Authors' Response to the Reviewer#2 Comments

Manuscript Ref.: hess-2017-384

Title: Population Growth – Land Use/Land Cover Transformations-Water Quality Nexus in Upper Ganga River Basin

Authors: Anoop Kumar Shukla, Chandra Shekhar Prasad Ojha, Ana Mijic, Wouter Buytaert, Shray Pathak, Rahul Dev Garg and Satyavati Shukla

We sincerely thank the reviewers for offering their critical comments and valuable suggestions that has helped to improve the manuscript. We hereby provide our responses to the reviewer's comments and highlight the changes made in the revised manuscript based on the comments provided. These have been incorporated in the revised manuscript as follows. The point wise replies of the comments of the Reviewer#2 are given below:

General comments:

The authors have a clear understanding of his topic area and have applied remote sensing and data collation techniques to answer investigate the impacts of demographic changes on water quality in the upper Ganga River Basin. However, I believe considerable work is needed to bring the paper to publishable standard.

We are grateful to the anonymous reviewer 2 for his/her efforts in providing extremely useful comments that has helped in improving our research work. The comments provided by all the three reviewers are duly complied in the revised manuscript which we believe have significantly improved the manuscript.

Major concerns are threefold:

Comment 1: What is it about this paper that is academically novel? Is it the application of existing methods to a new area? Or the evaluation of new methodology? The authors need to make this much clearer as the current introduction suggests that the paper aims to identify drivers; however the method takes drivers as given and the conclusions focus on the utility of the method…which aspect of the research are important to academia as a whole?

Response 1: Authors sincerely thank the reviewer for pointing out this very important question. We agree with reviewer that the previous manuscript required extensive restructuring and editing which is done in the revised manuscript. The main aim of the study is to analyse the *causative connection* (nexus) between the changing patterns of population, Land Use/Land Cover (LULC) and water quality of water stressed Upper Ganga River basin. In this study a comprehensive set of analyses are presented to assess and comprehend the current status of the population-LULC-water quality nexus in the study region, with respect to their changing patterns from 2001 to 2011. The present study is conducted at two different spatial scales i.e. (a) at river basin level (small scale), and (b) at district level (large scale) for three different seasons viz. pre-monsoon, monsoon and monsoon seasons. Such study is not done before for Upper Ganga River basin. Various methodologies are developed to study effects of LULC changes on water quality. But these methods cannot be applied directly to a region because of the differences in the data availability, climatic, topographic and LULC variations which may introduce errors. Hence, necessary modifications are made in the existing evaluation methodology as per the requirement. And a relationship is developed between LULC and Overall Index of Pollution (OIP) using multi linear regression. Findings from this research work may help engineers, planners, policy makers and different stakeholders for sustainable development in the river basin. The novelty of the work is discussed in detail in Response 4 of this draft.

Comment 2: There is an imbalance in the level of detail applied throughout. Some areas provide too much detail (I don't think the equation for growth rate needs five lines of explanation and an equation) whereas some key aspects of the methodology (e.g. radiometric correction) aren't described at all and there is little critical interegation of the results (why are they the way they are – what factors contribute and what doesn't). There is a need to work on structuring the data to make it make better sense.

Response 2: Authors are sincerely thankful to the reviewer for pointing it out. The revised manuscript is thoroughly edited and modified as required. The redundant information like extra figures, tables and texts are removed wherever they are suggested. The paragraph discussing the growth rate is trimmed down and the equation is removed. As suggested the methodology section including the radiometric correction is elaborated as a whole. It is described in detail in the Responses 1 to 6 (specific comments) of Reviewer#1. The data is restructured incorporating the comments from all the three reviewers and results are described in detail wherever necessary mainly focussing on cause and effects.

Comment 3: The paper needs a thorough restructure, as it is repetitive, provides considerable amounts of extraneous material and doesn't clearly signpost what is relevant to read. The language would benefit from a thorough proof read by a native English speaker also.

