Indian Journal of Ecology (2022) 49(6): 2370-2376
DOI: https://doi.org/10.55362/IJE/2022/3835

Manuscript Number: 3835
NAAS Rating: 5.79

Modelling and Forecasting Wheat Production in Punjab State of India using Hierarchical Time Series Models

Monika Devi, Pradeep Mishra[1*], Soumen Pal[2], Kanchan Sinha[2] and Chetna

Department of Mathematics & Statistics, CCS HAU, Hisar-125 004, India
[1]College of Agriculture, Jawaharlal Nehru Krishi Vishwa Vidyalaya, Rewa-486 001, India
[2]ICAR- Indian Agricultural Statistics Research Institute (IASRI), New Delhi-110 012, India
**E-mail: pradeepjnkvv@gmail.com*

Abstract: Due to the importance of wheat crop and Punjab being the leading wheat producer, this paper considers hierarchical time series data on wheat production in Punjab. The Punjab state wheat production data is organized in a hierarchy based on geographical regions. Top-down, bottom-up, middle-out, and optimal-combination approaches were used along with single series forecasting. An analysis of forecast performance shows that bottom-up approach outperforms other methods in terms of RMSE and MAE in out of sample forecast horizon. Finally, the state of Punjab has forecasted wheat production from 2019 to 2023 using a bottom-up approach.

Keywords: Bottom-up, Forecasting, Middle-out, Production, Wheat

Wheat (*Triticum astivum*) is a staple meal for nearly a third of the world's population, as well as a significant source of protein, niacin, and fibre in the diet. India is the world's second-largest wheat producer, blessed with a variety of agro-ecological conditions that give food and dietary security to the Indian people (Sharma and Sendhil 2015, Sendhil et al 2019), in particular in the recent past. Wheat has increased in this country to 99,70 million tons, 13.64 percent of global production, with a productivity level of about 3,371 kg/ha(www.fao.org.). Uttar Pradesh, Punjab, Haryana and Madhya Pradesh are major wheat producing states in India. (http://www.agricoop.nic.in). In Punjab, wheat production and productivity have increased consistently in the last five decades (Kaur et al 2015). Punjab has maintained the status of increased wheat productivity for many years (Sendhil et al 2019). Faridkot is top wheat growing district of Punjab with a production of 2635 thousand tons. Predicting the production behaviour of major crops is critical in tackling the global food safety crisis. Modeling and forecasting and application of phenomena, particularly in the agricultural sector, became significant in the latter half of the last century. Mishra et al (2015) considered the wheat productivity forecast time series model in India. With the introduction of hierarchical time series methodology, it has gained further boost, particularly in time series predictions. The hierarchy of time series models in various fields has been studied by Athanasopoulos et al (2009) and Moon et al (2012). Pal and Paul (2016) used hierarchical time-series to model and predict the production of sorghum in India and observed that the middle-out technology outperformed other approaches to hierarchical time series models and to traditional prediction methods. Mitra et al. (2017) use hierarchical time series models to predict the production of India's oilseeds and pulses and the bottom-up approach was superior to the other approaches of the hierarchical model of time series for modeling and forecasting the production of oilseeds in India, while the optimum combination for pulses production was the best. Gua and Xue (2014) applied Artificial Neural Network to compare spatial and temporal models of crop yield forecasting. Shrivastri et al (2022) compared the different time series model for forecasting of wheat production for India. In the present study the production of wheat has been predicted using hierarchical time series models in major regions and all districts of the state of Punjab.

MATERIAL AND METHODS

Hierarchical time-series: The level 0 (zero) refers to the aggregate series, level 1 the first stage, and the most disaggregated series is level K. A letter sequence identifies the individual series and the level of hierarchy. As in: A is for level 1 series A, AB is for level 2 series A, and so on (Fig. 1).

