मशीन लर्निंग के उपयोग से मौसम आधारित गेहूं की उपज का पूर्वानुमान

WEATHER BASED WHEAT YIELD PREDICTION USING MACHINE LEARNING

SHREYA GUPTA
Roll No. 21088

DIVISION OF AGRICULTURAL PHYSICS
ICAR- INDIAN AGRICULTURAL RESEARCH INSTITUTE
NEW DELHI - 110012

2020

Weather based wheat yield prediction using machine learning

A Thesis

By

Shreya Gupta

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

MASTER OF SCIENCE (AG.)
IN
Agricultural Physics

2020

Approved by:

Chairman : Dr. Ananta Vashisth

Co-chairman : Dr. P. Krishnan

Members : Dr. Shiv Prasad Prasad

: Dr. Achal Lama अचल लामा

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

Dr. Ananta Vashisth
Principal Scientist
ananta.iari@gmail.com

CERTIFICATE

This is to certify that the thesis entitled, **“Weather based wheat yield prediction using machine learning”** submitted to the **Faculty of the Post-Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi,** in partial fulfillment of the requirements for the award of **MASTER OF SCIENCE (AG.) IN AGRICULTURAL PHYSICS,** embodies the results of *bona fide* research work carried out by **Ms. Shreya Gupta, Roll No: 21088,** under my guidance and supervision and that no part of this thesis has been submitted for any other degree or diploma.

The assistance and help availed during the course of investigation as well as source of information have been duly acknowledged by her.

Vashisth

Place: New Delhi **(Dr. Ananta Vashisth)**

Date: October 1, 2020 **Chairperson**
Advisory Committee

ACKNOWLEDGEMENT

As a prelude to my thanks giving, at first, I wish to thank the almighty for giving me strength, courage and confidence…. after all he is the "greatest".

I wish to express my deepest sense of gratitude and indebtedness to Dr. Ananta Vashisth, Principal Scientist, Division of Agricultural Physics, as Chairperson of my Advisory Committee for her invaluable guidance, constant encouragement, scholarly suggestions, untiring enthusiasm, affectionate behavior, 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, Division of Agricultural Physics and co-chairperson 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 gratitude to Dr. V.K. Sehgal, Professor, División 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. Shiv Prasad , CESCRA, Dr. Achal Lama, Indian Agricultural Statistics Research Institute, for his concrete suggestions and commendable help throughout the experimental period.

I am expressing my sincere thanks to Agromet Field Unit (AMFU) Ludhiana, Hisar and Regional Met Center Chandigarh for providing daily weather and wheat yield data.

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, in monitoring the field experiments and collecting in-situ observations. I am extending my sincere gratitude to technical, administrative and supportive staff members of the Division of Agricultural Physics, IARI especially to Mrs. Rekha mam SRF, Arjun sir for their support which has helped me to complete my study in the division.

I articulate my appreciation to all my seniors especially, Aravind sir, Koushik sir, Brijesh sir, Madan sir, Pragya ma'am, Bhavani sir, Shafeeq sir, Sunny sir, Debasish sir, Sujan sir, Vimal sir , Priya maam, Sona sir, Koushik Bag sir I express my profuse thanks to my batchmates Tridiv Ghosh, Nandita, Sonia.. I also extend my heartiest thanks to my juniors *especially Sailja, Krishna, Abradeep, Bibhuti, Tarun for their company during my research work.*

It gives me an immense pleasure to extend my sincere thanks to the entire scientists' community of Agricultural Physics Division & NRL. I gratefully acknowledge the valued opinion and help rendered by my seniors and candid thanks to all my friends and juniors.

The endless love, affection, sacrifice and constant inspiration from mummy, papa have enabled me to reach the footstep of my long-cherished aspiration. Thanks to all my family members for their love. Also a heartfelt thanks to my friends Preeti Yadav, Akanksha Singh, Harshita, Sparsh, Shubham Sambhav, Suraj, Anupam, Krishna, Pragati, Saloni and Prashasti

Finally, I humbly and whole heartedly bow my head before Lord Almighty for blessing me all his grace to successfully complete my two years study and research in IARI. The financial assistance provided by the IARI in the form of Fellowship during the tenure is duly acknowledged.

Shreya

Place: New Delhi *(Shreya Gupta)*

Date: 01.10.2020

CONTENTS

LIST OF TABLES

LIST OF FIGURES

26b	Validation of Models developed for wheat yield prediction at grain filling stage for Ludhiana
27a	Calibration of Models developed for wheat yield prediction at grain filling stage for Patiala.
27b	Validation of Models developed for wheat yield prediction at grain filling stage for Patiala

1. *Introduction*

Wheat (*Triticum aestivum*) is one of the principal food crops of the country. According to 4th advance estimates total **food grain** production of major crops during 2018-19, in the country is assessed as 284.95 million tonnes. Among the total food grain rice constitute 116.42 million tonnes while wheat has a production of 102.19 million tonnes, similiarly Production of **wheat** estimated as record of 102.19 million tonnes, which is higher by 2.32 million tonnes as compared to wheat production of 99.87 million tonnes achieved during 2017-18. Moreover, the production of wheat during 2018-19 is also higher by 7.58 million tonnes compared to average wheat production of 94.61 million tonnes which is higher by 17.62 million tonnes than the previous five years' (2013-14 to 2017-18) average production of food grain. Thus predicting wheat production is an important step which should be carried out with accuracy to enhance the policy scheme, management, storage, import, export etc. Reliable and well-timed forecasts provide important and beneficial input for foresighted and informed planning in agriculture which is full of uncertainties.

According to agricultural and processed food products development authority (APEDA) wheat cultivation in India traditionally been conquered by the northern region of India. The northern states of Punjab and Haryana Plains in India have been prolific wheat producers and occupy large production area of the country. So Major wheat growing states in India are Punjab, Uttar Pradesh, Haryana, Madhya Pradesh, Rajasthan, Bihar and Gujarat. With a production accomplishment of ten times in past five years, India as today is the second largest wheat producer in the whole world. V. Demand of India's wheat in the whole world shows an expanding trend. Our country has exported 2,26,225 MT of wheat to the world attaining the worth of Rs. 424.94 crores during the year 2018-19. Demand of wheat is increasing in global scale due its gluten protein adhesive property and vioelastic nature, which makes the production of processed food items. Wheat provide vital nutrients and health benefit to diet and in developing countries it is contributing 70% total calories in diet (CIMMYT). Thus today, India is exporting ample quantity of all types of wheat to many countries of the world. The crop growth, development and its yield are controlled by the influence of different weather elements, agronomical management, improved crop variety, proper recommended fertilizer dosage, tillage operation, judicious application of irrigation, climate change, temperature, solar radiation, precipitation and extreme weather (Martinez *et al*., 2009 Nagai and Makino, 2009 and Osborne *et al*., 2013).

. Recent research reports shows that inter-annual variations in crop yields are normally associated and linked with climate variability and weather patterns (Wassmann et al. 2009). Thus, the impacts of climatic variability and change on food and fibres are of foremost growing concern globally (Rimi et al.2011). Wheat production is significantly influenced and controlled by climatic factors such as rainfall, temperature, solar radiation, and relative humidity (Ji et al. 2007). Evapotranspiration (ET) is also one of the most important factor affecting the water balance of crop, in turn associated with the water stress or surplus. It primarily controls several hydrological processes and its accurate assessment provides valuable information for water resources planning and management in the crops life cycle (Tabari et al. 2012). For estimation of evapotranspiration number of empirical equations have been developed for estimating ET using meteorological data but the Penman-Monteith FAO-56 combination equation (PMF-56) has been recommended by the Food and Agriculture Organization of the United Nations (FAO) as the sole standard equation for estimating ET. The PMF-56 requires a number of climatic parameters such as daily maximum and minimum temperature, relative humidity, solar radiation, and wind speed. Weather variability within or maybe between the seasons is intense source of variability in yields. Weather variables affect the crop differently during different stages of development. Thus extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on their distribution pattern of weather over the full crop season. Thus there is need to develop statistically sound objective forecasts of crop yield based on weather variables so that reliable forecasts can be obtained. (Laxmi *et al.,* 2011)

Crop yield forecast may be done by using three major methods (i) biometrical characteristics (ii) weather variables and (iii) agricultural inputs. These methodologies can be used individually or in combination to give a composite model. (Agrawal *et al.* 2001). Several studies have been done to forecast crop yield using weather parameters (Huda *et al.* (1975), Choudhary and Sarkar (1981), Kokate *et al.* (2000), etc.). However such forecast studies based on statistical models need to be done on continuing basis and for different agro-climatic zones, due to visible effects of changing environmental conditions and weather shifts at different locations and areas. Forecasting of crop yield has a growing status to ensure food security, optimize agro-management practices and resource use for sustainable development. Crop yield monitoring and forecasting are conducted with several sources of information such as field observation, climate data, satellite images and crop growth simulation models. In this regard, weather and climate have the main role on crop yield and productions. Climate extremes with intensity and frequency increasing can have an effect on crop yield, security and safety. The climate variables changes can affect global socio-economic and agricultural systems. The climate variables, for example temperature and precipitation, serve as direct inputs of agriculture whose changes have a significant impact on crop yield and related variation. Crop growth and development are affected by climate variables in different times related to growth cycle. Therefore, there is a need to develop area specific forecast models based on time series data with the help of machine learning to predict more accurately. These forecasting models have been based on either linear regression or more sophisticated non-linear approaches including modern machine learning methods.

Machine Learning is well-thought-out as a subfield of Artificial Intelligence and it is concerned with the development of techniques which enable the computer to learn. In simple terms development of algorithms which enable the machine to learn to perform tasks and activities. Over the period of time many techniques and methodologies have been developed for machine learning tasks. It can identify and learn correlated patterns between input data sets and corresponding target values through training. After training, it can be used to predict the outcome of new independent input data and have great capacity in predictive modelling, i.e. all the characters describing the situation can be presented to the trained machine learning models, and then prediction of agricultural system may be feasible. Machine learning algorithms are often classified as supervised and unsupervised where supervised machine learning algorithms can apply what has been learned in the past by training data sets to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors to modify the model accordingly. While the unsupervised machine learning algorithms are used when the information or the data used to train the system is neither classified nor is it labelled. Unsupervised learning studies can infer a function to describe a hidden structure from unlabelled data. The system doesn’t figure out the right output but it actually explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data. Various machine learning can include decision tree, artificial neural network, support vector machine etc.

Crop Advisor’ has been developed as a web page for predicting the influence of climatic parameters on the crop yield. Algorithm used to find out the most prompting climatic parameter on the crop yields of selected crops in some selected districts of Madhya Pradesh was C4.5. This provides a signal of relative influence of different climatic parameters on the crop yield. Other agro-input parameters responsible for crop yield were not considered in this tool. (*Veenadhari et. al* 2014.)Weather based crop yields estimate of pests and diseases were done by regression and ANN model. Weather based models were developed for premonition of important pests/diseases in various crops like rice, groundnut, mango, mustard, pigeon pea, sugarcane, potato and cotton at numerous locations using the practices like complex polynomials through GMDH technique, regression analysis (taking suitable functions of weather variables/indices as regressers), and Artificial Neural Network. (Agrawal and Mehta, 2001). Based on time series data of 27 years, the yield and the weather data rice yield forecast was done (w.e.f. 1981-82 to 2007-08) obtained from G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India, using adaptive neuro fuzzy inference system (ANFIS) technique. Thus developing crop yield forecast models to record relation between climatic data and crop yield (Pankaj Kumar, 2011).Under various weather conditions and management practices APSIM-maize model was used to capture the biomass and grain yield of summer maize in North China Plain. After calibration and validation, five scenarios were simulated using the APSIM model. The simulated results disclosed that weather factors including diurnal temperature range and the sunshine hours during the grain filling stage has positive effects on maize yield (Sun *et. al.* 2016). Das *et. al.* (2018) predicted rice yield using long-term weather data for west coast of India, using artificial neural network (ANN) merely and in grouping with principal components analysis (PCA), stepwise multiple linear regression (SMLR) and penalised regression models (like least absolute shrinkage and selection operator (LASSO) or elastic net (ENET) for rice yield prediction using long-term weather data. The result showed that R^2 and root mean square error (RMSE) of the models varied between 0.22–0.98 and 24.02–607.29 kg ha−1.

Yield is predicted mostly by statistical and crop simulation models but the use of machine learning is lagging behind for yield prediction. Estimation of multi stage wheat yield based on weather variables with machine learning have not been done in detail so far. Also use of hybrid ML (like SVR-LASSO, SVR MLR) model was not used so far for estimating multistage wheat yield.

Objectives

1. To develop and validate weather based models using machine learning for prediction of wheat yield
2. To predict wheat yield at multiple stages using the developed model

Agriculture is the most significant in Indian economy. More than 70 percent of rural livelihood depends on agricultural.. Among the numerous crops being cultivated in India wheat which is mostly the *Triticum aestivum* is one of the most important food crops of the country.. It accounts for the second most produced cereal after maize. Wheat is the most important *Rabi* cereal crop in North-West part of India.. The ideal climatic type required for wheat crop growth and development is semi-arid and sub-tropical, with a distinct wet season and a dry winter season. In India wheat is seeded mainly at the end of October to November and however it is often shifted to December. In North-West India wheat crop was grown after the harvest of rice crop.

Wheat production is reliant very much on climatic conditions, thus improving the ability to predict crop productivity, different climatic parameters like temperature, rainfall, humidity, cloud cover, vapour pressure, potential evapotranspiration etc. can support farmers and other stakeholders in making important judgements in terms of agronomy and crop choice. The prediction of agricultural yield is one of the most challenging and desirable task for every nation and especially in country like India which is a developing country and where there is large population to fight with hunger. Prediction of yields of any crop can be done grounded on historic crop cultivation data and meteorological. Crop yield prediction is the forecasting of harvest (crop yield) in kg per hectare. It can be used by Government, traders, policy makers, agro-based industries and agriculturists. Government use crop yield prediction in procurement, buffer-stocking, distribution, price fixation, import, export and marketing of agricultural commodities. Based on prior experience, Farmers cultivate crop, but nowadays due to the uncertainty increased in environment the accurate analysis of historic data of environment parameters should be done for successful farming. To get more harvest, we should also do the analysis of previous cultivation data.

Estimation of crop yield mainly cereals such as rice, wheat, corn has always been an fascinating research area for agro meteorologists, as these crops are important in national and international economic programming. Reliable crop yield forecasts which are early over large area would help policy makers as well as grain marketing agencies in planning for exports and imports business and local marketing. Any farmer is always attentive in knowing quantity of yield they are about to assume from their income generating farm. In past, mostly yield prediction was achieved by seeing farmer's experience on field and crop, which seems to be very cumbersome. The yield prediction is a major issue that remains to be solved based on available data. To some extent this issue can be solved through machine learning and artificial intelligence.

Weather parameters and yield prediction

The crucial climatic factor for crop growth and production is air temperature, light, rainfall and sunshine hours (Ekaputa, 2004; Orkwor, 1998). The global mean annual temperatures towards the end of 20th century were approximately 0.7^0 C compared to 19th century also is probable to surge further by means of 1.8 to 6.4 ^{0}C by 2100 AD (Kaur and Rajni, 2010). The most prominent change is inclining in the climatic temperature as a result of augmented levels of greenhouse gases in the atmosphere. Weather variables and indices like maximum temperature and minimum temperatures, rainfall and relative humidities were considered as input variables for application of Neural Networks (NNs) for crop yields forecast done at district level of Uttar Pradesh (*Ratna Raj Laxmi and Amrender Kumar; 2011*).Various weather and agricultural parameters, which includes NDVI, surface parameters (surface temperature and soil moisture) and rainfall data were used for crop yield estimation using a piecewise linear regression method with break point.(Singh et.al. 2010).The vital effect of high temperature is the augmented physiological changes, subsequently hasten maturation and low yield. The transformation of heat energy to dry matter in process photosynthesis depends on crop, sowing time and crop genetic factors (Sikder, 2009). The most important parameter among all the biological processes of yam production is the air temperature (Girijesh *et al*., 2011). It affects the phenology, growth, development and yield of yam most among all the other weather parameters in the tropical wet as well as dry climate (Maduakor *et al*., 1984; Okoh, 2004).

Dynamic simulation models of crop growth are extensively utilized for forecasting growth and harvestable yield of crops, but actually these are very comprehensive models with huge prerequisite of input data besides being highly difficult and complex in practical situation. Agro climatic models constructed on thermal indices along with weather parameters may possibly fulfil these goals. Therefore various efforts have been done in many fields to predict the crop phenology (Hundal *et al*., 1997), growth rate (Singh, 1996), leaf area index (LAI) (Benbi 1994), yield (Tripati *et al*., 1999), growth and crop yield (Hundal *et al*., 2003 a,b) by means of thermal based indices. Thermal time is the very important independent variable to explain plant development. It can also be used as an important means for characterizing thermal responses in crops. (Dwyer and Stewart, 1986).

Heat indices and yield prediction

The time interval of a specific growth phase is directly associated to temperature which is used to derive the various agro climatic indices, viz., helio-thermal units, growing degree days, photo-thermal units, phenol-thermal index and heat use efficiency (Wang, 1960). Information of GDD can provide an estimate of harvest date and crop development (Bonhomme, 2000; Wurr *et al*., 2002; Roy *et al*., 2005).

Growing degree days explain the thermal requirement of a crop. Nevertheless accumulation of growing degree days and photo thermal units are comparatively constant and autonomous for crop variety and sowing date but it can modify significantly for each developmental stage in crop (Phadnawis and Saini, 1992). (Pal *et al*,1996) reported that the growing degree days and photo thermal unit requirement differ from crop to crop. It also differ significantly from genotype to genotype. The ever-changing of sowing dates parallels to the deviations in temperature causing lengthening or shortening of growth period. Furthermore, it was also resulted that in a general the delay in sowing causes a reduction in GDD and PTU requirement of pheno-phases. (Tewari and Singh, 1993; Slafer and Rawson, 1994) witnessed that temperature has momentous impact on duration of different phonological phases and dynamics of leaf appearance. Predictive models for sweet potato (*Ipomoea batatas*) harvest in relative to growing degree days (GDD) was done using set of data of 116 planting dates using a combination of minimum CV, linear regression (LR), and several models in data mining (DM) mode to identify candidate methods of estimating relationships between GDD and harvest dates. These algorithms included regression trees, neural networks, multivariate adaptive regression splines, support vector machine, and generalized linear models. Models were based on calculated GDD and climate-related variables like maximum air temperature, mean relative humidity 20 days after transplanting (DAT), and maximum soil temperature. In this GDD method demonstrated high predictive accuracy as shown by mean square error (Villordon, *et.al,* 2009). (Bazgir, *et.al,* 2007) Yield prediction was done in Hamedan district during time period of 2003-04 and 2004-05 for Wheat using different variables of meteorology with meteorological indices. Result was interpreted according to statistical methods which gave the best subset of meteorological indices which were selected like daily minimum temperature, growing degree days (GDD), difference of maximum and minimum temperatures (TD), sunshine hours (SH), potential evapotranspiration (PET), and water vapour pressure deficit (VPD). Yield estimation was done two months in prior of harvesting. It showed that in the statistical models 83% of yield variability was accounted for variation in meteorological indices.

It can be assumed that solar radiations and temperature are the driving force of agriculture production and photothermal quotient (PTQ) use to describe effects on plant growth.Photo thermal index (PTI) is numerically equal to sum of the ratios of day length per 24 h period and multiplied by the total growing degree days (Masle *et al.*, 1989**).** Heat use efficiency (HUE) also significantly depends on genetic factors, crop type and time of sowing and has great application (Rao *et al.*, 1999). Helio thermal unit (HTU) helps in calculating and articulating the consequence of variable ambient temperature on the spell of phenological activities relating the crop reactions to the environmental temperature (Rajput, 1980).

Machine Learning

Machine learning methods are also named data-driven methods. It is a subtype of computer science and is classified as an artificial intelligence method. It bothers with the perfection of techniques which empower the computer to acquire. In simple terms evolution of algorithms which permit the machine to learn, accomplish tasks and actions. Its commonalities with statistics in many ways. Over the period of time many methodologies were developed for tasks of machine learning. It very useful may be utilized in a large areas and the benefit of this technique is that a model can tackle issues which are difficult to express by calculations. The AI models discover relations among information sources and yields (output) even if the representation is not possible. Hence this permit the utilization of AI models as a rule, for instance in design acknowledgment, characterization issues spam filtering, in information mining and anticipating.

Machine learning is concerned about the development and investigation of frameworks that can gain from informational data sets, enabling system to learn without being explicitly programmed. In the predictive learning issues, the framework comprises of an irregular “output” or response” variable y and a lot of arbitrary "input" or "logical" factors. They build a model from proof acquired from a set of information which is Input and Output. The preparation stage that is basically the training phase brings about a capacity that can be applied to new Input data so as to foresee the relating Outputs. The calculations can recognize the complicated patterns in the input with connection to yield or the output by combining simple components. In rice crop yield estimation, different order procedures were applied like support vector machine, naïve bayes, bayesnet, locally weighted learning, decision tree and so forth, to get increasingly noteworthy outcome and to show signs of improvement characterization precision. Bayesnet determines better precision of forecast of rice yield in future. This reasoned the more exact rice crop yield forecast model can be built by utilizing various data mining techniques (Sukhadia and Chaudhari, 2019).

Yield estimation Models using weather parameters

Crop yield depends on numerous factors such as temperature and precipitation, disease, soil conditions, and anthropogenic factors like fertilizers, irrigation. Climatic data are easily available also available in most of districts of India, but some of the factors can be difficult or may be impossible to quantify thus using weather parameter can boost the prediction of yield well in advance. Weather causes direct and indirect influence on crop growth development and hence its production in terms of yield. Models are simplified illustration of the complex relationships between weather or climatic elements and the performance of the crop at different stages of growth such as biomass, and yield with the help of mathematical or statistical methods (Baier, 1981).

There are a many yield forecast models and generally classified in two parts first one is Statistical Models and other being Crop Simulation Models. Lately, application of Artificial Intelligence (AI), Artificial Neural Networks (ANNs), Fuzzy Systems are upcoming. The simulation models intended to forecast crop yield use detailed crop ecology and requires widespread information on soil type, plant factors information and weather data which is related to the crop growth stage. For large area yield estimation,, the regression approach has been preferred over simulation because of trouble in getting all the data (for example, soil information, plant parameters, planting dates and certain agrometeorological information) which are expected to develop, test and run even a basic simulation model. (Walker 1989).A problem associated with regression, illustrated by Williams (1971) is multi-collinearity amid the independent variables. Simulation models are gaining popularity from past few decades, because it is process-based and requires intensive input data for better operation. Crop simulation models are specific and accurate thus can be useful for large spatio-temporal scales if sufficient input data is available.

Pioneer work in crop weather relationship was done by Fisher (1924) and Hendricks and Scholl (1943). Models developed by them require few parameters for assessind when dealing with trend of weather parameters over the crop growing period. Fisher statistical model was that impact of meteorological conditions on crop follows an orderly pattern according to mathematical law. The Fisher’s technique was modified by in 1943 by Hendricks and Scholl and later this methodology was modified by IASRI-New Delhi.

According to Baier (1981) crop-weather models are classified into three kinds: crop-weather analysis models, crop growth simulation models and empirical statistical models. Multiple regression analysis uses one or more climate variables for its analysis. Besides from meteorological variables, soil qualities, biometric perceptions, inferred meteorological parameters, accessible soil dampness, real evapotranspiration, water stress records, temperature-based list etc. have been associated with product yields. Independent variables comprised weather variables, agro-meteorological variables, soil characteristics or some derived indices of these variables. Some popular agro-climatic indices like Thermal Interception Rate Index (TIR), Growing Degree Days (GDD) were used in models. (Ramakrishna *et al*., 2003) used Southern Oscillation Index (SOI) with various other weather parameters or elements to forecast the crop yield. Various works for the development of crop weather yield models have been done by India Meteorological Department (IMD) (Sarwade 1988). They use long term yield data of district yields as dependent parameters whereas the independent variables were rainfall, temperature, relative humidity which are recorded daily, also some derived parameters like radiation, rainy days, mean temperature, evapo-transpiration, soil moisture, and yield moisture index were used. Highly correlated variable during the critical phenological stage of the crop were analyzed to establish multiple collinearity with yield.

Currently crop yield forecasting is done using linear statistical models, which are not able to account nonlinear relations in the data. The cropland ecosystem being very complex and most of the processes are non-linear. This is the limitation with statistical models, such as multiple linear regression (Jiang et al., 2004). Majority of the forecasting models in the past have been using multiple linear regressions (MLRs) for developing yield forecast (Rai *et al.*, 2013; Dhekale *et al.*, 2014; Kumar *et al.*, 2014). Multiple linear regression have very big disadvantage of over-fitting, when data have number of samples less than the number of variables along with multi-collinearity, when the independent predictors are correlated (Verma *et al.*, 2016). To overcome the various challenges in modelling, feature selection methods in statistical analysis are included, such as Least absolute shrinkage and selection operator (LASSO), Stepwise Multiple Linear Regression (SMLR), Support Vector Machine (SVM) etc. are being developed.

Neural networks which are non-linear models have been effectively used to predict crop yield by means of remotely sensed vegetation indices. Li et al. (2007) and Kaul et al. (2005) both of them have used artificial neural networks to predict soybean and corn yield in the USA. So the study attempted to use completely nonlinear models in the field of machine learning, like model-based recursive partitioning (MOB) (Zeileis et al., 2008). If the design a network or data can be arranged which correctly learns and understand the relations of effective and more contributing climatic factors on crop production, this can be used to envisage the crop yield in both the long and short term and along with enough and useful data can get a ANNs model for each area. Also using ANNs can detect the most effective factors on production of crop. So some of the factors for which the measurements are very difficult and only theoretically valid measurement are there can be ignored as there are many climatic or weather data which can be easily measured through the meteorological observatory or with the sensors. There are many other classification techniques like Naïve Bayes, Bayes Net, K Nearest Neighbour (KNN), Locally Weighted Learning (LWL) etc.

Ray *et al.*, (1994) for the yield prediction of cotton in Surendranagar district, 21 years (1972-1992) of fortnightly Actual Evapo-Transpiration (AET) was computed and these values were stepwise regressed through yield to get the final model. Developed model had coefficient of determination (R^2) value of 0.74, which predicted the yield of cotton lint 157.77 kg/ha for 1992-93. Kandiannan *et al.*, (2002) established weather based model for yield predicting in turmeric intended for Coimbatore district, Tamil Nadu with coefficient of determination 0.89. Frere and Popov (1979) found that FAO methodology involves developing an agro-meteorological index dependent on water surplus/water deficit in period of crop growth, having good potential for early crop yield assessment designed for rainfed crops. Kazmi *et al.*, (2012) done study on wheat yield prediction based on agrometeorological data in rainfed Potohar region of Pakistan.

Prediction of Rice yield in 13 districts of West Bengal using agrometeorological models was done and result obtained was a good relation between observed yield and predicted yield. Models developed have coefficient of determination ranging between 0.56 to 0.96 (Ghosh *et al.*, 2014). Rajegowda *et al.*, (2014) developed statistical models in coastal Karnataka for Ragi and rice using weather elements and yield data for long term. Kumari *et al.*, (2014) based on the weather data ordinal logistic model was developed for forecasting wheat yield in district Kanpur of Uttar Pradesh. They used weekly weather data on maximum temperature, minimum temperature, rainfall, relative humidity and for sixteen weeks of the crop cultivation for the period 1971-72 to 2009-10 along with the yield data of wheat crop.

Harvest yield estimates are prepared at different levels viz. National, State and District level under the project "Forecasting Agricultural output using Space, Agrometeorology and Land based observations (FASAL)" which operate under Ministry of Agriculture and Farmers Welfare, Govt. of India in cooperation with Space Application Centre (SAC), India Meteorological Department (IMD) and Institute of Economic Growth (IEG). Under FASAL venture, IMD in a joint effort with ICAR research institutes, Agromet Field units (AMFU) at many State Agricultural Universities (SAUs), IITs, develops functioning yield estimate for the kharif and *rabi* seasons utilizing statistical models. Inside IMD, notwithstanding the Agricultural Meteorology Division, all the Regional Meteorological Centres (RMCs) and Meteorological Centres (MCs), situated in distinctive states are also involved in this project.

Crop Prediction using stepwise regression model and regression model

In the regression approach, the relationship among yield and various variables affecting it is established through statistical methods like stepwise forward and stepwise backward multivariate analysis. The efficiency of these empirical or regression models mostly depends on the past data set which are used for the model development. The cause and effect relationship usually implicit and are used for the estimation of yields. Empirical or regression built crop models by using long terms yield data and weather are used to predict the crop production in terms of yield and still is being use. Simple regression techniques are widely used alternative to simulation model with long term weather data and crop yield data (Lobell and Burke 2010; Shi *et al*., 2013). The application of empirical models is that it is able to give better insight about the past yield of a targeted region, and also for updating weather-based models weather interactions can be used (Lobell and Burke 2010; Lobell *et al*., 2011; Basso *et al*., 2013). Dutta *et al*., (2001) reported good accuracy of pre-harvest district wise rice yield forecast for Bihar using weather data. They also noted that remarkable growth in productivity of wheat is observed in India from the last few decades, so for proper and efficient planning and policy making, crop yield forecasting is a vital tool, which helps to manage excess production.

Pandey *et al*., (2015) Studied the particular and combined effect of weather elements on yield of rice of eastern Uttar Pradesh. Also showed that the importance of these variables in crop yield variation. They also reported that bright sunshine hour has more effect on yield predicting tailed by wind velocity and rainfall with R^2 value of 67.57, 48.63 and 46.74%, respectively. The combined weather parameters such as wind velocity and rainfall, sunshine hour and rainfall, and wind velocity and sunshine hour were found to be more influencing on crop yield modelling with R^2 value of 82, 63, and 53.8% respectively. Singh *et al*., (2014) reported that 51 to 79% variability in yield of rice and 65 to 92% variability in the wheat yield can be explained through weather-based yield prediction models for the Eastern Uttar Pradesh districts. Kandiannan *et al*., (2002) predicted the first season rice yield by including solar radiation as one of the predictors at Coimbatore, Tamil Nadu. Ten years data from 1988 to 1997 were collected and used for the model development with seven predictors. Stepwise regression analysis were performed for rice yield forecast by MSTAT package. The full model regression without considering variable as solar radiation (Model I) had recorded only R^2 value to be 0.63, compared to second model which included solar radiation have enhanced the R^2 value to 0.94. In the third model they used 7 variables for stepwise regression analysis, and the developed final model which retained only four variables from the subjected input variable with an R^2 value of 0.92, this model was able to predicting the Coimbatore's rice yield.

Singh *et al*., (2014) developed yield prediction model for nine districts of Eastern Uttar Pradesh by using the weather and yield data of eighteen years (1991 to 2008), models were validated only for two years (2009 and 2010). Outcomes directed that models clarified 51 to 79% and 65 to 92% disparities for rice and wheat yield respectively. The percent Mean Bias Error ranged between -1.05 (Mau) to 6.17

(Mirzapur) districts for rice and from -6.56 (Mau) to 0.01 (Varanasi) in case of wheat crop. The percent Root Mean Square Error was in the range of 6.87 (Jaunpur) to 11.60 (Sant Ravidas Nagar) for rice and in case of wheat it was 5.52 (Mirzapur) to 11.11 (Mau). Thus they reported that the models can be applied to some extent for forecasting the yield in different districts of Eastern Uttar Pradesh.

Garde *et al.*, (2015) used discriminant function and Multiple Linear Regression (MLR) techniques for estimating wheat productivity for the district Varanasi in Uttar Pradesh. He concluded that stepwise multiple linear techniques can be used effectively for the pre-harvest wheat crop, which are more consistent in performance. Incorporating statistical indicators enhanced the precision of forecasting of wheat crop yield for both Adjusted R^2 and RMSE values. Similar work has also been done by several researchers viz. (Jain *et al.*, 1980) to forecast rice yield, developed pre-harvest model only after about two months of sowing. Agrawal *et al.*, (2001) reported that reliable forecasting of wheat yield could be obtained when the crops were at twelve week i.e. only about two months before harvest. This study was conducted on Vindhyanchal Plateau of Madhya Pradesh for wheat yield forecast. Agrawal *et al.*, (2012) used discriminant function analysis intended for developing wheat yield forecasting models for Kanpur. This methodology gave very trustworthy yield forecast about two months beforehand harvest.

Multiple linear regressions are very much appropriate for the short or intermediate term forecasting, using weather variables, weighted and unweighted weather indices to generate multiple linear regression forecasting models (Kumar *et al.*, 1999; Agrawal and Mehta, 2007; Chauhan *et al.*, 2009). Lee *et al.*, (2013) predictedwheat yield and wheat quality using weather predictors, using regression models to explain the consequences of weather on wheat yield, protein and test weight. These variables includes temperature and precipitation for growing period of wheat development stages. The forecasting efficacy of the models were boosted by adding a spatial lag effect. Wheat yield, test weight and protein showed strong correlation with the weather. Meant for developing models for the prediction of yield at different levels like district, state or national level, time series of crop yields are utilized. Trend analysis of the input-output data and Auto Regressive Integrated Moving Average (ARIMA) analysis technique is commonly used to develop model (Box and Jenkins, 1970). For modelling time series data gathered over the long run, ARIMA strategy is by and large utilized. One impediment of this approach is that the time arrangement under thought ought to be stationary or ought to be fit for turning out to be so by method for differencing or detrending. Suresh *et al*., (2011) predicted the sugarcane yield using the ARIMA model.

Yield estimation Models using machine learning

Time series is a process to analyse time on parametric series data to mine significant statistics and additional features of the data to transform into an information. Time series forecasting is the model to foresee upcoming values based on formerly observed data. New idea of crop yield in average climate situations is being developed and it is applied in time series methods on the older yield data to set up a predicting model.

Agricultural organization needs simple and precise estimation techniques for the prediction or estimation of rice yields in the scheduling process (Ji & Wan, 2007). Compulsion was to: (Washington Okori, 2011) recognize whether artificial neural network (ANN) could successfully predict rice yield for characteristic climatic circumstances of the mountainous region, (Dahikar *et.al*, 2014) assess performance of ANN model relative to the variations of parameters and compared the efficiency of multiple linear regression with ANN models. The Generalized Regression Neural Networks (GRNN) process was used for predicting production of given crop (Jin *et.al*, 2008). They reported that GRNN is a suitable technique for prediction grain production. It was stated that GRNN is appropriate for multi-objectives, non-linear, and multivariate predicting. Assessment of modified k-Means clustering set of rules in crop prediction was validated and the evaluation indicated the assessment of modified k-Means over k-Means and-Means++ clustering algorithm and also stated that the modified k-Means has attained the highest number of great superiority clusters along with accurate prediction of crop and greatest precise count (Narkhede, *et.al* , 2014). Model was developed for predicting yield of the sugarcane by means of fortnightly weather variable like average daily maximum temperature and minimum temperature, relative humidity in the morning and evening and total rainfall and the yield data in Coimbatore district (Suresh, *et.al*, 2009).

Computer software for machine learning

Machine learning methods can be performed in MATLAB, PYTHON, R software but R is one of the most preferred. It makes statistical computing very easy also graphs are easy to plot and depict in R. Advance statistical and machine learning packages are provided in R software along with various other packages and in built functions which makes statistical analysis very easy. It provide plots, effective data handling in huge amount and storage facility depending on or interest and use. R is very much helpful in predictive analytics, data pre-processing, statistical modelling, data visualization and deployment (Team, 2013).

Yield estimation Models using hybrid machine learning

Hybrid machine learning involves use of two machine learning algorithms in combination to each other for example SVM-LASSO, this includes to two techniques in combination to each other. Stas et.al (2016) Used two machine learning algorithms boosted Support Vector Machines (SVM) and Regression Trees (BRT) to predict regional winter wheat yields. The models are created through Normalized Difference Vegetation Indices (NDVI) drawn from SPOT vegetation imagery. Three types of NDVI-related forecasters were used: Single NDVI, Targeted NDVI and Incremental NDVI. BRT and SVM were initiallly used to select features with good significance for predicting the yield. After feature selection, BRT and SVM models were applied to the subset of selected features for yield forecasting. BRT outperform SVM. Forecast of rice yield through ARIMAX and proposed hybrid models using weather variables. Two hybrid approaches like ARIMAX-ANN and ARIMAX-SVM have been used for the rice yield along with weather variables of Aligarh in Uttar Pradesh. Based on the results obtained, performance of ARIMAX-SVM and ARIMAX-ANN models are close to each other but much superior to the conventional ARIMAX model for the considered data set. Performance of hybrid ARIMAX model was found to be quite encouraging. (Alam *et.al*, 2018).

Das *et al*., (2018) determined yield of rice by utilizing long-term weather data with six different statistical methods. These included stepwise multiple regression alone and in alliance with principal component analysis, similarly ANN individually and in combination with principal component analysis, elastic net (ENET) and least absolute shrinkage and selection operator (LASSO) for developing regression equations for calibration of model along with validating the accuracy of the model. During the calibration R^2 and root mean square error ranged between 0.22–0.98 and 24.02–607.29 kg ha^{-1} respectively. Independent dataset for validation resulted with the RMSE as 21.35–981.89 kg ha^{-1} and normalized root mean square error (nRMSE) as 0.98–36.7%. Ranking of the models depicted that LASSO (2.63) was the finest model, next best was ENET (3.07) while PCA-ANN (4.19) was the poorest model which was observed significant at $p < 0.001$. Due to the inhibition of overfitting and reducing the magnitude of regression coefficient by penalization decreases the model complexity, LASSO showed good performance. Pair-wise multiple comparison test was showed that LASSO was the finest model alike to SMLR and ENET..

Crop yield prediction model using LASSO

LASSO is an attractive method because it improves the quality of prediction by shrinking regression coefficient, when compared to prediction models fitted through unpenalized maximum likelihood methods. Tibshirani (1996) suggested LASSO, for utilizing in the crop yield forecasting technique. LASSO curtails residual sum of squares focus to the sum of the absolute value of the coefficient being less than a constant. This characteristic enable to produce some coefficient that are exactly zero and henceforth gives easier understandable models. It produces significant models alike subset selection and shows the stability of even the ridge regression. The idea of LASSO is relatively general and can be practical in variety of statistical models. Majority of the forecasting models in the past few decades have been using only multiple linear regressions (MLRs) for developing crop yield prediction (Rai *et al*., 2013; Dhekale 2014; Kumar *et al*., 2014). Multiple linear regression has the biggest disadvantage of over-fitting when quantity of samples is less than the quantity of variables. Also another disadvantage is the multi-collinearity when independent predictors are correlated (Verma *et al*., 2016). To combat these demerits feature selection methods in statistical analysis like Stepwise multiple linear regression (SMLR), least absolute shrinkage and selection operator, machine learning statistical technique can be adopted.

Crop yield prediction using SVM

Support Vector Machine (SVM) first perceived in 1992 and was familiarized by Boser, Guyon, and Vapnik. SVMs are supervised learning methods applied for regression and classification. They are the type of classification of generalized linear classifiers, or in new terms it is a regression prediction and classification means that practices machine learning theory to exploit analytical accuracy while spontaneously escaping over-fitting to the data. The SVM can be used equally for grouping and regression problems and it be able to indicate as a two-layered network where in first layer the weights are non-linear while in second layer it is linear. (Bray and Han, 2004). Support Vector Regression (SVR) is used in texts and in common language to describe the regression with SVM.

Support Vector machines can be demarcated as organizations which use hypothesis space of a linear functions in a high dimensional feature space. It is trained through a learning algorithm from optimization theory that instigates a learning bias resulting from statistical learning theory. It was firstly widespread with the NIPS community and now is a dynamic part of the machine learning research everywhere in the world. SVM becomes well-known when, using pixel maps as an input gives accuracy comparable to ultra-modern neural networks with expounded features in a calligraphy recognition. Nowadays it is moreover being operated in many applications, such as face analysis, hand writing analysis, regression based applications and pattern classification. The fundamentals of Support Vector Machines (SVM) have been elaborated by Vapnik and gained admiration due to countless promising and advanced features such as improved empirical performance, Structural Risk Minimization (SRM), which has been shown to be exclusive to traditional Empirical Risk Minimization (ERM) principle, used by conventional neural networks. SRM curtails an upper bound on the predictable risk, on the other hand ERM diminishes the faults on the training data. It is this modification which enables the SVM with a greater capability to oversimplify, which is the objective in statistical learning. SVM was developed to resolve the problem of classification, but in recent times they have been stretched to explain regression problems (Jakkula, 2006).

The generalized support vector machines (SVMs) have a two-stage neural network design. In the first stage, self-organizing feature map (SOM) is used as a clustering algorithm to divide the whole input space into several incoherent regions. A tree-structured architecture is taken in the partition to avoid the difficulty of predetermining the figure of partitioned regions. In the second stage, multiple SVM that optimum fit partitioned regions are prepared by concluding the most applicable kernel function and the optimal free parameters of SVMs. The simulation displays that it accomplish significant upgrading in the generalization performance in appraisal with the single models It is established on the distinctive principle of the structural risk minimization principle to evaluate functions by decreasing an upper bound of the generalization error, they are made known to be very resistant to the over-fitting problem, ultimately succeeding high generalization performance in resolving various time series estimating problems.

Another crucial property of SVMs is that training SVM is alike to solving a linearly quadratic programming problem so that the solution of SVMs is always exclusive and optimal. Distinct to other networks, training necessitates non-linear optimization with the risk of getting trapped into local minima. In the modelling of time series, two of the vital problems are non-stationarity and noise. The non-stationarity infers that the time series change their dynamics between diverse regions. This will way to gradual variations in the reliance between the input and output variables. The noisy characteristic denotes the inaccessibility of comprehensive information from the historical behaviour of the time series to entirely capture the dependency between the past and the future. The noise in the data could lead to the under-fitting or over-fitting problem. It is applied to construct nonlinear nonparametric forecasting models to be used in Crop yield forecast models for spring wheat, barley and canola grown on the Canadian Prairies were developed taking vegetation indices resulting from satellite data machine learning approaches (Johnson *et.al*, 2016).

To achieve the objectives of the present investigation, field long term weather data were collected from Regional met centre Chandigarh for Amritsar and Patiala, AMFU Ludhiana for Ludhiana, AMFU Hisar for Hisar and AMFU New Delhi for IARI, New Delhi. Wheat yield data were collected from Directorate of Economics & Statistics (DES) and state agricultural department. The details and specifics of the materials and method adopted during the course of investigation is presented in this chapter.

Data Collection from different locations

For the development of weather based machine learning models for the predicting wheat yield, daily weather data during crop growing period were collected for last thirty to forty years depending on the availability from following five location of northern India.

i. IARI, New Delhi (IARI Agro-met observatory)
ii. Hisar (AMFU-Hisar)
iii. Amritsar (RMC-Chandigarh)
iv. Patiala (RMC-Chandigarh)
v. Ludhiana (AMFU-Ludhiana)

Calculation of different heat indices

1. **Growing degree days (GDD)**

   Growing degree days is well-defined as the mean daily temperature above a certain threshold base temperature (T_{base}) accumulated on a daily basis over a period of time. Negative value of GDD is taken as zero. The base temperature varies crop to crop and its value is derived from the growth behaviors of each and specific crop. Base temperature is the temperature below which plant growth is zero. Wheat base temperature is taken as 5°C.

   GDD calculation

   It is calculated each day as maximum temperature plus minimum temperature divided by 2, then subtracting base temperature. GDDs are accumulated by adding each day.

$$GDD = \Sigma\{((T_{max} + T_{min}) / 2) - T_{base}\}$$

It is calculated from sowing up to harvest of the crop and if daily mean temperature below T_{base} then it is set to T_{base}.

2. **Helio-thermal units (HTU)**

   Helio thermal units is calculated on a daily basis and then accumulated over the desired period (sowing to harvesting) on daily basis it is product of Growing Degree Days (GDD) and the bright sunshine hours between the developmental thresholds for each day, i.e. it is the product of GDD and the mean daily hours of bright sunshine hours (SSH). Then accumulated over the desired period.

   Helio-thermal units (HTU) $=\sum(GDD \times SSH)$

3. **Heat use efficiency (HUE)**

   Heat use efficiency (HUE) is well-defined as the yield divided by growing degree days.

   Heat use efficiency = Yield/GDD

4. **Photo thermal index (PTI)**

   It is defined as the ratio of Growing Degree Days and crop growing days.

   Photo thermal index (PTI) =GDD/ crop growing day

Table 3.1 Methods for calculation of 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)	$=\sum(GDD \times SSH)$	(*Rajpu*t,1980)
3	Heat use efficiency(HUE)	=Yield/GDD	(*Haider et al*., 2003)
4	Photo thermal index (PTI)	GDD / crop growing day	(*Haider et al*., 2003)

Evapotranspiration (ET)

Evapotranspiration (ET) is the combination of two separate processes in which water is lost from the soil surface called evaporation and from the crop by transpiration. Both the processes Evaporation and transpiration occur simultaneously and there is no easy way of distinguishing between them. Evapotranspiration is normally expressed in millimetres (mm) per unit time. The rate expresses the amount of water lost from a cropped surface in the unit of depth of water.

Reference evapotranspiration (ETo) is the evapotranspiration from the reference surface. Reference surface is a hypothetical grass reference crop with an assumed crop height of 0.12 m, a fixed surface resistance of 70 s m-1 and an albedo of 0.23. The reference surface closely resembles an extensive surface of green, well-watered grass, actively growing and completely shading the ground. ETo can be calculated rom meteorological data using FAO Penman-Monteith method. This method is recommended as the standard method for the definition and computation of the reference evapotranspiration. It requires radiation, air temperature, air humidity and wind speed data.

FAO Penman-Monteith equation

The FAO Penman-Monteith method to estimate ETo can be derived from the following equation:

$$ET_o = \frac{0.408\,\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1+0.34\,u_2)}$$

Where,

ETo = reference evapotranspiration [mm day^{-1}],

Rn =net radiation at the crop surface [MJ m-2 day^{-1}]

G =soil heat flux density [MJ m-2 day^{-1}],

T =mean daily air temperature at 2 m height [°C],

u_2 =wind speed at 2 m height [m s^{-1}],

es = saturation vapour pressure [kPa],

ea = actual vapour pressure [kPa],

es-ea =saturation vapour pressure deficit [kPa],

Δ =slope vapour pressure curve [kPa °C^{-1}],

γ =psychrometric constant [kPa °C^{-1}]

Calculation procedures

Slope of saturation vapour pressure curve (Δ□)

For the calculation of evapotranspiration, the slope of the relation between saturation vapour pressure and temperature, is required. The slope of the curve at a given temperature is given by.

$$\Delta = \frac{4098\left[0.6108\exp\left(\frac{17.27\ T}{T+237.3}\right)\right]}{(T+237.3)^2}$$

Where,

Δ = slope of saturation vapour pressure curve at air temperature T [kPa °C-1],

T = air temperature [°C],

In the FAO Penman-Monteith equation, the slope of the vapour pressure curve is calculated using mean air temperature.

Mean saturation vapour pressure (es)

As saturation vapour pressure is related to air temperature. It can be calculated from the air temperature. The relationship is expressed as:

$$e^{o}(T) = 0.6108\exp\left[\frac{17.27\ T}{T + 237.3}\right]$$

where

e°(T) = saturation vapour pressure at the air temperature T [kPa]

T = air temperature [°C]

Hence saturated vapour pressure is calculated using following formula

$$e_s = \frac{e^{o}(T_{max}) + e^{o}(T_{min})}{2}$$

Actual vapour pressure (ea)

The actual vapour pressure has been derived from relative humidity data.

$$e_a = \frac{e^{o}(T_{min})\frac{RH_{max}}{100} + e^{o}(T_{max})\frac{RH_{min}}{100}}{2}$$

Where,

ea = actual vapour pressure [kPa],

e°(Tmin) = saturation vapour pressure at daily minimum temperature [kPa],

e°(Tmax) = saturation vapour pressure at daily maximum temperature [kPa],

RHmax = maximum relative humidity [%],

RHmin = minimum relative humidity [%].

Extraterrestrial radiation for daily periods (Ra)

The extraterrestrial radiation, Ra, for each day of the year and for different latitudes is estimated from the solar constant, the solar declination and the time of the year by:

$$R_a = \frac{24\ (60)}{\pi} G_{sc}\, d_r\, [\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega_s)]$$

Where,

Ra = extraterrestrial radiation [MJ m-2 day-1],

Gsc = solar constant = 0.0820 MJ m-2 min-1,

dr = inverse relative distance Earth-Sun,

ω = sunset hour angle [rad],

Φ = latitude [rad],

δ =solar declination [rad].

Ra is expressed in MJ m-2 day-1.

Solar radiation (Rs)

Solar radiation, Rs, is calculated using the Angstrom formula, which relates solar radiation to extraterrestrial radiation and relative sunshine duration through the equation given below:

$$R_s = \left(a_s + b_s \frac{n}{N} \right) R_a$$

Where,

Rs = solar or shortwave radiation [MJ m-2 day-1],

n = actual duration of sunshine [hour],

N = maximum possible duration of sunshine or daylight hours [hour],

n/N = relative sunshine duration,

Ra = extraterrestrial radiation [MJ m-2 day-1],

As = regression constant, expressing the fraction of extraterrestrial radiation reaching the earth on overcast days (n = 0),

as+bs = fraction of extraterrestrial radiation reaching the earth on clear days (n = N).

Rs is expressed in MJ m-2 day-1

Net solar or net shortwave radiation (Rns)

The net shortwave radiation is calculated and is given by:

$$R_{ns} = (1 - \alpha) R_s$$

Where,

Rns = net solar or shortwave radiation [MJ m-2 day-1],

α = albedo or canopy reflection coefficient, which is 0.23 for the hypothetical grass reference crop [dimensionless],

Rs = the incoming solar radiation [MJ m-2 day-1].

Rns = is expressed in the above equation in MJ m-2 day-1.

Net longwave radiation (Rnl)

Rnl calculated using the given formula:

$$R_{nl} = \sigma \left[\frac{T_{max,K}{}^4 + T_{min,K}{}^4}{2} \right] \left(0.34 - 0.14 \sqrt{e_a} \right) \left(1.35 \frac{R_s}{R_{so}} - 0.35 \right)$$

Where,

Rnl = net outgoing longwave radiation [MJ m-2 day-1],

σ = Stefan-Boltzmann constant [4.903 10-9 MJ K-4 m-2 day-1],

Tmax,K = maximum absolute temperature during the 24-hour period [K = °C + 273.16],

Tmin,K = minimum absolute temperature during the 24-hour period [K = °C + 273.16],

ea =actual vapour pressure [kPa],

Rs/Rso = relative shortwave radiation,

Rs = measured or calculated (Equation 35) solar radiation [MJ m-2 day-1],

Net radiation (Rn)

The net radiation (Rn) is the difference between the incoming net shortwave radiation (Rns) and the outgoing net longwave radiation (Rnl):

$$R_n = R_{ns} - R_{nl}$$

The magnitude of soil heat flux (G) **f**or day and ten-day periods, beneath the grass reference surface is relatively small, so it may be ignored (G=0).

Where no wind speed data are available within the region, a value of 2 m/s is used. This value is the average of over 2000 weather stations around the globe.

