CrossMark

ORIGINAL ARTICLE

Modeling of hotspot detection using cluster outlier analysis and Getis-Ord Gi* statistic of educational development in upper-primary level, India

Mrityunjoy Jana[1] · Nityananda Sar[2]

Abstract In recent years educational variability has threatened the sustainability of elementary level in Indian educational system. Systematic methodology to assess the educational development of the elementary sector to educational variability is currently not available. Towards this end, the present paper pays attention with the assessment of spatial structure of Indian educational development to elementary levels in 35 states including union territories (UTs) of all over country. For this purpose, a set of indicators was selected for each of the four components of educational development index (EDI). The data were arranged in the form of diagonal matrix and normalized them using functional relationship. Upon receiving normalized values, here researchers consider a method of unequal weights followed. The choice of the weights in this manner would ensure that large variation in any one of the indicators would not unduly dominate the contribution of the rest of the indicators and distort inter states as well as union territories (UTs) comparisons. It is well known that, in statistical comparisons, it is more efficient to compare two or more means after equalizing their variances. The principal objectives of this article to identification of Hotspot and cold-spot and delineate the spatial cluster of upper primary education level in India based on Getis-Ord Gi* statistic using fixed distance band in ArcGIS software.

Introduction

Universalization of Elementary Education (UEE) is a constitutional provision and a national commitment in India. Universalization implies educating all children up to the age of 14 which is equivalent to completion of upper primary level of education. Over the years, efforts by the government towards achieving the goal of UEE were focussed more on the upper primary level of education. With the expansion of enrolment at the upper primary level of education, the pressure for expansion of the upper primary level of education has increased. The index on educational development and highlight the interstate disparity in development of elementary school education (Raju et al. 2008). The present study was an effort to analyze implications for provision of upper primary education facilities to ensure that this level of education is provided to all eligible children. This implies creation of easy geographical access conditions to enroll all the eligible children in upper primary schools, provision reasonable levels of physical and infrastructural facilities in upper primary schools and teaching learning materials in the classrooms to facilitate meaningful curriculum transaction. Mere assurance of physical access to education cannot guarantee quality education (Sengupta et al. 2004). Universalization of upper primary education in India is normally discussed in terms of enrolling and retaining all children belonging to the age group 11–14. According to Census of India 2011, literacy in India is 74.04 %. In other words, nearly a quarter of parents are still illiterate. This seems more to be a desirable goal to be achieved in the long run than a realisable target at the present levels of

✉ Nityananda Sar
 sarnityananda@gmail.com

[1] Department of Education, North Orissa University, Baripada, Orissa, India

[2] Department of Geography, North Eastern Hill University, Shillong, Meghalaya, India

development of upper primary education. Enrollment is a function of the relevant age group at the upper primary level of education. However, enrolment in upper primary schools is more a function of upper primary education completion rates than a function of the relevant age group. It is logical to argue that all children in the relevant age group (11–14) cannot be enrolled in upper primary classes unless they complete upper primary level of education. In other words, all relevant age group children can be provided upper primary education only when all children of the upper primary school going age group are enrolled, retained and successfully complete the upper primary stage of education. Since upper primary education is not yet universalised, this implies that universalization of upper primary education means providing upper primary education for all children who have successfully completed upper primary level of education. The present study has adopted this as the operational definition of universalisation of upper primary education in India. However, once universalisation of upper primary education is attained, then there cannot be any difference between providing upper primary level of education to all age group children and those who complete upper primary stage of education. The effort at present needs to be to improve the inter stage transition ratios from primary to upper primary levels of education. DISE data set to generate a host of indicators directly related with educational performance Again (Sengupta and Pal 2012).

Materials

An effort has been made by the National University of Educational Planning and Administration (NUEPA) and the Government of India (MHRD, Department of School Education and Literacy, 2004) to compute an Educational Development Index (EDI) (Table 1), separately for Upper primary and Upper primary levels of education and also a composite index for the entire Elementary education which is exclusively based on the DISE data (DISE Flash Statistics, 2005–2006, 2006–2007, 2007–2008, 2008–2009, 2009–2010, 2010–2011, 2011–2012, 2012–2013, 2013–2014) for the nine consecutive academic sessions periods from 2005–06 to 2013–14. The first such attempt was made in 2005–06. A set of 24 indicators have been used in computing EDI which are regrouped into the four sub-groups, namely Access, Infrastructure, Teachers and Outcome indicators. The Indicators used for constructing EDI were pre-determined by a Working Group on EDI constituted by the MHRD during 2005–06 of which NUEPA was also one of the institutional members. There is no change in indicators this year as the same set of indicators used in 2012–13 has also been used in 2013–14. From last year, some indicators are newly introduced, some are deleted or modified. Percentage of habitations not served,