The observations were recorded at times t = 1,2,…n, and the objective is to forecast each series at each level at times t = n + 1, n +2,…,n + h. Suppose, the notation X is used to refer to a generic series within the hierarchy. Observation on series X are written as $y_{x,t}$. Thus, $y_{AB,t}$ is the value of series AB at time t. $y_t$denotes aggregate of all series at time t. Therefore,

$$y_t = \sum_i y_{ij,t},\ y_{i,t} = \sum_j y_{ij,t},\ y_{ij,t} = \sum_k y_{ijk,t},\ y_{ijk,t} = \sum_l y_{ijkl,t},$$

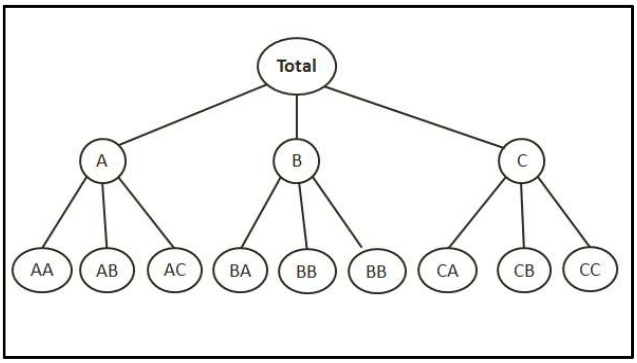

Fig. 1. A3-structure hierarchical tree diagram

The, observations at higher levels are obtained by adding up the series below.

Let m_i denotes the total number of series at level i (i=2,1,2,…,K) subject to the constraint, $m_i > m_{i-1}$, then the total number of series in the hierarchy is $m = m_0 + m_1 + \ldots + m_k$. In the above example m =13.

Let $y_{i,t}$ denotes the vector of all observations at level *i* and time t and $y_t = [y_t, y_{1,t},\ldots,y_{kt}]$

$y_t = Sy_{k,t}$ (1)

where S is a "summing" matrix of order $m \times m_K$ used to aggregate the forecasts of the lowest level series. In the above example, $y_t = [y_t, y_{A,t}, y_{B,t}, Y_{C,t}, Y_{AA,t}, Y_{AB,t},\ldots, y_{CC,t}]$ and the summing matrix is of order 13 × 9 and is given by

$$S=\begin{bmatrix}
1&1&1&1&1&1&1&1&1\\
1&1&1&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&1&1&1\\
1&0&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&1
\end{bmatrix}$$

The rank of S is m_k.

The $\hat{y}_{x,n}(h)$ be the h-step-ahead forecasts for the series y_x. A sample of t = 1,2,…, n is used to generate the forecasts. Therefore, $\hat{y}_{AA,n}(h)$ denotes the h-step-ahead base forecast of series y_{AA} using the sample $y_{AA,1}, y_{AA,2},\ldots,y_{AA,n}$. For level i, all h-step-ahead base forecasts can be represented by $\hat{y}_{i,n}(h)$ and the h-step-ahead base forecasts for the whole hierarchy are given as $\hat{y}_{,n}(h)$, which contains all of the base forecasts stacked in the same sequence as y_t.

Using this notation, all existing hierarchical methods can be represented by the general form

$$\tilde{y}_n(h) = SP\hat{y}_n(h) \qquad (2)$$

where S is the summing matrix of order $m \times m_k$, as in Eq.(1), and P is a matrix of order $m_k \times m$.

The form of P differs depending on the hierarchical forecasting approach.

