Detection of anomalies of magnetic field using clustering algorithm DBScan

Volzhina Lena
Saint Petersburg State University

Collecting magnetometric data

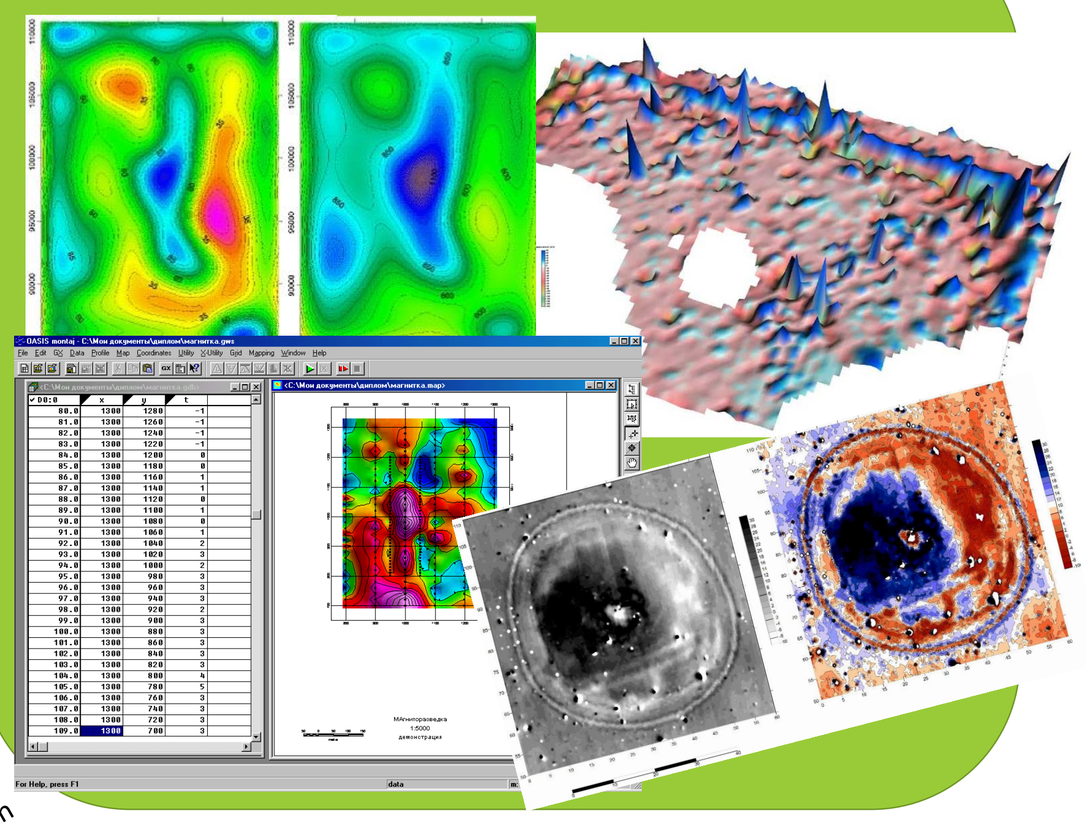
There is a lot of cases in archaeology and geology when scientists need to learn something about the contents of the ground without excavations. One of the common methods to solve this problem is to measure the value of the *magnetic field* at the earth's surface points and analyze these data.

The majority of objects of interest leave a mark there. For example, there can be traces of settlements, walls of buildings and other interesting things.

On the left you can see the process of magnetometric survey. Measurements are carried out on a grid, usually in increments of 0.5, 1 or 2 meters. Additionally, somewhere behind the scenes scientists measure constantly varying magnetic field of the earth to take it into account later.