percentage of schools with less than 2 teachers (in schools with more than 15 students) (upper primary schools only), percentage of schools with <3 teachers (upper primary schools/sections), repetition rate and ratio of exit class over Class VI enrolment (upper primary stage) have been removed. Newly introduced indicators are density of schools per 10 km^2, percentage of classroom-teacher ratio 1:1, percentage of schools with kitchen-shed (Government and Aided School), percentage of single-teacher schools, average number of instructional days, average working hours for teachers, percentage change in enrolment in Government schools over the previous year and participation of Muslim children. Percentage of schools with ramp has been modified to percentage of schools required and have ramp. NUEPA has been pursuing the goal of creating a reliable system of statistics on school education during the last two decades through the District Information System for Education (DISE) which provides the basis for assessing the progress under SSA and on status of implementation of the Right to Education Act. The importance of this has further increased with efforts to extend the policy of universal education to cover secondary education stage of schooling also. Keeping this in view NUEPA is making a concerted effort to provide a unified system of school education statistics for all levels of schooling from elementary to higher secondary education. The present publication is outcome of this effort through which information from about 1.52 million schools have been collected. It is, indeed, a great pleasure for me to present a summary data base on elementary education in India through present publication. Including this publication, NUEPA brings out on annual basis seven publications exclusively based on DISE data under the following titles (DISE data, 2005–06 to 2013–14): (a) DISE flash statistics: elementary education in India, progress towards UEE: (b) elementary education in India: where do we stand; district and state report cards; (c) analytical tables; and (d) elementary education in rural and Urban India.

Methodology

Cluster-outlier analysis

Cluster-outlier analysis was used to estimate spatial cluster among the states of India at upper–upper upper primary level educational development based on inverse distance weight (IDW) technique in ArcGIS environment (Mitchell 2005; Scott and Warmerdam 2005). The cluster-outlier field has been recognized with a statistically significant level (i.e., 0.05 level or so on); those are fell into cluster of high values (HH), cluster of low values (LL), outlier in which a higher values is surround by lower ones (HL), and outlier in which a lower values is surrounded by higher values (LH). The z scores and p values were then computed

Table 1 Educational development index (EDI) normalised value of upper-primary level in India from 2005–06 to 2013–14 (Flash statistics, NUEPA)

S. no.	State/UT	2005–06	2006–07	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13	2013–14
1	A and N Islands	0.522	0.645	0.733	0.771	0.802	0.772	0.637	0.632	0.604
2	Andhra Pradesh	0.663	0.680	0.775	0.731	0.753	0.792	0.630	0.584	0.594
3	Arunachal Pradesh	0.438	0.442	0.536	0.509	0.543	0.609	0.553	0.610	0.629
4	Assam	0.511	0.522	0.583	0.536	0.515	0.633	0.472	0.521	0.522
5	Bihar	0.325	0.344	0.437	0.453	0.492	0.515	0.488	0.523	0.522
6	Chandigarh	0.667	0.730	0.754	0.730	0.876	0.783	0.629	0.644	0.671
7	Chhattisgarh	0.557	0.529	0.562	0.621	0.545	0.641	0.548	0.578	0.596
8	Dadra and Nagar Haveli	0.555	0.567	0.704	0.645	0.738	0.701	0.644	0.637	0.612
9	Daman and Diu	0.601	0.650	0.766	0.795	0.818	0.837	0.729	0.693	0.630
10	Delhi	0.649	0.718	0.748	0.749	0.814	0.763	0.658	0.645	0.682
11	Goa	0.558	0.617	0.713	0.654	0.794	0.722	0.562	0.514	0.547
12	Gujarat	0.647	0.698	0.763	0.712	0.795	0.740	0.666	0.640	0.678
13	Haryana	0.524	0.614	0.750	0.785	0.782	0.800	0.662	0.634	0.651
14	Himachal	0.687	0.729	0.734	0.751	0.740	0.784	0.632	0.664	0.692
15	Jammu	0.616	0.659	0.702	0.671	0.646	0.724	0.566	0.608	0.574
16	Jharkhand	0.411	0.390	0.533	0.490	0.445	0.551	0.502	0.482	0.516
17	Karnataka	0.697	0.701	0.792	0.734	0.768	0.696	0.759	0.688	0.682
18	Kerala	0.689	0.762	0.820	0.804	0.822	0.839	0.634	0.619	0.655
19	Lakshadweep	0.591	0.693	0.809	0.836	0.867	0.870	0.750	0.754	0.712
20	Madhya	0.510	0.489	0.587	0.599	0.536	0.654	0.594	0.563	0.545
21	Maharashtra	0.635	0.701	0.758	0.740	0.738	0.743	0.694	0.672	0.638
22	Manipur	0.557	0.633	0.687	0.623	0.616	0.669	0.564	0.643	0.624
23	Meghalaya	0.533	0.509	0.590	0.530	0.536	0.619	0.499	0.552	0.573
24	Mizoram	0.669	0.653	0.728	0.754	0.785	0.780	0.599	0.653	0.635
25	Nagaland	0.506	0.554	0.672	0.669	0.666	0.686	0.483	0.561	0.589
26	Orissa	0.484	0.447	0.571	0.556	0.532	0.631	0.544	0.568	0.605
27	Pondicherry	0.691	0.760	0.789	0.861	0.870	0.888	0.677	0.721	0.735
28	Punjab	0.588	0.633	0.708	0.734	0.849	0.830	0.595	0.690	0.676
29	Rajasthan	0.589	0.618	0.695	0.646	0.662	0.647	0.580	0.627	0.591
30	Sikkim	0.607	0.606	0.636	0.645	0.786	0.769	0.674	0.694	0.703
31	Tamil	0.682	0.734	0.786	0.755	0.762	0.804	0.743	0.685	0.654
32	Tripura	0.542	0.543	0.641	0.592	0.582	0.676	0.574	0.562	0.605
33	Uttar Pradesh	0.466	0.509	0.621	0.595	0.474	0.593	0.480	0.480	0.451
34	Uttaranchal	0.621	0.643	0.673	0.700	0.599	0.702	0.601	0.586	0.615
35	West Bengal	0.448	0.394	0.435	0.433	0.485	0.575	0.506	0.475	0.491