Bottom-up approach: For the majority of disaggregated series at the bottom of the hierarchy, bottom-up forecasting generates independent base forecasts that are then aggregated to generate revised forecasts for the remaining series Figure 1. After obtaining the h-step-ahead independent forecasts for each of the bottom level series namely $\hat{y}_{AA,n}(h), \hat{y}_{AB,n}(h), \ldots, \hat{y}_{CC,n}(h)$ aggregate these forecasts upwards to obtain the h-step-ahead forecasts for the whole hierarchy as follows:

$$\tilde{y}_A(h) = \tilde{y}_{AA}(h) + \tilde{y}_{AB}(h) + \tilde{y}_{AC}(h)$$
$$\tilde{y}_B(h) = \tilde{y}_{BA}(h) + \tilde{y}_{BB}(h) + \tilde{y}_{BC}(h)$$
$$\tilde{y}_C(h) = \tilde{y}_{CA}(h) + \tilde{y}_{CB}(h) + \tilde{y}_{CC}(h)$$
$$\tilde{y}(h) = \tilde{y}_A(h) + \tilde{y}_B(h) + \tilde{y}_C(h)$$

For the bottom–level series the revised forecasts are same as the base forecasts (i.e. $\tilde{y}_{AA}(h) = \hat{y}_{AA}(h)$).

$$P = \left[0_{m_k \times (m-m_k)} / I_{m_k}\right] \qquad (3)$$

The general form of this approach is represented as

where $0_{i\times j}$ is the i × j null matrix. In this case, the role of is to aggregate the revised forecasts of all series in the hierarchy. For modeling, the bottom-up method uses the most disaggregated bottom level series data, so no information is lost.

Top-down approach: In the top-down approach, forecasts of the "Total" series y t are first generated, then disaggregated downward based on data proportions (Athanasopoulos et al 2009). This approach's general form is:

$$\mathbf{P} = \left[\mathbf{p} / \mathbf{0}_{m_k \times (m-1)}\right]$$

where $p = [p_1, p_2,\ldots, p_{mk}]$ are a set of proportions for the bottom level series. p gives the base forecasts for the "Total" series as revised forecasts for the hierarchy's bottom level. Her the top-level forecasts are disaggregated to obtain the lower-level series predictions. In this approach, the revised forecasts at the top level are equal to the highest base predictions, i.e., $\tilde{y}_t(h) = \hat{y}_t(h)$.

Middle-out approach: In the first step, intermediate hierarchical projections are generated and these forecasts disintegrate in order to achieve revised lower and aggregate forecasts at higher hierarchical levels for the purposes of

computing revised forecasts. Basic forecasts are first produced for all the selected mid-level series. These basic predictions are added to get a down-to-earth approach to the revised series above the mid-level predictions. Then the mid-level forecasts are divided down to get a top-down approach to the revised forecasts for the lower mid-level series.

Optimal forecasts using regression: Hyndman et al (2011)'s approach allows all h-step-ahead base predictions to be expressed by the linear regression model

$$\hat{\mathbf{y}}_n(h) = \mathbf{S}\boldsymbol{\beta}_n(h) + \boldsymbol{\varepsilon}_h$$

where $\beta_n(h) = E(y_{k,n+h} / y_1, y_2, \ldots y_n]$ is the unknown mean of the bottom level K and ε_h has zero mean and covariance-variance matrix $Var(\varepsilon_h) = \sum_h$. Then the $\beta_n(h)$ is estimated by considering Eq.(5) as a regression equation, and thus obtain forecasts for all levels in the hierarchy. If was $\sum_h$ known, one can use generalized least squares estimation to obtain the minimum variance unbiased estimate of $\beta_n(h)$ as

$$\hat{\boldsymbol{\beta}}_n(h) = \left(\mathbf{S}'\boldsymbol{\Sigma}^{\dagger}{}_h\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{\dagger}{}_h\hat{\mathbf{y}}_n(h) \qquad (5)$$

where $\sum^{+}_h$ is the Moore-Penrose generalized inverse of $\sum_h$. A generalized inverse is used because $\sum_h$ is often (near) singular due to aggregation involved in y_n. This results the following revised forecasts

$$\tilde{\mathbf{y}}_n(h) = \mathbf{S}\hat{\boldsymbol{\beta}}_n(h) = \mathbf{SP}\hat{\mathbf{y}}_n(h) \text{ and } \mathbf{P} = \left(\mathbf{S}'\boldsymbol{\Sigma}^{\dagger}{}_h\mathbf{S}\right)^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{\dagger}{}_h \quad (6)$$

For hierarchical time-series data under this study, modeling and forecasting have been done using hts package (Hyndman *et al.* 2020) in R softwarehttps://cran.r-project.org/web/packages/hts/hts.pdf.