Development of weather indices

Hendrick and Scholl (1943) model for distribution of weather element was modified at Indian Agricultural Statistics Research Institute (IASRI), where, the effects of changes in weather variables on yield in the wth week was expressed as second-degree polynomial in respective correlation coefficients between the yield and the weather variables (Agarwal *et al.*, 1980, 1983; Agarwal and Jain 1982; and Jain *et al.*, 1980). This relationship was elucidated in much better way as weather in different weeks receives suitable weightage. Agrawal *et al.* (1986) further modified this model by considering the fact that the impact exerted by variations in weather parameters in wth week on yield is having a linear function of corresponding correlation coefficients between the yield and the weather parameters. The significant impact of trend on yield was removed while calculating correlation coefficients of yield with weather elements to be used as weights. These studies on the effects of second-degree terms of weather variables showed that: (a) the models using correlation coefficients based on adjusted yield for trend effect was found to be better than the ones using only simple correlations (b) quadratic terms of weather variables as well as second power of correlation coefficients did not make any improvement in the model.

Weather data for crop growing period have been used for generating weather indices and developing the crop yield forecast model. Weather indices used for developing crop yield forecast model is given in Table 3.2.

Table 3.2 Weather indices used in models using composite weather variables

	Simple weather indices								Weighted weather indices							
	Tmax	Tmin	RF	RH I	RH II	SSH	EVP	ETo	Tmax	Tmin	RF	RH I	RH II	SSH	EVP	ETo
Tmax	Z10								Z11							
Tmin	Z120	Z20							Z121	Z21						
Rf	Z130	Z230	Z30						Z131	Z231	Z31					
RH I	Z140	Z240	Z340	Z40					Z141	Z241	Z341	Z41				
RH II	Z150	Z250	Z350	Z450	Z50				Z151	Z251	Z351	Z451	Z51			
SSH	Z160	Z260	Z360	Z460	Z560	Z60			Z161	Z261	Z361	Z461	Z561	Z61		
EVP	Z170	Z270	Z370	Z470	Z570	Z670	Z70		Z171	Z271	Z371	Z471	Z571	Z671	Z71	
ETo	Z180	Z280	Z380	Z480	Z580	Z680	Z780	Z80	Z181	Z281	Z381	Z481	Z581	Z681	Z781	Z81

$$Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq 1' = 1}^{p} \sum_{j=0}^{1} a_{ii'j} \, Z_{ii'j} + cT + e$$

$$Z_{ij} = \sum_{w=1}^{m} r_{iw}^{j} \, X_{iw} \; and \; Z_{ii'j} = \sum_{w=1}^{m} r_{ii'w}^{j} \, X_{iw} \, X_{ii'w}$$

Where,

X_w denotes the value of weather variable under study in the wth week. n is the number of weeks in the crop season and A_o, a_o, a_1 and a_2 are model parameters. These models were extended to study combined effects of weather variables and an additional variate T representing the year for time trend. Y is yield; riw/rii'w is the correlation coefficient of yield (adjusted for trend effect) with i-th weather variable (Xiw) /product of i-th and i'-th weather (Xiw/Xi’w) variables in w-th period; m is week of forecast, p is number of weather variables used and e is error term.

In this type of method, for each weather variable, two types of weather indices were developed. First one being the simple values of weather variable during crop growing period [un-weighted index -Zi0] and the second one as weighted [weighted index Zi1]. Weights being taken as correlation coefficients between yield and weather variable in respective periods. In the same way indices were also produced for interaction of weather variables by using weekly products of weather variables taking two at a time. Combination of various weather variables for Weather indices were generated are presented in Table 1. Weather parameters viz. maximum and minimum temperature, morning and evening relative humidity, rainfall, bright sunshine hour and evaporation were used for such model.

After development of weather indices following empirical models along with machine learning models were used to see the impact of various weather indices of the five different locations.

sStep-wise multiple regression model (SMLR)

SMLR model was first suggested by Fisher (1924) and then by Hendrick and Scholl (1943). They used small number of assessed parameters which are used by taking into consideration of distribution pattern of weather during the crop season. Fisher employed weekly weather data assuming that the effect of change in weather variable would not be abrupt or erratic in successive weeks but an orderly one following some mathematical law. It was also assumed that these influence as well as magnitude of the predictors in successive weeks are composed of the terms of the polynomial function with respect to time.

Fisher’s technique was modified by Hendrick and Scholl (1943) by bearing in mind a second-degree polynomial in week number which can able to express the effects in successive weeks. Based on this, Hendricks and Scholl suggested,

$$Y = A_0 + a_0 \sum_{w=1}^{n} X_w + a_1 \sum_{w=1}^{n} wX_w + a_2 \sum_{w=1}^{n} w^2 X_w + e$$

Where, X_w signifies value of weather variable under study in wth week; n is the number of weeks in the crop season and A_o, a_o, a_1 and a_2 are the model parameters. This model was also extended to study the combined effects of weather variables and an additional variate T which is representing the year for considering time trend.

Least absolute shrinkage and selection operator (LASSO)

Tibshirani proposed the Least Absolute Shrinkage and Selection Operator (LASSO). It is a model selection technique, used to overcome the shortcomings of ordinary least square (OLS) and ridge regression. Though residual mean square error can be minimised by OLS, but it has low biasness and large variance, that reduces the prediction accuracy. Stronger effect on interpretation of data with large number of predictors, smaller subset selection exhibit. Subset selection is discrete and variable process. Regressors are either retained or is eliminated from the model in order to provide the better interpretable model. Lasso estimators are used for consistent regression coefficient and automatic variable selection.

Suppose given data (x^i , y_i), i=1,2,……, N, where x^i= (x_{i1},…………, x_ip)^T are predictor variables and y_i are the yield responses. As in the usual regression, assuming the observations are independent or y_i’s are conditionally independent on given x_{ij}’s.

$$(\hat{\alpha}, \hat{\beta}) = \arg\min \left\{ \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\} \qquad \text{subject to} \sum_j |\beta_j| \leqslant t.$$

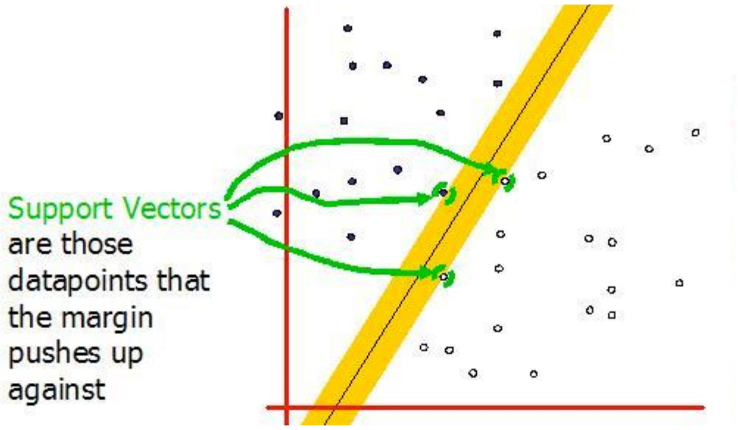
Here $t \geq 0$ is a tuning parameter, which controls the amount of shrinkage. β_j be the full least square estimates and $t_0 = \sum \beta_j$. Values of $t < t_0$ will cause shrinkage of solution towards 0 and some coefficients may be equal to 0. LASSO gives thin interpretable model with excellent prediction accuracy. An alternate formulation of LASSO to solve penalised likelihood problem is given as,

$$\min_{\beta} \frac{1}{n}(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) + \lambda \sum_{j=1}^{d} |\beta_j|.$$

Both given formulas are correspondent in sense. For any given $\lambda \in (0,\infty)$, there exists $t \geq 0$ such that the two problem have same solution and vice-versa.

Support Vector Machine(SVM)

Support vector machines (SVMs) have been proposed as a novel technique in recent time for time series forecasting or the prediction and estimation. It is a discriminative classifier defined by a separating hyperplane. i.e. given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new set of examples. In two dimentional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. It finds out a line/ hyper-plane (in multidimensional space that separate outs classes). Support vectors are data points that lie closes to the decision surface or hyperplane.



The generalized SVMs for time series forecasting have a two-stage neural network architecture. In the first stage self-organizing feature map (SOM) is used as the clustering algorithm to partition the whole input space into several disjointed zones. A tree-structured architecture is adopted in the partition to avoid the problem of pre-determining the number of partitioned regions. In the second stage, multiple SVMs, also called SVM experts, that best fit the partitioned regions are constructed by finding most appropriate kernel function along with the optimal free parameters of SVMs. SVMs experts also converge faster and use fewer support vectors. It is very specific type of learning algorithm characterized by the capacity control of the decision function. The use of kernel functions and the sparsity of the solution, established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error. It shows to be very resistant to the over-fitting problem, ultimately achieving high generalization performance in solving time series forecasting problems. Key property of SVMs is that training SVMs is equivalent to solving a linearly constrained quadratic programming problem so that the solution is always unique and globally optimal, unlike other networks' training which generally requires non-linear optimization with the danger of getting stuck into local minima.

In the modelling of time series, SVM tries to reduce the key problems which are noise and non-stationarity. In the first stage, the Isodata clustering algorithm is used to partition the whole input space into several disjointed regions. Then, in second stage, a mixture of experts is competed to solve partitioned regions. For each particular region, only the expert that best fits it is used for the final prediction. Using this strategy, the proposed method has an adaptive architecture in the sense any model can be chosen as the expert candidate.

Hybrid machine learning approach for developing crop yield estimation model

In hybrid machine learning approach combination of SMLR with LASSO, SMLR with SVM and LASSO with SVM approach was attempted

Wheat yield estimation model using Combination of SMLR-LASSO techniques

It is a combination of variable selection method used for the data analysis. SMLR selected variables from the data analysis is used as an input variable for LASSO. It is mainly used to reduce the multi-collinearity problem which arises from the weather variables. LASSO is shrinkage technique for the purpose of data reduction. This technique can reduce the number of regressors to be used in the model LASSO and hence reasonably precise estimation of wheat crop yield can be obtained by the given set of observation.

Wheat yield estimation model using combination of SMLR-SVM techniques

It is combination of variable selection method used for the data analysis. SMLR selected variables from the data analysis is used as an input variable for SVM. It is mainly used to reduce the multi-collinearity problem which arises from the weather variables. SMLR is used for the purpose of data reduction. This technique can reduce the number of regressors to be used in the SVM model and hence precise estimation of wheat yield may be obtained.

Wheat yield estimation model using combination of LASSO-SVM techniques

It is another type of combination of variable selection method used for the data analysis. In this method first variables are selected by LASSO techniques and these variables is used as an input variable for SVM. This is mainly done to reduce the multi-collinearity problem which arises from the weather variables. For data reduction LASSO is very effectively shrinkage technique. This technique can reduce the number of regressors to be used in the SVM model and hence give reasonably precise estimation of wheat yield from the given set of observation.

R software syntax for developing crop yield forecast model using LASSO techniques

```
library(HDCI)
set.seed(123857)
x<-as.matrix(IMD[,2:32])
y<-as.matrix(IMD[,1])
model_lasso<-Lasso(x, y, fix.lambda = FALSE)
yp<-model_lasso$beta0 +x%*%model_lasso$beta
RMSE<-sqrt(mean((y-yp)^2))
RMSE
write.csv(yp,"predicted values from LASSO.csv")
model_lasso$beta
write.csv(model_lasso$beta,"betas from LASSO.csv")
```

R software syntax for developing crop yield forecast model using SVM approach

```
train=read.csv(file.choose(), header=TRUE)

test=read.csv(file.choose(), header=TRUE)

head(train)

head(test)
```

```
install.packages("e1071")

library(e1071)

svm.model1 <- svm(MWD~.,data=train,scale =TRUE,kernel="linear")

svm.model2 <- svm(MWD~.,data=train,scale =TRUE,kernel="polynomial")

svm.model3 <- svm(MWD~.,data=train,scale =TRUE,kernel="sigmoid")

svm.model4 <- svm(MWD~.,data=train,scale =TRUE,kernel="radial")

svm.model1

svm.model2

svm.model3

svm.model4

fitted1<-predict(svm.model1,data=train)

fitted2<-predict(svm.model2,data=train)

fitted3<-predict(svm.model3,data=train)

fitted4<-predict(svm.model4,data=train)

predicted1<-predict(svm.model1,newdata=test)

predicted2<-predict(svm.model2,newdata=test)

predicted3<-predict(svm.model3,newdata=test)

predicted4<-predict(svm.model4,newdata=test)

write.csv(as.data.frame(fitted1), file="ZZL.csv")

write.csv(as.data.frame(fitted2), file="ZZP.csv")

write.csv(as.data.frame(fitted3), file="ZZS.csv")

write.csv(as.data.frame(fitted4), file="ZZR.csv")

write.csv(as.data.frame(predicted1), file="ZL.csv")

write.csv(as.data.frame(predicted2), file="ZP.csv")
```

```
write.csv(as.data.frame(predicted3), file="ZS.csv")

write.csv(as.data.frame(predicted4), file="ZR.csv")
```

Statistical test-

The data was analysed using R software. Statistical test was used to calculate the predicted model accuracy by several parameters as follows:

Mean Square Error (MSE)

Mean squared error (MSE) or mean squared deviation (MSD) is a measure of the average of the squares of the errors, that is, it is the average squared difference between the estimated values and real values. It is a risk function, which is corresponding to the expected value of the squared error loss. MSE is always positive (and not zero) because of randomness or because the estimator does not account for information that could produce a more accurate estimate

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

Where, Pi is the predicted value, O_i is the observed value, N is the given number of observations.

The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points *N* and multiplying it by 100% makes it a percentage error.

Root mean square error (RMSE)

This is most often used to measure the difference between predicted values from the model and actual observed values from the experiment which is being modelled. Through this, model performance during the calibration as well as validation period can be determined. It is also very helpful in comparing individual model performance with comparison to other predictive models.

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

Where, Pi is the predicted value, O_i is the observed value and N is the number of observations

Normalized mean square error (nRMSE)

Normalized mean square error is a measure (%) of the relative difference of estimated versus observed data. nRMSE is expressed in percentage and values close to zero indicates better model performance. nRMSE is considered excellent with the nRMSE <10 %, good if nRMSE value ranged between 10–20 %, fair if value ranged between 20–30 % and poor if value is >30 % (Jamieson *et al.*, 1991). nRMSE can be calculated by the formula given below.

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

Where, Pi, O_i, N and M are notated as predicted value, observed value, number of observations and mean of observed value respectively.

Percent Deviation

Percent deviation is the percentage of difference between measured and predicted values. For each station percent deviation was calculated using the formula given as:

$$\% \text{ Deviation} = \frac{Pi-Oi}{Oi} * 100$$

Pi is the predicted value,

O_i is the observed value

Objective 1: To develop and validate weather based models using machine learning for prediction of wheat yield

4.1 Heat indices and reference evapotranspiration during wheat crop growing period for IARI, New Delhi

4.1.1 Climatic conditions for New Delhi

New Delhi experiences sub-tropical and semi-arid with warm and dry summer and cold winters type of climate. It falls under the agro climatic zone named ‘Trans-Gangetic Plains which is highly fertile and abundant of water supply. During 1985 to 2017 annual mean maximum temperature was between 23.9 to 26.5 ^{o}C and annual mean minimum temperature ranged between 8.2 to 11.2 ^{o}C. Daily mean maximum temperature ranged between 40 to 46^{0}C in May and June and these months are hottest. September onwards temperature decreases. The coldest month in winter is observed in January with daily mean minimum temperature ranged between 5 to 7.5°C. Annual mean rainfall is 710 mm which is less compared to India average of about 1000 mm. Normal date of onset of monsoon in Delhi is 29th June and the wettest months are July and August. Around 80 percent of the annual rainfall is received from July to September showing the maximum rainfall due to south-west monsoon period. Climatic conditions for New Delhi is shown in (fig: 1.a).

4.1.2 Climatic conditions for Hisar

Hisar located at 29° 09' 14.28" N, 75° 43' 02.84" E and 215 meter above MSL. It comes under Trans Gangetic Plain region (VI), which belongs to Western Zone as per NARP classification. It has sub-tropical and semi-arid with warm and dry summer and cold winters. The annual mean rainfall is less which is around 455.1 mm, with a number of rainy days limited to 27 days. Normal onset of South-West monsoon is 1st week of July. 1985 to 2017 (34 years) average weather data of maximum temperature, minimum temperature, morning relative humidity, evening relative humidity, rainfall, bright sunshine hours, and evaporation are collected. Annual average maximum temperature is between 22.1 to 25.9 ºC, annual mean minimum temperature is ranged between 6.3 to 9.7 ºC. Annual mean morning and evening relative humidity was ranged between 88.6 to 55.1% and 79.5 to 30.4% respectively. Average value of bright sunshine hour was between 5.9 to 8.5 hours while range of evaporation rate was between 2.2 to 4.2mm. Climatic conditions for Hisar is shown in (fig: 1.b).

4.1.3 Climatic conditions for Amritsar

It is located at 31° 37' 50.20" N and 74° 52' 17.59" E. Amritsar is at the altitude of 254 meter above MSL. It comes under Trans Gangetic Plain region, belongs to Central Plain Zone as per NARP classification, with sub-tropical and semi-arid warm and dry summer as well as cold winters. Onset of SW monsoon is on 1st week of July. Annual mean rainfall of about 712.7 mm and number of rainy days being 42 days. During the period of 1971 to 2017 annual mean maximum temperature was between 18.9 to 24.7 ºC, annual mean minimum temperature within 5.5 to 9.5 ºC. The annual average morning relative humidity was 79.5-97.4%. Annually mean relative humidity of evening was ranged between 47.3 to 68.8%. Climatic conditions for Amritsar is shown in (fig: 1.c).

4.1.4 Climatic conditions for Ludhiana

30°54'20.87" N and 75°50'45.35" E is the latitude and longitude of Ludhiana. Its altitude is 247 meter above MSL and it comes under Trans Gangetic Plain region (VI), which belongs to Central Plain Zone (PB-3) as per the NARP classification; sub-tropical and semi-arid with warm and dry summer and cold winters. Onset of SW monsoon is at 4th week of June with the annual mean rainfall of around 749mm. number of rainy days about 40 days. Annual mean maximum temperature is between 20.91 ºC to 24.53 ºC. Total rainfall in wheat season was highest as 262.30 in 1982 and least was stated as 4.4 mm on 2001. Annual average values of maximum and minimum morning relative humidity was 95.2 and 58.6% respectively. Annual mean values of maximum and minimum evening relative humidity were 87.2 and 35.7% respectively Average value of bright sunshine hour was in between 5.4 to 8.9 hours. Climatic conditions for Ludhiana is shown in (fig: 1.d).

4.1.5 Climatic conditions for Patiala

Patiala being located at 30º19'49.59" N latitude and 76º23'41.23" E longitude having an altitude of 280 meter above MSL. As per NARP classification it comes under Trans Gangetic Plain region, belongs to Undulating Plain Zone (PB-2), Central Plain Zone (PB-3), Western Plain Zone (PB-4), sub-tropical and semi-arid with warm and dry summer and cold winters. Annual mean rainfall is around 788.3 mm. number of rainy days is about 39 days. Normal onset of SW monsoon is at 1st week of July. Major types of Soil found in this region is Fine loamy soils, coarse and fine loamy soil. Average annually maximum temperature ranged between 22.2-26.3ºC while average annual minimum temperature ranged between 8.6 to 12.0ºC.. Annual average morning relative humidity was ranged between 93.1-73.3%. Annual average evening relative humidity ranged between 42.5-66.3%. Climatic conditions for Patiala is shown in (fig: 1.e).

4.2.1 Heat indices and reference evapotranspiration during wheat crop growing period for IARI, New Delhi

Different heat indices and reference evapotranspiration were calculated during wheat growing period 1984 to 2017 for IARI, New Delhi. Value of growing degree days (GDD) calculated throughout wheat crop growing period was ranged between 1784.8 during 1996 to 2061 ^{0}C during 1987. The average value of GDD is 1953.5 ^{0}C. Value of helio thermal unit (HTU) calculated for IARI, New Delhi seen to be ranged between 9933.0 during 2010 to 16630.4 °C hour during 1987. Average value of HTU found was 12899.4 °C hour. Heat use efficiency for IARI, New Delhi ranged between 1.27 kg/ha/°C during 1984 to 2.51 kg/ha/°C during 2017. Average value of HUE is 1.85 kg/ha/°C. Value of PTI calculated during crop growing value for IARI, New Delhi ranged between 11.5 °C /day during 1996 to 13.4 °C /day during 2009. Average value of PTI seen is 12.6 °C /day. Cumulative value of reference evapotranspiration calculated during crop growing period for IARI, New Delhi was lowest (404.6 mm) during 2013 and highest (539.8 mm) during 1984. Average value of reference evapotranspiration during crop growing period was 478.8 mm. (Fig: 2.a)

4.2.2. Heat indices and reference evapotranspiration during wheat crop growing period for Hisar, Haryana

Different heat indices and reference evapotranspiration were calculated during wheat growing season 1970 to 2014 for Hisar. Value of growing degree days (GDD) calculated during wheat crop growing period remained between 1579.7 during 2004 to 3586.6 ^{0}C during 2002. The average value of GDD stood 1821.4 ^{0}C. Value of helio thermal unit (HTU) calculated for Hisar ranged between 10468.5 during 1997 to 28681.4 °C hour during 2002. Average value of HTU stood 14222.5 °C hour. Heat use efficiency for Hisar ranged between 0.18 kg/ha/°C during 1971 to 2.96 kg/ha/°C during 2012. Average value of HUE stayed 1.59 kg/ha/°C. Value of PTI calculated during crop growing value for Hisar ranged between 10.7 °C /day during 1981 to 23.2 °C /day during 2002. Average value of PTI stayed 11.8 °C /day. Cumulative value of reference evapotranspiration calculated during crop growing period for Hisar seen to be the lowest (283.9 mm) during 2012 and highest (477.4 mm) during 1975. Average value of reference evapotranspiration during crop growing period seen is 413.2 mm (Fig: 2.b).

4.2.3. Heat indices and reference evapotranspiration during wheat crop growing period for Amritsar, Punjab

Different heat indices and reference evapotranspiration were calculated during wheat growing season 1972 to 2015 for Amritsar. Value of growing degree days (GDD) calculated during the wheat growing period seen between 1450.9 during 1988 to 1747.0 ^{0}C during 1979. The average value of GDD was 1600.2 ^{0}C. Heat use efficiency for Amritsar ranged between 1.34 kg/ha/°C during 1975 to 3.97 kg/ha/°C during 2013. Average value of HUE stood 2.28 kg/ha/°C. Value of PTI calculated during crop growing value for Amritsar was between 8.0 °C /day during 2013 to 11.4 °C /day during 2001. Average value of PTI stood 10.4 °C /day. Cumulative value of reference evapotranspiration calculated during crop growing period for Amritsar was lowest (357.7 mm) during 2000 and highest (482.6 mm) during 1977. Average value of reference evapotranspiration during crop growing period was 418.6 mm (Fig: 2.c).

4.2.4 Heat indices and reference evapotranspiration during wheat crop growing period for Ludhiana, Punjab

Different heat indices and reference evapotranspiration were calculated during wheat growing season 1970 to 2015 for Ludhiana. Value of growing degree days (GDD) calculated during wheat crop growing period renged between 1512.5 during 1973 to 1878.4 °C during 2010. The average value of GDD was 1691.6 °C. Value of heilo thermal unit (HTU) calculated for Ludhiana ranged between 11930.4 during 1983 to 31752.4 °C hour during 1989. Average value of HTU stayed 21792.6 °C hour. Heat use efficiency for Ludhiana ranged between 1.84 kg/ha/°C during 1975 and 1980 to 3.20 kg/ha/°C during 2011. Average value of HUE was 2.44 kg/ha/°C. Value of PTI calculated during crop growing value for Ludhiana ranged between 9.8 °C /day during 1973 to 12.1 °C /day during 2010. Average value of PTI was 10.9 °C /day.

Cumulative value of reference evapotranspiration calculated during crop growing period for Ludhiana was the lowest (354.4 mm) during 2013 and the highest (546.6 mm) during 2010. The average value of reference evapotranspiration during the crop growing period was 449.6 mm (Fig: 2.d).

4.2.5. Heat indices and reference evapotranspiration during wheat crop growing period for Patiala, Punjab

Different heat indices and reference evapotranspiration were calculated during wheat growing season 1971 to 2012 for Patiala. Value of growing degree days (GDD) calculated during wheat crop growing period ranged between 1665.9 during 1981 to 2079.4 °C during 2003. The average value of GDD was 1866.0 °C. Heat use efficiency for Patiala ranged between 0.66 kg/ha/°C during 1971 to 2.52 kg/ha/°C during 2011. Average value of HUE was 1.98 kg/ha/°C. Value of PTI calculated during crop growing value for Patiala ranged between 10.7 °C /day during 1981 to 13.4°C /day during 2003 and 2009. Average value of PTI was 12.0 °C /day. Cumulative value of reference evapotranspiration calculated during crop growing period for Patiala was the lowest 321.7 mm during 2004 and 2012 and highest (453.1 mm) during 1976. Average value of reference evapotranspiration during crop growing period was 405.2 mm (Fig: 2.e).

Value of average wheat yield was found highest in Ludhiana (4139.1 kg/ha) followed by Patiala (3705.1 kg/ha), Amritsar (3625.3 kg/ha), IARI, New Delhi (3617.3 kg/ha) and 2859.9 kg /ha for Hisar (Fig.3). Value of growing degree days during crop growing period found to be the lowest for Amritsar (1600.2 ^{0}C) followed by Ludhiana (1691.6 ^{0}C), Hisar (1821.4 ^{0}C), Patiala (1866.0 ^{0}C) and (1953.5^{0}C) for IARI, New Delhi (Fig 4). Heat use efficiency was found highest in Ludhiana (2.44 kg/ha/^{0}C) followed by Amritsar (2.28 kg/ha/^{0}C), Patiala (1.98 kg/ha/^{0}C), IARI, New Delhi (1.85 kg/ha/^{0}C) and 1.59 kg/ha/^{0}C for Hisar (Fig 5). Photo thermal unit seen to be the lowest for Amritsar (10.4 °C /hour) followed by Ludhiana (10.9°C /hour), Hisar (11.8 °C /hour), Patiala (12.0°C /hour)and 12.6 °C /hour for IARI, New Delhi (Fig 6).Reference evaptranspiration during crop growing period was lowest for Patiala (405.2 mm) followed by Hisar (413.2 mm), Amritsar (418.6 mm), Ludhiana (449.6 mm) and 478.8 mm for IARI, New Delhi (Fig 7).

4.3.1 Weather based wheat yield prediction models for IARI, New Delhi using different techniques

Yield prediction models *of Rabi* wheat for IARI, New Delhi have been developed for crop yield prediction using long term crop yield data along with long period daily weather data from 46th to 15th standard meteorological week. The model was developed using stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Performances of the developed model during calibration and validation period are shown in Table 4.1.

Results showed that model developed by different techniques performed better with value of coefficient of determination (R^2) significant at 1% probability level for all the models. The models developed for predicting the yield had the value of coefficient of determination (R^2) during calibration ranged between 0.95 for SMLR to 0.99 for SVR. The maximum value of R^2 was found for model developed by SVR (R^2=0.99) followed by LASSO and LASSO-SVR (R^2=0.98), SMLR-SVR (R^2=0.96) and SMLR (R^2=0.95). RMSE for calibration was ranged between 44.0 and 98.2 kg/ ha for model developed using different techniques. The maximum value of RMSE during calibration occurred for SMLR-SVR (98.2 kg/ha) followed by SMLR (96.2 kg/ha), LASSO (72.0 kg/ha), LASSO-SVR (67.9 kg/ha) and SVR (44.02kg/ha). The mean square error for calibration was ranged between 1938 kg/ha for model developed by SVR techniques to 9650 kg/ha for model developed by SMLR-SVR techniques. The maximum value was found for SMLR-SVR followed by SMLR, LASSO, LASSO-SVR and SVR respectively. During calibration the value of nRMSE ranged between 1.26 to 2.81 %. The minimum value of nRMSE was found in the model developed by SVR (1.26%) techniques followed by LASSO-SVR (1.94%), LASSO (2.06%), SMLR (2.77%) and SMLR-SVR (2.81%) techniques. The calibration and validation of the model developed by different techniques for IARI, New Delhi are shown in (fig 8a & 8b).

The most important weather parameter identified for wheat crop yield prediction model developed by SMLR techniques for IARI, New Delhi are Z281 (Minimum temperature*evapotranspiration) Z581(minimum relative humidity*evapotranspiration)and Z671 (sunshine hour*evaporation) while the important weather parameter identified for wheat crop yield prediction model developed by LASSO are Z10 (maximum temperature), Z51 (minimum relative humidity), Z141(maximum temperature*morning relative humidity), Z151 (maximum temperature*minimum relative humidity), Z240 (minimum temperature*morning relative humidity), Z260(minimum temperature*bright sunshine hours), Z461(morning relative humidity*bright sunshine hours), Z561(evening relative humidity*bright sunshine hours), Z671(bright sunshine hours*evaporation), Z181(maximum temperature*evapotranspiration), Z581(evening relative humidity *evapotranspiration), GDD, PTI. Equation for wheat crop model developed by SMLR and LASSO techniques are shown in Table 4.2. The Parameters for wheat crop yield prediction model developed by SVR was SVM-Type: eps-regression, SVM-Kernel: linear kernel function with cost as 1, gamma as 0.01282051 and epsilon as 0.1 and Number of Support Vectors are 21. For yield prediction model developed by LASSO-SVR the parameters are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.05882353, epsilon: 0.1 and Number of Support Vectors: 19 and for model developed by SMLR-SVR the parameters recognized are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.2, epsilon: 0.1 and Number of Support Vectors was 17.

The nRMSE for validation ranged between 1.15 to 5.11%. The lowest value of nRMSE was 1.15% for the model developed by LASSO techniques followed by SMLR-SVR (1.54%), LASSO-SVR (1.88%)), SVR (2.39%) and SMLR (5.11%). According to nRMSE calculated for validation, model developed for wheat predictions for IARI, New Delhi using different techniques were excellent having nRMSE value < 5 %. The RMSE for validation ranged between 50.9 and 225.8 kg / ha. During validation, the highest RMSE was recorded 225.8 kg/ha for model developed by SMLR followed by SVR (105.5 kg/ha), LASSO-SVR (83.1 kg/ha), SMLR-SVR (68.3kg/ha) and LASSO (50.9 kg/ha). The mean square error for validation ranged between 2594 kg/ha for model developed by LASSO techniques to 50968 kg/ha for model developed by SMLR techniques. The maximum value was found for model developed by SMLR techniques followed by SVR, LASSO-SVR, SMLR-SVR and LASSO techniques. Among the different model developed for wheat crop prediction for IARI, New Delhi, modal developed by LASSO techniques performed best followed by SMLR-SVR, LASSO-SVR, SVR and SMLR techniques

Table 4.1 Weather based wheat yield prediction models for IARI, New Delhi using different techniques

S.No.	Techniques used for developing model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation		
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)
1	SMLR	0.95	9254	96.2	2.77	50969	225.8	5.11
2	SVR	0.99	1938	44.0	1.26	11192	105.5	2.39
3	LASSO	0.98	5177	72.0	2.06	2594	50.9	1.15
4	LASSO-SVR	0.98	4612	67.9	1.94	6913	83.1	1.88
5	SMLR-SVR	0.96	9650	98.2	2.81	4659	68.3	1.54

Table 4.2. Equation for Yield prediction model for IARI, New Delhi using SMLR and LASSO techniques.

S. No.	Techniques used for developing model	Equation
1.	SMLR	y=1635.55+57.38*time+1.84*Z671+4.71*Z281+0.66*Z581
2.	LASSO	Y=58.53*time+0.16*Z10+7.24*Z51+0.06*Z141+0.05*Z151+Z240*0.21+ Z260*0.15+Z461*0.28+Z561*0.105+Z671*2.74+GDD*0.35+PTI*0.13+ Z181*0.945+0.6*Z581

4.3.2 Weather based wheat yield prediction models for Hisar, Haryana using different techniques

Yield prediction models for wheat crop during *Rabi* season for Hisar have been developed for crop yield prediction using long term crop yield data along with long period daily weather data from 46th to 15th standard meteorological week. The model was developed using stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Modal accuracy parameter during calibration and validation are shown in Table 4.3.

The models developed using different techniques have nRMSE value during calibration between 2.87% to 10.95%. The maximum value of nRMSE was found in the model developed by SMLR-SVR (10.95%), followed by SVR (5.72%), LASSO (4.45), LASSO-SVR (3.88) and SMLR (2.87%). The coefficient of determination (R^2) was significant at 1% probability level for all the models. The models developed for predicting the yield had the value of coefficient of determination (R^2) during calibration ranged between 0.97 to 0.99. The maximum value of R^2=0.99 was found for model developed by LASSO, SVR and LASSO-SVR followed by SMLR (R^2=0.98) and SMLR -SVR (R^2=0.97). RMSE for calibration was ranged between 69.8 and 271.7 kg/ ha for model developed using different techniques. The maximum value of RMSE during calibration occurred for SMLR-SVR (271.7 kg/ha) followed by SVR (141.9 kg/ha), LASSO (110.4 kg/ha), LASSO-SVR (96.3 kg/ha) and SLMR (69.8 kg/ha). The mean square error for calibration was ranged between 4872 kg/ha for model developed by SMLR techniques to 73826 kg/ha for model developed by SMLR-SVR techniques. The maximum value was found for SMLR-SVR followed by SVR, LASSO, LASSO-SVR and SMLR respectively.

The important weather parameter identified for wheat yield prediction model developed by SMLR techniques for Hisar are Time, Z21 (minimum temperature*maximum temperature), Z461 (morning relative humidity* bright sunshine hours) and Z581 (evening relative humidity *evapotranspiration) and the weather elements identified by LASSO techniques are time, Z61 (bright sunshine hours* maximum temperature), Z120 (maximum temperature*minimum temperature), Z121 (maximum temperature*minimum temperature), GDD, HUE and PTI. Equation for wheat crop model developed by SMLR and LASSO techniques are shown in Table 4.4. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01298701, epsilon: 0.1 and number of Support Vectors: 24. The Parameters for LASSO-SVR are SVM-Type: eps-regression, SVM-

Kernel: linear, cost: 1, gamma: 0.1428571, epsilon: 0.1 and number of Support Vectors: 6 while the parameters for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.03333333, epsilon: 0.1 and number of Support Vectors was 30. The calibration and validation of the model developed by different techniques for Hisar are shown in (fig 9a & 9b).

The RMSE for validation ranged between 123.1 and 509.4 kg / ha. During validation, the highest RMSE was recorded 509.4 kg/ha for model developed by SMLR-SVR followed by SMLR (394.1 kg/ha), SVR (374.8 kg/ha), LASSO-SVR (145.9kg/ha) and LASSO (123.1 kg/ha). The mean square error for validation ranged between 15158 kg/ha for model developed by LASSO techniques to 259510 kg/ha for model developed by SMLR-SVR techniques. The maximum value was found for model developed by SMLR-SVR techniques followed by SMLR, SVR, LASSO-SVR, and LASSO techniques. The nRMSE for validation ranged between 2.82 to 11.65%. The maximum value of nRMSE was found for the model developed by SMLR-SVR (11.65%), followed by SMLR (9.06%), SVR (8.57%), LASSO-SVR (3.34%) and LASSO (2.82%) techniques. According to nRMSE calculated for validation, model developed for wheat predictions for Hisar using all techniques were excellent having nRMSE value < 10 % except for SMLR-SVR techniques having nRMSE value 11.65%. Among the different model developed for wheat crop prediction for Hisar, modal developed by LASSO techniques performed best followed by LASSO-SVR, SVR, SMLR and SMLR-SVR techniques

Table 4.3 Weather based wheat yield prediction models for Hisar, Haryana using different techniques

S. No.	Techniques used for developing model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation		
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)
1.	SMLR	0.98	4872	69.8	2.87	155331	394.1	9.06
2.	SVR	0.99	20142	141.9	5.72	140440	374.8	8.57
3.	LASSO	0.99	12191	110.4	4.45	15158	123.1	2.82
4.	LASSO-SVR	0.99	9264	96.3	3.88	21281	145.9	3.34
5.	SMLR-SVR	0.97	73826	271.7	10.95	259510	509.4	11.65

Table 4.4. Equation for Yield prediction model for Hisar, Haryana using SMLR and LASSO techniques

S. No.	Techniques used for developing model	Equation
1.	SMLR	Y=4804.87+Time*111.73* Z461*2.55+ Z21*107.83+ Z581*0.407
2.	LASSO	Y=time*2.49+Z61*1.05+Z120* Z121*0.02+GDD*1+ HUE*1680+PTI*1.01

4.3.3 Weather based wheat yield prediction models for Amritsar, Punjab using different techniques

Yield prediction models for wheat crop during *Rabi* season for Amritsar have been developed for crop yield prediction using long term crop yield data along with long period daily weather data from 46th to 15th standard meteorological week. The model was developed using stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Performance of the modal developed using different techniques for wheat yield forecast for Amritsar during calibration and validation are shown in Table 4.5.

The coefficient of determination (R^2) was significant at 1% probability level for all the models. The models developed for predicting the yield had the value of coefficient of determination (R^2) during calibration between 0.67 to 0.99. The maximum value of R^2=0.99 was found for model developed by LASSO, LASSO-SVR, SVR and SMLR-SVR techniques. Modal developed by SMLR had value R^2=0.97. RMSE for calibration was ranged between 41.2 and 69.1 kg/ ha. The maximum value of RMSE during calibration occurred for modal developed by SVR (69.1 kg/ha) followed by LASSO-SVR (61.8 kg/ha), LASSO (58.5 kg/ha), SMLR-SVR (45.3 kg/ha) and SLMR (41.2 kg/ha). The mean square error for calibration was ranged between 1696 kg/ha for model developed by SMLR techniques to 4773 kg/ha for model developed by SVR techniques. The maximum value was found for SVR followed by LASSO-SVR, LASSO, SMLR-SVR and SMLR respectively. Value of nRMSE ranged between 1.22 kg/ha to 2.05 kg/ha. Maximum value of nRMSE was found for model developed by SVR (2.05 %) techniques followed by LASSO-SVR (1.84 %), LASSO(1.74%), SMLR-SVR (1.35%) and SMLR (1.22%) techniques.

The most important weather parameter identified by SMLR for Amritsar are Z581 (evening relative humidity*evapotranspiration), HUE and PTI. While the important weather parameter identified by LASSO are time, Z120 (maximum temperature*minimum temperature), Z140 (maximum temperature*morning relative humidity), Z151 (maximum temperature*evening relative humidity), Z241 (minimum temperature*morning relative humidity), Z250 (minimum temperature* evening relative humidity), Z81 (evapotranspiration), Z581 (evening relative humidity*evapotranspiration), GDD, HUE, PTI. Equation for wheat crop model developed by SMLR and LASSO techniques are shown in Table 4.6. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02173913, epsilon: 0.1 and number of Support Vectors: 19. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.09090909, epsilon: 0.1 and number of Support Vectors: 10 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and the number of Support Vectors was 4. The calibration and validation of the model developed by different techniques for Amritsar are shown in (fig 10a & 10b).

The mean square error for validation ranged between 6431 kg/ha for model developed by LASSO techniques to 63399 kg/ha for model developed by SVR techniques. The maximum value was found for model developed by SVR techniques followed by SMLR, SMLR-SVR, LASSO-SVR, and LASSO techniques. The RMSE for validation ranged between 802 and 251.8 kg / ha. During validation, the highest RMSE was recorded 251.8 kg/ha for model developed by SVR followed by SMLR (207.0 kg/ha), SMLR-SVR (190.4 kg/ha), LASSO-SVR (162.1 kg/ha) and LASSO (80.2 kg/ha). The nRMSE for validation ranged between 1.81 to 5.70%. The maximum value of nRMSE was found for the model developed by SVR (5.70%), followed by SMLR (4.68%), SMLR-SVR (4.31%), LASSO-SVR (3.67%) and LASSO (1.81%) techniques. According to nRMSE calculated for validation, model developed for wheat predictions for Amritsar using all techniques were excellent having nRMSE value < 10 %. Among the different model developed for wheat crop prediction for Amritsar, modal developed by LASSO techniques performed best followed by LASSO-SVR, SVR, SMLR and SMLR-SVR techniques

Table 4.5 Weather based wheat yield prediction models for Amritsar, Punjab using different techniques

S.No.	Techniques used for developing model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation		
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)
1	SMLR	0.67	1696	41.2	1.22	42851	207.0	4.68
2	SVR	0.99	4773	69.1	2.05	63399	251.8	5.70
3	LASSO	0.99	3424	58.5	1.74	6431	80.2	1.81
4	LASSO-SVR	0.99	3813	61.8	1.84	26270	162.1	3.67
5	SMLR-SVR	0.99	2052	45.3	1.35	36268	190.4	4.31

Table 4.6. Equation for Yield prediction model for Amritsar, Punjab using SMLR and LASSO techniques

Techniques used for developing model	Equation
SMLR	Y=3010.16+0.141*Z581+1595*HUE+304.71*PTI
LASSO	Y=9.05*time-0.01*Z120-0.01*Z140+0.13*Z151+0.01*Z241-0.01*Z250+2.14*GDD+1313.81*HUE+6.88*PTI+ 13.41*Z81+13.42*Z581

4.3.4 Weather based wheat yield prediction models for Ludhiana, Punjab using different techniques

Yield prediction models *of Rabi* wheat for Ludhiana have been developed for crop yield prediction using long term crop yield data along with long period daily weather data from 46th to 15th standard meteorological week. The model was developed using stepwise multi linear regression (SMLR), support vector regression

(SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Performances of the developed model during calibration and validation period are shown in Table 4.7.

The coefficient of determination (R^2) was significant at 1% probability level for all the models. The models developed for predicting the yield had the value of coefficient of determination (R^2) ranged between 0.93 for, SMLR, LASSO, LASSO-SVR, SMLR-SVR to 0.97% for SVR. The nRMSE for calibration ranged between and 3.07% to 6.37%. The maximum value of nRMSE was found for the model developed by LASSO (6.37%), followed by SMLR-SVR (4.67%), SMLR (4.63%), LASSO-SVR (3.46%) and SVR (3.07%) techniques. RMSE for calibration was ranged between 119.2 and 247.2 kg/ ha for model developed using different techniques. The maximum value of RMSE during calibration occurred for LASSO (247.2 kg/ha) followed by SMLR-SVR (181.5 kg/ha), SMLR (179.9 kg/ha), LASSO-SVR (134.4 kg/ha) and SVR (119.2kg/ha). The mean square error for calibration was ranged between 14207 kg/ha for model developed by SVR techniques to 61111 kg/ha for model developed by LASSO techniques. The maximum value was found for LASSO followed by SMLR-SVR, SMLR, LASSO-SVR and SVR respectively. The calibration and validation of the model developed by different techniques for Ludhiana are shown in (fig 11a & 11b).

The important weather parameter identified by SMLR for Ludhiana are time, Z141 (maximum temperature*morning relative humidity) and Z151 (maximum temperature*evening relative humidity). While the weather elements identified by LASSO are time, Z41 (morning relative humidity*maximum temperature) , Z141 (maximum temperature*morning relative humidity) and Z151 (maximum temperature*evening relative humidity), Z251 (Minimum temperature*evening relative humidity) and Z361 (rainfall*bright sunshine hours). Equation for wheat crop model developed by SMLR and LASSO techniques are shown in Table 4.8. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.0212766, epsilon: 0.1 and number of Support Vectors: 28. The various parameters for LASSO- SVR are as follows SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.1, epsilon: 0.1, number of Support Vectors: 23

while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and number of Support Vectors was 24.

The mean square error for validation ranged between 53080 kg/ha for model developed by LASSO techniques to 295003 kg/ha for model developed by SMLR-SVR techniques. The maximum value was found for model developed by SMLR-SVR techniques followed by SMLR, SVR, LASSO-SVR, and LASSO techniques. The RMSE for validation ranged between 230.4 and 543.1 kg / ha. During validation, the highest RMSE was recorded 543.1 kg/ha for model developed by SMLR-SVR followed by SMLR (502.5 kg/ha), SVR (489.0 kg/ha), LASSO-SVR (470.3 kg/ha) and LASSO (230.4 kg/ha). The nRMSE for validation ranged between 4.83 to 11.39 %. The maximum value of nRMSE was found for the model developed by SMLR-SVR (11.39%), followed by SMLR (10.54%), SVR (10.26%), LASSO-SVR (9.87%) and LASSO (4.83%) techniques. According to nRMSE calculated for validation, model developed for wheat predictions for Ludhiana were excellent for model developed by LASSO and LASSO-SVR having nRMSE value < 10 % and very good for model developed by SVR, SMLR and SMLR-SVR techniques. Among the different model developed for wheat crop prediction for Ludhiana, modal developed by LASSO techniques performed best followed by LASSO-SVR, SVR, SMLR and SMLR-SVR techniques.

Table 4.7 Weather based wheat yield prediction models for Ludhiana, Punjab using different techniques

S.No.	Techniques used for developing model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation		
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)
1	SMLR	0.93	32378	179.9	4.63	252524	502.5	10.54
2	SVR	0.97	14207	119.2	3.07	239146	489.0	10.26
3	LASSO	0.93	61111	247.2	6.37	53080	230.4	4.83
4	LASSO-SVR	0.93	18052	134.4	3.46	221144	470.3	9.87
5	SMLR-SVR	0.93	32941	181.5	4.67	295003	543.1	11.39

Table 4.8. Equation for Yield prediction model for Ludhiana, Punjab using SMLR and LASSO techniques

Techniques used for developing model	Equation
SMLR	Y=-138.47+54.45*time+0.54*Z141+0.97*Z151
LASSO	Y=35.85*time+14.07*Z41+0.19*Z141+0.52*Z151+1.02*Z251+0.26* Z361

4.3.5 Weather based wheat yield prediction models for Patiala, Punjab using different techniques

Yield prediction models *of Rabi* wheat for Patiala have been developed for crop yield prediction using long term crop yield data along with long period daily weather data from 46th to 15th standard meteorological week. The model was developed using stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Performances of the developed model during calibration and validation period are shown in Table 4.9.

During calibration nRMSE values ranged between 0.77 to 4.81 % having highest value for the model developed by LASSO-SVR (4.81%) techniques followed by SVR (2.45%), SMLR-SVR (1.72%), LASSO (0.88%) and SMLR (0.77%) techniques. The models developed for predicting the yield had the value of coefficient of determination (R^2) ranged between 1.00 % for, SMLR, LASSO, SMLR-SVR, 0.99 for SVR and 0.94 % for LASSO-SVR. RMSE for calibration was ranged between 26.8 and 166.1 kg/ ha for model developed using different techniques. The maximum value of RMSE during calibration occurred for LASSO-SVR (166.1 kg/ha) followed by SVR (84.6 kg/ha), SMLR-SVR (59.3 kg/ha), LASSO (30.5 kg/ha) and SMLR (26.8 kg/ha). The mean square error for calibration was ranged between 716 kg/ha for model developed by SMLR techniques to 27593 kg/ha for model developed by LASSO-SVR techniques. The maximum value was found for LASSO-SVR followed by SVR, SMLR-SVR, LASSO and SMLR respectively. The calibration and validation of the model developed by different techniques for Patiala are shown in (fig 12a & 12b).

The various weather parameter identified by SMLR for Patiala are time, Z81 (evapotranspiration *maximum temperature), Z241 (minimum temperature*morning relative humidity), GDD and HUE while the parameters identified by LAASO are time, Z81 (evapotranspiration *maximum temperature), Z140 (maximum temperature*morning relative humidity), Z250 (minimum temperature* evening relative humidity), Z281 (Minimum temperature* evapotranspiration), GDD and HUE. Equation for wheat crop model developed by SMLR and LASSO techniques are shown in Table 4.10. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02173913, epsilon: 0.1 and number of Support Vectors: 22; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02173913, epsilon: 0.1 and number of Support Vectors: 22, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.2, epsilon: 0.1 and number of Support Vectors was 6. The calibration and validation of the model developed by different techniques for Patiala are shown in (fig 5a & 5b).

During validation value of RMSE ranged between 29.9 and 421.2 kg / ha. Maximum value of RMSE was 421.2 kg/ha for model developed by LASSO-SVR followed by SVR (283.5 kg/ha), SMLR-SVR (108.8 kg/ha), SMLR (30.3 kg/ha) and LASSO (29.9 kg/ha). The mean square error for validation ranged between 893 kg/ha for model developed by LASSO techniques to 177433 kg/ha for model developed by LASSO-SVR techniques. The maximum value was found for model developed by LASSO-SVR techniques followed by SVR, SMLR-SVR, SMLR and LASSO techniques. The nRMSE for validation ranged between 0.65 to 9.09 %. The maximum value of nRMSE was found for the model developed by LASSO-SVR (9.09%), followed by SVR (6.12%), SMLR-SVR (2.35%), SMLR (0.65%) and LASSO (0.65 %) techniques. Based on nRMSE value during validation, model developed by all techniques for wheat predictions for Patiala performed excellent having nRMSE value < 10 %. Among the different model developed for wheat crop prediction for Patiala, modal developed by LASSO and SMLR techniques performed best followed by SMLR-SVR, SVR and LASSO-SVR techniques.

Table 4.9 Weather based wheat yield prediction models for Patiala, Punjab using different techniques

S.No.	Techniques used for developing model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation		
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)
1	SMLR	1.00	716	26.8	0.77	919	30.3	0.65
2	SVR	0.99	7164	84.6	2.45	80369	283.5	6.12
3	LASSO	1.00	929	30.5	0.88	893	29.9	0.65
4	LASSO-SVR	0.94	27593	166.1	4.81	177433	421.2	9.09
5	SMLR-SVR	1.00	3517	59.3	1.72	11829	108.8	2.35

Table 4.10. Equation for Yield prediction model for Patiala, Punjab using SMLR and LASSO techniques

Techniques used for developing model	Equation
SMLR	Y=-3049.12+ 3.91*time+ 2.21*Z81+ 0.04*Z241+ 1.75*GDD+ 1723.57*HUE
LASSO	Y=4.06*time+ 0.43*Z81+ 0.002*Z140+0.03*Z250+ 0.01*Z281+ 1.72*GDD+ 1713.2*HUE

Table 4.11 RMSE and nRMSE during validation for different model for different districts

Techniques used for developing model	**IARI, New Delhi**		**Hisar, Haryana**		**Amritsar, Punjab**		**Ludhiana, Punjab**		**Patiala, Punjab**	
	RMSEv (kg/ha)	nRMSEv (%)	RMSEv (kg/ha)	nRMSEv (%)	RMSEv (kg/ha)	RMSEv (kg/ha)	RMSEv (kg/ha)	nRMSEv (%)	RMSEv (kg/ha)	nRMSEv (%)
SMLR	225.8	5.11	394.1	9.06	207.0	4.68	502.5	10.54	30.3	0.65
SVR	105.5	2.39	374.8	8.57	251.8	5.70	489.0	10.26	283.5	6.12
LASSO	**50.9**	**1.15**	**123.1**	**2.82**	**80.2**	**1.81**	**230.4**	**4.83**	**29.9**	**0.65**
LASSO-SVR	83.1	1.88	145.9	3.34	162.1	3.67	470.3	9.87	421.2	9.09
SMLR-SVR	68.3	1.54	509.4	11.65	190.4	4.31	543.1	11.39	108.8	2.35

4.3.6. Performance of different prediction model:

The performance of different models for wheat crop prediction for different district of north west region of India based on RMSE and nRMSE during validation are sown in (Table 4.11). Based on performance for all districts LASSO performed best among all the five models having excellent performance for all districts having nRMSE < 5 %. For IARI, New Delhi all model performed excellent having nRMSE < 10 %. LASSO performed best followed by SMLR-SVR, LASSO-SVR, SVR and SMLR. For Hisar LASSO model was best followed by LASSO-SVR, SVR, SMLR and SMLR-SVR model. For Hisar all model performed excellent having nRMSE<10 % except SMLR-SVR performed good having nRMSE value 11.65. For Amritsar all model performed excellent having nRMSE value <10 %. For Amritsar LASSO model was best followed by LASSO-SVR, SMLR-SVR, SMLR and SVR. For Ludhiana LASSO performed best followed by LASSO-SVR, SVR, SMLR and SMLR- SVR. Model developed for wheat yield prediction for Ludhiana performed excellent for LASSO and LASSO-SVR having nRMSE < 10 % and good for SVR, SMLR and SMLR- SVR having nRMSE < 15 %. For Patiala all the model performed excellent having nRMSE < 10 %. For Patiala LASSO and SMLR was best followed by SMLR-SVR, SVR and LASSO-SVR.

