Chapter 18
Mapping the CN Ratio of the Forest Litters in Europe-Lessons for Global Digital Soil Mapping

F. Carré, N. Jeannée, S. Casalegno, O. Lemarchand, H.I. Reuter, and L. Montanarella

Abstract The Carbon/Nitrogen ratio (CN) of forest soils is one of the best predictors for evaluating the soil functions mainly involved in climate change issues.

The CN ratio of forest litters depends generally on tree species and forest management which are local factors, but also on broader environmental factors. Thus, the European forest litter CN ratio map is predicted using: (a) punctual CN ratio measurements collected systematically every 16 km in European forests and analyzed according to a common European laboratory method; (b) spatially continuous information on tree species abundance (derived from interpolation) and climate, landform and lithology at 1 km resolution.

The spatial modeling of the CN ratio is done according to complementary approaches: first, a classical kriging approach done on the CN ratio measurements; and second, a neural network approach using a set of nonlinear equations on the environmental predictors. Other multivariate geostatistical approaches were tested but not retained for final results due to lack of correlation between environmental factors.

Twenty percent of CN ratio measurements are kept for validation purpose. The two approaches are compared using coefficient of determinations and Root Mean Square Errors on the validation dataset. Surprisingly, the best approach is the classical kriging, meaning that the spatial structure and variability of CN ratio cannot be explained by the environmental factors, which show high local variation. This leads to a discussion of the quality of the data and to envisage possible risks for global digital soil maps.

Keywords CN ratio · Kriging · Neural network

F. Carré (✉)
INERIS, Scientific Division, Parc technologique Alata, BP 7, 60550 Verneuil en Halatte, France
e-mail: Florence.CARRE@ineris.fr

J.L. Boettinger et al. (eds.), *Digital Soil Mapping*, Progress in Soil Science 2,
DOI 10.1007/978-90-481-8863-5_18,

18.1 Introduction

As an indicator of soil mineralization processes, the Carbon/Nitrogen (CN) ratio of the forest soils is one of the best predictors for evaluating soil functions such as biomass production and carbon storage capacity of forest soils. When integrated into risk assessment, these functions can serve for modeling scenarios of soil sustainability with climate change issues (gas fluxes emissions, biofuel production . . .). For instance, for a soil having a relatively high CN ratio, the mineralization process tends to be slower and the weak leaching of nitrogen results in a weak quantity of N gas flux emission.

The CN ratio is very dependent on the forest species (see Section 10.3.3 for more explanations on processes in the litter), management, and environmental factors (Burke et al., 1989) such as climate, relief, soil type and parent material. Furthermore, since forest management is done locally, the CN ratio variability has to be analyzed locally. The final objective is to model and map the local spatial distribution of the forest soil CN ratios for the entire area of Europe based on an European dataset. To this aim, different soil inference systems are tested (spatial and aspatial) based on the European CN ratio database and environmental data. In order to analyze the consistency of the data for such a large extent, a preliminary data analysis is performed. Then, three complementary soil inference systems are developed and compared. The results are then discussed relative to the dataset.

18.2 Material

18.2.1 The CN Data

There are 5,289 CN ratio measurements (see Fig. 18.1) available from the 1st Forest Focus campaign database (the official date is 1996 but the measurements are from 1987 up to 1995). The plot sampling and analysis are done according to the common manual which must be followed by the Member States (UNECE, 2003). The plots are analyzed systematically, usually every 16 km where forest exists. The plots are described according to 5 strata: the soil litter, the topsoil horizon (0–10 cm) and lower horizons (10–20 cm; 20–40 cm; 40–80 cm). Each strata measurement is a composite of at least 4 soil samples about 50 m from each other. For this study, we focus only on the litter dataset.

Out of the 5,289 measurements, 1,929 were lacking either the CN ratio information, or the coordinates. Moreover, due to possible errors of laboratory analysis, the 3,360 measurements were analyzed using classical statistics and an indicator of the spatial variability (variogram cloud analysis). Twenty outliers were detected and removed. At the end, there were only 3,340 remaining measurements of CN ratio for the litter.