Forecasting accuracy was measured by RMSE and MAE respectively (Mishra et al 2021) of five forecasting methods for different forecasting horizon (h=1 to 6) using eq. (7) and (8).

$$MAE = 1/h\Sigma_{i=1}^{h}\left|y_{t+i} - \hat{y}_{t+i}\right| \qquad (7)$$

$$RMSE = [1/h_{i=1}^{h}\{(y_{t+i} - \hat{y}_{t+i})^2\}^{1/2} \qquad (8)$$

To meet the aim of present investigation, data was obtained from Open Government Data (OGD) Platform of India related to wheat production in the districts of Punjab state for the period from 1973 to 2018.The condition of Punjab is classified according to homogeneity, rainfall distribution, soil texture, crop pattern, etc into three agro-climate zones. These areas are sub-mountainous, central and south-west zones, also called wheat-maize areas, wheat-paddy and wheat-cotton (Table 1). Presently whole Punjab is divided into 22 districts. However, before 1992, there were 12 districts in Punjab state Table 1. Later, these 12 districts have been further divided into other districts. As the district wise wheat production time series data, before 1992, is consisted of 12 districts only, this many number of districts are considered in the present study. 1992 onwards, data series of newly formed districts have been merged into their respective old districts to obtain data for those 12 districts (Table 2). Punjab state is at the top level of this structure. Level 1 consists of the three wheat productivity zones. Level 2 consists of 12 nodes representing 12 districts of Punjab. For example, node AA represents the Ferozepur district. In

Table 2.

Total level		
1	Total	Punjab
Level 1: Zone		
2	A	South West (Zone I)
3	B	Central Zone (Zone II)
4	C	Sub Mountainous Zone (Zone III)
Level 2: District		
South West (Zone I)		
5	AA	Ferozepur+Fazilka (2011)
6	AB	Faridkot+Moga(1995)+Sri Muktsar Shahib(1995)
7	AC	Bathinda+Mansa (1992)
Central (Zone II)		
8	BA	Amritsar+Tarantaran (2006)
9	BB	Kapurthala
10	BC	Jalandhar
11	BD	Ludhiana
12	BE	Sangrur+Barnala (2006)
13	BF	Patiala+Fatehgarh Sahib
Sub Mountainous (Zone III)		
14	CA	Gurdaspur+Pathankot (2011)
15	CB	Hoshiarpur+Shaheed Bhagat Singh Nagar (1995)
16	CC	Rupnagar+SAS Nagar (2006)

*Figures in bracket represent the year of formation of different district of Punjab

Table 1. Wheat agro-climatic zones in Punjab

Name of zone	Productivity	Districts
South west (Zone I)	Mid productivity	Ferozepur, Faridkot, Bathinda
Central zone (Zone II)	High productivity	Amritsar, Kapurthala, Jalandhar, Ludhiana, Sangrur, Patiala , Fatehgarh Sahib
Sub mountainous zone (Zone III)	Low productivity	Gurdaspur, Hoshiarpur, Rupnagar

2011, Fazilka district came out as the result of division of original Ferozepur district into these two separate districts

RESULTS AND DISCUSSION

Level 1 shows production of wheat in 3 different zones separately of Punjab state. Time-series of total wheat production in Punjab is displayed in level 0. For each of the graphical representation, y-axis represents production in '000 tonne and x-axis indicates the time period (year). Tables 3 and 4 present forecasting accuracy (RMSE and MAE, Mishra et al 2021) respectively of five forecasting methods for different forecasting horizon (h=1 to 6). It the bottom-up approach outperformed the other methods in case of both RMSE and MAE. On an average (over the forecast horizon), we estimated a MAE=209.19 for bottom-up, followed by the middle-out with an MAE. According to the RMSE accuracy statistics, the bottom-up gives a value of 234.6 followed by the middle-out and the optimal method. Consequently, select the bottom-up method for forecasting the next five years' (2019-2023) wheat production in Punjab, 3 different productivity zones and in 12 districts separately (Table 5).

