# Editor

**Where possible authors should design all graphics with greyscale reproduction in mind, being aware that greyscale reproduction of colour images can lead to ambiguous use of greys.**

*Thank you for your comment, however, we feel that the use of colour figures is justified, and that it makes a lot of concepts easier to understand If deemed necessary, we can update the figures into a grayscale form in the later version.*

**PLEASE check your English spelling and grammar. The paper will be returned if this is not done.**

*We conducted a thorough English spelling and grammar check and made changes throughout the manuscript.*

**Reviewer 2 also notes that RF is not an AI method. This is correct - it is a machine learning method.**

*We now only say that random forest is a machine learning method.*

**Please try to avoid making the paper longer (keep in mind that the average length of papers in this journal is 18-20 published pages).**

*We have rewritten several paragraphs more concisely to try to keep the length of the paper down. The length of the paper including references has been reduced from 25 to just over 22 pages despite the introduction of a couple of new figures, and the word count has dropped from 6128 to 5360.*

# Reviewer 1

**I suggest not to declare the unknown location of ditches a 'research gap', maybe this rather can be phrased as a research objective**

*Good point, this has now been amended by rephrasing the paragraph 2 of the Introduction section. We have also added some additional motivations for the research objective by including ditch/channel mapping statistics for Sweden:*

*Method 2.1. The need for better small scale channel mappings:*

*entire paragraph*

*Results and Analysis - paragraph 1:*

*entire paragraph*

*Discussion - paragraph 1:*

*line 1*

*Discussion - paragraph 3:*

*entire paragraph*

*A new figure (Figure 10) has also been added.*

**p2/17ff .. the performance assessment of methods should be based on references or own empirical evidence**

*We have moved these explanations from the Introduction to the “Extracting ditches with digital terrain indices” section and clarified that our understanding of the strengths and weaknesses of the individual indices are based on our own observations of their performance for ditch detection.*

*This is clarified in Extracting ditches with digital terrain indices, opening paragraph:*

*last sentence*

*The individual descriptions of the digital terrain indices have been updated with the weaknesses of each index:*

*2.4. Extracting ditches with digital terrain indices, (Sky View Factor - line 3, Impoundment Index - last sentence, and High Pass Median Filter - line 3)*

**ground truth: more information on length/depth/width of ditches would be desirables, beyond 'on average wider than 0.5m'**

*We have expanded on this in 2.3 Digitising the ground truth - paragraph 2:*

**aren't road ditches typically on both sides of a road?**

*We need some clarification on what needs to be corrected with this comment (what paragraph etc.). It is true that road ditches are typically on both sides of the road but sometimes road ditches are only dug on one side of the road if the other side slopes steeply away from the road. However, we can’t find any inconsistencies with this in our article.*

**any depression is considered a ditch, no distinction from the concept of a sink is made, which might be helpful with expressing the linear character of a ditch**

*Several approaches were taken to prune sinks incorrectly predicted as ditches.*

*We use gabor filters for several input variables. Gabor filters are specifically designed to detect linear features in images, and features using this technique proved to be very effective, as can be seen in the feature importance table (Table 4).The gabor filter explanation has been clarified in section 2.5.1. Processing the digital terrain indices - paragraph 4:*

*last sentence*

*Finally, at the post processing phase, our technique substantially removed sinks and small “clusters” of incorrectly predicted ditch pixels. We looked at both the size and shape of connected clusters of ditch predictions and corrected where the clusters were too small, or the shape was not elongated. We included these clarifications in the manuscript. This post processing step is also illustrated graphically in Figure 8.*

*Clarified explanation in 2.6.3. Cluster removal:*

*last two sentences*

**would sink filling assist with developing a suitable index?**

*Sink filling could probably work as a post/pre processing method. However, because of our own cluster removal algorithm (see previous answer), we do not deem the inclusion of a sink filling algorithm necessary. Our cluster removal algorithm is tailored to the specific problem of correctly detecting ditches, whereas a sink filling function is more general, and could possibly fill ditches erroneously.*

**p6/15 typo argGIS**

*This mistake has been fixed!*

**why is not curvature index used?**

*We have made other approaches at removing streams from the predictions. This is described in 2.5.1. Processing the digital terrain indices paragraph 5 (previously paragraph 7). As can be seen in the feature importance table (Table 4), these input variables work quite well. They do, however, have the downside of sometimes removing deep ditches as well, which can be seen in Figure 11 (f).*

*Comparing curvature index with our own, and other, stream removal methods would be an interesting addition in a future study, but was out of scope for this study. We talk about this potential future improvement (shape index) in section 4. Discussion, last paragraph.*

