Mapping Topsoil Organic Carbon Content in Grand-Duchy of Luxembourg

Antoine Stevens & Bas van Wesemael
Georges Lemaître Centre for Earth and Climate Research,
Earth and Life Institute,
3 Place Pasteur - 1348 Louvain-la-Neuve,
Belgium
email: antoine.stevens@uclouvain.be

Simone Marx & Lionel Leydet
Ministère de l'Agriculture, de la Viticulture et de la Protection des consommateurs,
Administration des services techniques de l'agriculture,
Service de Pédologie,
72, avenue L. Salentiny - 9080 Ettelbruck,
Grand Duchy of Luxembourg
email: simone.marx@asta.etat.lu

September 22, 2014

Abstract

Providing localized predictions of soil properties is needed to assist soil surveyors and land managers, and inform the political debate with quantified estimates of the status and change of the soil resource. Maps can be produced with data originating either from purpose-built soil monitoring networks (SMN) or from previous soil measurements exercises such as soil testing for farmers by commercial/institutional soil laboratories. Although SMN's are likely to provide better and less biased estimates of soil attributes because of their optimal sampling strategy, SMN's are costly to establish, maintain and re-visit. Data gathered from other sources, on the other hand, are often more numerous which might favor greater accuracy in a geostatistical context. Another advantage is that these data are often acquired continuously so that derived-maps can be rapidly and easily improved or updated with upcoming data. In this study, we produce a map of the topsoil Organic Carbon (OC) content of croplands, grasslands, vineyards and forest land of the Grand-Duchy of Luxembourg using more than 2,000 samples analyzed for OC (by dry combustion) in 2012-2014 by the national laboratory of Ministry of Agriculture, Viticulture and Consumer Protection in Luxembourg. We also provide based on this dataset simple OC statistics for combinations of land cover and soil associations, to set benchmarks OC levels to which future analyses can be compared. To model OC content, this study relies on a set of spatial covariates with a resolution of 90 m, including elevation and its derivatives, land cover, soil texture, climate and livestock intensity. Different spatial prediction models were developed separately for each of the four land cover classes using either Generalized Additive Models (GAM). Carbon content in cropland and forest soils show overall a clear south to north increase in OC in Luxembourg due to a climatic gradient, the North being wetter and colder than the south. In grasslands, OC distribution is highly related to soil types, with more OC in clay-rich soils. Carbon content in vineyards can vary greatly in fields close to each others and show very little spatial structure. For cropland, soils, the model is characterized by a $R^2 = 0.66$ and RMSE = 5.5 g C kg^{-1} as computed on a validation set. Models developed for the other land covers were less successful ($R^2 < 0.40$). This could be explained either by the relatively low number of sampling points (grasslands) or with the fact that a large part of OC variation occurs over short distances or without any important relationship with the covariates (forest, vineyards). A final, map combining maps of the four land covers and covering almost 90 % of the territory provides detailed information on the current status of OC content in soils of Luxembourg.

Contents

Chapter 1

Introduction

Since the Industrial Revolution, human actions have become the main working force behind global environmental changes, pushing the Earth System outside the stable environmental conditions prevailing during the *Holocene* [1]. The same applies to the soil resource, that human activities are rapidly altering, influencing not only the global carbon budget [2, 3] but also triggering soil degradation [4] and loss of soil quality over large extent [5], thereby threatening the soil capacity to provide future ecosystem and social services or support biomass production. The quantity and quality of Soil Organic Carbon (OC) plays a key role in mitigating these degradation processes and, hence the need for monitoring OC in space and time has intensified in order to target effective remediation measures or political instruments [6, 7]. Soil OC is indeed at the cross-roads of many soil functions and services through its influence on soil structure and resilience to erosion, nutrient holding capacity and biological activity [8, 9]. In European agricultural lands, soil OC is under pressure and tends to decrease due to ploughing and a decreasing supply of carbon to arable soils with the intensification of agricultural management [10, 11]. In Europe, the decline in OC is identified by the European Commission as one of the main threat to the soil resource in its proposal of a Soil Framework Directive to protect and preserve soil and its associated functions [12]. There are still debates on the lower limits under which OC cannot fall but there are indications that these thresholds depends highly on percentage of clay and climatic conditions [13].

The state and change of the soil resource over large extent are often assessed through the establishment of soil inventories or Soil Monitoring Networks (SMN) and digital soil maps. A SMN can be defined as a set of locations where changes in soil properties are periodically measured [14]. Several types of SMN's can be distinguished [14]: i) SMN's build on the purpose of monitoring the state and change of the soil attributes, ii) resampling of locations where previous measurements have been made for other purposes, and iii) SMN's based on soil databases gathering routine soil analyzes for farmers or other applied experiments. There are several examples of national-scale SMN's already implemented in Europe or across the globe [15].

Using numerical modeling (geostatistics) based on SMN's databases, digital maps are produced to give a representation of the spatial distribution of the soil properties. There are several ways to create a map of a soil property. Spatial prediction models can be broadly classified into mechanical (or empirical) models and statistical (or probability) models [16]. The first category includes techniques such as *Thiessen polygons*, *Inverse Distance Interpolation*, *Splines Interpolation* and for which no strict assumptions of the variability of the studied variable exist. In the second category, model parameters are estimated using the probability theory and predictions are computed with their associated errors. Statistical models includes for instance *kriging*, *environmental regression* and *hybrid techniques* (e.g. regression-kriging). Kriging uses the information contained in the spatial auto-correlation of observed points/polygons to interpolate data to unknown locations. Environmental regression, on the other hand, produce spatial prediction equation by using sampling locations for which both inputs (covariates) and outputs (soil variable) is known. Provided that inputs are available as spatially-continuous layers, the environmental regression technique is able to infer soil properties at un-sampled locations.

One of the simplest approach to environmental regression is to divide the territory into LandScape Units (LSU) having the same soil characteristics and land use. Then, the mean (or another robust estimate) of the observed values in each LSU is computed and assigned to the LSU's (this is actually equivalent to a simple linear regression with a categorical variable). Such approach has for instance been followed by [17] to create a map of topsoil OC in Belgium. To produce more realistic results, however, a multiple regression can be fitted to observed data using a set of soil covariates[18]. Soil covariates

are environmental factors that are controlling soil formation and explaining the spatial variability of the target variable. Such regression approach has been popularized by McBratney et al. [19] and formalized in the so-called *scorpan* equation (1.1):

$$P_x = f\,(s, c, o, r, p, a, n) \tag{1.1}$$

where P_x is the target variable at location x, f is the spatial prediction equation including factors related to soil (s), climate (c), organisms, land use and human effects (o), relief (r), parent material (p), age (a) and spatial position (n). The environmental regression approach is getting more popular as these layers of soil covariates are becoming widely available with the rise of airborne and spaceborne sensors producing space- and time-resolved geophysical and ecosystemic data. Basically, the function f can be any of type of statistical model such as linear regression, stepwise multiple linear regression, Generalized Additive Models (GAM). With the increase of the size of the SMN's and in the number of available soil covariates, models coming from the field of machine learning such as boosted regression trees [20], decision trees [21], random forests [22], etc are increasingly exploited. Kriging and environmental regression can be combined into hybrid model strategies, which can take into account both deterministic and stochastic sources of soil variation [23].

Creating digital maps of soil properties from these large SMN's can be particularly challenging because the observed variation in the soil properties are the result of driving factors acting over very different spatial scales [14]. The accuracy of digital maps depend not only the sampling density of the SMN network [24] but also on its spatial extent and spatial resolution of input data [25]. Organic carbon, as other soil attributes, vary greatly in space as a function of *scorpan* factors. The effects of some factors that can been observed at small scales disappears when considering OC variations of large geographical extent and inversely. For instance, the well-documented effect of land clearing on OC levels in Brazilian Amazon based on local studies could not be observed clearly at regional scales [26]. Inversely, the influence of climate on OC stocks in Northeast China is weakening with decreasing spatial scale [27]. Hence, environmental correlations with the target variables vary from place to place and from one scale to another[16].