to measure of statistical significance. A high positive z score would prefer the surrounding features that have similar values (either high values or low values) and low negative z score (critical value as ±1.96 or 2.58) indicating a statistically significant (0.05 level or 0.10 level) spatial outlier consistence.

Hotspot analysis

Hotspot and cold-spot analysis is performed to delineate the spatial cluster of upper primary education level in India based on Getis-Ord Gi* statistic using fixed distance band in ArcGIS software (Mitchell 2005). The resultant Z score identified the states having the high or low values of cluster spatially. For statistically significant The positive and larger Z scores indicated the more intense the clustering of high values (hot spot) and negative and the smaller the Z score signified the more intense the clustering of low values (cold spot). A z score near zero indicates no apparent spatial clustering. The hot spot analysis tool calculates the Getis-Ord Gi* statistic for each feature in a dataset. The resultant Z score tells you where features with either high or low values cluster spatially. This tool works by looking at each feature within the context of neighbouring features.
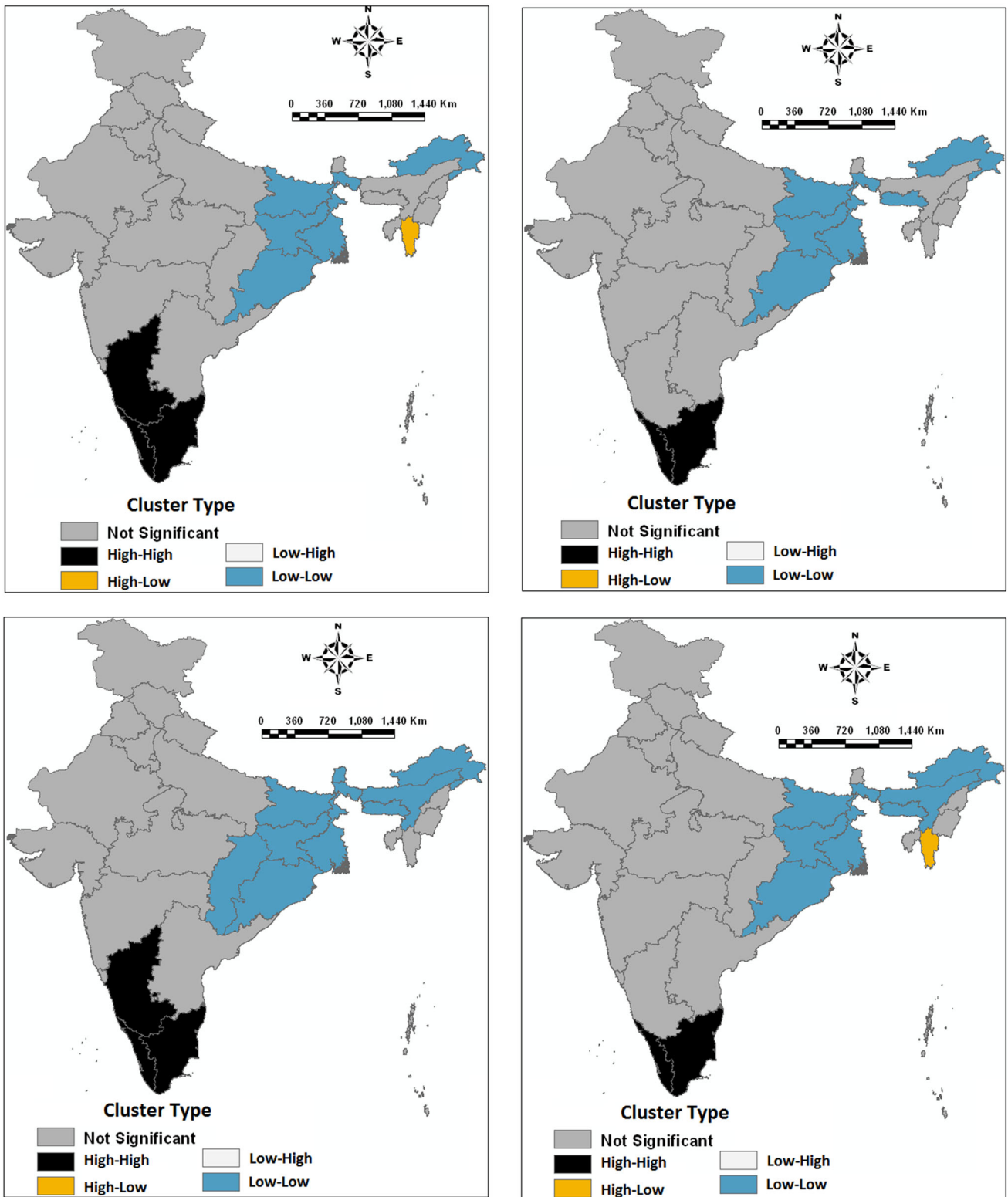
Fig. 1 Spatial cluster among the states of India at upper–upper upper primary level educational development based on inverse distance weightage of the academic year 2005–06 to 2013–14
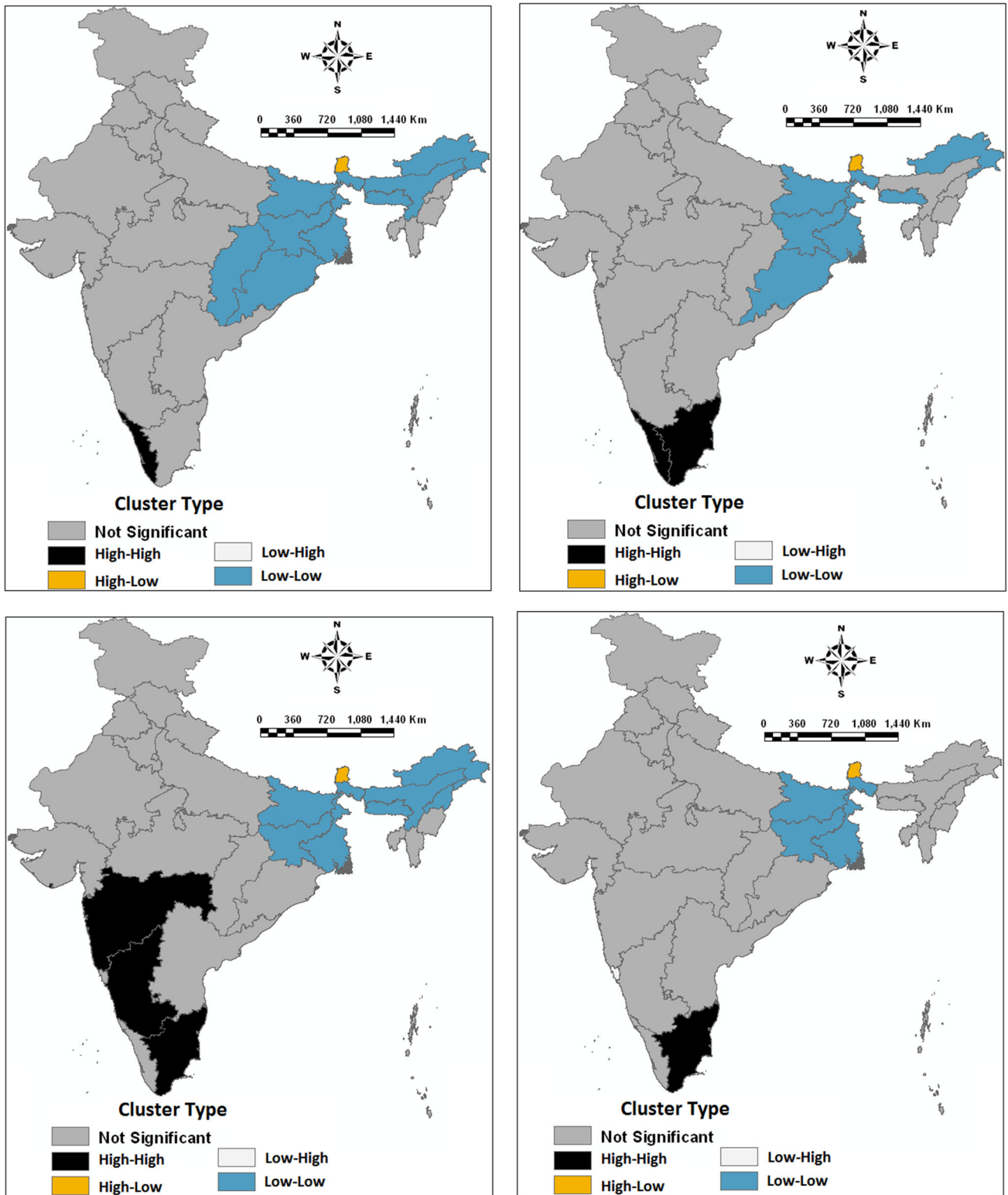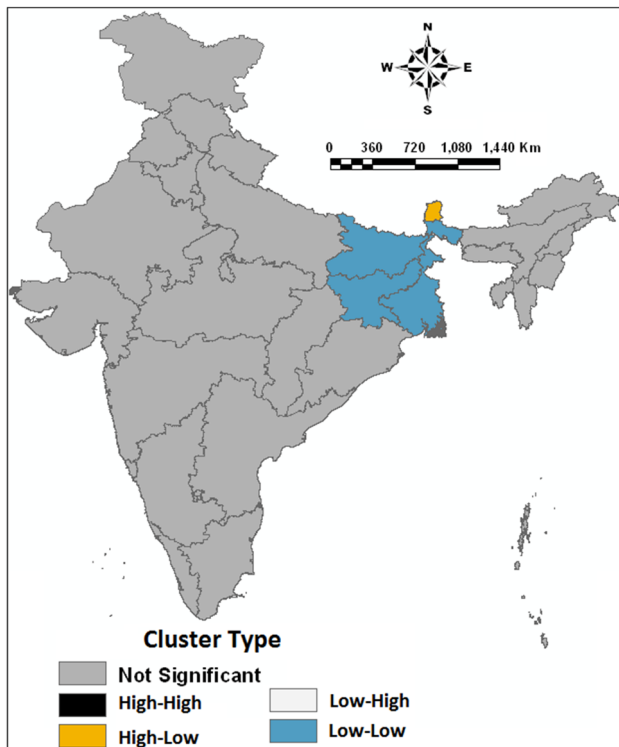
Fig. 1 continued

