1528-1530.

Choudhary, D., Patel, H.R. and Pandey, V. (2014a). Calibration and validation of InfoCrop model v.1.0 for yield and yield attributing characters of *Rabi* maize in middle Gujarat region. *Annals of Biology*, **30**(4): 658-660.

Collins, L. M., Schafer, J. L. and Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, **6**:330–351.

Collobert, R., Bengio, S. and Williamson, R. C. (2001). SVM torch: Support vector machines for large-scale regression problems. *Journal of Machine Learning Research*, **1**:143– 160.

Cossani, C. and Matthew, R. (2012). Physiological Traits for Improving Heat Tolerance in Wheat. *Plant physiology*, **160** (4): 1710-18.

Curry, R. B., Baker, C. H. and Streeter, J. G. (1975). SOYMOD I: A dynamic simulator for soybean growth and development, *Transactions of the ASAE*, **18**: 963 - 968.

Das, B., Nair, B., Reddy, V. K. and Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalized regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology*, **62**(18) 10.1007/s00484-018-1583-6.

Das, D. K., Chakravarty, N. V. K. and Vashisth, A. (2009). Annual report 2008-2009. Division of Agricultural Physics, IARI, New Delhi. pp. 57-62.

Das. B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Chinnusamy, V., Sehgal, V. K. and Gupta, V. K. (2017). Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. *Biosystem Engineering*, **160**:69–83.

Dastidar, K. K. G. and Patra, M. M. (2004). Character association for seed yield components in Indian mustard (*Brassica juncea* L. Czern and Coss). *Journal of Inter academicia*, **8**(2): 155-160.

De Wit, C.T. (1978). Simulation of assimilation, respiration and transpiration of crops, Simulation Monographs, PUDOC, Wageningen, *The Netherlands*, pp-140.

DES (2018). Directorate of Economics and Statistics, Ministry of Agricultural Cooperation and Farmers Welfare, Government of India.

Dey, U. K., Masud, A. K. and Uddin, M. N. (2017). Rice Yield Prediction Model Using Data Mining Techniques. *International Conference on Electrical, Computer and Communication Engineering*, pp 321- 326.

Dhakar, R., Sehgal, V. K., Chakraborty, D., Mukherjee, J. And Kumar, N. (2019). Evaluating InfoCrop model for growth, development and yield of spring wheat at farmers' field in semi-arid environment. *Journal of Agrometeorology,* 21(3) : 254-261

Dhoble, M. V., Mulgir, K. S. and Bhale, V. M. (1997). Effect of sowing dates on growth and yield of different varieties of mustard. *Journal of Maharashtra Agricultural University*, **22**(2): 190-192.

Duncan, W. G., Loomis, R. S., Williams, W. A. and Hanau, R. (1967). A model for simulating photosynthesis in plant communities, *Hilgardia*, **38** (4): 181-205.

Dutta S, Patel, N. K. and Srivastava, S. K. (2001). District wise yield models of rice in Bihar based on water requirement and meteorological data. *Journal of Indian Society of Remote Sensing* **29**:175–181.

Eitzinger, J., Trnka, M., Hösch, J., Žalud, Z., and Dubrovský, M. (2004). Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. *Ecological Modelling*, **171**(3): 223-246.

Emamgholizadeh, S., Parsaeian, M., and Baradaran, M. (2015). Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*, **68**, 89-96.

Everingham, Y., Sexton, J., Skocaj, D. and Bamber, G. I. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, **36**(27): 1-29.

Fick, G. W. (1981). ALSIM I (Level 2) user's manual, Agronomy Mimeograph 81-35, Department of Agronomy, Cornell University, Ithaca, USA.

Fisher, R. A. (1924). The influence of rainfall on the yield of London wheat at Rothamsted. Phil. Trans. Roy. Soc., B213, 89-142.

Fukuda, S., Spreer, W., Yasunaga, E., Yuge, K., Sardsud, V. and Muller, J. (2013). Random Forests modelling for the estimation of mango (Mangifera indica L. cv. Chok Anan) fruit yields under different irrigation regimes. *Agricultural and Water Management*, **116**: 142-150.

Gandhi, N., Armstrong, L. J., Petkar, R. O. and Kumar A.(2016). Rice crop yield prediction in India using support vector machines. *International Joint Conference on Computer Science and Software Engineering*, Khon Kaen, Thailand, pp-1-5.

Garde, Y. A., Dhekale, B. S. and Singh, S. (2015). Different approaches on pre harvest forecasting of wheat yield. *Journal of Applied and Natural Science.* **7** (2): 839 – 843.

Gare, B. N., Khade, K. K. and Bhoi, P. G. (1996). Performance of mustard cultivars at different dates of sowing. *Journal of Maharashtra Agriculture University*, **21**(1): 147-148.

Gaudart, J., Giusiano, B. and Huiart, L. (2004). Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data. *Computational statistics & data analysis*, **44**(4): 547-570.

George, R. K., Rammohan, S., Kulshretha, M. S., Shekh, A.M. and Jaita, H. (2001). Prediction of soil temperature using artificial neural network. *Journal of Agrometeorology,* **3 (1&2):** 169-173.

Ghosh, R. K. and Chatterjee, B. N. (1988). Effect of date of sowing on the productivity of Indian mustard (*Brassica juncea* sub sp. *juncea*) in short and midwinter of the gangetic plains of West Bengal. *Indian Journal of Agricultural Sciences,* **58**(8): 589-596.

Gifford, R. M., Thorne, J. H., Witz, W. D. and Giaquinta, R. T. (1984). Crop productivity and photo-assimilate partitioning. *Science* **225**: 801-808.

Gill, K. K., Kaur, B., Sandhu, S. S. and Kaur, P. (2016). Simulating mustard crop phenology and productivity with InfoCrop-Mustard model under temperature increase scenarios. *Journal of Agrometeorology*, **18**(2): 240-244.

Giri, A. K, Bhan, M. and Agrawal, K. K. (2017). District wise wheat and rice yield prediction using meterolgical variables in eastern Madhya Pradesh. *Journal of Agrometerology*,. **19**(4): 367-369.

Giri, G. (2001). Effect of irrigation and nitrogen on performance of Indian mustard (*Brassica juncea*) and sunflower (*Helianthus annuus*) under two dates of sowing. *Indian Journal of Agronomy*, **46**(2): 304-308.

Giunta, F. and Motzo, R.(2004). Grain yield, dry matter, and nitrogen accumulation in the grains of durum wheat and spring triticale cultivars grown in a Mediterranean environment. *Australian Journal of Agricultural Research*, **56**(1): 25-32.