This report presents results of a study aiming at the digital mapping of topsoil OC in Grand-Duchy of Luxembourg (GDL), commissioned by the *Administration des Services techniques de l'Agriculture* (ASTA) of the *Ministère de l'Agriculture, de la Viticulture et de la Protection des consommateurs*. Surface OC maps have been already produced in neighbouring countries such as in Belgium [28, 29], France [30] and Netherlands [31] but it is the first time such a mapping exercise is realized in GDL. Moreover, such map is needed in the context of the Soil Framework Directive for which the European Commission asked State Members to provide, on a voluntary basis, a map of topsoil OC content which will contribute to the *European Environment Information and Observation Network for soil* (EIONET-SOIL) project [31].

Land cover	Area (ha^-1)	Relative area (%)
Artificial land	28440	10.97
Cropland	56779	21.90
Grassland	70392	27.15
Vineyard	1359	0.52
Other Agricultural land	3202	1.23
Forest	97832	37.73
Wetland	218	0.08
Water	1042	0.40
all	259264	100.00

Table 1.1: Area and relative area of land cover classes in GDL (source: OBS)

This study adopts on the *scorpan* methodology to create a state-wide map of OC content in four different land use classes (cropland, grassland, forest and vineyards) representing approx. 87 % of the GDL territory (Table 1.1). The modeling is based on a soil database collected by the *Service de Pédologie de l'ASTA* including more than 2,000 samples analyzed for OC (by dry combustion) in the 2012-2013 period (which we name afterwards ASTA-SOC) and a set of soil covariates that has been gathered from various sources. Additionally, the report will also present a descriptive analysis of the soil database giving likely ranges of observed OC content in soils of GDL. ASTA-SOC mainly includes samples that have been analyzed for testing soil parameters for fertilizer advices and therefore their locations have not been selected following an optimal sampling strategy. Some locations for instance might appear

strongly clustered. However the number of samples in ASTA-SOC is relatively large, which might favor greater accuracy in a geostatistical context. Moreover soil data are acquired continuously at ASTA so that derived-maps can be rapidly and easily improved or updated with upcoming data.

Chapter 2

Methodology

2.1 Geo-pedological description of the GDL

The Grand-Duchy of Luxembourg (GDL) is a country of 2,586 km^2 that shares its borders with France in the south, Germany in the east, Belgium in the west. It benefits of a temperate semi-oceanic climate with mean temperature ranging from 7.5 °C to 9.5 °C . The country can be divided in two main natural regions (Figure 2.1). In the north, the Oesling, like the Ardennes in Belgium and Eiffel in Germany, is a massif of the Primary Period made of Lower Devonian slate and quartzite that were highly folded during the hercynian orogeny. The Oesling is now an sub-horizontal peneplain with deeply incised valleys and a mean altitude of ca. 500 m. In the centre and south of the country, the Gutland is a more heterogeneous region characterized by a south-west -facing cuesta topography which developed on monoclinal Triassic and Jurassic sediments. Rocks formed during the same period can be found in the Gaume region in Belgium, north of the Lorraine in France and Bitburger Gutland in Germany. Triassic deposits are made of marls, sandstone and dolomites, all containing mineral dolomite while Jurassic sediments are made of sandstone and marls with calcium carbonates. The Oesling is predominantly covered by shallow stony loam soils (Cambisols in the WRB classification [32]) while soils of the Gutland are mainly Luvisols. The soil association map of GDL (1:100,000) shows 26 soil associations that we further regrouped into 10 classes (Table 2.1; Figure 2.2) representing variations in texture and mineralogy.

#	Soil association	Region	Geologic Period	Provisional WRB classification [32]
1	Oesling	Oesling	Lower Devonian	skeletic dystric Cambisol (siltic)
2	Argiles lourdes des schistes bitumineux	Gutland	Jurassic	vertic calcaric Cambisol (clayic)
3	Argiles lourdes du Keuper	Gutland	Triassic	vertic dolomitic Cambisol (clayic)
4	Argiles du Lias Inf. et Moyen	Gutland	Jurassic	gleyic/stagnic endocalcaric Luvisol (loamic)
5	Dépôts limoneux sur Grès	Gutland	Jurassic	haplic Luvisols (loamic)
6	Grès de Luxembourg	Gutland	Jurassic	haplic Luvisols (arenic), Arenosols
7	Calcaires du Bajocien	Gutland	Jurassic	leptic calcaric Cambisol (loamic)
8	Dolomies du Muschelkalk	Gutland	Triassic	leptic dolomitic Cambisol (loamic)
9	Bundsandstein	Gutland	Triassic	endodolomitic Luvisol (loamic)
10	Autres	Gutland/Oesling	Alluvium, colluvium	Fluvisols, Cambisols, Regosols

Table 2.1: Soil associations and corresponding WRB classification.

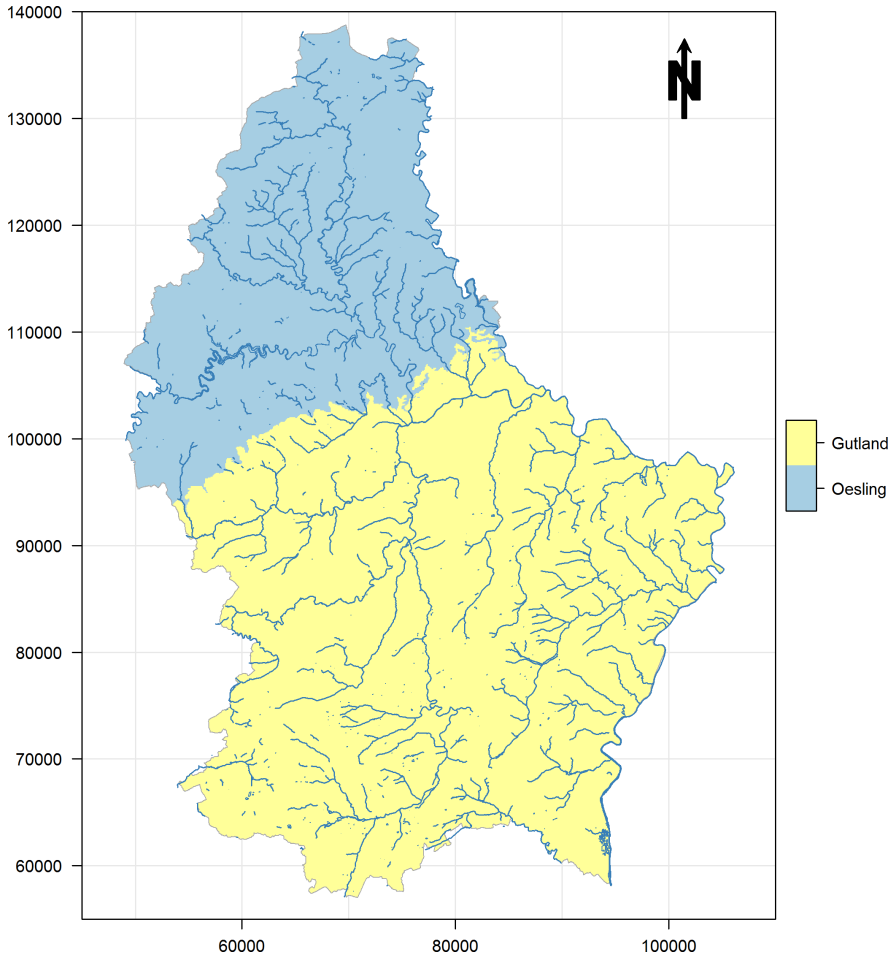

Figure 2.1: The Oesling and Gutland regions, and the river network.

