HESS Research Article

hess-2017-99

Groundwater impacts on surface water quality and nutrient loads in lowland polder catchments: Monitoring the greater Amsterdam area

In the present work, the authors attempt to characterize the impacts of groundwater seepage on the polder network around Amsterdam by exploiting data from the dense network of groundwater and surface water monitoring in this area. The authors combine water quality monitoring data with other biophysical characteristics of 144 polders and take a statistical approach to bettering our understanding of sources, transport mechanisms, and pathways in this area. They conclude that groundwater is a major source of nutrients in this mixed urban/agricultural catchment. In particular, they note that elevated nutrient and bicarbonate concentrations in the groundwater seepage originate from decomposition of organic matter in subsurface sediments coupled to sulfate reduction and possibly methanogenesis. Their results suggest that groundwater-surface water interactions are important to nutrient dynamics in urbanizing delta regions.

The current work is important, as it attempts to tease out the relative importance of natural and anthropogenic sources of nutrients within the region and to elucidate why implementation of nutrient management practices may not effectively reduce surface water concentrations to target levels, particularly in urban areas. The approach used in the paper, which combines correlation analysis between surface water and groundwater quality, as well as statistical analysis of relationships between landscape characteristics provides an interesting perspective on the drivers of various solute concentrations in surface water.

The study does, however, leave some questions unanswered. First, in the abstract it is claimed that “land use” is used as a variable in the multiple linear regression, which attempts to identify the strongest drivers of surface water nutrient concentrations. In the analysis, however, the only land-use variable that I see is “paved area.” As the authors mention more than once that agriculture in the polder catchments could be driving surface water nutrient concentrations (and I would agree), I find it puzzling that this is not used as a potential variable for the regression analysis. Second, the authors average groundwater data taken over a period of more than 100 years, but do not discuss how groundwater levels may have change over time, and how these trajectories may have differed from place to place, thus affecting use of the GW data in the spatial analysis. Finally, it is unclear how issues of collinearity impact the results of the correlation analysis and development of the multiple linear regression model. A more complete treatment and subsequent discussion of possible collinearity between independent variables would strengthen the analysis.

**Specific comments:**

p. 6, ll. 8-12 You describe here the variables used in your analysis, but do not include any land-use variable other than “paved area.” Clearly, agricultural area is a major factor driving concentrations in your study area, so it seems a large omission to not include it in your analysis. Is it simply that the agricultural area was not included in the database that you utilized? If so, could you obtain that information through other sources of land-use data? It is possible that including agricultural area in your analysis would significantly change the findings of your analysis regarding significant drivers of surface water concentrations.

p. 6, ll. 14-18 In your methods, you mention that for each well, you average concentrations for each monitoring well (at individual monitoring screens) for all sampling dates. You also mention that the groundwater data is from the period 1910-2013—more than 100 years. I would assume that there could have been significant changes in groundwater quality over that period, and that the temporal patterns of change could have differed across the study period. Accordingly, is it correct to combine all sampling data across this 100-year period, or in doing so are you conflating spatial and temporal differences across the study area?

p. 8, ll. 20-30 You do not discuss here how you dealt with issues of collinearity among the explanatory variables. For example, there are clearly high correlations ($r>0.60$) among some of the groundwater solute concentrations (particularly with regard to $HCO_3$). With this being the case, how do you (from a quantitative perspective) make decisions regarding inclusion in the multiple linear regression model? For example, in your MLR equation for TP, you include both $HCO_{3(GW)}$ and $TP_{(GW)}$, although your correlation table in Table 1 shows a reasonably high collinearity ($r=0.68$) between these two variables. How do you justify use of both of them in the MLR equation?

p. 12, ll. 22-23 You say here that ammoniums correlates more strongly with TN than nitrate and conclude that ammonium is therefore likely the main form of TN in the study area. When I look at Fig. 5, however, it appears that nitrate is likely the dominant form of N in the ice-pushed ridge area (5) and possibly the Vecht Lakes area (4). It might be more useful to discuss the actual variations among locations (and reasons why), rather than just to cite the simple regression results.

p. 13, ll. 8-21 You discuss the results of the MLR analysis here, but do not reference the table that contains the results. Please include the table reference here.

Fig. 5 It is very difficult to understand the variations in concentrations of solutes among locations in these figures due to the different concentration ranges from site to site. For example, for TN, all of the concentration ranges look very similar, simply because you scale the y-axis to include all of the outlier values for site #5. Is it important to include all of the outliers? I would recommend plotting these in such a way that you allow the reader to understand differences in median and interquartile range values, rather than prioritizing the representation of outliers.

Table 1 For your correlation analysis, you should include the 1.0 values to show perfect correlation between two identical variables. This will help add structure to the table and make it easier to understand.