Goyal, A., Das, D., Sehgal, V., Vashisth, A., Datta, S., Mukherjee, J., and Singh, R. (2018). Effect of row direction and cultivar on micrometeorological and biophysical parameters of oil seed Brassica. *Journal of Agrometeorology,* **20**(special issue 1): 85-90.

Gromping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *American Statistical Association*, **63**(4): 308–19.

Gupta M., Sharma, C., Sharma, R., Gupta, R. and Khushu, M. K. (2017). Effect of sowing time on productivity and thermal utilization of mustard (*Brassica juncea*) under sub-tropical irrigated conditions of Jammu. *Journal of Agrometeorology*, **19** (2) : 137-141.

Gupta, S., Singh, A., Kumar, A., Shahi, U. P., Sinha, N. K. and Roy, S. (2018). Yield forecasting of wheat and mustard for western Uttar Pradesh using statistical model. *Journal of Agrometeorology*, **20**(1): 66-68.

Hagan, M. T. and Menhaj, M. (1994). Training feed forward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, **5(6)**: 989–993.

Haider, S. A., Alam, M. Z., Alam, M. F. and Paul, N. K. (2003). Influence of different sowing dates on the phenology and accumulated heat units in wheat. *Journal of Biological Sciences*, **3**: 932-939.

Hall, A. J., Connor, D. J. and Whitfield, D. M. (1989). Contribution of pre-anthesis assimilates to grain-filling in irrigated and water- stressed sunflower crops. I. Estimates using labelled carbon. *Field Crops Research*, **20**: 95-112.

Han, H., Li, Z., Ning, T., Zhang, X., Shan, Y. and Bai, M. (2008). Radiation use efficiency and yield of winter wheat under deficit irrigation in North China. *Plant Soil Environment,* **54**: 313–319.

Hao, Yu. and Wilamowski, B. M. (2011). Levenberg Marquardt Training, [web page] http://www.e ng.auburn.e du/~wi l ambm/ pa p/ 2011/K10149_C012.pdf.

Hatfield, J. L., Boote, K, J., Fay, P., Hahn, L., Izaurralde, R, C., Kimball, B, A., Mader, T., Morgan, J., Ort, D., Polley, W., Thomson, A. and Wolfe, D. (2008). Agriculture In: The Effects of Climate Change on Agriculture, Land Resources, Water Resources, and Biodiversity in the United States. *Chapter 6 Agriculture*.

Hatfield, J., Takle, E., Grotjahn, R., Holden, P., Izaurralde, R. C., Mader, T. and Liverman, D. (2014). The Effects of Climate Change on Agriculture, Land Resources, Water Resources, and Biodiversity in the United States. The Third National Climate Assessment, U.S. Global Change Research Program, 150-174.

Haykin, S. (2001). Neural Networks – A Comprehensive Foundation. IEEE Press, New York.

Hebbar, K. B., Venugopalan, M. V., Seshasai, M. V. R., Rao, K. V., Patil, B. C., Prakash, A. H., Kumar, V., Hebbar, K. R., Jeyakumar, P., Bandopadhyay, K. K., Rao, M. R. K., Khadi, B. M. and Aggarwal, P. K. (2008). Predicting cotton production using Infocrop-cotton simulation model, remote sensing and spatial agro-climatic data. *Current Science*, **95**: 1570–1579.

Hendrick, W.A. and Scholl. J. E. (1943).Technique in measuring joint relationship the joint effects of temperature and precipitation on crop yield. North Carolina Agriculture Experimental Station, Tech. Bull., 74.

Hijmans, R. J., Guiking-Lens, I. M. and Van Dipen, C. A. (1994). WOFOST 6.0: user's guide for the WOFOST 6.0 crop growth simulation model, Technical Document 12, DLO Winand Staring Centre, Wageningen, *The Netherlands,* 146 p.

Ho, S. L., Xie, M. and Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction, *Computers & Industrial Engineering*, **42**: 371–375.

Hodges, T., Johnson, S. L. and Johnson, B. S. (1992). A modular structure for crop simulation models: Implemented in SIMPOTATO model, *Agronomy Journal*, **84**: 911 - 915.

Hoogenboom, G., Jones, J. W., Wilkens, P. W., Porter, C. H., Boote, K. J., Hunt, L. A., Singh, U., Lizaso, J. L., White, J. W., Uryasev, O., Royce, F. S., Ogoshi, R., Gijsman, A. J., Tsuji, G. Y. and Koo, J. (2012). Decision Support System for Agrotechnology Transfer (DSSAT) Version4.5 [CD-ROM]. University of Hawaii, Honolulu, Hawaii.

Hoogenboom, G., White, J. W., Jones, J. W. and Boote, K. J. (1994). BEANGRO: A process oriented dry-bean model with a versatile user interface. *Agronomy Journal*, **86**: 182 - 190.

Horie, T., Yajima, M. and Nakagawa, H. (1992). Yield forecasting. *Agricultural and Forest. Meteorology*, **40**: 211 – 236.

Hsiao, C. and Wan, S. K. (2013). Is there an optimal forecast combination? *Journal of Econometrics*, **178**: 294–309.

Hund, L., Schroeder, B., Rumsey, K. and Huerta, G. (2018). Distinguishing between model- and data-driven inferences for high reliability statistical predictions. *Reliability Engineering and System Safety*, **180:** 201–210.

IBSNAT. (1984). Experimental design and data collection procedures for IBSNAT. The Minimum Data Set for Systems Analysis and Crop Simulation, first ed. IBSNAT Technical Report no 1 http://pdf.usaid.gov/pdf_docs/PNABK919.pdf.

Iwata F (1984). Heat unit concept of crop maturity, Physiological aspects of dry land farming, Gupta U.S. Eds, Oxford and IBH publishers, New Delhi, 351–370.

Jackson, B. S., Arkin, G. F., and Hearn, A. B. (1988). The cotton simulation model "COTTAM": Fruiting model calibration and testing. *Transactions of the ASAE*, **31**: 846 - 854.

Jamieson, P. D., Porter, J. R. and Wilson, D. R. (1991). A test of computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. *Field Crops Reaserch*, **27**: 337–350.

Jamieson, P. D., Semenov, M. A., Brooking, I. R. and Francis, G. S. (1998). SIRIUS: A· mechanistic model of wheat response to environmental variation. *European Journal of Agronomy*, **8 (3-4):** 161-179.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K. and Butler, E. E. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLoSONE*, **11**(6): 156571.