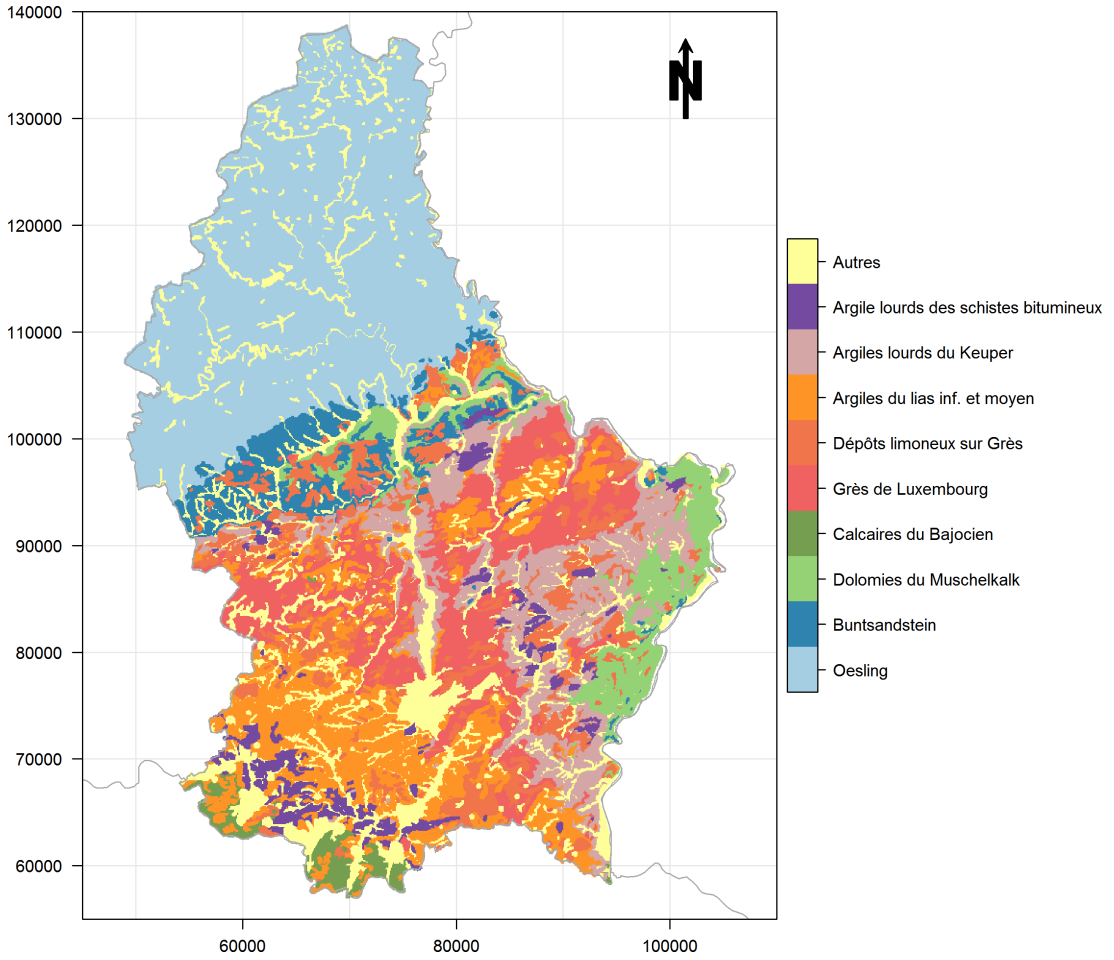

Figure 2.2: Soil associations used in this study.

2.2 ASTA-SOC

2.2.1 Preamble

ASTA-SOC is a compilation of OC data analyzed at the soil analytical laboratory of ASTA. Originally, the database contained more than 8,000 OC analyzes for agricultural lands and about 1,000 analyzes for forest land that were measured during the 2008-2013 period. Various checks were carried out to ensure data quality: duplicated observations and outlying values were removed, etc. Samples were dried at 95 °C, sieved by 2 mm and reduced to powder in a mortar before SOC analysis. At ASTA, samples were analyzed during the last 5 years with different instruments and slightly different protocols (Table 2.2) which might induce temporal discrepancies between OC analyzes due to inter-instrumental errors. An analysis of the replicate error between the OC analyzes produced with the TruSpec CN (LECO Corporation, Michigan, USA) and the Multi EA 4000 (Analytik Jena AG, Germany) is given in Annex A. The replicate errors were found too large (between 3 and 7 g C kg^{-1}) and therefore only samples measured with the Multi EA 4000 instrument were kept in the database. In total, after data cleansing, a total of 3,492 samples were included in ASTA-SOC, including 984 samples taken in cropland, 323 samples in grasslands, 1,145 samples in vineyards and 1,011 samples in forests (Figure 2.3).

Year	Date of the last sample	Intrument & Methodology	Method†
2008	-	LECO RC 412	TC = dry Combustion at 500°C TOC = TC
2009	-	LECO RC 412	TC = dry Combustion at 500°C TOC = TC
2010	21/02/2011	LECO TruSpec CN	TC = dry combustion at 900°C TIC = dry combustion after heating at 500 °C for 5h TOC = TC-TIC
2011	20/01/2012	LECO TruSpec CN	
2012	23/01/2013	Analytik Jena Multi EA 4000	TC = dry combustion at 1200°C TIC =Measure of CO2 after treatment to phosphoric acid 40% TOC = TC-TIC
2013	06/09/2013	Analytik Jena Multi EA 4000	

† TC = Total Carbon; TOC = Total Organic Carbon; TIC = Total Inorganic Carbon

Table 2.2: Methods of OC analysis used at the ASTA soil laboratory

2.2.2 Soil database for cropland, grassland and vineyards

Sample were collected in agricultural fields all over GDL for soil testing. The sampling unit is the parcel (polygons) for which one composite sample is taken to estimate its (average) soil properties. The depth at which samples were collected differs from one land cover to another: 0-25 cm in croplands, 0-10 cm in grasslands and 0-30 cm in vineyards. Each parcel has an identifier (named *FLIK*) corresponding to a unique agricultural parcel in the the official land parcel information system (LPIS) of GDL, with a mean area of 1.8 ha, which allowed to retrieve the location of the soil analyzes by merging the soil data with the parcel plan. The position of the samples were defined as the centroid of the parcels. We should note that converting areas (parcels) to points using polygon centroids is a great simplification and is not strictly appropriate as it assumes that the spatial support is constant in shape and size (see [33]) but this will greatly facilitate the subsequent spatial modeling of OC. Only the last soil analyzes done for one *FLIK* were kept in the database. Metadata associated with the soil analyzes were available such as cultivation type and soil texture in four broad classes: light (L), moderately light (M), heavy soils (S) and a special class for samples from the Oesling area (OM)[1].

[1] These classes correspond to texture classes of the Luxemburgish soil classification: L = {S+Z}; M = {L+P+E+A}; OM = G; S = U

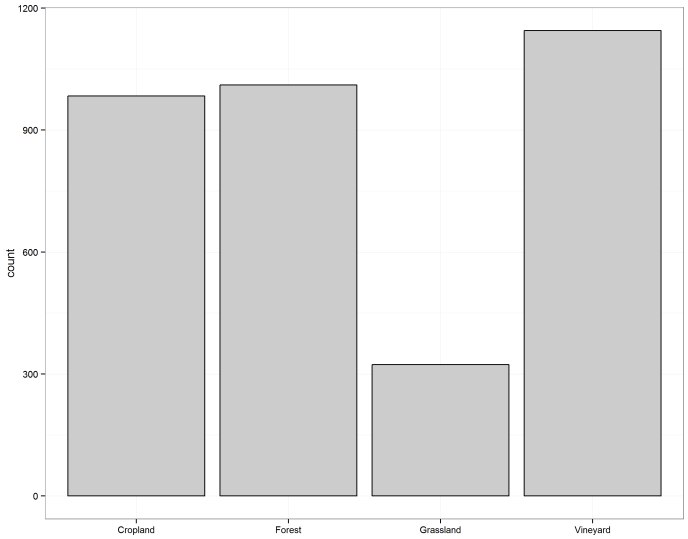

Figure 2.3: Total number of samples in the database per land cover class

2.2.3 Soil database for forests

Samples were collected from 1998 to 2001 in the framework of *Inventaire Forestier National* (IFL) [34], a survey aiming at providing information on the state of the forest resources in GDL during which 1,600 sampling points located on a grid (500 m by 1000 m) were visited. One of the aspects of the survey concerned forest soils. Composite soil samples were taken at two depths (0-20 and 20-40 cm) within circular sampling plots of max. 18 m of radius located on the national grid. Soil analyzes are still ongoing. Around 1,000 samples for the 0-20 cm layer have been already measured by ASTA and includes the following attributes: pH_{CaCl_2}, pH_{water}, P, N, Na, Mg and OC. Carbon content was analyzed with the LECO TrueSpec CN instrument from 2011 to 2013. In addition to geographical coordinates, the IFL database contains several environmental variables such as forest stand, canopy cover, stand age, type of humus and texture, etc that can be linked to the soil database.

