Eurasian Journal of Soil Science

Journal homepage : http://fesss.org/eurasian_journal_of_soil_science.asp

Continuous Mapping of Soil pH Using Digital Soil Mapping Approach in Europe

Ciro Gardi *, Yusuf Yigini

European Commission Joint Research Centre, Institute for Environment and Sustainability Land Resources Management Unit, Via Enrico Fermi, Ispra (VA), Italy

Abstract

Soil pH is one of the most important chemical parameters of soil, playing an essential role on the agricultural production and on the distribution of plants and soil biota communities. It is the expression of soil genesis that in turns is a function of soil forming factors and influences all the chemical, physical and biological processes that occur in the soil. Thus it shapes the entire soil ecosystem. Due to any of the above reasons, mapping of soil pH becomes very important to provide harmonised soil pH data to policy makers, public bodies and researchers. In order to obtain a continuous mapping of soil pH for Europe, adopting the digital soil mapping approach, a set of continuously distribute covariates, highly correlated with pH, were selected. The estimate of soil pH was realized using a regression procedure, coupled with the kriging of the residuals. More than 30.000 points on top soil pH ($CaCl_2$) were used, and 27 covariates were tested as predictors. The similar approach was already applied with 12.333 samples to produce a pH map of Europe using European Soil Profile Data in 2008 which compiles several databases from 11 different sources (Reuter et al. 2008). Our study was conducted to update the previous data and maps based on LUCAS (EUROSTAT - Land Use/Cover Area frame statistical Survey), BIOSOIL (Hiederer and Durrant, 2010) and merged database which was used to produce previous soil pH map of Europe (Reuter et al. 2008). We used a compilation of more than 30.000 soil pH measurements from 13 different sources to create a continuous map of soil pH across Europe using a geostatistical approach based on regression-kriging. Regression was based on the use of 27 covariates in the form of raster maps at 1km resolution to explain the differences in the distribution of soil pH in $CaCl_2$ and we added the kriged map of the residuals from the regression model.

Article Info

Received : 17.01.2012
Accepted : 03.08.2012

Keywords: digital soil mapping, LUCAS, BIOSOIL, Europe, Geostatistics

Introduction

Soil pH is a fundamental property of soil systems and many inferences can be made regarding the chemical properties of soils based on its value (Hossner, 2008). Soil pH is considered a master variable, because it controls other physical, chemical and biological soil properties (Bolan and Kandaswamy, 2005). It is related to soil fertility level, concentrations of most heavy metals, activities of microorganisms and fauna, C/N ratio, and humus form. The availability of essential plant nutrients in soil, and in turn the distribution of several living organisms, is closely related to soil pH (Hossner, 2008). Soil pH has also implication related to some important EU policies, such as the definition of Less Favoured Areas (Van Orshoven et al., 2012). For the reasons explained above, soil pH is one of the most common parameters used for both soil characterization

* Corresponding author.

European Commission Joint Research Centre, Institute for Environment and Sustainability Land Resources Management Unit, Via Enrico Fermi 2749 21027 Ispra (VA), Italy

Tel.: +39 332 785015 Fax: +39 332 786394 E-mail address: ciro.gardi@jrc.ec.europa.eu

ISSN: 2147-4249

and soil monitoring programmes. Also within two important soil monitoring projects, coordinated by European Commission Joint Research Centre, soil pH have been analysed on more than 26000 soil samples. These projects are respectively BioSoil, mostly focused on forest and other semi-natural land uses, and Lucas soil, focused mainly on agricultural lands. With the objective of fully exploit this data legacy, and to produce a continuous estimate of soil pH across Europe, a digital soil mapping approach have been adopted. The digital soil mapping methodologies (McBratney et al., 2003) relies heavily upon modeling relationships between soil properties and easily measured, spatially exhaustive environmental covariates (e.g. digital elevation values, land use/land cover, meteorological data, etc.). In the research presented in this paper a geostatistical approach based on regression kriging, aimed to the estimate of soil pH a t 1 km resolution, is presented.

Material and Method

The objective of the present research was to obtain a continuous mapping of soil pH for Europe, with the aim of improving the previous soil pH map realized by Reuter et al. (2008).

Compared to this latter map, that was based on the use of 12,333 data points; we could rely on a larger and more updated dataset. This dataset is composed by the following data:

- Lucas: the entire dataset is composed by around 22,000 soil samples, 19,865 of which fully validated. For the objectives of the present research, only the values of pH in $CaCl_2$ have been used.
- Biosoil: the dataset is composed by more than 4,000 sampling points, but after a quality check procedures, only 3,395 have been retained. The sampling points were selected mainly in forest, also in this case, only the value of soil pH in $CaCL_2$ has been used.
- Other datasets: 10,633 data on soil pH in $CaCl_2$, selected from the database (12,333 data) previously used by Reuter et al., have also been included in the present analysis. These data derives in turn, from the aggregation of data collected within several projects, such as Spade, Forest Focus, Wise, etc.

