Soil erosion detection research

Soil erosion detection is a global problem that people have tried solving for decades(since the 1960s). The field has shown significant progress thanks to the development of artificial intelligence and especially satellite images. In this paper, I will convey 3 analyzed papers on the related theme to give insights into how the detection works, which methods are used, and where to move with the problem in the future.

Having analyzed the related paper on the topic, I have determined the factors influencing the problem's solution based on different approaches. Different authors used different techniques, but I will focus on one of the traditional methods (paper published in 2019), the method which combines machine learning and actual soil investigation (paper published in 2023), and the method that only uses satellite data and deep learning approach (paper published in 2023).

Improved semantic segmentation model for anthropogenically disturbed parcels with soil erosion from remote sensing images

I believe that according to the task, the last method is the most suitable, so I will elaborate on the techniques the authors used there the most. The authors implemented an improved semantic segmentation model for anthropogenically disturbed parcels with soil erosion from remote sensing images. They implemented an improved model based on Unet++ with additional data engineering and some additional techniques.

At the very start, the publishers review the existing models and identify the problem they solve. The existing methods fail to integrate global pixels well, potentially resulting in information loss. In addition, those methods were unable to work appropriately with high-resolution images(as we are working with satellite data), which results in the inability to segment large structures, especially regarding their boundaries. Lastly, the methods have salt and pepper noise.

The authors introduced a few techniques they used to improve the model:

1. The model that solves the problem of blurry boundaries and significant structure identification introduced
2. To obtain more accurate boundaries, they used the PCA method implemented into the DL algorithm
3. The problem depends on context information, so they used BCJHDC, which improves segmentation accuracy.

Preprocessing of the data is done in a couple of stages:

1. Satellite images collected from the same regions (there are existing techniques that allow the developer to identify the percentage of similarity between different land scenes)

2. Image preprocessing, such as alignment, radiometric calibration, atmospheric correction, cropping, and filtering, was performed to obtain more accurate image information (previously done in different studies).
3. The data was then cropped to the extent where the area of interest contained 10% of all the data.

As mentioned above, U-net++ was used for this problem with different adjustments, which contain the solution to the upsampling problem. The problem contributed largely to the inability of the model to detect boundaries clearly. As a solution was used modified Hybrid Dilated Convolution (HDC) and PointRend algorithm, which allowed to achieve a higher boundary smoothness, so it does not suddenly break or fade.
BCJHDC tied with the methods described above, was used to create the output suitable for insertion into CNN.

The authors also used the Self-Attention Module, which is called Polarized Self-Attention. It is used to filter the most critical features from not important.

Finally, the model can capture more essential features, as it is more profound due to the techniques mentioned above, meanwhile focusing on the most critical details.

The model was properly tested against similar approaches and yielded the best result on the training dataset with F1 score equal to about 88%. However, the model works worse on the test dataset, even though it is better than all the competitors, with F1-score equal to approximately 77%.

The key results of the model are that it yields no noise in combination with better learning of the feature from a high-resolution image and better segmentation of the areas of interest with clear boundaries.

The authors also discuss some cons, such as a drop in significance between test data and train. The model also might misclassify the areas with buildings, for example, where no erosion is visible.

To summarize, while working with data, we need to pay careful attention to the distribution of the land scenes, as two different scenes might have a bit different landscapes and soil color, resulting in poor efficiency. Additionally, the model was trained using the data from the region of China, so the data preparation must be done accordingly to the region of interest. The model also identifies large regions better than all the competitors, but it often overlooks small details and, generally, a lot of tiny regions, which might be helpful to identify for preventing erosion. It is an important factor, as, for example, if the task is to identify soil erosion in forest areas, the model might not be able to solve it correctly. It is especially important to find those regions in forests, as trees serve as obstacles that cover the land. So, for this problem, one might consider different techniques. It is also important to mention that the images must be from a similar area and taken during the same period, as the land color and cover might change depending on the season.

