# 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.**

*We feel that the use of colour figures is justified, and that it makes a lot of concepts easier to understand, instead of using scales of gray or dotted/dashed lines etc. If deemed necessary, we can update the figures into a grayscale form.*

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

*Answer*

**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.*

# 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.*

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

*Answer*

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

*Answer*

**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. However, we can’t find any inconsistencies with this this fact 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 try to prune sinks incorrectly predicted as ditches. One input variable (that has subsequently been removed in the feature space reduction experiment) attempted to fill small gaps in ditches by looking in opposing directions of a pixel, and updating their values if a suspected ditch was found in both directions.*

*We also use gabor filters for several input variables. Gabor filters are specifically designed to detect linear features in images, and this feature proved to be very effective, as can be seen in the feature importance table (LÄGG TILL KORREKT TABLE). The gabor filter explanation has been clarified in section 2.4.1. Processing the digital terrain indices:*

*“30 Gabor filters, which were rotated in different angles and with different frequencies, were combined to detect lines, amplifying ditches by utilising the fact that ditches have a linear elongated shape (Figure 3: d and g).”*

*Finally, our post processing method makes substantial efforts to remove sinks and small “clusters” of incorrectly predicted ditch pixels. We both look at the size of connected clusters of ditch predictions, as well as the shape of the clusters. If the clusters are small or not of an elongated shape. we remove these clusters from the prediction. This is described in detail in section 2.6.3. Cluster removal, as well as in Figure 7.*

**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 2.4.1. Processing the digital terrain indices paragraph 7 (now 6). As can be seen in the feature importance table (LÄGG TILL KORREKT TABLE), 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 9 (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?**

*Answer*

**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 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). This helped both in determining the correct machine learning algorithm, as well as the best feature space to use in the final experiment.*

*We first determined broad intervals by shifting from 81, 50, 35, 25, 20, 15, and 10 input variables. We then performed a more detailed search with smaller intervals around 50 input variables (the best result from the first iteration) and found the best number of input variables to be 53. The Random Forests feature importance from the previous input variable experiment was used to determine what features to include in the following test. With this approach, we could remove unnecessary features that dragged the performance down.*

*The final experiment was then rerun using only these 53 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.:*

*2.4.1. 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 2.6.1. 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:*

*“Overall, 81 input variables were extracted from the terrain indices. We then conducted a sub-experiment to find the best input variables, as well as the optimal number of variables to use. Random Forest's feature importance was used to select the most valuable input variables for differen*t variable amounts. It was found that using 53 input variables produced the best results.*”*

*Of course, this also means new result metrics for the end experiment, as well as new feature importances.*

**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)**

*temp: en feature försöker att åtgärda detta (om vi har kvar den featuren). Vi försöker även att åtgärda detta med post processing (ta bort icke-avlånga klumpar).*

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

*Answer*

**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?**

*Answer*

# 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.**

*Answer*

**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 “2.4.2 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 was also added to graphically explain the cross validation approach with the 11 hold-out subsections.*

*We feel that subsection “2.4.2 Training and validation datasets - paragraph 2” explains clearly that the upper 10 zones are not used in the final experiment: “The remaining 10 subsections were used solely before the experiment to develop the model to test different ways of preparing the data. This allowed the model to be evaluated on unseen data to strengthen the validity of the experiment. Figure 1 shows which subsections were used for development and evaluation respectively.”.*

*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.**

*Answer*

**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.**

*Answer*

**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.**

*Answer*

**Page13, line 44-45. Why 6 \* 6 grid zone, and probability of 35 % are used?**

*Answer*

**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.**

*Answer*

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

*Answer*

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

*Answer*

**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: XGBoost, 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). This helped both in determining the correct machine learning algorithm, as well as the best feature space to use in the final experiment. Random Forests was shown to be the best performing algorithm with respects to the Cohen’s Kappa metric.*

*We feel that a more proper place in the report to talk about this is in the Introduction section as a reason to select Random Forests as the algorithm, rather than in the Discussion section. The following sentence was added to Introduction, paragraph 4: “In a pilot study, we compared several different algorithms (Random Forests, XGBoost, Naive Bayes, and Support Vector Machines), and it was found that Random Forests produced the most accurate results.”*

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

*Answer*

## Change log

## **2.3. Extracting ditches with digital terrain indices - paragraph Slope**

ArgGIS ⇒ ArcGIS

## **1. Introduction - paragraph 2**

A major research gap is that the locations of many ditches are unknown. A comparison between a field inventory of ditches (within the National Inventory of Landscapes in Sweden (NILS)) and current maps, shows that approximately 90 \% of the ditch networks are missing on current maps (Ågren, unpublished). To better prioritise the restoration of ditched wetlands, it is necessary to know where the ditches are located, but field inventories are often too costly. A solution to this can be to automatically detect ditches from high-resolution digital elevation models (DEMs) using digital terrain indices.