2.3 Soil covariates

A set of spatial layers in raster format were prepared with a resolution of 90 meters and with the same grid topology. While some of the soil covariates were initially available at a smaller resolution (e.g. the digital elevation model has a 5 m pixel resolution), we resampled all rasters to the resolution corresponding to the one of the raster with the maximum resolution. These operations and generally all the manipulations related to spatial data were realized with the raster [35] and sp [?] R packages. We describe thereafter how we created a stack of soil covariates for the spatial modeling topsoil OC in GDL.

2.3.1 Digital Elevation Model (DEM) and its derivatives

We used the DEM from the *Base de Données TOPO/CARTO* (BD-L-TC) altimetric product of the *Administration du Cadastre et de la Topographie*, with a resolution of 5 m. Using the SAGA-GIS software [36], we derived from the DEM a series of morphometric and hydrologic variables including: slope, aspect, profile, plan and general curvature, topographic position index (TPI [37]), flow accumulation, upstream slope length, LS factor of the RUSLE equation and wetness index (Figures 2.4 and 2.5).

Following Zar [38], we converted aspect (in degrees) into two separate continuous variables according to Eq. 2.1 and 2.2:

$$eastness = \sin\left(aspect \cdot \pi/180\right) \tag{2.1}$$

$$northness = \cos\left(aspect \cdot \pi/180\right) \tag{2.2}$$

Northness and eastness represent the degree to which aspect is close to the north or to the east and take values in the range [-1; +1].

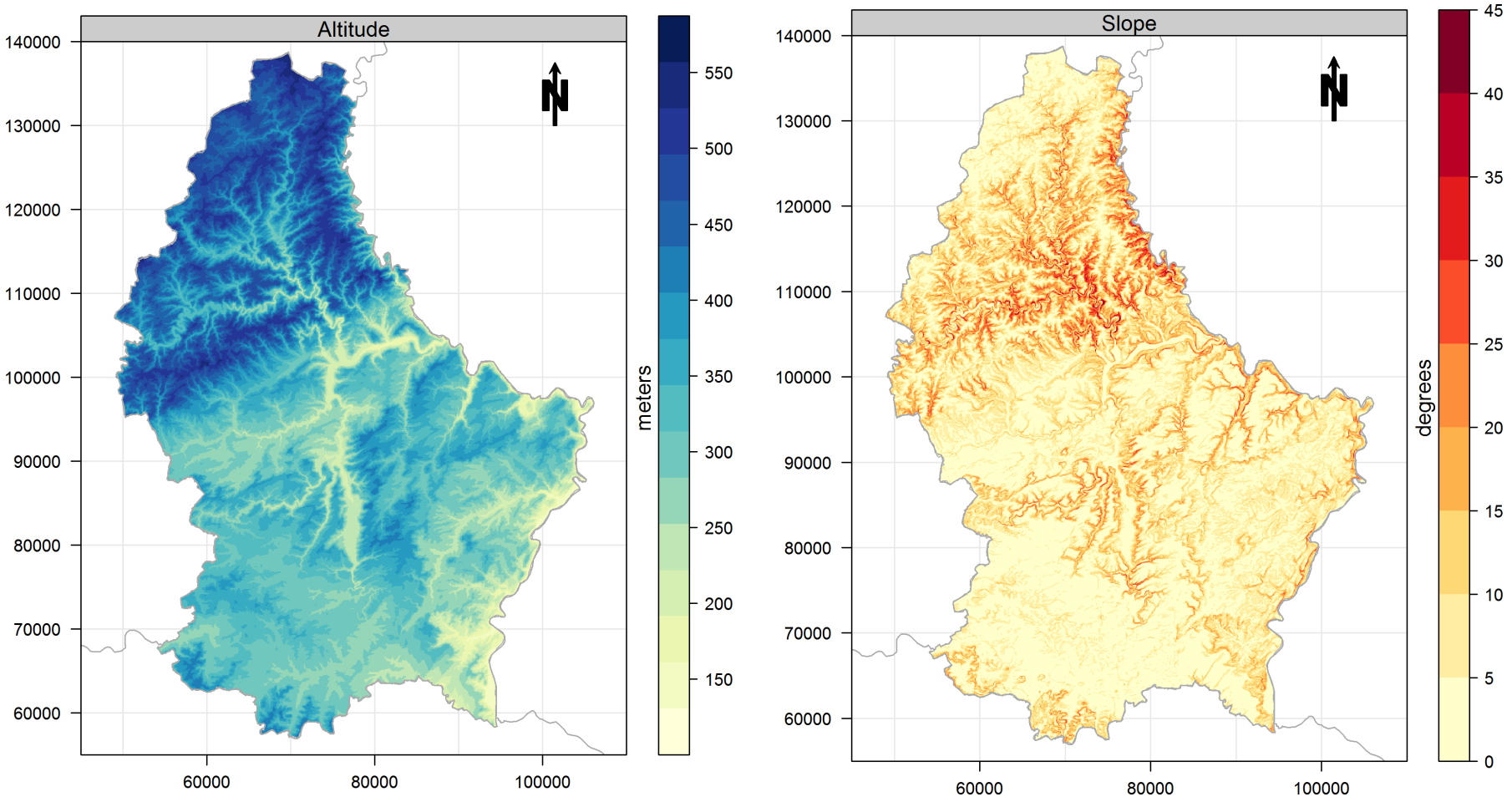

Figure 2.4: Digital Elevation Model (/meters) and slope (/degrees).

