Spectral analysis of sediment sources in the Isábena catchment in NE – Spain

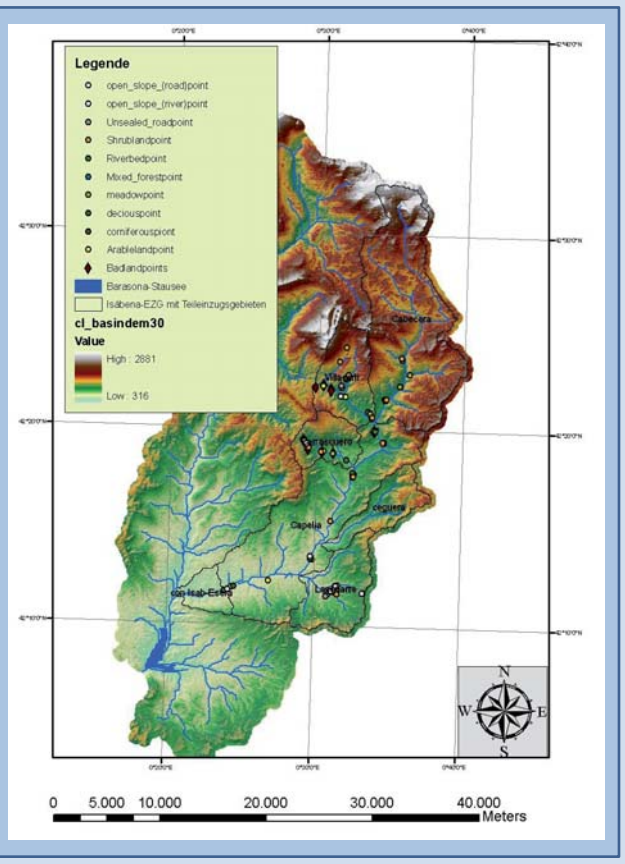
Arne Brauer, Benjamin Kayatz
Institute of Earth and Environmental Science, University of Potsdam

In semiarid climate water reservoirs are highly important for securing agricultural and domestic water supply. In these climate regions upland erosion has a great impact on the lifespan as well as on the efficiency of reservoirs. To assess this problem the quantitative understanding of erosion and the identification of main erosion sources are crucial. Previous studies have indicated that spectroscopy can be successfully applied in erosion studies. Our aim was to find out if field spectroscopic data is applicable at our study site to discriminate different land use areas using random forest.

Study site - Isabena Catchment

- Area: 445 km²
- Serves as water source for Barasona reservoir
- Pre-Pyrenees, NE-Spain
- Altitude: 430 – 2200 m a.s.l.
- 750 to 800 mm/a mean rainfall
- Divers land use and lithology
- High erosion due to badland formations

Fig, 1.: Isabena catchment landscape: badland, cropland, forest



Field work

- September/October 2010
- Measuring device: ASD FieldSpec 3 with contact probe, 350 to 2500nm
- 86 sampling sites with each 5 measurements
- 11 different land uses
- Collected data: spectra, soil colour, surface roughness, GPS data, vegetation cover, land use, slope, exposition etc.

Fig. 2.: field measurements

Methods

1. Preparing data: determine 1st derivation and continuum removal as well as different indices
2. Applying Wilcoxon-Mann-Whitney U test to find out for which wavelengths most land uses are significantly different to one another
3. Applying random forest to establish a classification model with different configurations of input data
4. Analysing the results