Jin, Z., Zhuang, Q., Tan, Z., Dukes, J. S., Zheng, B. and Melillo, J. M.(2016). Do maize models capture the impacts of heat and drought stresses on yield? Using algorithm ensembles to identify successful approaches. *Global Change Biology*, 22(9): 3112-26.

Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression. *Applied Statistics*, **31(**3): 300-303.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M. and Baret, F. (2004). Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, **121**(1-2): 19–35.

Jones, C. A. and Kiniry, J. R. (1986). CERES-Maize: A Simulation Model of Maize Growth and Development. Texas A&M University Press. *College Station, Texas*.

Jones, C. A., Wegener, M. K., Russell, J. S., McLeod, I. M. and Williams, J. R. (1989). AUSCANE: Simulation of Australian sugar cane with EPIC, Technical Paper. 29, Division of Tropical Crops and Pastures, CSIRO, Australia.

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., and Keating, B. A. (2017). Brief history of agricultural systems modeling. *Agricultural systems*, **155**: 240-254.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A. and Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of agronomy*, **18**(3-4): 235-265.

Jones, J. W., Keating, B. A. and Porter, C. H. (2001). Approaches to modular model development. *Agricultural Systems*, **70**(2-3): 421-443.

Jones, M. B., Leafe, E. L. and Stile, W. (1980). Water stress in field grown perennial rye grass, Its effect on growth, canopy photosynthesis and transpiration. *Annals of Applied Biology,* **2**: 87-101.

Jun, Z., Wancang, S., Yaxiong, M., Huiling, F., Yue, G., Jian, Y. And Jianmin, D. (2007). The effect of sowing date on plant survive rate and seed yield of winter rapeseed (*B. rapa*) in Northwest China. In: *Proceedings of 12th International rapeseed congress* Wuhan, China, March 26-30, **III**, pp. 186-187.

Kahle, K. W., Long Jr., J. H. and Todd, T. C. (2003). Long-term crop rotations affect soybean yield, seed weight, and soil chemical properties. *Field Crops Research* **83**: 41-50.

Kandiannan, K., Chandaragiri, K. K., Sankaran, N., Balasubramanian, T. N. and Kailasam, C. (2002). Empirical model for turmeric (*Curcuma domestica* Val.) yield prediction. *Journal of Spices and Aromatic Crops,* **10** (1): 59-61.

Kanth, R. H., Shah, M. H. and Bali, A. S. (2000). Effect of different sowing dates on heat-unit requirement for different phenophases of brown sarson (*Brassica campestris*) varieties and dynamics of aphid (*Lipaphis erysimi*) population. *Indian Journal of Agronomy,* **45**: 170-173.

Kar, G. and Chakravarty, N.V.K. (2001). Biomass partitioning and crop water requirements of *Brassica* as influenced by sowing dates and cultivars. *Annals of Agricultural Research,* **22**(1): 49-52.

Kar, G., and Kumar, A. (2015). Sowing dates and irrigation regimes on growth, productivity and radiation use efficiency of winter maize (*Zea mays*). *Indian Journal of Soil Conservation*, **43**(3): 255-261.

Karimi, Y., Prasher, S. O., Madani, A. and Kim, S. (2008). Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. *Canadian Biosystems Engineering*, **50**: 7.14-7.20.

Kazmi, D. H. and Rasul, G. (2012) Agrometeorological wheat yield prediction in rainfed Potohar region of Pakistan. *Agricultural Sciences*, **3**(2): 170-177.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., and McLean, G. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of agronomy*, **18**(3-4): 267-288.

Keerthi, P., Singh, R., Dhaka, A. K., and Divesh, C. (2017). Growth and yield prediction of Indian mustard using InfoCrop model at Hisar, Haryana. *Journal of Agrometeorology*, **19**(3): 259-261.

Kipkorir, E. C., Raes, D., Bargerei, R. J. and Mugalavai, E. M. (2007). Evaluation of two risk assessment methods for sowingmaize in Kenya. *Agricutural and Forest Meteorology*, **144**: 193-199.

Krishnamurthy, L. and Bhatnagar, V. B. (1998). Growth analysis of rainfed mustard *(Brassica juncea* (L.) Coss and Czerh.) cv. Varuna. Crop Res. Hissar, **15** (1) : 43-53.

Krishnan, P., and Aggarwal, P. (2018). Global sensitivity and uncertainty analyses of a web based crop simulation model (web InfoCrop wheat) for soil parameters. *Plant and soil*, **423**(1-2): 443-463.

Krishnan, P., Sharma, R. K., Das, A., Kukreja, A., Srivastava, R., Singhal, R. J., Bandyopadhyay, K. K., Khajanchi, L., Manjaiah, K. M., Chhokar, R. S. and Gill, S. C. (2016). Web-based crop model: web InfoCrop – wheat to simulate the growth and yield of wheat. *Computer Electronics in Agriculture*, **127**:324–335.

Kropff, M. J., Vclnlaar, H. H. and Mathews, A. B. (1994). ORYZA-1: An eco-physiological model for irrigated rice production, SARP Research Proceedings, IRRI, Los Banos, Philippines, 110 pp.

Kuhn, M. (2008). Building predictive models in R using caret package. *Journal of Statistical Software*, **28**: 1–6.

Kumar, A., Pandey, V., Shekh, A. M. and Kumar, M. (2008). Radiation use efficiency and weather parameter influence during life cycle of soybean (*Glycine max* L.) production as well accumulation of dry matter. *American-Eurasian Journal of Agronomy*, **1:** 41–44.

Kumar, A., Tripathi, P., Yadav, S. B., Singh, K. K. and Mishra, S. R. (2015). Validation of InfoCrop model for rice cultivars under eastern plain zone of Utter Pradesh. *Journal of Agrometeorology*, **17** (1): 80-83.

Kumar, R., Nagamani, K. and Babu, G. A. (2020). A Framework of Dimensionality Reduction Utilizing PCA for Neural Network Prediction. *Advances in Data Science and Management,* pp. 173-180.

Kumar, R., Singh, D. and Singh, H. (2000). Growth and yield of Brassica species as influenced by sulphur application and sowing dates. *Indian Journal of Agronomy*, **47** (3):417-421

Kumar, R., Singh, R. P. and Yeshpal, A. K. (2008). Yield and quality of *Brassica* species as influenced by different dates of sowing and varieties. *Pantnagar Journal of Research,* **6**(1): 6-11.

Kumar, S. N., Bai, K. V., Rajagopal, V. and Aggarwal, P. K. (2008). Simulating coconut growth, development and yield with the InfoCrop-coconut model. *Tree Physiology*, **28:** 1049–1058.