Response 3: As per the reviewer suggestions, repetitive and extraneous material is removed and the write up is trimmed down just showing relevant information and to the point explanation of the data presented. Authors have corrected the missing words and grammar, improved the phraseology and checked the English in whole manuscript suitably. To properly restructure the manuscript and for better description of the results, some subsections are added as required. Hence, review suggestions regarding modification of structure of the paper are duly considered in the revised manuscript. Our endeavour will be that the revised paper is much better than the current version.

Technical/Specific comments:

Comment 1: There is too great a level (e.g. population figures to the person on page 18. Round to the nearest 1000?)

Response: As suggested by the reviewers, all the population figures used on page 18 are rounded off to the nearest 1000 and updated in the revised manuscript.

Comment 2: Check English.

Response 2: The whole manuscript have been suitably modified and the English is duly checked and corrected wherever it was required.

Comment 3: Identify a clear research question.

Response 3: The primary research question answered in this work is as follows: "What is the causative connection (nexus) between the changing patterns of population, Land Use/Land Cover (LULC) and water quality of water stressed Upper Ganga River basin"?

Comment 4: Explain why it is novel – this is very important and does not come across well in the current draft.

Response 4: Ganga River is extremely significant to its inhabitants as it supports various important services such as: (i) source of irrigation for farmers in agriculture and horticulture; (ii) provides water for domestic and industrial purposes in urban areas; (iii) source of hydro-power; (iv) serves as a drainage for waste and helps in pollution control; (v) acts as support system for terrestrial and aquatic ecosystems, (vi) provides religious and cultural services; (vii) helps in navigation; (viii) supports fisheries and other livelihood options, etc. (Amarasinghe et al. 2016; SoE report, 2012; Watershed Atlas of India, 2014). However, for the past few decades Upper Ganga River basin has experienced rapid growth in population, urbanization, industrialization, infrastructure development activities and agriculture. Due to these changes, maintaining the acceptable water quality for various uses is being challenged. Therefore, there is a need to study the causative connection (nexus) between the changing patterns of population, Land Use/Land Cover (LULC) and water quality at both river basin (small scale) and at districts level (large scale) for three different seasons. Such study is yet to be done for this large river basin. OIP developed specifically for Indian context is used in this study to assess the status of water quality across the study area. Due to unavailability of the continuous population, satellite and water quality data at desired interval, establishing the interrelationship between these factors is not trivial. Hence, in order to achieve the objectives a comprehensive set of analyses are performed in this study. Between LULC and OIP a 2-time slice analysis is done for the years 2001 and 2012 with seasonal component. A relationship is developed between LULC and OIP (Indian WQI) using correlation and multi linear regression analyses. Further, trend (Mann-Kendall method) analysis was performed on monthly water quality parameters of the monitoring stations from 2001 to 2012 to understand their temporal variations over the years. Also, it was interesting to see the effects of seasonal variations on status of water quality. Hence, finally the results were inferred from these comprehensive set of analyses to understand nexus between population-LULC-water quality of Upper Ganga River basin which is our main contribution.

Comment 5: Find a clear argument that flows throughout the paper and only select figures and data that make it easier for the reader to understand this argument. E.g. Remove superfluous data such as the city populations on page 18.

Response 5: As per the reviewer suggestion, whole manuscript write up has been improved wherever required. Our endeavor will be that the revised manuscript is much better than the current version. From page 18 (old manuscript) superfluous data have been removed and used rounding to the nearest values. Authors are grateful to the reviewer for giving suggestions to follow a clear argument throughout the manuscript. This point greatly helped to improve the revised manuscript. All across the manuscript the following argument was followed: "There is a causative connection between the population-LULC-water quality of Upper Ganga River basin and they are interrelated with each other". Only those tables and figures are used that justify the given argument and helped to explain the results. City population data given on page 18 is removed as suggested.

Comment 6: Section 5.1 could be summarised in a paragraph of text. I over-simplify but much of it can be covered by the following sentence: "Growth rates for urban and rural areas were calculated from official statistics (Figure 3)". Is the individual city data relevant? How does it fit to the overall argument? Would spatial/mapped data be more useful or relevant when compared with RS data? Figure 3 simply repeats statistics shown in the text.