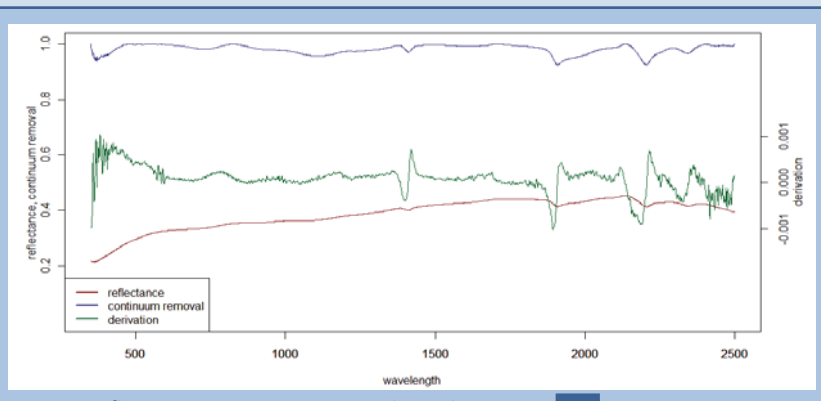


Fig. 3: reflectance, continuum removal, 1st derivation

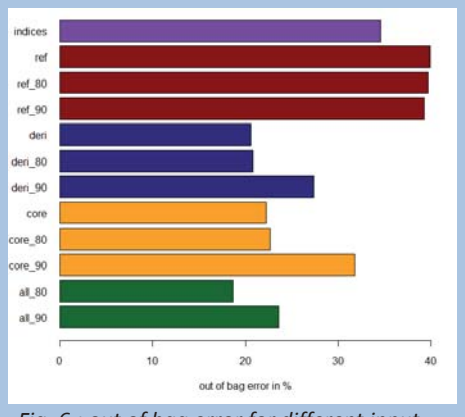


Fig. 6.: out of bag error for different input data configurations

```
> rf_all_80

Call:
 randomForest(formula = landuse ~ ., data = all_80[, -1], ntree = 500)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 53

        OOB estimate of  error rate: 18.69%
Confusion matrix:
                 Arable land Badland Coniferous forest Deciduous forest Meadow Mixed forest open slope (river) open slope (road) Riverbed Shrubland
Arable land               61       0                 1                 0      0            0                 0                 0        0        14
Badland                    0      58                 0                 0      0            0                 0                 0        3         0
Coniferous forest          1       0                 4                 1      0            3                 0                 0        0         1
Deciduous forest           2       0                 1                48      2            3                 0                 0        0         3
Meadow                     3       0                 0                11     13            0                 0                 0        0         3
Mixed forest               1       0                 2                 6      1            9                 0                 0        0         0
open slope (river)         0       0                 0                 0      0            0                15                 0        0         1
open slope (road)          0       0                 0                 0      0            0                 0                19        0         1
Riverbed                   0       2                 0                 0      0            0                 1                 0       29         2
Shrubland                  5       0                 0                 0      2            0                 0                 0        0        81
Unsealed road              0       0                 0                 0      0            0                 0                 0        0         4
                 Unsealed road class.error
Arable land                  0  0.19736842
Badland                      0  0.04918033
Coniferous forest            0  0.60000000
Deciduous forest             0  0.18644068
Meadow                       0  0.56666667
Mixed forest                 0  0.52631579
open slope (river)           0  0.06250000
open slope (road)            0  0.05000000
Riverbed                     0  0.14705882
Shrubland                    0  0.07954545
Unsealed road               11  0.26666667
```

Fig. 5.: Random Forest output for all wavelengths which have more than 80 significant comparisons

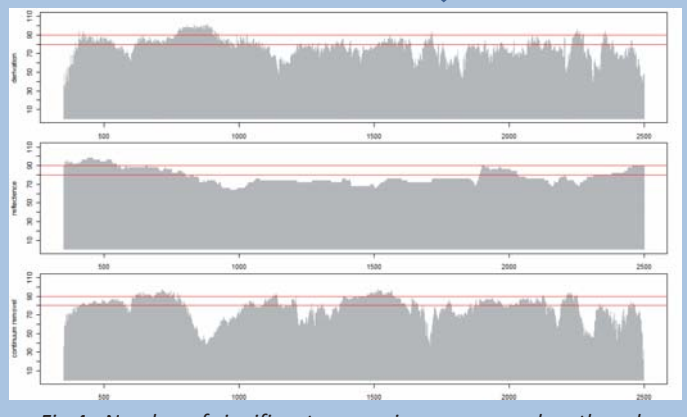


Fig 4.: Number of significant comparisons per wavelength and the two set thresholds of 80 and 90

Conclusion

- To improve results we need to extend field measurements
- Each land use should be tested the same amount to improve explanatory power
- Wilcoxon-Mann-Whitney test is a great way for data exploration, however there is a big difference between the results of Wilcoxon and random forest concerning distinguishability of land uses
- Using just wavelengths that are considered as important by Wilcoxon test helps to safe a lot of time but won´t improve the results if you using to few wavelengths
- The results of wavelengths that have more than 80 significant comparisons have almost the same classification accuracy as the entire data set
- The best results can be observed if we use reflectance, 1st derivation and continuum removal 80 as input data
- Given that 11 different land uses are a lot, random forest is a great classification tool and is applicable at our study site

Acknowledgements:
We want to thank Dr. Saskia Förster, Arlena Brosinsky, Carsten Neumann and Dr. Till Francke for their great support.
This project is funded by the Deutsche Forschungsgemeinschaft (DFG)

References:
Liaw A. und Wiener M. (2002): Classification and Regression by randomForest, R News Vol. 2/3,18-22
Martinez-Carreras N. et al (2010): The use of sediment colour measured by diffuse reflectance spectrometry to determine sediment sources: Application to the Attert River catchment, Journal of Hydrology 382, 49 - 63
Neumann, C.(2010): Synthese von ökologischer Gradientenanalyse und hyperspektraler Fernerkundung zum Monitoring naturschutzfachlich bedeutsamer Offenlandschaften, Potsdam

Contact: Arne Brauer: arnebrauer@gmx.de; Benjamin Kayatz: kayatz@uni-potsdam.de