Kumar, Y., Kumar, R. S. A. and Kumar, S. (2017). Sensitivity analysis of InfoCrop model for Indian-mustard varieties in western Haryana. *Journal of Oilseed Brassica*, **81**(2): 175-180.

Kumari, A., Singh, R. P. and Yeshpal (2012). Productivity, nutrient uptake and economics of mustard hybrid (*Brassica juncea*) under different planting time and row spacing. *Indian Journal of Agronomy*, **57** (1): 61-67.

Kumari, P., Mishra, G. C. and Srivastava, C. P. (2013). Forecasting of Productivity and Pod Damage by *Helicoverpa armigera* using Artificial Neural Network Model in Pigeonpea (*Cajanuscajan). International Journal of Agriculturre, Enviroment & Biotechnology*, **6(2):** 187-193.

Kumari, P., Mishra, G. C. and Srivastava, C.P. (2016). Statistical models for forecasting pigeonpea yield in Varanasi region. *Journal of Agrometeorology,* **18**(18): 306–310.

Kumari, P., Mishra, G. C., Pant, A. K., Shukla, G. and Kujur, S. N. (2014a). Autoregressive Integrated Moving Average (ARIMA) Approach for Prediction of Rice (*Oryza Sativa* L.) yield in India. *The Bioscan*, **9**(3): 1063-1066.

Kumari, V. and Kumar, A. (2014). Forecasting of wheat (*Triticum aestivum*) yield using ordinal logistic regression. *The Indian Journal of Agricultural Sciences,* **84** (6): 691-694.

Kurmi, K. (2002). Influence of sowing date on the performance of rapeseed and mustard varieties under rainfed situation of Southern Assam. *Journal of Oilseeds Research*, **19**(2): 197–198.

Lajevardy, P., Parand, F. A. and Rahimi, H. (2015). A novel hybrid approach to improve the accuracy of load forecasting by IOWA. In *2015 International Aegean Conference on Electrical Machines & Power Electronics (ACEMP), 2015 International Conference on Optimization of Electrical & Electronic Equipment (OPTIM) & 2015 International Symposium on Advanced Electromechanical Motion Systems (ELECTROMOTION)* (pp. 500-505).

Lallu, R. S., Baghel, V. S. and Srivastava, S. B. L. (2010). Assessment of mustard genotypes for thermo tolerance at seed development stage. *Indian Journal of Plant Physiology*, **15**(1): 36-43.

Lawler, J. J., White, D., Neilson, R. P. and Blaustein, A. R. (2006). Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*, **12**(8):1568–84.

Laxmi, R. R. and Kumar, A. (2011). Weather based forecasting model for crops yield using neural network approach. *Statistics and Applications*, **9 (**1&2): 55-69.

Lenka, S., Singh, A. K. and Lenka, N. (2008). Water and nitrogen interaction on soil profile water extraction and ET in maize–wheat cropping system. *Agricultural Water Management* **96**: 195–207.

Li, Q. Q., Chen, Y. H., Liu, M. Y., Zhou, X. B., Yu, S. L. and Dong, B. D. (2008a). Effects of irrigation and planting patterns on radiation use efficiency and yield of winter wheat in North China. *Agricultural Water Management*, **95**: 469–476.

Lindquist, J. L., Arkebauer, T. J., Walters, D. T., Cassman, K. G. and Dobermann, A. (2005). Maize radiation use efficiency under optimal growth conditions. *Agronomy Journal* **97**: 72-78.

Liu, J., Pattey, E. and Admiral, S. (2013). Assessment of in situ crop LAI measurement using unidirectional view digital Photography. *Agricultural and Forest Meteorology*, **169**:25-34.

Liu, S., Wang, X., Liu, M. and Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Journal of LATEX Templates*, **1**: 48–56.

Liyong, Hu., Zhiquang, Lu. and Tingdong, Fu. (2007).The effects of sowing date on fatty acid synthesis of rapeseed (*B. napus*. L). In: *Proceeding of 12th International Rapeseed Congress*. Wuhan, China. March 26-30. **III**, pp. 80-84.

Lobell, D. B. Hammer, G. L., McLean, G., Messina, C., Roberts, M. J. and Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Natural Climate Change*, **3:** 1–5.

Lobell, D. B., and Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental research letters*, **2**(1): 1-7.

Lobell, D. B., Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, **150**: 1443–1452.

Lobell, D. B., Cassman, K. G. and Field, C. B. (2009). Crop yield gaps: their importance, magnitudes, and causes. *Annual review of environment and resources*, **34:** 179-204.

Lobell, D. B., Schlenker, W. and Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, **333**(6042): 616-620.

Long, S. P., Zhu, X., Naidu, S. L. and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant, Cell and Environment* **29**: 315-330

Major, D. J., Johnson, D. R., Tanner, J. W. and Anderson, I. C. (1975). Effect of day length and temperature on soybean development. *Crop Science*, **15**: 174-179.

Makridadis, S., Wheelwright, S. C. and Hyndman, R. J. (1998). Forecasting, Methods and Applications. New York, John Wiley and Sons.

Mandal, K. G., Hati, K. M., Misra, A. K. and Bandyopadhyay, K. K., (2006). Assessment of irrigation and nutrient effects on growth, yield and water use efficiency of Indian mustard (*Brassica juncea*) in central India. *Agricultural Water Management*, **85**: 279-286.

Marani, A. and Baker, D. N. (1981). Development of a predictive dynamic simulation model of growth and yield of Acala Cotton, US-Israel Binational Science Foundation Proposal No. 1708/78, Technical Research Report, 177 pp.

McCarl, B. A. (1984). Model validation: an overview with some emphasis on risk models. *Review of Marketing and Agricultural Economics*, **52**: 153-173.

McClung, C. R. and Davis, S. J. (2010). Ambient thermometers in plants: from physiological outputs towards mechanisms of thermal sensing. *Current Biology*, **20**:1086-1092.

McCown, R. L., Hammer, G. L., Hargreaves, J. N. G., Holzworth, D. P. and Freebairn, D. M. (1996). APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. *Agricultural systems*, **50**(3): 255-271.

Mccree, K. J. and Davis, S. D. (1974). Effect of water stress and temperature on leaf epidermal cells in grain sorghum. *Crop Sciences*, **5**: 751-755.

Meena, P. K., Khare, D. and Nema, M. K. (2016). Constructing the downscale recipitation using ANN model over the Kshipra river basin, Madhya Pradesh, *Journal of Agrometeorology*,**18** (1): 113-119.

