The paper uses a large data set of soil ammonium and nitrate concentrations and attempts to correlate these values with various soil properties and with land use. Although the paper does not represent novel concepts, it uses a state of the art analysis and a large data set. Given the size of the surface under investigation, the data set is however relatively sparse (as acknowledged by the authors). Nonetheless, the study represents a contribution to scientific progress and an important basis for the investigation of large-scale drivers of soil N, thus in my opinion warranting publication. I have however a couple of main concerns with the paper, which should be addressed before publication. The analyses are clearly outlined and assumptions seem to be valid (with one exception, as noted below). The paper is well structured and generally well written, although there are several cases where the text is unclear; these exceptions are outlined below. Additionally, the conclusion section contains some terms that are not addressed in the paper and I have suggested their removal (see below).

Main comments

I have two main concerns with the manuscript.

Firstly, the authors state that they are investigating drivers of soil ammonium and nitrate. This is indicated by the language used throughout the study ('controls' and 'drivers'), including in the aims section. The study is however an observational study, meaning cause and effect cannot be derived from these results; the 'drivers' of soil ammonium and nitrate cannot be identified from such a study, only correlated variables (or 'patterns'). It would be acceptable to state that this study aims to identify candidate soil properties that might be considered, following further study, as 'controls' or 'drivers'; the study cannot however identify these controls and drivers itself. Such wording would agree with the fact that this study represents a basis for further studies –as indeed the authors state several times.

The text throughout the manuscript (including the title) needs to be corrected to reflect this. Words such as 'controls' and 'drivers' need to be avoided.

Secondly, it is unclear to me why the authors have split the soil samples into the two regions ('agricultural' and 'ecological'), given that within the agricultural region there seems to be a large variation in the intensity of land management:

- Would it not make more sense to use the actual land use of the 469 sites as a factor, or to derive a scale of land management intensity from the 28 land use types, and examine the correlation of this with the soil attributes?
- The authors state that they aim to assess the relative importance of large-scale drivers, which I assume is the purpose of using two large regions (agriculture and ecological). However, I suspect these two regions are inadequate to do this: given that the two regions encompass very large climatic and geological variation, important large-scale potential 'drivers' are not addressed by this method either. The authors could consider incorporating broad climatic information into the analysis. This may reduce some of the noise and thus improve the outcome of the analyses.

I have a number of additional minor concerns with the paper that need to be addressed:

P3 L15-16: I appreciate that the sampling design of the BASE project is described elsewhere, but it would be useful to have a little more detail on this, including the number of soil samples taken from each 25m x 25m 'site'.

P3 L 27-33: The soil maps used were a result of spatial modelling. The outcome of a model cannot be considered as data; please therefore change the word 'data' (L27 and L32) to 'values' or 'information'.

P5 L8-10: The values given in these lines (sum of NH_4^+ and NO_3^-, NH_4^+ and NO_3^-) seem to refer to mean values (for the first of this set of values, it is indeed stated so). Would it not make more sense to give median values here, given that median values are what are shown on the corresponding graphs (figure 2)? The mean average value of a population that is not normally distributed is not particularly informative. Additionally, given that the NH_4^+ and NO_3^- concentrations are not normally distributed, stating the standard deviation of these data is misleading, as the use of a SD value to convey information assumes the population is more or less normally distributed.

P5 L 19-13: Why carry out an analysis omitting the samples for which NO_3^- concentration was below the detection rate? If a soil sample has a NO_3^- concentration below the detection level, this does not equate with 'no data', but rather means that the NO_3^- concentration is simply very low (as the authors indeed assume). Unless something specific is being tested, which I do not think is the case here, an analysis with these points removed is uninformative.

This subject is re-visited in the discussion (P7 L11-12) but as the text is written, I still do not understand what information this analysis brings. The results of this extra analysis do however indicate that the agricultural soils have a bimodal distribution with respect to NO_3^- concentration, i.e. many soils have very low concentration and many soils have a very high concentration (indicative of high addition rates of NO_3^-). If this additional analysis was carried out to illustrate this, the authors need to make this clear in the discussion, and indeed expand this point in the discussion.