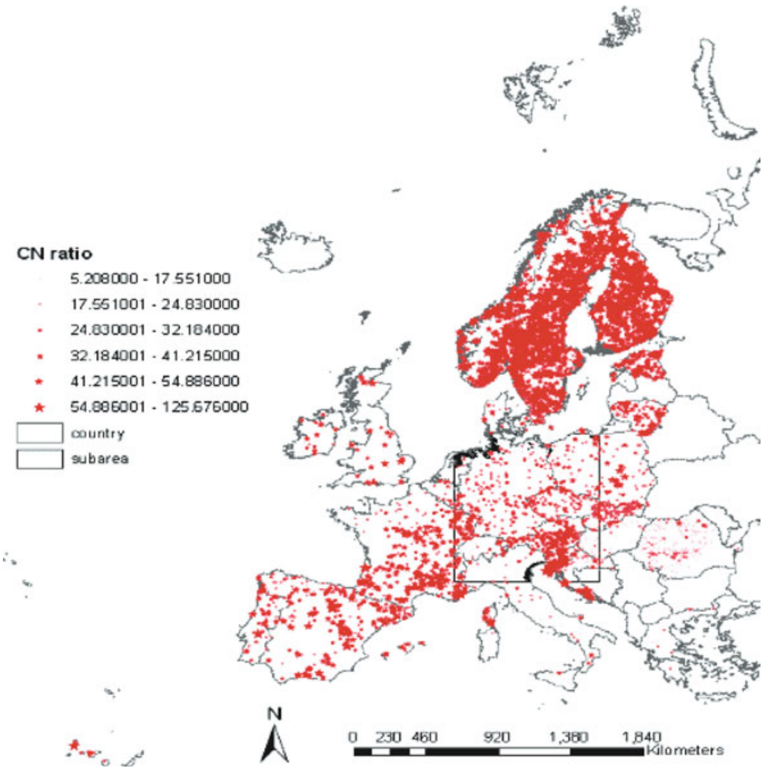

Fig. 18.1 The sample locations of CN ratio across Europe and the location of the study subarea

18.2.2 The Environmental Dataset

In order to improve the spatial prediction of the CN ratio for the forest soil litters, 39 soil covariates which can influence the CN ratio were considered. These are:

– The 23 main abundant tree species in Europe (representing 98% of the total number of trees) as of 1996. The tree species were punctual information for which only 58% had correspondence with soil sample locations. They were then transformed into raster grid cell by an Inverse Distance Weighted Interpolation, consistent with the European Forest species map estimation of 2004 (Casalegno, 2009). For the tree species prediction we also tried to use the FAPAR (Fraction of Absorbed Photosynthetically Active Radiation) of Europe for the period 1997–2000 (Gobron et al., 2006). Unfortunately, there was no correlation. We prefered a bottom-up approach compared to a top-down approach for predicting forest vegetation cover;
– Landform attributes derived from the SRTM (mean altitude, standard deviation altitude in a 1km raster grid cell, slope, aspect, curvature and moisture index-see Section 20.2.3 for more explanation on covariates);