Micheletto, S., Rodriguez-uribe, L., Hernandez, R., Richins, R. D., Curry, J. and Connell, M. A. O. (2007). Comparative transcript profiling in roots of Phaseolus acutifolius and P. vulgaris under water deficit stress. *Plant Science,* **173**: 510–520.

Mishra, G. C. and Singh, A. (2013). A study on forecasting price of groundnut oil in Delhi by ARIMA Methodology and Artificial Neural Network. *AGRIS On-line Papers in Economics and Informatics*, **5(3):** 25-34.

Monteith, J. L. (1994).Validity of the correlation between intercepted radiation and biomass. *Agricultural and Forest Meteorology* **68**: 213-220.

Monteith, J. L., Huda, A. K. S. and Midya, D. (1989). RESCAP: A resource capture model for sorghum and pearl millet. *Modeling the Growth and Development of Sorghum and Pearl Millet,* **12**: 30-34.

Muchow, R. C., Robertson, M. J. and Pengelly, B. C. (1993). Radiation-use efficiency of soybean, mugbean and cowpea under different environmental conditions. *Field Crops Research* **32**: 1-16.

Mutsaers, H. J. W. (1984). "KUTIN": A morphological model for cotton (*Gossypium hirsutum* L.). *Agriculture System*, **14**: 229 - 257.

Nain, A.S., Dadhwal, V.K., Singh, T.P., 2004. Use of CERES-Wheat model for wheat yield forecast in central Indo-Gangetic Plains of India. *The Journal of Agricultural Science (Cambridge),* **42** (1): 59-70.

Nanda, R., Bhargava, S. C., Tomar, D. P. S. and Rawson, H. M. (1996). Phenological development of *Brassica campestris*, *B. juncea, B. napus* and *B. carinata* grown in controlled environments and from 14 sowing dates in field. *Field Crops Research*, **46**(1-3): 93-103.

Narayanan, S., Aiken, R.M., Vara Prasad, P.V., Xin, Z. and Yu, J. (2013). Water and radiation use efficiencies in sorghum. *Journal of Agronomy and Crop Science*, **105**: 649–656.

NASS, (2006). The yield forecasting program of NASS by the Statistical Methods Branch, Estimates Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. NASS Staff Report No. SMB 06-01.

Nazeri, P., Shirani Rad, A. H., ValadAbadi, S. A., Mirakhori, M. and Hadidi, M. E. (2018). Effect of sowing dates and late season water deficit stress on quantitative and qualitative traits of canola cultivars. *Outlook on Agriculture*, **47**(4): 291-297.

Neogi, P., Chakravarty, N. V. K., Srivastava, A. K., Goutom, B., Katiyar, R. K. and Singh, H. B. (2005). Thermal time and its relationship with seed yield and oil productivity in *Brassica* cultivars. *International Journal of Brassica*, **7**(1 & 2): 63-70.

Ng, E. and Loomis, R. S. (1984). Simulation of growth and yield of the potato crop, Simulation Monographs Series, PUDOC, Wageningen, *The Netherlands*.

Nishad, R. (2017). Relation of micrometeorological parameters with biophysical characters and aphid infestation in mustard under temperature stress. M.Sc. Thesis, I.A.R.I., New Delhi.

Olden, J. D., Lawler, J. J. and Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, **83**(2):171–193.

Oppenheimer, H. R. (1960). Adaptation to drought, xerophytisim. In "Plant water relationships in Arid and Semi-arid conditions. Proc. Marid syp. UNESCO. Paris, pp.105-138.

Ozer, H. (2003). Sowing data and nitrogen rate effects on growth, yield and yield components of two summer rapeseed cultivars. *European Journal of Agronomy*, **19**: 453-463.

Pal, M. (2009). Kernel methods in remote sensing: A review. *ISH Journal of Hydraulic Engineering*, **15**: 194–215.

Pal, S., Singh, D. and Rao, V. (1996). Yield parameter of mustard cultivars under various environments. *Annals of Biology*, **12**(2): 356-360.

Palanivel, K. and Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, **10(**3): 110-118.

Panda, B. B., Bandyopadhyay, S. K. and Shivay, Y. S. (2004). Effect of irrigation level, sowing dates and varieties on yield attributes, yield, consumptive water use and water-use efficiency of Indian mustard (*Brassica juncea*). *Indian Journal of Agricultural Sciences*, **74**(6): 339-342.

Pandey, K. K., Rai, V. N., Sisodia, B. V. S. and Singh, S. K. (2015). Effect of weather variables on rice crop in eastern Uttar Pradesh, India. *Plant Archives*, **15**: 575–579.

Pandey, P. C., Dadhwal, V. K., Sahai, B., and Kale, P. P. (1992). An optimal estimation technique for increasing the accuracy of crop forecasts by combining remotely sensed and conventional forecast results. *International Journal of Remote Sensing*, **13**(14): 2735-2741.

Patel, H. R., Lunagaria, M. M., Karade, B. L., Vyas, P., Yadav, S. B., Shah, A. V., Rao,V. U. M. and Kumar, N. S. (2013). Impact of projected climate change on groundnut in Gujrat. *Journal of Agrometeorology*, **(Special issue)**: 81-84.

Patel, K. G., Rathore, S. S., Shiyani, G. L. and Raman, S. (1994). Response of different mustard varieties to dates of sowing and levels of nitrogen in South Gujarat. *Journal of Madras Agriculture*, **81**(6): 348-349.

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S. and Cassman, K. G. (2004) Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences,* **101**(27): 9971-75.

Penning de Vries, F.W.T., Jansen, D. M., Ten Berge, H. F. M. and Bakema, A. (1989). *Simulation of ecophysiological processes of growth in several annual crops* (Vol. 29). International Rice Research Institute, Phillipines.

Penning de Vries, F.W.T., van Laar, H.H., Kropff, M.J. (Eds.), 1991. Simulation and systems analysis for rice production (SARP) 1991. PUDOC, Waneningen, *The Netherlands*, pp. 369.

Petr, J. (1991). Weather and Yield. *Developments in Crop Science No. 20*. Amsterdam, Elsevier. pp- 288.

Plenet, D., Mollier, A. and Pellerin, S. (2000). Growth analysis of maize field crops under phosphorus deficiency. II. Radiation-use efficiency, biomass accumulation and yield components. *Plant and Soil* **224** (2): 259-272.

Porter, J. A. (1993). AFRCWHEAT2: A model for the growth and development of wheat incorporating responses to water and nitrogen, *European Journal of Agronomy*., **2**(2): 69 - 82.