**would the concept of negative/positive terrain be helpful?**

*We use a high pass median filter (HPMF) to capture the curvature in the terrain. This is similar to negative/positive terrain. We feel that we cover this adequately in 2.4 Extracting ditches with digital terrain indices - High Pass Median Filter:*

*first sentence*

**p11/13 Python in upper case**

*This mistake has been fixed!*

**table 1: was there any sensitivity analysis done re inclusion/exclusion of variables? (ok, I see this partially answered in table 4)**

*We conducted an experiment where we used two subsections (that are not in the final experiment) and a cross validation method to evaluate each algorithm with varying feature spaces (testing several feature spaces from 81 down to 10 features). This helped in determining the correct machine learning algorithm. The algorithms that were examined are: Extreme Gradient Boosting, Naive Bayes, Support Vector Machines, and Random Forests. Random Forests was shown to be the most suitable algorithm.*

*To determine what input variables to use in the final experiment, we performed a more detailed experiment with Random Forests, in which we also included the post processing steps. The input variables were reduced by 2 at a time, from 81 down to 2. Random Forest’s feature importance was used to determine which variables to exclude from the previous iteration. This approach showed us that using 40 input variables yielded comparable results to using 60 or 70 variables. With this approach, we could remove unnecessary features that dragged the performance and speed down.*

*With these 40 variables, we also performed a detailed parameter tuning which included the post processing steps. The final experiment was then rerun using only these 40 input variables (down from 81), producing completely new results for the study. Adjustments have been made in several parts of the article due to this change:*

*2.5.1. Processing the digital terrain indices, paragraph 5:*

*This paragraph was removed, due to the removal of this input variable. Consequently, a reference to this section in 2.7.1. Noise reduction and gap filling, paragraph 2 was updated to function without referencing the removed paragraph.*

*We added a paragraph to explain the sub-experiment for variable selection in 2.5.1. Processing the digital terrain indices - paragraph 6:*

*entire paragraph*

*Of course, this also means new result metrics for the end experiment, as well as new feature importances and result illustrations. These results are better for some metrics and worse for other metrics. This is in part due to using fewer input variables, but also using a different random seed for the Random Forest algorithm. In the previous experiment, no random seed was set, meaning that the results were not completely reproducible. We now use a random seed to ensure the validity of the results.*

**overall, ditches typically have rather stable directional characteristics. This seems not to be considered in the chosen parametrisation (higher probability for a pixel representing a ditch on one direction over others)**

*See answer to the comment about any depression being considered a ditch, and no distinction being made from the concept of a sink (Reviewer 1).*

**similarly, is (vector) topology a potential useful approach to fill gaps and connect ditches?**

*Thank you for this interesting idea, but we feel that it is beyond the scope of this paper. We would certainly consider this in a future study.*

**p15/15 replace chance with random?**

*Chance has been replaced with random when referring to the Cohen’s Kappa metric.*

**in several places, replace 'cavity' with 'sink' or 'concavity'?**

*Good point, the word cavities has been replaced with the word sinks throughout the text.*

**it would be interesting to compare the ML/RF approach with using simple morphographic indicators, whether there is any significant improvement?**

*This was done in the study: we compare our method to the Sky View Factor, Impoundment Index, High Pass Median Filter, as well as Slope indices. Both Slope and HPMF (positive/negative terrain) are morphographic indicators. see Result section: paragraph 2 (entire paragraph), and Table 2.*

*We talk about this in the introduction, last paragraph. We have clarified this slightly by adding “compared to if a single index would be used”:*

*first sentence*

*We also talk about this in the discussion, first paragraph. We have also here clarified slightly by adding “than if indices are used separately”:*

*line 5*

# Reviewer 2

**The authors combine the information from all the digital terrain indices using Random Forests to improve the detection of the ditches. Why we should combine terrain indices and RF? Authors need to emphasize the significance and originality of this work in the introduction because terrain indices, and RF for ditch detection have done by other researchers.**

*We agree that the approach of the study needed clarification in the Introduction. We updated the introduction to highlight that the originality of our work is the development of custom input variables for the model, as well as custom post processing steps:*

*We feel that we have already motivated why we should combine indices and use Random Forests in Introduction, last paragraph:*