⇒

A comparison between a field inventory of ditches (within the National Inventory of Landscapes in Sweden (NILS)) and current maps, shows that approximately 90 \% of the ditch networks are missing on current maps (Ågren, unpublished). To better prioritise the restoration of ditched wetlands, it is necessary to know where the ditches are located, but field inventories are often too costly. The research objective of this study is therefore to automatically detect ditches from high-resolution digital elevation models (DEMs) using digital terrain indices.

## **1. Introduction - paragraph 3**

From the high resolution DEM, we use artificial intelligence in the form of machine learning to locate the ditch networks.

⇒

From the high resolution DEM, we use machine learning to locate the ditch networks.

## **2.5. Building the random forests model - paragraph 5**

## python ⇒ Python

## **2.7. Evaluation - paragraph 1**

Cohen's κ index measures how much better a prediction is compared to a prediction based purely on chance, where chance would yield a value of zero (Sim and Wright 2005). With our data, a κ value close to zero would be attained by predicting 2 % of the occurrences as ditch pixels completely at random. Values above zero are better than chance and values below zero are worse than chance.

⇒

Cohen's κ index measures how much better a prediction is compared to a completely random prediction, where random would yield a value of zero (Sim and Wright 2005). With our data, a κ value close to zero would be attained by predicting 2 % of the occurrences as ditch pixels completely at random. Values above zero are better than random, and values below zero are worse than random.

**2.7. Evaluation - paragraph 2**

chance ⇒ random

**4. Discussion - paragraph 3**

This means that our model performed substantially better than one based purely on chance.

⇒

This means that our model performed substantially better than a completely random model.

## **1. Introduction - paragraph 2** & **2.4.1 Processing the digital terrain indices - paragraph 5** & **2.6.3 Cluster removal - paragraph 1** & **4. Discussion - paragraph 4**

cavities ⇒ sinks

**2.4. Processing the digital terrain indices - paragraph 2**

This facilitated finding obscurities in the neighbouring areas around pixels.

⇒

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.

**1. Introduction - paragraph 4**

We reproduce these neighbouring area variables when building the model in our study. Random Forests also computes Mean Decrease in Impurity, or Gini importance, highlighting what input variables are the most important for a given prediction (Menze *et*  *al.* 2009). Because of this, and the fact that it proved to be suitable for Roelens *et al.* (2018), a Random Forests model is used to locate the ditches.

⇒

We reproduce these neighbouring area variables when building the model in our study. In a pilot study, we compared several different algorithms (Random Forests, XGBoost, Naive Bayes, and Support Vector Machines), and it was found that Random Forests produced the most accurate results. Random Forests also computes Mean Decrease in Impurity, or Gini importance, highlighting what input variables are the most important for a given prediction (Menze *et al.* 2009). Because of these factors, and the fact that it proved to be suitable for Roelens *et al*.(2018), Random Forests models is used to locate the ditches.

**2.4.1 Processing the digital terrain indices - paragraph 6 (now 5)**

Both HPMF and Sky View Factor were used with the image processing gabor filter,which can be used to detect lines of a certain orientation in an image (Hong *et al.* 1998). To detect lines in all directions, 30 Gabor filters, which were rotated in different angles and with different frequencies, were combined to amplify ditches, see Figure 3:d and g.

⇒

Both HPMF and Sky View Factor were used with the image processing gabor filter,which can be used to detect lines of a certain orientation in an image (Honget *al.*1998). 30 Gabor filters, which were rotated in different angles and with different frequencies, were combined to detect lines, amplifying ditches by utilising the fact that ditches have a linear elongated shape (Figure 3: d and g).