preprocessing & filtering

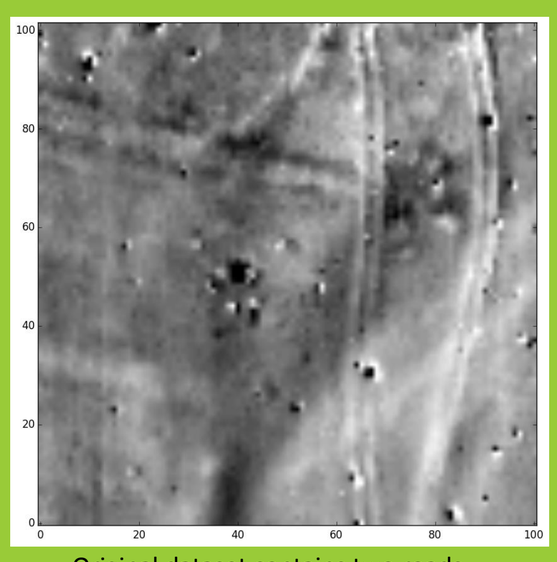
Manually data studying



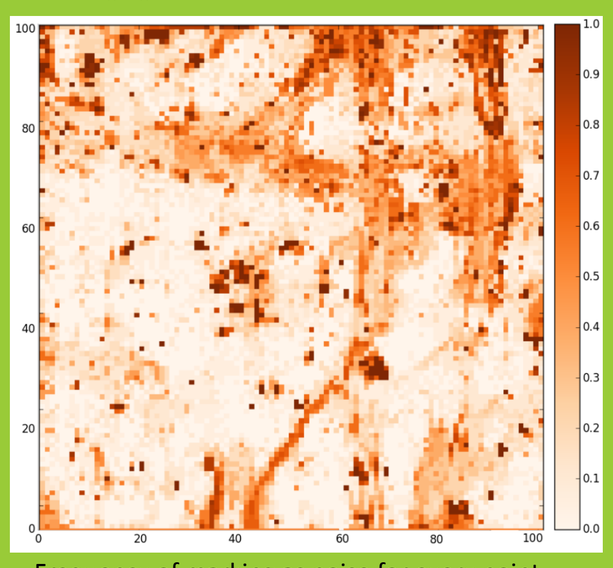
Depressingly often further research is carried out in manual mode: scientists use a variety of filtering, and then just look at the pictures closely and try to find some deviations and anomalies. But wait, isn't there a *huge pile* of other, more efficient ways to identify anomalies in data?

clustering with DBScan

Multiple DBScan application



Original dataset contains two roads, something like a ditch and also many small anomalies (may be metal objects)



Frequency of marking as noise for every point (can be considered as the probability of presence of anomaly in the point)

My goal was to find these anomalies automatically using clustering algorithms. In this problem it seems pretty convenient to define the cluster as homogeneous area without anomalies, and after that to mark as abnormal those points which frequently did not fall into cluster. After trying to use the most common and simple algorithms like k-means and c-means I got some good results using **DBScan**.

So I have run this algorithm multiple times with different parameters and counted how often points are marked as noise points. On the left you can see comparison of original data and obtained frequencies.

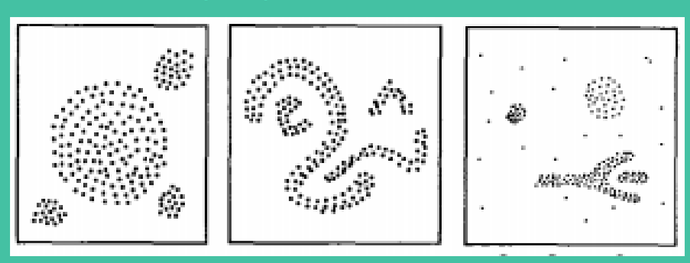
Now we can roughly estimate the probability of each point belonging to some anomaly. This already could help archaeologists and geologists in their work.

Next, becomes possible to select individual anomalies and work with them as *separate objects*. For example, it would be useful to classify anomalies according to their form and intensity, because these classes can match different types of objects in the real world. This problem can be also solved with clustering algorithms, or some other types of supervised or unsupervised learning.

DBScan

DBScan is a density-based clustering algorithm, which assumes that intra-cluster density is about equal and also higher than outside of the cluster. It works very good on clusters with complex shapes (as the images below) as well as on the simple ones.

DBScan requires specifying two parameters determining the desired density, and the clustering results are heavily dependent on them.



• • •

PROFIT!