While in the Lucas soil project, the sampling scheme was designed for the collection of a composed top soil sample (30 upper cm, of mineral soil), in Biosoil the sampling scheme was based on fixed depth layers. In order to make these two datasets comparable, the weighted average of soil pH, for the upper 30-40 cm, was calculated for Biosoil data. Data of the third, composed dataset were already referred to the topsoil. The aggregation of the datasets described above, enable us to have in total, 33,893 soil pH values. 3,388 of these values (10% of the total, randomly selected), has been saved in a separate database in order to be used for the validation process. The database used for the regression (pH_REGR) was composed by 30,505 values.

Covariates

The list and the main characteristics of the covariates used in this research are reported in table 1.

Table 1. Main Characteristics of the Covariates

Covariates	Spatial Resolution	Data Source
Climate		
T- mean	1000 m	WordClim
T-Mean July	1000 m	
T-Mean January	1000 m	
Total Precipitation	1000 m	
Topography/Geography		
Elevation	1000 m	SRTM
Slope	1000 m	
Latitude	1000 m	
Human/Management		
Land cover *Artificial area, Arable lands, Irrigated arable lands, Rice Fields, Vineyards, Fruits, Olive, Pastures, Agro-forestry Areas, Broad-leave Forest, Coniferous Forest, Mixed Forest, Grasslands, Moors & Heath-lands, Sclerophyllus Vegetation, Transitional Woodlands Shrubs, Other Natural Areas, Wetlands, Marine Wetlands, Water Surfaces*	1000 m	CORINE 2006 and 2000

With the pH_REGR database, a point shape file has been produced in ArcGIS. For each point the correspondent values of the covariates has been extracted using the "Extract Multi Values to Points" command (pH_REGR_COV). This new database, consisting of 28 parameters (pH + 27 covariates – table 1), was then imported in SPPS for the statistical analysis and for the calculation of the multiple regression equation.

Statistics and Geostatistical methods

Explorative statistics on soil pH values and the covariates and Multiple Linear Regression was calculated using SPSS (SPSS for Windows, Version 16). The spatial structures of the differences between measured and estimated soil pH data (residuals) was analysed calculating semivariograms. The ordinary kriging was applied in order to obtain a continuous estimate of the residuals over the investigated area. Semivariograms and ordinary kriging were calculated using the geostatistical tools of ArcGIS 10.0 (ESRI, 2011).