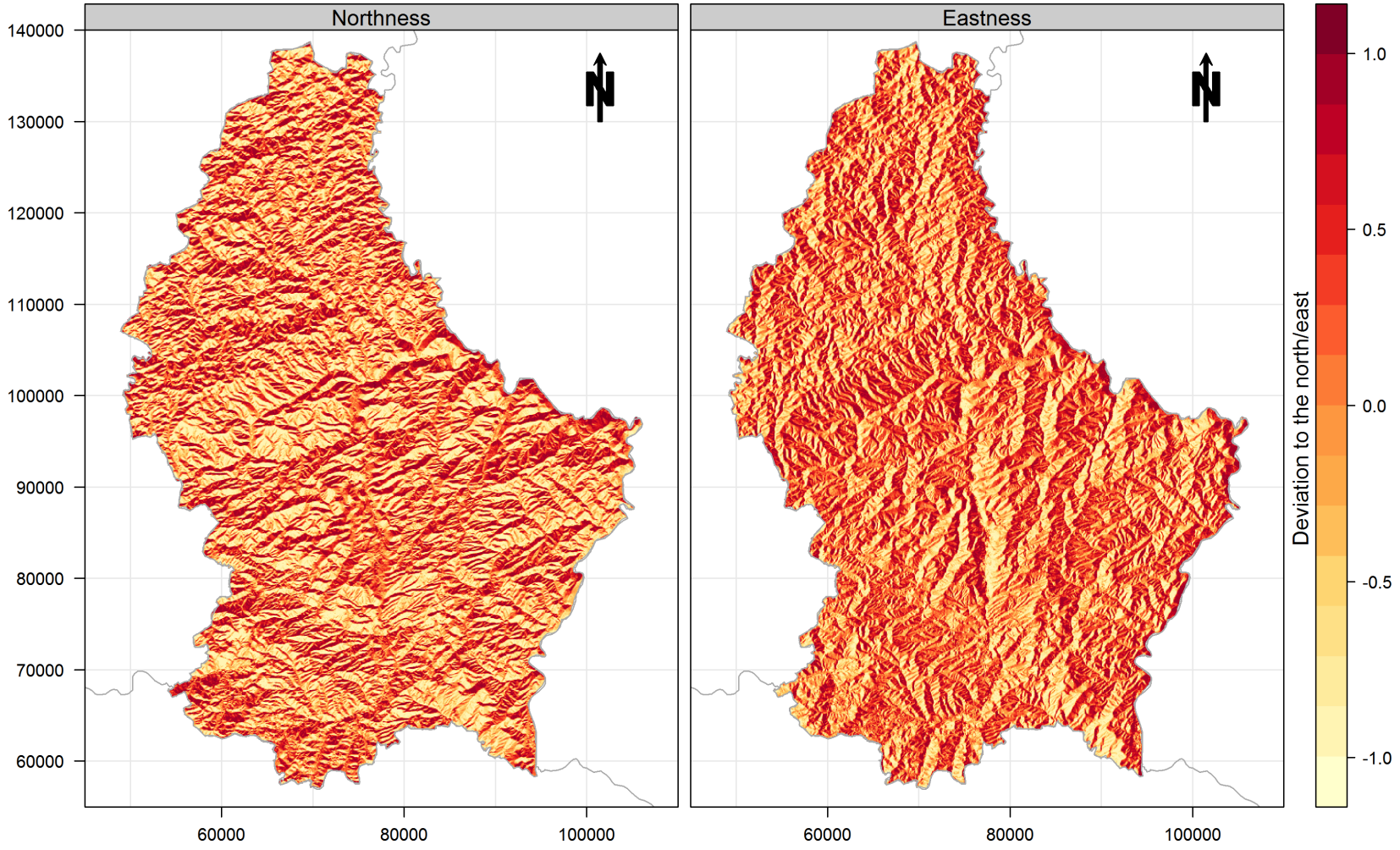

Figure 2.5: Eastness and northness derived from the DEM.

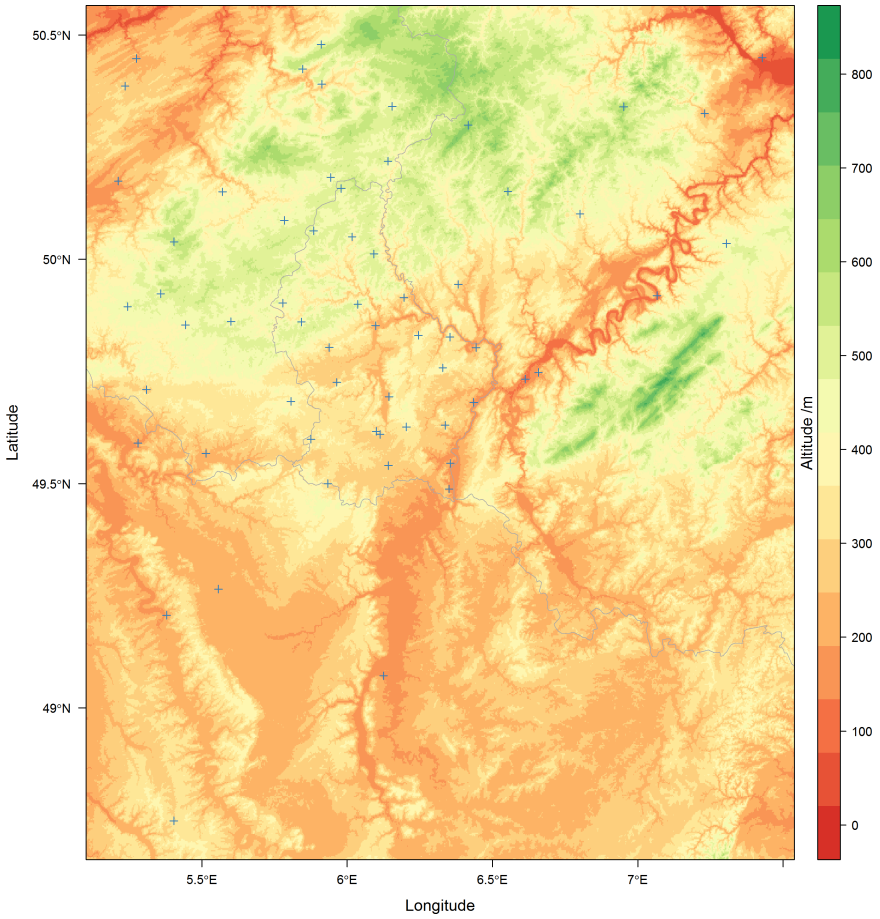

Figure 2.6: Location of meteorological stations (+) used to map average temperature and precipitation patterns in GDL.

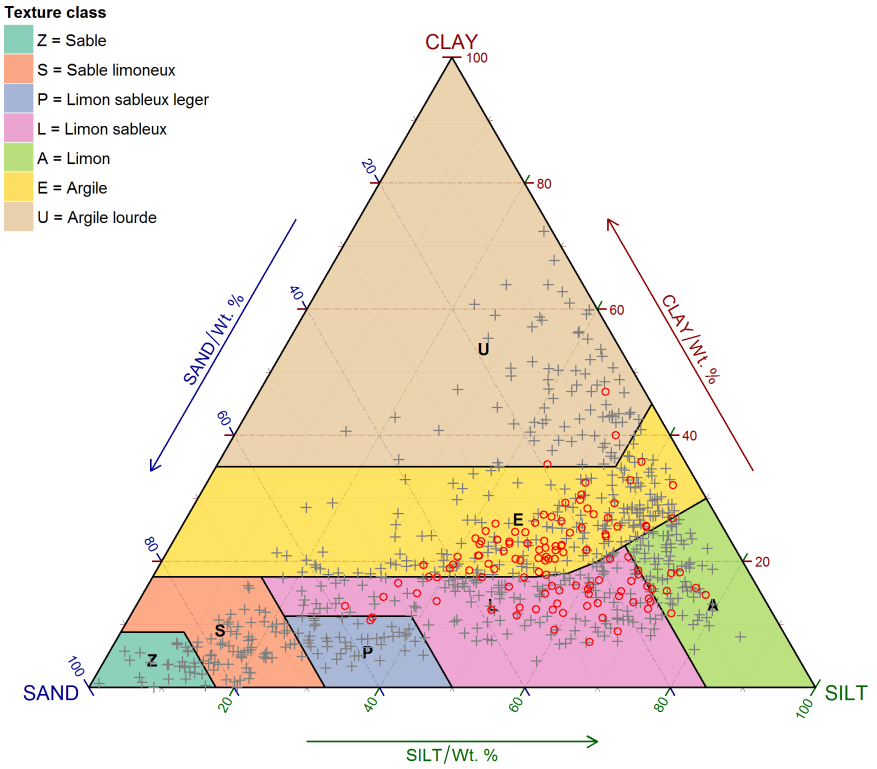

Figure 2.7: Samples of BDSOL1964-1973 (+ = Gutland; o = Oesling) in the soil texture classification triangle of GDL.