Response 6: Authors are sincerely thankful to reviewer for suggesting the summary of Section 5.1 as: "Growth rates for urban and rural areas were calculated from official statistics (Figure 3)". It is incorporated in Section 5.1 and further suitable updations and modifications were done in this section. The statistics of Figure 3 which is repeated in the text is removed and justifications are given on why these changes are occurring. Write up is improved suitably wherever it is required.

The water quality of a monitoring station is highly influenced by proximity of a city due to domestic and industrial discharges into the river from urban areas. The water quality status of these monitoring stations is dependent on the type of activities undergoing in and around the city. For e.g. water quality of Ankinghat monitoring station is affected mainly by discharge of effluents from tanning industries, municipal discharges and solid waste disposal into the river. Source of water quality pollutants are both point source and non-point source. Pollutants from both urban and rural areas affect the water quality. In urban areas water quality is mainly affected by municipal discharges and industrial effluents. However, in rural areas it is mainly affected by agricultural activities. Harmful chemical compounds are used in the agricultural lands in the form of herbicides, pesticides, weedicides, etc. During heavy rainfall, runoffs generated from these fields discharge directly into the nearby streams and carry these chemicals to the stream which causes water pollution. Therefore, just considering city data is not very rational approach. In a study, buffer zones of different thresholds were created surrounding a water quality monitoring station to determine the dominant LULC class that affects the water quality of that particular station (Kibena et al. 2014). In this study our argument is to determine the causative connection between changing patterns of population, Land Use/Land Cover (LULC) and water quality of the study area. The population data is available at district level not at buffer level. Reviewer#1 has suggested using district specific LULC statistics to establish relationship with OIP.

Reviewer#1 (Comment 8): *I would recommend the addition of district specific land use change maps to help support your discussion. At present, it is impossible to visually relate the pattern of land use change to the water quality and population statistics because the scale of the mapping in figure 4 is too small.*

Districts selected in this study consist of both urban and rural areas. District wise LULC change is extremely helpful in comprehending the water quality changes at the local scale and to identify source of pollutants at a particular monitoring station. Whereas LULC changes at the basin can only give a broad outlook on the status of water quality of the river basin which is also very useful for some applications. Hence, districts were chosen as a unit and district wise population and LULC were related to OIP of the monitoring station to comprehend the causative connection between them. Yes, the spatial/mapped data are more useful or relevant when compared with remote sensing data. But monitoring stations in the Upper Ganga River basin are scarce. Therefore, over a relatively large study area the interpolation maps generated using OIP are not likely to provide very good comparison results with LULC changes. If the number of monitoring stations are sufficient or large then this method can be used. Hence, with the type of data available for this river basin, it will be better to relate OIP, with population and LULC of complete river basin and districts under study.

Comment 7: The remote sensing work seems well carried out. However more detail is needed on the interpretation of the confusion matrices etc. (what is confused with what?; why?; what does this mean for the interpretation of the results?)

Response 7: Authors are sincerely thankful to reviewers for appreciating our remote sensing work. With the suggestions from all the three reviewers we have further tried to improve it. In thematic mapping of remotely sensed data, the term accuracy is used typically to express the degree of correctness of a classified map (Foody 2002). The confusion matrix based accuracy assessment is a widely used approach that includes a simple cross-tabulation of the mapped class label against that observed on the ground (or reference data) for a sample of cases at specified locations. A simple random sampling of 649 pixels belonging to corresponding image objects were selected and verified against reference data (GCPs). As a rule of thumb, Congalton (1991) recommends a minimum of 75-100 sample points per category. In LULC classification, a good classification depends on the separability of the spectral signatures while preparing the training sets. It was found that it is difficult to distinguish between built up and wastelands. To avoid this confusion the built up is first masked out and then it is classified independently. Similarly, it is difficult to distinguish between forest and agricultural lands. But the uncertainties in their classification can be reduced by increasing the number of training samples. Snow and glaciers can be confused with wastelands or waterbodies sometimes. The spectral confusion may occur between any classes. Spectral confusion in the classes introduces uncertainties and errors in the classification. Therefore, selection of training sets should be done with care. After image classification, accuracy assessment results are presented in confusion matrix showing characteristic coefficients viz. User's accuracy, Producer's accuracy, Overall accuracy and Kappa coefficients. The User's and Producer's accuracy express the accuracy of each LULC types whereas the overall accuracy estimates the overall mean of user accuracy and producer accuracy. The Kappa coefficient denotes the agreement between two datasets corrected for the expected agreement (Gebremicael et al. 2017). Results from this study showed a good overall accuracy of 90.14% and Kappa coefficient of 0.88. All the LULC classes were classified with accuracy more than 83% and they were in almost similar range. Comparatively highest inaccuracy was observed in forest class (Table 7).