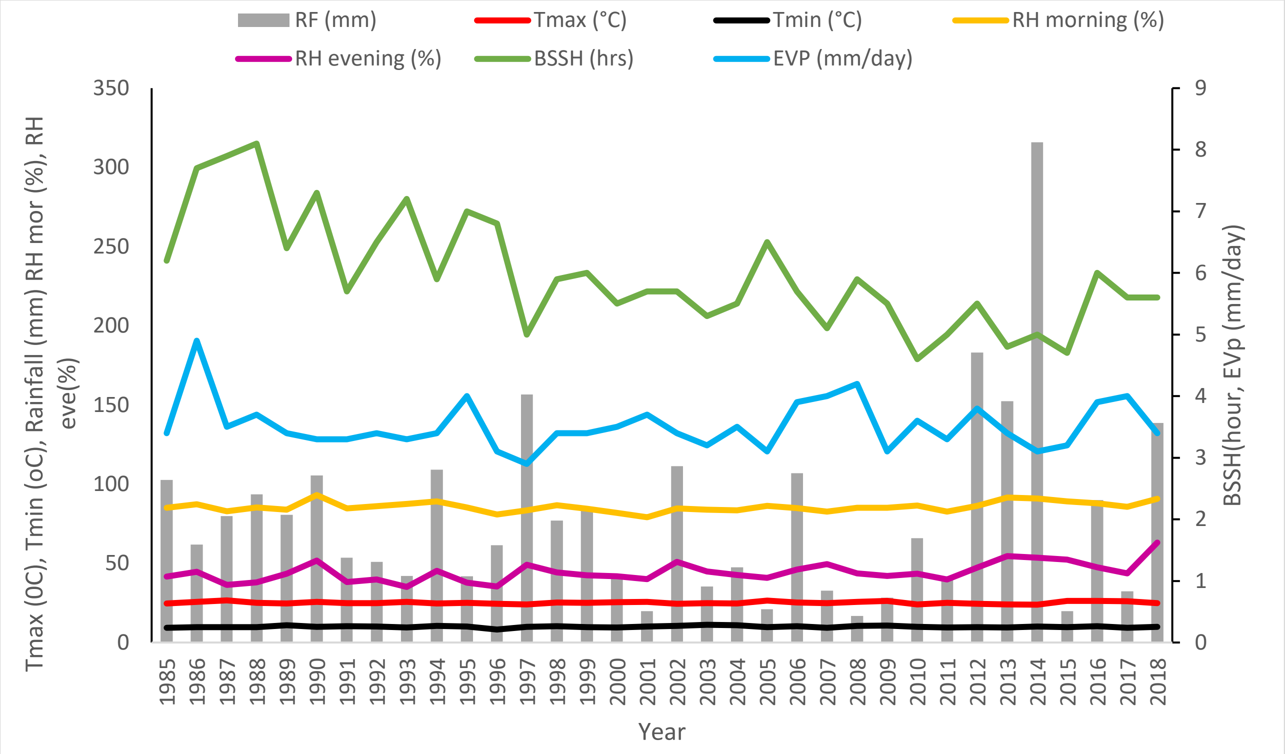


Fig.1.a Climatic conditions for IARI, New Delhi

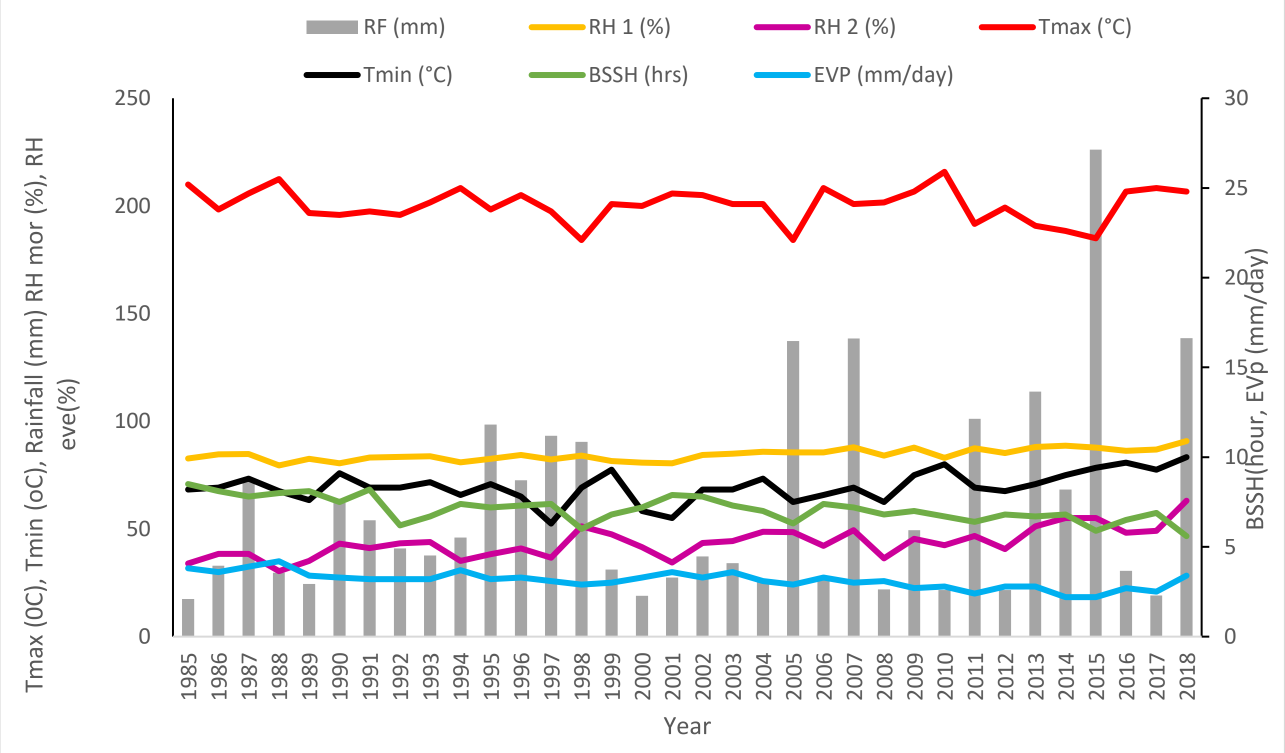


Fig.1.b Climatic conditions for Hisar, Haryana

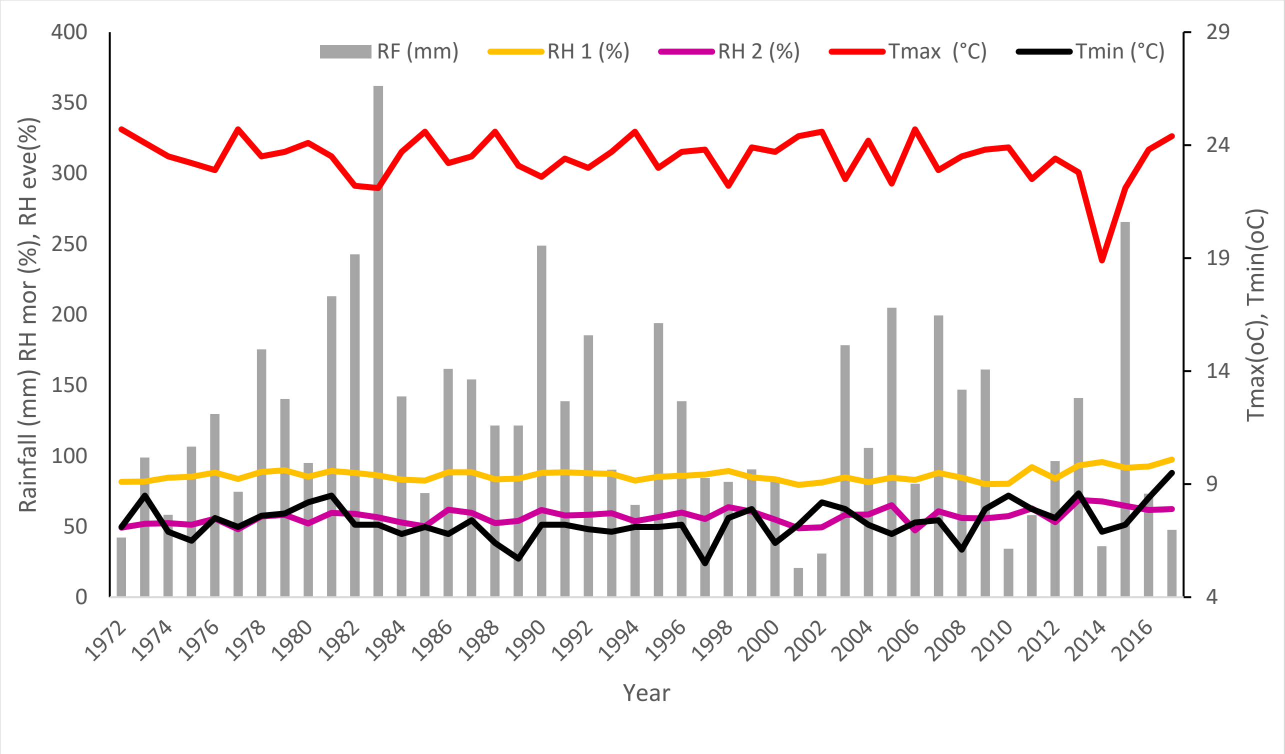


Fig.1.c. Climatic conditions for Amritsar, Punjab

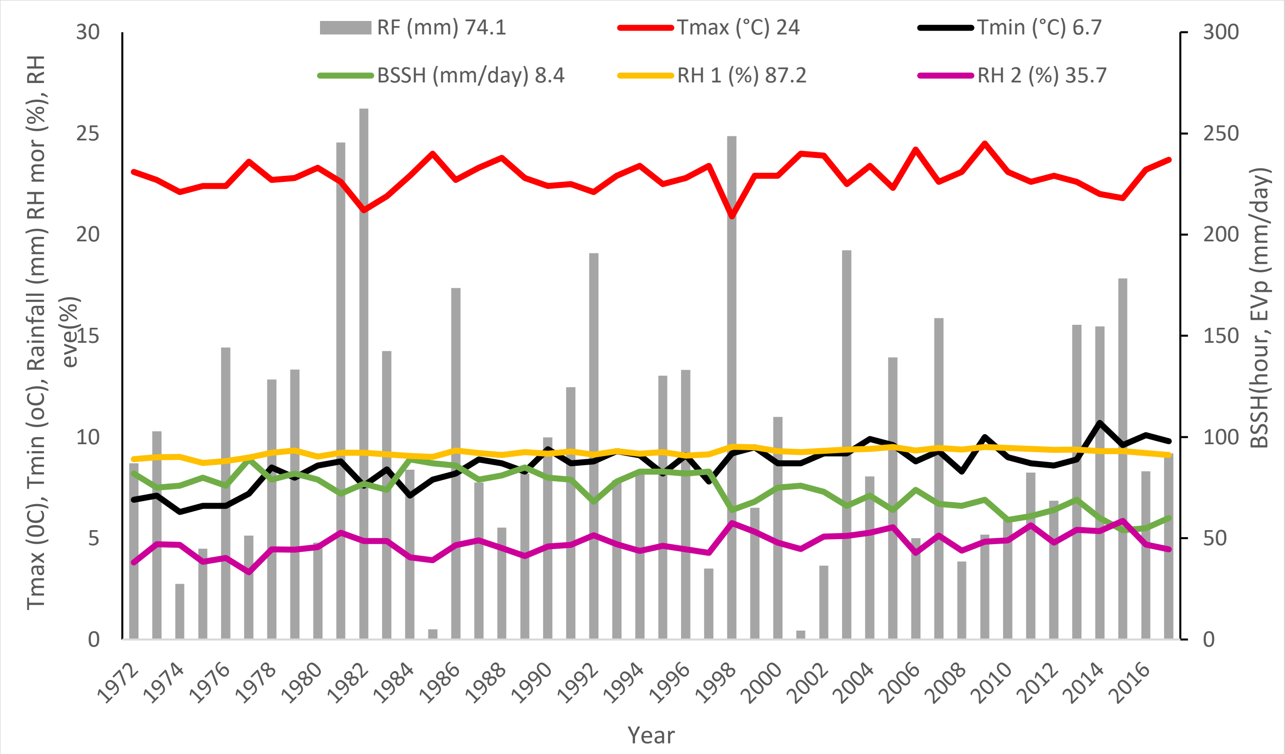


Fig.1.d. Climatic conditions for Ludhiana, Punjab

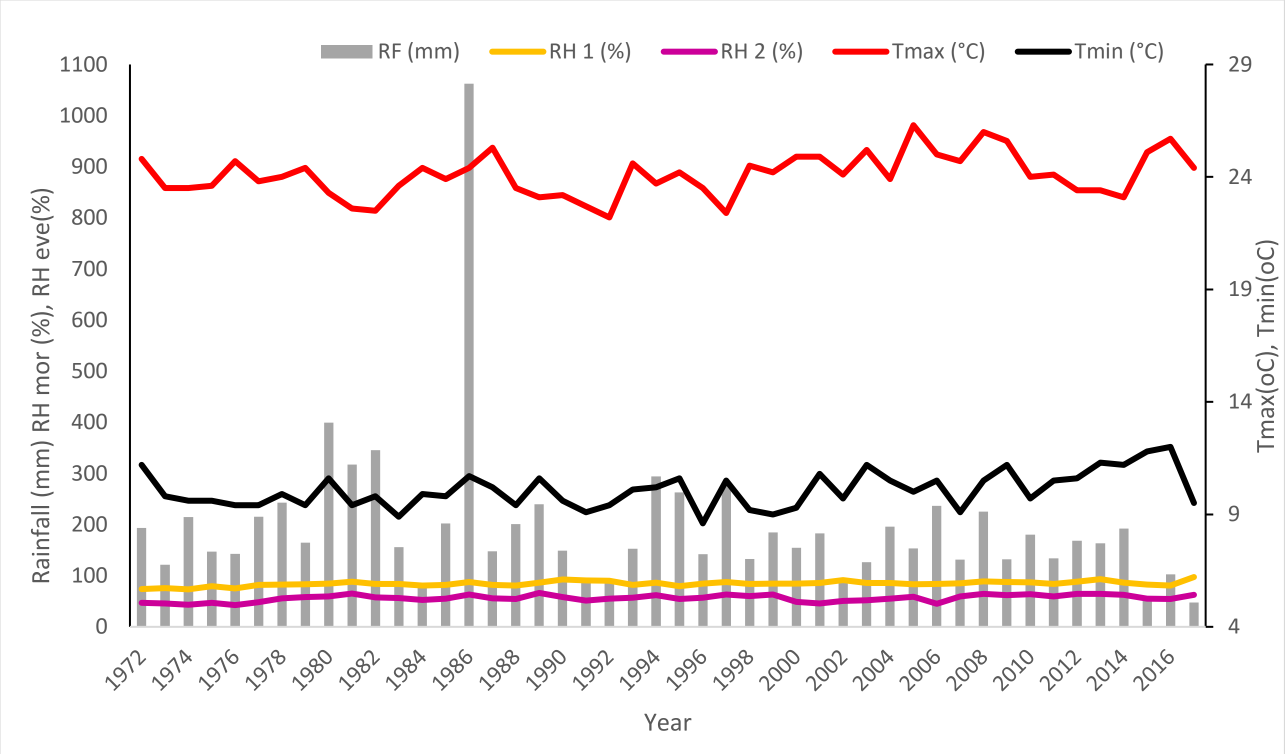


Fig.1.e. Climatic conditions for Patiala, Punjab

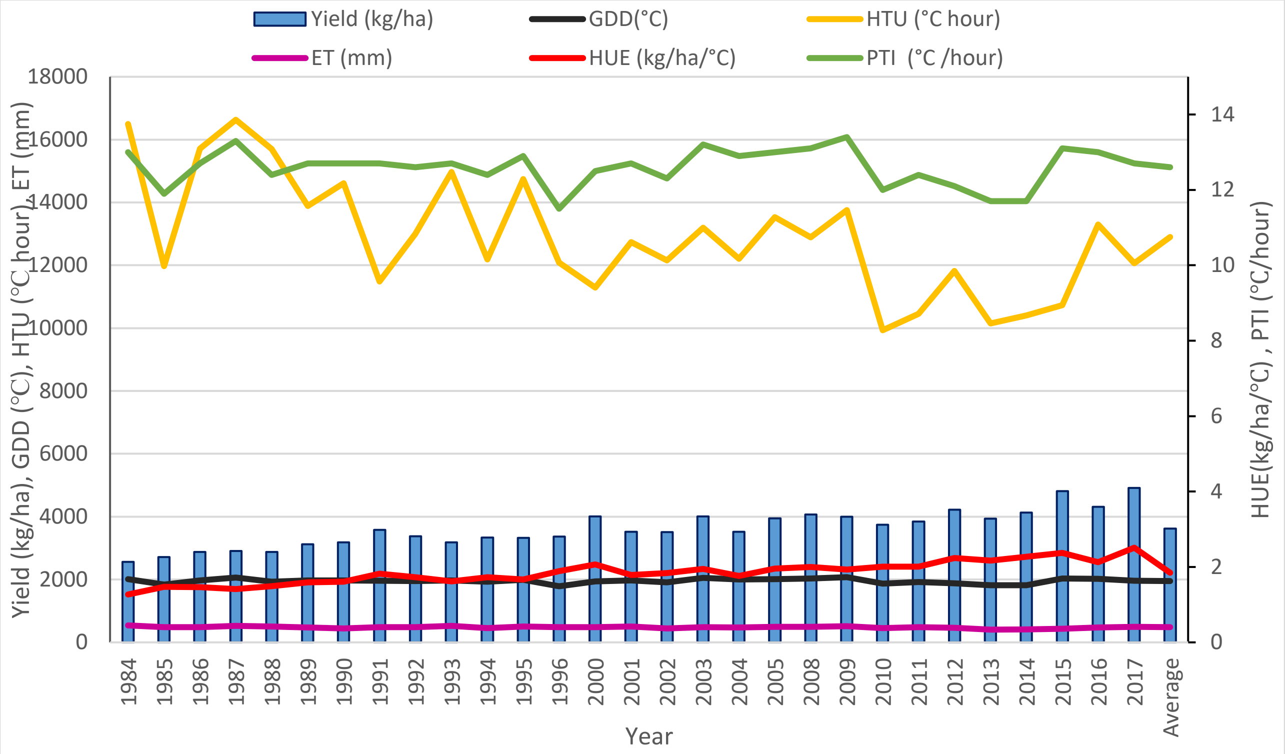


Fig.2.a. Heat indices and reference evapotranspiration during wheat crop growing period for IARI, New Delhi

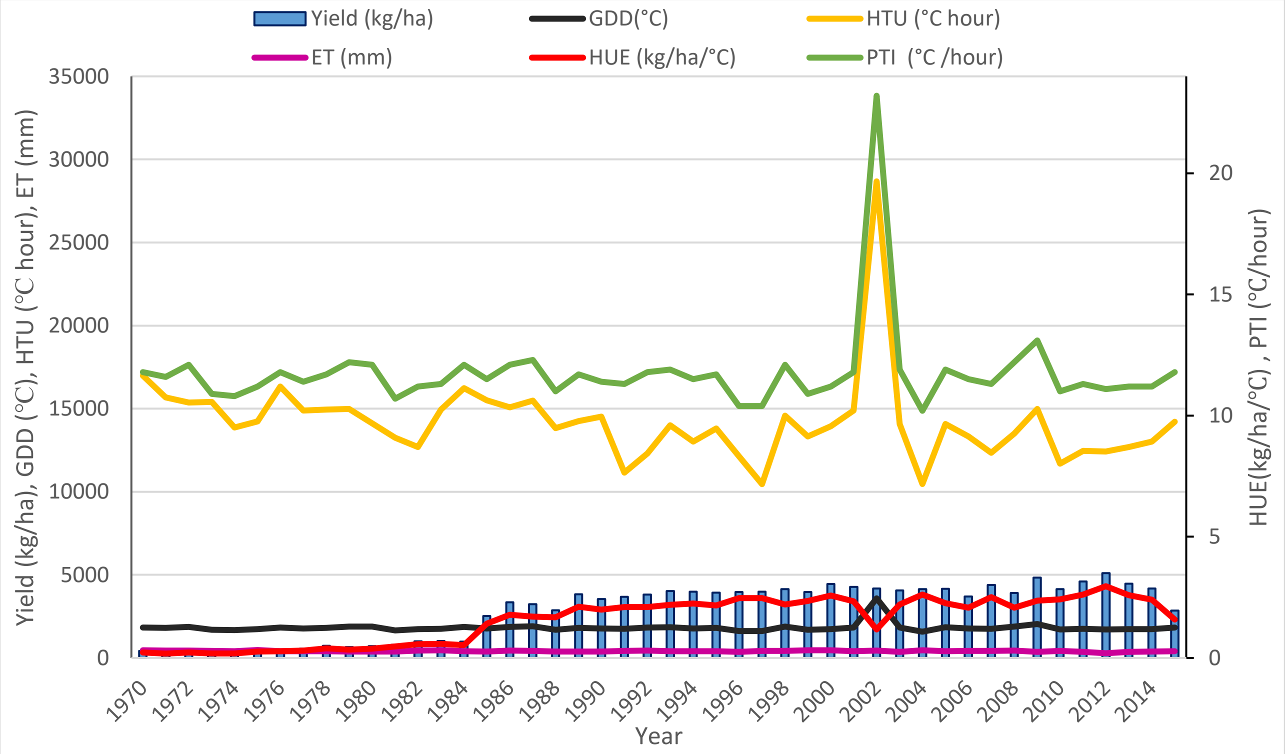


Fig.2.b Heat indices and reference evapotranspiration during wheat crop growing period for Hisar

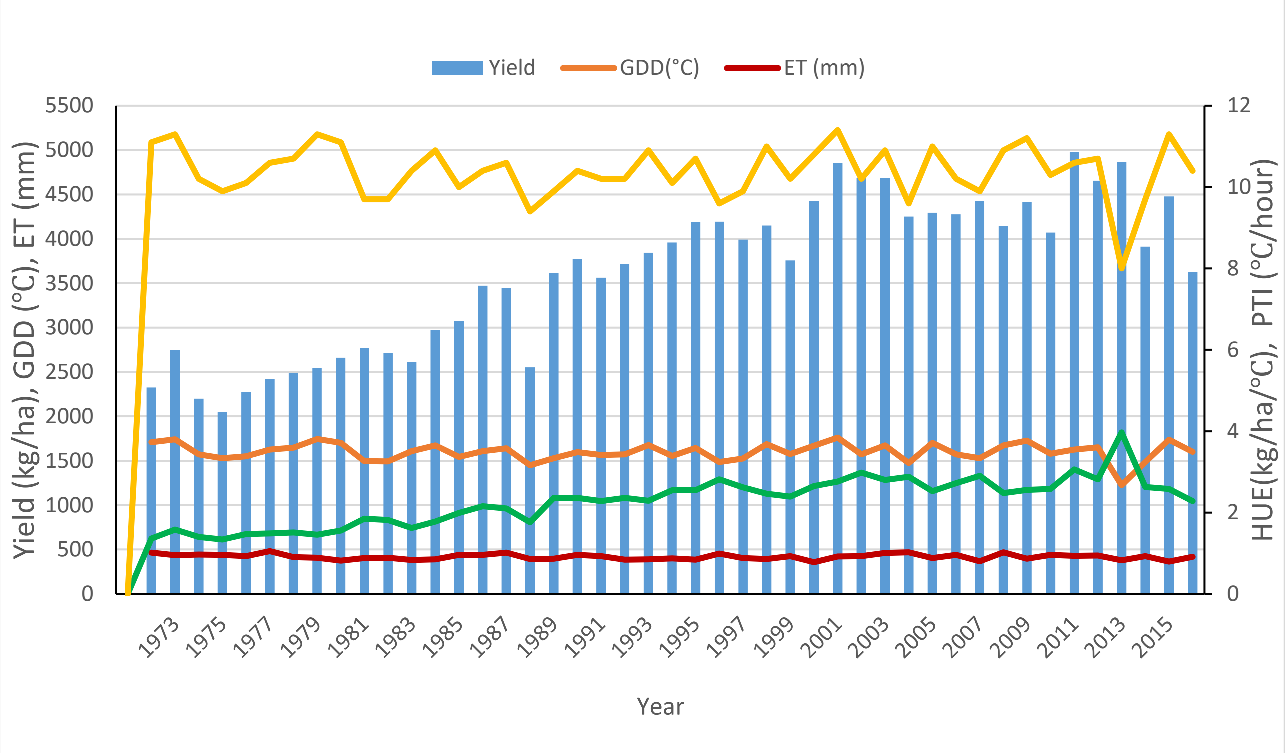


Fig.2.c Heat indices and reference evapotranspiration during wheat crop growing period for Amritsar

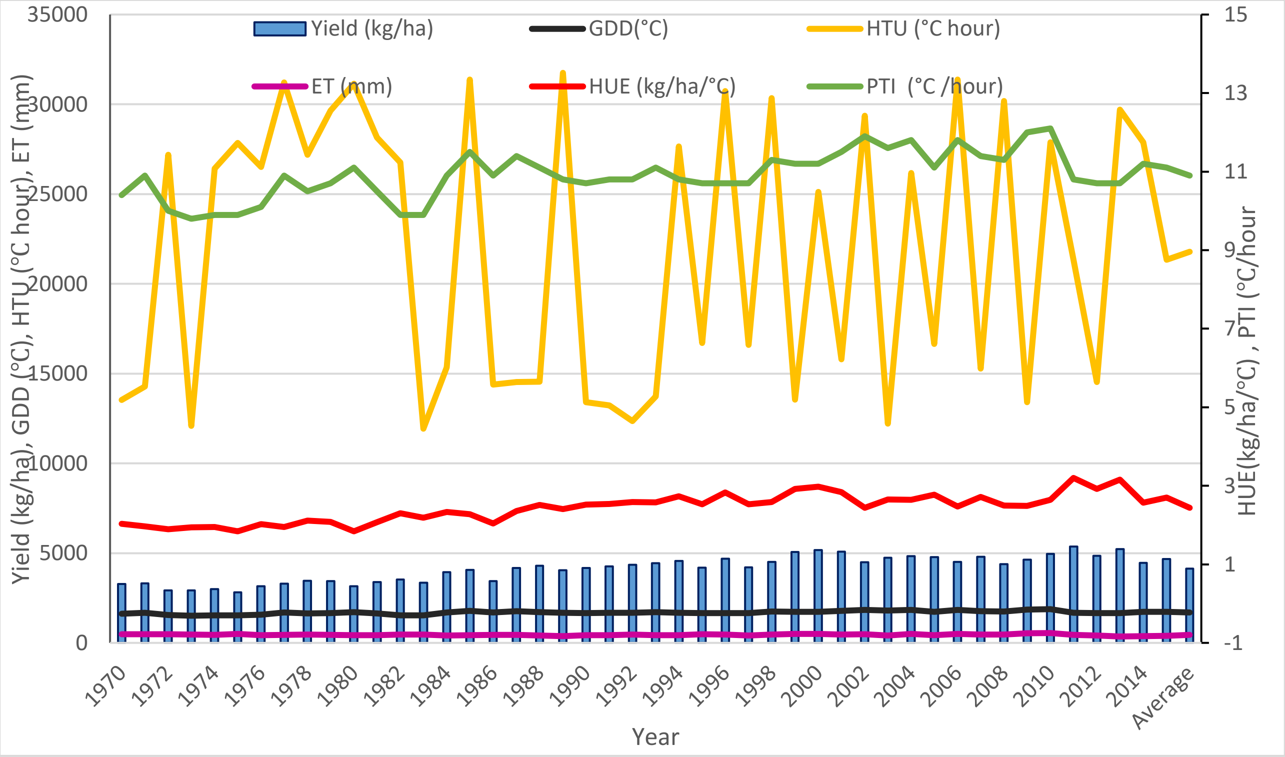


Fig.2.d Heat indices and reference evapotranspiration during wheat crop growing period for Ludhiana

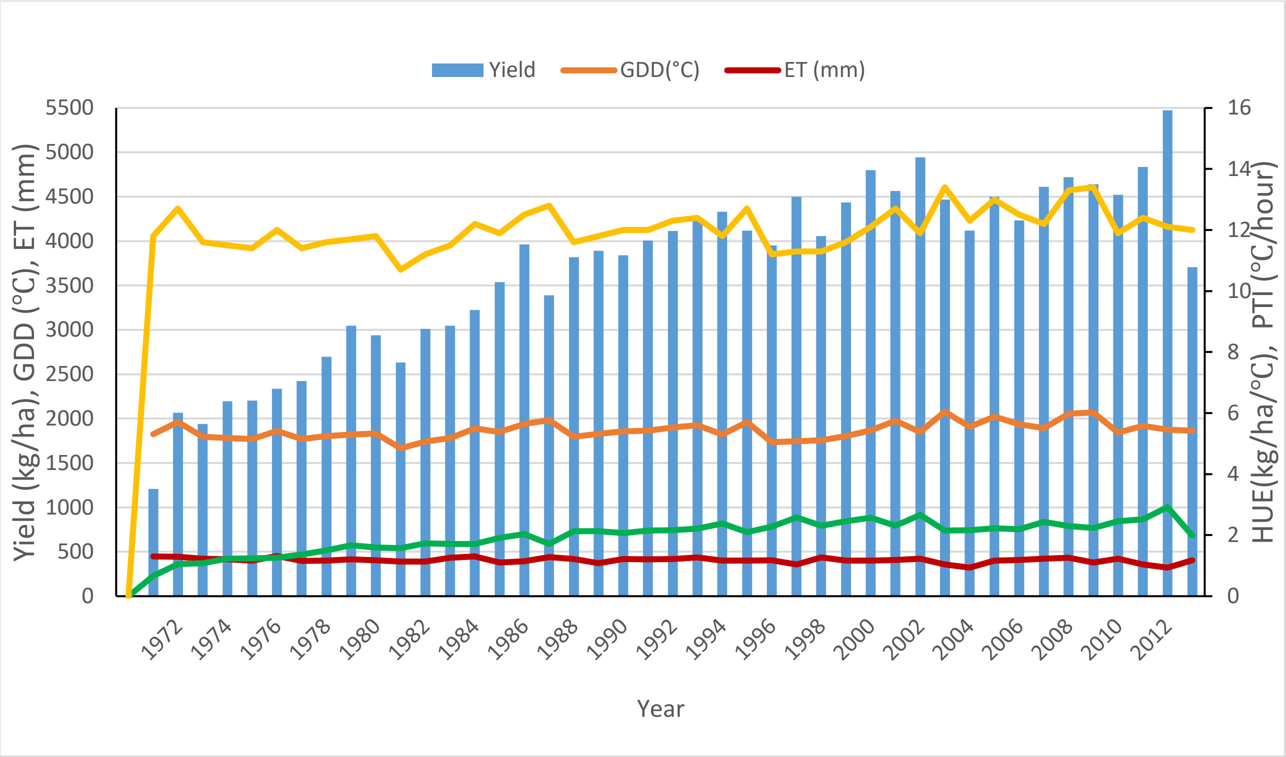


Fig.2.e Heat indices and reference evapotranspiration during wheat crop growing period for Patiala

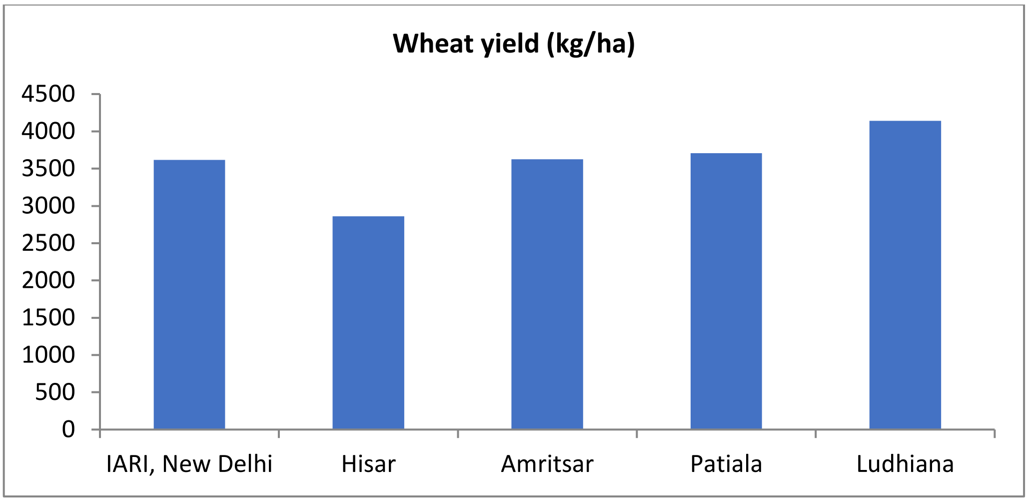


Fig.3 Average Wheat yield (kg /ha) for different station

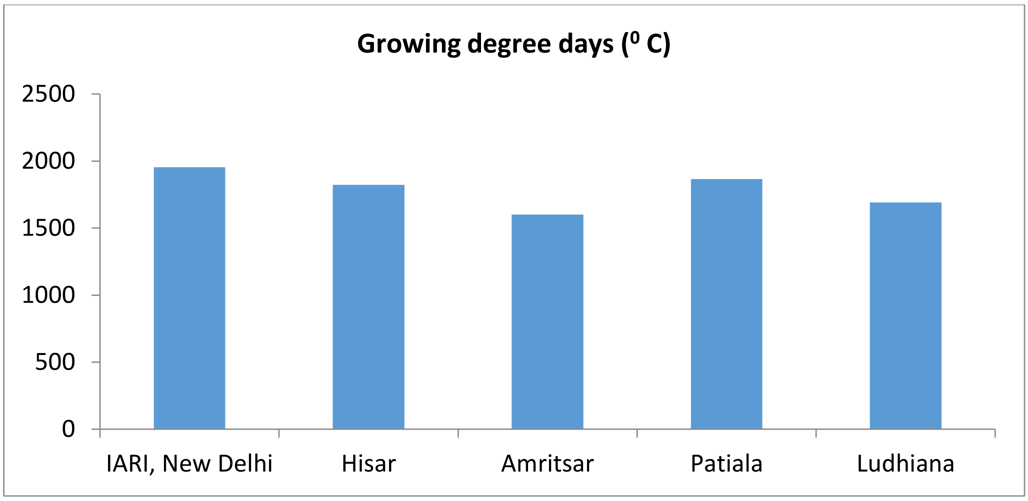


Fig 4 Average Growing degree days (^{0} C) during wheat crop growing period for different station

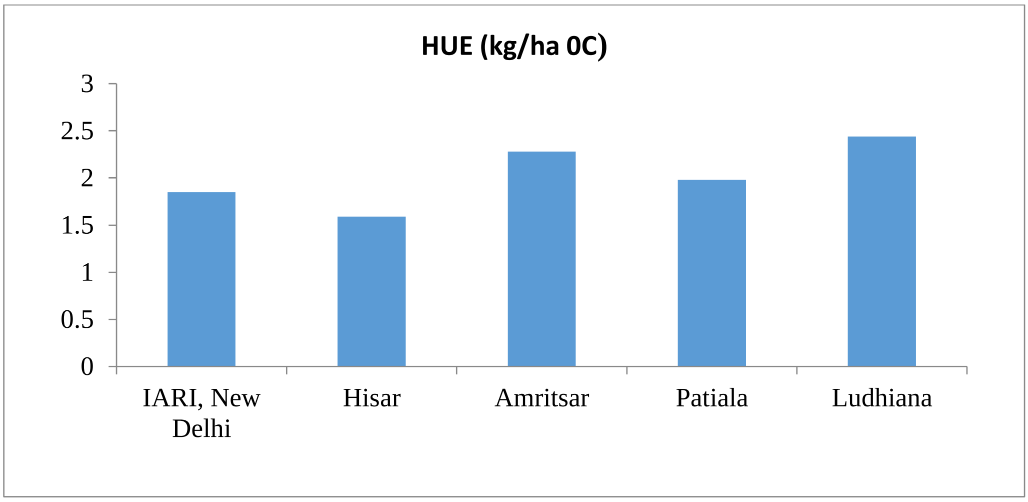


Fig 5. Average Heat use efficiency of wheat for different station

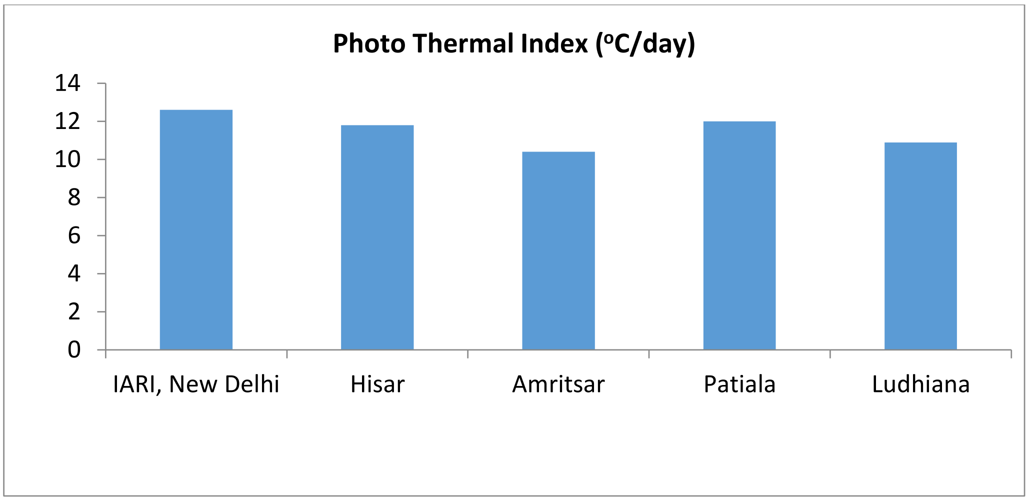


Fig. 6 Average Photo Thermal Index (ºC/day) of wheat crop for different station

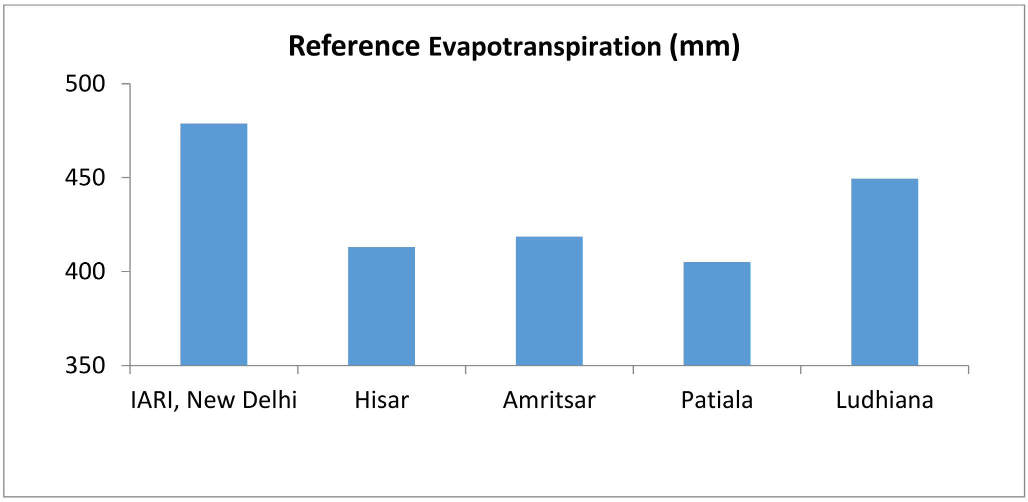


Fig. 7 Average Reference Evapotranspiration during wheat growing period for different station

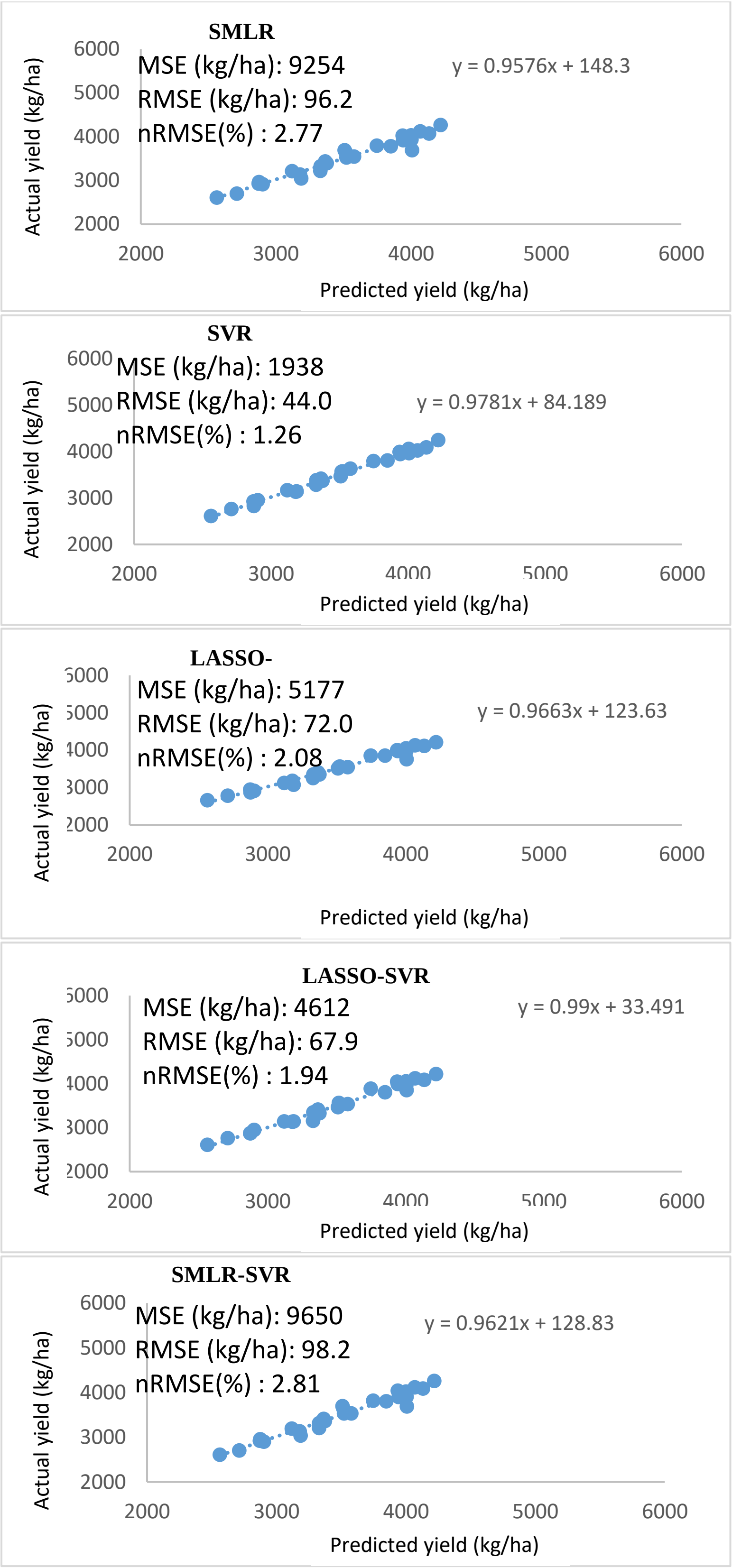


Fig.8.a Calibration of Models developed for wheat yield estimation for IARI, New Delhi.

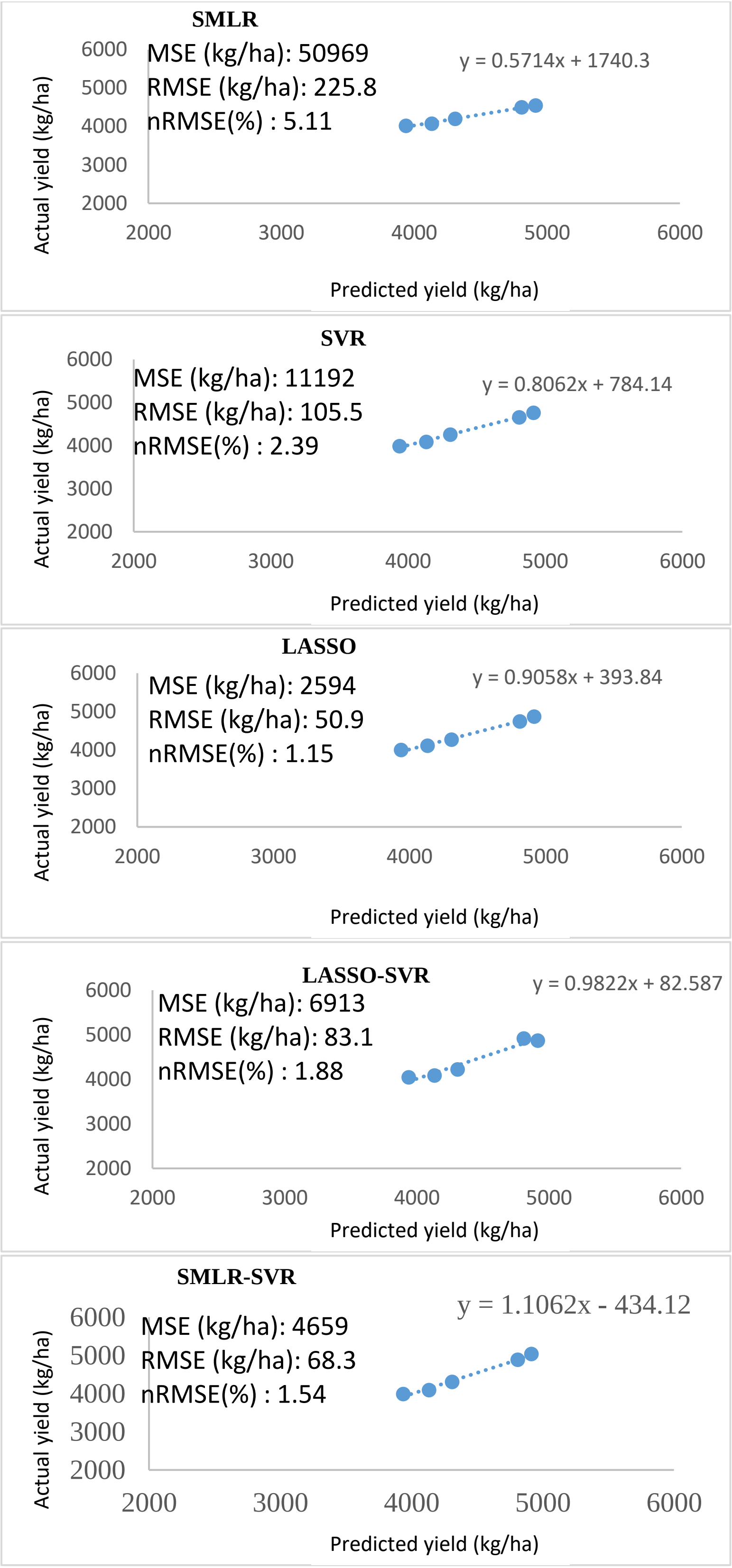


Fig.8.b Validation of Models developed for wheat yield estimation for IARI, New Delhi.

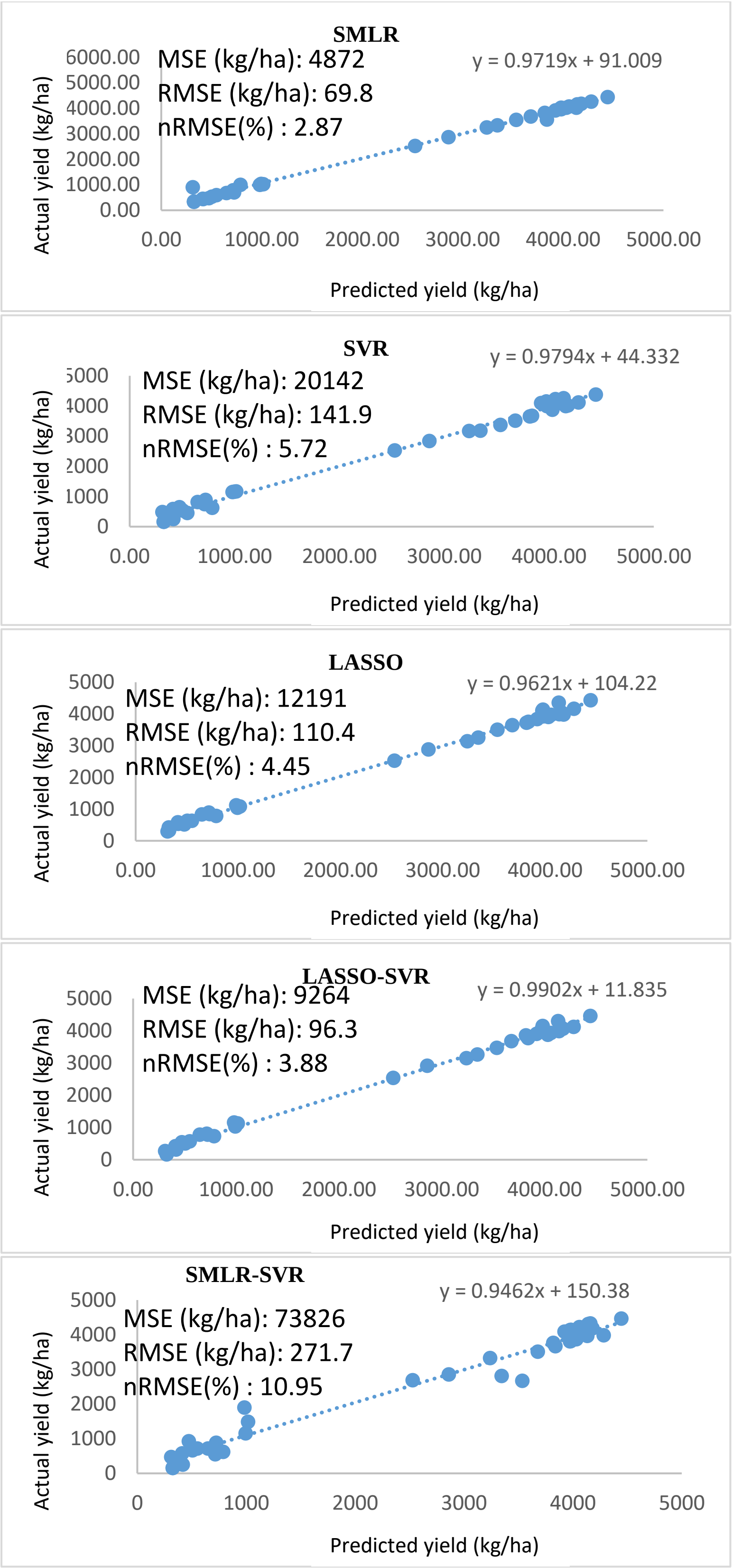


Fig.9.a Calibration of Models developed for wheat yield estimation for Hisar.