Poureisa, M. and Nabipour, M. (2007). Effect of planting dates on canola phenology, yield and yield components. In: *Proceedings of 12th International rapeseed congress*. Wuhan, China. 26-30th March, 2007. **III**, 97-101.

Pradhan, S., Sehgal, V. K., Bandyopadhyay, K. K., Panigrahi, P., Parihar, C. M. and Jat, S. L. (2018). Radiation interception, extinction coefficient and use efficiency of wheat crop at various irrigation and nitrogen levels in a semiarid location. *Indian Journal of Plant Physiology*, **23**: 416-25.

Pradhan, S., Sehgal, V.K., Das, D.K., Jain, A.K., Bandyopadhyay, K.K., Singh, R. and Sharma, P.K. (2014). Effect of weather on seed yield and radiation and water use efficiency of mustard cultivars in a semi-arid environment. *Agricultural and Water management*, **139**: 43-52.

Prakash, A. H., Bandyopadhyay, K. K., Gopalakrishnan, N. and Kumar, A.A. (2010). Interaction of thermal time and nitrogen levels on growth and productivity of cotton (*Gossypium hirsutum*). *Indian Journal of Agricultural Sciences*, **80**: 704–709.

Pritchard, F. M., Eagles, H. A., Norton, R. M., Salisbury, S. A. and Nicolas, M. (2000). Environmental effects on seed composition of Victoriana canola. *Australian Journal of Experimental Agriculture*, **40**: 679-685.

Quiring, S. M., and Legates, D. R. (2008). Application of CERES-Maize for within-season prediction of rainfed corn yields in Delaware, USA. *Agricutural and Forest Meteorology*, **148**(6-7): 964-975.

Rajegowda, M. B., Soumya, D. V., Padmashri, H. S., Janardhana Gowda, N. A. and Nagesha, L. (2014). Ragi and groundnut yield forecasting in Karnataka–statistical model. *Journal of Agrometeorology*. **16** (2): 203-206.

Rajput, R. P. (1980). Response of soybean crop to climate and soil environments. Ph D thesis. Indian Agricultural Research Institute, Pusa, New Delhi.

Rane, J., Pannu, R, K., Sohu, V,S., Saini, R, S., Mishra, B., Shoran, J., Crossa, J., Vargas, M. and Joshi, A, K. (2017). Performance of yield and stability of advanced wheat genotypes under heat stress environments of the Indo-Gangetic plains. *Crop sciences*, **47** (4): 1561-73.

Ranganathan, A., (2004). The Levenberg-Marquardt Algorithm. http://ananth.in/docs/lmtut.pdf.

Rao, T. R. (1992). Growth and development of mustard and pearl millet in relation·· to light interception and photosynthesis and modelling for Brassica spp., Ph.D. thesis, Division of Agricultural Physics, Post Graduate School, Indian Agricultural Research Institute, New Delhi, India.

Rao, Y. V. S., Garg, R. N. and Das, D. K. (1999). Prediction of wheat yield under different management practices using remote sensing parameters, *Annals of Agricultural Research*, **20**(2): 202-205.

Rickman, R. W., Waldman, S. E. and Klepper, S. L. (1996). MODWht3: A development driven winter wheat growth simulation model. *Agronomy Journal*, **86**: 325 - 331.

Ritchie, J. T. (1985). Description and performance of CERES wheat: A user-oriented wheat yield model. *ARS wheat yield project*, 159-175.

Ritchie, J. T. and Alagarswamy, G. (1989). Simulation of sorghum and pearl millet phenology (eds), *Modelling the growth and development of sorghum and pearl millet, Research Bulletin No. 12*, ICRISAT, Patancheru, India, pp. 2426.

Ritchie, J. T. and Otter, S. (1984). Description and performance of CERES-wheat: a user-oriented wheat yield model. In: Wheat Yield Project, A.R.S. (Ed.), ARS-38. National Technical Information Service, Springfield, Missouri, pp. 159–175.

Ritchie, J. T., Alocilija, E. C. and Uehara, G. (1986). *IBSNAT/CERES Rice Model Agro-technology Transfer*, **3**:1-5.

Robertson, M. J., Holland, J. F. and Bambach, R. (2004). Response of canola and Mustard to sowing date in the grain belt of north-eastern Australia. *Australian Journal of Experimental Agriculture*, **44**: 43-52.

Robertson, M. J., Holland, J. F., Bambach, R. and Cawthray, S. (2002). Response of canola and Indian mustard to sowing date in risky Australian environments. *Australian Journal of Agricultural Research*, **53**(10): 1155- 1164.

Rosenthal, W. D., Vanderlip, R. L., Jackson, B. S. and Arkin, G. F. (1989). SORKAM: A grain sorghum crop growth model, Texas Agriculture Experiment Station, Texas A&M University, College Station, Computer Software documentation services, MP 1669, 205 pp.

Roy, S., Meena, R. L., Sharma, K. C., Kumar, V., Chattopadhyay, C., Khan, S. A. and Chakravarthy, N. V. K. (2005). Thermal requirement of oilseed *Brassica* cultivar at different phenological stages under varying environmental conditions. *Indian Journal of Agricultural Sciences,* **75:** 717–721.

Sani, B, M., Oluwasemire, K, O. and Mohammed, H, I. (2008). Effect of irrigation and plant density on the growth, yield and water use efficiency of early maize in the Nigerian Savanna. *ARPN Journal of Agricultural and Biological Science,* **3**(2): 33-40.

Schirrmann, M., Hamdorf, A., Giebel, A., Dammer, K. H., and Garz, A. (2015). A mobile sensor for leaf area index estimation from canopy light transmittance in wheat crops. *Biosystems Engineering*, **140**: 23-33

Sehgal, A., Sita, A., Kadambot, H. M. S., Kumar, Rakesh, Sailaja, B. and Varshney, R. K. (2018). Drought or/and Heat-Stress Effects on Seed Filling in Food Crops: Impacts on Functional Biochemistry, Seed Yields, and Nutritional Quality. *Frontiers in Plant Science,* **9**: 1–19.

Shabnam, Bansal, S. K. and Dabas D.S. (2013). Use of time-series data of temperature and yield to assess the impact of climate change on crop yield using mustard in Haryana as an Example. *International Research Journal of Social Sciences,***2**(4): 31-33.

Sharma, S. K., Mendhe, S. N., Kotle, H. S., Rajput G. R. and Yenpreddiwar, M. D (2006): Effect of sowing and irrigation management on growth and yield of mustard (*Brassica juncea*). *Journal of soil and crop*, **16**: 455-459.

Sharma, V., Rai, S. and Dev, A. (2012). A Comprehensive Study of Artificial Neural Networks. *International Journal of Advance Research in Computer Science & Software Engineering*, **2** (10): 278-284.