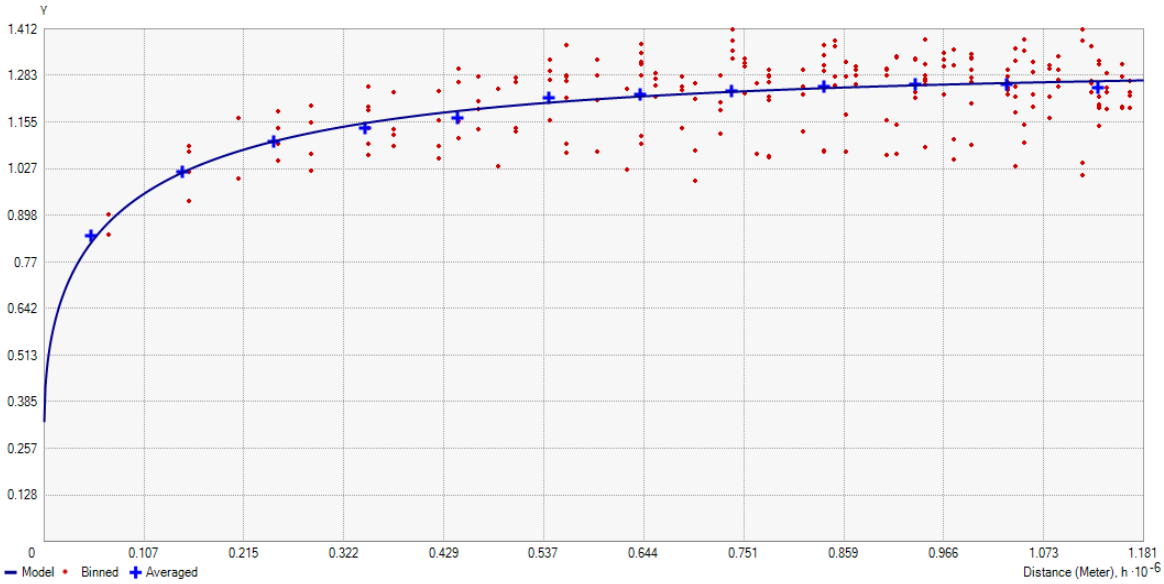

Figure 1. Semivariogram of the residuals

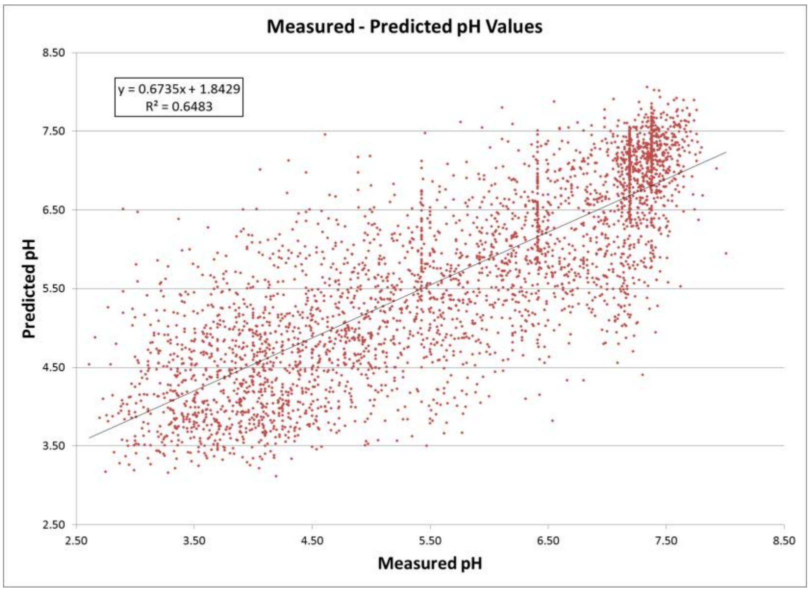

Figure 2. Plot of measured versus predicted values of pH ($CaCl_2$) in the validation dataset.

Results and Discussion

Among the 27 covariates analysed, 24 were significant ($p<0.05$). Latitude and Coniferous Land Cover (dummy variable derived from CLC 2006) were the two covariates showing the highest *r* values (-0.55 and -0.39 respectively). All the significant covariates have been selected for the calculation of the regression equation; 2 covariates (Tmean and Tmean January) however, have been excluded due to collinearity. The regression equation adopted (1) has a r^2 of 0.46:

$$pH\ (CaCl_2) = 5.699 + Total\ Precipitation^{*} \cdot -0.01 + Tmean7^{*}0.013 + Slope^{*}0.001 + Latitude^{*}{-0.081} + Landcover^{*}Factor^{\#} \quad [1]$$

[#]*Artificial area (4.471), Arable lands (4.744), Irrigated arable lands (4.912), Rice Fields (4.486), Vineyards (5.032), Fruits (4.689), Olive (4.995), Pastures (4.363), Agro-forestry Areas (4.461), Broadleave Forest (3.998), Coniferous Forest (3.637), Mixed Forest (3.818), Grasslands (4.381), Moors & Heath-lands (3.930), Sclerophyllus Vegetation (4.603), Transitional Woodlands Shrubs (4.040), Other Natural Areas (4.674), Wetlands (3.986), Marine Wetlands (4.716), Water Surfaces (4.222)*

This equation was used in “map algebra” tools (ArcGIS), in order to continuously estimate the soil pH over Europe. The next step was the extraction of the estimated pH value, in correspondence of the sampled points, and then the calculation of the residuals (Equation 2).

$$Residuals = Measured\ pH - Estimated\ pH \quad [2]$$

The spatial structure of the residuals was analysed calculating the semivariograms. The semivariogram (Figure 1) was characterized by a nugget value of 0.328, a sill of 0.969 and a range of 852. The value of the nugget should be interpreted as the existence of variance, not explained with the model that is generally related to the existence of a spatial variability at a finer scale, compared to the scale adopted in the sampling scheme. The parameters of the semivariogram were used for the application of ordinary kriging for the estimate of residuals. Using the raster calculator the values of the residuals were algebraically added to the values of the estimated pH, allowing producing the final map of soil pH (Figure 3).