The, Zone-II would be leading among the zones of Punjab in wheat production with 8834.07 thousand tones in year 2023-24. By the same period, wheat production of whole Punjab state would reach 19037.1 thousand tones. Faridkot would be leading district with 2826.07 thousand tones production of wheat in the year 2023-24. An increasing future trend of wheat production in district-wise and zone-wise has

Fig. 2. Hierarchical time-series of wheat production at level 0

Table 3. Root mean square error (RMSE) for each forecast horizon

Methods	RMSE						
	Forecast horizon (h)						
	1	2	3	4	5	6	Average
Bottom-up	66.43	196.05	153.40	341.63	278.81	371.29	234.60
Top-down	76.78	212.83	161.03	340.22	286.94	473.30	258.52
Middle-out	68.96	226.99	176.49	333.55	289.22	363.55	243.13
Optimal	71.05	211.81	160.80	338.08	284.94	406.18	245.48
Independent	403.72	1059.24	617.77	1787.97	1494.33	1943.00	1217.67

Table 4. Mean absolute error (MAE) for each forecast horizon

Methods	MAE						
	Forecast horizon (h)						
	1	2	3	4	5	6	Average
Bottom-up	66.43	194.73	144.54	300.07	226.64	322.71	209.19
Top-down	76.78	208.37	142.86	299.13	237.89	434.35	233.23
Middle-out	68.96	220.94	159.32	291.64	241.60	317.91	216.73
Optimal	71.05	208.88	147.77	296.47	235.34	361.79	220.22
Independent	403.72	1058.59	615.82	1593.59	1214.42	1695.89	1097.00