Shi, W., Tao, F. and Zhang, Z. (2013). A review on statistical models for identifying climate contributions to crop yields. *Journal of Geographical Sciences*, **23**:567–576.

Sinclair, T.R., Muchow, R.C., 1999. Radiation use efficiency. *Advances in Agronomy* **65:** 215–265.

Singh, J. P., Govindakrishnan, P. M., Lal, S. S. and Aggarwal, P. K. (2005). Increasing the efficiency of agronomic experiments in potato using Infocrop-potato model. *Potato Research*, **48**: 131–152.

Singh, K. K., Baxla, A. K., Chattopadhyay, N., Balasubramanian, R., Singh, P. K., Rana, M., BorpatraGohain, G., Vishnoi, L. and Singh, P. (2017). Crop Yield forecasting under FASAL (Forecasting Agricultural output using Space Agrometeorology and Land based observations). FASAL *Technical Report-2017*, pp 1-52.

Singh, P. and Virmani, S. M. (1996). Modelling growth and yield of chickpea (*Cicer arietinum* L.). *Field Crops Research*, **46**: 41-59.

Singh, R. S., Patel, C., Yadav, M. K. and Singh, K. K. (2014). Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *Journal of Agrometeorology*, **16**:199–202.

Singh, R., Patidar, M. and Singh, B. (2001). Response of Indian mustard (Brassica juncea) cultivars to different sowing time. *Indian Journal of Agronomy*, **46**(2): 292- 295.

Singh, R., Rao, V. U. M. and Singh, D. (2002). Biomass partitioning in *Brassica* as affected by sowing dates. *Journal of Agrometeorology*, **4**(1): 59-63.

Singh, S. K. and Singh, G. (2002). Response of Indian mustard (*Brassica juncea*) varieties nitrogen under varying sowing dates in eastern Uttar Pradesh. *Indian Journal of Agronomy*, **47** (2): 242-248.

Smola, A. J. and Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, **14**: 199–222.

Song, H., Zhang, R., Zhang, Y., Xia, F. and Miao, Q. (2011). Energy consumption combination forecast of Hebei province based on the IOWA operator. *Energy Procedia*, **5:** 2224-2229.

Spitters, C. J. T. (1987). An analysis of variation in yield among potato cultivars in terms of light absorption, light utilization and dry matter partitioning. *Agrometeorology of the Potato Crop*, **214**: 71-84.

Stapper, M., and Arkin, G. F. (1980). CORNF: A dynamic growth and development model for the Maize (*Zea mays* L.), Texas Experiment Station, Program and Model Documentation No. 80-2, College Station, TX, USA.

Stockle, C. O., Van Evert, F., Nelson, R. and Campbell, G. S. (1991). CropSyst, a cropping systems model with GIS links, Agronomy Abstracts, *American Society of Agronomy*, Madison, WI, USA.

Stone, P. (2001). The effects of heat stress on cereal yield and quality. In AS Basra, ed, Crop Responses and Adaptations to Temperature Stress. Food Products Press, Binghamton, NewYork, pp 243–291.

Su, Y. X., Xu, H. and Yan, L. J. (2017). Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi Journal of Biological Science*, **24**(3):537–547.

Surekha, E. and Reddy, R. G., (1996). Performance of mustard varieties under different date of sowing. *Journal of Oilseeds Research*, **13**(1): 70-73.

Suzuki, M., Shibahara, T. and Muragaki, Y. (2020). A Method to Extract Feature Variables Contributed in Nonlinear Machine Learning Prediction. *Methods of Information in Medicine*, **59**(1):1-8.

Thomas, P. (2013). Enhancing water productivity in wheat by optimization of irrigation management using ‘AquaCrop’ model. M.Sc. Thesis, I.A.R.I., New Delhi.

Thurling, N. (1974). Morphological determinants of yield of rapeseed and mustard (*Brassica campestris* and *B. juncea*). Growth and morphological character. *Australian Journal of Agricultural Research*, **25**(5): 692-710.

Thurling, N. and Das, L. D. V. (1980). The relationship between pre-anthesis development and seed wild of spring rape (*Brassica napus* L.) Australian. *Journal of Agricultural Research*, **31**:25-36.

Tobe, A., Hokmalipour, S., Jafarzadeh, B. and Darbandi, M. H. (2013). Effect of sowing date on some phenological stages and oil contents in spring canola (*Brassica napus*, L.) Cultivars. *Middle-East Journal of Scientific Research*, **13**: 1202–1212.

Todorovic, M., Albrizio, R., Zivotic, L., Saab, M. T. A., Stöckle, C., and Steduto, P. (2009). Assessment of AquaCrop, CropSyst, and WOFOST models in the simulation of sunflower growth under different water regimes. *Agronomy Journal*, **101**(3): 509-521.

Tomar, R. K. (1997). Potentiality of Indian mustard under late sowed condition in Bundelkhand zone of Madhya Pradesh. *Advance Agricultural Research in India*, **8**: 81-85.

Tsuji, G. Y., Hoogenboom, G., and Thornton, P. K. (Eds.). (1998). *Understanding options for agricultural production* (Vol. 7). Springer Science & Business Media.

Tulbure, M. G., Wimberly, M. C., Boe, A. and Owens, V. N. (2012). Climatic and genetic controls of yields of switch grass, a model bioenergy species. *Agriculture, Ecosystems & Environment*, **146**(1): 121-129.

Turhan, H., Gul, K. M., Egesel, C. O. and Kahriman, F. (2011). Effect of sowing time on grain yield, oil content, and fatty acids in rapeseed (*Brassica napus* subsp. oleifera). *Turkish Journal of Agriculture and Forestry,* **35**(3): 225-34.

Tuteja, S. S., Lakpale, R. and Tripathi, R. S. (1996). Effect of date of sowing, nitrogen levels and nitrogen splitting on mustard (*Brassica juncea*). *Advance in plant science,* **9**(1): 167-168.

Uehara, G. and Tsuji, G.Y. (1998). Overview of IBSNAT. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Chapter 1, pp. 1–7. Understanding Options for Agricultural Production.Springer, *The Netherlands*.

Umburanas, R. C., Yokoyama, A. H., Balena, L., Dourado-Neto, D., Teixeira, W. F., Zito, R. K., Reichardt, K., and Kawakami, J. (2019). Soybean yield in different sowing dates and seeding rates in a subtropical environment. *International Journal of Plant Production*, **13**:117-128.