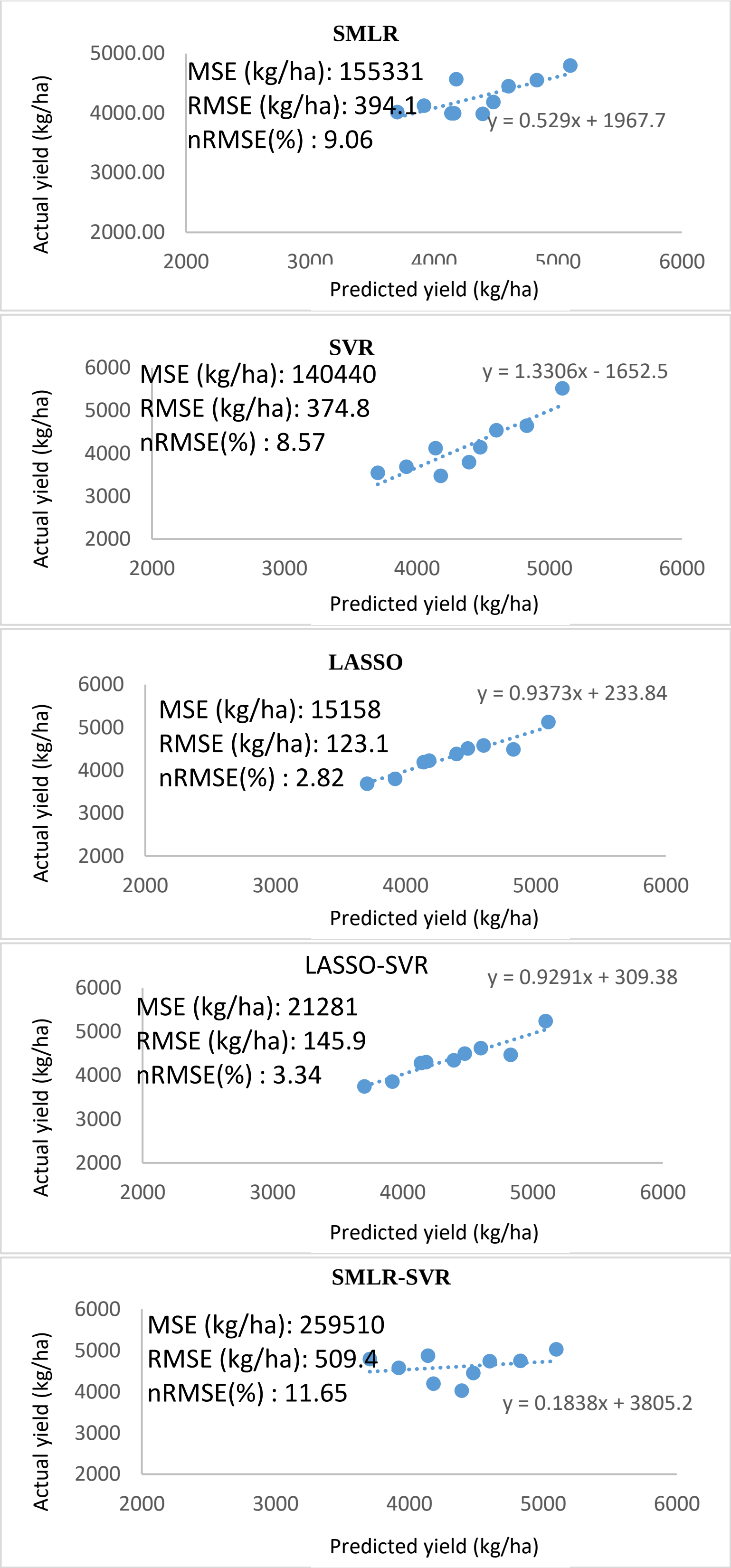


Fig.9.b Validation of Models developed for wheat yield estimation for Hisar.

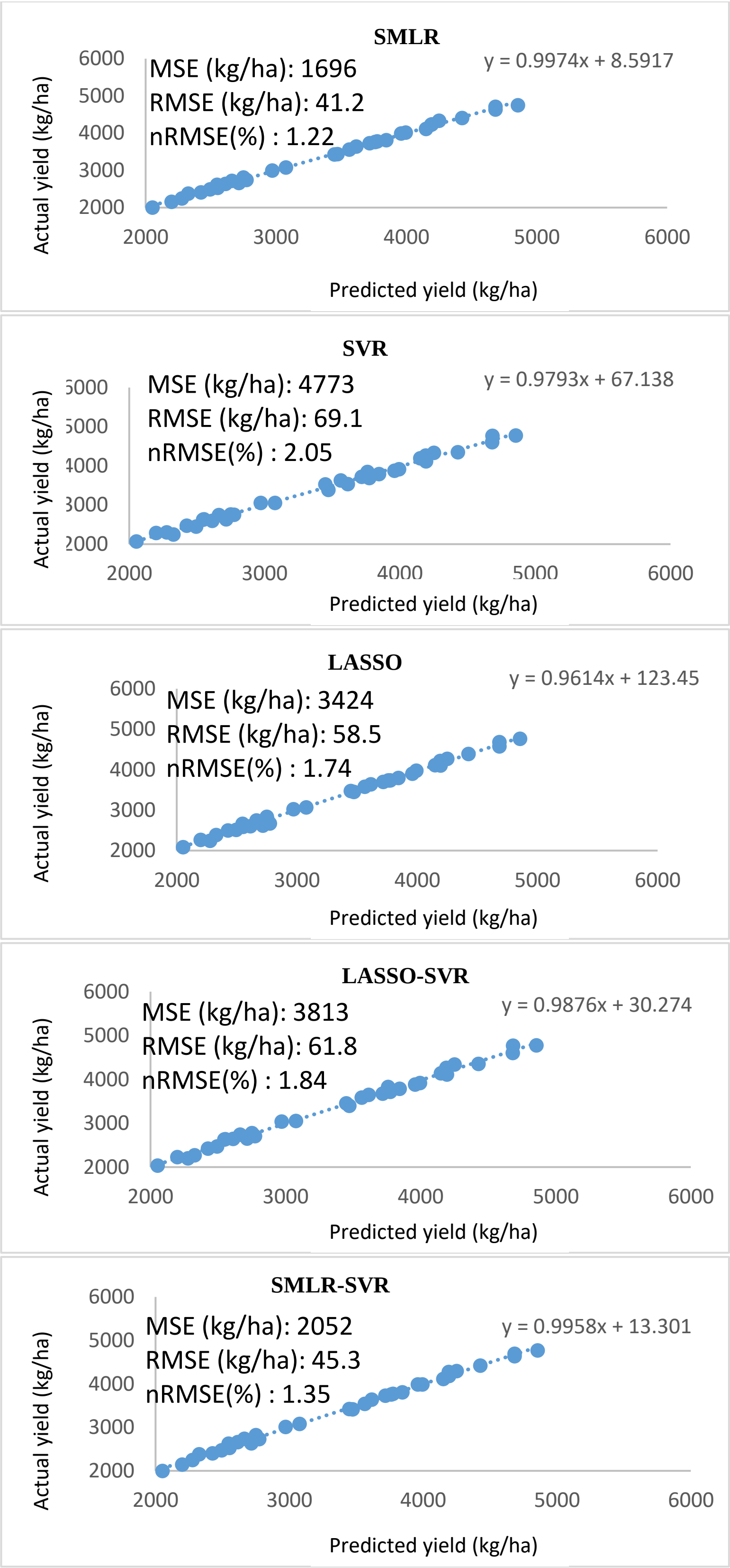


Fig.10.a Calibration of Models developed for wheat yield estimation for Amritsar.

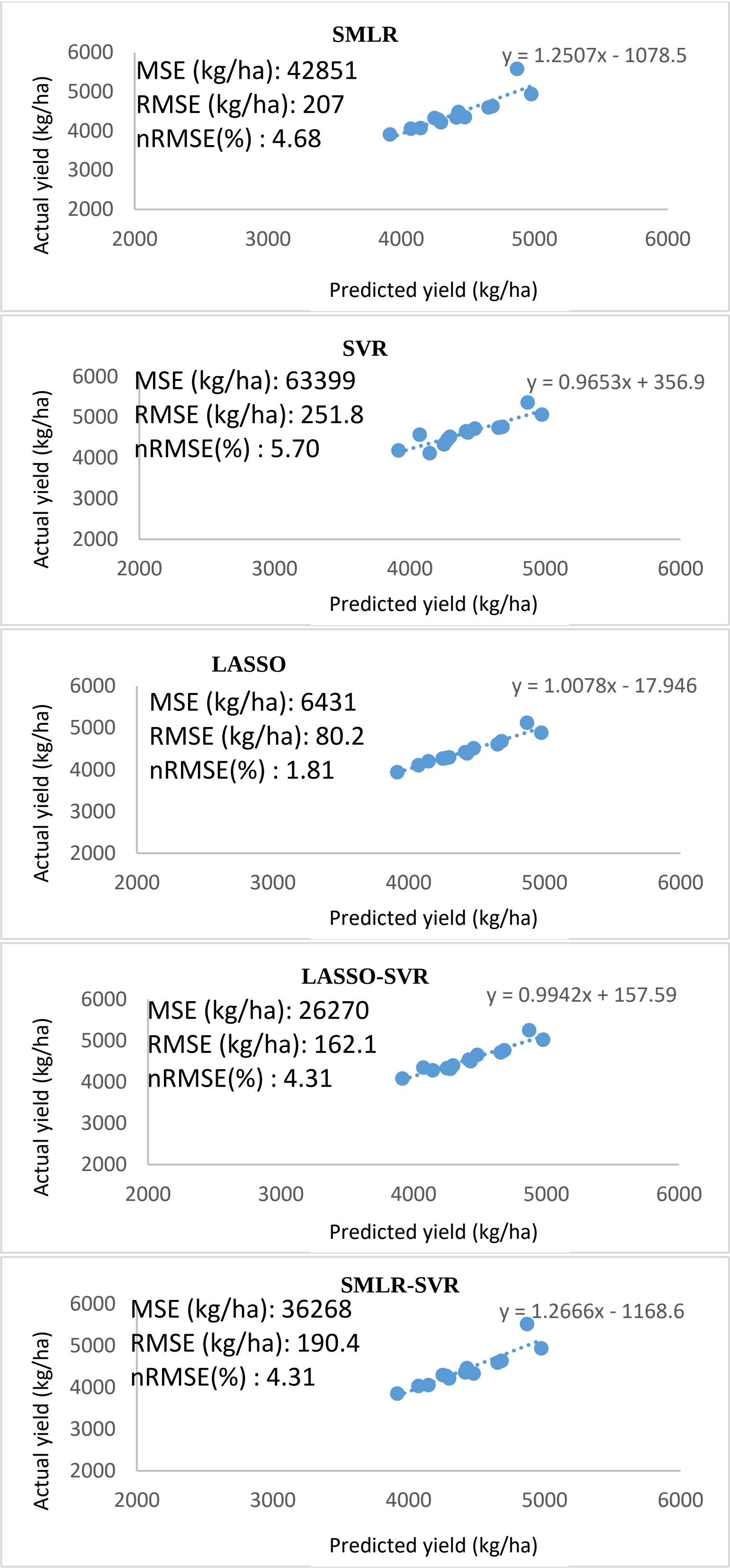


Fig.10.b Validation of Models developed for wheat yield estimation for Amritsar.

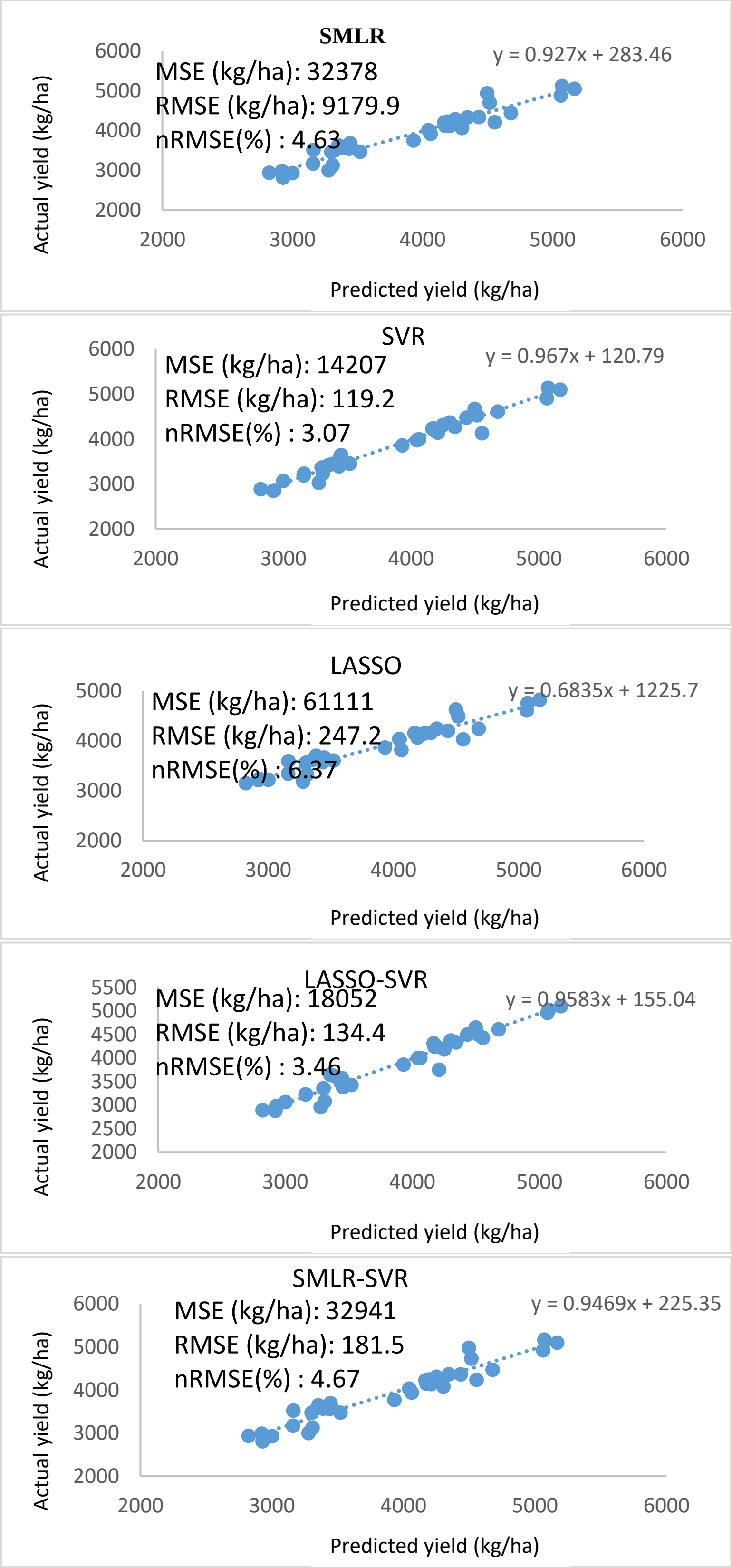


Fig.11.a Calibration of Models developed for wheat yield estimation for Ludhiana.

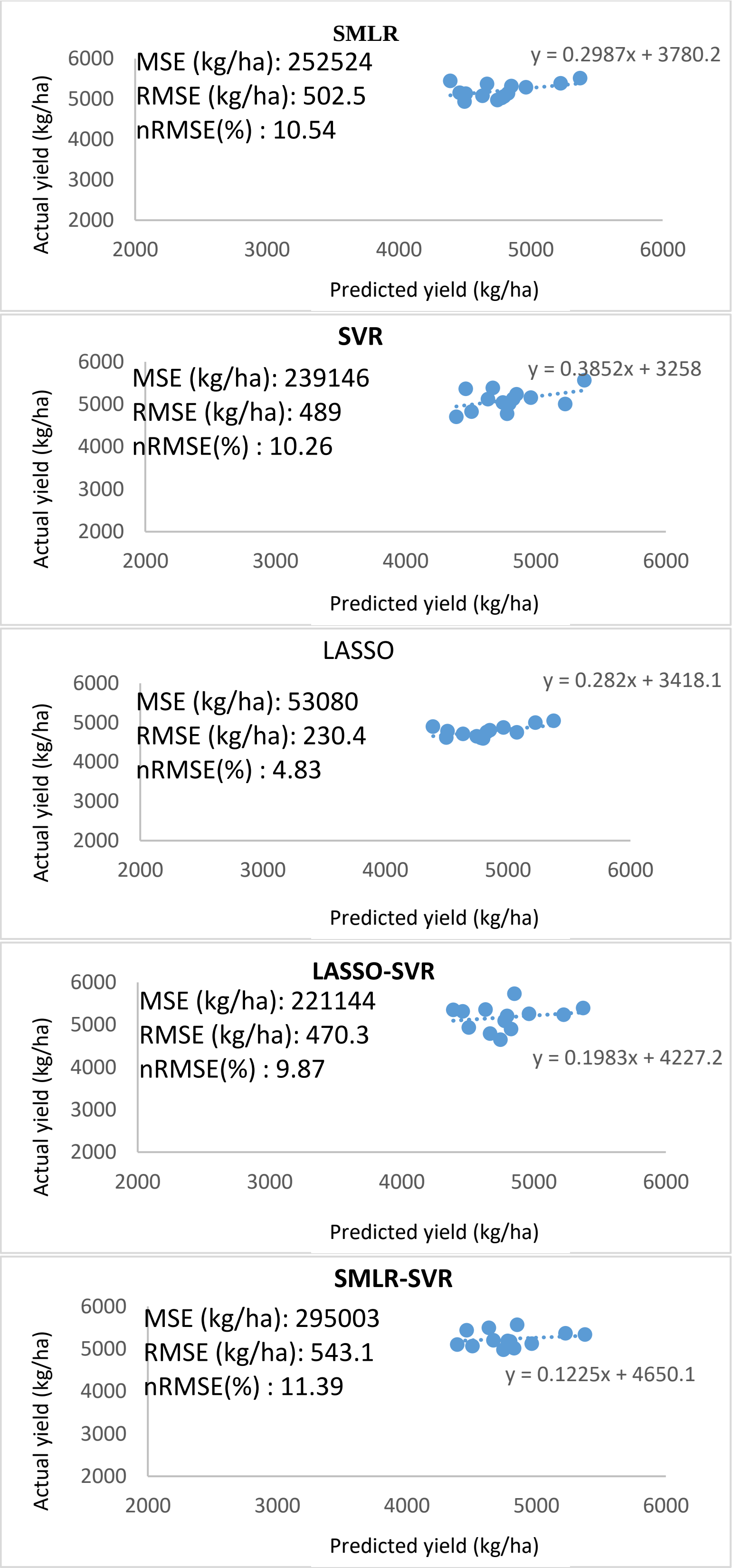


Fig.11.b Validation of Models developed for wheat yield estimation for Ludhiana.

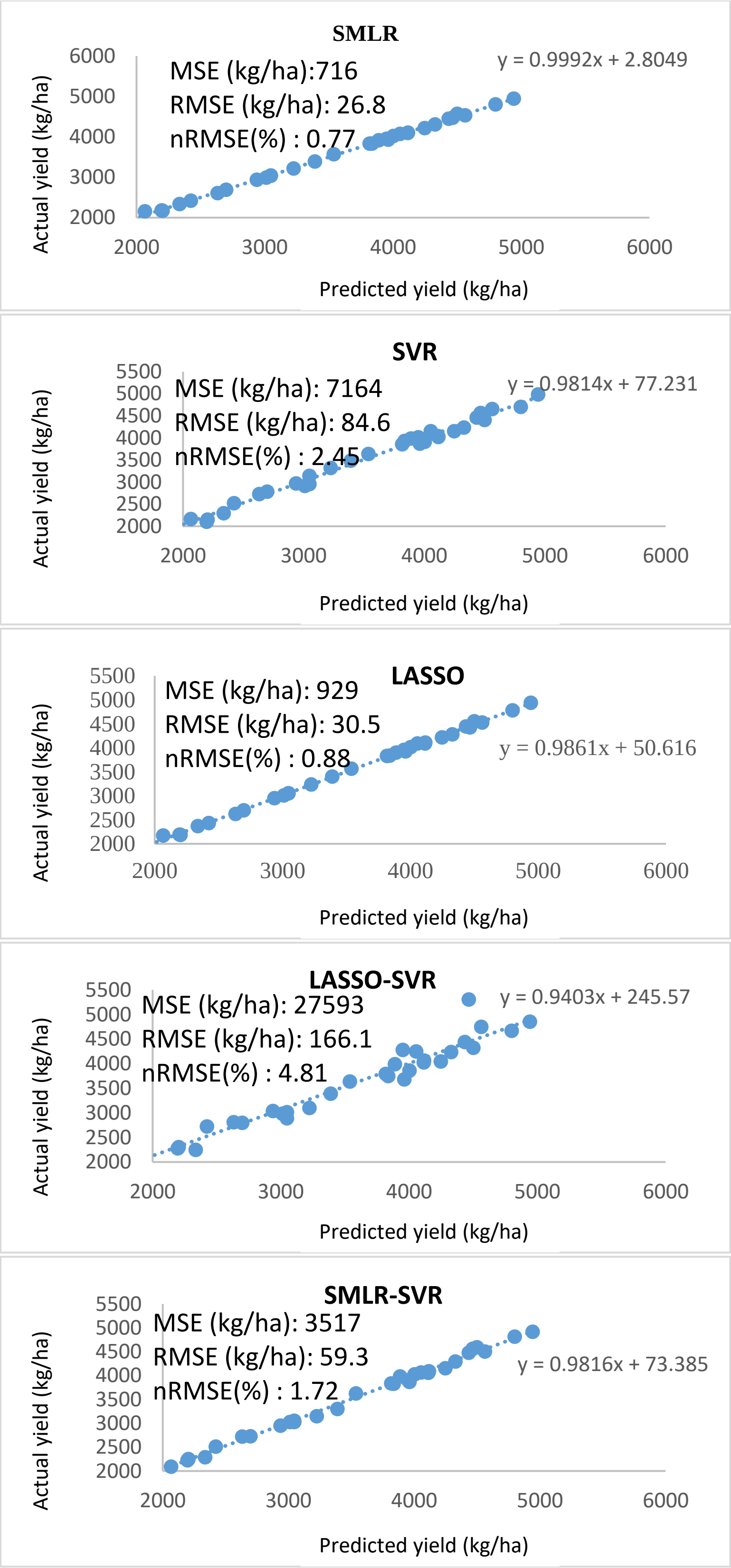


Fig.12.a Calibration of Models developed for wheat yield estimation for Patiala.

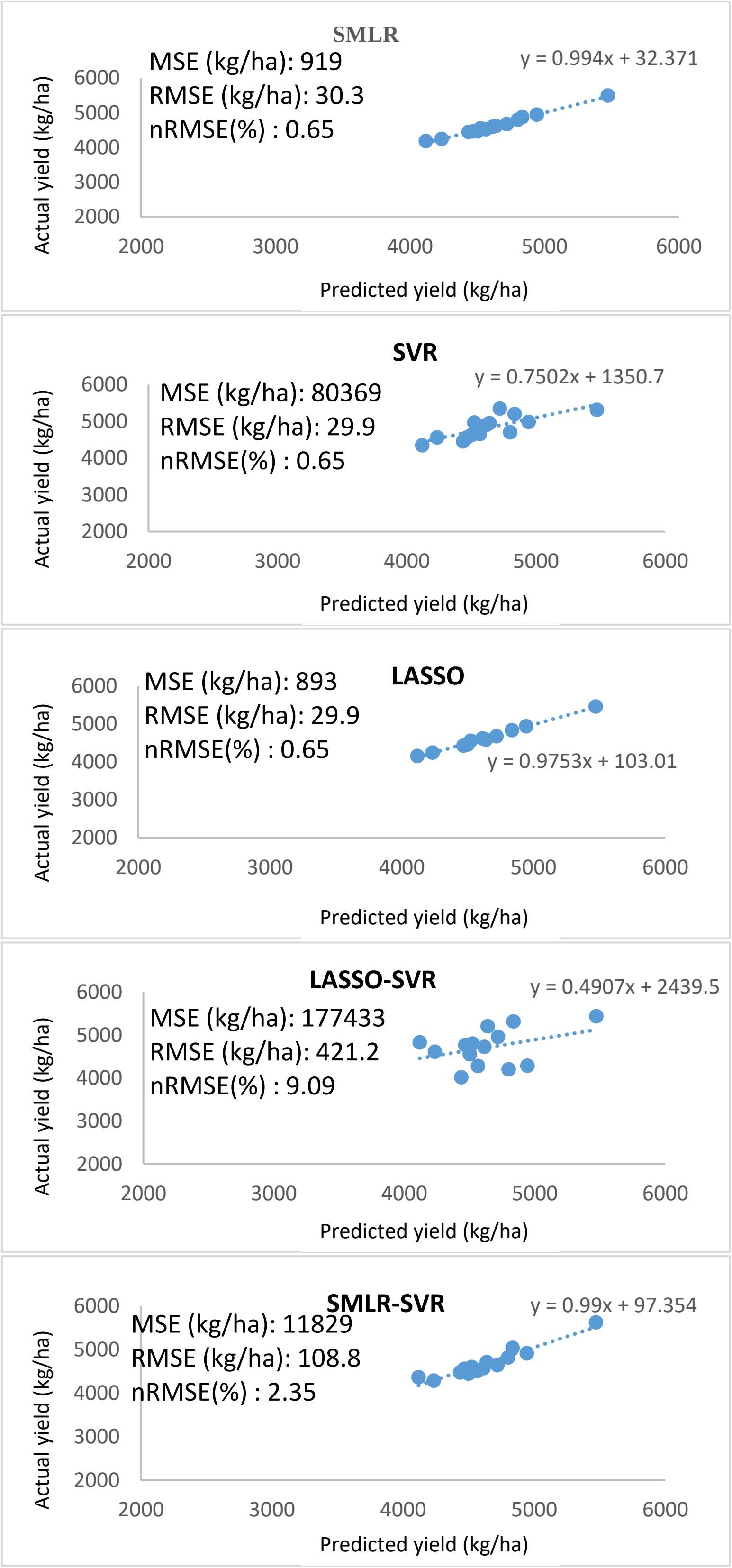


Fig.12.b Validation of Models developed for wheat yield estimation for Patiala.

Objective 2: To predict wheat yield at multiple stages using the developed model

4.4 Wheat Yield prediction at tillering stage for different district of North West India

Model for predicting wheat yield at tillering stage for different district of North-West India have been developed using long term crop yield data as well as long period daily weather data from 45th to 4th standard meteorological week for respective districts.

4.4.1 Wheat Yield prediction at tillering stage by different model for IARI, New Delhi

Model for predicting wheat yield at the tillering stage by different techniques for IARI, New Delhi have been developed using long term crop yield data as well as long period daily weather data from 46th to 4th standard meteorological week for IARI, New Delhi. The performance of the model for wheat crop yield prediction at tillering stage for IARI, New Delhi is shown in the (Table 4.12).

During calibration value of R^2 ranged between 0.89 to 0.99% having the highest value 0.99 % for SVR, 0.92 % for LASSO-SVR, 0.92% for LASSO, 0.89 % for SMLR and 0.96 for SMLR-SVR model respectively. The mean square error for calibration ranged between 1806.82 kg/ha for SVR model to 18572.32 kg/ha for SMLR model. The maximum value was found for SMLR modal followed by LASSO, LASSO-SVR, SMLR-SVR, and SVR modal. The RMSE for calibration was lowest for SVR modal (42.51 kg/ha) followed by SMLR-SVR (91.13 kg /ha), LASSO-SVR (133.41 kg/ha), LASSO (124.35kg/ha) and SMLR (136.28 kg/ha). During calibration nRMSE ranged between 1.23 to 3.95 % having the lowest value for SVR modal followed by SMLR-SVR, LASSO-SVR, LASSO and SMLR model.

The most important weather parameter identified by SMLR are time, Z281 (Minimum temperature* evapotranspiration) and Z671 (sunshine hour*evaporation) while parameters for LASSO are time, Z51 (minimum relative humidity* maximum temperature), Z151 (maximum temperature*minimum relative humidity), Z261 (Minimum temperature *bright sunshine hours), Z281 (Minimum temperature* evapotranspiration), Z671 (bright sunshine hours*evaporation) and Z581 (evening relative humidity*evapotranspiration). Equation developed for wheat yield prediction at tillering stage for IARI, New Delhi by SMLR, LASSO is given in Table 4.13.The parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01351351 and number of Support Vectors: 20. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.125, epsilon: 0.1 and number of Support Vectors: 20 and for SMLR-SVR the parameters are SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.04761905, epsilon: 0.1, and number of Support Vectors: 26.

The mean square error for validation ranged between 32469.20 kg/ha for LASSO-SVR model to 67872 kg/ha for SMLR-SVR model. The maximum value was found for SMLR-SVR followed by SMLR, SVR, LASSO and LASSO-SVR model respectively. The value of RMSE during validation was lowest for LASSO-SVR (2180.19 kg/ha) model followed by LASSO (188.91 kg/ha), SVR (208.59 kg/ha) SMLR (254.36 kg/ha), and SMLR-SVR (260.52 kg/ha) model.

The nRMSE value for validation ranged between 4.25 to 6.14%. Maximum value of nRMSE was found for SMLR-SVR (6.14%) followed by SMLR (6.00%), SVR (4.92%), LASSO (4.45%), and minimum for LASSO-SVR (4.25%). respectively. Results showed that the performance of wheat yield prediction for IARI, New Delhi done at tillering stage was excellent for all the models having nRMSE < 10. The calibration and validation for all models are shown in (fig 13a & 13b).

The percentage deviation of yield estimation done at tillering stage for IARI, New Delhi by observed yield using the model developed was lowest for SMLR-SVR (-3.12 %) model followed by LASSO-SVR (-4.71%), LASSO (-5.33 %), SMLR (-6.62 %) and SVR (-7.41%). The negative sign indicates underestimation of the wheat crop yield at the tillering stage for IARI, New Delhi.

Table 4.12 Wheat yield prediction at tillering stage by different modal for IARI, New Delhi

S.No.	NAME OF MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation (%)
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.89	18572.32	136.28	3.95	64700.57	254.36	6.00	-6.62
2	SVR	0.99	1806.82	42.51	1.23	43511.68	208.59	4.92	-7.41
3	LASSO	0.92	18050.04	134.35	3.84	35687.30	188.91	4.45	-5.33
4	LASSO-SVR	0.92	17799.52	133.41	3.85	32469.20	180.19	4.25	-4.71
5	SMLR-SVR	0.96	8303.98	91.13	2.63	67872.30	260.52	6.14	-3.12

Table 4.13: Equation for wheat yield prediction at tillering stage for IARI, New Delhi

Model	Equation
SMLR	Y=3564.33+51.96*time+ 5.39*Z281+ 12.84*Z671
LASSO	Y=57.76*time+ 2.81*Z51+ 0.29*Z151+0.94*Z261+ 6.49*Z671+ 3.44*Z281+ 0.30*Z581

4.4.2 Wheat Yield prediction at tillering stage by different model for Hisar, Haryana

Model for predicting the wheat yield at the tillering stage by different techniques for Hisar have been developed using long term crop yield data as well as long period daily weather data from 46th to 4th standard meteorological week for Hisar. The performance of the model for wheat crop yield prediction at tillering stage for Hisar is shown in the (Table 4.14).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During calibration value of R^2 ranged between 0.92 to 0.99% having the highest value 0.99 % for SVR, 0.94 % for LASSO-SVR, 0.93 for LASSO and 0.92 % for SMLR and SMLR-SVR model respectively. The mean square error for calibration was ranged between 38198 kg/ha for SVR model to 226758 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, SMLR, LASSO-SVR and SVR modal. The RMSE for calibration was lowest for SVR modal (195.4 kg/ha) followed by LASSO-SVR (419.1 kg/ha), SMLR (466.0 kg/ha), SMLR-SVR (470.5 kg /ha) and LASSO (476.2 kg/ha). During the calibration nRMSE ranged between 7.87 to 19.18 % having the lowest value for SVR modal followed by LASSO-SVR, SMLR, SMLR-SVR and LASSO model.

The various weather parameter identified by SMLR for Patiala are time, Z10 (maximum temperature) and Z61 (sunshine hours*maximum temperature), while the parameters identified by LAASO are time, Z61 (sunshine hours*maximum temperature), Z351 (rainfall*evening relative humidity), Z121 (maximum temperature*minimum temperature), Z141 (maximum temperature*morning relative humidity), Z181 (maximum temperature*evapotranspiration), Z231 (Minimum temperature*rainfall), Z251 (Minimum temperature*evening relative humidity), Z261 ((Minimum temperature *bright sunshine hours), Z461 (morning relative humidity*bright sunshine hours), Z560 (evening relative humidity*bright sunshine hours) Z571 (evening relative humidity*evaporation) and Z681 (sunshine hours*evapotranspiration). Equation developed for wheat yield prediction at tillering stage for Hisar by SMLR, LASSO is given in Table 4.15.

The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01369863, epsilon: 0.1 and number of Support Vectors: 31; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.07142857, epsilon: 0.1 and number of Support Vectors was 27, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.333333, epsilon: 0.1 and number of Support Vectors was 30.

The value of RMSE during validation was lowest for LASSO (338.8 kg/ha) model followed by SVR (729.4 kg/ha), LASSO-SVR (818.3 kg/ha) SMLR-SVR (932.7 kg/ha) and SMLR (988.5 kg/ha) model. The mean square error for validation ranged between 114795 kg/ha for LASSO model to 977116 kg/ha for SMLR model. The maximum value was found for SMLR followed by SMLR-SVR, LASSO-SVR, SVR and LASSO model respectively.

The nRMSE value for validation ranged between 7.79 to 22.72%. The maximum value of nRMSE was found for SMLR (22.72%) followed by SMLR-SVR (21.44%), LASSO-SVR (18.81%), SVR (16.77%), and the minimum for LASSO (7.79%) respectively. Results showed that the performance of wheat yield prediction done at the tillering stage for Hisar was excellent for LASSO model having nRMSE < 10%, good for SVR and LASSO-SVR having nRMSE < 20%, fair for SMLR-SVR and SMLR having nRMSE value < 25 %. The calibration and validation for all models are shown in (fig 14a & 14b).

The percentage deviation of yield estimation done at tillering stage for Hisar by observed yield using the model developed was the lowest for % SMLR (7.58%) model followed by SVR (10.15), LASSO (12.07%), LASSO-SVR (14.97 %), SMLR-SVR (20.62%) respectively.

Table 4.14 Wheat yield prediction at tillering stage by different modal for Hisar

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage deviation %
		R² (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.92	217192	466.0	18.78	977116	988.5	22.72	7.58
2	SVR	0.99	38198	195.4	7.87	532053	729.4	16.77	10.15
3	LASSO	0.93	226758	476.2	19.18	114795	338.8	7.79	12.07
4	LASSO-SVR	0.94	175617	419.1	16.88	669586	818.3	18.81	14.97
5	SMLR-SVR	0.92	221375	470.5	18.96	869873	932.7	21.44	20.62

Table 4.15: Equation for wheat yield prediction at tillering stage for Hisar

Model	Equation
SMLR	Y=-5103+ 138.79*time+ 24.00*Z10+ 295.47*Z61
LASSO	Y=96.89*time+ 19.71*Z61+ 1.51*Z121+ 0.05*Z141+ 2.39*Z231+ 0.007*Z251+ 16.46*Z261+ 0.05*Z351+ 0.0007*Z461- 0.03*Z5606+ 0.08*Z561+3.50*Z571+ 0.88*Z181+ 8.09*Z681

4.4.3 Wheat Yield prediction at tillering stage by different model for Amritsar, Punjab

Model for predicting wheat yield at tillering stage by different techniques for Amritsar have been developed using long term crop yield data as well as long period daily weather data from 46th to 4th standard meteorological week for Amritsar. The performance of the model for wheat crop yield prediction at tillering stage for Amritsar is shown in the (Table 4.16).

During the calibration value of R^2 ranged between 0.92 to 0.97% having the highest value 0.99 % for SVR, 0.92 % for LASSO, LASSO-SVR, SMLR and SMLR-SVR model respectively. The mean square error for calibration was ranged between 22369 kg/ha for SVR model to 76397 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR, SMLR-SVR, LASSO-SVR and SVR modal. The RMSE for calibration was the lowest for SVR modal (149.6 kg/ha) followed by LASSO-SVR (195.0 kg/ha), SMLR-SVR (203.6 kg /ha), SMLR (242.4 kg/ha) and LASSO (276.4 kg/ha). During calibration nRMSE ranged between 4.45 to 8.22 % having the lowest value for SVR modal followed by LASSO-SVR, SMLR-SVR, SMLR and LASSO model.

The most important weather parameter identified by SMLR for are time, Z151 (maximum temperature*evening relative humidity) and Z481 (morning relative humidity *evapotranspiration) and while parameters for LASSO are time, Z31 (morning relative humidity*maximum temperature),Z81 (evapotranspiration *maximum temperature), Z131 (maximum temperature*rainfall), Z141 (maximum temperature*morning relative humidity), Z151 (maximum temperature*evening relative humidity), and Z451 (morning relative humidity * evening relative humidity). Equation developed for wheat yield prediction at tillering stage for Amritsar by SMLR, LASSO is given in Table 4.17. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02702703, epsilon: 0.1 and number of Support Vectors: 24. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.1428571, epsilon: 0.1 and number of Support Vectors was 21 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and number of Support Vectors was 19.

The value of RMSE during validation was lowest for LASSO (217.0 kg/ha) followed by SMLR (359.4 kg/ha), SVR (374.4 kg/ha), LASSO-SVR (433.0 kg/ha), SMLR-SVR (497.7 kg/ha) and. The mean square error for validation ranged between 47102 kg/ha for LASSO to 247680 kg/ha for SMLR-SVR. The maximum value was found for SMLR-SVR followed by LASSO-SVR, SVR, SMLR and LASSO respectively. The nRMSE value for validation ranged between 4.91 to 11.26% With the maximum value for maximum value for SMLR-SVR (11.26%) followed by LASSO-SVR (9.80%), SVR (8.47%), SMLR (8.13%),and the minimum for LASSO (4.91%). respectively.

Results showed that the performance of wheat yield prediction for Amritsar done at tillering stage was excellent for all the models having nRMSE < 10% except performed good for SMLR-SVR having nRMSE value 11.26 < 15 %. The calibration and validation for all models are shown in (fig 15a & 15b).

The percentage deviation of yield estimation done at tillering stage for Amritsar by observed yield using the model developed was lowest for LASSO (-0.29%) model followed by SMLR-SVR (-2.82%), LASSO-SVR (-5.90 %), SVR (-12.11%) and SMLR (12.65%) modal respectively. The negative sign indicates the under estimation of wheat yield at tillering stage for Amritsar.

Table 4.16 Wheat yield prediction at tillering stage by different modal for Amritsar

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.92	58738	242.4	7.21	129135	359.4	8.13	12.65
2	SVR	0.97	22369	149.6	4.45	140200	374.4	8.47	-12.11
3	LASSO	0.92	76397	276.4	8.22	47102	217.0	4.91	-0.29
4	LASSO-SVR	0.92	38033	195.0	5.80	187518	433.0	9.80	-5.90
5	SMLR-SVR	0.92	41451	203.6	6.05	247680	497.7	11.26	-2.82

Table 4.17: Equation for wheat yield prediction at tillering stage for Amritsar

Model	Equation
SMLR	Y=2676.07+62.72*time+0.941*Z151+0.681*Z481
LASSO	Y=52.11*time+0.53*Z31+0.07*Z81+0.29*Z131+0.78*Z141+0.52* Z151+0.46*Z451

4.4.4 Wheat Yield prediction at tillering stage by different model for Ludhiana, Punjab

Model for predicting wheat yield at tillering stage by different techniques for Ludhiana have been developed using long term crop yield data as well as long period daily weather data from 46th to 4th standard meteorological week for Ludhiana. The performance of the model for wheat crop yield prediction at tillering stage for Ludhiana is shown in the (Table 4.18).

During the calibration value of R^2 ranged between 0.91 to 0.97% having the highest value 0.97 % for SVR, 0.94 % for LASSO-SVR, 0.93 % for LASSO, 0.91 % for SMLR and SMLR-SVR model respectively. The mean square error for calibration was ranged between 11696 kg/ha for SVR model to 39847 kg/ha for SMLR model. The maximum value was found for SMLR modal followed by LASSO, LASSO-SVR SMLR-SVR and SVR modal. The RMSE for calibration was the lowest for SVR modal (108.2 kg/ha) followed by SMLR-SVR (126.2 kg /ha), LASSO-SVR (159.4 kg/ha), LASSO (180.7 kg/ha) and SMLR (199.6 kg/ha). During the calibration nRMSE ranged between 2.78 to 5.14 % having lowest value for SVR modal followed by SMLR-SVR, LASSO-SVR, LASSO and SMLR model.

The most important weather parameter identified by SMLR are time and Z161 (maximum temperature*bright sunshine hours), while parameters for LASSO are time, Z10 (maximum temperature), Z11 (maximum temperature* minimum temperature), Z41 (morning relative humidity*maximum temperature), Z21 (minimum temperature* maximum temperature), Z151 (maximum temperature*evening relative humidity), Z161 (maximum temperature*bright sunshine hours), Z230 (Minimum temperature*rainfall), Z251 (Minimum temperature*evening relative humidity), Z261 (Minimum temperature *bright sunshine hours), Z361 (rainfall*bright sunshine hours), Z561 (evening relative humidity*bright sunshine hours) and Z360 (rainfall*bright sunshine hours). Equation developed for wheat yield prediction at tillering stage for Ludhiana by SMLR, LASSO is given in Table 4.19. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 31. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.2325581, epsilon: 0.1 and number of Support Vectors was 31 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.5, epsilon: 0.1 and number of Support Vectors was 23.

The mean square error for validation ranged between 18935 kg/ha for LASSO model to 311716 kg/ha for SMLR-SVR model. The maximum value was found for SMLR-SVR followed by SVR, SMLR, LASSO-SVR and LASSO model respectively. The value of RMSE during validation was lowest for LASSO (137.6 kg/ha) model followed by LASSO-SVR (464.0 kg/ha), SMLR (474.0 kg/ha), SVR (539.7 kg/ha) and SMLR-SVR (558.3 kg/ha) model.

The nRMSE value for validation ranged between 2.89 to 11.71%. Maximum value of nRMSE was found for SMLR-SVR (11.71%) followed by SVR (11.32%), SMLR (9.94%), LASSO-SVR (9.73%), and the minimum for LASSO (2.89%). respectively. Results showed that performance of wheat yield prediction for Ludhiana done at tillering stage was excellent for LASSO, LASSO-SVR and SMLR models having nRMSE < 10%, performed good for SVR and SMLR-SVR having nRMSE value < 15 %. The calibration and validation for all models are shown in (fig 16a & 16b).

The percentage deviation of yield estimation done at tillering stage for Ludhiana by observed yield using the model developed was the lowest for LASSO (0.62%) model followed by SMLR (5.23%), LASSO-SVR (9.09 %), SVR (9.59%) and SMLR-SVR (10.33%) modal respectively.

Table 4.18 Wheat yield prediction at tillering stage by different modal for Ludhiana

S.No.	MODEL	CALIBERATION				VALIDATION			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.91	39846.98	199.62	5.14	224666.40	473.99	9.94	5.23
2	SVR	0.97	11695.95	108.15	2.78	291248.85	539.67	11.32	9.59
3	LASSO	0.93	32634.23	180.65	4.65	18935.16	137.61	2.89	0.62
4	LASSO-SVR	0.94	25413.53	159.42	4.10	215293.42	464.00	9.73	9.09
5	SMLR-SVR	0.91	15920.23	126.18	3.25	311715.81	558.32	11.71	10.33

Table 4.19: Equation for wheat yield prediction at tillering stage for Ludhiana

Model	Equation
SMLR	Y=1562.73+66.82*time+3.62*Z161
LASSO	Y=46.77*time+2.69*Z10+2.82*Z11+8.0*Z211+8.46*Z41+0.78*Z151+ 2.48*Z161+0.07*Z231+0.69*Z251+0.35*Z261+0.08*Z350-0.64*Z360+ 2.00*Z361+0.07*Z561

4.4.5 Wheat Yield prediction at tillering stage by different model for Patiala, Punjab

Model for predicting wheat yield at tillering stage by different techniques for Patiala have been developed using long term crop yield data as well as long period daily weather data from 46th to 4th standard meteorological week for Patiala. The performance of the model for wheat crop yield prediction at tillering stage for Patiala is shown in the (Table 4.20).

During the calibration value of R^2 ranged between 0.95 to 0.96% having the highest value 0.96 % for SVR and SMLR, 0.95 % for LASSO, LASSO-SVR and SMLR-SVR model respectively. The mean square error for calibration was ranged between 37512 kg/ha for SVR model to 53792 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, SMLR, LASSO-SVR and SVR modal. The RMSE for calibration was lowest for SVR modal (193.7 kg/ha) followed by LASSO-SVR (205.6 kg/ha), SMLR (207.6 kg/ha), SMLR-SVR (211.9 kg /ha) and LASSO (231.9 kg/ha). During the calibration nRMSE ranged between 5.61 to 6.72 % having lowest value for SVR modal followed by LASSO-SVR, SMLR, SMLR-SVR and LASSO model.

The various weather parameter identified by SMLR for Patiala are time, Z41(morning relative humidity*maximum temperature) and Z381 (rainfall*evapotranspiration) while the parameters identified by LAASO are time, Z11 (maximum temperature),Z41 (morning relative humidity*maximum temperature), Z140 (maximum temperature*morning relative humidity), Z180 (maximum temperature*evapotranspiration), Z241 (minimum temperature*morning relative humidity), Z281 (Minimum temperature* evapotranspiration), Z351(rainfall* evening relative humidity), Z381 (rainfall*evapotranspiration), Z451 (morning relative humidity* evening relative humidity) and Z481 (morning relative humidity*evapotranspiration). Equation developed for wheat yield prediction at tillering stage for Patiala by SMLR, LASSO is given in Table 4.21.

The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 25; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors was 25, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.333333, epsilon: 0.1 and number of Support Vectors was 22.

The value of RMSE during validation was the lowest for LASSO (233.4 kg/ha) model followed by SVR (378.8 kg/ha), LASSO-SVR (448.9 kg/ha) SMLR (526.8 kg/ha) and SMLR-SVR (563.7 kg/ha) model. The mean square error for validation ranged between 54452 kg/ha for LASSO model to 317740 kg/ha for SMLR-SVR model. The maximum value was found for SMLR-SVR followed by SMLR, LASSO-SVR, SVR and LASSO model respectively.

The nRMSE value for validation ranged between 5.06 to 12.17%. The maximum value of nRMSE was found for SMLR-SVR (12.17%) followed by SMLR (11.37%), LASSO-SVR (9.69%), SVR (8.18%), and the minimum for LASSO (5.06%) respectively. Results showed that the performance of wheat yield prediction for Patiala done at tillering stage was excellent for LASSO, SVR and LASSO-SVR having nRMSE < 10%, performed good for SMLR and SMLR-SVR having nRMSE value < 15 %. The calibration and validation for all models are shown in (fig 17a & 17b).

The percentage deviation of yield estimation done at tillering stage for Patiala by observed yield using the model developed was the lowest for SMLR (1.39%) model followed by SMLR-SVR (2.31%) SVR (3.94 %), LASSO (-5.36%) and LASSO- SVR (8.02%) and modal respectively. Negative sign indicates the under estimation of wheat yield at tillering stage for Patiala by LASSO model.

Table 4.20 Wheat yield prediction at tillering stage by different modal for Patiala

S.No.	NAME OF MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage deviation
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.96	43114	207.6	6.01	277548	526.8	11.37	1.39
2	SVR	0.96	37512	193.7	5.61	143474	378.8	8.18	3.94
3	LASSO	0.95	53792	231.9	6.72	54452	233.4	5.06	-5.36
4	LASSO-SVR	0.95	42266	205.6	5.95	201501	448.9	9.69	8.02
5	SMLR-SVR	0.95	44891	211.9	6.14	317740	563.7	12.17	2.31

Table 4.21: Equation for wheat yield prediction at tillering stage for Patiala

Model	Equation
SMLR	Y=-690.66+ 80.44*time+ 6.52*Z41+ 0.02*Z381
LASSO	Y=57.98*time+ 44.94*Z11+9.23*Z41- 0.05*Z180+0.30* Z241- 2.26*Z281+ 0.003*Z351+ 0.01*Z381- 0.02*Z481

4.5 Wheat Yield Estimation at flowering stage for different district of North West India

Model for estimating wheat yield at flowering stage for different district of North-West India have been developed using long term crop yield data as well as long period daily weather data from 45th to 8th standard meteorological week for respective districts.

4.5.1 Wheat Yield prediction at flowering stage by different model for IARI, New Delhi

Model for predicting wheat yield at flowering stage by different techniques for IARI, New Delhi have been developed using long term crop yield data as well as long period daily weather data from 46th to 8h standard meteorological week for IARI, New Delhi. The performance of the model for wheat crop yield prediction at flowering stage for IARI, New Delhi is shown in the (Table 4.22).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During the calibration value of R^2 ranged between 0.95 to 0.99% having the highest value 0.99 % for SVR, LASSO, LASSO-SVR, SMLR-SVR and 0.95 % for SMLR model respectively. The mean square error for calibration was ranged between 1584 kg/ha for LASSO-SVR model to 9837 kg/ha for SMLR model. The maximum value was found for SMLR modal followed by LASSO, SMLR-SVR, SVR and LASSO-SVR modal. The RMSE for calibration was the lowest for LASSO-SVR modal (39.8 kg/ha) followed by SVR (44.0 kg /ha), SMLR-SVR (44.5 kg/ha), LASSO (47.9 kg/ha), and SMLR (99.2 kg/ha). During the calibration nRMSE ranged between 1.15 to 2.84% having the lowest value for LASSO-SVR modal followed by SVR, SMLR-SVR, LASSO and SMLR model.

The most important weather parameter identified by SMLR are time, Z41 (morning relative humidity*maximum temperature), Z80 (evapotranspiration *maximum temperature), and Z281 (Minimum temperature* evapotranspiration), while parameters for LASSO are time, Z10 (maximum temperature), Z21 (minimum temperature*maximum temperature) Z41(morning relative humidity*maximum temperature) Z121(maximum temperature*minimum temperature), Z150 (maximum temperature*minimum relative humidity), Z151 (maximum temperature*evening relative humidity), Z231(Minimum temperature*rainfall), Z261((Minimum temperature *bright sunshine hours), Z271((Minimum temperature*evaporation), Z371(rainfall*evaporation), Z451(morning relative humidity*evening relative humidity), Z461 (morning relative humidity*bright sunshine hours), Z560 (evening relative humidity*sunshine hours), Z671 (sunshine hour*evaporation), Z281 (Minimum temperature* evapotranspiration), and Z581 (evening relative humidity*evapotranspiration). Equation developed for wheat yield prediction at flowering stage for IARI, New Delhi by SMLR, LASSO is given in Table 4.23. The parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01351351, epsilon: 0.1 and number of Support Vectors: 21. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.05, epsilon: 0.1 and number of Support Vectors was 14 and for SMLR-SVR the parameters are SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.03846154, epsilon: 0.1 nnd number of Support Vectors was 21.

The value of RMSE during validation was the lowest for LASSO (55.1 kg/ha) followed by LASSO-SVR (131.0 kg/ha), SVR (157.9 kg/ha), SMLR (240.5 kg/ha) and SMLR-SVR (289.7 kg/ha). The mean square error for validation ranged between 3038 kg/ha for LASSO to 83914 kg/ha for SMLR-SVR. The maximum value was found for SMLR-SVR followed by SMLR, SVR, LASSO-SVR and LASSO respectively. The nRMSE value for validation ranged between 1.30 to 6.83% with the maximum value for SMLR-SVR (6.83%) followed by SMLR (5.67%), SVR (3.72%), LASSO-SVR (3.09%) and minimum for LASSO (1.30 %) model respectively. Results showed that the performance of wheat yield prediction done at flowering stage for IARI, New Delhi was exceptional for all models having nRMSE < 10%. The calibration and validation for all models are shown in (fig 18a & 18b).