P7 L18-19: What is meant here by this sentence, particularly the term "agricultural soil"? Do the authors mean that not every site in the agricultural region is under agriculture? If so, please change accordingly and change the terminology to 'soil under agriculture'. If not, please explain the term 'agricultural soil'.

P8 L2-3: This sentence is either incorrect (as I have understood it) or imprecisely written: I understand from this sentence that the NH_4^+ concentrations between the soils from the agricultural and ecological regions are different. According to figure 2 and text in the results section (P5 L8-9) however, NH_4^+ concentrations are similar. I suspect a more complex pattern is meant by the authors; this needs to be more clearly explained.

P9 L 18-25: The text here is difficult to understand. I have a few suggestions that might help:

L 19: Replace "by each specific region" by "within each region" (if I have understood correctly).

L20: Replace "on each region specific basis" with 'for each region" (if I have understood correctly).

The sentence L20-22 is very unclear. Is it referring to the higher prediction error for the high concentrations of NO_3^- (in the ecological region in particular)? Rewrite.

L23: "presence of small values" is too vague. Do the authors mean that the high frequency of samples with very small NO_3^- concentrations is the cause of the limited overall model performance? Needs to be explained.

L24: Replace "most of model errors" with "much of the model error"

L24-26: This sentence is too vague. This sentence relates to the presence of small values of What is meant be "the limited data set" exactly?

P10 L7-9: The authors here imply that they have identified a process in the results, the "potential to maintain or increase NH4+ concentrations". This process is a possible explanation of the results they have found, but is not in itself a result. Please change text accordingly.

P10 L33-34: The term "management" should be avoided here, unless the authors specify what they mean by management, as a term separate from and in addition to 'land use'.

P 11 L 2: The term "human modification" should be removed; this term implies some sort of scale or land use intensity (e.g. nutrient input levels), but this has not explicitly been investigated in this study.

Figure 1: Three main regions are shown here, whereas two are considered in the text. I recommend that the number of regions considered should be consistent. Alternatively, if the three ecological regions were considered distinct enough to warrant their separation on the map, why not use three regions in the analysis?

Technical corrections

P1 L11: It is unclear what the 'other' ecological region refers to.

P3 L17-22: The sentences from L17 to 22 need to be moved out of this section; I suggest to the discussion.

P5 L5-6: In the first sentence of the results, the mean NH4+ and NO3- concentrations are stated, referring to figure 2. However, in figure 2, the median concentrations are given. Please correct text (or change figure 2) accordingly.

P5 L25: Change "large" to "high"

P5 L26: "in that" needs to be inserted between the words "…..environments or" and "used mainly…..".

P6 L16-17: This sentence belongs in the discussion.

P7 L8: Remove "In our case".

P7 L7: Change "which may suggest" to "which suggests"

P7 L22-23: Change "…which was considerably low in the soil from the agricultural region compared to the ecological region" to "… which was considerably lower in the soil from the agricultural region compared to that in the ecological region".

P10 L6: Change "difficulties to directly measure" to "difficulties in directly measuring"

P10 L10: Change "also indication that' to 'also an indication that"

P10 L11: Unclear. What depends mostly on soil mineralogy? Soil input stoichiometric ratios or final soil elemental ratios?

P10 L28-29: Unclear. I think this sentence needs to be re-written as: "Therefore, the importance of soil elemental interactions in determining the variation of mineral N at different spatial scales across and within various cropping and ecological conditions needs to be estimated." I have possibly interpreted this sentence incorrectly. If this is the case, what is "it" (L28)?

Figure 2: Please clarify, do the error bars shown represent the standard deviation of the data values i.e. population, or the standard error of the mean?

Figure 4: Is 'standard error of the (estimated) mean' meant here?