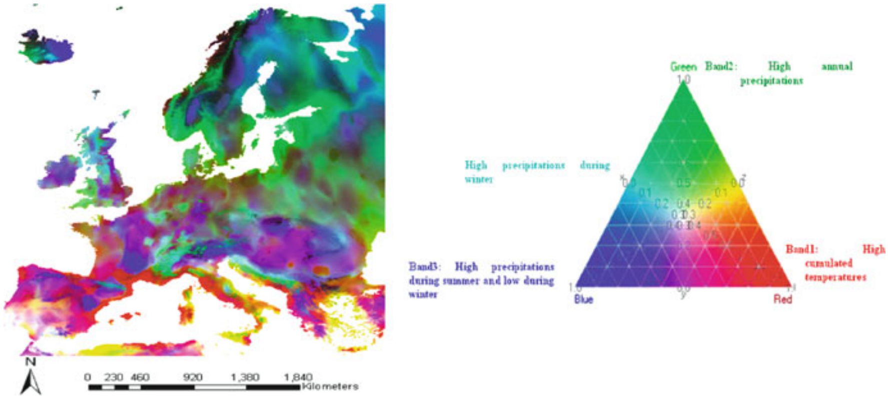

Fig. 18.2 (**a**) The bioclimatic map of Europe (1 km resolution); (**b**) the associated legend

- Principal components (see Fig. 18.2) on climatic data over the 1950–2000 period (annual mean temperature, mean diurnal range of temperature, temperature seasonality, isothermicity, annual rainfall) and on derived bioclimatic variables (ombrothermic index, drought stress index, thermicity index, continentality);
- Mean annual evapotranspiration, cumulated evapotranspiration and box moisture index, also derived from climatic data;
- Latitude and longitude coordinates.

All the raster grids were at the resolution of 1 km × 1 km and projected according to the European INSPIRE standards (ETRS89 Lambert Azimuthal Equal Area).

18.3 Methods for CN Ratio Modeling and Mapping

The modeling efficiency of the CN ratio using spatially continuous covariates was assessed on a subarea having transboundaries (Fig. 18.1 – see Central Europe), which contain 739 CN ratio measurements. 80% of the measurements were randomly selected for modeling (616 measurements) and 20% for validation (123). Different approaches were tested and developed.

(a) The first approach simply consisted of a classical ordinary kriging approach (Wackernagel, 1995) on the CN ratio measurements. The modeling was based on the following equation:

$$\mathrm{CN}(x_0) = \sum \lambda_i \, \mathrm{CN}(x_i)$$

Where λ_i are the kriging weights, $\mathrm{CN}(x_i)$ the measurements at location x_i and $\mathrm{CN}(x_0)$, the prediction at location x_0.

(b) The second approach was done using a neural network approach (see Section 13.2.5.1) on the CN ratio predictors. This two-stage classification approach consisted of a set of nonlinear equations that predicted the CN ratio from the soil covariate predictors in a flexible way, using 8 node layers of linear regressions and S-shaped ($1/1 + \mathrm{e}^{-x}$) functions. The modeling was based on the following equation:

$$\mathrm{CN} = \beta_0 + \sum \beta_i H_i$$

with $H_i = 1/(1 + \mathrm{e}^{-xi})$ are the node layers (hidden layer)

$$x_i = \alpha_{i,0} + \sum \alpha_{i,j}{}^{*} p_j$$

$p_1, \ldots, p_\mathrm{n}$ are the predictors and $\alpha_1, \ldots, \alpha_n$ are the regression coefficients.

Other multivariate geostatistical approaches were tested, such as cokriging and Min/Max Autocorrelation Factors, first using a reduction of the number of environmental predictors and then a kriging with external drift on the resulting MAF factors (Desbarats and Dimitrakopoulos, 2000). They were not retained for further analysis due to weak correlation between CN ratio and environmental factors and high small scale variability of the predictors.

The first approach (a) is spatial whereas the second one (b) is aspatial. For comparing the efficiency of the two approaches and to reach a conclusion on the need of using spatial or aspatial approaches, the percentage of good prediction (R^2 adjusted coefficient of determination) and RMSE were computed on the validation dataset.

Final mapping was limited to the areas covered by more than 10% of forests in Europe to get the final distribution of the CN ratios in the forest litters. Forest occupation data was derived from the Pan European Forest/Non-Forest map 2000 (Pekkarinen et al., 2009). All maps followed the standards of the predictors (INSPIRE resolution and projections).

18.4 Results

The three approaches are discussed and then compared.