Ustuner, M., Sanli, F. B. and Dixon, B. (2015). Application of support vector machines for land use classification using high-resolution rapid eye images: A sensitivity analysis. *European Journal of Remote Sensing*, **48:** 403–422.

Van den Berg, M. and Driessen, P. M. (2002). Water uptake in crop growth models for land use systems analysis: I. A review of approaches and their pedigrees. *Agriculture, ecosystems & environment*, **92**(1): 21-36.

Van Diepen, C. A., Wolf, J., Van Keulen, H. and Rappoldt, C. (1989). WOFOST: A simulation model of crop production. *Soil Use and Management*, **5**: 16 - 24.

Van Keulen, H. and Seligman, N. G. (1987). Simulation of water use, nitrogen nutrition and growth of a spring wheat crop, Simulation Monographs, PUDOC, Wageningen. *The Netherlands*, 310 pp.

Van Keulen, H., Penning de Vries, F.W.T. and Drees, E. M. (1982). A summary model for crop growth, In: Penning de Vries, F.W.T. and van Laar, H.H. (eds.), Simulation of plant growth and crop production, Simulation Monographs, PUDOC, Wageningen, *The Netherlands*, pp. 87 - 97.

Van Kraalingen, D.W.G. (1995). The FSE system for crop simulation, version 2.1 (No. 1). ABDLO.

Vapnik V. (1998) The Support Vector Method of Function Estimation. In: Suykens J. A. K., Vandewalle J. (eds) Nonlinear Modeling. Springer, Boston, MA. pp 55-85. https://doi.org/10.1007/978-1-4615-5703-6_3

Vashisth, A., Chakravarty, N. V. K., Goutam, B. and Sharma, P.(2011). Effect of weather variability on growth characteristics of brassica crop. In book: Challenges and Opportunities in Agrometeorology, 519-28.

Vashisth, A., Choudary, M., Joshi, D. K. and Baloda, R. (2015). Pre harvest yield forecast in maize-mustard cropping system under semi arid region using crop simulation model. *International Journal of Tropical Research*, **33**(2):1059-1067.

Vashisth, A., Goyal, A. and Roy, D. (2018). Pre harvest maize crop yield forecast at different growth stage using different model under semi arid region of India. *International Journal of Tropical Agriculture*, **36**(4): 915-920.

Vashisth, A., Krishanan, P. and. Joshi, D. K. (2019). Multi Stage Wheat Yield Estimation Using Different Model Under Semi Arid Region Of India. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-3/W6, 2019 ISPRS-GEOGLAM-ISRS Joint Int. Workshop on “Earth Observations for Agricultural Monitoring”, New Delhi, India.

Vashisth, A., Singh, R. and Choudary, M. (2014). Crop yield forecast at different growth stage of wheat crop using statistical model under semi-arid region. *Journal of Agroecoogy and Natural Resource Management*, **1**(1): 1-3.

Verma ,U., Piepho, H. P. and Goyal, A. (2016). Role of climatic variables and crop condition term for mustard yield prediction in Haryana. *International Journal of Agricultural and Statistics Sciences,* **12**: 45–51.

Verma, U. (2015). Linear mixed modeling for mustard yield prediction in Haryana state (India). *Journal of Mathematics and Statistical Science,* **1** (3): 96-105.

Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G. A. and Torricelli, P. (2011). Application of a Random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy. *Ecological Model*, **222**(8):1471–1478.

Viscarra-Rossel, R.A., Walvoort, D. J. J., Mcbratney, A. B., Janik, L. J. and Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, **131**:59-75.

Wall, G. W., Amthor, J. S. and Kimball, B. A. (1994). COTC02: A cotton growth simulation model for global change. *Agricultural Forest Meteorology*, **70**: 289 342.

Wallach, D., Goffinet, B., Bergez, J. E., Debaeke, P., Leenhardt, D., and Aubertot, J. N. (2001). Parameter estimation for crop models. *Agronomy journal*, **93**(4): 757-766.

Waring, R. H., and Landsberg, J. J. (2011). Generalizing plant–water relations to landscapes. *Journal of Plant Ecology*, **4**(1-2): 101-113.

Warner, B. and Misra, M. (1996). Understanding neural networks as statistical tools. *The American Statistician*, **50**(4): 284-293.

Weerakoon, S. R. and Somaratne, S. (2011). Seasonal variation of growth and yield performance of mustard [*Brassica juncea* (L.) Czern and Coss)] genotypes in Sri Lanka. *Journal of Agricultural Sciences,* **6**(1): 1-15.

Weir, A. H., Bragg, P. L., Porter, J. R., and Rayner, J. H. (1984). A winter wheat crop simulation model without water or nutrient limitations. *The Journal of Agricultural Science*, **102**(2): 371-382.

Welles, J. M. and Norman, J. M. (1991). Instrument for indirect measurement of canopy architecture. *Agronomy Journal*, **83**: 818-25.

Were, K., Bui, D. T., Dick, O. P. and Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators,* **52:** 394–403.

Wilhelm, W. W., McMaster, G. S., Rickman, R. W. and Klepper, B. (1993). Above-ground vegetative development and growth of winter wheat as influenced by nitrogen and water availability. *Ecological Modelling*, **68**(3-4): 183-203.

Wilkerson, G. G., Jones, J. W., Boote, K. J., Ingram, K. T. and Mishoe, J. W. (1983). Modeling soybean growth for crop management. *Transactions of the ASAE,* **26**(1): 63-73.

Williams, J. R., Jones, C. A., Kiniry, J. R., and Spanel, D. A. (1989). The EPIC crop growth model. *Transactions of the ASAE,* **32**(2): 497-511.

Xing, L., Li, L., Gong, J., Ren, C., Liu, J. and Chen, H. (2018). Daily soil temperatures predictions for various climates in united-states using data-driven model. *Energy,* **160**: 30–440.

Xiong, W., Matthews, R., Holman, I., Lin, E. and Xu, Y. (2007). Modelling China's potential maize production at regional scale under climate change. *Climatic change*, **85**(3-4): 433-451.

Yang, H.S., Dobermann, A., Lindquist, J.L., Walters, D.T., Arkebauer, T.J. and Cassman, K.G. (2004). Hybrid-maize-a maize simulation model that combines two crop modeling approaches. *Field Crops Research,* **87**: 131-154.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of forecasting*, **14**(1): 35-62.

Zhang, T., Zhu, J. and Yang, X. (2008). Non-stationary thermal time accumulation reduces the predictability of climate change effects on agriculture. *Agricultural and Forest Meteorology,* **148**(10): 1412–1418.

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of Translational Medicine,* **4**(7): 126.