The percentage deviation of yield estimation done at flowering stage for IARI, New Delhi by observed yield using the model developed was the lowest for LASSO-SVR (-0.85 %) model followed by LASSO (-0.91 %), SVR (-2.05%), SMLR-SVR (-3.60%) and SMLR (-7.56%)modal respectively. Negative sign indicates the under estimation of wheat yield at flowering stage for IARI, New Delhi.

Table 4.22 Wheat yield prediction at flowering stage by different modal for IARI, New Delhi

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Pecentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.95	9837	99.2	2.84	57830	240.5	5.67	-7.56
2	SVR	0.99	1938	44.0	1.27	24918	157.9	3.72	-2.05
3	LASSO	0.99	2297	47.9	1.37	3038	55.1	1.30	-0.91
4	LASSO-SVR	0.99	1584	39.8	1.15	17157	131.0	3.09	-0.85
5	SMLR-SVR	0.99	1983	44.5	1.27	83914	289.7	6.83	-3.60

Table 4.23: Equation for wheat yield prediction at flowering stage for IARI, New Delhi

Model	Equation
SMLR	Y=2814.20+ 48.36*time+ 12.18*Z41+ 28.17*Z80+ 3.78*Z281
LASSO	Y=58.75*time+ 0.71*Z10+ 8.34*Z21+ 28.34*Z41+ 0.58*Z121+ 0.04*Z150+ 0.16*Z151+ 0.09*Z231+.1.22*Z261+ 1.69*Z271+ 2.81*Z371+ 0.07*Z451+ 0.02*Z560+ 6.03*Z671+ 0.04*Z181+ 1.85*Z281- 0.15*Z481+1.77*Z581

4.5.2 Wheat Yield prediction at flowering stage by different model for Hisar

Model for predicting wheat yield at flowering stage by different techniques for Hisar have been developed using long term crop yield data as well as long period daily weather data from 46th to 8h standard meteorological week for Hisar. The performance of the model for wheat crop yield prediction at flowering stage for Hisar is shown in the (Table 4.24).

During the calibration value of R^2 ranged between 0.93 to 0.99% having the highest value 0.99 % for SVR, 0.94 % for LASSO and LASSO-SVR, 0.93 % for SMLR and SMLR-SVR model respectively. The mean square error for calibration was ranged between 33677 kg/ha for SVR model to 220446 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, SMLR, LASSO-SVR and SVR modal. The RMSE for calibration was the lowest for SVR modal (183.5 kg/ha) followed by LASSO-SVR (403.7 kg/ha), SMLR (434.9 kg /ha), SMLR-SVR (441.1 kg/ha) and LASSO (469.5 kg/ha). During the calibration nRMSE ranged between 7.39 to 18.92 % having lowest value for SVR modal followed by LASSO-SVR, SMLR, SMLR-SVR and LASSO model.

The various weather parameter identified by SMLR for Patiala are time, Z271 ((Minimum temperature*evaporation) and Z461 (morning relative humidity*bright sunshine hours), while the parameters identified by LAASO are time, Z21(minimum temperature*maximum temperature), Z251 (Minimum temperature*evening relative humidity), Z261 ((Minimum temperature *bright sunshine hours), Z361 (rainfall*bright sunshine hours), Z460 (morning relative humidity*bright sunshine hours), Z461 (morning relative humidity*bright sunshine hours), Z671 (bright sunshine hours*evaporation) and Z681 (bright sunshine hours*evapotranspiration). Equation developed for wheat yield prediction at flowering stage for Hisar by SMLR, LASSO is given in Table 4.25.

The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01369863, epsilon: 0.1 and number of Support Vectors: 32; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.0111111, epsilon: 0.1 and number of Support Vectors was 26, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.333333, epsilon: 0.1 and number of Support Vectors was 28.

The value of RMSE during validation was the lowest for LASSO (363.8 kg/ha) followed by SVR (589.7 kg/ha), LASSO-SVR (758.8 kg/ha), SMLR-SVR (857.7 kg/ha) and SMLR (942.4 kg/ha). The mean square error for validation ranged between 132326 kg/ha for LASSO to 888152 kg/ha for SMLR. The maximum value was found for SMLR followed by SMLR-SVR, LASSO-SVR, SVR and LASSO respectively.
The nRMSE value for validation ranged between 8.36 to 21.66% With the maximum value for maximum value for SMLR (21.66%) followed by SMLR-SVR (19.72%), LASSO-SVR (17.44%), SVR (13.56%), and minimum for LASSO (8.36%). respectively. The results showed that performance of wheat yield prediction for Hisar done at flowering stage was excellent for LASSO models having nRMSE < 10% good for SVR, LASSO-SVR and SMLR-SVR having nRMSE value < 20 %, fair for SMLR with nRMSE value 21.66 < 25 %. The calibration and validation for all models are shown in (fig 19a & 19b).
The percentage deviation of yield estimation done at flowering stage for Hisar by observed yield using the model developed was the lowest for SMLR (6.38%), model followed by SVR (9.94) LASSO-SVR (12.87 %) LASSO (14.44 %),, and SMLR-SVR (17.54%) modal respectively. The negative sign indicates the under estimation of wheat yield at flowering stage for Hisar.

Table 4.24 Wheat yield prediction at flowering stage by different modal for Hisar

S.No.	NAME OF MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.93	189122.00	434.88	17.52	888152.49	942.42	21.66	6.38
2	SVR	0.99	33677.24	183.51	7.39	347748.40	589.70	13.56	9.94
3	LASSO	0.94	220446.41	469.52	18.92	132326.04	363.77	8.36	14.44
4	LASSO-SVR	0.94	162953.12	403.67	16.26	575784.91	758.80	17.44	12.87
5	SMLR-SVR	0.93	194562.24	441.09	17.77	735610.24	857.68	19.72	17.54

Table 4.25: Equation for wheat yield prediction at flowering stage for Hisar

Model	Equation
SMLR	Y=73.60+ 132.86*time+ 51.68*Z271 + 2.67*Z461
LASSO	Y=95.47*time+ 17.07*Z21+ 0.15*Z251+11.19*Z261+ 3.18*Z361- 0.02*Z460+ 0.85*Z461+3.22*Z671+ 5.73*Z681

4.5.3 Wheat Yield prediction at flowering stage by different model for Amritsar, Punjab

Model for predicting wheat yield at flowering stage by different techniques for Amritsar have been developed using long term crop yield data as well as long period daily weather data from 46th to 8h standard meteorological week for Amritsar districts. The performance of the model for wheat crop yield prediction at flowering stage for Amritsar is shown in the (Table 4.26).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During the calibration, value of R^2 ranged between 0.91 to 0.98% having the highest value 0.98 % for SVR , 0.92 % for LASSO, LASSO-SVR, SMLR, and 0.91 % for SMLR-SVR model respectively. The mean square error for calibration was ranged between 16038 kg/ha for SVR model to 72106 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR, SMLR-SVR, LASSO-SVR, and SVR modal. The RMSE for calibration was the lowest for SVR modal (126.6 kg/ha) followed by LASSO-SVR (203.3 kg /ha), SMLR-SVR (210.8 kg/ha), SMLR (237.4 kg/ha), and LASSO (268.5 kg/ha). During the calibration nRMSE ranged between 3.77 to 7.98% having the lowest value for SVR modal followed by LASSO-SVR, SMLR-SVR, SMLR and LASSO model.

The most important weather parameter identified by SMLR for are time and Z181 (maximum temperature*evapotranspiration), while parameters for LASSO are time, Z11 (maximum temperature), Z151 (maximum temperature*evening relative humidity), Z181 (maximum temperature*evapotranspiration) and Z481 (morning relative humidity *evapotranspiration). Equation developed for wheat yield prediction at flowering stage for Amritsar by SMLR, LASSO is given in Table 4.27. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 25. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.2, epsilon: 0.1 and number of Support Vectors: 21 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.5, epsilon: 0.1 and number of Support Vectors was22.

The mean square error for validation ranged between 46892 kg/ha for LASSO to 370644 kg/ha for SMLR. The maximum value was found for SMLR followed by SMLR-SVR, SVR, LASSO-SVR and LASSO respectively. The value of RMSE during validation was the lowest for LASSO (216.5 kg/ha) followed by LASSO-SVR (406.0 kg/ha), SVR (410.7 kg/ha), SMLR-SVR (472.4 kg/ha) and SMLR (608.8 kg/ha). The nRMSE value for validation ranged between 4.90 to 13.78% with the maximum value for SMLR (13.78%) followed by SMLR-SVR (11.09%), SVR (9.29%), LASSO-SVR (9.19%) and the minimum for LASSO (4.90%). Results showed that performance of wheat yield prediction done at flowering stage for Amritsar was excellent for LASSO, LASSO-

SVR and SVR model having nRMSE < 10% and performed good having nRMSE <15% for SMLR-SVR and SMLR. The calibration and validation for all models are shown in (fig 20a & 20b).

The percentage deviation of yield estimation done at flowering stage for Amritsar by observed yield using the model developed was the lowest for LASSO (2.93 %) model followed by LASSO-SVR (-5.36%), SMLR-SVR (-9.02%), SMLR (9.22%)and SVR (-10.18%) modal respectively. The negative sign indicates the under estimation of wheat yield at flowering stage for Amritsar.

Table 4.26 Wheat yield prediction at flowering stage by different modal for Amritsar

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.92	56365	237.4	7.06	370644	608.8	13.78	9.22
2	SVR	0.98	16038	126.6	3.77	168673	410.7	9.29	-10.18
3	LASSO	0.92	72106	268.5	7.98	46892	216.5	4.90	2.93
4	LASSO-SVR	0.92	41333	203.3	6.04	164814	406	9.19	-5.36
5	SMLR-SVR	0.91	44417	210.8	6.27	223156	472.4	11.09	-9.02

Table 4.27: Equation for wheat yield prediction at flowering stage for Amritsar

Model	Equation
SMLR	$Y=1605.06+74.96*time+1.63*Z181$
LASSO	$Y=52.87*time+33.93*Z11+0.74*Z151+0.33*Z181+0.51*Z481$

4.5.4 Wheat Yield prediction at flowering stage by different model for Ludhiana

Model for predicting wheat yield at flowering stage by different techniques for Ludhiana have been developed using long term crop yield data as well as long period daily weather data from 46th to 8h standard meteorological week for Ludhiana. The performance of the model for wheat crop yield prediction at flowering stage for Ludhiana is shown in the (Table 4.28).

During the calibration value of R^2 ranged between 0.93 to 0.99% having the highest value 0.99 % for SVR, 0.95 % for SMLR, LASSO-SVR, SMLR-SVR and 0.93 % for LASSO model respectively. The mean square error for calibration was ranged between 5707 kg/ha for SVR model to 36693 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, SMLR, LASSO-SVR and SVR modal. The RMSE for calibration was lowest for SVR modal (75.5 kg/ha) followed by LASSO-SVR (132.4 kg /ha), SMLR (149.9 kg/ha), SMLR-SVR (151.2 kg/ha) and LASSO (191.6 kg/ha). During the calibration nRMSE ranged between 1.95 to 4.93 % having the lowest value for SVR modal followed by LASSO-SVR, SMLR SMLR-SVR and LASSO model.

The most important weather parameter identified by SMLR are time, Z141 (maximum temperature*morning relative humidity) and Z151 (maximum temperature*minimum relative humidity)while parameters for LASSO are time, Z21 (minimum temperature*maximum temperature), Z41 (morning relative humidity), Z121 (maximum temperature*minimum temperature), Z141 (maximum temperature *morning relative humidity), Z161 (maximum temperature*bright sunshine hours), Z460 (morning relative humidity*bright sunshine hours), Z361 (rainfall*bright sunshine hours), Z461 (morning relative humidity*bright sunshine hours) and Z561 (evening relative humidity*bright sunshine hours). Equation developed for wheat yield prediction at flowering stage for Ludhiana by SMLR, LASSO is given in Table 4.29. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors was 26. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.1, epsilon: 0.1 and number of Support Vectors was 23 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and number of Support Vectors was 23.

The mean square error for validation ranged between 38178 kg/ha for LASSO to 288293 kg/ha for SMLR-SVR. The maximum value was found for SMLR-SVR followed by SMLR, SVR, LASSO-SVR and LASSO respectively. The value of RMSE during validation was the lowest for LASSO (195.4 kg/ha) followed by LASSO-SVR (423.5

kg/ha), SVR (494.5 kg/ha), SMLR (498.9 kg/ha) and SMLR-SVR (536.9 kg/ha). The nRMSE value for validation ranged between 4.10 to 11.26% with the maximum value for SMLR-SVR (11.26%) followed by SMLR (10.47%), SVR (10.37%), LASSO-SVR (8.88%) and minimum for LASSO (4.10 %) model respectively. Results showed that performance of wheat yield prediction done at flowering stage for Ludhiana was excellent for LASSO and LASSO-SVR models having nRMSE < 10% and good for SVR, SMLR and SMLR-SVR model with nRMSE value < 15 % . The calibration and validation for all models are shown in (fig 21a & 21b).

The percentage deviation of yield estimation done at flowering stage for Ludhiana by observed yield using the model developed was the lowest for SVR (0.07 %) model followed by LASSO (4.43 %), LASSO-SVR (6.99%), SMLR (7.31%), and SMLR-SVR (9.86%) modal respectively.

Table 4.28 Wheat yield prediction at flowering stage by different modal for Ludhiana

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.95	22459	149.9	3.86	248867	498.9	10.47	7.31
2	SVR	0.99	5707	75.5	1.95	244568	494.5	10.37	0.07
3	LASSO	0.93	36693	191.6	4.93	38178	195.4	4.10	4.43
4	LASSO-SVR	0.95	17540	132.4	3.41	179340	423.5	8.88	6.99
5	SMLR-SVR	0.95	22873	151.4	3.89	288293	536.9	11.26	9.86

Table 4.29: Equation for wheat yield prediction at flowering stage for Ludhiana

Model	Equation
SMLR	Y=1921.22+63.79*time+0.476*Z141+0.79*Z151
LASSO	Y=47.66*time+12.95*Z21+17.66* Z41+0.51*Z121+0.01*Z141+ 0.86*Z161+0.21*Z241+0.61*Z361+0.19*Z461+0.07*Z561

4.5.5 Wheat Yield prediction at flowering stage by different model for Patiala

Model for predicting wheat yield at flowering stage by different techniques for Patiala have been developed using long term crop yield data as well as long period daily weather data from 46th to 8h standard meteorological week for Patiala. The performance of the model for wheat crop yield prediction at flowering stage for Patiala is shown in the (Table 4.30).

During the calibration value of R^2 ranged between 0.96 to 0.98% having the highest value 0.98 % for SVR, 0.97 % for SMLR, LASSO-SVR and 0.96 % for LASSO and SMLR-SVR model respectively. The mean square error for calibration was ranged between 23176 kg/ha for SVR model to 53568 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, LASSO-SVR, SMLR and SVR modal. The RMSE for calibration was the lowest for SVR modal (152.2 kg/ha) followed by SMLR (162.5 kg /ha), LASSO-SVR (175.4 kg/ha), SMLR-SVR (181.7 kg/ha) and LASSO (231.5 kg/ha). During calibration nRMSE ranged between 4.41 to 6.70 % having lowest value for SVR modal followed by SMLR, LASSO-SVR, SMLR-SVR and LASSO model.

The various weather parameter identified by SMLR for Patiala are time, Z11 (maximum temperature), Z30(rainfall), Z151 (maximum temperature*evening relative humidity) and Z381 (rainfall*evapotranspiration), while the parameters identified by LAASO are time, Z11 (maximum temperature), Z21(minimum temperature*maximum temperature), Z41 (morning relative humidity*maximum temperature), Z181 (maximum temperature*evapotranspiration), Z241 (minimum temperature*morning relative humidity), Z351(rainfall* evening relative humidity), Z480(morning relative humidity*evapotranspiration) andZ581 (evening relative humidity*evapotranspiration). Equation developed for wheat yield prediction at flowering stage for Patiala by SMLR, LASSO is given in Table 4.31.

The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 24; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors was 24, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.2, epsilon: 0.1 and number of Support Vectors was 19.

The value of RMSE during validation was the lowest for LASSO (227.8 kg/ha) followed by SVR (330.0 kg/ha), LASSO-SVR (348.1 kg/ha), SMLR-SVR (409.3

kg/ha) and SMLR (575.2 kg/ha). The mean square error for validation ranged between 51890 kg/ha for LASSO to 330896 kg/ha for SMLR. The maximum value was found for SMLR followed by SMLR-SVR, LASSO-SVR, SVR and LASSO respectively. The nRMSE value for validation ranged between 4.94 to 12.42% with the maximum value for SMLR (12.42%) followed by SMLR-SVR (8.83%), LASSO-SVR (7.51%) SVR (7.12 %) and minimum for LASSO (4.94 %) model respectively. Results showed that the performance of wheat yield prediction for Patiala done at flowering stage was exceptional for all models having nRMSE < 10% except performed good for SMLR having nRMSE value (12.52) < 15 %. The calibration and validation for all models are shown in (fig 22a & 22b).

The percentage deviation of yield estimation done at flowering stage for Patiala by observed yield using the model developed was the lowest for LASSO (0.28 %) model followed by LASSO -SVR (5.87 %), SMLR (7.29%), SVR (7.32%), and SMLR-SVR (7.67%) modal respectively.

Table 4.30 Wheat yield prediction at flowering stage by different modal for Patiala

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Pecentage Deviation %
		R² (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.97	26392	162.5	4.70	330896	575.2	12.42	7.29
2	SVR	0.98	23176	152.2	4.41	108912	330.0	7.12	7.32
3	LASSO	0.96	53568	231.5	6.70	51890	227.8	4.94	0.28
4	LASSO-SVR	0.97	30760	175.4	5.08	121152	348.1	7.51	5.87
5	SMLR-SVR	0.96	33001	181.7	5.26	167502	409.3	8.83	7.67

Table 4.31: Equation for wheat yield prediction at flowering stage for Patiala

Model	Equation
SMLR	Y=792.32+78.14*time+ 52.04*Z11+ 4.45*Z30+ 0.13*Z151+ 0.68*Z381
LASSO	Y=50.73*time+ 19.03*Z11+ 7.75*Z21+ 6.67*Z41+0.27*Z181+ 0.11*Z241+0.003*Z351+ 0.005*Z480+0.32*Z581

4.6 Wheat Yield Prediction at grain filling stage for different district of North West India

Model for wheat yield prediction at grain filling stage for different district of North-West India have been developed using long term crop yield data as well as long period daily weather data from 45th to 11th standard meteorological week for five districts of North West India.

4.6.1 Wheat Yield prediction at grain filling stage by different model for IARI, New Delhi

Model for predicting wheat yield at grain filling stage by different techniques for IARI, New Delhi have been developed using long term crop yield data as well as long period daily weather data from 46th to 11th standard meteorological week for IARI, New Delhi. The model developed is shown in the (Table 4.32).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. The models developed for predicting the yield had the value of R^2 ranged between 0.80 to 0.99%. During calibration value of RMSE was the lowest for the SMLR-SVR (40.0 kg/ha) followed by LASSO (60.6 kg/ha), LASSO-SVR (67.9 kg /ha), SMLR (159.8 kg/ha) and SVR (205.5 kg/ha). The mean square error for calibration was ranged between 3674 kg/ha for LASSO model to 42230 kg/ha for SVR model. The maximum value was found for SVR modal followed by SMLR, LASSO-SVR, LASSO and SMLR-SVR modal. During calibration nRMSE ranged between 1.15 to 5.92 % having the lowest value for SMLR-SVR modal followed by LASSO, LASSO-SVR, SMLR and SVR model.

The most important weather parameter identified by SMLR are time, Z281 (Minimum temperature* evapotranspiration) and Z671 (sunshine hour*evaporation), while parameters for LASSO are time, Z10 (maximum temperature), Z21 (minimum temperature*maximum temperature), Z40 (morning relative humidity), Z51 (evening relative humidity*maximum temperature), Z71(evaporation* maximum temperature), Z120 (maximum temperature*minimum temperature), Z151 (maximum temperature*evening relative humidity), Z141 (maximum temperature*morning relative humidity), Z150 (maximum temperature*minimum relative humidity), Z170 (maximum temperature* evaporation), Z171(maximum temperature* evaporation), Z460(morning relative humidity*bright sunshine hours), Z461 (morning relative humidity*bright sunshine hours), Z560 (evening relative humidity*sunshine hours), Z180 (maximum temperature*evapotranspiration), Z181 (maximum temperature*evapotranspiration) and Z581 (evening relative humidity *evapotranspiration).

Equation developed for wheat yield prediction at grain filling stage for IARI, New Delhi by SMLR, LASSO is given in Table 4.33. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 25. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.09090909, epsilon: 0.1 and number of Support Vectors: 10 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.03846154, epsilon: 0.1 and number of Support Vectors was 17.

The value of RMSE during validation with independent dataset, was the lowest for LASSO (51.8 kg/ha) followed by LASSO-SVR (83.1 kg/ha), SVR (190.1 kg/ha), SMLR (269.2 kg/ha) and SMLR-SVR (281.2 kg/ha). The mean square error for validation was ranged between 2685 kg/ha for LASSO to 79052 kg/ha for SMLR-SVR. The maximum value was found for SMLR-SVR followed by SMLR, SVR, LASSO-SVR and LASSO respectively. The nRMSE value for validation ranged between 1.22 to 6.63% with the maximum value for SMLR-SVR (6.63%) followed by SMLR (6.35%), SVR (4.41%) , LASSO-SVR (1.88%) and minimum for LASSO (1.22%). This shows that all the models performed excellent for wheat yield estimation at grain filling stage for IARI, New Delhi having nRMSE < 10%. The calibration and validation for all models are shown in (fig 23a & 23b).

The percentage deviation of yield prediction done at grain filling stage for IARI, New Delhi by observed yield using the model developed was lowest for LASSO (-0.02 %) model followed by SVR (-0.67%), LASSO -SVR (-1.03 %), SMLR-SVR (-

2.24%), and SMLR (-7.60) modal respectively. The negative sign here indicates under estimation of wheat yield for IARI. New Delhi at grain filling stage.

Table 4.32 Wheat yield prediction at grain filling stage by different modal for IARI, New Delhi

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation (%)
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.93	25539	159.8	4.57	72478	269.2	6.35	-7.60
2	SVR	0.80	42230	205.5	5.92	36133	190.1	4.41	-0.67
3	LASSO	0.98	3674	60.6	1.73	2685	51.8	1.22	-0.02
4	LASSO-SVR	0.98	4612	67.9	1.94	6913	83.1	1.88	-1.03
5	SMLR-SVR	0.99	1596	40.0	1.15	79052	281.2	6.63	-2.24

Table 4.33: Equation for wheat yield prediction at grain filling stage for IARI, New Delhi

Model	Equation
SMLR	Y=2626.54+ 53.57*time+ 2.26*Z281+ 10.47*Z671
LASSO	Y=93.34*Time+ 62.57*Z10+ 36.68*Z21+ 0.04*Z40+ 13.64*Z51+ 0.05*Z71+ 3.53*Z151+ 0.22*Z141+ 0.40*Z150+ 0.06*Z170+ 0.76*Z171+ 2.47*Z581

4.6.2 Wheat Yield prediction at grain filling stage by different model for Hisar, Haryana

Model for predicting wheat yield at grain filling stage by different techniques for Hisar have been developed using long term crop yield data as well as long period daily weather data from 46th to 11th standard meteorological week for Hisar districts. The performance of the model for wheat crop yield prediction at grain stage for Hisar is shown in the (Table 4.34).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During calibration value of R^2 ranged between 0.93 to 0.99% having the highest value 0.99 % for SVR and SMLR-SVR model, 0.95 % for LASSO, 0.94 % for LASSO-SVR and 0.93 % for SMLR respectively. The mean square error for calibration ranged between 33677 kg/ha for SMLR-SVR model to 172707 kg/ha for SMLR model. The maximum value was found for SMLR modal followed by LASSO, LASSO-SVR, SVR and SMLR-SVR modal. The RMSE for calibration was the lowest for SMLR-SVR modal (183.5 kg/ha) followed by SVR (187.7 kg/ha), LASSO-SVR (383.5 kg /ha), LASSO (394.2 kg/ha) and SMLR (415.6 kg/ha). During calibration value of nRMSE ranged between 7.39 to 16.74 %, having the lowest value for SMLR-SVR modal followed by SVR, LASSO-SVR, LASSO and SMLR model.

The important weather parameter identified by SMLR for Hisar are time, Z271 ((Minimum temperature*evaporation) and Z461 (morning relative humidity* bright sunshine hours), weather elements identified by LASSO are time Z21 (minimum temperature*maximum temperature), Z61 (bright sunshine hours* maximum temperature), Z61 (bright sunshine hours* maximum temperature), Z460 (morning relative humidity* bright sunshine hours) Z461 (morning relative humidity* bright sunshine hours), Z571 (evening relative humidity*evaporation), Z581 (evening relative humidity*evapotranspiration) and Z671 (bright sunshine hours*evaporation). Equation developed for wheat yield prediction at grain filling stage for Hisar by SMLR, LASSO is given in Table 4.35. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.01369863, epsilon: 0.1 and number of Support Vectors: 32. The Parameters for LASSO-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.08333333, epsilon: 0.1 and number of Support Vectors was 29 while the parameters for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.03333333, epsilon: 0.1 and number of Support Vectors was 30.

The mean square error for validation ranged between 66106 kg/ha for LASSO to 846468 kg/ha for SMLR. The maximum value was found for SMLR followed by SVR, LASSO-SVR, SMLR-SVR and LASSO respectively. The value of RMSE during validation was the lowest for LASSO (257.1 kg/ha) followed by SMLR-SVR (589.7 kg/ha), LASSO-SVR (633.3 kg/ha), SVR (870.2 kg/ha), and SMLR (920.0 kg/ha). The nRMSE value for validation ranged between 5.91 to 21.15% with the maximum value for SMLR (21.15%) followed by SVR (20.00%), SMLR-SVR (13.56%), LASSO-SVR (14.56%) and minimum for LASSO (5.91%). Results showed that performance of wheat yield prediction done at grain filling stage for Hisar was excellent for LASSO modal having nRMSE < 10%, good for SMLR-SVR, LASSO-SVR and SVR model having nRMSE <20% and fair for SMLR model having nRMSE value 21.15 %. The calibration and validation for all models are shown in (fig 24a & 24b).

The percentage deviation of yield estimation done at grain filling stage for Hisar by observed yield using the model developed was the lowest for SVR (-4.48%), model followed by SMLR (5.36 %), LASSO (6.09%), SMLR-SVR (9.94) and LASSO -SVR (12.81%) modal respectively. Negative sign here indicates under estimation of wheat yield for Hisar at grain filling stage by SVR modal.

Table 4.34 Wheat yield prediction at grain filling stage by different modal for Hisar

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R² (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.93	172707	415.6	16.74	846468	920.0	21.15	5.36
2	SVR	0.99	35242	187.7	7.56	757158	870.2	20.00	-4.48
3	LASSO	0.95	155359	394.2	15.88	66106	257.1	5.91	**6.09**
4	LASSO-SVR	0.94	147063	383.5	15.45	401087	633.3	14.56	12.81
5	SMLR-SVR	0.99	33677	183.5	7.39	347748	589.7	13.56	9.94

Table 4.35: Equation for wheat yield prediction at grain filling stage for Hisar

Model	Equation
SMLR	Y=-924.86+ 139.21*Time+ 43.72*Z271+ 2.21*Z461
LASSO	Y=93.32*time+ 62.57*Z21+ 36.69*Z61+ 0.04*Z150+ 13.64*Z261+ 0.05*Z351+ 3.53*Z361+ 0.22*Z460+ 0.40*Z461+ 0.06*Z571+ 0.76*Z581+ 2.48*Z671

4.6.3 Wheat Yield prediction at grain filling stage by different model for Amritsar, Punjab

Model for predicting wheat yield at grain filling stage by different techniques for Amritsar have been developed using long term crop yield data as well as long period daily weather data from 46th to 11th standard meteorological week for Amritsar districts. The model developed is shown in the (Table 4.36).

The coefficient of determination (R^2) was significant at 1% probability level for all developed models. Value of coefficient of determination R^2 for models developed by different techniques for predicting the wheat crop yield at grain filling stage was 0.94% for model developed by SMLR-SVR techniques, 0.95 % for modal developed by SMLR and LASSO techniques, 0.96 % for modal developed by SMLR-SVR techniques and 0.96% for model developed by LASSO-SVR techniques. The RMSE for calibration was lowest for SVR modal (82.9 kg/ha) followed by SMLR-SVR(121.9 kg/ha), LASSO-SVR (123.7 kg /ha), SMLR (186.4 kg/ha) and LASSO (222.9 kg/ha). The mean square error for calibration was ranged between 6870 kg/ha for SVR model to 49690 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR, LASSO-SVR, SMLR-SVR and SVR modal. During calibration nRMSE ranged between and 2.46 to 6.63% having the lowest value for SVR modal followed by SMLR-SVR, LASSO-SVR, SMLR and LASSO model .

The most important weather parameter identified by SMLR for Amritsar are time, Z121 (maximum temperature*minimum temperature) and Z131 (maximum temperature*rainfall) while parameters for LASSO are time, Z11 (maximum temperature), Z121 (maximum temperature*minimum temperature), Z151 (maximum temperature*evening relative humidity), Z230 (Minimum temperature*rainfall) Z281

(Minimum temperature* evapotranspiration), Z481(morning relative humidity *evapotranspiration) and Z581 (evening relative humidity*evapotranspiration). Equation developed by SMLR and LASSO model for wheat crop yield prediction at grain filling stage are given in table 4.37. The various parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors was 25. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.09090909, epsilon: 0.1 and number of Support Vectors: 10 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and number of Support Vectors was 22.

The mean square error for validation was ranged between 33897 kg/ha for LASSO modal to 363639 kg/ha for SMLR modal. The maximum value was found for SMLR model followed by SVR, SMLR-SVR, LASSO-SVR and LASSO modal respectively. RMSE value during validation with independent dataset, was maximum for SMLR **(**603.0 kg/ha) followed by SVR (526.4 kg/ha), SMLR-SVR (429.9 kg/ha), LASSO-SVR (422.0 kg/ha) and LASSO (184.1 kg/ha). Value of nRMSE during validation ranged between 4.17% to 13.64% with the maximum value for maximum value for SMLR (13.64%) followed by SVR (11.91%), SMLR-SVR (9.73%), , LASSO-SVR (9.55%) and minimum for LASSO (4.17%). The performance of wheat yield prediction at grain filling stage for Amritsar was excellent with nRMSE value <10% for LASSO, LASSO-SVR, SMLR-SVR model and good for SVR and SMLR model having nRMSE value <15 %. The calibration and validation for all models are shown in (fig 25a & 25b).

The percentage deviation of yield estimation done at grain filling stage by observed yield for Amritsar using developed model was lowest for LASSO (-0.29 %) model followed by SMLR-SVR (-2.82%), LASSO -SVR (-5.90%), SVR (-12.11%),) and SMLR (12.65) modal respectively. The negative sign here indicates under estimation of wheat yield for Amritsar at grain filling stage.

Table 4.36 Wheat yield prediction at grain filling stage by different modal for Amritsar

S.No.	Model	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation (%)
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.95	34760	186.4	5.54	363639	603.02	13.64	12.65
2	SVR	0.99	6870	82.9	2.46	277126	526.43	11.91	-12.11
3	LASSO	0.95	49690	222.9	6.63	33897	184.11	4.17	-0.29
4	LASSO-SVR	0.96	15297	123.7	3.68	178062	421.97	9.55	-5.90
5	SMLR-SVR	0.94	14856	121.9	3.62	184775	429.85	9.73	-2.82

Table 4.37: Equation for wheat yield prediction at grain filling stage for Amritsar

Model	Equation
SMLR	Y=1746.54+73.25*time+2.51*Z121+0.729*Z131
LASSO	Y=52.13*time+10.66*Z11+1.29*Z121+0.67*Z151+0.86*Z230+0.92*Z281+ 0.49*Z481+0.15*Z581

4.6.4 Wheat Yield prediction at grain filling stage by different model for Ludhiana, Punjab

Model for predicting wheat yield at grain filling stage by different techniques for Ludhiana have been developed using long term crop yield data as well as long period daily weather data from 46th to 11th standard meteorological week for Ludhiana districts. The performance of the model for wheat crop yield prediction at grain stage for Ludhiana is shown in the (Table 4.38).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During calibration value of R^2 ranged between 0.93 to 0.98% having the highest value 0.98 % for SVR followed by 0.96 % for SMLR, 0.95 % for SMLR-SVR, 0.94 % for LASSO and 0.93 for LASSO-SVR model respectively. The mean square error for the calibration was ranged between 7083 kg/ha for SVR model to 32707 kg/ha for LASSO model The maximum value was found for LASSO modal followed by LASSO-SVR, SMLR, SMLR-SVR and SVR modal. The RMSE for calibration was lowest for SVR modal (84.2 kg/ha) followed by SMLR-SVR (114.0 kg/ha), SMLR (141.3 kg/ha), LASSO-SVR (144.5 kg /ha), and LASSO (180.9 kg/ha). During calibration nRMSE ranged between 2.17 to 4.66%, having the lowest value for SVR modal followed by SMLR-SVR, SMLR, LASSO-SVR, and LASSO model.

The most important weather parameter identified by SMLR are time, Z151 (maximum temperature*evening relative humidity), Z141 (maximum temperature*morning relative humidity and Z361 (rainfall*bright sunshine hours) while parameters for LASSO are time, Z41 (morning relative humidity*maximum temperature), Z121 (maximum temperature*minimum temperature), Z141 (maximum temperature*morning relative humidity), Z151 (maximum temperature*evening relative humidity), Z261 (Minimum temperature *bright sunshine hours), Z161 (maximum temperature*bright sunshine hours), Z361 (rainfall*bright sunshine hours) and Z561(evening relative humidity*sunshine hours). Equation developed for wheat yield prediction at grain filling stage for Ludhiana by SMLR, LASSO is given in Table 4.39. The parameter for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 30. For LASSO-SVR the parameters recognized SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.1666667, epsilon: 0.1 and number of Support Vectors: 22 and for SMLR-SVR the parameters recognized are as follows SVM-Type: eps-regression, Kernel: linear, cost: 1, gamma: 0.25, epsilon: 0.1 and number of Support Vectors was 23.

The mean square error for validation was ranged between 39022 kg/ha for LASSO to 302381 kg/ha for SMLR. The maximum value was found for SMLR followed by SMLR-SVR, SVR, LASSO-SVR and LASSO respectively.

The value of RMSE during validation with independent dataset, was lowest for LASSO (197.5 kg/ha) followed by LASSO-SVR (453.1 kg/ha), SVR (500.8 kg/ha), SMLR-SVR (540.8 kg/ha) and SMLR (549.9 kg/ha). The nRMSE value for validation ranged between 4.14 to 11.54% with the maximum value for SMLR (11.54%) followed by SMLR-SVR (11.35%), SVR (10.51%), LASSO-SVR (9.51%) and minimum for LASSO (4.14%). The performance of wheat yield prediction done at grain filling stage for Ludhiana was excellent having nRMSE < 10% for LASSO and LASSO-SVR and good having nRMSE < 15% for SVR, SMLR-SVR and SMLR model.The calibration and validation for all models are shown in (fig 26a & 26b).

The percentage deviation of yield estimation done at grain filling stage for Ludhiana by observed yield using the model developed was the lowest for SVR (-1.70 %) model followed by LASSO (4.48 %), SMLR (6.23), LASSO -SVR (8.23%) and SMLR-SVR (10.12%) modal respectively. The negative sign here indicates under estimation of wheat yield by SVR model at grain filling stage for Ludhiana

Table 4.38 Wheat yield prediction at grain filling stage by different modal for Ludhiana

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation (%)
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.96	19966	141.3	3.64	302381	549.9	11.54	6.23
2	SVR	0.98	7083	84.2	2.17	250835	500.8	10.51	-1.70
3	LASSO	0.94	3270	180.9	4.66	39022	197.5	4.14	4.48
4	LASSO-SVR	0.93	20865	144.5	3.72	205274	453.1	9.51	8.23
5	SMLR-SVR	0.95	13001	114.0	2.94	292471	540.8	11.35	10.12

Table 4.39: Equation for wheat yield prediction at grain filling stage for Ludhiana

Model	Equation
SMLR	Y=2138.25+64.70*time+1.716*Z361+0.33*Z141+0.76*Z151
LASSO	Y=49.95*time+22.63*Z41+0.60*Z121+0.55*Z141+1.71*Z151+ 0.70*Z261+0.01*Z161+0.35*Z361+0.023*Z561

4.6.5 Wheat Yield prediction at grain filling stage by different model for Patiala, Punjab

Model for predicting wheat yield at grain filling stage by different techniques for Patiala have been developed using long term crop yield data as well as long period daily weather data from 46th to 11th standard meteorological week for Patiala districts. The performance of the model for wheat crop yield prediction at grain stage for Patiala is shown in the (Table 4.40).

The coefficient of determination (R^2) was significant at 1% probability level for all the models. During calibration value of R^2 ranged between 0.95 to 0.98% having highest value 0.98 % for SVR, 0.96 % for LASSO, LASSO-SVR, SMLR, and 0.95 % for SMLR-SVR model respectively. The mean square error for calibration was ranged between 22906 kg/ha for SVR model to 45228 kg/ha for LASSO model. The maximum value was found for LASSO modal followed by SMLR-SVR, SMLR, LASSO-SVR, and SVR modal. The RMSE for calibration was the lowest for SVR modal (151.4 kg/ha) followed by LASSO-SVR (189.5 kg /ha), SMLR (199.0 kg/ha), SMLR-SVR (200.7 kg/ha), and LASSO (212.7 kg/ha). During calibration nRMSE ranged between 4.38 to 6.16%. Having the lowest value for SVR modal followed by LASSO-SVR, SMLR, SMLR-SVR and LASSO model.

The various weather parameter identified by SMLR for Patiala are time, Z31 (morning relative humidity*maximum temperature) and Z141 (maximum temperature*morning relative humidity), while the parameters identified by LAASO are time, Z11 (maximum temperature), Z21(minimum temperature*maximum temperature) Z31 (morning relative humidity*maximum temperature), Z41 (morning relative humidity*maximum temperature), Z120 (maximum temperature*minimum temperature), Z141 (maximum temperature*morning relative humidity), Z181 (maximum temperature*evapotranspiration), Z280 (Minimum temperature* evapotranspiration), Z451(morning relative humidity* evening relative humidity) and Z581 (evening relative humidity*evapotranspiration). Equation developed for wheat yield prediction at grain filling stage for Patiala by SMLR, LASSO is given in Table 4.41. The Parameters for SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors: 25; for LASSO-SVR these are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.02325581, epsilon: 0.1 and number of Support Vectors was 25, while for SMLR-SVR are SVM-Type: eps-regression, SVM-Kernel: linear, cost: 1, gamma: 0.3333333, epsilon: 0.1 and number of Support Vectors was 24.

The value of RMSE during validation with independent dataset, was the lowest for LASSO (221.8 kg/ha) followed by SVR (412.3 kg/ha), LASSO-SVR (415.5 kg/ha), SMLR-SVR (420.5 kg/ha) and SMLR (596.6 kg/ha). The mean square error for validation ranged between 49210 kg/ha for LASSO to 355923 kg/ha for SMLR. The maximum value was found for SMLR followed by SMLR-SVR, LASSO-SVR, SVR and LASSO respectively. The nRMSE value for validation ranged between 4.81 to 12.88% with the maximum value for SMLR (12.88%) followed by SMLR-SVR (9.07%), LASSO-SVR (8.97%), SVR (8.90%) and minimum for LASSO (4.81%). Results showed that performance of wheat yield prediction done at grain filling stage for Patiala was excellent for all the model having nRMSE < 10% except SMLR performed good having nRMSE value 12.88%. The calibration and validation for all models are shown in (fig 27a & 27b).

The percentage deviation of yield estimation done at grain filling stage by observed yield for Patiala using developed model was the lowest for SMLR (1.74) model followed by LASSO (-2.03 %), SMLR-SVR (2.06%), SVR (6.71%) and LASSO -SVR (7.75%) modal respectively. The negative sign here indicates under estimation of wheat yield at grain filling stage for Patiala by LASSO modal.

Table 4.40 Wheat yield prediction at grain filling stage by different modal for Patiala

S.No.	MODEL	Modal accuracy parameter during calibration				Modal accuracy parameter during validation			Percentage Deviation %
		R^2 (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	MSE (kg/ha)	RMSE (kg/ha)	nRMSE (%)	
1	SMLR	0.96	39604	199.0	5.76	355923	596.6	12.88	1.74
2	SVR	0.98	22906	151.4	4.38	169982	412.3	8.90	6.71
3	LASSO	0.96	45228	212.7	6.16	49211	221.8	4.81	-2.03
4	LASSO-SVR	0.96	35926	189.5	5.49	172670	415.5	8.97	7.75
5	SMLR-SVR	0.95	40293	200.7	5.81	176781	420.5	9.07	2.06

Table 4.41: Equation for wheat yield prediction at grain filling stage for Patiala

Model	Equation
SMLR	Y=-252.23+ 77.58*time+ 0.21*Z31+ 0.21*Z141
LASSO	Y=49.66*time+ 9.58*Z11+ 37.58*Z21+ 0.12*Z31+ 4.91*Z41- 0.11*Z120+ 0.55*Z141+ 0.01*Z280+ 0.31*Z451

4.7.1 Percentage deviation of predicted yield done at tillering stage by observed yield for different district using different models

Percentage deviation of predicted yield at tillering stage for different district using different models by observed yield are sown in (Table 4.42).

LASSO had percentage deviation < 5% for four station Ludhiana, Amritsar, IARI, New Delhi, Patiala and 12.07% for Hisar. LASSO-SVR had percentage deviation < 10% for Ludhiana, Amritsar, IARI, New Delhi, Patiala, <15% for Hisar (14.97%). SMLR model had percentage deviation < 10 % for Patiala, Ludhiana, IARI, New Delhi and Hisar, more than 10% for Amritsar (11.21 %). SVR had percentage deviation < 10 % for all stations. SMLR-SVR model had percentage deviation < 10 % for IARI, New Delhi, Patiala, Amritsar and Ludhiana, more than 20% for Hisar (20.62).

Table 4.42 Percentage deviation of predicted yield done at tillering stage by observed yield for different district using different models

Name of district	SMLR			SVR			LASSO		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4590.68	-6.62	4916	4551.80	-7.41	4916	4654.18	-5.33
Hisar	4162	4477.48	7.58	4162	4584.44	10.15	4162	4664.35	12.07
Amritsar	4478	4979.98	11.21	4478	4036.92	-9.85	4478	4611.44	2.98
Ludhiana	4670	4914.24	5.23	4670	5117.85	9.59	4670	4698.95	0.62
Patiala	5472	5548.06	1.39	5472	5687.60	3.94	5472	5178.70	-5.36

Name of district	LASSO-SVR			SMLR-SVR		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4684.59	-4.71	4916	4762.621	-3.12
Hisar	4162	4785.05	14.97	4162	5020.20	20.62
Amritsar	4478	4250.07	-5.09	4478	4164.99	-6.99
Ludhiana	4670	5094.50	9.09	4670	5152.41	10.33
Patiala	5472	5910.85	8.02	5472	5598.40	2.31

4.3.22 Percentage deviation of predicted yield done at flowering stage by observed yield for different district using different models

Percentage deviation of predicted yield at flowering stage for different district using different models by observed yield are sown in (Table 4.43).

LASSO had percentage deviation < 5% for four station Patiala, IARI, New Delhi, Amritsar, Ludhiana, and 14.44% for Hisar. SVR had percentage deviation < 5% for Ludhiana and IARI, New Delhi <10 % for Hisar and Patiala, more than 10 % for Amritsar (10.18). LASSO-SVR model had percentage deviation < 10 % for IARI, New Delhi, Amritsar, Ludhiana and 1287% for Hisar. SMLR model had percentage deviation < 10 % for Patiala, Ludhiana, Amritsar, Hisar and IARI, New Delhi. SMLR-SVR had percentage deviation < 10 % for Patiala, IARI, New Delhi, Amritsar and Ludhiana, more than 10 % for Hisar (17.54 %).

Table 4.43. Percentage deviation of predicted yield done at flowering stage by observed yield for different district using different models

Name of district	SMLR			SVR			LASSO		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4544.13	-7.56	4916	4815.31	-2.05	4916	4871.13	-0.91
Hisar	4162	4427.54	6.38	4162	4575.70	9.94	4162	4762.99	14.44
Amritsar	4478	4890.87	9.22	4478	4022.14	-10.18	4478	4609.21	2.93
Ludhiana	4670	5011.38	7.31	4670	4673.27	0.07	4670	4876.88	4.43
Patiala	5472	5870.91	7.29	5472	5872.55	7.32	5472	5487.32	0.28

Name of district	LASSO-SVR			SMLR-SVR		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4874.36	-0.85	4916	4739.02	-3.6
Hisar	4162	4697.65	12.87	4162	4892.01	17.54
Amritsar	4478	4237.98	-5.36	4478	4074.08	-9.02
Ludhiana	4670	4996.43	6.99	4670	5130.93	9.87
Patiala	5472	5793.21	5.87	5472	5891.70	7.67

4.3.21 Percentage deviation of predicted yield done at grain filling stage by observed yield for different district using different models

Percentage deviation of estimated yield at grain filling stage using different models by observed yield for different district are sown in (Table 4.44)

LASSO had percentage deviation < 5% for four station IARI, New Delhi, Amritsar, Patiala, Ludhiana, and 6.09 % for Hisar. SVR had percentage deviation < 5% for IARI, New Delhi and Ludhiana and Hisar < 10 % for Patiala and 12.11 % for Amritsar. LASSO-SVR model had percentage deviation < 10 % for IARI, New Delhi, Amritsar, Ludhiana and Patiala, < 15 % for Hisar. SMLR model had percentage deviation < 10 % for IARI, New Delhi, Hisar, Ludhiana and Patiala and, < 15% for Amritsar. SMLR-SVR had percentage deviation < 5 % for Patiala, Amritsar and IARI, New Delhi, lees than 10 % for Hisar (9.94 %) followed by Ludhiana (10.12 %).

Table 4.44 Percentage deviation of predicted yield done at grain filling stage by observed yield for different district using different models

Name of district	**SMLR**			**SVR**			**LASSO**		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4542.48	-7.6	4916	4883.25	-0.67	4916	4915.25	-0.02
Hisar	4162	4385.08	5.36	4162	3975.54	-4.48	4162	4415.47	6.09
Amritsar	4478	5044.47	12.65	4478	3935.71	-12.11	4478	4465.01	-0.29
Ludhiana	4670	4960.94	6.23	4670	4590.61	-1.7	4670	4879.22	4.48
Patiala	5472	5567.21	1.74	5472	5839.17	6.71	5472	5360.92	-2.03

Name of district	**LASSO-SVR**			**SMLR-SVR**		
	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)	observed yield(kg/ha)	predicted yield(kg/ha)	Percent deviation (%)
IARI,New Delhi	4916	4865.25	-1.03	4916	4805.81	-2.24
Hisar	4162	4695.15	12.81	4162	4575.70	9.94
Amritsar	4478	4213.80	-5.9	4478	4351.72	-2.82
Ludhiana	4670	5054.34	8.23	4670	5142.60	10.12
Patiala	5472	5896.08	7.75	5472	5584.72	2.06

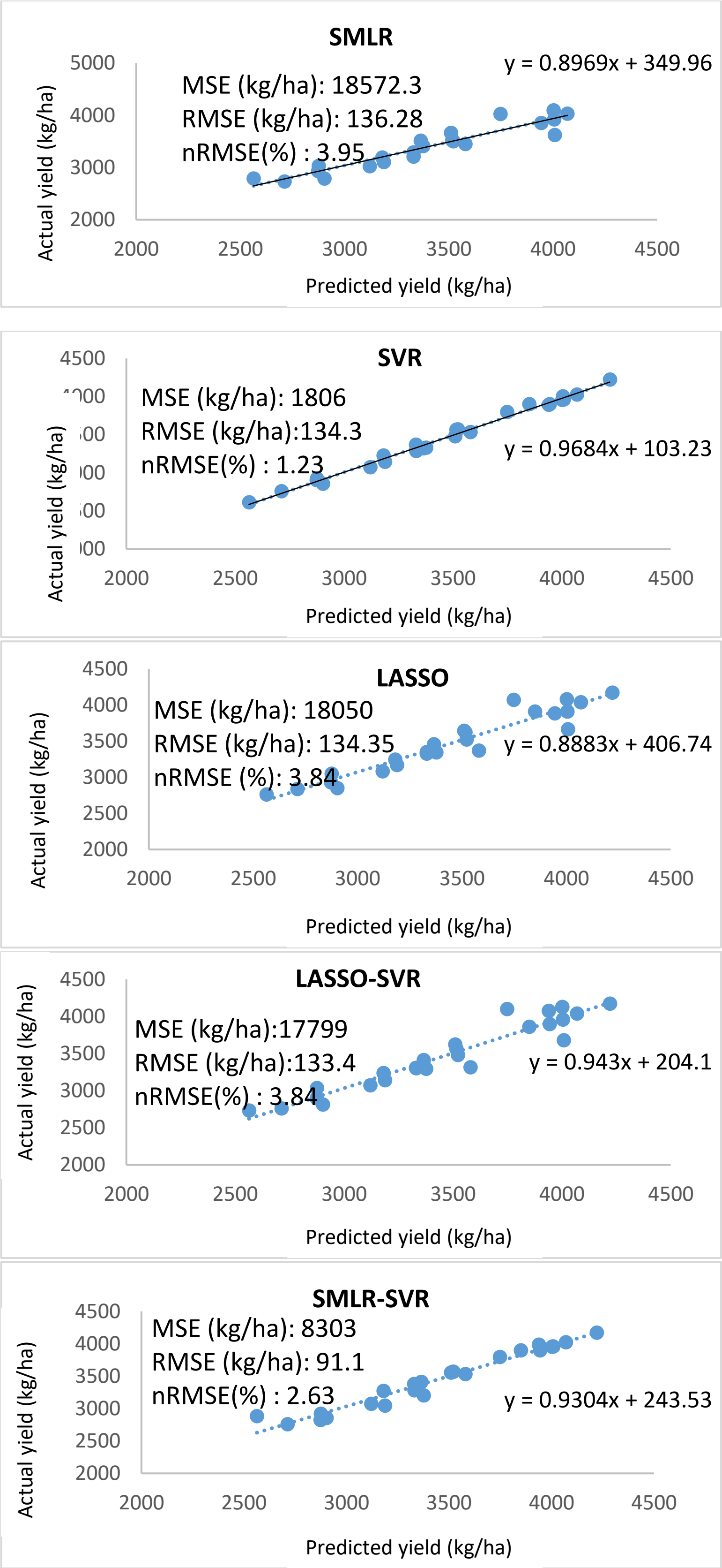


Fig.13.a Calibration of Models developed for wheat yield prediction at tillering stage for IARI.