18.4.1 The Classical Kriging Approach

The CN ratio variogram (Fig. 18.3) was obtained after a preliminary Gaussian transformation to reduce the impact of data variability. The variogram shows important small scale variability (50% of total variability), attributable to local forest management and potential measurement errors and/or inconsistencies in measurement procedures between adjacent countries (this point will be discussed later). Moreover, there was a clear spatial structure with a range approximately equal to 250 km (maximum distance of correlation).

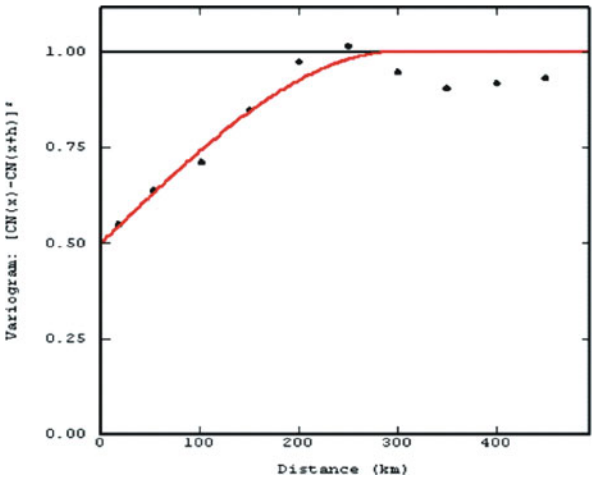

Fig. 18.3 The variogram of Gaussian transformed CN ratio measurements in the subarea. Experimental points in *black*, modeled variogram in *red*. Statistical variance is also displayed (*black line*)

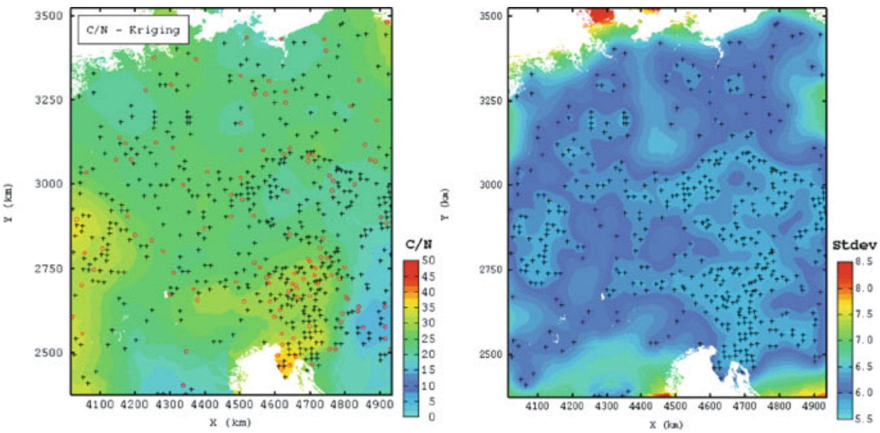

Fig. 18.4 (**a**) Map of the CN prediction derived from classical kriging; (**b**) map of the standard deviations of the kriged errors. *Red rings* are the validation points, *Black crosses* are the modeling points

The resulting maps of the kriging (CN ratio prediction in Fig. 18.4a) show a smoothed CN ratio prediction as expected, increased by the modeled nugget effect. The standard deviation error map allows for future possible sampling areas (Fig. 18.4b).

The adjusted coefficient of determination between predicted data and measurements was 0.60 and the Root Mean Square Error was 4.91 (Fig. 18.6a).

