Article

Choosing the Best Model for Crop Yield Prediction by the Means of Regression Analysis on the Example of "Crop Yield – Water Use" Simulations Through Best Subsets Approach

Pavlo Lykhovyd [1,*], Raisa Vozhehova [1], Olena Piliarska [1] and Sergiy Zaiets [1]

[1] Institute of Climate-Smart Agriculture; icsanaas@ukr.net
* Correspondence: pavel.likhovid@gmail.com; Tel.: +380660629897

Abstract: Crop yield prediction is relevant subject of current agricultural science. There are various mathematical approaches to crop yield prediction, and regression analysis, notwithstanding the fact that it is somewhat outdated, is still one of the most used ones in this purpose. The quality of predictive model is of great importance, and it is strongly dependent on the rational choice of the target function. The goal of this study is to find out the best regression model for winter wheat, soybeans, and grain corn yield prediction depending on the crops' water use. The data on true crops' yields and water use were collected within 1970-2020 at the experimental fields of the Institute of Climate-Smart Agriculture, Kherson region, Ukraine. In total, 145 data pairs were processed by the best subsets regression to find out the best model in terms of fitting quality (assessed by the Pearson's coefficient of correlation), and prediction accuracy (assessed by the values of the minimum and maximum absolute errors and mean average percentage error). As a result, it was established that the best fitting quality for all the studied crops is attributed to cubic function, while the best accuracy is recorded for hyperbolic (reverse) function in soybeans (mean absolute percentage error is 12.27%), quadratic and hyperbolic functions in winter wheat (mean absolute percentage error is 20.54%), and cubic function in grain corn (mean absolute percentage error is 14.92%). To sum up the results of the study, it is recommended to apply cubic regression function for modeling crops' yields in agricultural studies.

Keywords: agricultural modeling; fitting quality; function; grain corn; prediction accuracy; soybeans; winter wheat

1. Introduction

Crop modeling is an important branch of modern agricultural science. Current crop modeling is deeply integrated into modern systems of decision support, and includes such directions as yield prediction, phenological modeling, growth and development simulation, natural resources use simulation, environmental impacts modeling and forecasting, agrotechnological modeling, etc. [1]. Among the mentioned above directions, yield prediction is one of the most important one for science and practice. For example, it is extremely useful for food security, because nowadays decisions on food distribution policy are taken with accordance to crop yields forecasts. Accurate predictions are essential for rational trading decisions on the global grain and food market, and they also impact fair distribution of food among the global population [2]. Besides, individual crop producers also rely upon yield predictions in their everyday routine to take rational management decisions on agrotechnology during the process of cultivation to obtain the highest possible yields under the best out pay of technological expenses and the least losses of crops' productivity [2, 3]. Therefore, it is difficult to overestimate the importance of scientifically sound yield predictions in modern agriculture.

At the time, there are different methodological approaches to solving the problem of crops' yield modeling and prediction. Generally, current predictive models could be classified as deterministic and stochastic [4]. Deterministic models are based on a concrete initial input dataset, which must be properly organized and structured, while stochastic models can handle unorganized datasets and the data with some degree of uncertainty. Stochastic models have low practical use, while deterministic ones are the prevalent type of mathematical models in agriculture, and could be subdivided into statistical, mechanistic, and functional. Functional models, as well as mechanistic, often are complex and are based on empirical calculations related to the well-established impacts of agrotechnology or environment on crops. Such models are used in decision support systems, and one of the most known and internationally recognized agricultural functional model is the equation of Penman-Monteith for the reference evapotranspiration assessment [5, 6].

Statistical models are relatively old, but they still are of great demand and importance. These models are based on the results of applying various statistical processing methods to the analysis of the results of field experiments. First statistical models were quite simple and of average or even low accuracy. But as novel, more perfect algorithms appear in mathematical statistics, statistical models in agriculture become more complicated and accurate [4].