**2.4.1 Processing the digital terrain indices - paragraph 5**

The *Sky View Factor Conic filter* was developed to attempt to detect and fill gaps in ditches. This was done by calculating the mean of all the pixels covered by a cone-shaped mask with a radius of 5 metres, which expanded outwards from the examined pixel point in eight directions. If two opposing mean values were both below a threshold, the pixel was given a lower value. This meant that only pixels with strong ditch indicative values in two opposing directions were updated, allowing the filter to avoid updating pixels in cavities or hollows, focusing only on linear geographical properties.

⇒

*removed*

**2.6.1 Noise reduction and gap filling - paragraph 2**

The third step involved taking measures to try to fill gaps in ditches that the model failed to correctly predict. A similar method to the one described in skyviewconic was employed to calculate the mean of cone masks expanding outwards in different directions from the examined pixel. This step also amplified some of the noise that was left, but filling the gaps in the ditches was judged to be more important to help make the next step more effective (Figure 5 b).

The third step involved taking measures to try to fill gaps in ditches that the model failed to correctly predict. This was done by aggregating pixels covered by cone masks expanding outwards in different directions from the examined pixel. This step also amplified some of the noise that was left, but filling the gaps in the ditches was judged to be more important to help make the next step more effective (Figure 5 b).

**2.4.2 Training and validation datasets - paragraph 2**

Using the 11 subsections in the hold-out data for the final experiment, a process called k-fold cross validation was employed (11-fold cross validation in our case). K-fold cross validation is a method where you divide your dataset into folds of similar size and traina model on all but one of your folds (subsections). You then use that subsection to evaluate the results (Wong 2015). Using this technique, shifting which subsection to leave out from the training, allowed us to train 11 different Random Forests classifying models with a large amount of data from the remaining 10 subsections in the hold-out data, producing 11 sub-experiments to evaluate the method on.

**⇒**

Using the 11 subsections in the hold-out data for the final experiment, a process called k-fold cross validation was employed (11-fold cross validation in our case). K-fold cross validation is a method where you divide your dataset into folds of similar size and traina model on all but one of your folds (subsections). You then use that subsection to evaluate the results (Wong 2015) (Figure 4). A new model is trained using the training folds for each iteration (Figure 4). Using this technique, shifting which subsection to leave out from the training, allowed us to train 11 different Random Forests classifying models with a large amount of data from the remaining 10 subsections in the hold-out data, producing 11 independent sub-experiments to evaluate the method on.

Figure 4: new figure to make the cross validation approach more understandable from the perspective of our 11 subsections.

**Multiple places throughout the text**

model ⇒ models and classifier ⇒ classifiers (when referring to the final experiment)

**Multiple places throughout the text**

model ⇒ ditch locator (when referring to the entire ditch detection process (i.e. input variables, model, and post-processing))

**1. Introduction - paragraph 5**

This mapping was used to both train the Random Forests model and to ground-truth the models.

⇒

This mapping was used to both train the Random Forests models and to produce the ground-truth.

**2.2 Digitising the ground truth - figure 1**

The 10 subsections with a white border were used for developing the model before the experiment.

⇒

The 10 subsections with a white border were used for developing input variables and optimising parameters before the experiment.

**2.4.1. Processing the digital terrain indices - paragraph 1**

Developing the Random Forests model involved examining how different kinds of input variables affected the prediction.

⇒

To improve the performance of the Random Forests models, we examined how different kinds of input variables affected the prediction.

**2.5 Building the Random Forests model - Headline**

Building the Random Forests model

⇒

Developing the Random Forests model

**4. Discussion - paragraph 5**

However, small sinks or hilly areas incorrectly classified as ditches by the model were generally removed in the post-processing.

⇒

However, small sinks or hilly areas incorrectly classified as ditches by the Random Forests models were generally removed in the post-processing.

**2.4.1 Processing the digital terrain indices - last paragraph**

*added:*

Overall, 81 input variables were extracted from the terrain indices. We then conducted a sub-experiment to find the best input variables, as well as the optimal number of variables to use. Random Forest's feature importance was used to select the most valuable input variables for different variable amounts. It was found that using 53 input variables produced the best results.

**3. Results and analysis - Table 2**

New metrics after rerun experiment with fewer input variables.

**3. Results and analysis - Table 4**

New feature importances after rerun experiment with fewer input variables.

**Multiple places throughout the text**

81 input variables ⇒ 53 input variables

*Dear Editor*

*...*