18.4.2 The Neural Network Approach

The neural network, contrary to the kriging, shows high local variability of the CN ratios (Fig. 18.5), respecting in some parts, boundaries of tree species and relief. The

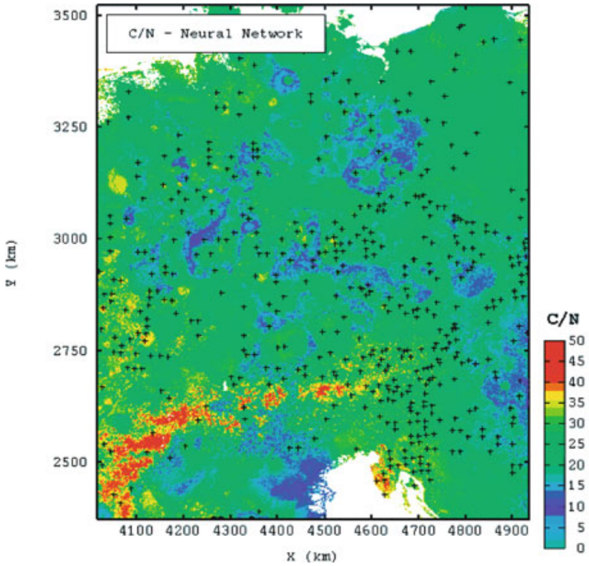

Fig. 18.5 Map of the CN ratio prediction derived from the neural network approach

adjusted coefficient of determination between predicted data and measured data was 0.40 and the Root Mean Square Error was 4.85 (Fig. 18.6b).

The comparison of the two maps (Figs. 18.4a and 18.5) reveals global trends in the CN ratios of Central Europe. The CN ratio is globally medium in this area but high in the northern part of Croatia and the eastern French border. However, the neural network map clearly shows shapes of environmental factors, particularly landform, but also some round spots which represent artifacts in the prediction of tree species. This map gives more importance to local variability than the kriged map. Both maps are defendable from a methodological point of view. For an assessment of the CN ratio, we would recommend the kriging method due to the reasons in the following discussion.

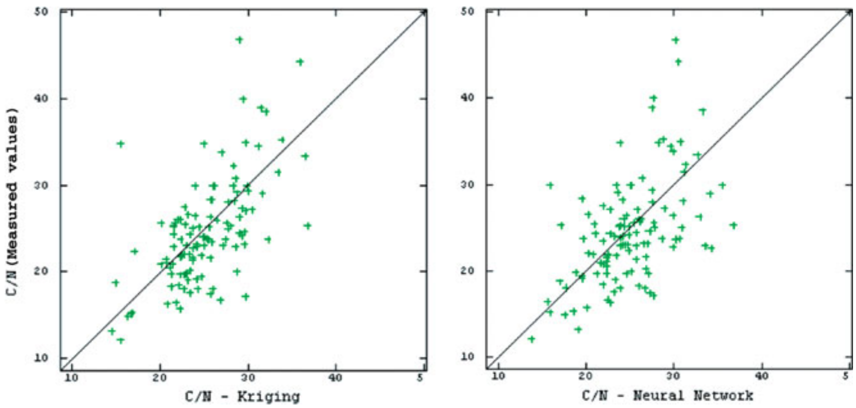

Fig. 18.6 (**a**) Scatter plot of validation against predicted values for the ordinary kriging; (**b**) scatter plot of validation against predicted values for the neural network

18.5 Discussion

The results of both methodologies are not satisfying and should be improved for risk assessments. Moreover, it is interesting to observe that taking into account spatial variability of the CN ratio without considering potential predictors sometimes gives better results than taking into account the predictors without considering spatial trends. This could be due to four main reasons:

(a) High local variability of the CN ratio which can be modeled at 1 km resolution. The four sampling repetitions are then not sufficient;
(b) The measurement errors due to sampling and laboratory analysis as shown by the variogram analysis, increased by an expected lack of consistency between sampling/analysis protocols on adjacent countries (see hereafter);
(c) The problem of Inverse Distance Weighted (IDW) interpolation of the tree species abundances. Indeed, tree abundances show a high spatial variability, and as a consequence, the IDW approach probably gives poor estimates of tree abundances outside of data locations.
(d) The models may not be efficient enough to predict the CN ratio.

These different issues are now studied. The prediction of tree species for the year 2004 shows that regression trees on environmental predictors should be preferably used (Casalegno et al., 2010). Concerning the study of local variability and other model tests, they should be studied once the quality of the data has been estimated.