The most widespread type of statistical models in agriculture is regression models. They vary depending on the scope of the experimental work, dataset size, number of inputs, purposes of modeling, and utilize different regression analysis approaches, including those from the simplest one represented by linear pairwise regression and ending with the complex multiple fuzzy regression functions [7]. Notwithstanding the fact that regression models are sometimes referred to as not the best choice for agricultural modeling with the advent of deep learning data processing, this statement is controversial, and a great share of modern crops' yields predictions are still successfully made using regression approaches, especially non-linear ones [8, 9]. However, it is stressed that the quality and accuracy of regression modeling is strongly dependent on the rational choice of the target function, and great difference between different regression approaches is claimed and proved in the recent studies [10, 11]. Besides, it also depends on the right choice of the inputs [12]. For example, if we are talking about crops' yield modeling in the South of Ukraine, one of the decisive factors affecting yields of the major crops is water availability, as this region of the country is under high risks of drought and aridity increase [13]. Therefore, it is rational to take this factor, expressed in different ways, as the input.

The main goal of this study is to find out the best approach among common regression analysis techniques, used nowadays in crops' yield modeling, for the yield prediction of major crops, cultivated in the South of Ukraine, by the volumes of their water use. Such models are of great importance for sustainable crop production in the region, as it is one that belongs to the area of highly risky agriculture because of constant lack of natural water supply [14].

2. Materials and Methods

The study is based on the retrospective statistical data on the yields of winter wheat, grain corn and soybeans, cultivated in the irrigated and non-irrigated experimental fields of the Institute of Climate-Smart Agriculture (former Institute of Irrigated Agriculture) in the period 1970-2020. The dataset for winter wheat included 45 "yield – water use" pairs (collected within the period 1971-2016); the dataset for grain corn included 47 "yield – water use" pairs (collected within the period 1970-2016); the dataset for soybeans included 53 "yield – water use" pairs (collected within the period 1981-2020). Yielding data in the field experiments were established through direct harvesting of the crops, with further recalculation of the yield to the standard moisture: 14% for grain of winter wheat and corn, and 12% for soybeans. The water use of the studied crops was established by the common methodology, explained in the work [15], using the Eq. (1):

$$WU = ER + SM + IR \tag{1}$$

where WU is water use, m^3/ha; ER – effective rainfall (precipitation exceeding 50 m^3/ha); SM – soil moisture used by the crops, m^3/ha; IR – irrigation rate applied during the crop growing season, m^3/ha.

Effective rainfall was measured in the field conditions using the rain gauge. The value of effective rainfall in mm was recalculated to m^3/ha multiplying the first one by 10. The soil moisture, used by the crops, was calculated as the difference between the moisture at the time of sowing and harvesting. The soil moisture was determined by gravimetric method [16].

For more details on initial datasets, used in the study, see Appendix A.

Mathematical modeling of the crops' yields based on the values of their water use was performed by the means of the best subsets regression analysis procedure within BioStat v.7 software [9, 17, 18]. The procedure embraced nine regression functions, described in the Table 1.

Table 1. Regression functions used within the best subsets regression analysis to predict the crops' yields by the values of their water use.

Function type	Equation
Linear	$Y=ax+b$
Quadratic	$Y=ax^2+bx+c$
Cubic	$Y=ax^3+bx^2+cx+d$
Stepwise (Power)	$Y=ax^b$
Exponential-1	$Y=ae^{bx}$
Hyperbolic (Reverse)	$Y=a+b/x$
Logarithmic	$Y=a+b\ln(x)$
Exponential-2	$Y=ab^x$
Sigmoid	$Y=e^{a+b/x}$

The fitting quality of different regression models was evaluated by the value of Pearson's correlation coefficient (R; the greater, the better), while the accuracy was assessed by the values of mean absolute percentage error (MAPE; the less, the better) and the values of the maximum absolute error (MAE; the less, the better) and the amplitude of the absolute errors (A; the less, the better) [19, 20]. The best model must have the lowest MAPE and A, while the value of R should be the highest.

The final decision on the model quality was made through the calculation of total score for each studied model. The total points "for" were calculated as the sum of the best values of the studied statistical indices for each studied regression function, every index is taken as 1 point. If the models have equal total score, the model with the highest R and the lowest MAPE should be preferred.