Fig. 1 continued

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \overline{X} \sum_{j=1}^{n} w_{i,j}}{\sqrt{\frac{s\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\overline{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\overline{X} \right)^2} \quad (3)$$

The G_i^* statistic is a z-score so no further calculations are required.

## Results and discussion

Aggregate of states with lower or higher values are easily detected in the local cluster analysis. Significant small clusters of educational development index were spread around the entire district (high–high); however, results also showed no significant spatial auto-correlation or a pattern

Table 2 Analysis of high/low clustering for different years using Getis-Ord Gi* statistics

Year	Variance	Z score	P value	Pattern
2006	0.0006	10.94884	<0.01	Clustered
2007	0.00309	9.11725	<0.01	Clustered
2008	0.00127	4.76506	<0.01	Clustered
2009	0.00314	4.35325	<0.01	Clustered
2010	0.00068	9.84952	<0.01	Clustered
2011	0.00158	3.71121	<0.01	Clustered
2012	0.00140	6.64918	<0.01	Clustered
2013	0.00152	5.98105	<0.01	Clustered
2014	0.00435	9.48813	<0.01	Clustered

of clustering of low values (low–low) in India during the study period shows in Fig. 1. There are other identified clusters, all associated with p-values greater than 0.05, which are not presented in the results shows in Table 2. The presence of extreme p values only in this current study (less than 0.01) show clearly that statistically significant clusters are identified. The result of our analysis showed Tamilnadu, Andhra Pradesh and Kerala facilitated by highly developed stage of educational condition; whereas, Bihar, Jharkhand, Odisha, West Bengal and Arunachal Pradesh showed very backward stage of