*first sentence*

**Random forest is a commonly used machine learning algorithm and in most cases are worked. So that be ok to use RF to detect ditches. But my concern is that I notice the author used 81 features to train the RF model and my question is that do all these features worked or not? More specifically, in table 4, the author list top 20 important input variables. And we can easily find out from the table that, the importance score of the top one feature (Impoundment mean 3) is much higher than the 20th feature (Slope non-ditch amplification) in the list. So, I guess the score of the least important feature could be extremely low (maybe 0.1 or 0.05 and something like that). Therefore, in that case, I doubt that it’s unnecessary to use all 81 features. Authors need to make more analysis about this problem. You perhaps can reference the following paper:**

**Georganos S , Grippa T , Vanhuysse S , et al. Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application[J]. GIScience & Remote Sensing, 2017.**

*We agree. See the answer to Reviewer 1’s comment regarding inclusion/exclusion of variables.*

**Section 2.4.2, you manually divided the dataset into 21 subsections. All the upper subsection was used to develop RF model (figure 1), and all the lower was used to validate. I doubt that using all the upper may cause systematic bias for the model. In general, the subsection must be divided and selected randomly for RF model, not manually. Please revise your model or explain why.**

*We probably have not explained our methodology clearly enough. The upper 10 subsections were not used to build a model that was evaluated on the 11 lower subsections. Rather, we used a cross validation approach, where each of the lower 11 subsections were used to evaluate on once with a model trained from the remaining subsections (i.e. 11 different models were trained for the experiment). The cross validation process has been clarified in subsection “2.5.2 Training and validation datasets - paragraph 2” by adding the sentence: “* A new model is trained using the training folds for each iteration (Figure 4).*”. A figure (Figure 4) was also added to graphically explain the cross validation approach with the 11 hold-out subsections.*

*We feel that subsection “2.5.2 Training and validation datasets - paragraph 2” explains clearly that the upper 10 zones are not used in the final experiment:*

*last three sentences*

*However, we changed the use of the word “model” to “models” when referencing the models used in the final experiment in a lot of places throughout the text to make it clear that multiple models are built for the final experiment.*

**the title can not represent this work well. A lot of methods for detection ditch from DEM have proposed. Perhaps you need to emphasis your feature.**

*We have now changed the title to include the word machine learning, to more accurately represent the approach of our study.*

**Page 2, line 29. You said you used artificial intelligence. Which AI are you used? RF is not AI.**

*We now only say that random forest is a machine learning method.*

**In section 2.3 (page 5). do you have any basis for the window selection of the terrain indices? Such as radius of 10 m for Sky View Factor, dam length of 3 m for Impoundment Index.**

*We based the window selections for terrain indices on our own expert judgment. We tested many different thresholds manually and it seems to be a big compromise to find reasonable numbers. Most ditches are not wider than 3 metres, and the impoundment index radius needed to cover the ditch sufficiently, without looking too far. The sky view factor radius was not as sensitive, but it also needed to cover the ditch. We set the limit to 10 metres to avoid the index being affected by steep hills in close proximity to ditches. Computational challenges (it takes two days to generate the sky view factor for our study area) also limited how much we could test different parameter thresholds.*

*We have updated the motivations slightly in 2.4 Extracting ditches with digital terrain indices:*

*Sky View Factor:*

*last sentence*

*and*

*Impoundment Index:*

*second last sentence*

**Section 2.4.1, page6, line 44. “This facilitated finding obscurities in the neighbouring areas around pixels”. The ‘obscurities’ is hard to understand. What is mean?**

*We agree. This sentence has now been rewritten to:*

*“Using these statistical aggregations with the help of neighbouring areas around pixels aided in pruning pixels with outlier values, often smoothing out the data to represent ditches more accurately on a per-pixel basis.”*

**Page 12, line 52. What is the custom function? Please specify.**

*We have removed the reference to “custom function” when rewriting in a more concise manner. This function is described in 2.7.1. Noise reduction and gap filling - paragraph 1:*

*last two sentences*

**Page 13, line 44-45. Why 6 \* 6 grid zone, and probability of 35 % are used?**

*We evaluated several different zone sizes (from 2\*2 to 12\*12 pixels), but 6\*6 showed the best results overall. We also noticed a numerical error in this paragraph: a probability threshold of 40 % (not 35 %) was used in the experiment. The probability threshold, similarly to the grid zone size, was evaluated in a pilot study, and 40 % showed the best results.*

*We have updated the article to clarify this:*

*2.7.2. Binarisation with grid zones:*

*second sentence*

**Page 16, line 31-38 and figure 8. You modified the evaluation labels, which will affect the accuracy assessment. I disagree to modify original label data. you used the data to train the RF model and then you modify it when you assess accuracy. This contradicts itself. For your concerns, perhaps you can convert the raster results to vector form. Then, comparing it with your digitized label ditch. When the detection result falls in a suitable buffer zone of the label ditch, you can regard it as correct detection.**