3. Results

As a result of the crops' yielding data statistical processing, in total 27 mathematical models for the yield prediction depending on the crops' water use were developed (Table 2; for initial datasets configuration see Appendix A). The statistical indices for the models, which were selected for the purpose of their fitting quality and accuracy evaluation, are presented in the Table 3. The best regression function is chosen by the best total score, presented in the correspondent graph of the Table 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 March 2023 doi:10.20944/preprints202303.0372.v1

Table 2. Regression models for the crops' yield prediction depending on their water use

Function type	Equations of the regression models		
	Grain corn	**Winter wheat**	**Soybeans**
Linear	$Y=1.1392\times10^{-4}x+3.4934$	$Y=6.4135–1.7234\times10^{-4}x$	$Y=0.2058+0.0006x$
Quadratic	$Y=6.4029\times10^{-8}x^2+7.4585\times10^{-4}x$ $–11.74$	$Y=3.2958+1.3053\times10^{-3}x$ $–1.5903\times10^{-7}x^2$	$Y=0.1813\times10^{-7}x^2$ $+0.0019x–1.8718$
Cubic	$Y=3.0951\times10^{-11}x^3+4.0807\times10^{-7}x^2$ $+1.6047\times10^{-3}x+26.492$	$Y=12.927+1.3394\times10^{-2}x$ $–3.0184\times10^{-6}x^2+2.1646\times10^{-10}x^3$	$Y=0.3277\times10^{-11}x^3$ $–5.3722\times10^{-8}x^2+0.0031–2.9201$
Stepwise (Power)	$Y=1.9388\times10^{-3}x^{0.72189}$	$Y=1.619x^{-0.12658}$	$Y=4.5518\times10^{-7}x^{1.5859}$
Exponential-1	$Y=4.3889e^{0.0001433x}$	$Y=6.5592e^{-0.000034952x}$	$Y=0.345e^{0.0005x}$
Hyperbolic (Reverse)	$Y=14.86–27646/x$	$Y=5.6093+815.47/x$	$Y=4.1027–5102.03/x$
Logarithmic	$Y=39.592+5.7366\ln(x)$	$Y=9.2518–0.4104\ln(x)$	$Y=13.6679+1.9734\ln(x)$
Exponential-2	$Y=4.3889\times1.0001^x$	$Y=6.5592\times0.99997^x$	$Y=0.345\times1.0005^x$
Sigmoid	$Y=e^{2.91 – 3482.7/x}$	$Y=e^{1.6352 – 366.81/x}$	$Y=e^{2.0212 – 4249.6163/x}$

Table 3. Regression statistics of the models for the crops' yield prediction depending on the water use (the best values of the index are marked with bold font).

Crop	Statis-tics	Function type								
		Linear	**Quad-ratic**	**Cubic**	**Power**	**Exp-1**	**Reverse**	**Logarith-mic**	**Exp-2**	**Sig-moid**
Grain corn	R	0.40	0.46	**0.47**	0.40	0.38	0.43	0.42	0.39	0.42
	MAPE	15.59%	15.32%	**14.92%**	15.80%	15.94%	15.44%	15.47%	24.32%	15.68%
	MAE	4.62	**4.33**	4.42	4.80	4.86	4.49	4.56	6.42	4.73
	A	4.56	**4.26**	4.42	4.78	4.84	4.45	4.55	6.38	4.70
Winter wheat	R	0.08	0.14	**0.22**	0.05	0.08	0.03	0.06	0.02	0.03
	MAPE	21.11%	**20.54%**	20.59%	90.38%	21.26%	**20.54%**	20.66%	20.67%	26.57%
	MAE	3.52	3.81	3.59	8.40	**3.48**	3.75	3.73	3.87	4.29
	A	3.45	3.80	3.53	6.93	**3.44**	3.70	3.66	3.86	4.26
Soy-beans	R	0.79	**0.87**	**0.87**	0.74	0.63	0.87	0.85	0.63	0.83
	MAPE	16.44%	12.28%	13.12%	20.41%	37.82%	**12.27%**	13.66%	37.78%	15.19%
	MAE	0.92	1.00	1.14	1.60	4.15	0.98	0.89	4.14	**0.85**
	A	0.87	1.00	1.12	1.56	4.06	0.97	0.88	4.06	**0.82**
Total pts. "for"		**0**	**4**	**4**	**0**	**2**	**2**	**0**	**0**	**2**

As a result of the models' evaluation, it was found out that the best general quality of the models (both in terms of fitting quality and prediction accuracy) is for soybeans, and the worst – for winter wheat. This fact could be explained by higher homogeneity of the input dataset for soybeans (the crop varieties varied less, all the crops were irrigated), and the highest heterogeneity for winter wheat (many different varieties, cultivation in the irrigated and rainfed conditions).