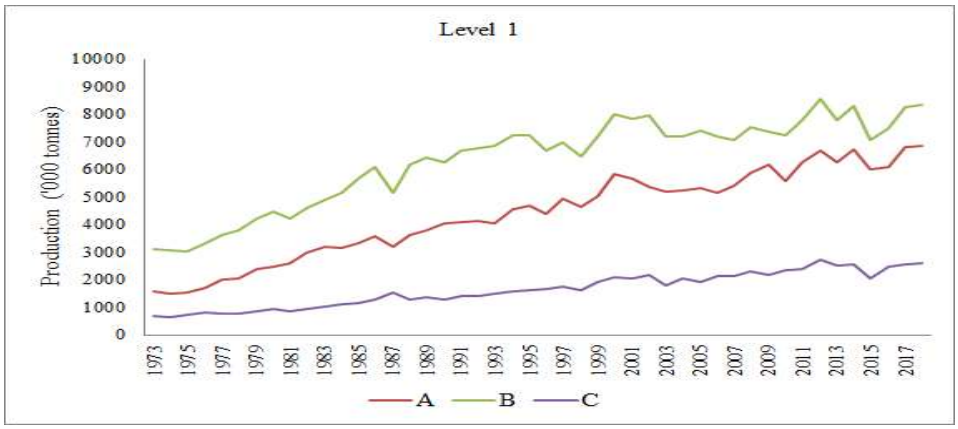

Fig. 3. Hierarchical time-series of wheat production at level 1

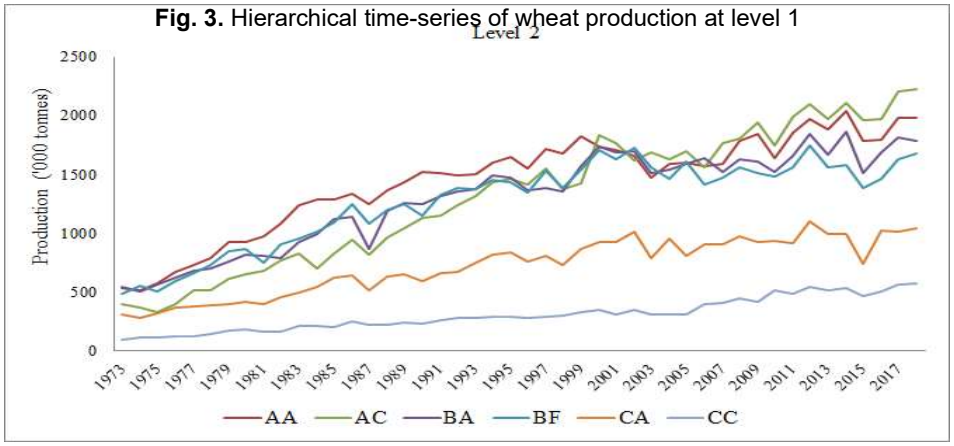

Fig. 4. Hierarchical time-series of wheat production at level 2 for selective districts of Punjab state

Table 5. Forecasting of wheat production at all levels during 2019-2023

Level		2019	2020	2021	2022	2023
Top level						
1	Total	17954.7	18225.4	18495.9	18766.5	19037.1
Level 1						
2	A	6900.43	7018.89	7137.35	7255.81	7374.27
3	B	8400.27	8508.72	8617.17	8725.62	8834.07
4	C	2654.01	2697.76	2741.39	2785.08	2828.74
Level 2						
5	AA	2014.25	2046.73	2079.2	2111.67	2144.14
6	AB	2647.63	2692.24	2736.85	2781.46	2826.07
7	AC	2238.54	2279.92	2321.3	2362.68	2404.06
8	BA	1823.54	1851.8	1880.04	1908.27	1936.51
9	BB	528.059	536.468	544.876	553.284	559.692
10	BC	861.23	862.12	863.52	865.01	867.3
11	BD	1288.63	1300.12	1311.12	1322.01	1334.12
12	BE	2225.55	2259.35	2293.17	2327.01	2360.83
13	BF	1673.27	1698.86	1724.44	1750.03	1775.61
14	CA	1044.74	1060.99	1077.24	1093.49	1109.74
15	CB	1024.89	1041.73	1058.58	1075.42	1092.27
16	CC	584.383	595.039	605.574	616.162	626.727