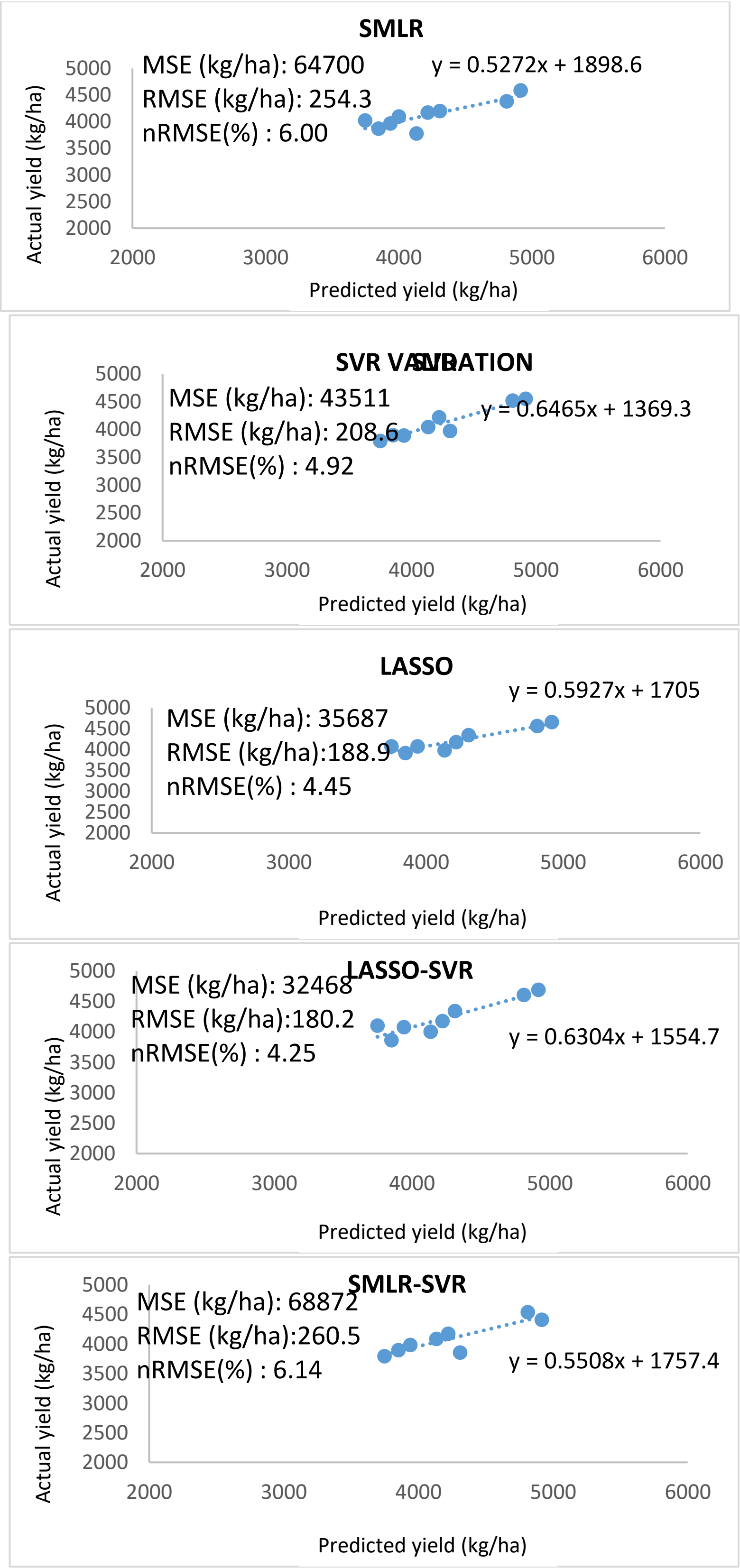


Fig.13.b Validation of Models developed for wheat yield prediction at tillering stage for IARI.

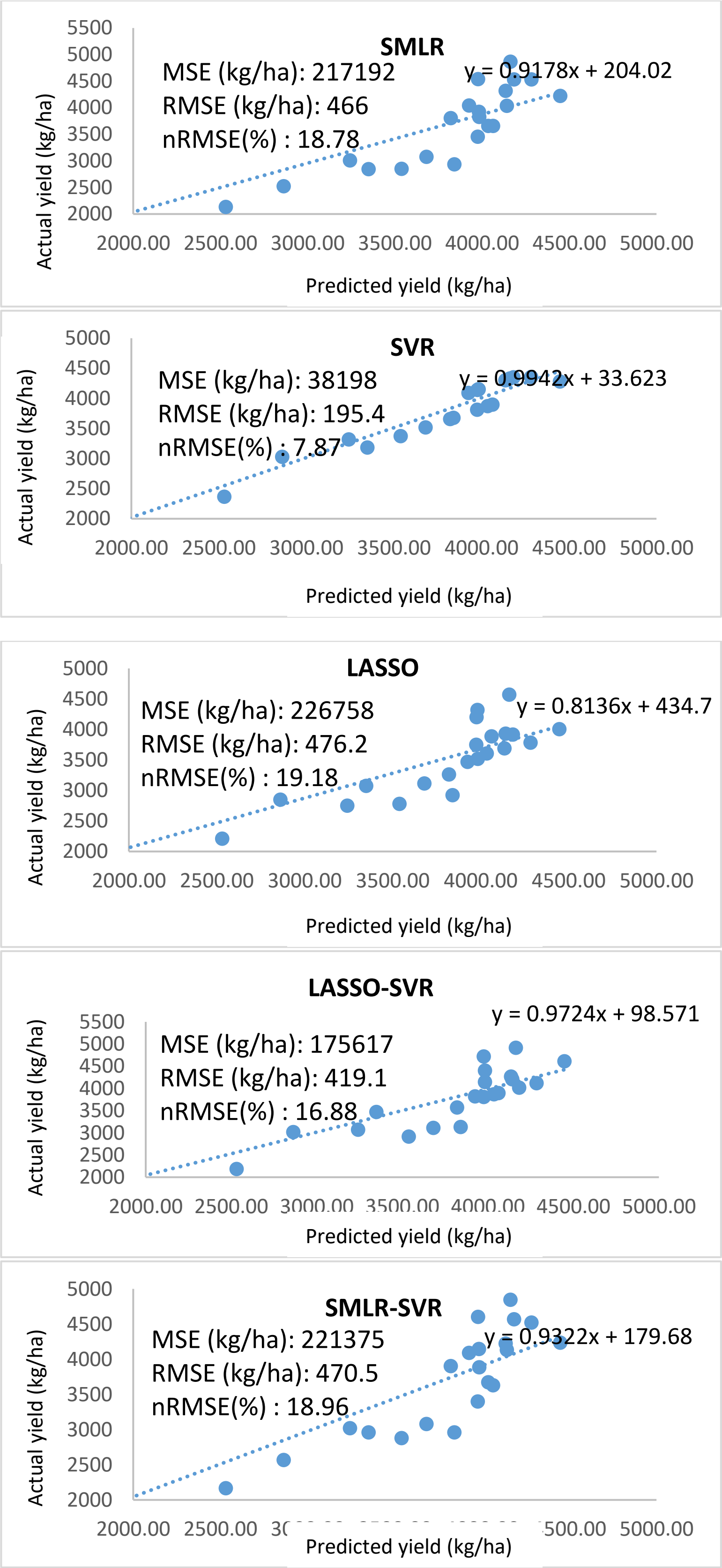


Fig.14.a Calibration of Models developed for wheat yield prediction at tillering stage for Hisar.

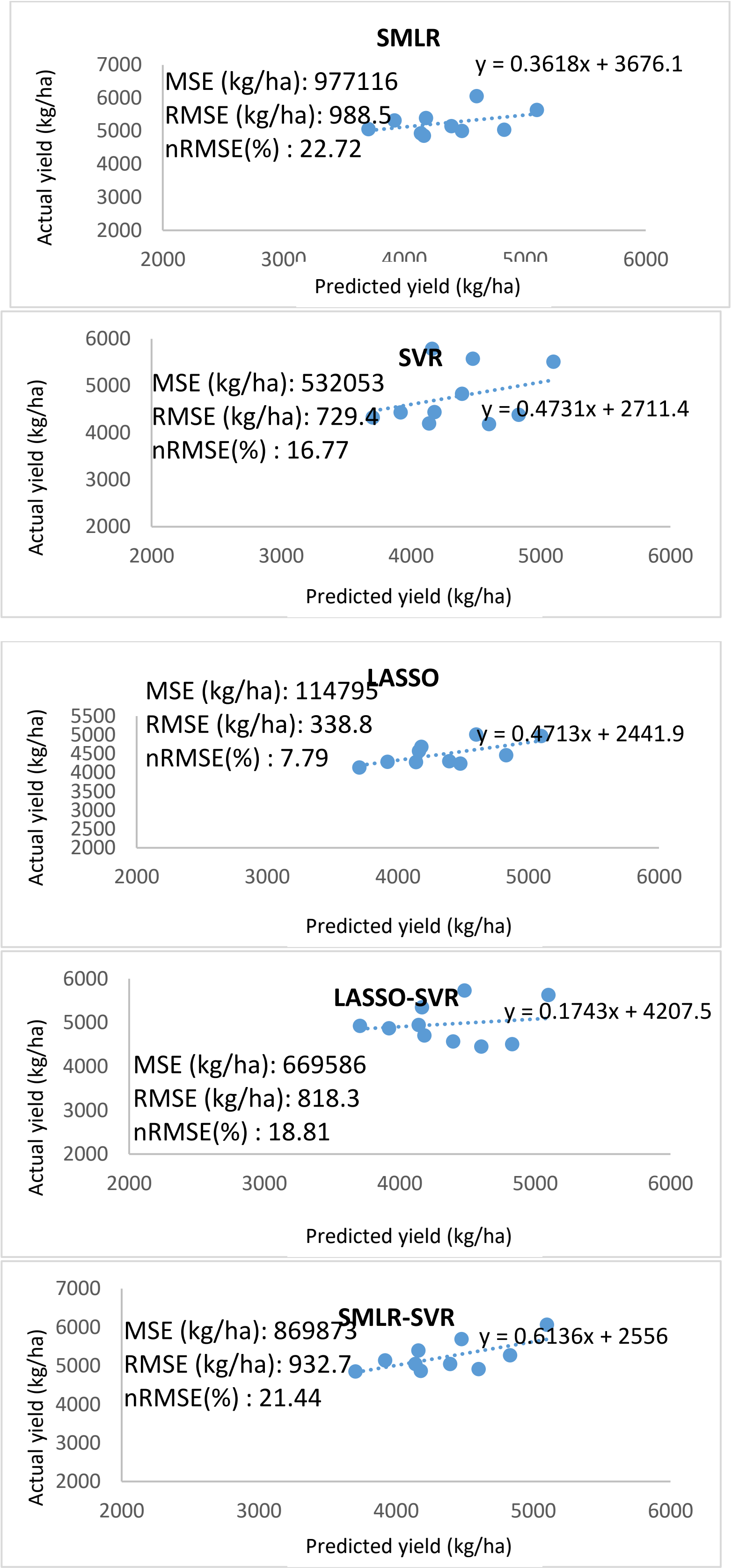


Fig.14.b Validation of Models developed for wheat yield prediction at tillering stage for Hisar.

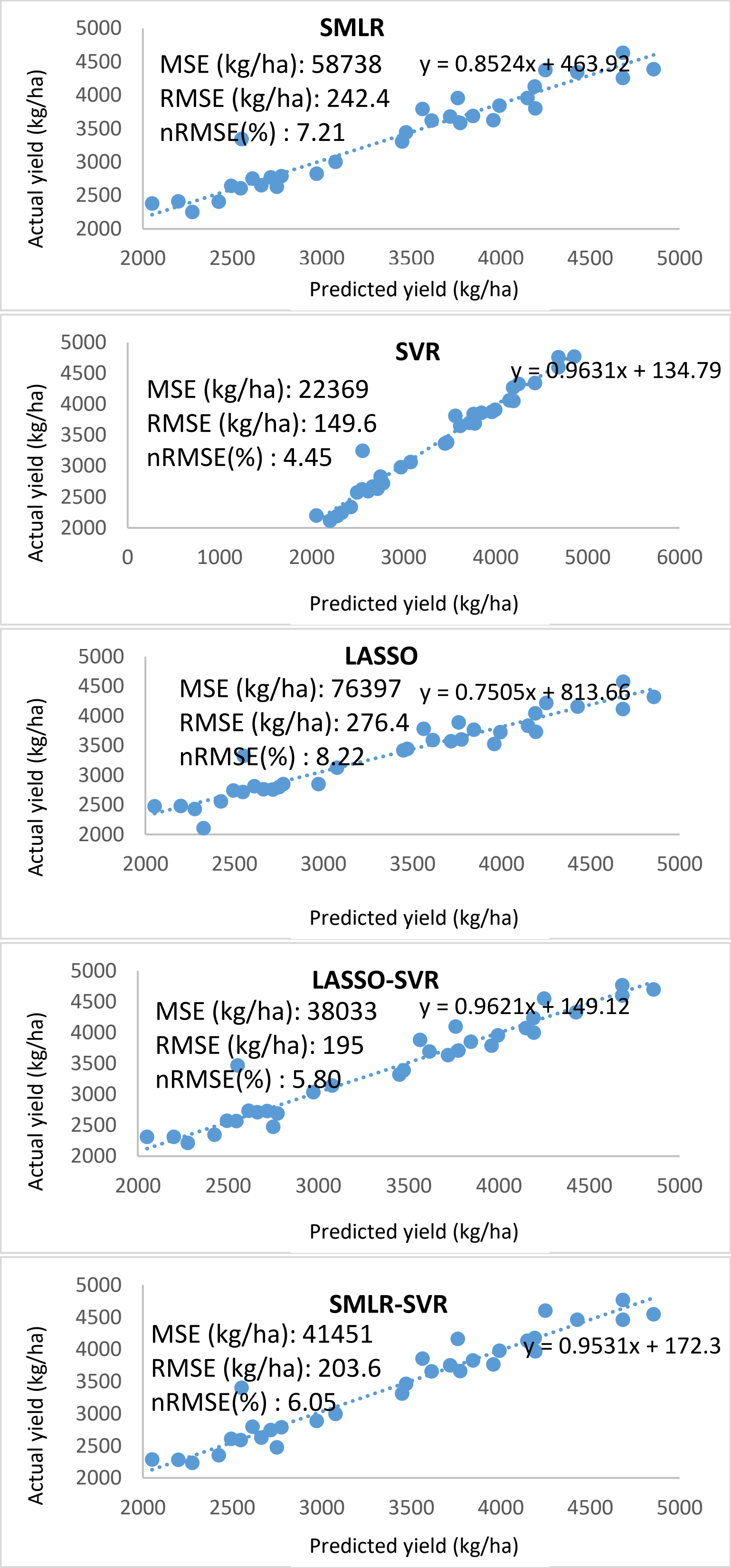


Fig.15.a Calibration of Models developed for wheat yield prediction at tillering stage for Amritsar.

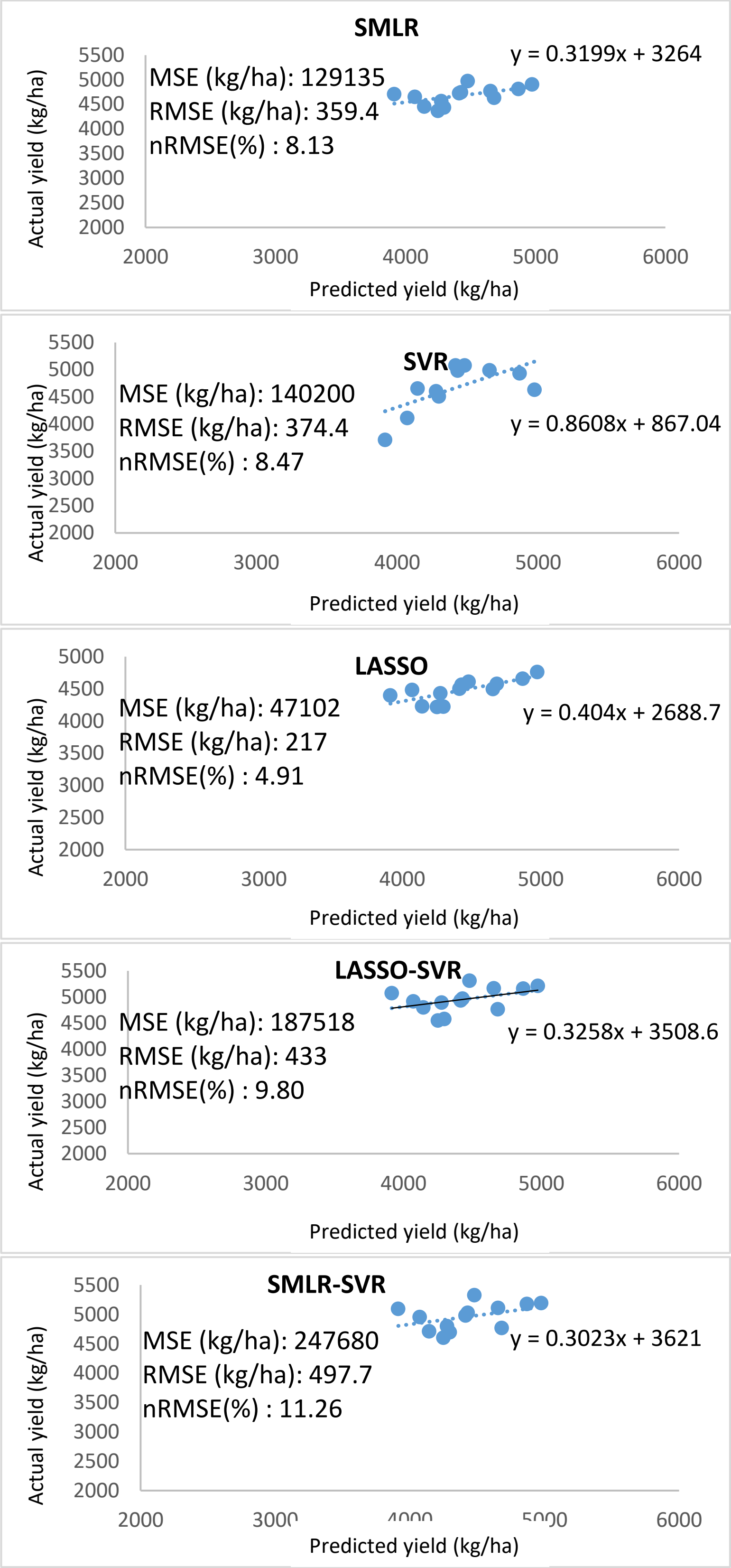


Fig.15.b Validation of Models developed for wheat yield prediction at tillering stage for Amritsar.

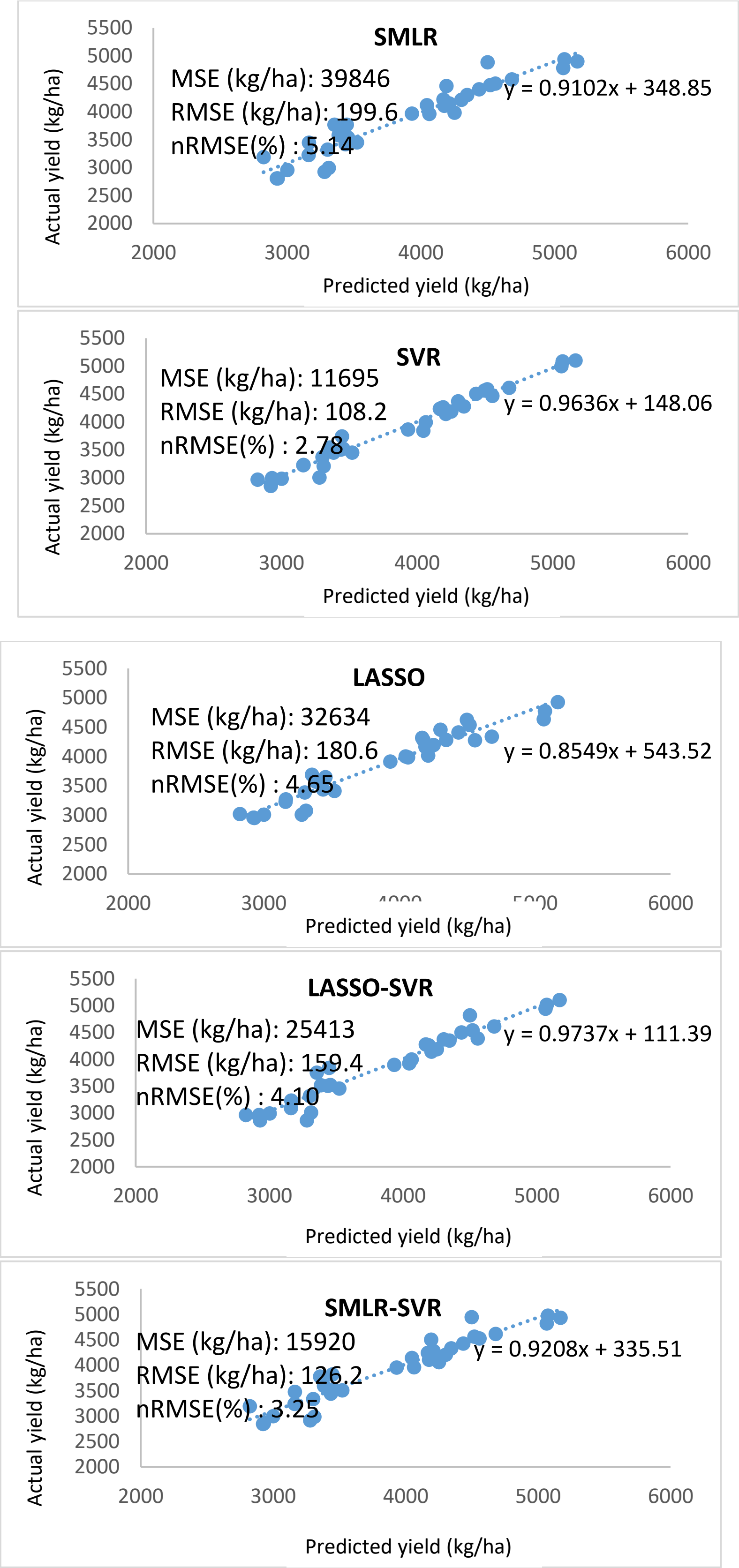


Fig.16.a Calibration of Models developed for wheat yield prediction at tillering stage for Ludhiana.

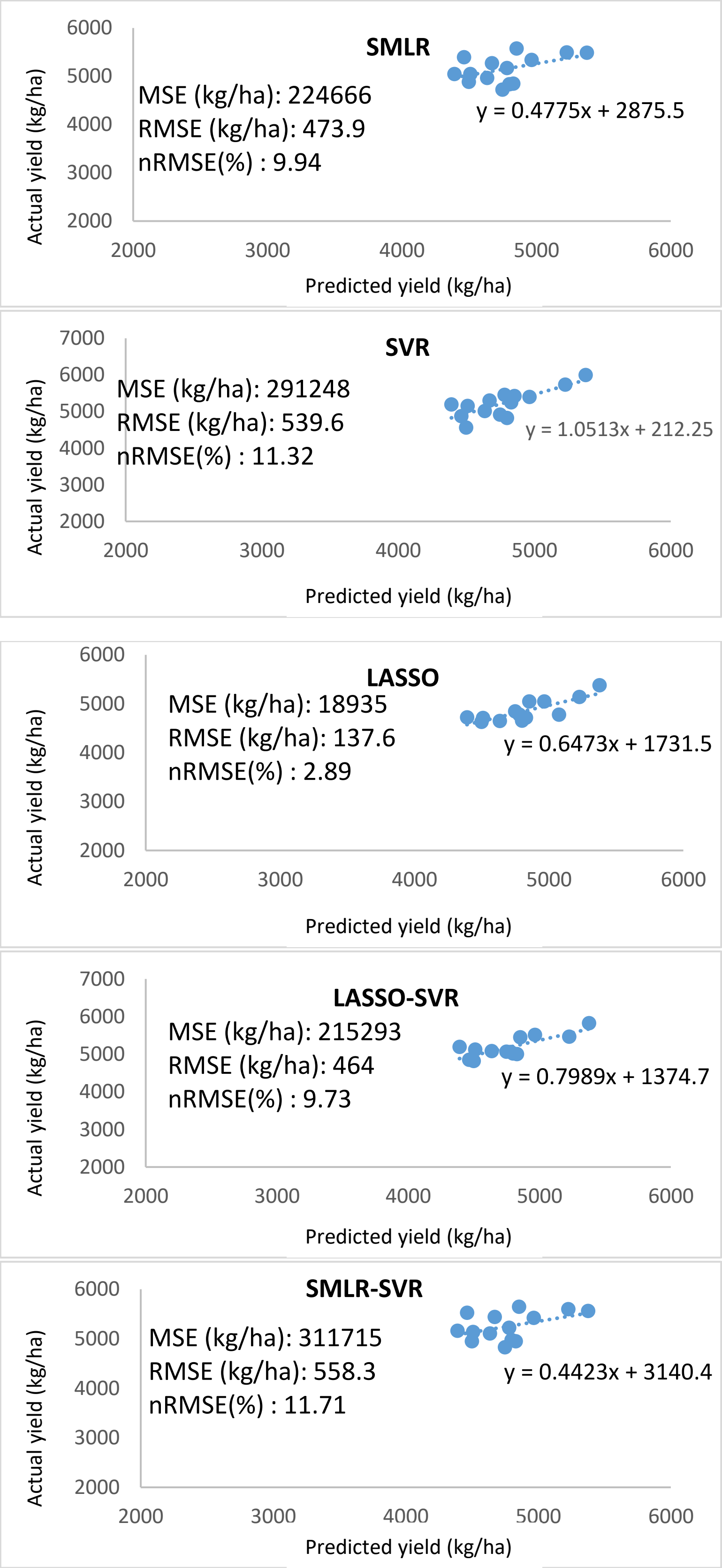


Fig.16.b Validation of Models developed for wheat yield prediction at tillering stage for Ludhiana.

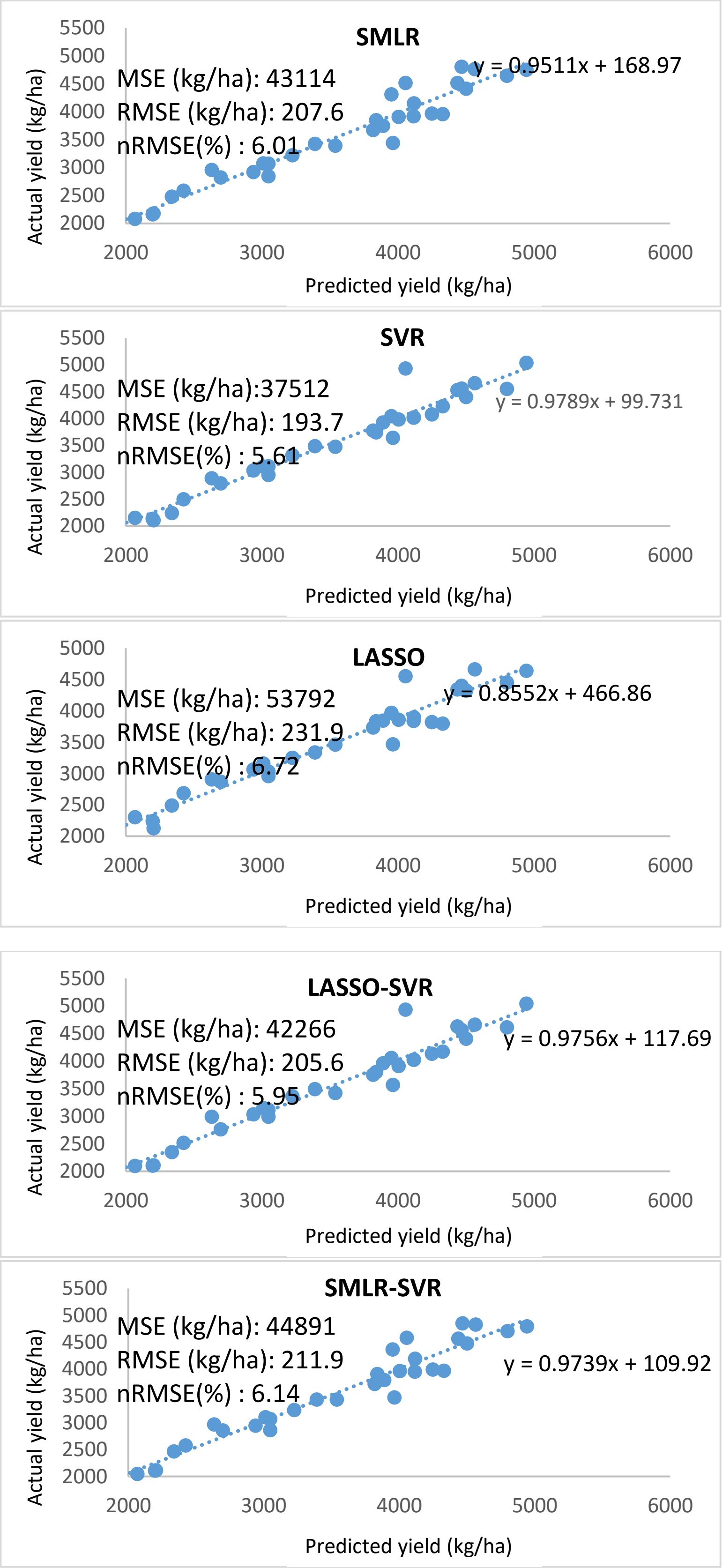


Fig.17.a Calibration of Models developed for wheat yield prediction at tillering stage for Patiala.

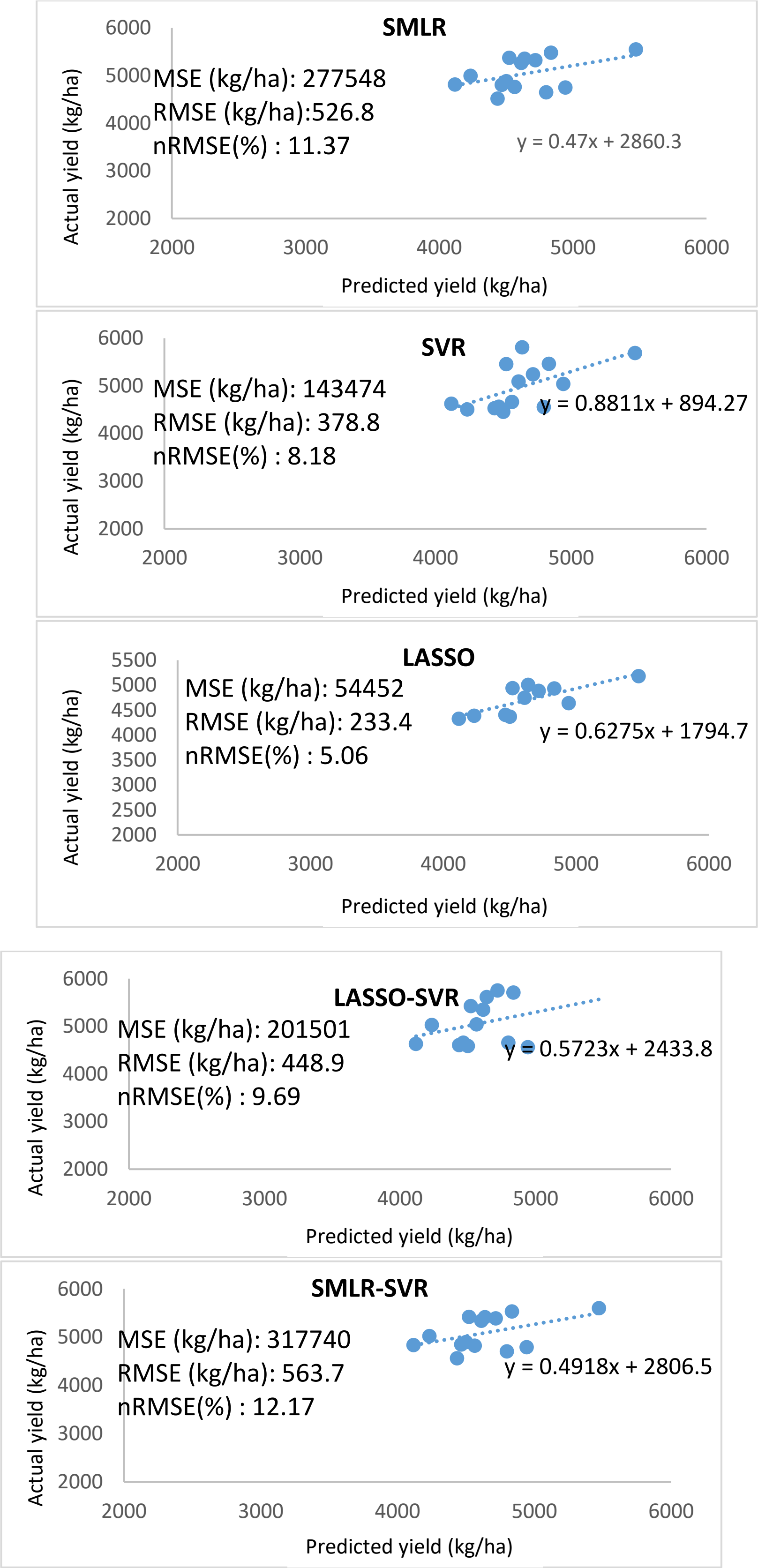


Fig.17.b Validation of Models developed for wheat yield prediction at tillering stage for Patiala

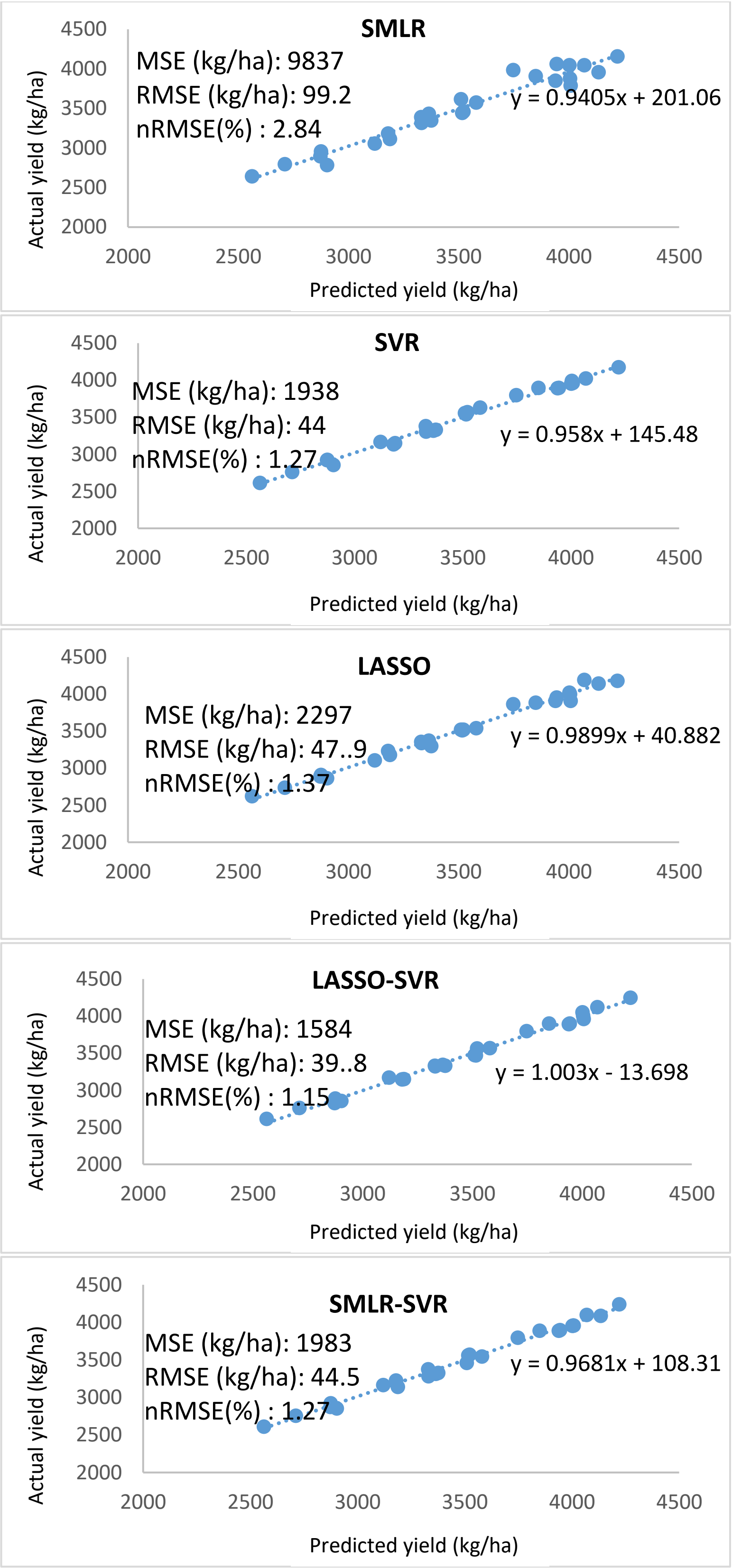


Fig.18.a Calibration of Models developed for wheat yield prediction at flowering stage for IARI.

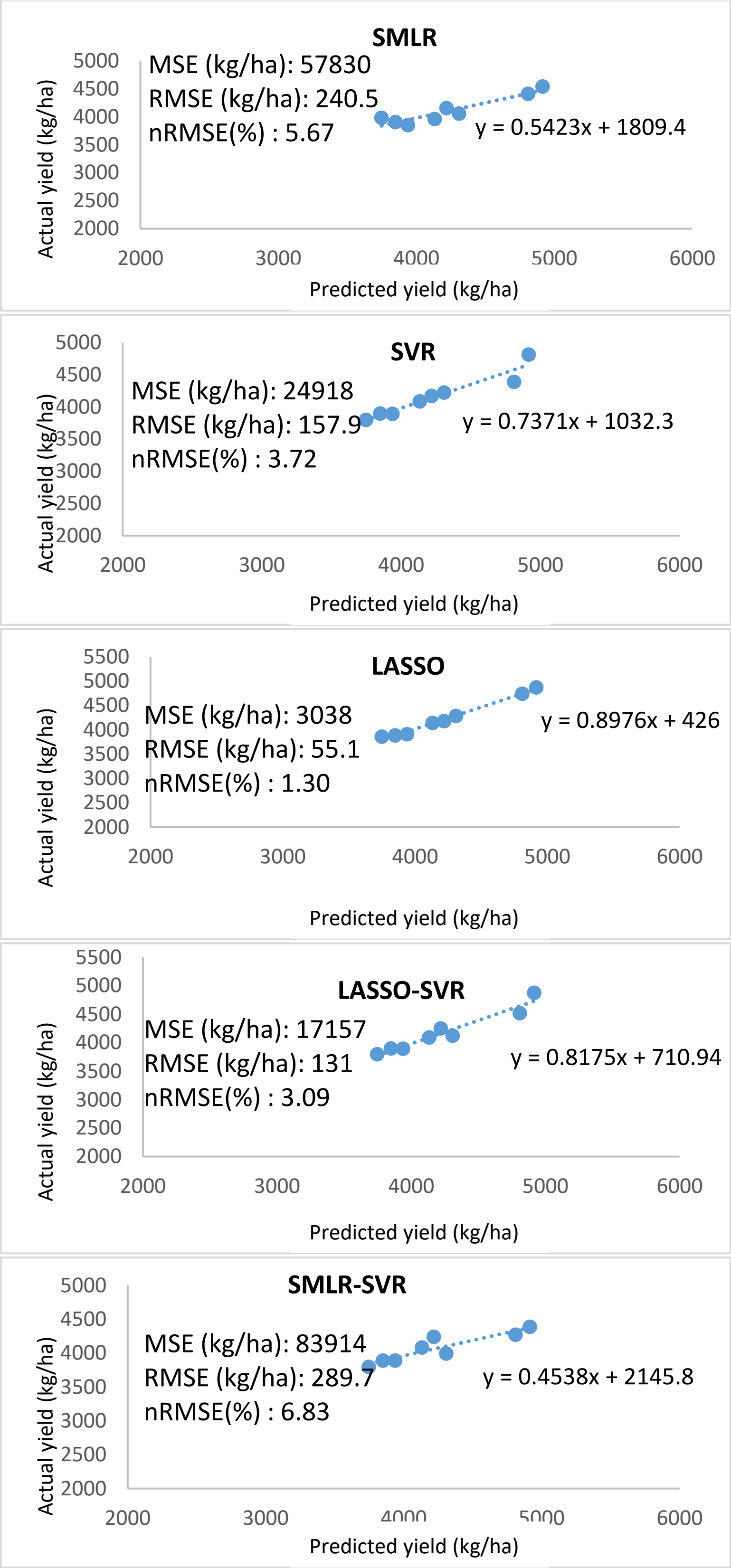


Fig.18.b Validation of Models developed for wheat yield prediction at flowering stage for IARI.

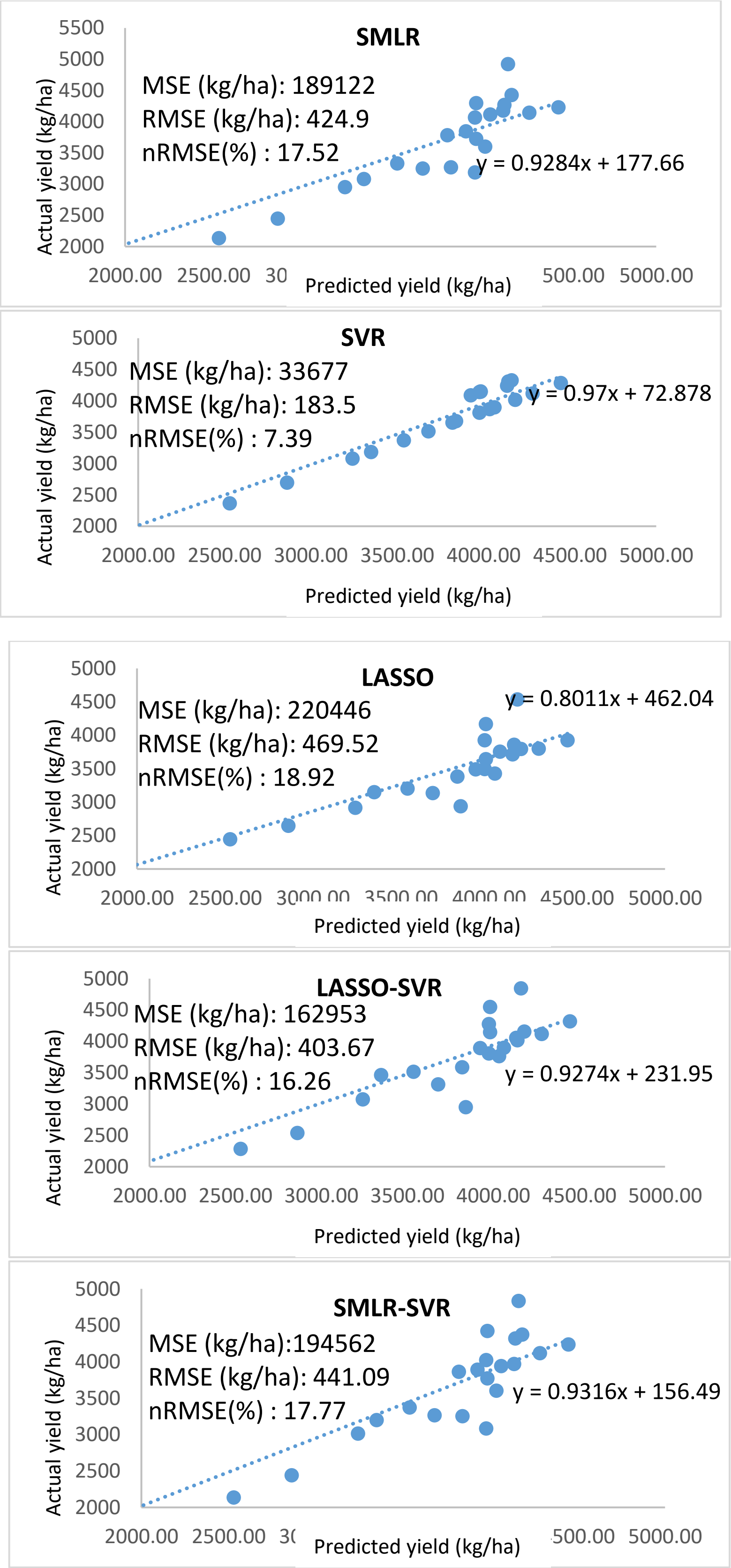


Fig.19a Calibration of Models developed for wheat yield prediction at flowering stage for Hisar

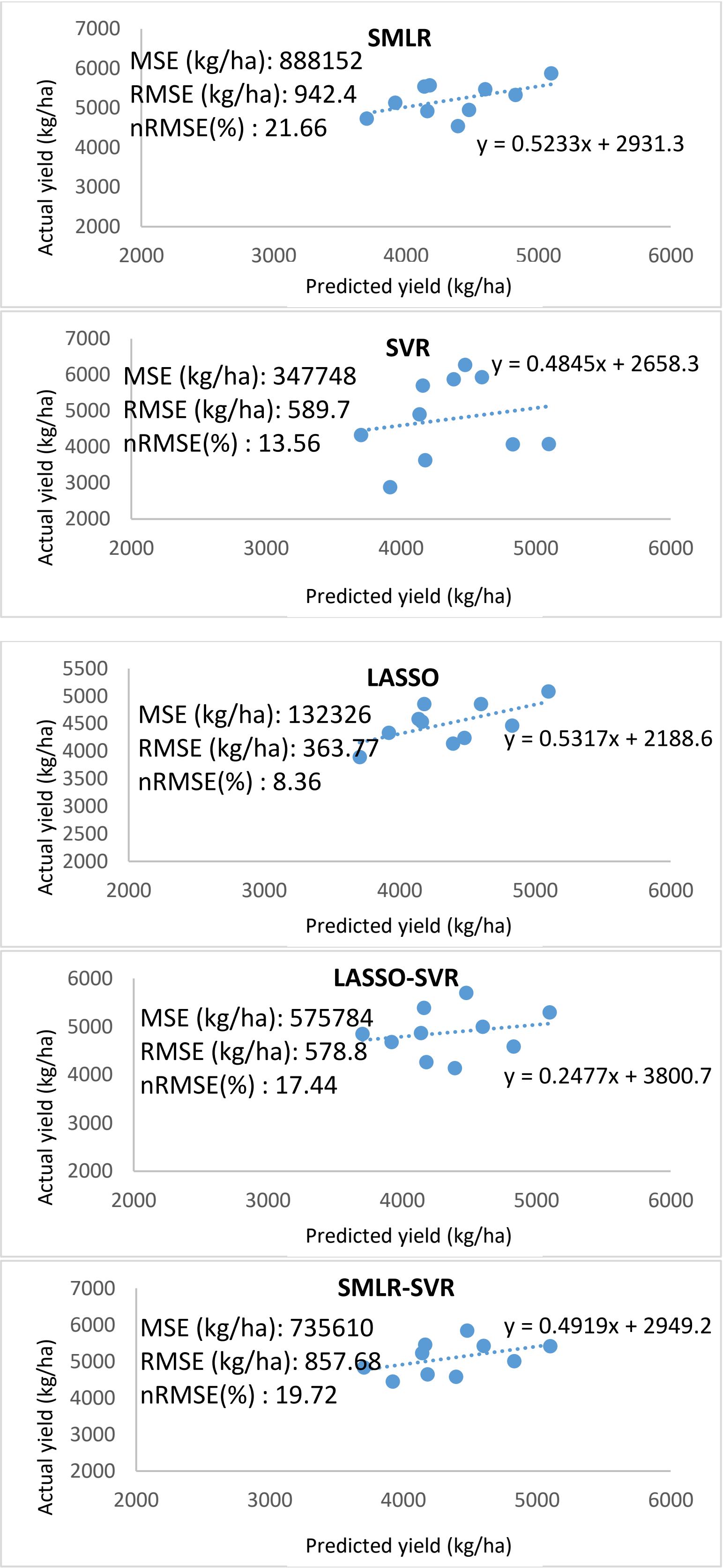


Fig.19.b Validation of Models developed for wheat yield prediction at flowering stage for Hisar.

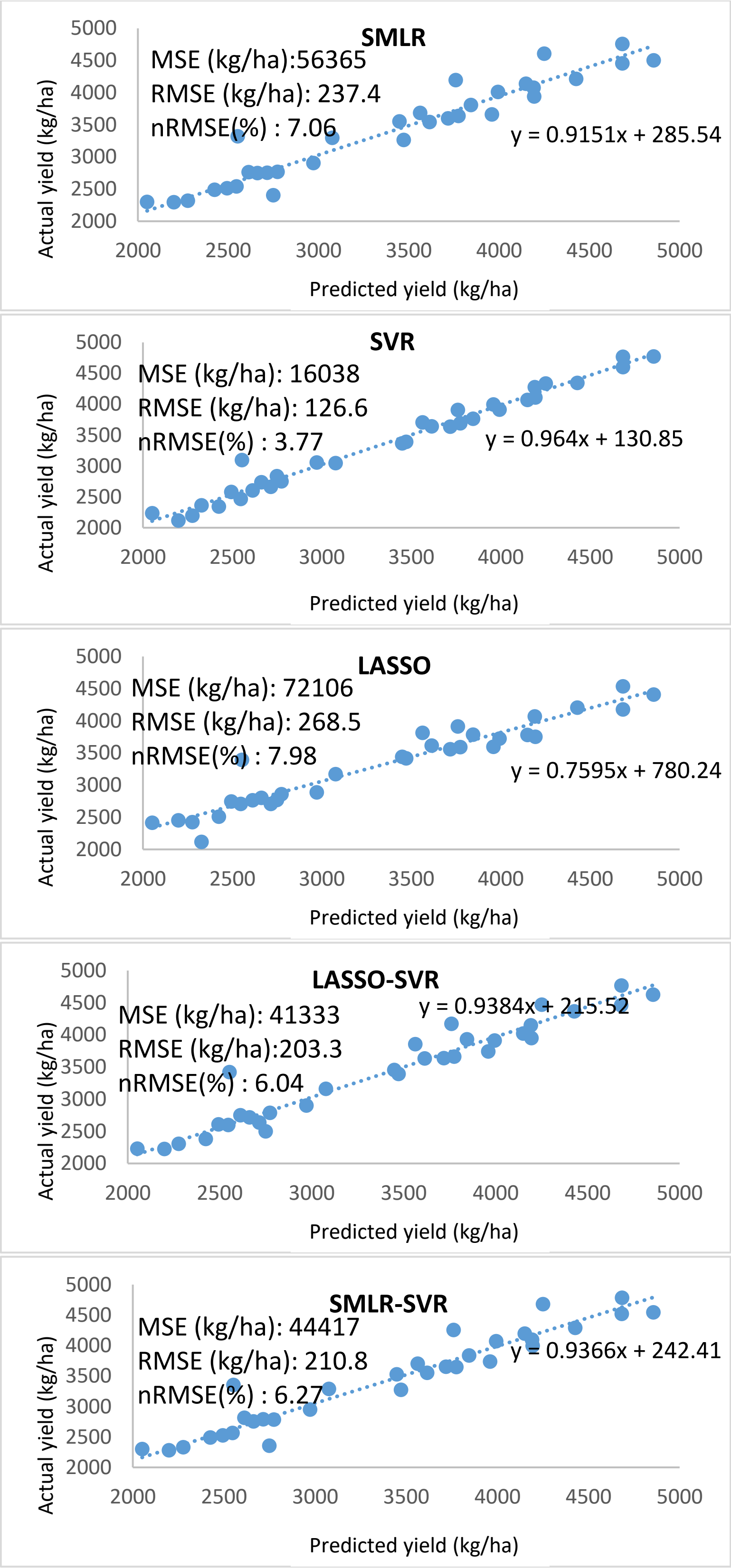


Fig.20.a Calibration of Models developed for wheat yield prediction at flowering stage for Amritsar.