Besides, there is a difference in terms of the best response to regression functions. Thus, grain corn and soybeans tend to have the best response for polynomial functions (quadratic and cubic), while winter wheat – to exponential-1 and reverse functions.

Evaluating the final score for each model, we have found out an equal score of "4" points for quadratic and cubic functions. Considering that cubic function will have higher score if we neglect less important indices of the maximum absolute error and the amplitude of the absolute errors, we suggest that cubic function should be the first choice for

 doi:10.20944/preprints202303.0372.v1

agricultural modeling. We suggest that linear, power (stepwise), logarithmic, and exponential-2 functions should not be used unless there are strong reasons to implement these functions in crop modeling. Of course, this study has its limitations, and the suggestion is true for the models, developed on the basis of medium-size datasets (45-55 input pairs), as in this study.

4. Discussion

Prediction of crop yields is relevant and, at the same time, challenging task of modern agricultural science. As soon as the importance of crop yields prediction was admitted, scientists all over the globe started the search for appropriate mathematical methods to adopt them for the mentioned task [21]. Among statistical methods, regression analysis has been the first one to serve for crop yield prediction [22].

Starting with simple linear function, regression approach grew more and more important and increased the complexity of the applied mathematical functions. In the course of time, regression analysis became a dominant one in crop yield prediction by various inputs (climate models, results of the field experiments, remote sensing data, etc.), with numerous computation techniques and target functions involved, e. g., polynomial functions, multiple and multivariate regression, stepwise, logistic regression, etc. [23, 24, 25, 26]. Besides, novel regression approaches are still being developed and introduced, e.g., an interaction regression model, fuzzy regression, which were proved to be quite reliable and accurate in performing the task of yield analysis [27, 28]. At this point, right choice of the approach applied to yielding data analysis became crucial, and the focus of researchers should also be changed to studying the best statistical method for yield prediction in terms of fitting quality and prediction accuracy [29, 30].

The studies, devoted to the issue, raised above, are quite limited. The study [31] is somewhat like ours in terms of comparing different regression techniques for crop yields prediction on the basis of rainfall income. Another study [32] investigated the efficiency of Lasso and conventional polynomial regression techniques in crop yields prediction. Comprehensive evaluation of modern popular techniques for crop yield prediction, including multiple linear regression and stepwise linear regression, was performed by Gonzalez-Sanchez et al. [33]. All the referred studies provided different and valuable insights into the art of choosing right modeling approach, and our study just supplements them in terms of pure intra-comparison between regression analysis methods. The results of each study are somewhat different, but general idea is still the same: the more appropriate regression function is chosen, the better quality of crops' yield prediction is received.

A few words should be said about more recent mathematical methods of crops' yield analysis, in particular, about artificial neural networks, which are of growing popularity among scientific community. Convolutional neural networks alongside with long-short term memory and deep neural networks are of the greatest demand and are increasing their importance in various models for crop yields prediction [34, 35]. Often, especially when we are talking about big datasets, these mathematical approaches appear to be somewhat more accurate than conventional ones, including regression analysis [36, 37, 38]. However, the latter approach is not completely lost in the shade of artificial neural networks.

Deep learning algorithms, combined with multiple regression analysis in different ways, are proposed as a reasonable solution for precision crop yields prediction, and such an approach has been successfully implemented in several studies [21, 39]. Such combined models solve the issue of getting clear equation of yield prediction, because it is almost impossible to get an idea on how particular neural network had achieved its modeling results, due to so called "black box nature" of the latter. Besides, artificial neural networks sometimes are prone to overfitting, while in regression analysis this is under the control of a researcher [40]. Thus, regression analysis did not lose its value until now, when it is deeply integrated with novel data analysis methods. And as in case of using regression analysis as a self-sufficient data analysis method, it is highly important to choose right

 doi:10.20944/preprints202303.0372.v1

function for good combination with deep learning techniques. This issue will be the subject of our further investigations.

