Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-169-RC3, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Determination of vadose and saturated-zone nitrate lag times using long-term groundwater monitoring data and statistical machine learning" *by* Martin J. Wells et al.

Sophie Ehrhardt (Referee)

sophie.ehrhardt@ufz.de

Received and published: 14 July 2020

The new paper "Determination of vadose and saturated-zone nitrate lag times using long-term groundwater monitoring data and statistical machine learning" by Wells et al., presents an innovative approach to estimate vadose-zone and saturated-zone lag times using long-term groundwater nitrate data. The use of statistical machine learning could be an alternative to expensive groundwater age-dating techniques and has the computational power to uncover nonlinear trends. Both are convincing arguments for the application of the Random Forest analyses and provide valuable information for

groundwater management. General comments Clearly written, but with some missing further information (see specific comments). Nicely explained method section! Especially for natives in ML very instructive. Interesting analysis and connection to previous results from that study area. Thank you for your work! Abstract Line 16: Could you add some information about which area/time/well number you averaged the mean? And you did not mention the name or location of the study area in the abstract to which all numbers correspond to. Try to add this to make it more precise and enable the reader to set the study in space. Line 27: Mention that denitrification plays no major role in the study area. Otherwise diffuse recharge could be affected by this process. Introduction Line 37: Please add a few sentences why research for nitrate contamination is important. Line 63/64: The explanations "vadose (unsaturated)" and "groundwater (saturated zone)" could be earlier in the paragraph e.g. Line 38. Methods Line 107: In which depths are shallow, intermediate and deep groundwaters? Even more important than the screen length. Line 123: I did not check the paper, but how can the mean recharge stay the same, if 88% of the rates decrease? Because of highly positive outliers? Line 133-174: Really nice explanation of the method and its principles! Line 203: How strong was the relation between "Area of planted corn" and "fertilizer aplication rates"? R2? Should be really high as you substitute the Ninput mass by an area. Line 204: More information on the reduction- perhaps in bracktes "from... to..." or "by ....%" to estimate the effect (or its potential as marker in case of drastic drop). Line 230: I am not sure, how to imagine the "apparent" travel time as I only know about distributions (gamma or log-normal) of TTs. Your TT is the peak TT without any parts of it travelling faster or slower? So, you don't assume a mixed signal stemming from TTs from different ages (e.g. in 2010 10% signal/NO3 load from 1990, 40% signal from 1991, 50%...)? Line 234: Please, define shallow! Line 252-255: And the fertilizer input (Nsurplus) of 1990? Isn't this the most important input variable? Perhaps already cleared by Line 203, when adding R2. Line 263: "historical nitrate groundwater concentrations" or do you mean historical Ninput data? Results Line 292: I struggle to understand you differentiation between TTs and evolution of NO3. You don't use

NO3 as tracer to derive TTs and therefore you can correlate both? Or don't you use NO3 to derive transport rates? If you calculate one variable based on the other, isn't the correlation useless? Sorry for my confusion. You concept of TTs is quite different from ours. Line 332: Doesn't your canal leakage has also high NO3 from time to time, based on surface runoff from fertilized fields directly (pipes and drainages)? And can you add some information on the canal system previously? Is it also to drain the fields? Line 332: Why does influence of canals extends further from the canal? Isn't its influence decreasing with distance? Line 337: "nitrate reduction" add (also known as denitrification)? Line 338: "The partial dependence plot" add (Fig. 7) Line 342: I am surprised about your conclusion regarding the rapid aquifer response. You mention stratification and a groundwater age of 7years. Doesn't this account for a dampening of changing signals? Or what time do you assume with "rapid"? Or does this only correspond to the shallow, unstratified groundwater? Line 355: Do you have a recommendation how many data (stations) we need or how long time series should be to use your ML approach? Line 361: Isn't your "may be biased" a bit to optimistic? How can you distinguish a vanished NO3 imprint after denitrification from "stored somewhere in the upper soil"?

Figures Line 584: Is this pattern clockwise? Don't you need to switch the lower plots then? Line 594-595: Thanks for the explanation again! Line 597-600: Is there a difference between %inc and %Inc? It is not consistent in all figures. Line 622: Is there a space missing at "bData required further analyses"? Line 625: Why only "some models were ultimately based on <1049 obs"? According to your table all models fit the condition "<= 1049" and some "= 1049 observations".

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-169, 2020.