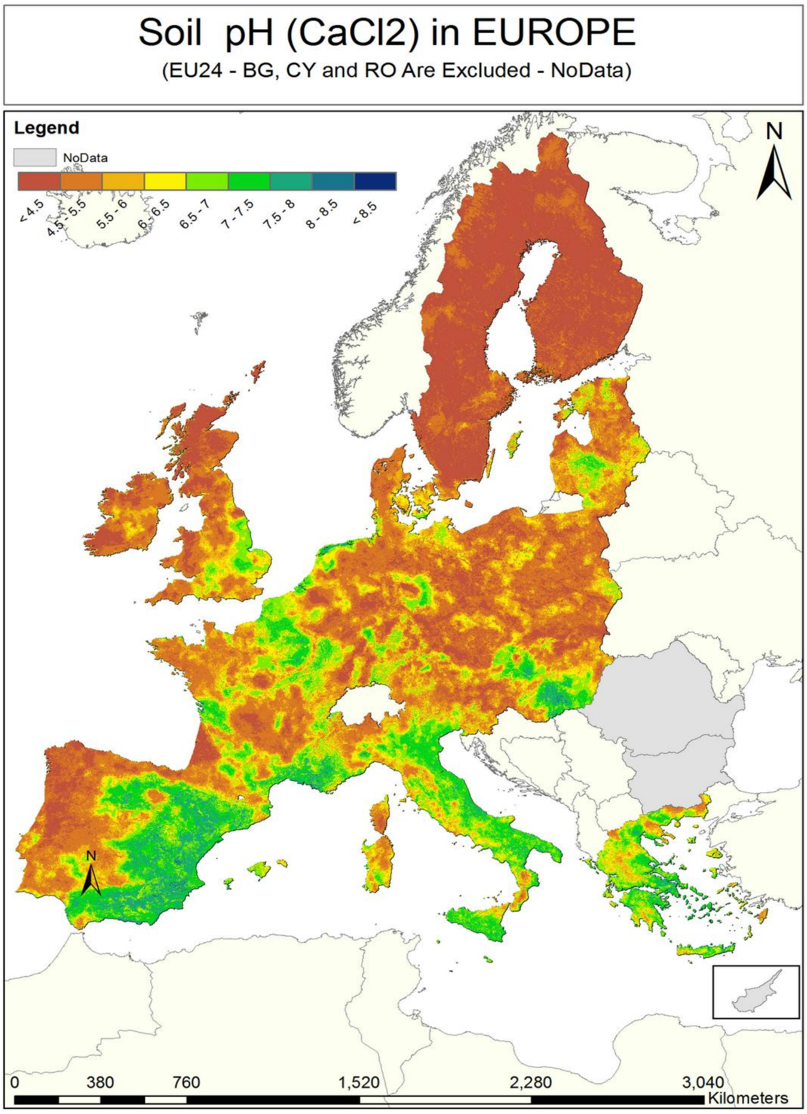

Figure 3. Estimated Values of pH ($CaCl_2$) for the EU24 Member States (BG, RO and CY Are Excluded - NoData)

Validation

3388 (10%) points were chosen randomly from the original dataset and this subset was used only for validation. The validation was performed comparing the pH values of the subset, with the estimated data, and calculating the Correlation Coefficient and the RMSE. These values were 0.805 and 0.18 respectively. On the base of validation process we found that 81 % of the errors were less than 1 point of pH (overestimate or underestimate), and 57% were less than 0.5.

Conclusions

The soil pH map produced showed a better accuracy and higher spatial resolution compared to the previous one. However some limitations are still present in the datasets we have used. Lucas datasets was mainly focused on agricultural areas, below 1000 m of elevation. In order to fill this gap we decided to integrate this dataset with other soil legacy data.

Further analysis on this datasets, using different DSM approaches are already planned, and will allows to compare the quality of the final soil pH map of Europe and possibly to improve the final results.

Acknowledgments

The authors would express their gratitude for the advices, suggestions and support provided by Cristiano Ballabio, Arwyn Jones and Roland Hiederer.

References

ESRI, 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. WORLDCLIM, Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978

Hiederer, R., Durrant, T., 2010. Evaluation of BioSoil Demonstration Project - Preliminary Data Analysis. EUR 24258 EN. Luxembourg: Office for Official Publications of the European Communities. 126pp.

Montanarella, L., Toth, G., Jones, A., 2011. Soil Component in the 2009 LUCAS Survey. Land Quality and Land Use Information In the European Union, Eds. Gergely Toth, Tamas Nemeth, Keszthely, LB-NA-24590-EN-C.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3-52.

Reuter, H.I, Rodriguez Lado, L., Hengl, T., Montanarella, L., 2008. Continental Scale Digital Soil Mapping using European Soil Profile Data: soil pH. - in Eds: Böhner,J. ,Blaschke,T. and Montanarella,L. SAGA – Seconds Out,; Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, Heft 19, Universität Hamburg, Institut für Geographie, ISSN: 1866-170X

SPSS for Windows; Version 16, SPSS Inc.; Chicago, Illinois

Van Orshoven J., Tarres J., Toth T., 2012. Updated common bio-physical criteria to define natural constraints for agriculture in Europe Definition and scientific justification for the common biophysical criteria; Technical Factsheets. Jrc Technical and Scientific Reports, EUR 25203 EN - 2012