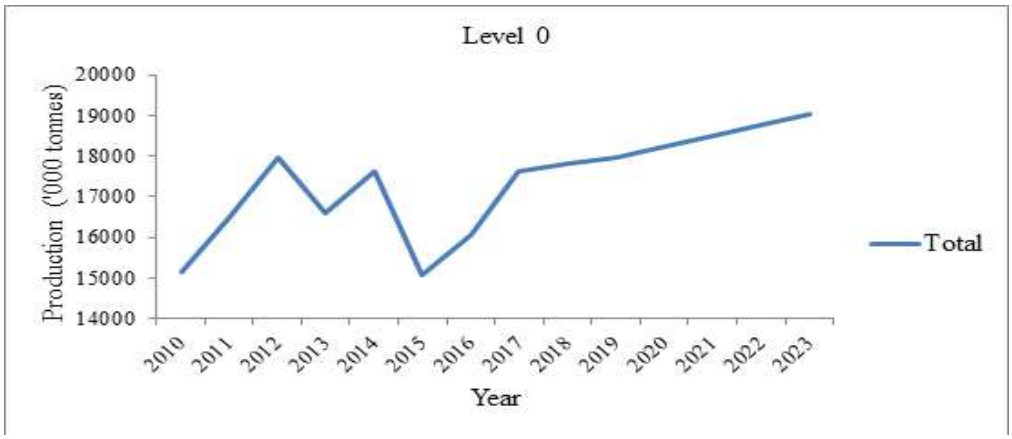

Fig. 5. Hierarchical forecasting of wheat production at level 0

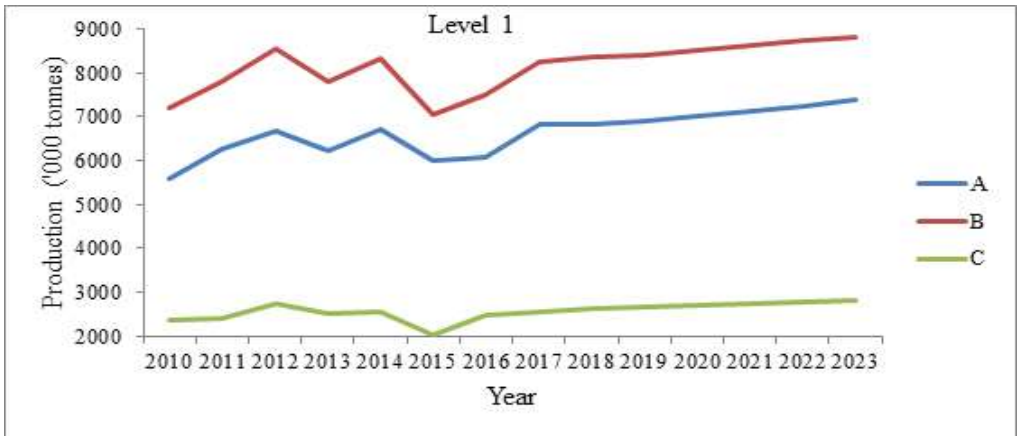

Fig. 6. Hierarchical forecasting of wheat production at level 1

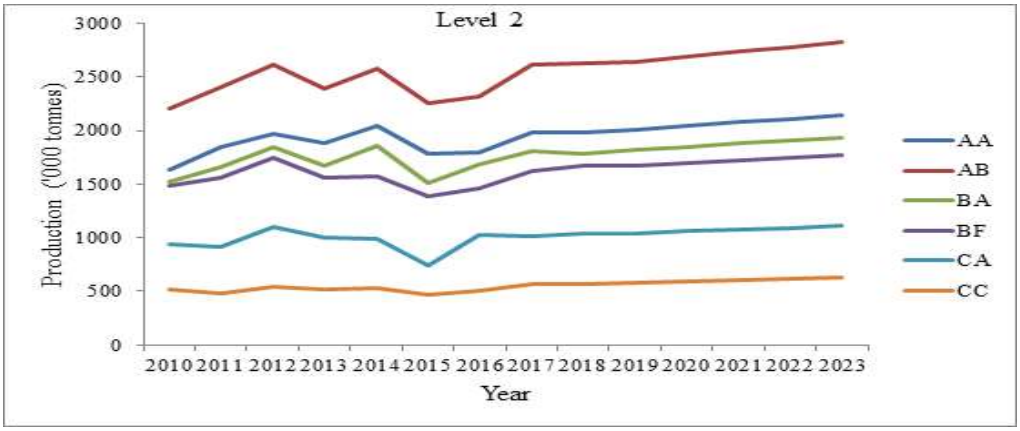

Fig. 7. Hierarchical forecasting of wheat production at level 2

been observed for Punjab (Figure 5, 6 and 7). The bottom-most level i.e. level 2 exhibits forecast value of wheat production for each of the districts, however, for clarity of graphical representation, only few of those are labeled. Level 1 shows forecasted production of wheat in different zones of Punjab. Predicted future values of total wheat production in Punjab are displayed in level 0.