5. Conclusions

Although, regression analysis is looked upon as an outdated one, it is widely implemented in agricultural sciences. Success of yield modeling is strongly dependent not only on the quality and quantity of the model's inputs, but also on the right choice of the target function. Having analyzed regression statistics of the models for the three studied crops' yield prediction, it is established that the best choice for the medium-sized pair datasets is a cubic function. Linear, power (stepwise), logarithmic, and exponential-2 regression should be avoided as it is highly unlikely to receive good prediction and fitting quality with these functions.

Supplementary Materials: The supporting information is presented in the Appendix A to this manuscript.

Author Contributions: Conceptualization, P.L.; methodology, S.Z. and P.L.; software, P.L.; validation, P.L. and R.V.; formal analysis, P.L.; investigation, S.Z., R.V., O.P. and P.L.; resources, O.P.; data curation, O.P.; writing—original draft preparation, P.L.; writing—review and editing, P.L.; visualization, P.L.; supervision, R.V.; project administration, R.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All the necessary data, required for understanding of the study's concept and conduction, are presented in the manuscript itself, both in the main body and in the appendix. Additional data could be provided for reasonable request to the authors.

Conflicts of Interest: The authors declare no conflict of interest.

 doi:10.20944/preprints202303.0372.v1

Appendix A

The initial datasest used un the study

No. of pair	Grain corn		Winter wheat		Soybeans	
	Yield, t/ha	Water use, m^3/ha	Yield, t/ha	Water use, m^3/ha	Yield, t/ha	Water use, m^3/ha
1	8.91	5290	6.65	5862	3.68	4620
2	8.21	4580	5.54	5521	3.13	5060
3	7.44	5324	5.52	4762	3.15	4265
4	8.23	4357	6.01	4068	3.72	5310
5	9.14	4730	4.69	3177	3.66	5130
6	8.78	5110	6.60	4198	3.30	4850
7	8.62	4580	5.18	4244	2.47	3398
8	8.04	5060	6.74	4479	3.56	5440
9	8.66	4490	6.48	5184	1.90	4178
10	10.37	4770	5.90	4231	3.19	4019
11	9.82	4360	7.94	4572	2.91	5024
12	13.40	4644	6.82	5599	2.85	4960
13	10.61	5430	5.90	4276	2.88	4719
14	9.36	4570	6.54	3100	2.73	5246
15	9.70	4170	5.24	3946	3.67	4341
16	10.30	5080	7.76	4937	2.26	4762
17	9.12	3730	6.05	5663	3.74	4723
18	6.76	4005	7.39	3914	3.29	3763
19	8.11	3767	8.97	4016	2.27	4382
20	5.54	4210	8.24	4608	2.76	3104
21	5.38	4280	6.44	4279	2.58	3769
22	8.61	4302	4.30	4347	2.10	3846
23	9.54	4576	7.36	4016	3.04	5137
24	9.62	4140	4.76	4506	3.67	4538
25	4.38	3640	3.56	4929	2.67	4297
26	6.41	4570	3.37	5189	3.44	4783
27	3.83	4138	4.12	6341	2.91	6036
28	12.35	5300	5.86	5340	3.00	5950
29	7.74	4617	4.94	5126	2.80	4625
30	7.09	5214	5.59	3971	3.74	5221
31	8.47	4605	5.98	5351	2.33	4613
32	7.26	3562	2.02	5062	2.90	3569
33	9.46	4012	7.64	6138	2.72	4021
34	8.42	4012	5.79	6261	2.24	4020
35	11.67	4012	4.24	5122	3.09	4024
36	9.44	4109	5.25	4401	2.88	4118
37	6.56	3590	8.40	5347	2.60	4913
38	7.37	4012	3.44	5868	2.75	4922
39	8.69	4628	3.25	4849	2.85	4935
40	9.35	6177	4.00	5266	3.21	5362
41	8.54	6821	5.74	2913	3.38	5369
42	8.57	4755	4.82	2295	3.48	5382
43	12.29	5800	5.47	3316	2.84	5416
44	10.15	5299	5.86	5786	3.05	5421
45	9.38	4985	8.24	4637	3.11	5428
46	10.57	5978	–	–	0.28	1343
47	12.95	5196	–	–	0.30	1268
48	–	–	–	–	0.26	1415
49	–	–	–	–	0.28	1342
50	–	–	–	–	2.04	2954
51	–	–	–	–	2.54	2946
52	–	–	–	–	2.82	3200
53	–	–	–	–	2.43	3033