educational development during 2005 and 2006. In between 2006 and 2007, Kerala and Tamilnadu portrayed highly developed stage of educational condition. Consequently, Bihar, Jharkhand, Odisha, West Bengal, Arunachal Pradesh and Meghalaya depicted backward stage of educational development. In 2007–2008, entire part of north-east India illustrated backward stage of educational development. During the period between 2011 and 2012, Andhra Pradesh, Tamilnadu and Maharashtra portrayed highly developed stage of educational development. However, in 2013–2014, Bihar, Jharkhand and West Bengal illustrated low level of educational development. There were some outstanding spatial clusters of educational development indices covering specific locations. Each hotspot analysis of educational development indices showed statistically significant hotspots and cold spots (P < 0.01) shows in Fig. 2. The result of the analysis showed clear spatial patterns of highly educational development condition that were mostly spread south and south-west part of India since 2005–2010. Some hotspots were also portrayed in the southern part of India from 2012 to 2014. However, most of the cold spots were found in the north east part of India. Some of the cold spots were also portrayed on the northern part of India during the period between 2007–2008 and 2010–2011. The map for 2005–2007 shows hotspot in the southern and in small pocket of western part; however, the
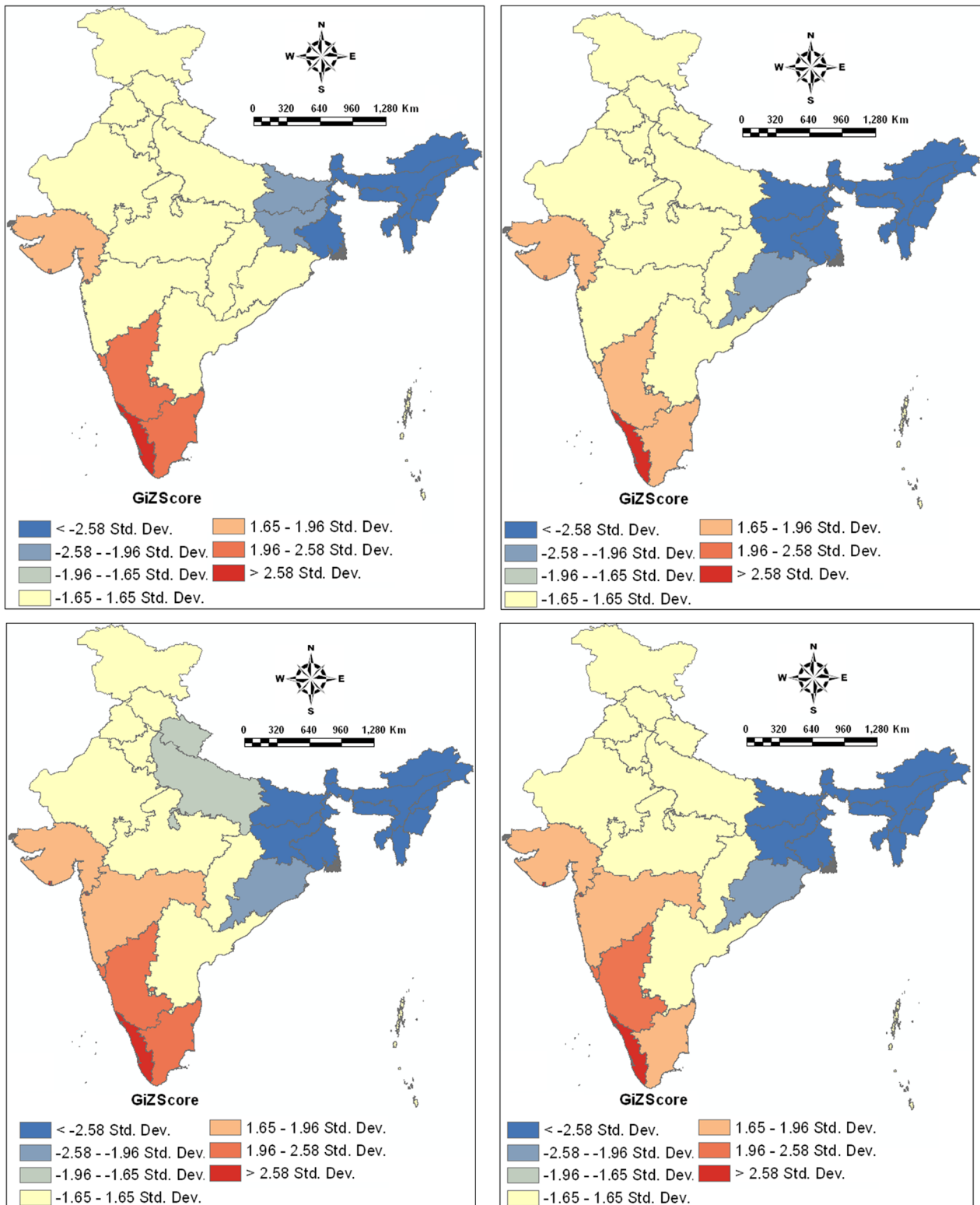
Fig. 2 Maps showing the hotspot and cold spot over different years (2007–2011) in India

cold spots were found in the north-east part of India. The map for 2007–2008 showed most of the hotspots were found in the west and south-west part of India, however,