As suggested by the reviewer, the Section 5.3 on accuracy assessment is elaborated in the revised manuscript. Further, characteristic coefficients are described in details in Response 5 of Reviewer#3.

Table 7. Accuracy assessment of the 2012 LULC map produced from Landsat Enhanced Thematic Mapper Plus (ETM+) data, representing both the confusion matrix and the Kappa statistics

Classified Data	***Reference Data***						***Row Total***	***User's Accuracy (%)***	***Overall Kappa Statistics***
	Agricultural Land	**Built Up**	**Forest**	**Snow & Glacier**	**Wastelands**	**Water Bodies**			
Agricultural Land	**128**	0	6	0	3	0	137	93.43	0.88
Built Up	2	**96**	2	5	1	0	106	90.57	
Forest	11	0	**88**	3	0	3	105	83.81	
Snow & Glacier	0	4	1	**103**	2	1	111	92.79	
Wastelands	1	2	0	7	**82**	2	94	87.23	
Water Bodies	0	0	1	1	6	**88**	96	91.67	
Column Total	142	102	98	119	94	94	**649**		
Producer's Accuracy (%)	90.14	94.12	89.80	86.55	87.23	93.62			
Overall Classification Accuracy (%)	90.14								

Comment 8: There is too much detail in some areas (e.g. full description of equation for population growth rate; detail of full Mann-Kendall method etc.) Only add detail like this if it is needed to help the reader understand the method, or if there is new method development else use references. Much of the Mann-Kendal work in 5.4.1 should be in methods not results.

Response 8: Authors are sincerely thankful to reviewers for their suggestions. The full description of equation for population growth rate and details of full Mann-Kendall method have been removed. The paragraphs having the details of these methods are modified. As suggested, only appropriate references are given for their methods. Section 5.1 and 5.4.1 (old manuscript) have been summarised in a one paragraph and write up has been improved wherever required. Section 5.4.1 have been removed from results section and included in methodology section in the modified manuscript.

References

Amarasinghe, U. A.; Muthuwatta, L.; Smakhtin, V.; Surinaidu, L.; Natarajan, R.; Chinnasamy, P.; Kakumanu, K. R.; Prathapar, S. A.; Jain, S. K.; Ghosh, N. C.; Singh, S.; Sharma, A.; Jain, S. K.; Kumar, S.; Goel, M. K. 2016. Reviving the Ganges water machine: potential and challenges to meet increasing water demand in the Ganges River Basin. Colombo, Sri Lanka: International Water Management Institute (IWMI). 42p. (IWMI Research Report 167). doi: 10.5337/2016.212.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1), 35-46.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. Remote sensing of environment, 80(1), 185-201.

Gebremicael, T. G., Mohamed, Y. A., van der Zaag, P., & Hagos, E. Y. (2017). Quantifying longitudinal land use change from land degradation to rehabilitation in the headwaters of Tekeze-Atbara Basin, Ethiopia. Science of the Total Environment.

Kibena, J., Nhapi, I., & Gumindoga, W. (2014). Assessing the relationship between water quality parameters and changes in landuse patterns in the Upper Manyame River, Zimbabwe. Physics and Chemistry of the Earth, Parts A/B/C, 67, 153-163.

SoE report, 2012: http://www.ucost.in/document/publication/books/env-books.pdf. Accessed on 12 March, 2018.

Watershed Atlas of India, 2014, Ministry of Water Resources, Govt. of India. Accessed on 10 March, 2018.
http://www.indiawris.nrsc.gov.in/Publications/WatershedSubbasinAtlas/Watershed%20Atlas%20of%20India.pdf