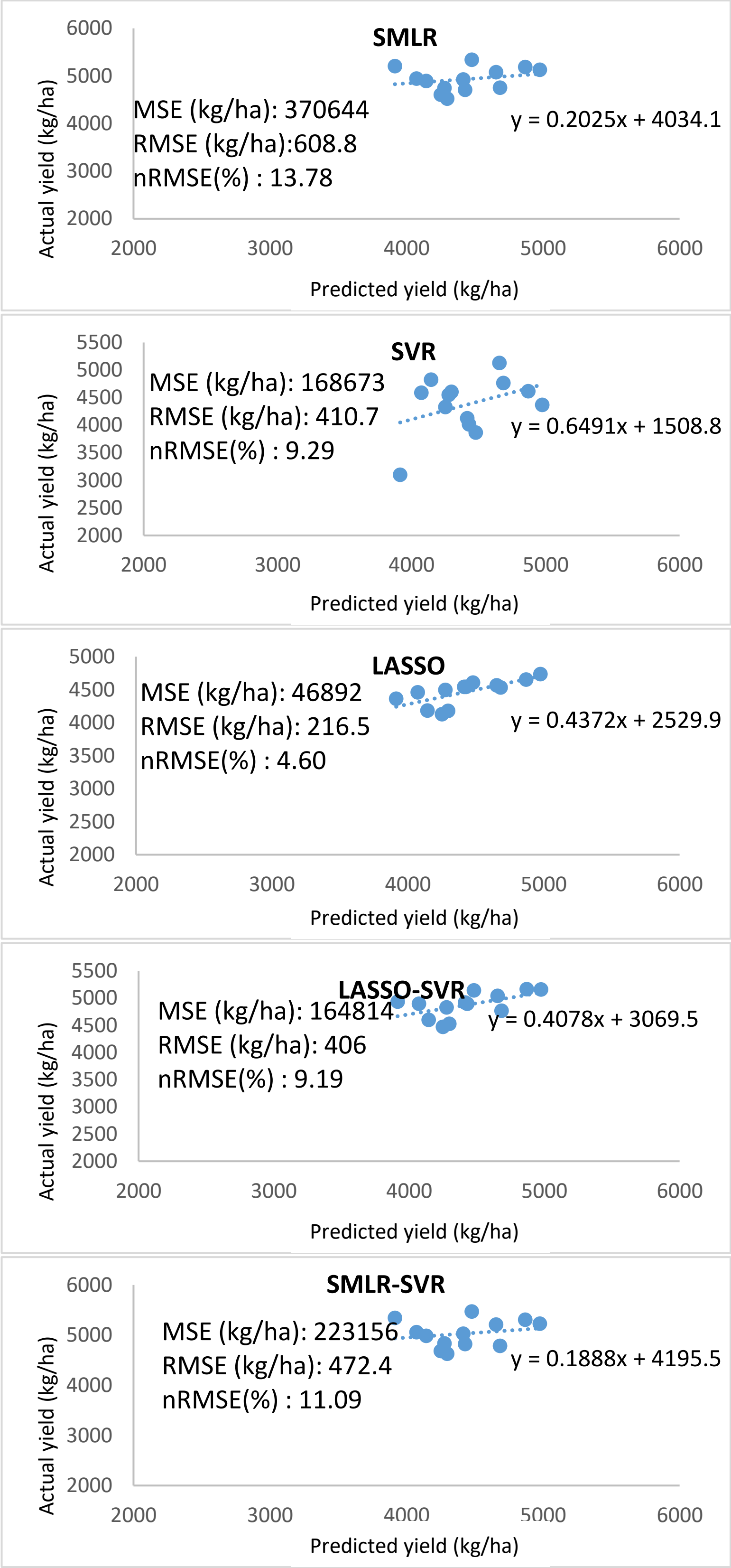


Fig.20.b Validation of Models developed for wheat yield prediction at flowering stage for Amritsar.

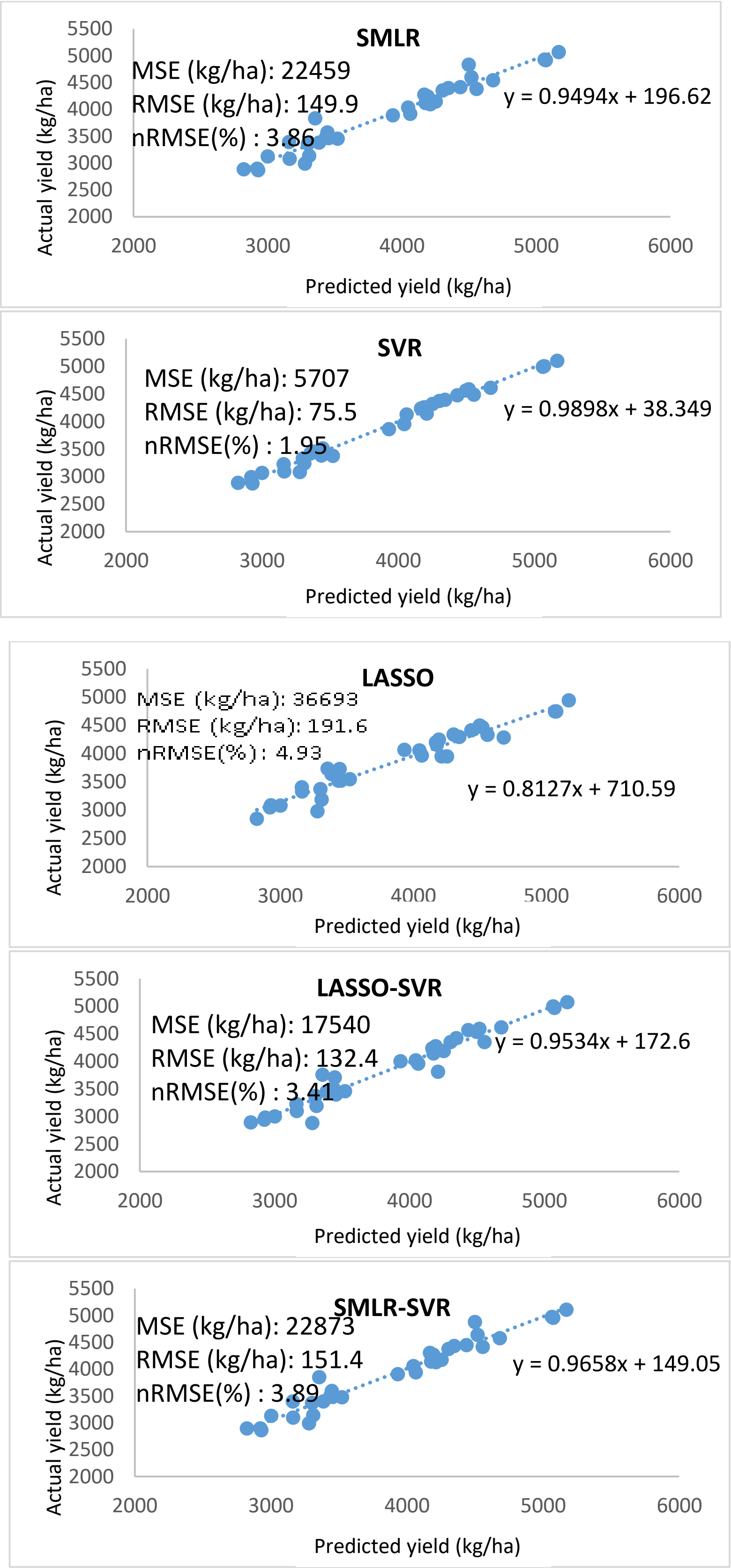


Fig.21.a Calibration of Models developed for wheat yield prediction at flowering stage for Ludhiana.

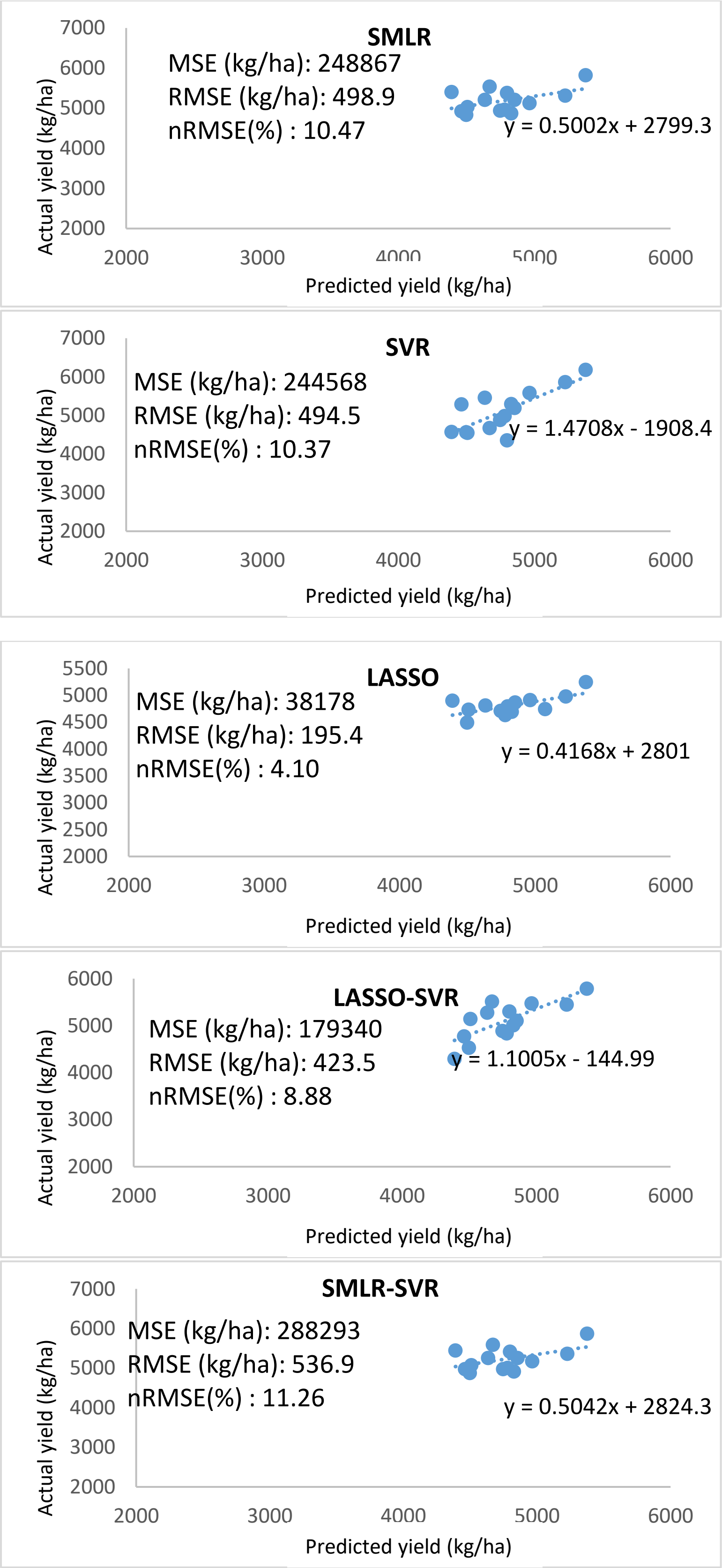


Fig.21.b Validation of Models developed for wheat yield prediction at flowering stage for Ludhiana.

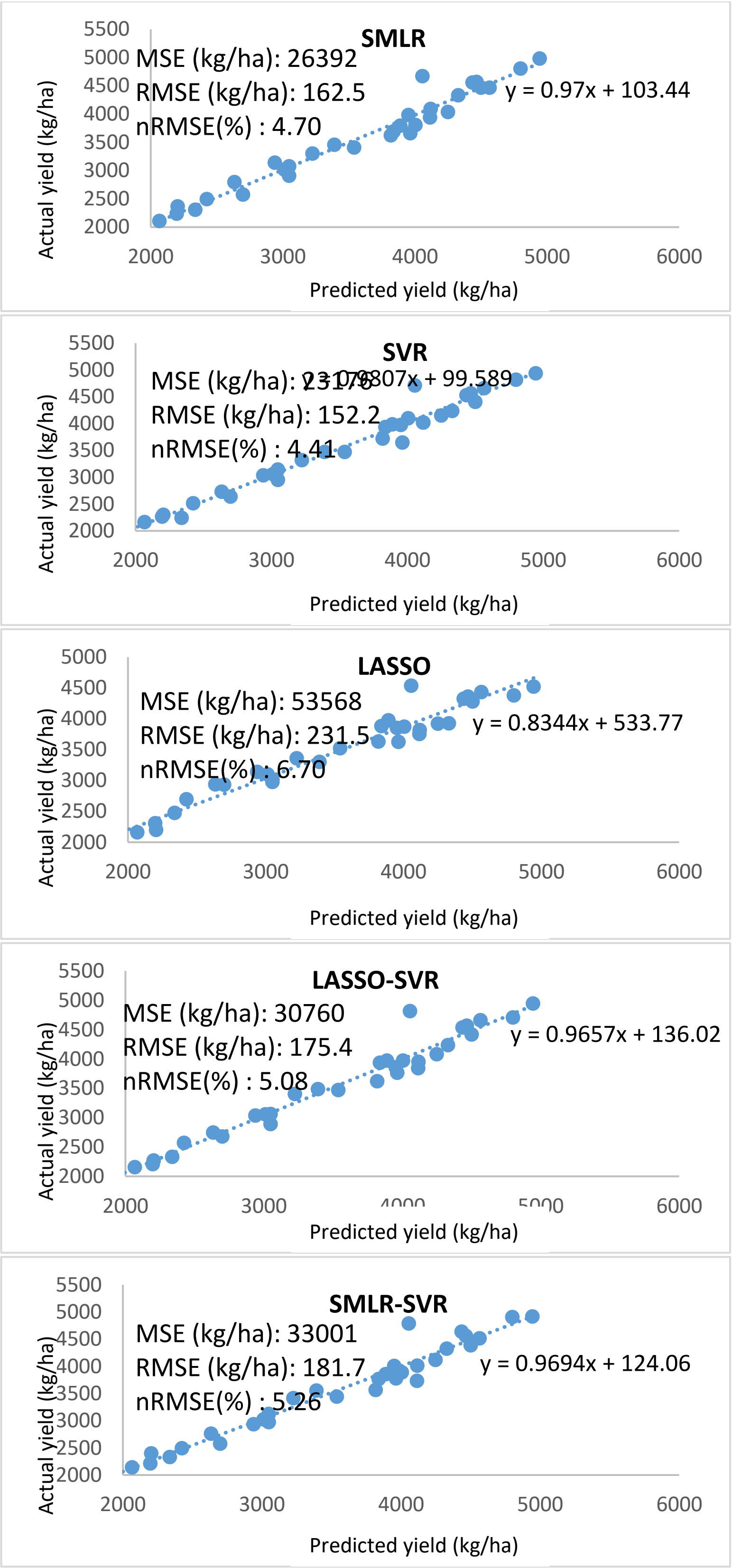


Fig.22.a Calibration of Models developed for wheat yield prediction at flowering stage for Patiala.

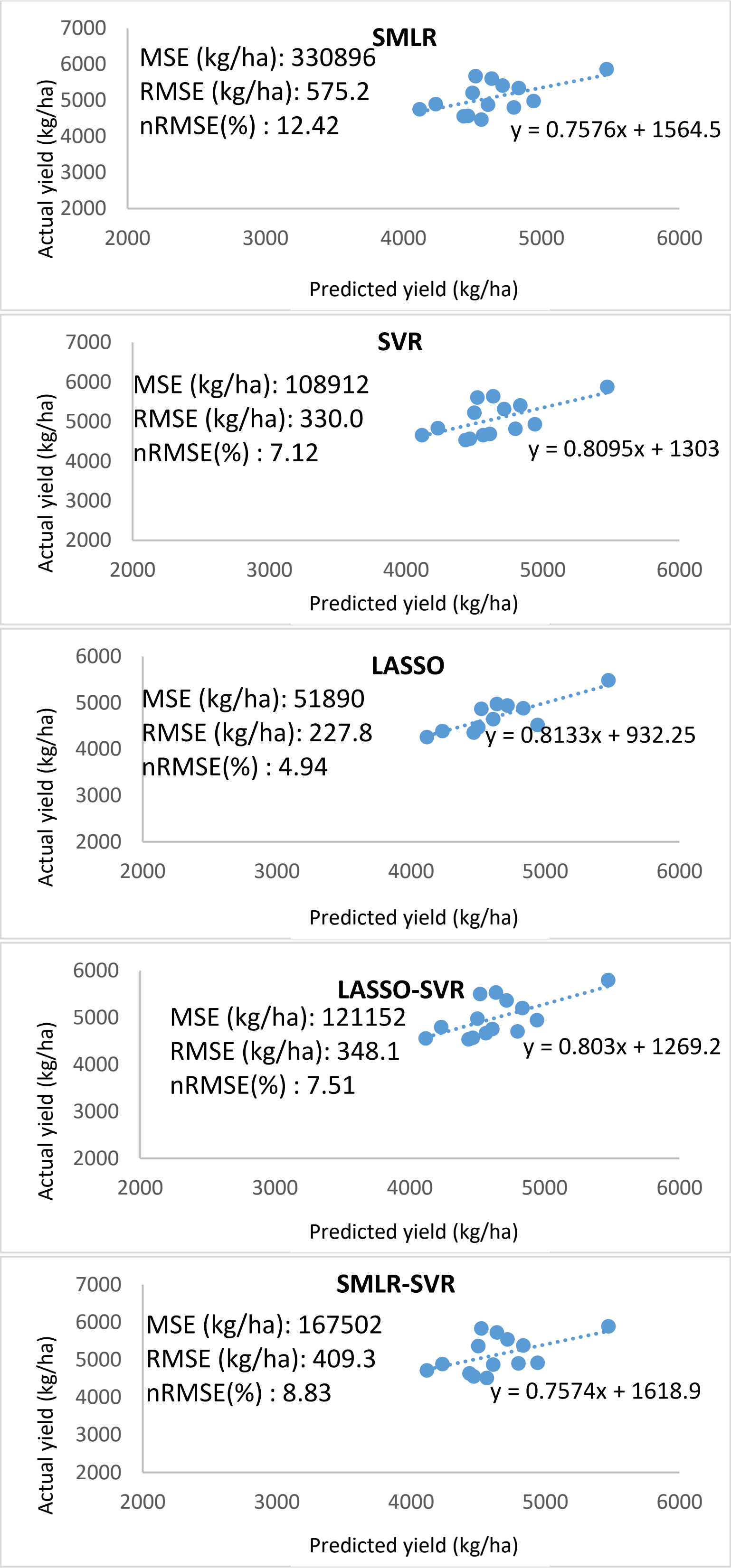


Fig.22.b Validation of Models developed for wheat yield prediction at flowering stage for Patiala.

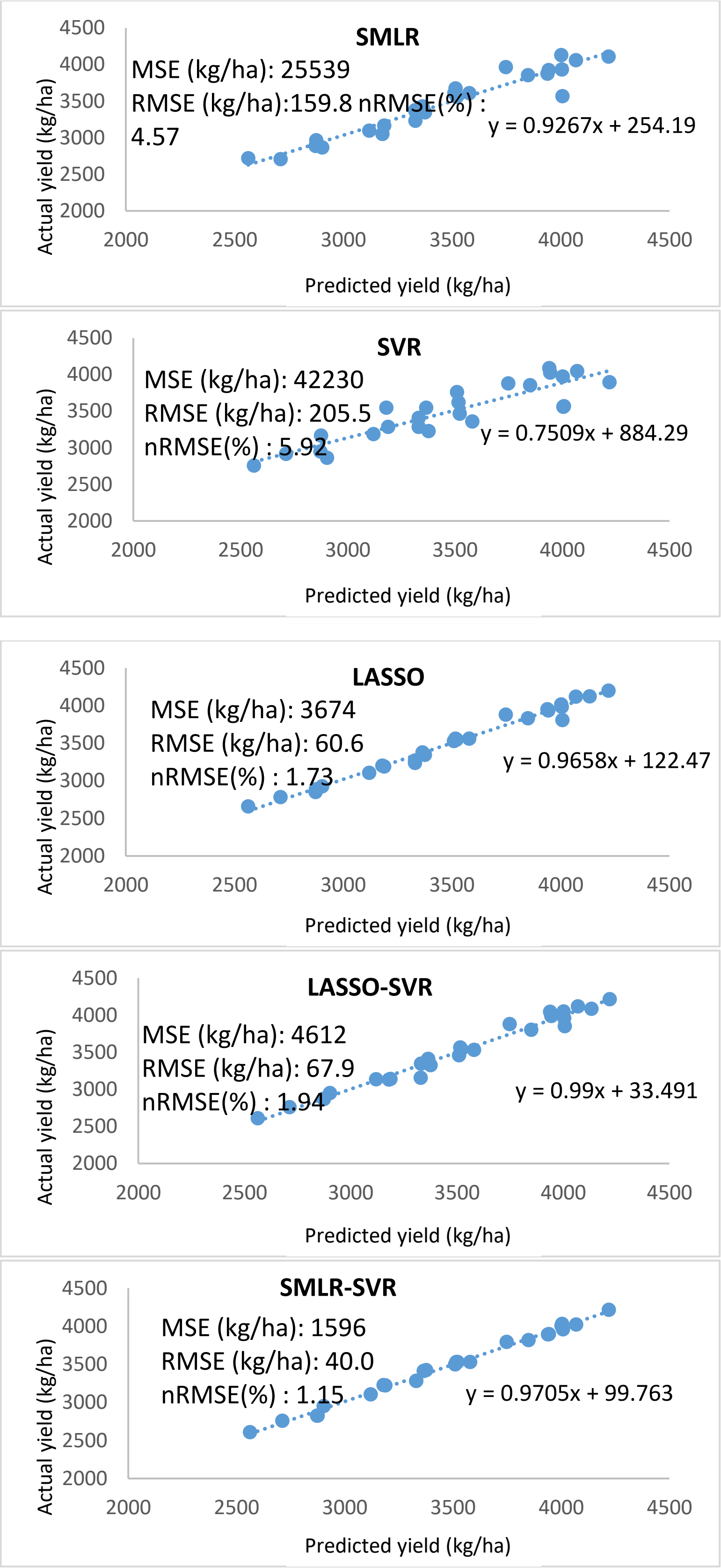


Fig.23.a Calibration of Models developed for wheat yield prediction at grain filling stage for IARI.

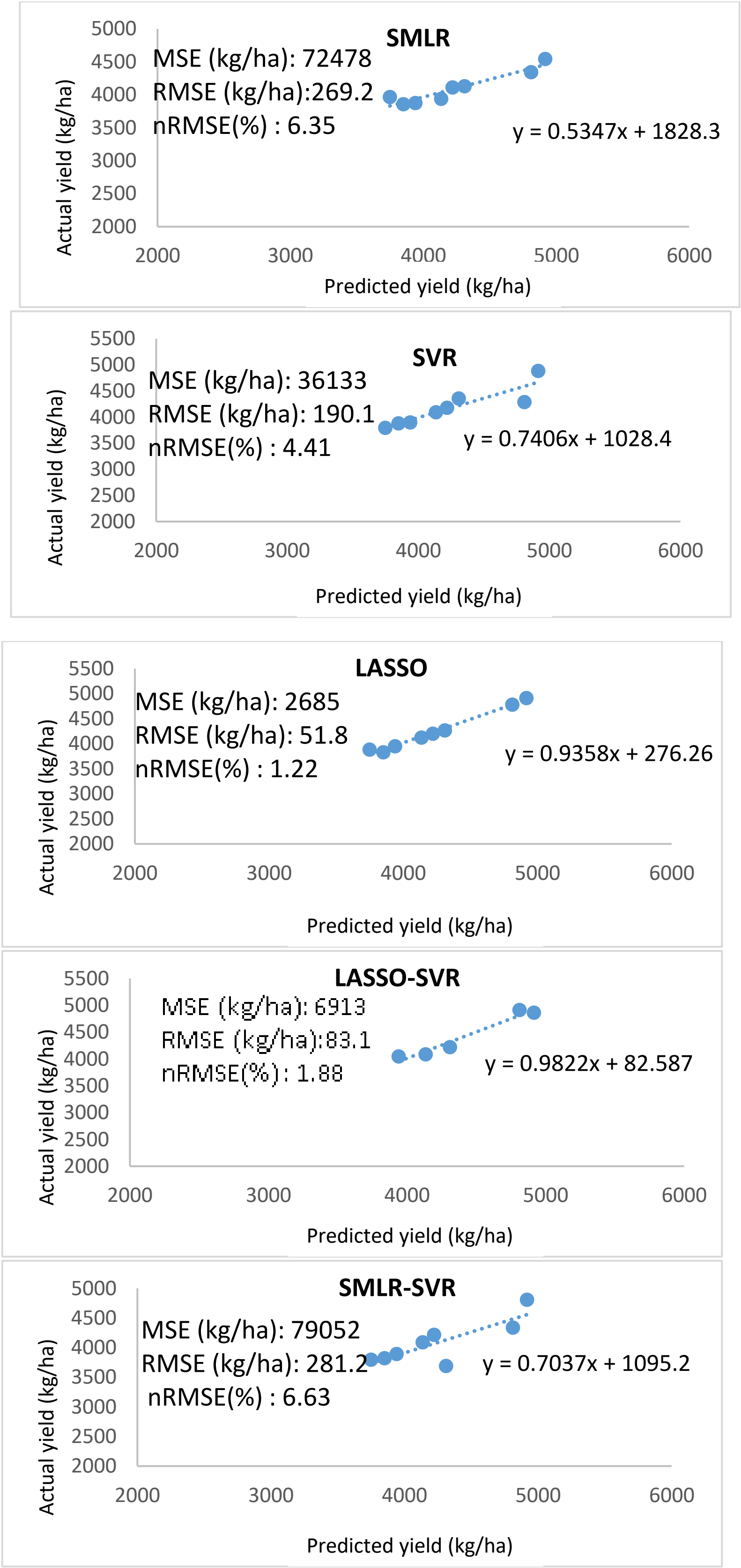


Fig.23.b Validation of Models developed for wheat yield prediction at grain filling stage for IARI.

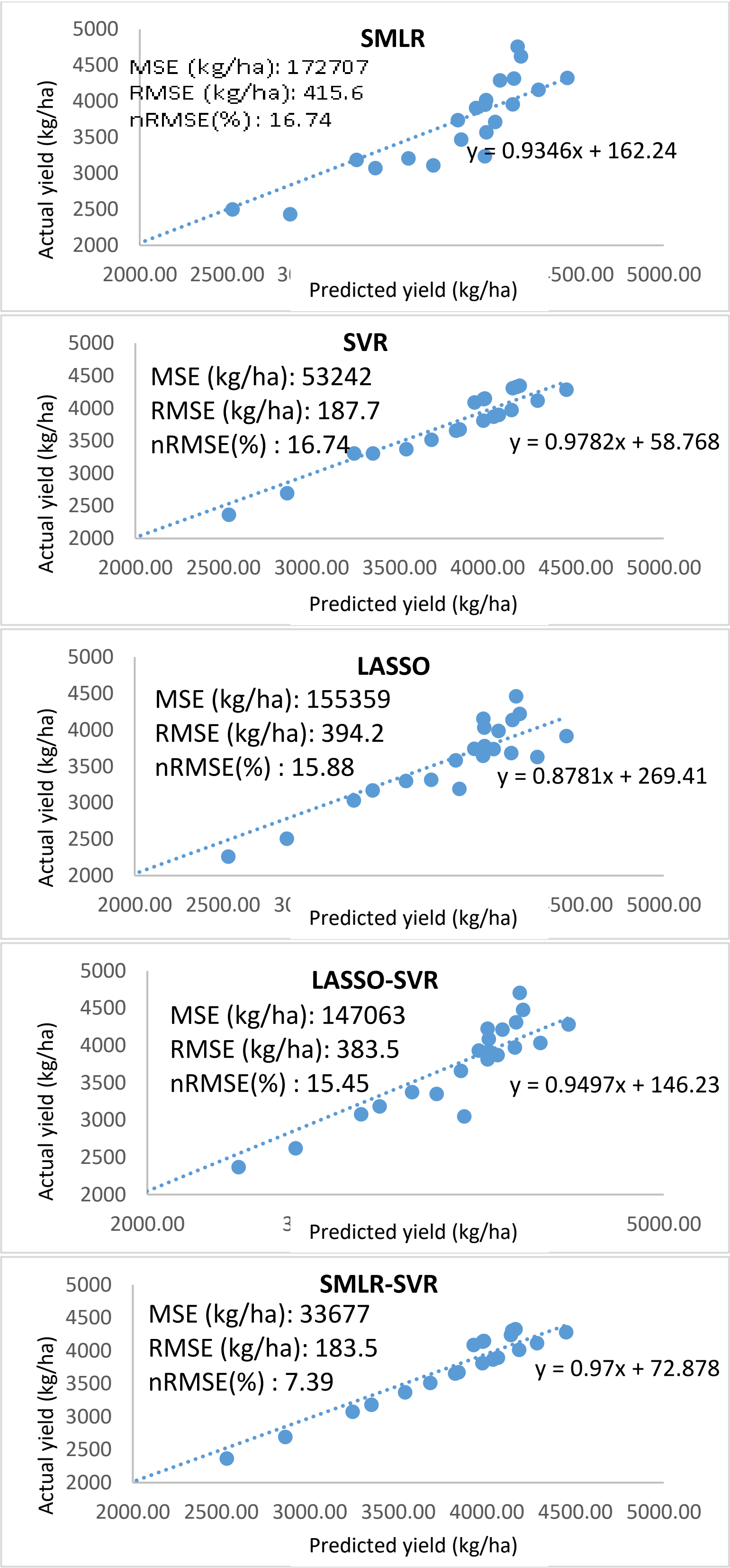


Fig.24.a Calibration of Models developed for wheat yield prediction at grain filling stage for Hisar.

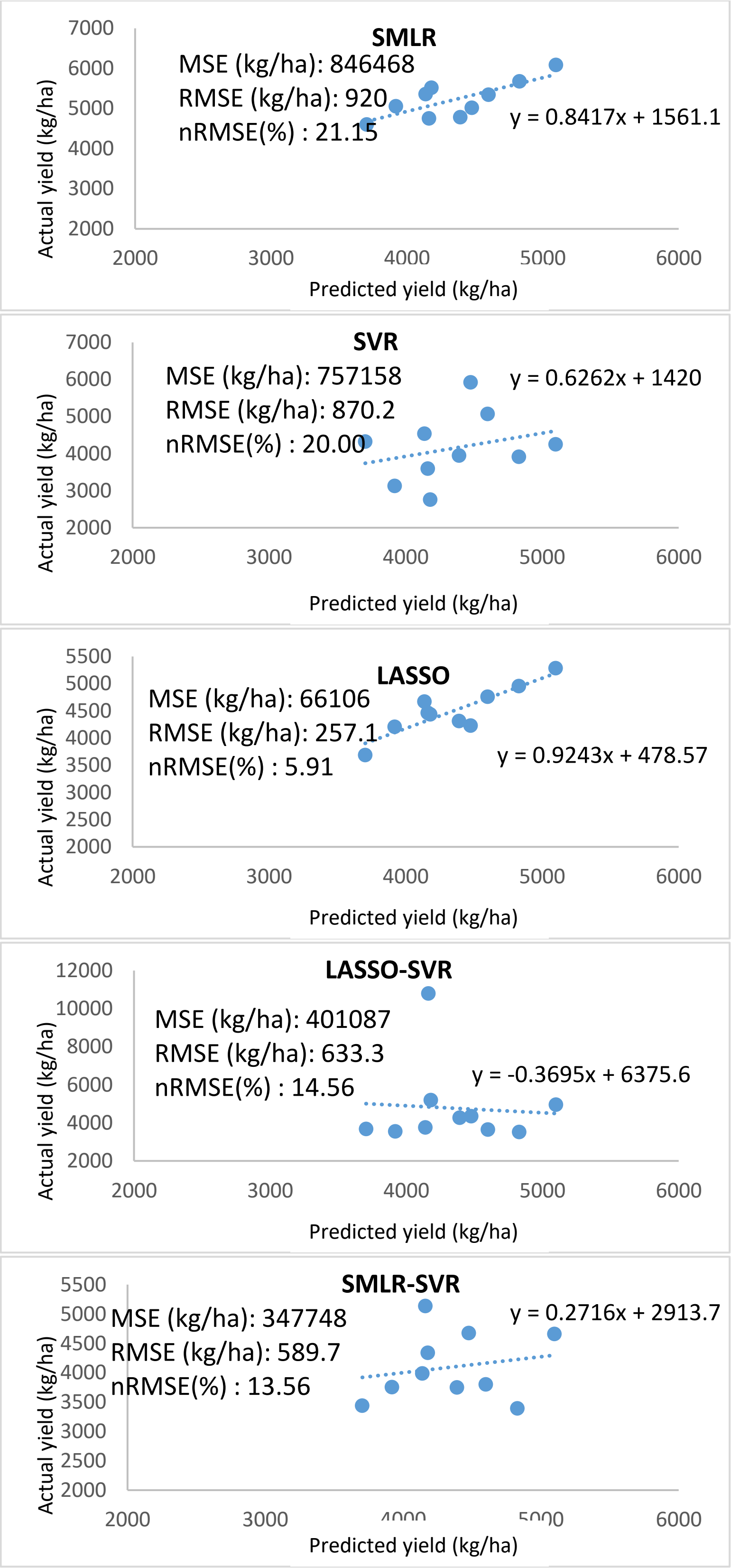


Fig.24.b Validation of Models developed for wheat yield prediction at grain filling stage for Hisar.

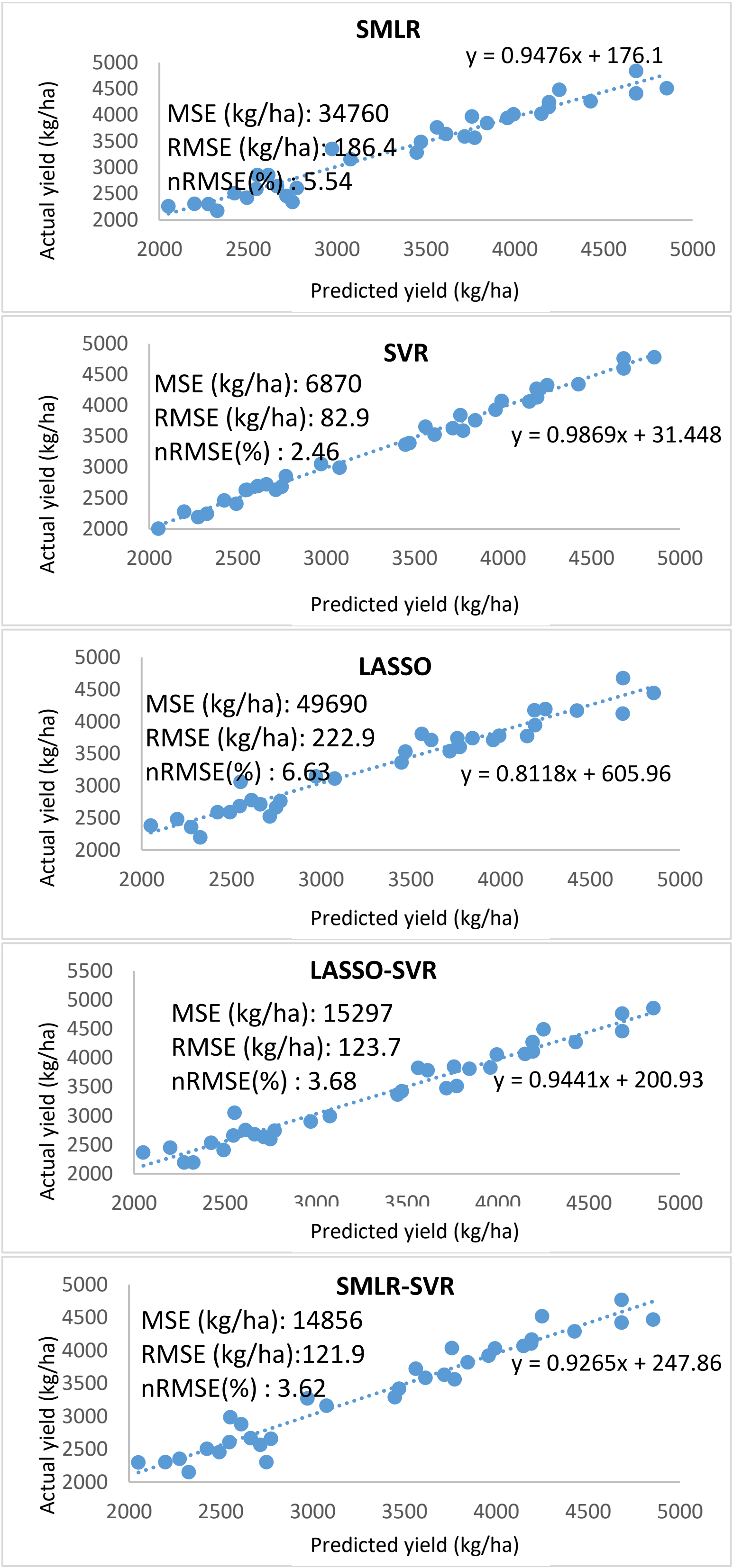


Fig.25.a Calibration of Models developed for wheat yield prediction at grain filling stage for Amritsar.

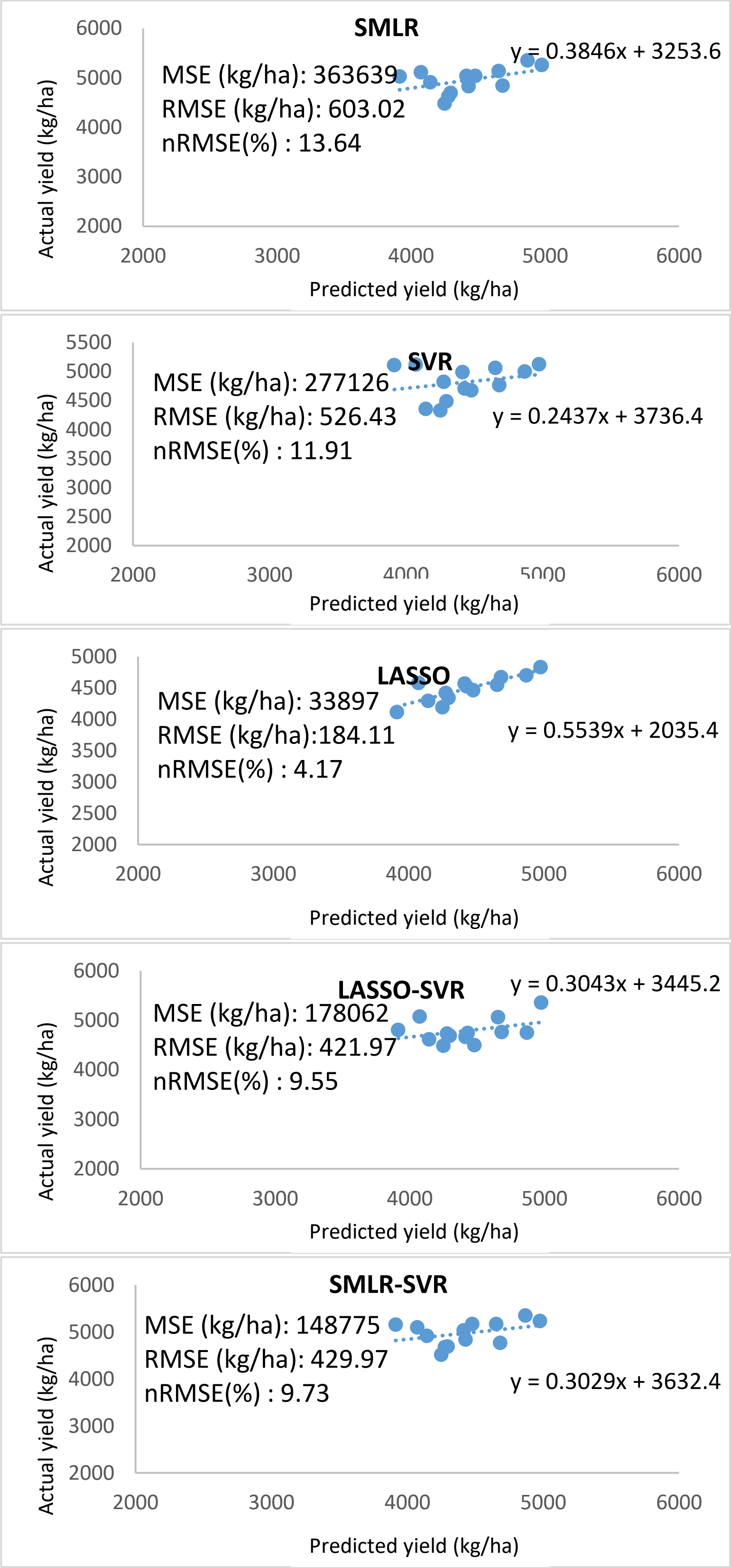


Fig.25.b Validation of Models developed for wheat yield prediction at grain filling stage for Amritsar.

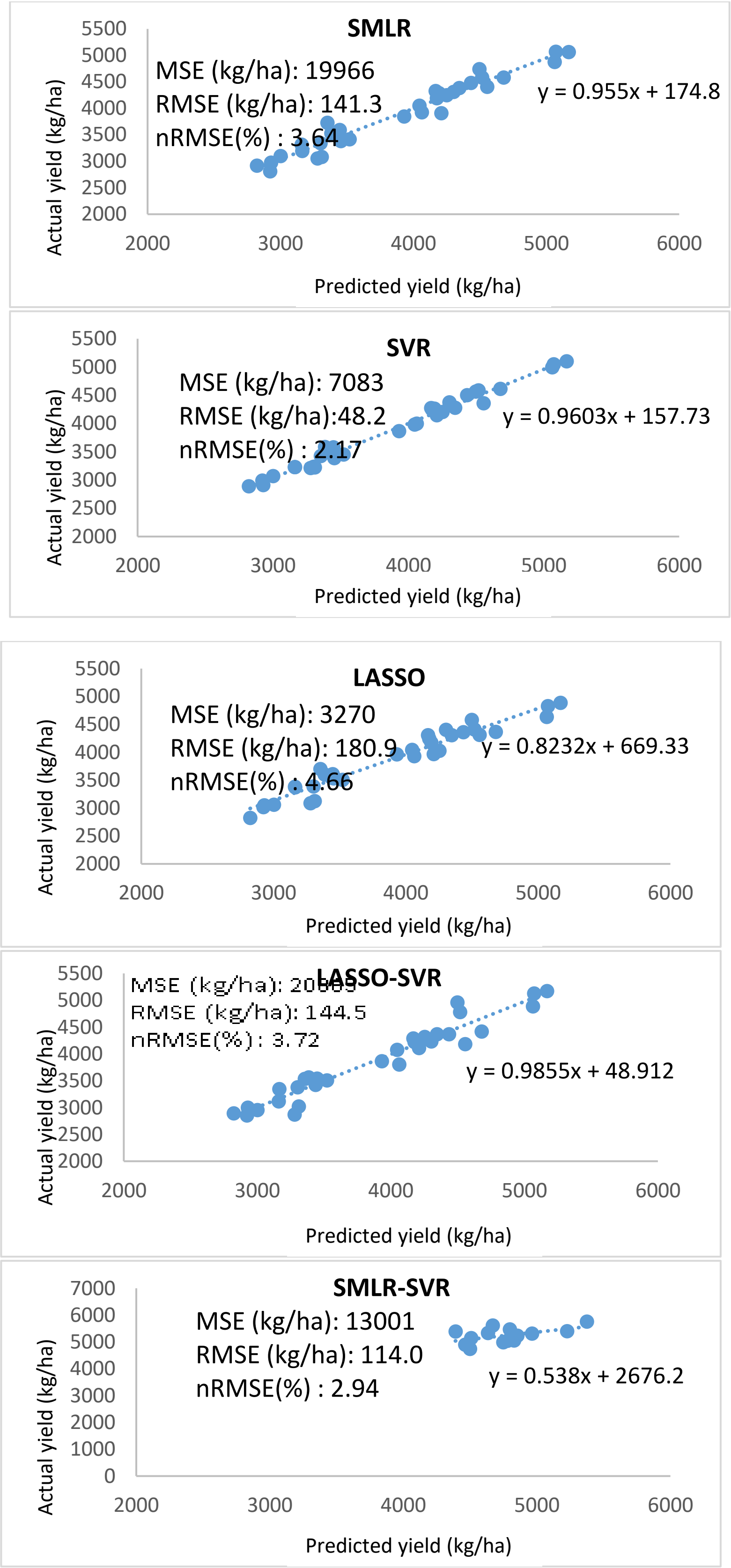


Fig.26.a Calibration of Models developed for wheat yield prediction at grain filling stage for Ludhiana.

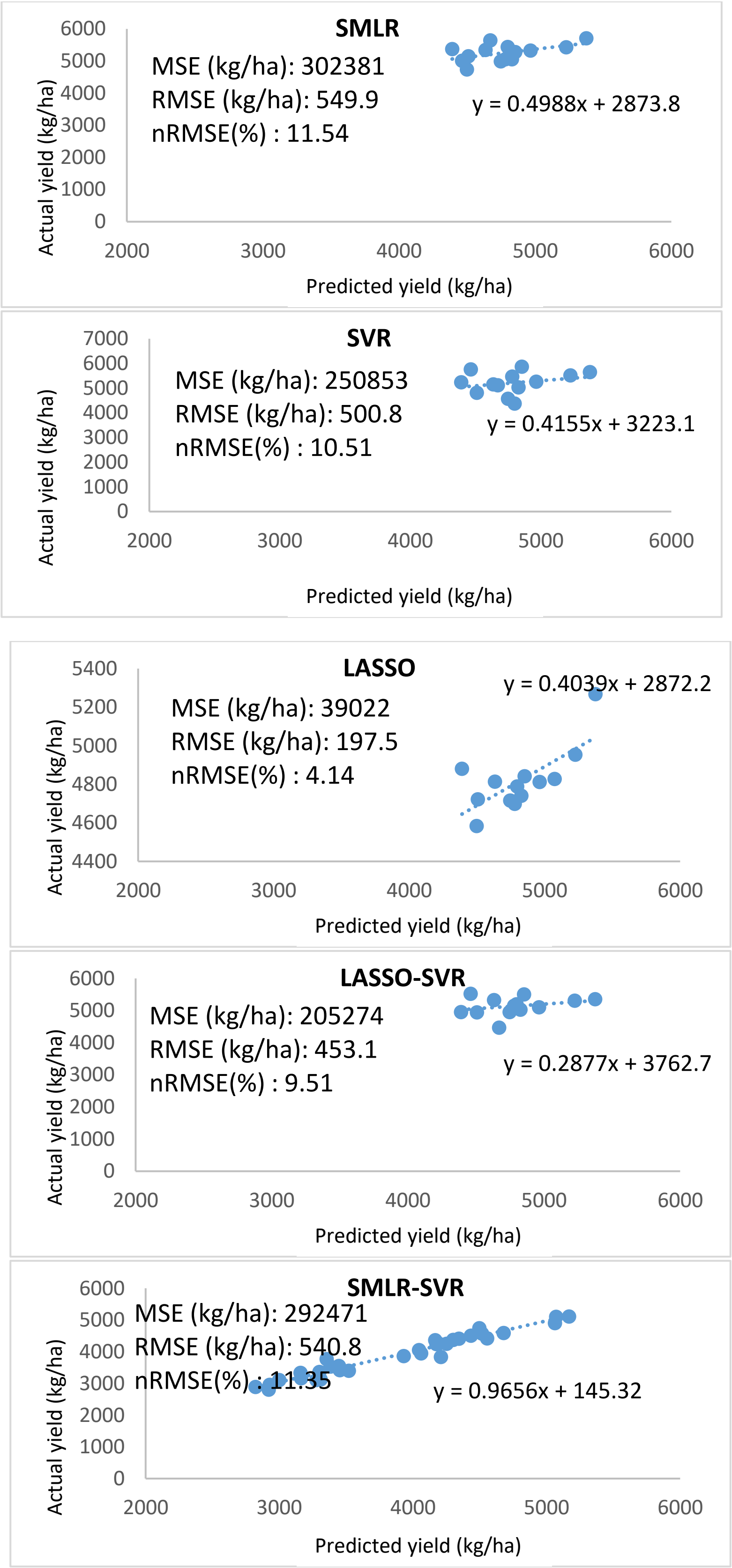


Fig.26.b Validation of Models developed for wheat yield prediction at grain filling stage for Ludhiana

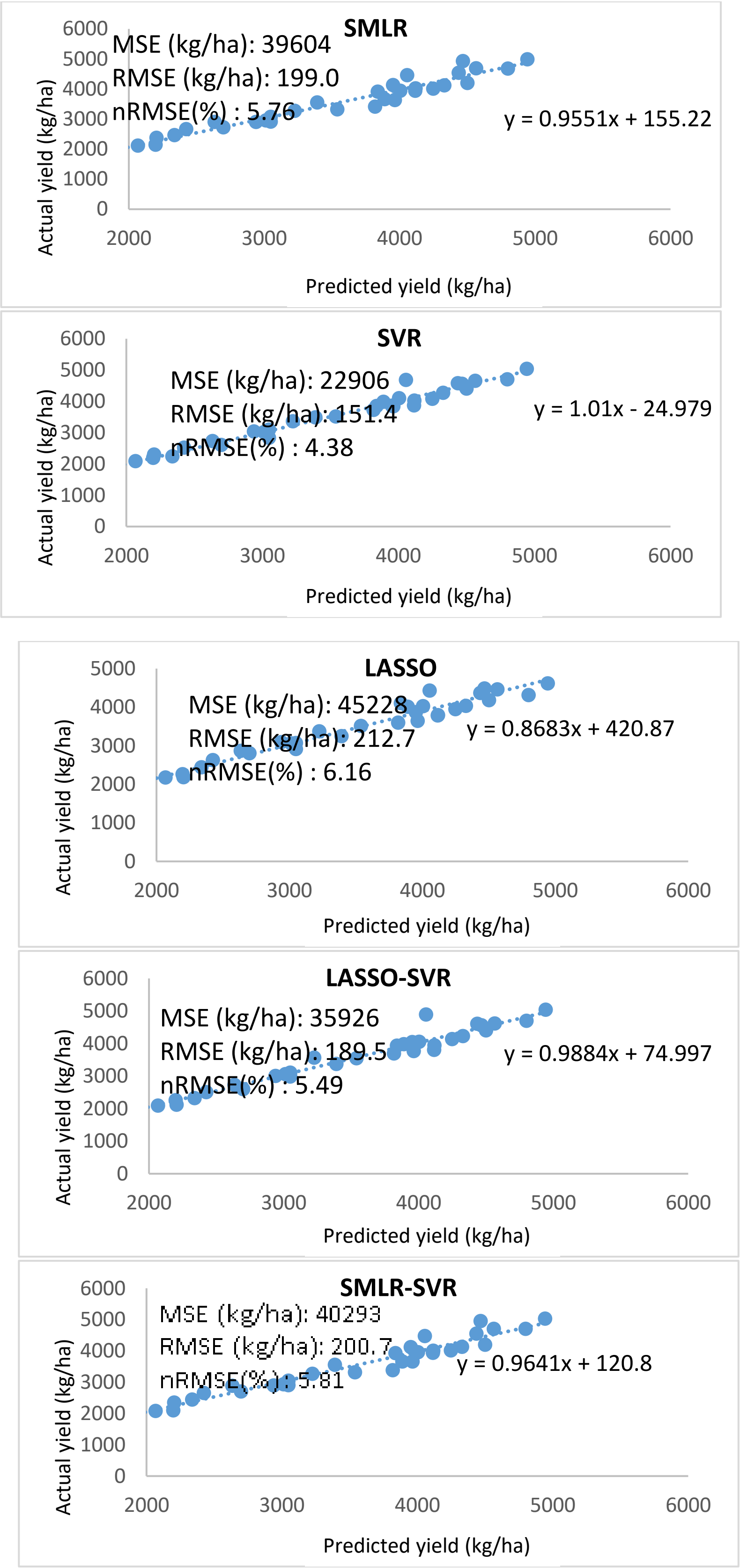


Fig.27.a Calibration of Models developed for wheat yield prediction at grain filling stage for Patiala.