References

1. White, J.W. *Crop Modeling and Decision Support*. Tsinghua University Press, 2009. DOI: 10.1007/978-3-642-01132-0
2. Horie, T.; Yajima, M.; Nakagawa, H. Yield forecasting. *Agric. Syst.* **1992**, *40*, 211–236. DOI: 10.1016/0308-521X(92)90022-G
3. Choudhury, A.; Jones, J. Crop yield prediction using time series models. *J. Econ. Econ. Educ. Res.* **2014**, *15*(3), 53–67. DOI:
4. Vozhehova, R.; Kokovikhin, S.; Lykhovyd, P.V.; Balashova, H.; Lavrynenko, Yu.; Biliaieva, I.; Markovska, O. Statistical yielding models of some irrigated vegetable crops in dependence on water use and heat supply. *J. Water Land Dev*. **2020**, *45*(IV–VI), 190–197. DOI: 10.24425/jwld.2020.133494
5. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements – FAO Irrigation and Drainage Paper 56*. FAO: Rome, Italy, 1998; D05109.
6. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT cropping system model. *Europ. J Agron*. **2003**, *18*(3-4), 235–265. DOI: 10.1016/S1161-0301(02)00107-7
7. Chachi, J.; Taheri, S.M. Multiple fuzzy regression model for fuzzy input-output data. *Iran. J. Fuzzy Syst*. **2016**, *13*(4), 63–78.
8. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric*. **2018**, *47*, 70–90. DOI: 10.1016/j.compag.2018.02.016
9. Archontoulis, S.V.; Miguez, F.E. Nonlinear regression models and applications in agricultural research. *Agron. J*. **2015**, *107*(2), 786–798. DOI: 10.2134/agronj2012.0506
10. Manly, B.F.J.; de Almeida Machado, A.; Vencovsky, R.; Demétrio, C.G.B.; Ferreira, D.F. Statistical models in agriculture: biometrical methods for evaluating phenotypic stability in plant breeding. *Cerne* **2006**, *12*(4), 373–388.
11. Ruß, G. Data mining of agricultural yield data: A comparison of regression models. In Advances in Data Mining. Applications and Theoretical Aspects: 9th Industrial Conference, ICDM 2009, Leipzig, Germany, July 20-22, 2009. Proceedings 9 (pp. 24–37). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-03067-3_3
12. Sindelar, R.; Babuska, R. Input selection for nonlinear regression models. *IEEE Trans. Fuzz. Syst*. **2004**, *12*(5), 688–696. DOI: 10.1109/TFUZZ.2004.834810
13. Morozov, V.; Ushkarenko, V.; Lazer, P. Integrate water resources management on the irrigated lands of the South of Ukraine in the global climate changes conditions. In BALWOIS, Ohrid, Republic of Macedonia, May 25, 29, 2010. (pp. 1–3).
14. Lykhovyd, P. Irrigation needs in Ukraine according to current aridity level. *J. Ecol. Eng*. **2021**, *22*(8), 11–18. DOI: 10.12911/22998993/140478
15. Ushkarenko, V.O.; Vozhehova, R.A.; Holoborodko, S.P.; Kokovikhin, S.V. *Methodology of the Field Experiment (Irrigated Agriculture).* Hrin D.S.: Kherson, Ukraine, 2014.
16. Reynolds, S.G. The gravimetric method of soil moisture determination Part IA study of equipment, and methodological problems. *J. Hydrol*. **1970**, *11*(3), 258–273. DOI: 10.1016/0022-1694(70)90066-1
17. Frost, J. *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*. Statistics By Jim Publishing., 2019.
18. Frost, J. *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*. Statistics By Jim Publishing., 2019.
19. Evans, J.D. *Straightforward Statistics for the Behavioral Sciences*. Thomson Brooks/Cole Publishing Co., 1996.
20. Blasco, B.C.; Moreno, J.J.M.; Pol, A.P.; Abad, A.S. Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema* **2013**, *25*(4), 500–506. DOI: 10.7334/psicothema2013.23
21. Gopal, P.M.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric*. **2019**, *165*, 104968. DOI: 10.1016/j.compag.2019.104968
22. Basso, B.; Liu, L. Seasonal crop yield forecast: Methods, applications, and accuracies. *Adv. Agron*. **2019**, *154*, 201–255. DOI: 0.1016/bs.agron.2018.11.002
23. Dharmaraja, S.; Jain, V.; Anjoy, P.; Chandra, H. Empirical analysis for crop yield forecasting in India. *Agric. Res*. **2020**, *9*, 132–138. DOI: 10.1007/s40003-019-00413-x
24. Shah, A.; Dubey, A.; Hemnani, V.; Gala, D.; Kalbande, D. R. Smart farming system: Crop yield prediction using regression techniques. In Proceedings of International Conference on Wireless Communication: ICWiCom 2017 (pp. 49-56). Springer Singapore. DOI: 10.1007/978-981-10-8339-6_6
25. Kumar, S.; Attri, S.D.; Singh, K.K. Comparison of Lasso and stepwise regression technique for wheat yield prediction. *J. Agrometeorol*. **2019**, *21*(2), 188–192. DOI: 10.54386/jam.v21i2.231
26. Kumari, V.; Agrawal, R.; Kumar, A. Use of ordinal logistic regression in crop yield forecasting. *Mausam*. **2016**, *67*(4), 913–918. DOI: 10.54302/mausam.v67i4.1419
27. Ansarifar, J.; Wang, L.; Archontoulis, S.V. An interaction regression model for crop yield prediction. *Sci. Rep*., **2021**, *11*(1), 17754. DOI: 10.1038/s41598-021-97221-7
28. Garg, B.; Aggarwal, S.; Sokhal, J. Crop yield forecasting using fuzzy logic and regression model. *Comput. Electric. Eng*. **2018**, *67*, 383–403. DOI: 10.1016/j.compeleceng.2017.11.015
29. Sharif, B.; Makowski, D.; Plauborg, F.; Olesen, J.E. Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. *Europ. J. Agron*. **2017**, *82*, 11–20. DOI: 10.1016/j.eja.2016.09.015
30. Cerrato, M.E.; Blackmer, A.M. Comparison of models for describing; corn yield response to nitrogen fertilizer. *Agron. J*. **1990**, *82*(1), 138–143. DOI: 10.2134/agronj1990.00021962008200010030x
31. Antony, B. Prediction of the production of crops with respect to rainfall. *Environ. Res*. **2021**, *202*, 111624. DOI: 10.1016/j.envres.2021.111624

