Thursday Exercise 1

In this course, statistical validation is typically concerned with model outputs that are measured on a numerical scale. However, many models produce output that is measured on a categorical scale (e.g. land use, or tree type), for which the ME, RMSE and MEC cannot be calculated. In this exercise we assess the accuracy of such model outputs by comparison of model predictions with independent observations, using soil type prediction as an example.

Figure 1 shows the old and updated soil map of the Dutch province of Drenthe. The old map was made in the early 1990s using traditional soil surveying techniques. However, soils change over time, for example by oxidation of peat soils. Therefore, the soil map was updated using more recent data and a multinomial logistic regression approach[1].

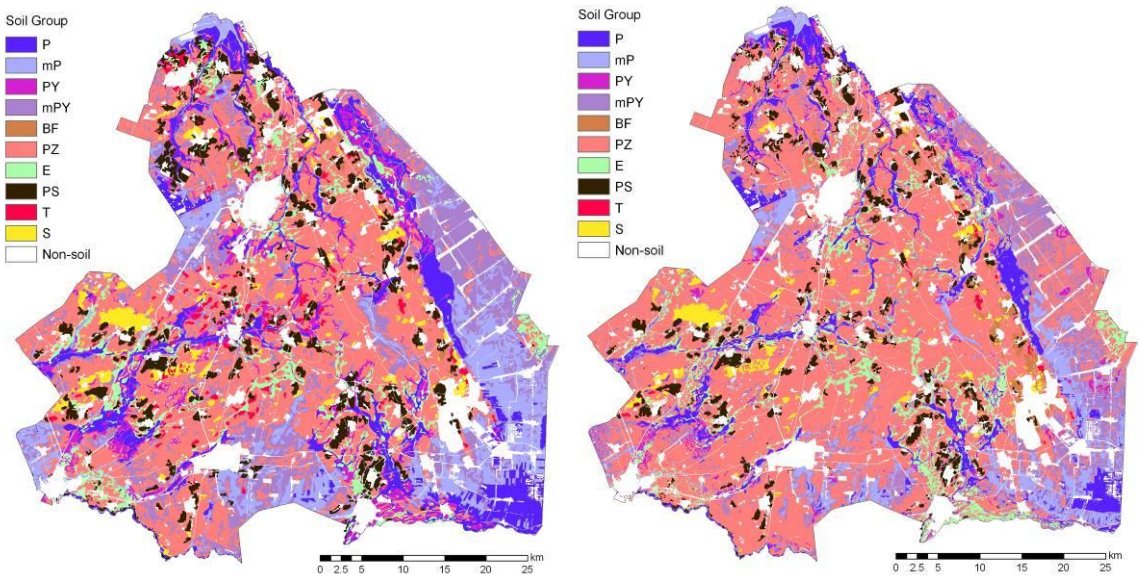

Figure 1. Old (left) and updated (right) soil maps of the Dutch province of Drenthe. P=thick peat; mP=thick peat soil with mineral surface horizon; PY=thin peat soil; mPY=thin peat soil with mineral surface horizon; BF=brown forest soil; PZ=podzol; E=earth soil; PS=plaggen soil; T=till soil; S=sandy vague soil.

In 2008 the true soil type was observed at 150 locations in Drenthe. See file 'soil_obs_drenthe.csv'.

Categorical data can be compared using a *contingency table*, which lists all combinations of classes (i.e., the measured and predicted soil types), counts the number of occurrences of each combination and puts these in a two-dimensional table. In the remote sensing literature this is known as an *error* or *confusion* matrix. The diagonal of the matrix represents the cases where observed and predicted soil type agree. The *purity* is a single measure of the map accuracy and is defined as the sum of the diagonal elements divided by the total number.

Calculate the purity for the old and updated soil maps. Which of the two maps is more accurate?

[1] https://doi.org/10.1016/j.geoderma.2009.04.023