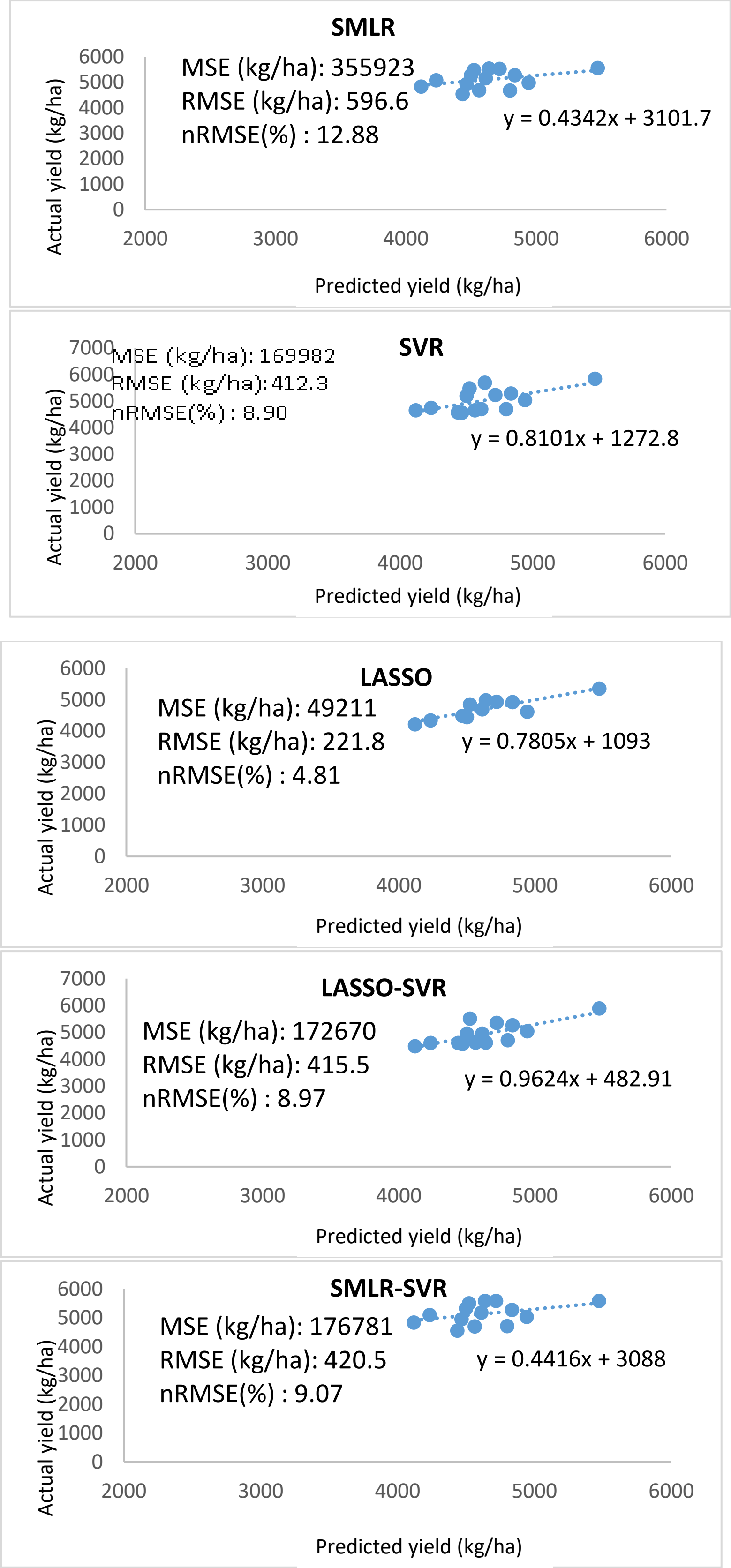


Fig.27.b Validation of Models developed for wheat yield prediction at grain filling stage for Patiala.

Weather variables affect the crop differently during different stages of development.. Weather is dynamic, continuous and multi-dimensional, these unfavourable properties make weather prediction a difficult challenging task for developing the model for crop yield forecast. Thus, there is need to develop crop yield forecast model based on weather variables so that consistent forecasts can be obtained.

Wheat yield prediction using weather indices

Crop physiological process dependent on weather parameters in which temperature is an important weather parameter that affects plant growth, development and yield. Winter crops are vulnerable to high temperature during reproductive stages and differential response of temperature change to various crops has been noticed under different production environments (Kalra, 2008).. For quantifying the thermal relation of crops, thermal units approach is widely used (Ramteke et al., 1996) and has been further modified to include photothermal units and heliothermal units (Rao et al., 1999). The photothermal unit concept provides a reliable index for the progress of the crop that can be used to predict the yield of any crop. with an increase in average seasonal air temperature (17 to 18.6° C) caused reduction in yield (Khichar and Niwas, 2007) by 13 to 26.1 percent in the year 2007-08, whereas, 14.6 to 29.3 percent in 2008-09 (due to increase in average seasonal air temperature from 18.8 to 20.4 °C). . Due to several reasons, the response of crop yields to temperature variations may depend on the relative warming of minimum and maximum temperature (Stone, 2001; Peng *et al*., 2004). In our studies yield prediction model depend upon different weather indices developed by weather parameter maximum, minimum temperature, rainfall, morning, evening relative humidity, bright sun shine hour, GDD, HTU, HUE, PTI and evapotranspiration. Weather parameters have profound influence on wheat yield. Temperature affects the crop yield by changing the rate of photosynthesis, respiration, spikelet sterility and length of growing season (Rai *et al*. 2012; Akinbile *et al*.2015).

Pal et al (2013) used Agro-meteorological indices to predict plant stages and yield of wheat for foot hills of western himalayas. He reported that wheat crop sown on 20th November (normal) required less photothermal unit as well as heliothermal unit, while, 09th January (late) sowing accounted higher values of PTU and HTU during crop growth period. Timely sown wheat crop (20th November) produced highest yield, while, with every 25 days delay in sowings reduction in yield was accounted by 13 to 26.1 percent in the year 2007-08, whereas, 14.6 to 29.3 percent in 2008-09. The productivity of wheat is highly variable and mainly affected by rainfall, temperature and solar radiation. Crop is exposed to a variety of weather conditions during its different phenophases of growth, resulting in large variations in growth rate and yield. Villordon, *et.al, (*2009) observed that GDD method demonstrated high predictive accuracy as shown by mean square error. Bazgir, *et.al, (*2007) did wheat yield prediction for Hamedan district during time period of 2003-04 and 2004-05 using different variables of meteorology with meteorological indices.

Weather based wheat yield prediction models using different techniques

In our study SMLR modal developed for wheat yield modal during validation performed excellent with nRMSE value <10 % for IARI, New Delhi, Hisar, Amritsar and Patiala, good for Ludhiana having nRMSE value 10.54 %. This result is in line with previous findings by Singh *et al.* (2014). He used eighteen years weather data and yield data of rice and wheat for nine districts of Eastern Uttar Pradesh for developing yield prediction equations. He indicated that models explained 51 to 79 percent variations for rice yield and 65 to 92 percent variations for wheat yield in different districts. Garde *et al*., (2015) used discriminant function and multiple linear regression (MLR) techniques for estimating wheat productivity for the district Varanasi in Uttar Pradesh.

In our study SVR during validation performed excellent with nRMSE value <10 % for IARI, New Delhi, Hisar, Amritsar and Patiala, good for Ludhiana having nRMSE value 10.26 %. Parviz (2020) observed that support vector machine (SVM) with capturing the nonlinearity of time series, could improve barley yield estimation, with the minimum UI for yazd province. He showed that SVM is able to have high efficiency to model the climate effect on crop yield. Support vector machines (Vapnik 1998, Schölkopf and Smola 2002) can be used as a very robust and efficient nonlinear multivariate regression tool. The basic idea is what might be a complex non-linear relationship between some variables could be ascribed to a linear relationship in a higher dimensional space. Thus, the variables of interest are projected into a high-dimensional space where linear optimization techniques are applied, and the resulting regression function is then projected back to the original low dimensional phase space of the observed variables. Thus, SVM are now increasingly used as very efficient and effective non-linear regression models. In our study the value of nRMSE of the model developed using LASSO during validation performed excellent having value of nRMSE <10 % for all the location. The selection of different weather parameters as an important variable affecting the wheat yield using LASSO is in line with previous studies (Zhang *et al.*, 2010; Yang *et al.*, 2015; Oguntunde *et al.*, 2018). Tibshirani (1996) proposed the method LASSO for shrinkage and selection for regression and generalized regression problems. He reported that LASSO does not focus on subsets but rather it defines a continuous shrinking operation that can produce coefficient that are exactly to zero.

Based on the value of RMSE and nRMSE during validation, LASSO model performed excellent for all the station and best among all the model developed. Our result is in line with previous findings results reported by (Das *et al*., 2018). He used six model SMLR, PCA-SMLR, ANN, PCA-ANN, LASSO and Elastic Net for prediction of rice yield based on weather parameters for west cost of India.He found that LASSO performed best followed by Elastic Net. Due to the prevention of overfitting of model and reducing the magnitude of regression coefficient with feature selection by penalization decreases the model complexity, LASSO showed good performance for different selected locations. These penalised models give better computational advantage over SMLR as the features with zero coefficients can be eliminated from the model. The feature selection algorithms like LASSO performed better than methods utilising all the weather indices as feature selection reduces over fitting and avoids multicollinearity present in the dataset. He suggested that LASSO is worthy competitor to subset selection and ridge regression. Kumar *et al.* (2019) evaluate the performance of stepwise and Lasso regression technique in variable selection and development of wheat forecast model for crop yield using weather data and wheat yield for the period of 1984-2015 for IARI, New Delhi. He reported that statistical parameters viz. R^2, RMSE, and MAPE were 0.81, 195.90 and 4.54 per cent respectively with stepwise regression and 0.95, 99.27, 2.7 percentage, respectively with Lasso regression. Forecast models were validated during 2013-14 and 2014-15. Prediction errors were -8.5 and 10.1 per cent with stepwise and 1.89 and 1.64 percent with the Lasso. He showed that performance of Lasso regression is better than stepwise regression. In our study LASSO-SVR during validation performed excellent having nRMSE <10 % for all the location, however SMLR-SVR modal performed excellent for IARI, New Delhi, Amritsar and Patiala having nRMSE < 10 %, good for Hisar and Ludhiana with nRMSE value 11.67 % for Hisar and 11.39 % for Ludhiana respectively. The model performance of SVR is increased if hybrid modal LASSO-SVR and SMLR-SVR is used. However, hybrid LASSO-SVR model performed better than SMLR-SVR for wheat yield prediction. Alam *et.al, (*2018) used ARIMAX and two hybrid approaches like ARIMAX-ANN and ARIMAX-SVM for the rice yield forecast using weather variables of Aligarh in Uttar Pradesh. Based on the results obtained, performance of ARIMAX-SVM and ARIMAX-ANN models are close to each other but much superior to the conventional ARIMAX model for the considered data set. He observed that performance of hybrid ARIMAX model was found to be quite encouraging. He concluded that on the basis of significant reduction in MAPEs application of hybrid ARIMAX using ANN and SVM is recommended.

Multi stage crop yield estimation using weather-based model.

In our study wheat yield prediction done at different crop stage by using different weather based model had percentage deviation between 0.62 to 12.07 % by LASSO, 3.94 to 10.15 % by SVR, 1.39 to 11.21 % by SMLR, -4.71 to 14.97 % by LASSO-SVR, 2.31 to 20.62 by SMLR-SVR during tillering stage, 0.28 to 14.44 % by LASSO, 0.07 to 12.87 % by LASSO-SVR, -2.05 to -10.18 % by SVR, -3.60 to 17.54 by SMLR-SVR and 6.38 to 9.22 by SMLR during flowering and -0.02 to 6.09 % by LASSO, -0.67 to -12.11 % by SVR, -1.03 to 12.81 % by LASSO-SVR, -2.24 to 10.12 % by SMLR-SVR and 5.36 to 12.65% by SMLR during grain filling stage respectively. Dutta *et al*. (2001) had developed district wise yield model for rice in Bihar using meteorological data and concluded that models were able to predict pre-harvest crop yield with good accuracy. Garde *et al*., (2015) concluded that stepwise multiple linear techniques can be used effectively for the pre-harvest wheat crop, which are more consistent in performance. Incorporating statistical indicators enhanced the precision of forecasting of wheat crop yield for both Adjusted R^2 and RMSE values. Similar work has also been done by several researchers. Jain *et al*., (1980) did rice yield forecast, developed pre-harvest model only after about two months of sowing. Agrawal *et al*., (2001) reported that reliable forecasting of wheat yield could be obtained when the crops were at twelve weeks i.e. only about two months before harvest. This study was conducted on Vindhyanchal Plateau of Madhya Pradesh for wheat yield forecast. Agrawal *et al*., (2012) used discriminant function analysis intended for developing wheat yield forecasting models for Kanpur. This methodology gave very trustworthy yield forecast about two months beforehand harvest. Vashisth et. al. (2018) reported that percentage deviation of estimated yield by actual yield of maize crop done at flowering stage and at grain filling stage was 10.3 and 7.1 % by weather based statistical model. In our study based on percentage deviation of estimated yield by observed yield done at tillering, flowering and grain filling stage by SVR model was ranged between 1.03 to 17.54 %. Parviz (2020) observed that minimum correlation coefficient between the observed and simulated yield for barley was found in Gilan province using Support vector machine.

LASSO modal performed excellent for yield prediction done at tillering, flowering and grain filling stage having nRMSE value ranged between 0.02 at grain filling stage for IARI, New Delhi to 8.36 % for Hisar at flowering stage. Based on percentage deviation of predicted yield done at tillering, flowering and grain filling stage by observed yield, LASSO model performed best for all the station having value <6 % except for Hisar at tillering and flowering stage it was 12.07 and 14.44 % respectively**.** Kumar *et al* (2019) used Stepwise and LASSO regression variable selection techniques and weather parameters for developing regression forecast model forty-five days before harvest. Based on forecast model results he found that stepwise forecast model over fit, whereas LASSO performs better fit model. Also, the per cent error by LASSO regression model was less than Stepwise regression. He inferred that LASSO variable selection method performed better than stepwise.

The developed weather-based models using different methods showed that results of multi stage in season yield forecasting are closer to observed yield at the pre-harvest stage as compared to flowering and tillering stage. Model developed using different methods using weather parameters had lower value of nRMSE and root mean square error (RMSE) for the yield forecast done by the model at grain filling stage as compared to flowering and tillering stage. This indicates better performance of the model at the grain filling stage. This work is line of the pre-harvest forecast models for several crops based on time series data on crop yield and weekly data on weather variables developed by various research workers (Pandey *et al*., 2014; Azfar *et al*., 2015; Yadav *et al*., 2015). Vasthisth *et al.* (2014) reported that the percentage deviation of observed yield by simulated wheat crop yield forecast done at forty-five days and twenty-five days before harvest using statistical model was less than 10 %.

6. summary and conclusions

Wheat yield is highly affected by the influence of weather variables. Weather variables affect the crop during different stages of development. Thus, there is need to develop and validate weather-based models using machine learning so that reliable prediction of wheat yield can be obtained at multiple stages.

For the development of district wise wheat yield prediction models, long term weather and wheat yield data were collected from IARI, New Delhi, Hisar, Amritsar, Ludhiana and Patiala station in the North-West India. Model was developed for wheat crop yield prediction using SMLR, SVR, LASSO, LASSO-SVR and SMLR-SVR techniques for all the five station. Wheat yield estimation was done at tillering, flowering and grain filling stage by weather-based model developed using SMLR, SVR, LASSO, LASSO-SVR and SMLR-SVR techniques for all the five station. Salient findings of our study are given below.

- Value of average wheat yield was found highest in Ludhiana (4139.1 kg/ha) followed by Patiala (3705.1 kg/ha), Amritsar (3625.3 kg/ha), IARI, New Delhi (3617.3 kg/ha) and 2859.9 kg /ha for Hisar.
- Value of growing degree days during crop growing period was found lowest for Amritsar (1600.2 ^{0}C) followed by Ludhiana (1691.6 ^{0}C), Hisar (1821.4 ^{0}C), Patiala (1866.0 ^{0}C) and (1953.5^{0}C) for IARI, New Delhi.
- Heat use efficiency was found highest in Ludhiana (2.44 kg/ha/^{0}C) followed by Amritsar (2.28 kg/ha/^{0}C), Patiala (1.98 kg/ha/^{0}C), IARI, New Delhi (1.85 kg/ha/^{0}C) and 1.59 kg/ha/^{0}C for Hisar.
- Photo thermal unit was lowest for Amritsar (10.4 °C /hour) followed by Ludhiana (10.9°C /hour), Hisar (11.8 °C /hour), Patiala (12.0°C /hour) and 12.6 °C /hour for IARI, New Delhi.
- Reference evaptranspiration during crop growing period was lowest for Patiala (405.2 mm) followed by Hisar (413.2 mm), Amritsar (418.6 mm), Ludhiana (449.6 mm) and 478.8 mm for IARI, New Delhi.
- All model developed for wheat crop prediction for IARI, New Delhi performed excellent. Model developed by LASSO techniques performed best having

lowest value of nRMSE (1.15%) followed by SMLR-SVR (1.54%), LASSO-SVR (1.88%)), SVR (2.39%) and SMLR (5.11%) techniques

- Model developed for wheat predictions for Hisar using different techniques were excellent except for SMLR-SVR it performed good. LASSO performed best having lowest nRMSE value 2.82% followed by LASSO-SVR (3.34%), SVR (8.57%), SMLR (9.06%) and SMLR-SVR (11.65%) respectively.
- All model developed for wheat predictions for Amritsar performed excellent. Modal developed by LASSO techniques performed best having lowest nRMSE value 1.81% followed by LASSO-SVR (3.67%), SMLR-SVR (4.31%), SMLR (4.68 %) and SVR (5.70%) respectively.
- For Ludhiana model developed for wheat predictions were excellent by LASSO and LASSO-SVR and good for SVR, SMLR and SMLR-SVR techniques. Modal developed by LASSO performed best having lowest nRMSE value 4.83 % followed by LASSO-SVR (9.87%), SVR (10.26%), SMLR (10.54 %) and SMLR-SVR (11.39%) respectively.
- For Patiala model developed by all techniques performed excellent. Modal developed by LASSO and SMLR performed best having lowest nRMSE value 0.65 % followed by SMLR-SVR (2.35 %), SVR (6.12%) and LASSO-SVR (9.09%) respectively.
- Based on the value of RMSE and nRMSE during validation, LASSO model performed excellent for all the station and best among all the model developed.
- Performance of wheat yield prediction for IARI, New Delhi done at tillering stage was excellent for all the models having nRMSE < 10. The percentage deviation of yield estimation done at tillering stage for IARI, New Delhi by observed yield was lowest for SMLR-SVR (-3.12 %) model followed by LASSO-SVR (-4.71%), LASSO (-5.33 %), SMLR (-6.62 %) and SVR (-7.41%) respectively.
- For Hisar performance of wheat yield prediction done at tillering stage was excellent for LASSO, good for SVR and LASSO-SVR, fair for SMLR-SVR and SMLR. The percentage deviation of yield estimation done at tillering stage for Hisar by observed yield was lowest for SMLR (7.58%) model followed by SVR (10.15), LASSO (12.07%), LASSO-SVR (14.97 %), SMLR-SVR (20.62%) and modal respectively.

- Performance of wheat yield prediction for Ludhiana done at tillering stage was excellent for LASSO, LASSO-SVR and SMLR good for SVR and SMLR-SVR. Percentage deviation of yield estimation done at tillering stage for Ludhiana by observed yield was lowest for LASSO (0.62%) model followed by SMLR (5.23%), LASSO-SVR (9.09 %), SVR (9.59%) and SMLR-SVR (10.33%) modal respectively.
- For Patiala performance of wheat yield prediction done at tillering stage was excellent for LASSO, SVR and LASSO-SVR, good for SMLR and SMLR-SVR. Percentage deviation of yield estimation done at tillering stage for Patiala by observed yield was lowest for SMLR (1.39%) followed by SMLR-SVR (2.31%) SVR (3.94 %), LASSO (-5.36%) and LASSO- SVR (8.02%) respectively.
- Performance of wheat yield prediction done at flowering stage for IARI, New Delhi was excellent for all models. The percentage deviation of yield estimation done at flowering stage for IARI, New Delhi by observed yield was lowest for LASSO-SVR (-0.85 %) followed by LASSO (-0.91 %), SVR (-2.05%), SMLR-SVR (-3.60%) and SMLR (-7.56%) respectively.
- For Hisar, performance of wheat yield prediction done at flowering stage was excellent for LASSO, good for SVR, LASSO-SVR and SMLR-SVR, fair for SMLR. Percentage deviation of yield estimation done at flowering stage for Hisar by observed yield was lowest for SMLR (6.38%), model followed by SVR (9.94) LASSO-SVR (12.87 %), LASSO (14.44 %) and SMLR-SVR (17.54%) modal respectively.
- Performance of wheat yield prediction done at flowering stage for Amritsar was excellent for LASSO, LASSO-SVR and SVR model, good for SMLR-SVR and SMLR. Percentage deviation of yield estimation done at flowering stage for Amritsar by observed yield was lowest for LASSO (2.93 %) model followed by LASSO-SVR (-5.36%), SMLR-SVR (-9.02%), SMLR (9.22%) and SVR (-10.18%) modal respectively.
- For Ludhiana, performance of wheat yield prediction done at flowering stage was excellent for LASSO and LASSO-SVR models, good for SVR, SMLR and SMLR-SVR model. Percentage deviation of yield estimation done at flowering

stage for Ludhiana by observed yield was lowest for SVR (0.07 %) model followed by LASSO (4.43 %), LASSO-SVR (6.99%), SMLR (7.31%), and SMLR-SVR (9.86%) modal respectively.

- Performance of wheat yield prediction for Patiala done at flowering stage was excellent for all models except good for SMLR model. Percentage deviation of yield estimation done at flowering stage for Patiala by observed yield was lowest for LASSO (0.28 %) model followed by LASSO -SVR (5.87 %), SMLR (7.29%), SVR (7.32%), and SMLR-SVR (7.67%) modal respectively.
- Wheat yield prediction done at grain filling stage for IARI, New Delhi, all models performed excellent. The percentage deviation of yield estimation done at grain filling stage for IARI, New Delhi by observed yield was lowest for LASSO (-0.02 %) model followed by SVR (-0.67%), LASSO -SVR (-1.03 %), SMLR-SVR (-2.24%), and SMLR (-7.60) modal respectively.
- Performance of wheat yield prediction done at grain filling stage for Hisar was excellent for LASSO, good for SMLR-SVR, LASSO-SVR and SVR model and fair for SMLR model. Percentage deviation of yield estimation done at grain filling stage for Hisar by observed yield was lowest for SVR (-4.48%), model followed by SMLR (5.36 %), LASSO (6.09%), SMLR-SVR (9.94) and LASSO -SVR (12.81%) modal respectively.
- For Amritsar wheat yield prediction at grain filling stage was excellent for LASSO, LASSO-SVR, SMLR-SVR model, good for SVR and SMLR model. Percentage deviation of yield estimation done at grain filling stage by observed yield for Amritsar was lowest for LASSO (-0.29 %) model followed by SMLR-SVR (-2.82%), LASSO -SVR (-5.90%), SVR (-12.11%),) and SMLR (12.65) modal respectively.
- Wheat yield prediction done at grain filling stage for Ludhiana was excellent for LASSO and LASSO-SVR, good for SVR, SMLR-SVR and SMLR model. Percentage deviation of yield estimation done at grain filling stage for Ludhiana by observed yield was lowest for SVR (-1.70 %) model followed by LASSO (4.48 %), SMLR (6.23), LASSO-SVR (8.23%) and SMLR-SVR (10.12%) modal respectively.
- Performance of wheat yield prediction done at grain filling stage for Patiala was excellent for all the model except good for SMLR model. Percentage deviation

of yield estimation done at grain filling stage by observed yield for Patiala was lowest for SMLR (1.74) model followed by LASSO (-2.03 %), SMLR-SVR (2.06%), SVR (6.71%) and LASSO -SVR (7.75%) modal respectively.

- Based on model performance and percentage deviation of estimated yield by observed yield LASSO model performed best for yield prediction done at tillering, flowering and grain filling stage for all the station having value <6 % except for Hisar at tillering and flowering percentage deviation was 12.07 and 14.44 % respectively.
- Out of five models developed, based on model performance and percentage deviation, LASSO models were found to be excellent in estimating wheat yield for selected locations.
- The model performance of SVR is increased if hybrid machine learning (in combination with LASSO and SMLR) is applied.
- Combination of SVR with LASSO i.e. hybrid machine learning with LASSO-SVR has shown more improvement in SVR model compared with hybrid machine learning SMLR-SVR

ABSTRACT

Wheat crop is highly affected by the influence of weather parameter, adverse weather causes drastically reduced in wheat yield. Thus, there is need to develop and validate weather-based models using machine learning so that reliable prediction of wheat yield can be obtained at multiple stages. Wheat yield and weather data during crop growing period (46th to 15th SMW) for last 35 years were collected from IARI, New Delhi, Hisar, Amritsar, Ludhiana and Patiala. The wheat yield prediction model was developed using stepwise multi linear regression (SMLR), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), variable section by LASSO and SVR (LASSO-SVR), variable section by SMLR and SVR (SMLR-SVR) techniques in R software. Analysis were carried out by fixing 70% of the data for calibration and remaining dataset for validation. Weather indices, GDD, HTU, HUE, PTI and reference evapotranspiration were calculated for each station during crop growing period. There indices along with time were used as input for developing yield production modal. Results showed that based on the value of RMSE and nRMSE during validation, LASSO model performed excellent for all the station and best among all the model developed. During validation nRMSE value was between 0.65 to 4.83 % for LASSO, 1.88 to 9.87% for LASSO-SVR, 2.39 to 10.26% for SVR, 0.65 to 10.54% for SMLR and 1.54 to 11.65 % for SMLR-SVR respectively. Multistage wheat yield prediction was done at tillering, flowering and grain filling stage for wheat crop by considering 46th to 4th, 46th to 8th and 46th to 11th SMW for model development. On examining these models for stage-wise prediction of wheat yield, percentage deviation was found between 0.62 to 12.07 % by LASSO, 3.94 to 10.15 % by SVR, 1.39 to 11.21 % by SMLR, -4.71 to 14.97 % by LASSO-SVR, 2.31 to 20.62 by SMLR-SVR during tillering stage, 0.28 to 14.44 % by LASSO, 0.07 to 12.87 % by LASSO-SVR, -2.05 to -10.18 % by SVR, -3.60 to 17.54 by SMLR-SVR and 6.38 to 9.22 by SMLR during flowering and -0.02 to 6.09 % by LASSO, -0.67 to -12.11 by SVR, -1.03 to 12.81 % by LASSO-SVR, -2.24 to 10.12 % by SMLR-SVR and 5.36 to 12.65 by SMLR during grain filling stage respectively. LASSO modal performed excellent for yield prediction done at tillering, flowering and grain filling stage having nRMSE value ranged between 0.02 at grain filling stage for IARI, New Delhi to 8.36 % for Hisar at flowering stage. Based on percentage deviation of predicted yield done at tillering, flowering and grain filling stage by observed yield, LASSO model performed best for all the station having value <6 % except for Hisar at tillering and flowering stage it was 12.07 and 14.44 % respectively. The model performance of SVR is increased if hybrid model in combination with LASSO and SMLR is applied. Hybrid

model LASSO-SVR has shown more improvement in SVR model compared with SMLR-SVR. From this study it may be concluded that LASSO model based on weather parameters can be used for district level wheat yield forecast at different crop growth stage.

मशीन लर्निंग का उपयोग करके मौसम आधारित गेहूं के उपज का पूर्वानुमान

गेहूं की फसल मौसम की विविधता के प्रभाव से अत्यधिक प्रभावित होती है, प्रतिकूल मौसम के कारण गेहूं की पैदावार में भारी कमी आती है। इसलिए, मौसम आधारित मॉडल को विकसित और सत्यापन करने की आवश्यकता है ताकि गेहूं की उपज का विश्वसनीय पूर्वानुमान कई चरणों में प्राप्त किया जा सके | भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली, हिसार, अमृतसर, लुधियाना और पटियाला से पिछले 35 वर्षों के फसल की बढ़ती अवधि (46 वें से 15 वें मानक मौसमीय सप्ताह) के दौरान मौसमीय आँकड़े तथा गेहूं की उपज के आंकड़े एकत्र किए गए। गेहूं की उपज के पूर्वानुमान मॉडल को स्टेपवाइज मल्टी लीनियर रिग्रेशन (एसएमएलआर), सपोर्ट वेक्टर रिग्रेशन (एसवीआर), लासो, लासो- एसवीआर और एसएमएलआर -एसवीआर तकनीकी का उपयोग करके आर सॉफ्टवेयर में विकसित किया गया। अंशांकन के लिए 70% तथा सत्यापन के लिए शेष डेटा का प्रयोग करके विश्लेषण किया गया। फसल अवधि के दौरान प्रत्येक स्टेशन के लिए मौसम सूचकांकों, जी डी डी, एच टी यु, एच यु ई, पी टी आई और संदर्भ वाष्पीकरण की गणना की गई। उपज उत्पादन मॉडल के विकास के लिए इनपुट के रूप में समय के साथ इन सूचकांकों का उपयोग किया गया। परिणामों से पता चला कि सत्यापन के दौरान आरएमएसई और एनआरएमएसई के मूल्य के आधार पर, लासो मॉडल सभी स्टेशन के लिए उत्कृष्ट है और सभी विकसित मॉडल में सबसे अच्छा है। सत्यापन के दौरान एन आर एम एसई की सीमा लासो के लिए 0.65 से 4.83% के बीच, लासो -एसवीआर के लिए 1.88 से 9.87% के बीच, एसवीआर के लिए 2.39 से 10.26% के बीच, एसएमएलआर के लिए 0.65 से 10.54% के बीच और एसएमएलआर -एसवीआर के लिए 1.54 से 11.65% के बीच क्रमश: रही।

मॉडल विकास के लिए 46 वें से 4 वें, 46 वें से 8 वें और 46 वें से 11 वें मानक मौसमीय सप्ताह लेकर फसल के कल्ले फूटने, फूल और अनाज भरने पर गेहूं की उपज का पूर्वानुमान किया गया | इन मॉडल की जांच करने पर यह पाया गया कि गेहूं की पैदावार की विभिन्न चरणों के दौरान पूर्वानुमान तथा वास्तविक उपज में प्रतिशत विचलन कल्ले फूटने के दौरान लासो द्वारा 0.62 से 12.07% के बीच, एसवीआर द्वारा 3.94 से 10.15% के बीच, एसएमएलआर द्वारा 1.39 से 11.21% के बीच, लासो -एसवीआर द्वारा -4.71 से 14.97% % के बीच, एसएलआरआर-एसवीआर द्वारा 2.31 से 20.62 के बीच

क्रमश: रही । फूल खिलने के दौरान गेहूं की पैदावार के पूर्वानुमान तथा वास्तविक उपज में प्रतिशत विचलन लासो द्वारा 0.28 से 14.44% के बीच, लासो -एसवीआर द्वारा 0.07 से 12.87% के बीच, एसवीआर द्वारा -2.05 से -10.18% के बीच, एसएमएलआर-एसवीआर द्वारा -3.60 से 17.54 के बीच और एसएमएलआर द्वारा 6.38 से 9.22 के बीच क्रमश: रही । अनाज भरने के दौरान गेहूं की पैदावार के पूर्वानुमान तथा वास्तविक उपज में प्रतिशत विचलन लासो द्वारा -0.02 से 6.09% के बीच, एसवीआर द्वारा -0.67 से -12.11 के बीच, लासो -एसवीआर द्वारा -1.03 से 12.81% के बीच, एसएमएलआर-एसवीआर द्वारा -2.24 से 10.12% के बीच, एसएमएलआर द्वारा 5.36 से 12.65 के बीच क्रमश: रही । कल्ले फूटने, फूल खिलने तथा अनाज भरने के दौरान दिए गये पूर्वानुमान में लासो मॉडल की सटीकता अन्य मॉडलो की तुलना में बेहतर पाई गई। एन आर एम एसई की सीमा कल्ले फूटने, फूल खिलने तथा अनाज भरने के दौरान 0.02 से 8.36% के बीच, भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली के लिए अनाज भरने के दौरान तथा हिसार के लिए फूल खिलने के दौरान क्रमश:रही । कल्ले फूटने, फूल खिलने तथा अनाज भरने के दौरान दिए गये पूर्वानुमान का वास्तविक उपज से प्रतिशत विचलन के आधार पर, लासो मॉडल अन्य मॉडलो की तुलना में बेहतर पाया गया, प्रतिशत विचलन का मान सभी स्टेशन के लिए 6% से कम पाया गया केवल हिसार के लिए कल्ले फूटने के दौरान 12.07 % तथा फूल खिलने के दौरान 14.44% क्रमशः रही । एसवीआर मॉडल को यदि लासो और एसएमएलआर के साथ संयोजन में हाइब्रिड मॉडल बनाया जाता है तो एसवीआर मॉडल की सटीकता बेहतर हो जाती है | लासो -एसवीआर हाइब्रिड मॉडल में एसएमएलआर-एसवीआर हाइब्रिड मॉडल की तुलना में एसवीआर मॉडल में अधिक सुधार पाया गया । इस अध्ययन से यह निष्कर्ष निकाला जा सकता है कि मौसम के मानकों के आधार पर लासो मॉडल का उपयोग विभिन्न फसल विकास के स्तर पर जिला स्तर पर गेहूं की उपज के पूर्वानुमान के लिए किया जा सकता है।

Agrawal, R. and Jain, R.C. (1982). Composite model for forecasting rice yield, *Indian Journal of Agricultural Sciences,* 52(3): 189-194

Agrawal, R., Jain, R.C., Jha, M. P. and Singh, D. (1980): Forecasting of rice yield using climatic variables, *Indian Journal of Agricultural Sciences,* 50(9):680-684.

Agrawal, R., Jain, R.C. and Jha, M. P. (1983). Joint effects of weather variables on rice yield, *Mausam*, 37(1):67-70.

Agrawal R., Chandrahas and Kaustav A. (2012). Use of discriminant function analysis for forecasting crop yield, *Mausam,* 63, 3 (July 2012), 455-458

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-c climatic zone basis. *Indian Journal of Agricultural Sciences* 71(7):487-490

Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pests and diseases-IASRI models. *J. Indian.Society. Agricultural. Statistics.,* 61(2): 255-263.

Akinbile, C. O., Akinlade, G. M., & Abolude, A. T. (2015). Trend analysis in climatic variables and impacts on rice yield in Nigeria. *Journal of Water and Climate Change*, *6*(3), 534-543.

Alam, W., Ray, M., Kumar, R. R., Sinha, K., Rathod, S., and Singh, K. N. (2018). Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables. *Indian Journal of Agricultural Sciences, 88*(12), 101-105.

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

Baier, w.(1981). Crop-weather analysis models.In: Applicaions of remote sensing to agricultural production forecasting (ed breg A.), Joint research Centre, ISPRA, Italy. PP105-118.

Basso,B., Cammarano, D. and Carfagna, E. (2013). Review of crop yield forecasting methods and early warning systems. In: Intergovernmental panel on climate change (ed) climate change 2013 - the physical science basis. Cambridge University Press, Cambridge, pp 1–30.

Bazgir, S., Kamali, G., and Mortazavi, A. (2007). Wheat yield prediction through agrometeorological indices for Hamedan, Iran.

Benbi, D.K. (1994). Prediction of leaf area indices and yield of wheat. *Journal of Agricultural Sciences Cambridge* **122**: 13-20.

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

Bonhomme, R. (2000). Beware of comparing RUE values calculated from PAR vs. solar radiation or absorbed vs. intercepted radiation. *Field Crops Research* **68**: 247-252.

Box, G.E.P. and Jenkins, G. (1970). Time series analysis, Forecasting and control, Holden-Day, San Francisco, CA.

Bray, M., and Han, D. (2004). Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics*, *6*(4), 265-280.

Chauhan, V.S., Shekh, A.M., Dixit ,S.K., Mishra, A.P. and Kumar, S. (2009), Yield prediction model of rice in Bulsar district of Gujarat, *Journal of. Agrometeorology*, **11 (2):** 162-168.

D, Alexandridis, T., Whetton, R. L., and Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, *121,* 57-65.

Dahikar, S. S., and Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2*(1), 683-686.

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 Indian Society Remote Sensing* 29:175–181

Das, B., Nair, B., Reddy,K.V. and Venkatesh, P.(2018). Evaluation of multiple linear, network and penalised regression models for prediction of rice yield based on weather parameters foe west cost of India. *Int J Biometerol* 1583-6

Dhekale, B.S., Mahdi SS, Dalvi TP, and Sawant PK (2014) . Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. *Journal of Agrometeorology* 16:238–239

Dwyer, L.M. and Stewart, D.W. (1986). Leaf area development in field grown maize. *Journal of Agronomy* **78**: 334-348.

Ekaputa, U.N. (2004). *Agricultural systems*. Port Harcourt: UNIK Publishers. pp. 126-134.

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

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.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S.,Chattopadhyay, N., Singh, K. K. and Rathore, L. S.(2014). Development of crop yield forecast models under FASAL – a case study of *kharif* rice in West Bengal. *Journal of Agrometeorology*, 16 (1): 1-8.

Girijesh, G.K., Kumara wamy, A.S., Sridhara, S., Dinesh Kumar, M., Vageesh, T.S and Nataraju, S.P. (2011). Heat use efficiency and helio-thermal units for maize genotypes as influenced by dates of sowing under southern transitional zone of karnataka state. *International Journal of Science and Nature* **2**(3): 529-533.

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 Agric. Exp.Sta. Tech.Bull.*, 74.

Hundal, S.S., Singh, R. and Dhaliwal, L.K. (1997). Agro-climatic indices for predicting phenology of wheat (*Triticum aestivum*) in Punjab. *Indian Journal of Agricultural Sciences* **67** (6): 265-8.

Hundal, S.S. Kaur, P. and Malikpuri, S.D.S. (2003a). Agro-climatic models for prediction of growth and yield of Indian mustard (*Brassica juncea*). *Indian journal of agricultural science*.**73** (3): 142-144.

Hundal, S.S., Singh, H., Prabhjyot, K. and Dhaliwal, L.K. (2003b). Agro-climatic models for growth and yield of soybean (*Glycine max.*). *Indian Journal of Agricultural Sciences.* **73**(12): 668-70.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University, 37.*

Jain, R.C., Agrawal, R. and Jha, M.P. (1980). Effect of climatic variables on rice yield and its forecast. *Mausam,*31(4):591-596

Ji, B., Sun Y, Yang S and Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. J Agric Sci 145:249–261

Jin, M., & Jin, C. (2008, September). Forecasting agricultural production via generalized regression neural network. In *2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE)* (pp. 1-3). IEEE.

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., and Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and forest meteorology, 218,* 74-84.Pantazi, X. E., Moshou,

Jain, R.C., Agrawal, R. and Jha, M.P. (1980). Effect of climatic variables on rice yield and its forecast. *Mausam,*31(4):591-596

Jain, R. C., Agrawal, R., and Singh, K. N. (1992). A within year growth model for crop yield forecasting. *Biometrical journal*, *34*(7), 789-799.

Jamieson, P.D., Porter,J.R. and Wilson, D.R .(1991). A test of the computer simulation and technology, 49, 32-35.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*.

Kandiannan, K., K. K. Chandaragiri, N. Sankaran, T. N. Balasubramanian and C. Kailasam (2002). Crop-weather model for turmeric yield forecasting for Coimbatore District, Tamil Nadu, India. *Agricultural and Forest Meteorology*, **112**, 133- 137.

Kaur, R. and Rajni (2010). Impact of increasing CO_2 on crop productivity. *Indian Farming*. June issue. 23-7.

Kalra, N., Chakraborty, D., Sharma, A., Rai, H. K., Jolly, M., Chander, S. and Lal, M. (2008). Effect of increasing temperature on yield of some winter crops in northwest India. *Current science*, 82-88.

Khichar, M. L., and Niwas, R. (2007). Thermal effect on growth and yield of wheat under different sowing environments and planting systems. *Indian Journal of Agricultural Research*, *41*(2), 92-96.

Kumar, N., Pisal, R., Shukla, S.P. and Pandye, K.K. (2014) Regression technique for South Gujarat. *Mausam* ,65:361–364

Kumar R., Gupta B.R.D., Athiyaman B., Singh K.K. and Shukla R.K. (1999).Stepwise regression technique to predict pigeonpea yield in Varanasi district, *Journal of Agrometeorology*, **1(2):** 183-186.

Kumar, S., Attri, S. D., and Singh, K. K. (2019). Comparison of Lasso and stepwise regression technique for wheat yield prediction. *Journal of Agrometeorology*, *21*(2), 188-192.

Kumar, S., Attri, S. D., & Singh, K. K. (2019). Comparison of Lasso and stepwise regression technique for wheat yield prediction. *Journal of Agrometeorology*, *21*(2), 188-192.

Kumari, Prity., Mishra, G. C., Pant, A. K., Shukla, G. and Kujur, S. N. 2014b. Comparison of forecasting ability of different statistical models for productivity of rice (oryza Sativa l.) in India. The Ecoscan. 8(3&4): 193 -198.

Lee, J. H., and Ha, S. H. (2009). Recognizing yield patterns through hybrid applications of machine learning techniques. *Information Sciences, 179*(6), 844-850.

Lee, Byoung-Hoon, Phil Kenkelb and Wade Brorsenb, B. (2013). Pre-harvest forecasting of county wheat yield and wheat quality using weather information. *Agricultural and Forest Meteorology*, 168, 26– 35

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

Lobell DB, Schlenk er W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* 333:616–620.

Maduakor, H.O., Lal, R. and Opara-Nadi, O.A. (1984). Effects of methods of seedbed preparation and mulch on the growth and yield of white yam (*Dioscorea rotundata* L.) on an ultisol in south -east Nigeria. *Field Crop Research* **9**: 119-130.

Masle, J., Doussinault, G., Farquhar, G.D. and Sun, B. (1989). Foliar stage in wheat correlates better to photothermal time than to thermal time. *Plant Cell Environ.***12**:235-247.

Nagai, T., and Makino, A. (2009). Differences between rice and wheat in temperature responses of photosynthesis and plant growth. *Plant and Cell Physiology*, *50*(4), 744-755.

Narkhede, U. P., and Adhiya, K. P. (2014). Evaluation of Modified K-Means Clustering Algorithm in Crop Prediction. *International Journal of Advanced Computer Research*, *4*(3), 799.

Oguntunde, P. G., Lischeid, G., & Dietrich, O. (2018). Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. *International journal of biometeorology*, *62*(3), 459-469.

Okoh, C.A. (2004). The effect of mulching on soil physico-chemical properties and the yield of white yam. *Tropical Journal of Root Tuber Crops* **4**(2): 24-31.

Okori, W., and Obua, J. (2011, July). Machine learning classification technique for famine prediction. In *Proceedings of the world congress on engineering* (Vol. 2, pp. 991-996).

Orkwor, G.C. and Ekanayake, I.J. (1998). Growth and Development. In: Orkwor, G.C., Asiedu, R. and Ekanayake, I.J. (Eds).*The progress in yam Research*. Ibadan, Nigeria: IITA. pp. 39-62.

Osborne, T., Rose, G., and Wheeler, T. (2013). Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agricultural and Forest Meteorology*, *170*, 183-194.

Pandey, K.K., Rai, V.N. and Sisodia, B.V.S. (2014). Weather variable based rice yield forecasting models for Faizabad district of eastern Uttar Pradesh. *Internat. J. Agric. &*

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 Arch* 15: 575–579.

Parviz, L. (2019, June). Assessing accuracy of barley yield forecasting with integration of climate variables and support vector regression. In *Annales Universitatis Mariae Curie-Sklodowska, sectio C–Biologia* (Vol. 73, No. 1, pp. 19-30).

Peng, S., Huang, J., Sheehy, J.E., Laza, R.C., Visperas, R.M., Zhong, X., Centeno, G.S., Khush, G.S., Cassman, K.G., (2004). Rice yields decline with higher night temperature from global warming. Proc. *National. Academy of sciences*. U.S.A. 101 (27), 9971–9975.

Phadnawis, N.B. and Saini, A.D. (1992). Yield models in wheat based on sowing time and phonological development. *Annals of Plant Physiology* **6**: 52-59.

Rai, A., Joshi, M. K., & Pandey, A. C. (2012). Variations in diurnal temperature range over India: under global warming scenario. *Journal of Geophysical Research: Atmospheres*, *117*(D2).

Rai, K.K., N.P.V., Bharti, B.V.S. and SA K (2013) Pre -harvest forecast models based on weather variable. Adv Biores 4:118–122

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.

Ramakrishna, Y.S., Singh, H. P. and Nageswara, R.G. (2003). Weather based indices for forecasting food grain production in India. *Journal of. Agrometeorology*,5(1):1-11.

Ramteke, S. D., Chetti, M. B., and Salimath, P. M. (1996). Heat unit requirement of chickpea genotypes for various phenological stages during kharif and rabi seasons. *Annals of Plant Physiology, 10*, 176-181.

Rao, V.U.M., Singh, D. and Singh, R. (1999). Heat use efficiency of winter crops in Haryana. *Journal of Agrometeorology* **1**(2): 143-8.

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

Rao, V.U.M., Singh, D. and Singh, R. (1999). Heat use efficiency of winter crops in Haryana. *Journal of Agrometeorology* **1**(2): 143-8.

Rimi RH, Rahman SH, Karmakar S, Hussain SG (2011) Trend analysis of climate change and investigation on its probable impacts on rice production at Satkhira, Bangladesh. Pak J Meteorol 6:37–50

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 cultivars at different phenological stages under varying environmental conditions. *Indian J. Agric. Sci.* **75**(11): 717-21.

Sarwade,G.S.(1998).Meteorological yield models, Technical Note: IRS-UP/SAC/CYM/TN/17/88, Space Applications Centre, Ahmadabad, India.

Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Shi W, Tao F, Zhang Z (2013) A review on statistical models for identifying climate contributions to crop yields. *J Geogr Sci* 23:567–576.

Singh RS, Patel C, YadavMK, Singh KK (2014) Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *Journal of. Agrometeorology* 16:199– 202

Sikder, S. (2009). Accumulated heat unit and phenology of wheat cultivars as influenced by late sowing heat stress condition. *Journal of Agriculture & Rural Development* **7**(1&2).57-64.

Singh, S. (1996). Growth and yield response of different crop species to low light and high temperature-humidity stress. *Indian Journal of plant Physiology* **2**:151-155.

Slafer, G.A. and Rawson, H.M. (1994). Sensitivity of wheat phasic development to major environmental factors: A re-examination of some assumptions made by physiologists and modellers. *Australian Journal of Plant Physiology* **21**: 393-426.

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

Stone, P. J., D. R. Wilson, J. B. Reid, and R. N. Gillespie. "Water deficit effects on sweet corn. I. Water use, radiation use efficiency, growth, and yield." *Australian Journal of Agricultural Research* 52, no. 1 (2001): 103-113.

Sukhadia, K., and Chaudhari, M. B. (2019). A Survey on Rice Crop Yield Prediction in India Using Improved Classification.

Suresh, K. K., and Krishna Priya, S. R. (2009). A study on pre-harvest forecast of sugarcane yield using climatic variables. *Stat. Appl. 7&8 (1&2)(New Series)*.

Suresh, K.K. and Krishna Priya, S.R.(2011). Forecating Sugarcane Yield of Tamilnandu Using ARIMA Models. *Sugar Tech.,* 13(1): 23-26.

Tabari H, Kisi, O, Ezani, Hosseinzadeh, A. and Talaee, P. (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. Journal of Hydrology. 444–445:78–89

Team, R. C. (2013). R: A language and environment for statistical computing.

Tewari, S.K. and Singh, M. (1993).Yielding ability of wheat at different dates of sowing: a temperature development performance. *Indian Journal of Agronomy* **38**: 204-209.

Tibshirani, R. (1996). Regression shrinkage and selection via lasso**.** *J Roy StatSoc B*, **58:**267–288.

Tripati, P., Tomar, S. K. and Singh, A.K. (1999). Crop weather models to predict the growth and yield of mustard and wheat under mustard-wheat cropping system. In: Proceedings of the National Workshop on Dynamic Crop Simulation Modelling for Agro-met Advisory Services, National Center for Medium range Forecasting, New Delhi. pp. 285

Vapnik, V. (1998). The support vector method of function estimation. In *Nonlinear Modeling* (pp. 55-85). Springer, Boston, MA.

Vashisth, Ananta, Singh R. and Manu Choudary (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* : 1-3

Vashisth, Ananta, Goyal, Avinash and Roy, Debasish (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

Verma ,U., Piepho, H.P. and Goyal, A. et al (2016) .Role of climatic variables and crop condition term for mustard yield prediction in Haryana. *Int J Agric Stat Sci* 12:45–51

Villordon, A., Clark, C., Ferrin, D., and LaBonte, D. (2009). Using growing degree days, agrometeorological variables, linear regression, and data mining methods to help improve prediction of sweetpotato harvest date in Louisiana. *HortTechnology*, *19*(1), 133-144.

Walker, G. K. (1989). Model for operational forecasting of western Canada wheat yield. *Agricultural and Forest Meteorology, 44*(3-4), 339-351.

Wang, J.Y. (1960). A critique of the heat units approach to plant response studies. *Ecology* **41**: 785-90.

Wassmann R, Jagadish S.V.K., Surnfleth, K., Pathak, H., Howell, G., Ismail, A,,Serraj, R., Redoña, E., Singh, R.K.and Heuer, S. (2009) Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. Adv Agron 102:99–133

Williams, G. D. V. (1971). Geographical variations in yield-weather relationships over a large wheat growing region. *Agricultural Meteorology, 9*, 265-283.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wurr, D.C.E., Fellows, J.R. and Phelps, K. (2002). Crop scheduling and prediction – Principles and opportunities with field vegetables. *Advances in Agronomy***76**: 201-234.

Yang, D., Ye, Z., Lim, L. H. I., and Dong, Z. (2015). Very short term irradiance forecasting using the lasso. *Solar Energy, 114*, 314-326.

Yadav, R.R. and Sisodia, B.V.S. (2015). Predictive models for Pigeon-pea yield using weather variables. *Internat. J. Agric. & Stat. Sci.*, **11**(2): 462-472.

Zhang, T., Wang, Z., Yin, Y., Cai, R., Yan, S., and Li, W. (2010). Starch Content and Granule Size Distribution in Grains of Wheat in Relation to Post-Anthesis Water Deficits. *Journal of Agronomy and crop science*, **196** (1), 1-8.