32. Rai, S., Nandre, J., & Kanawade, B. R. (2022, June). A Comparative Analysis of Crop Yield Prediction using Regression. In 2022 2nd International Conference on Intelligent Technologies (CONIT) (pp. 1-4). IEEE. DOI: 10.1109/CONIT55038.2022.9847783
33. Gonzalez-Sanchez, A.; Frausto-Solis, J.; Ojeda-Bustamante, W. Attribute selection impact on linear and nonlinear regression models for crop yield prediction. *The Sci. World J*. **2014**, *2014*, 509429. DOI: 10.1155/2014/509429
34. Nevavuori, P.; Narra, N.; Lipping, T. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.*, **2019**, *163*, 104859. DOI: 0.1016/j.compag.2019.104859
35. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric*/ **2020**, *177*, 105709. DOI: 10.1016/j.compag.2020.105709
36. Li, A.; Liang, S.; Wang, A.; Qin, J. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. *Photogramm. Eng. Remote Sens*., **2007**, *73*(10), 1149–1157.
37. Comrie, A.C. Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manag*. **1997**, *47*(6), 653–663. DOI: 10.1080/10473289.1997.10463925
38. Seya, H.; Shiroi, D. A comparison of residential apartment rent price predictions using a large data set: Kriging versus deep neural network. *Geogr. Anal*. **2022**, *54*(2), 239–260.
39. Lavrenko, S.; Lykhovyd, P.; Lavrenko, N.; Ushkarenko, V.; Maksymov, M. Beans (Phaseolus vulgaris L.) yields forecast using normalized difference vegetation index. *Intern. J. Agric. Technol*. **2022**, *18*(3), 1033–1044.
40. Tu, J.V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epid*. **1996**, *49*(11), 1225–1231.