the similar backward stage of educational development indices cold spots were observed as in earlier period and also extend towards the northern direction. In 2009–2010,
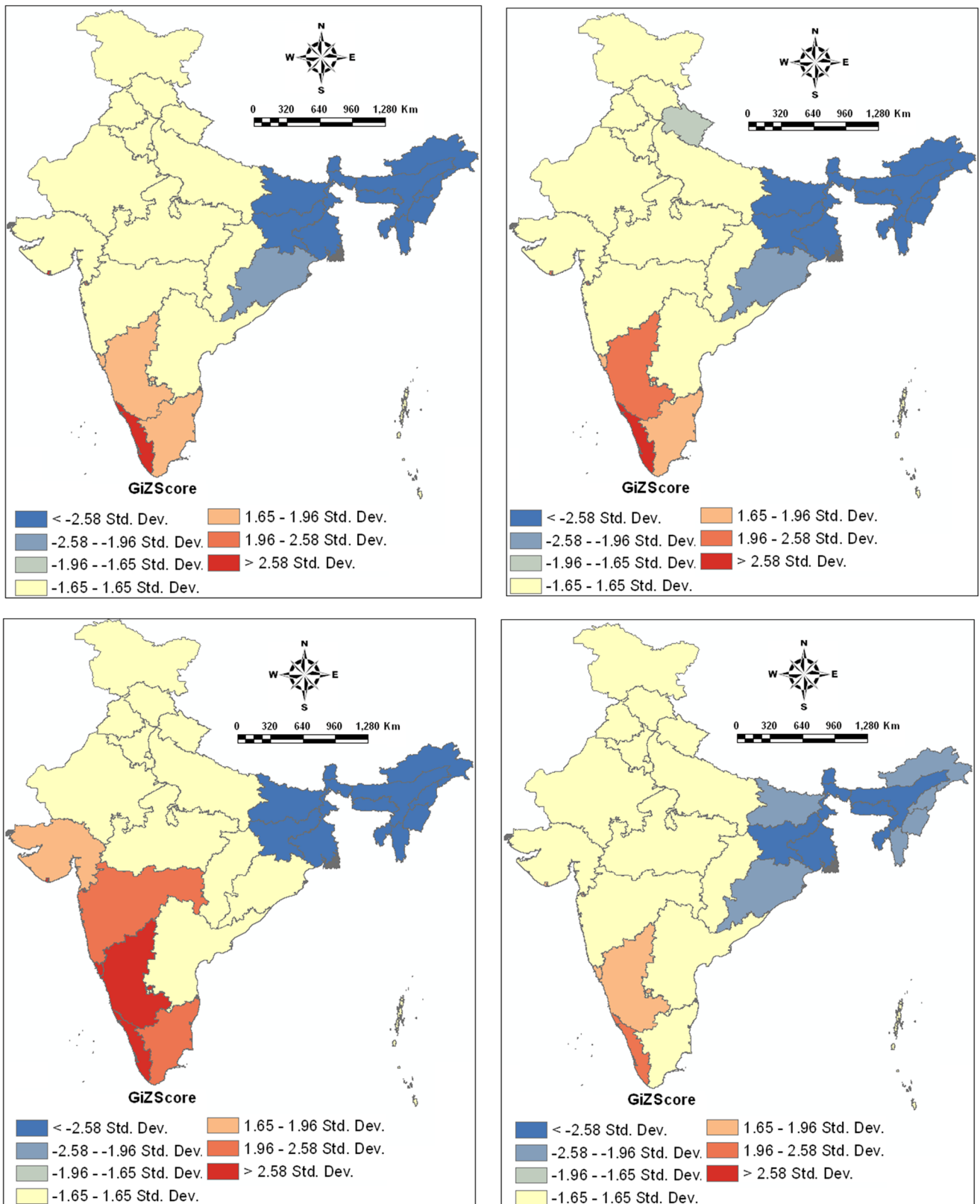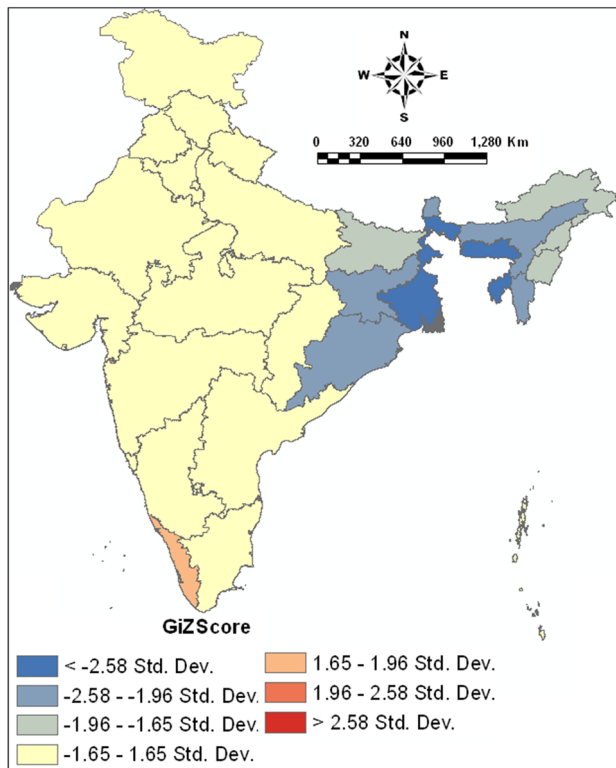
Fig. 2 continued