*We understand this concern. However, we feel that the modified evaluation labels are still justified. As proposed, a buffer zone is a good approach to avoid punishing the model’s prediction incorrectly. We feel that our evaluation method solves the incorrect punishment of ditch pixels adjacent to ditches with a “custom buffer”. This is an overall problem with classifying pixels, when the labelled data is derived from a vector format without knowledge of the width of individual ditches.*

*The Random Forests model, with our method, will receive the correct label for the majority of the pixels during the training phase, and therefore should learn to correctly detect ditches fairly well, despite some pixel’s labels being incorrectly around the edges of ditches. When predicting, however, using these incorrectly labelled pixels as a ground truth, would do the model an injustice, producing metrics that do not accurately describe the performance of the model.*

*If we had converted our predictions to vector format, many more assessment difficulties would have been created: We would not be able to use evaluation metrics such as Cohen’s Kappa or Area under the precision/recall curve. It would also be difficult to find a fair way to assess partial ditch detection, or incorrect ditch detection if vectors were used. For instance, if a ditch of incorrect shape and location was predicted, should it be punished more than other incorrect predictions? How is it turned into a vector if it does not have a linear shape?*

**Figure 8. why do you use 3 m2 zone?.**

*This is another numerical error, we have corrected this. The zones are 9 m2 (3 \* 3 metres). This is the same zone size that has been used when producing the final prediction, as described in section “2.7.2. Binarisation with grid zones”.*

**Some important figures, such as Figure 5 to 8, should be added legend instead of making too many descriptions in figure title.**

*We have revised the figure descriptions. We have tried to follow the style guides in the LaTeX template: “Template for International Journal of Geographical Information Science”, as well as used other published articles as a reference point.*

*Instead of adding legends to the figures, we have shortened the description of each sub-figure to only act as a headline for the sub-figure, and moved up some of the descriptions to the overall figure description, instead of as a sub-figure description.*

*For example in Figure 9 (previously Figure 8):*

*“Illustration of the modified evaluation labels where* ***a*** *shows the original result, and* ***b*** *shows the modified result. Green marks true positives, red marks false negatives, and blue marks false positives. False positives and false negatives that lay within one grid zone (9 m2) of a ditch label were evaluated as true positives and true negatives.”*

*was changed to:*

*“Illustration of the modified evaluation labels. Green marks true positives, red marks false negatives, and blue marks false positives. False positives and false negatives that lay within one grid zone (9 m2) of a ditch label were evaluated as true positives and true negatives.* ***a:*** *original results,* ***b:*** *modified results.”*

*Figures 6-8 (previously 5-7) had the headline for the sub-figures shortened.*

**In the discussion part, I suggest the author make some comparative analysis between RF and other machine learning methods such as SVM, ANN, etc., which are also very commonly used.**

*We conducted an experiment to ensure that the best algorithm for the task was selected. We compared four algorithms (as described in the answer to Reviewer 1’s comments regarding inclusion/exclusion of variables): Extreme Gradient Boosting, Random Forests, Naive Bayes, and Support Vector Machines. We used two zones (that are not in the final experiment) and a cross validation method to evaluate each algorithm with varying feature spaces (testing several feature spaces from 81 down to 10 features). Random Forests was shown to be the best performing algorithm with respect to the Cohen’s Kappa metric.*

*The code for this pilot study is added as a link to figshare in the Data and codes availability statement.*

*We feel that a more proper place in the report to talk about this is in the Methods section:*

*2.5.1. Processing the digital terrain indices:*

*entire last paragraph*

**The conclusion needs to be more accurate of a description of the work.**

*We have rewritten large parts of the conclusion to more accurately reflect the queries raised in the Abstract and Introduction. We removed some parts that felt shoehorned in, and added more suggestions for future work that have been hinted at throughout the article.*

*Dear Professor Laffan*

*Here is our revised copy of the manuscript, including a change highlight document, as well as our answers to the reviewers’ comments. We have made an effort to bring down the size of the manuscript by writing more concisely where possible, and by removing redundant paragraphs. We have done our best to improve language, and to remove any spelling errors. We feel that the use of colour figures is justified where we have them, but if required, we can remake these figures in grayscale. We hope that our revisions are sufficient, and that the manuscript will now be accepted.*

*Sincerely,*

*Niklas Lavesson (corresponding author)*

*On behalf of the authors*