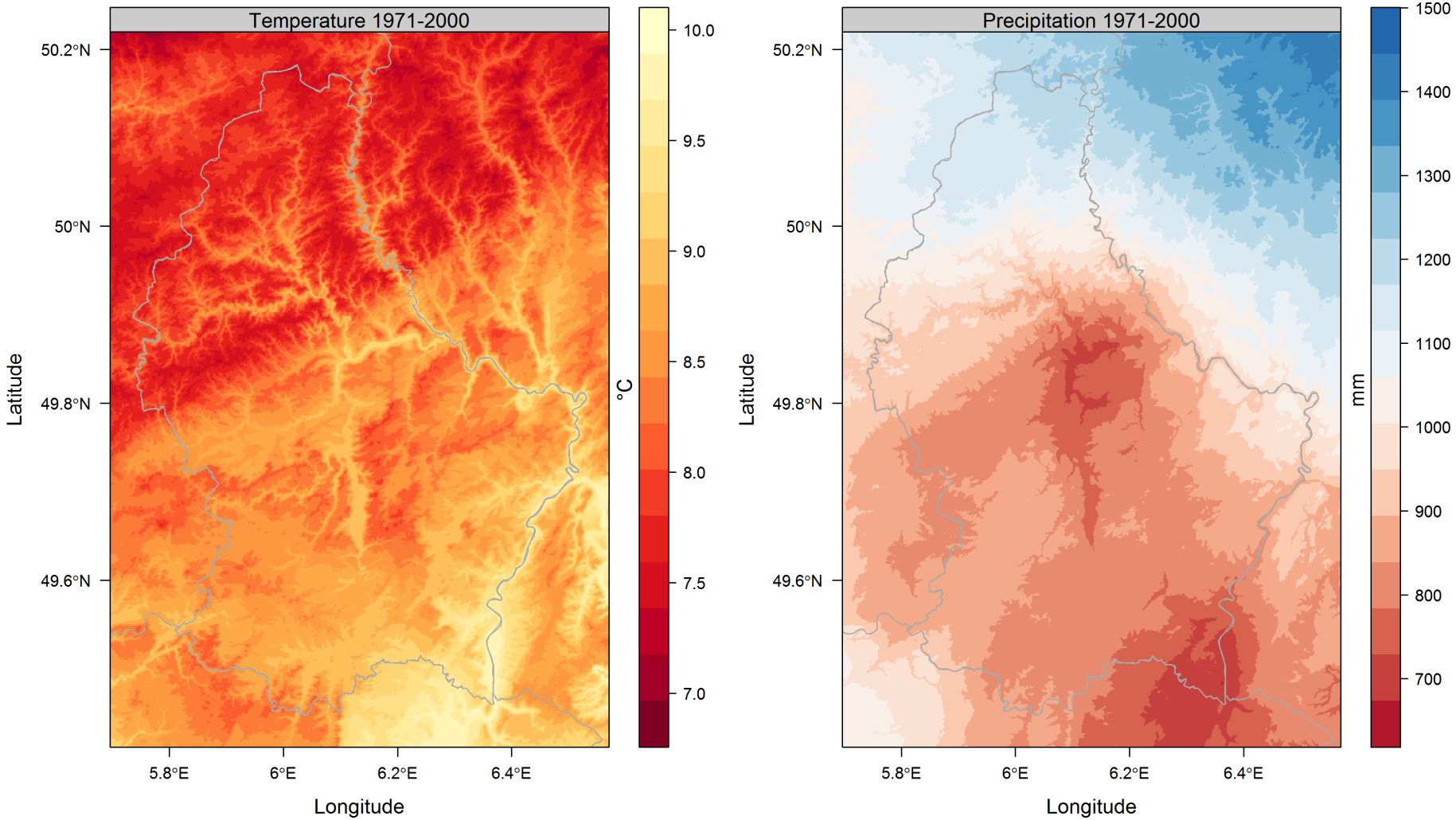

Figure 2.8: Mean temperature (/°C) and precipitation (/mm) during the 1971-2000 period.

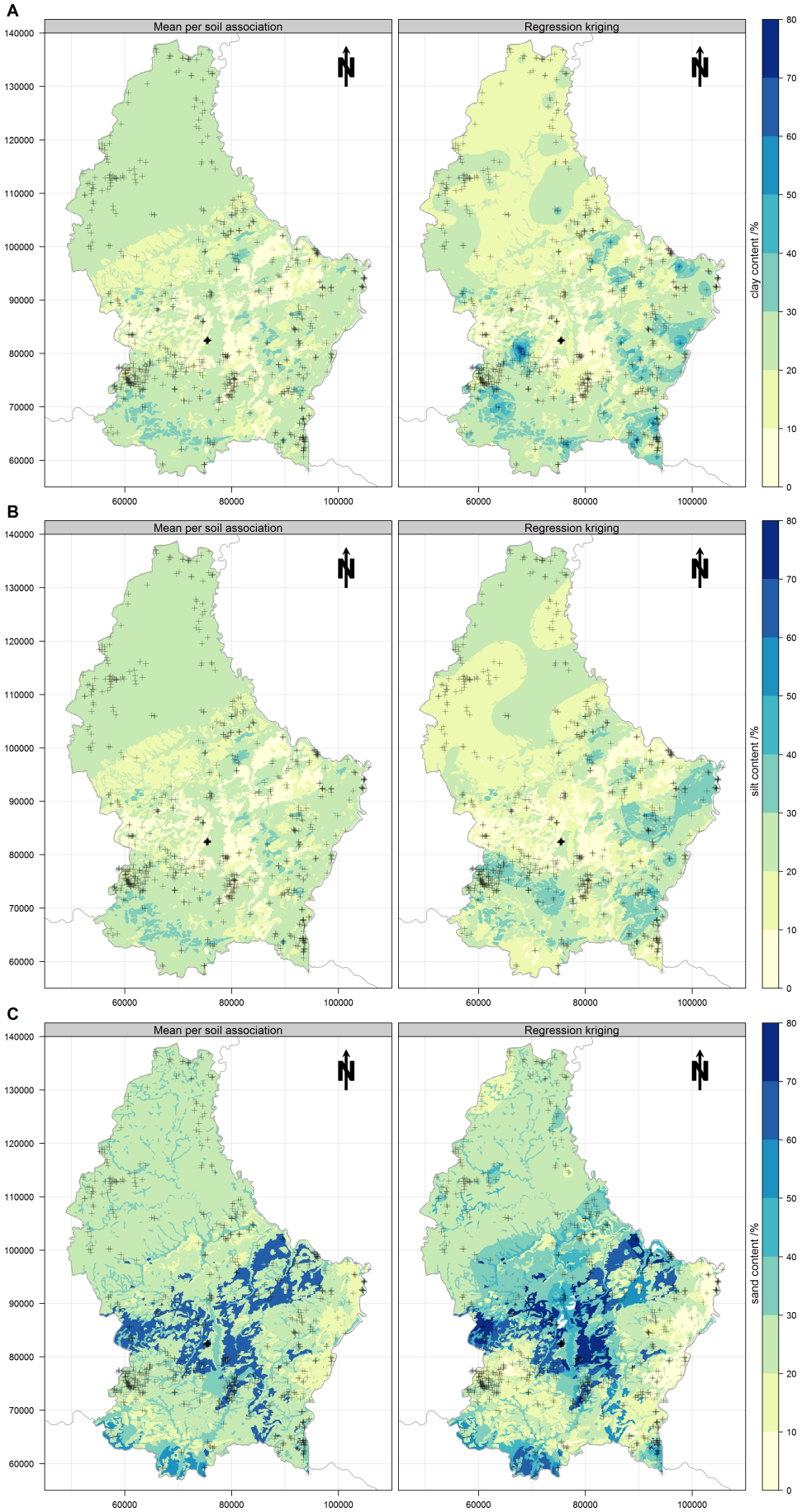

Figure 2.9: Map of (A) clay, (B) silt and (C) sand content as produced by assigning mean values per soil associations (left panels) or regression kriging (right panels). Location of the profiles are indicated with crosses (+).