Fig. 2 continued

there was a slight decrease of hotspot of educational development condition states in southern part. In 2011–2012, there were increased of hot spot in the south and western part of India. In between 2012 and 2014, there were dramatic decreases of hotspot and cold spot areas in the southern and north-east part of India respectively.

Conclusion

The result of this study show that developmental measures can be used to make out high and low risk areas more efficiently which aids in enriching developmental plans and policies in upper primary educational level. The present study provides planners with a more sophisticated tool to differentiate risk patterns of educational development using spatial statistical approaches rather than relying on annual cumulative incidence alone, so that high-risk areas can be expansively recognized untimely in the low educational development based on their integrated spatial–temporal profiles. Some limitation of this study should also be highlighted, viz. despite the relatively high coverage of geo coding of educational development cases; it is likely to have been worse in peripheral states of the country. Second, whole analysis has been performed based on the state wise data. Despite these limitations, the study showed GIS

and GIS based spatial statistical techniques may provide an opportunity to clarify and quantify the development situation of development on upper primary level within re-develop the low developed areas, and lay a foundation to pursue future investigations into the socio-economic factors responsible for the educational development index (EDI). Spatio-temporal diffusion patterns and hotspot detection may offer useful information to sustain low EDI to control and predict development on upper primary level spread over critical hotspot areas only rather than for a whole region. Further, the methodology is based on notions on general principles of spatial statistics may employ the model to plan a strategy to control development on upper primary level by the information received on distribution and hotspots for various annually.

References

Census of India 2011; Household Schedule-Side A (2011) Government of India. Retrieved 22 January 2011

DISE Flash Statistics: Elementary Education in India: progress towards UEE for the year, 2005–2006. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi. India. http://www.dise.in/Downloads/Publications/Publications%202005-06/Flash%20Statistics.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year, 2006–2007. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi-110016, India. http://www.dise.in/Downloads/Publications/Publication%20200607/Flash%20Statistics%202006-07.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year, 2007–2008. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publications%20200708/Flash%20statistics2007-08.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year, 2008–2009. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publications%20200809/Flash%20Statistics%202008-09.pdf

DISE Flash Statistics: Elementary Education in India: progress towards UEE for the year. 2009–2010. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publications%20200910/Flash%20Statistics%202009-10.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year. 2010–2011. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publications%202010-11/Flash%20Statistics-2010-11.pdf

DISE Flash Statistics: Elementary Education in India: progress towards UEE for the year. 2011–2012. National University of Educational Planning and Administration (NUEPA) 17-B, Sri

Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publications%202011-12/Flash%202011-12.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year. 2012–2013. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Documents/Flash%20Statistics%202012-13.pdf

DISE Flash Statistics: Elementary Education in India: Progress towards UEE for the year. 2013–2014. National University of Educational Planning and Administration (NUEPA) 17-B, Sri Aurobindo Marg, New Delhi—110016, India. http://www.dise.in/Downloads/Publications/Publication2013-14/FlashBook2013-14.pdf

MHRD, Department of School Education and Literacy (2004) http://mhrd.gov.in/

Mitchell A (2005) The ESRI guide to GIS analysis, 2nd edn. ESRI Press, Redlands

Raju BMK, Singh A, Bansal (2008) Development of educational development index (EDI). Research Project Report, National Council of Educational Research and Training, New Delhi

Scott L, Warmerdam N (2005) Extend crime analysis with ArcGIS spatial statistics tools in arc user online, April–June 2005

Sengupta A, Pal NP (2012) Assessing the primary schools—a multi-dimensional approach: a school level analysis based on Indian data. Int J Educ Develop 32(2):264–272

Sengupta A, Sengupta M, Ghosh M (2004) Economics of learning: a village level study based on midnapore (west) district, West Bengal. Paper presented at National Seminar, Vidyasagar University, West Bengal, 30–31 March 2004

National University of Educational Planning and Administration (NUEPA), established by the Ministry of Human Resource Development, Government of India. http://www.nuepa.org/