CONCLUSION

Hierarchical time series modelling and forecasting of wheat production in Punjab state of India has been carried out. Among five different methods, bottom-up approach has outperformed the other methods in terms of forecast accuracy measure criteria viz., RMSE and MAE in out of sample forecast horizon. The next five years' forecast is also

calculated for all three levels which show a growing trend in wheat production and this forecasting approach is ideal for short forecast time, since predictive accuracy is used to decrease as the forecast horizon increases. The projection can provide direct support in the formulation of national agriculture policies and provide good food security decision-making well in advance. The approaches utilized here can be used for forecasting the production of other crops for which hierarchical time series data are available.

REFERENCES

Anonymous 2003. *Agriculture at a Glance. Directorate of Economics and Statistics.* Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India. Available from http://www.agricoop.nic.in. Last accessed on 20 October 2020.

Anonymous 2010. *Agricultural Statistics at a Glance. Directorate of Economics and Statistics.* Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India. Available from http://www.agricoop.nic.in. Last accessed on 20 October 2020.

Athanasopoulos G, Ahmed RA and Hyndman RJ 2009. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* **25**(1): 146-166.

Guo WW and Xue H 2014. Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. *Mathematical Problems in Engineering*. http://dx.doi.org/10.1155/2014/857865.

Hyndman RJ, Ahmed RA, Athanasopoulos G and Shang HL 2011. Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis* **55**(9): 2579-2589.

Hyndman R, Lee A, Wang E and Wickramasuriya S 2020. hts: Hierarchical and Grouped Time Series. R package version 6.0.1.URLhttps://CRAN.R-project.org/package=hts.

Kahn KB 1998. Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting* **17**(2): 14-19.

Received 13 August, 2022; Accepted 26 November, 2022

Mishra P, Sahu PK, Dhekale BS and Vishwajith KP 2015. Modeling and forecasting of wheat in India and their yield sustainability. *Indian Journal of Economics and Development* **11**(3): 637-647.

Mishra P, Ray S. Abotaleb M. Abdullah MG, Al Khatib, Tiwari S Badr A and Balloo R 2021. Estimation of fish production in India using ARIMA, Holt's, Linear, BATS and TBATS Models. *Indian Journal of Ecology* **48**(5): 1254-1261.

Mishra P, Yonar A, Yonar H, Kumari B, Abotaleb M, Das SS and Patil SG 2021. State of the art in total pulse production in major states of India using ARIMA techniques. *Current Research in Food Science* **4:** 800-806.

Mitra D, Paul RK and Pal S 2017. Hierarchical time-series models for forecasting oilseeds and pulses production in India. *Economic Affairs* **62**(1): 103-111.

Moon S, Hicks C and Simpson A 2012. The development of ahierarchical forecasting method for predicting spare partsdemand in the South Korean Navy-A case study. *International Journal of Production Economics* **140**(2): 794–802.

Pal S and Paul RK 2016. Modelling and forecasting sorghum (Sorghum bicolor) production in India using hierarchical time-series models. *The Indian Journal of Agricultural Sciences* **86**(6): 803-808.

Kaur P, Singh H, Rao VUM, Hundal S, Sandhu S, Nayyar S, Rao B and Kaur A 2015 Agrometeorology of wheat in Punjab state of India, https://doi.org/10.13140/ RG.2.1.5105.6721.

Sendhil R, Kumar TK and Singh GP 2019. Wheat Production in India: Trends and Prospects. *Global Wheat Production. Intech Open Limited* London, 16-19.

Sharma I and Sendhil R 2015. Domestic production scenario of wheat. *Souvenir of Roller Flour Millers Federation of India Platinum Jubilee Celebration* 18-20.

Shrivastri S, Alakkari KM, Lal P, Yonar A and Yadav S 2022. A comparative study between (ARIMA-ETS) Models to Forecast Wheat Production and its Importance's in Nutritional Security. *Journal of Agriculture, Biology and Applied Statistics* **1**(1): 25-37

Sharma I, Sendhil R and Chatrath R 2015. Regional disparity and distribution gains in wheat production. *Souvenir of 54th AIW&B Workers Meet;* Gujarat: Sardar Krishnagar Dantiwada Agricultural University.