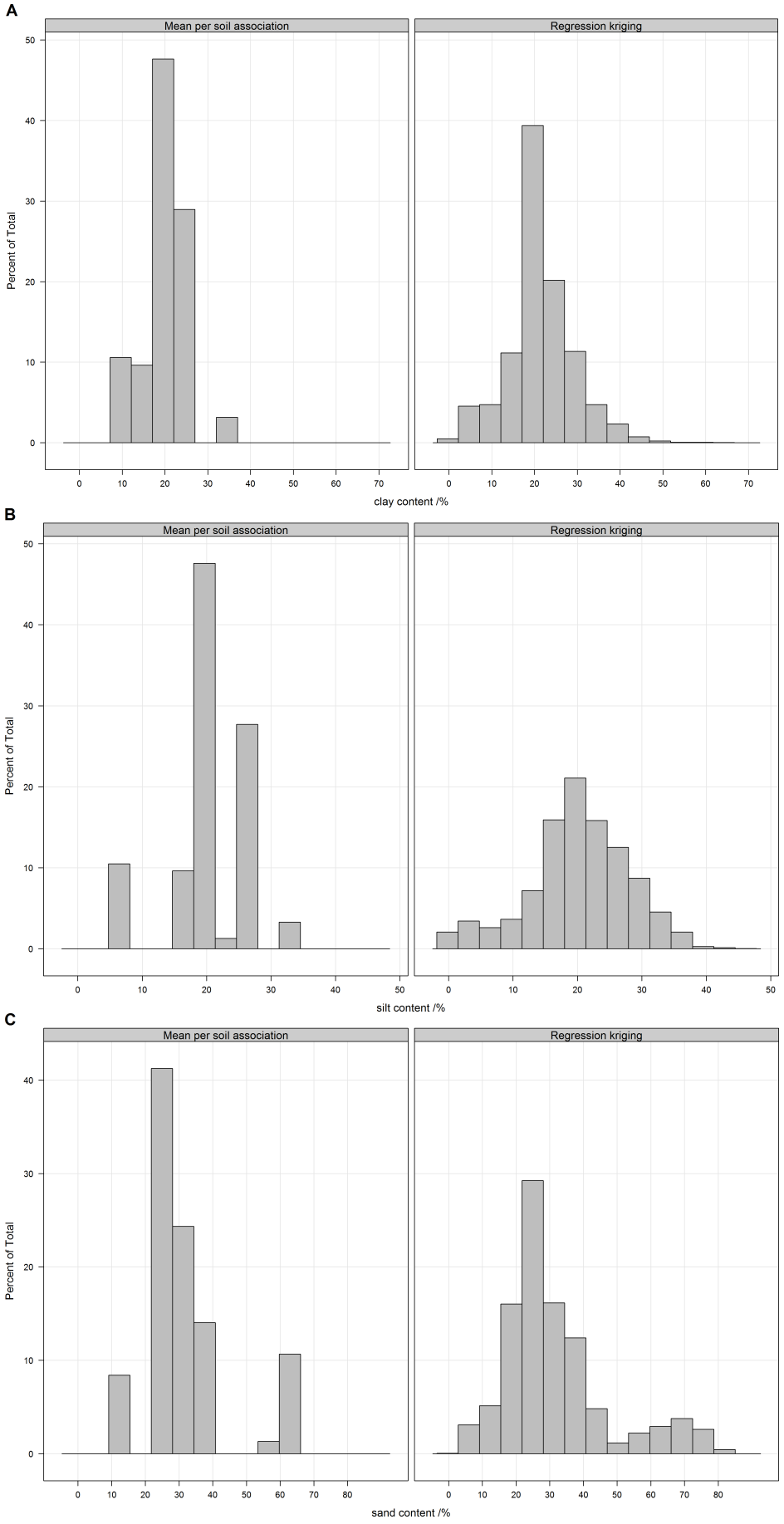

Figure 2.10: Histogram of pixel values of the (A) clay, (B) and (C) sand maps as predicted by the mean per soil association method (left panels) and regression kriging (right panels).

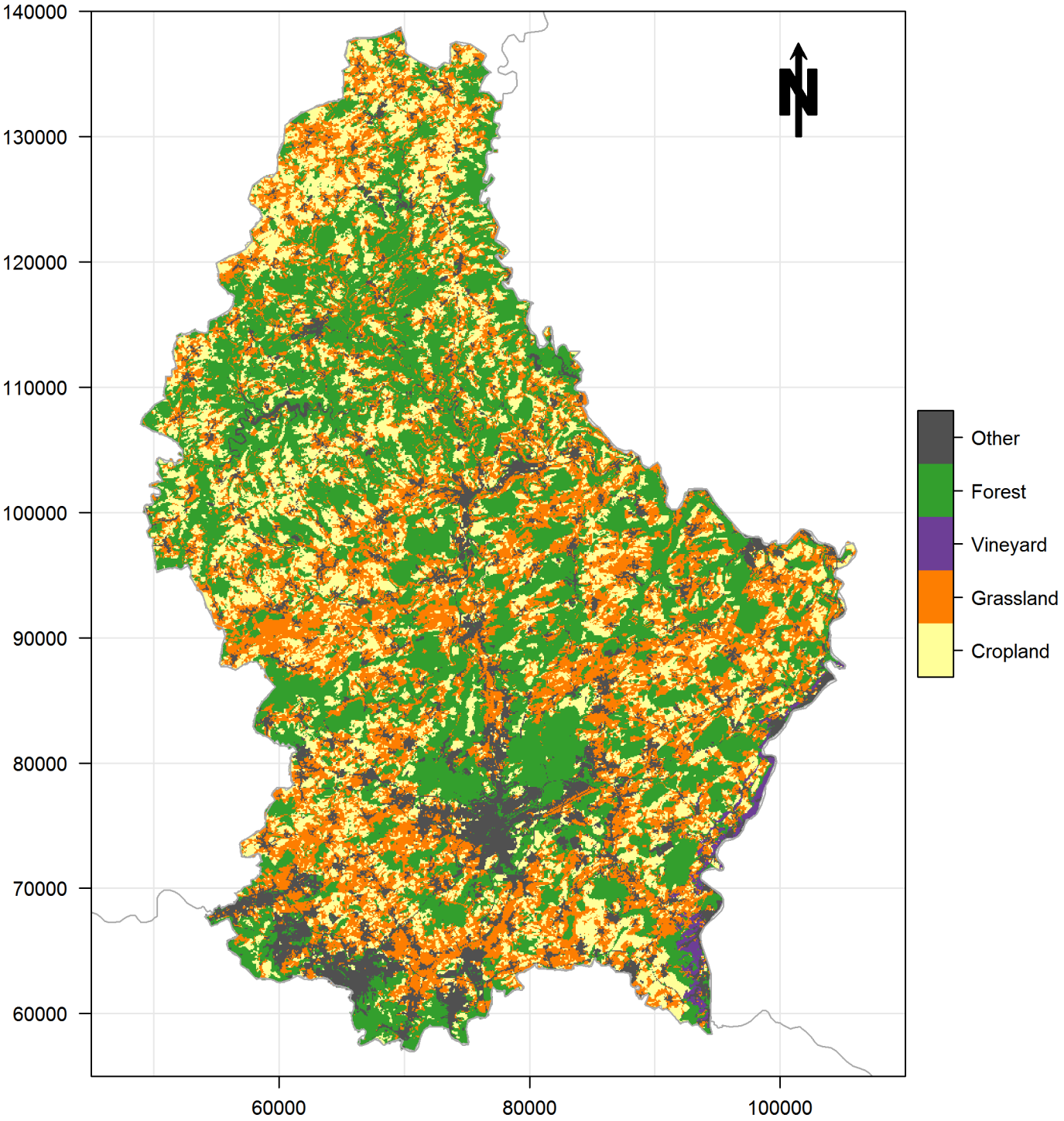

Figure 2.11: Reclassification of land cover classes from the OBS dataset.