Source: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10042988

A Remote Sensing Based Method to Detect Soil Erosion in Forests

The paper addresses the problem of soil erosion detection in the forest, which is often overlooked and has not been considered an important issue compared to croplands. However, much area is covered with plants, so it is important to identify the source of spreading and vulnerable areas.

In different studies, the authors used RUSLE models that used statistical data, such as rainfall, soil erodibility, and others. The authors of the current paper took a different approach to derive all the data from the satellite images.

Data preparation consists of a couple of steps:

1. Aerial images collection (the authors used the data from Landsat 8)
2. Radiometric correction (one of the standard methods in the field)
3. Selection of predictor factors

The last point refers to the combination of parameters used in the model, which are Fractional vegetation coverage (FVC) and slope - the standard parameters used in similar studies. As the problems are quite different from the related studies due to the forest cover, the authors decided to introduce two additional parameters: the health status of vegetation (the soil erosion is related to the plant's health, so the trees around the area might give some insight into the state of the land) and soil exposure degree (according to others studies this is an important factor of erosion, that also sheds some light on the problem solution).
FVC is calculated in two forms - linear and quadratic. The method works differently on different datasets, so ideally, both of them must be tested.

There are two metrics related to the plants' health - Nitrogen Reflectance Index (NRI), which is calculated based on reflection, and Yellow Leaf Index (YLI), which is also calculated from the satellite data.

After retrieving all the parameters, the authors normalize the data and use two approaches to form the model - Principal Component Analysis and multiplication approach. PCA works much better, so I will only focus on the first one.
The calculated data was then inserted into the SEUFM model:
SEUFM1 = PC1(FVC, NRI, YLI, NDSI, Slope)

The output of those five factors is then used for building the image, which can be used to identify the intensity of erosion, as it is a binary picture. The authors claim that the model has an efficiency of 90% tested on the data collected by the local scientists and locally marked for erosion.

However, the biggest disadvantage is that the model highly depends on the amount and quality of the numerical data. If the amount of data is insufficient, no satisfactory results will be received. The model was tested only on the Northern part of China, and the authors do not state they have used it for different regions of the world. However, similar studies yielded comparable results.

Detection of Soil Erosion Hotspots in the Croplands of a Typical Black Soil Region in Northeast China: Insights from Sentinel-2 Multispectral Remote Sensing

The paper addresses the problem of different extents of severity erosion detection(low, moderate, severe). The authors hypothesized that land with different types of erosion is characterized by different distinct spectral signatures and different landscapes (differences in minerals, oxides, and others). As erosion progresses, different parts with different extents of erosion produce distinct spectral signals, thus forming the basis for spatial classification and mapping of soil erosion intensity.

The data for the study was collected in two ways. The first one involves Sentinel-2 images. The second method involves a manual collection of the data. The data was collected from a cropland region of China and was taken in three different slopes: summit, mid-slope, and foot-slope, which correspond to the different erosive states of the land.

After the collection, the scientists ensured that the data was similar and had no drastic differences.

After the collection, the samples were taken to the laboratory for spectra acquisition and analysis. The scientists conducted the tests to obtain visible and near-infrared (VNIR) spectral regions of soil. The combination of PCA and LDA was conducted on the VNIR data to obtain proof of the spectral separability of soils.

Having the thresholds, they were tested on the satellite dataset, which is taken during cloudless weather. The results were tested against each other, along with the comparisons. The PCA-LDA method demonstrated clear inter-class spectral separability to discriminate soils under varying erosion influences, enabling the authors to develop the thresholds for future detection and classification.

In summary, the method showed very great results on the datasets. The authors were able to develop the thresholds using the data collected and naked manually and from a satellite. However, there are several problems with this approach. The first problem is the laboratory, resources, and gadgets needed to recreate the experiment.
The second problem is that the method is still empirical. In order to achieve a high-efficient classification of the soil, the thresholds must be developed for each specific region.
There is also a problem with the classes chosen, as the authors only work with three types of clearly distinct data.