For that, we launched inquiries asking each Member State the final methodology used for collecting and analyzing data. These inquiries demonstrated that each Member State followed its own methodology, the one they traditionally used concerning sampling strategy, even when common standards should be followed. Furthermore, the laboratory measurements are also very dependent on the calibration of the materials. It is then difficult to compare transboundary samples and also national samples, since sometimes different laboratories were used for the chemical analyses. Some work should be done on transboundary corrections of the measurements. To this aim, the new BIOSOIL project (IES, 2008) aims to collect data from existing and new samples in Member States and to analyze them at a central laboratory, in order to cross-check the previous measurements and to study soil and forest biodiversity in Europe. But, even using a central laboratory, possible errors due to transport and instability of soil samples could be introduced.

18.6 Conclusion

The spatial predictions of the CN ratio over European forest litters was done from CN ratio measurements across Europe and potential predictors related to climate, relief and tree species. Climate and relief are spatially continuous information whereas tree species have been interpolated by inverse distance weighting from punctual information. Two predictions were completed: one using a classical kriging on the CN ratio measurements and another one using a neural network approach on the environmental predictors. Both methods are quite similar and show globally the same spatial trends (same order of magnitude of the predictions and higher values in the southern and western parts of Central Europe). Therefore, the use of spatial predictors is not necessary when kriging. This allows for discussing and reanalyzing the quality of the data. A first check on data quality demonstrates that even with a common base methodology and common criteria, errors are introduced in the measurements, due mainly to cultural heritage, but also due to technical issues (material calibration, transport and instability of the samples . . .). The issue Europe is facing is quite representative of potential problems we can face during the building of global digital soil maps. The solution provided here for a worldwide mapping of soil attributes with legacy data is then to deal with prior information, containing prior metadata (when existing). Then make a first spatial prediction which allows for identifying errors or gaps in the metadata and then make a final prediction. Rescuing the data appears to be the most difficult problem for the final mapping. In terms of economics, it should be necessary to check if the rescue cost is not more than cost of getting new data.

References

Burke, I.C., Yonker, C.M., Parton, W.J., Cole, C.V., Flach, K.D., and Schimel, S., 1989. Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils. Soil Science Society of America Journal 53:800–805.

Casalegno, S., 2009. Forest species and forest types current distribution and habitat suitability under climate change. Retrieved from the European Commission Joint Research Centre website: http://forest.jrc.ec.europa.eu/climate-change. Data downloadable at: http://efdac.jrc.ec.europa.eu/climate (Last accessed 22 April 2010).

Casalegno, S., Amatulli, G., Camia, A., Nelson, A., and Pekkarinen, A., 2010. Vulnerability of Pinus cembra L. in the Alps and the Carpathian mountains under present and future climates. Forest Ecology and Management 259:70–761.

Desbarats, A. J., and Dimitrakopoulos, R., 2000. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. Mathematical Geology 32:919–942.

IES, 2008. The Biosoil project. http://forest.jrc.ec.europa.eu/ForestFocus/biosoil.html (last verified 10 September 2008).

Gobron, N., Pinty, B., Melin, F., Taberner, M., Verstraete, M., Robustelli, M., and Widlowski, J.L., 2006. Evaluation of the MERISIENVISAT fAPAR product. Advanced Space Research 37, doi: 10.101 6lj.asr.2006.02.048.

Pekkarinen, A., Reithmaier, L., and Strobl, P., 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. ISPRS Journal of Photogrammetry and Remote Sensing 64:171–183.

UNECE, 2003. ICP Forests Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests, 2003. Part IIIa Sampling and Analysis of Soil and Part IIIb Soil Solution Collection and Analysis. United Nations Commission for Europe Convention on Long-Range Transboundary Air Pollution, International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests.

Wackernagel, H., 1995. Multivariate Geostatistics – An Introduction with Applications, Springer, Berlin.