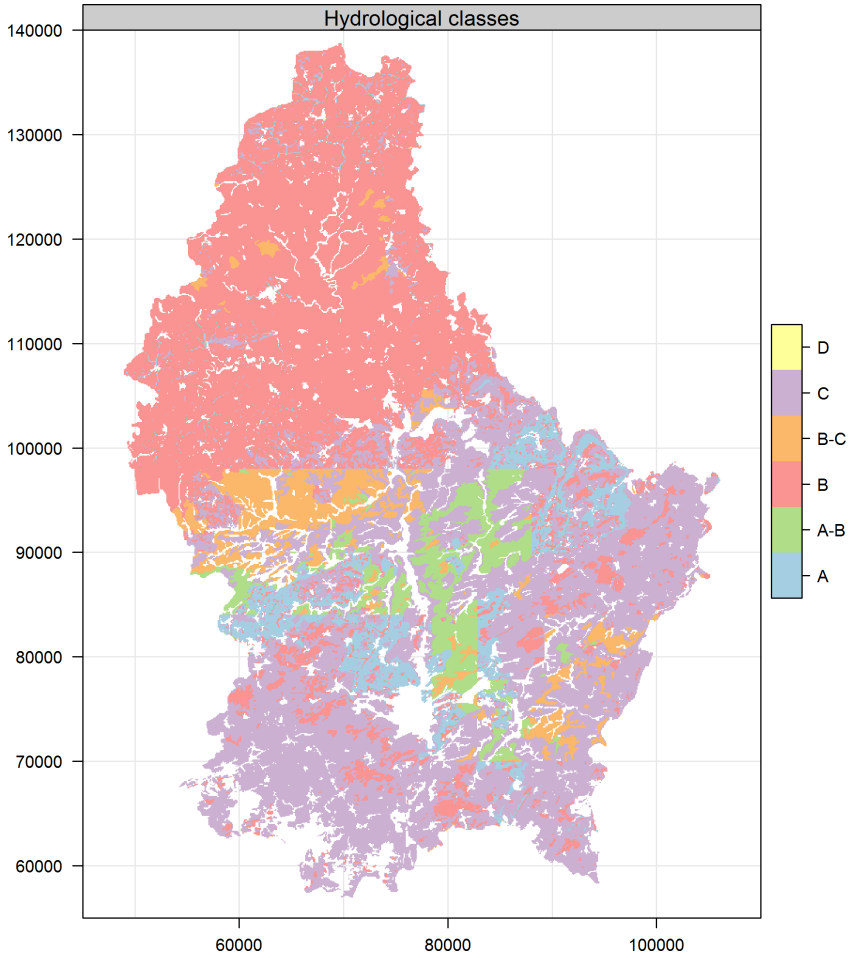

Figure 2.12: Hydrological classes as defined in the ERRUISSOL project.

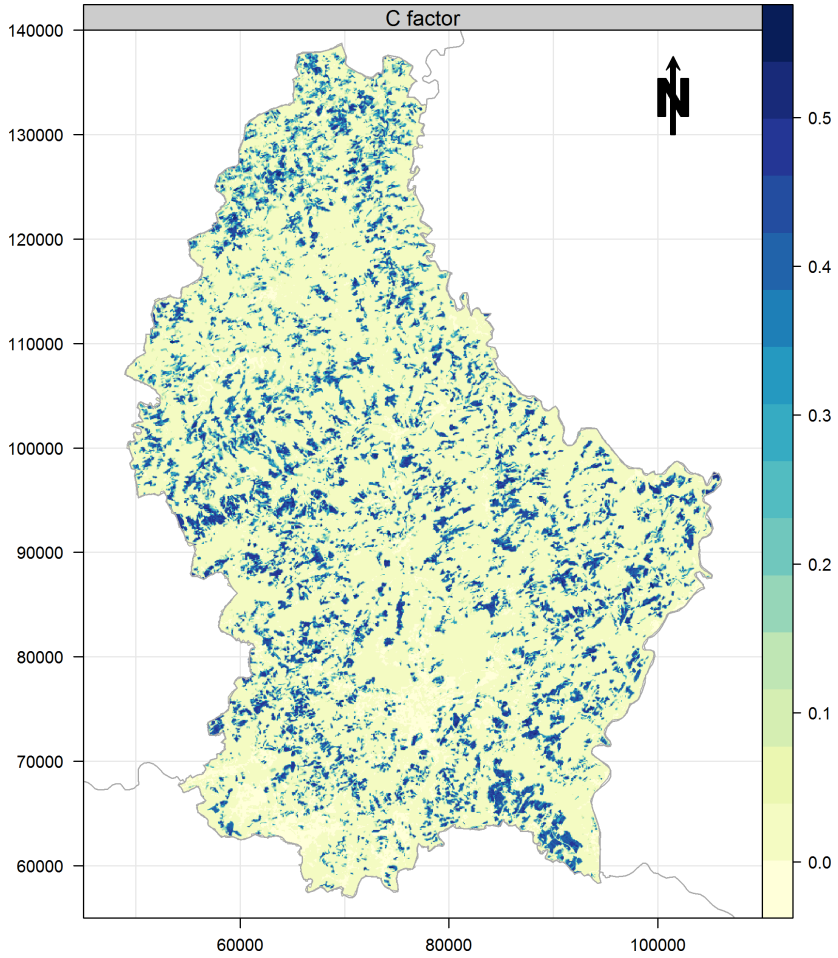

Figure 2.13: C factor from the ERRUISSOL project.

2.3.2 Temperature and precipitation

Spatial layers representing climatic variations in GDL were created by spatial interpolation of temperature and precipitation average for the period 1971-2000 data from weather stations in GDL and neighboring countries (Figure 2.6). Aggregated meteorological data of Luxembourg weather stations were obtained from the *Observatory for Climate and Environment, Department of Environmental and Agro-Biotechnologies of the Centre de Recherche Public - Gabriel Lippmann.* This dataset includes precipitation records of 25 stations and temperature records for 7 stations. We combined this dataset with weather data of Belgium (Koninklijk Meteorologisch Instituut, KMI), France (Meteo France) and Germany (Deutscher Wetterdienst, DWD) obtained from Dr Jeroen Meersmans (Exeter University) and that he gathered for creating precipitation and temperature maps of Belgium [39]. Climatic maps were created by modeling temperature and precipitation with altitude using thin-plate splines regression. Using altitude as covariate for mapping climatic variables can improve predictions dramatically [40]. The smoothing parameter is chosen automatically by generalized cross-validation. Elevation data were derived from the Shuttle Radar Topography Mission (SRTM) mission of the NASA [41]. The resulting temperature and precipitation maps are given in Figure 2.8.

2.3.3 Soil

Three maps of soil texture (sand 50 μm-2mm, silt 2-50μm, clay $<$ 2 μm) were created based on a database of 766 historical soil profiles from the national soil database (BDSOL1964-1974) analyzed for granulometry. Only the first soil layer (0-10cm) was kept in the database. The texture of these samples is shown in Figure 2.7. We adopted two approaches for the spatialization of the observations. First, we computed the mean sand, silt and clay content per soil association (Figure 2.2) and assigned the resulting values to the corresponding spatial polygons. In a second approach, we further computed the residuals between the observations and the mean and kriged the residuals in space (i.e. regression kriging). Maps of texture obtained with the two strategies are shown in Figure 2.9. The regression kriging maps show generally more detailed spatial patterns. Also, the histogram of the pixel values of the three maps indicate a more realistic outcome for regression kriging approach (Figure 2.10). In addition to soil texture, we used a spatial layer representing the hydrological status of soils as estimated by a project also commissioned by ASTA called ERRUISSOL[2] aiming at the mapping of risks of erosion and runoff in GDL. The hydrological status of soils is divided into four classes, having high (A) to low (D) infiltration capacity (Figure 2.12) and was defined as a function of texture, depth, stratification and drainage of soils and are based on the the 1:25,000 soil map of GDL and the soil association map for areas that are still not mapped at the 1:25,000 scale.

2.3.4 Land cover & human influence

Land cover data was obtained from the *Occupation Biophysique du Sol* (OBS), an inventory of the land cover in GDL at the scale of 1:10,000 based on aerial photography dating from 1999. The 76 different land cover and biotope classes were reclassified into five main land cover: cropland, (permanent) grassland, vineyard, forest and other, as shown in Figure 2.11. The data is in polygon format and was therefore converted to raster with a 90 m resolution. We also incorporated in our study a raster map of the C (crop) factor of the Universal Soil Loss Equation (Figure **??**), computed through an analysis of the crop rotation systems 2010-2012 realized during the ERRUISSOL project (Figure 2.13). Finally, we included in the spatial modeling of OC the livestock intensity of 2012 expressed in fertilizing units per ha (UF = *unité fertilisante* / 80 kg N ha^{-1}) aggregated at the level of GDL municipalities.

2.4 Modeling ASTA-SOC and mapping surface OC in GDL

Attributing the values of the covariates (independent variables) to OC observations (dependent variable) in ASTA-SOC is required to develop the spatial models and produce a map. For soils under cropland, grassland and vineyards land cover, the sampling units are the parcels. The mean values of the covariates within each parcel was computed and attributed to the corresponding observations in ASTA-SOC through their *FLIK* number (see section 2.2.2). For IFL data, the values of the covariates under each sample location was simply extracted and attached to observations of ASTA-SOC.

[2] *Etablissement d'une carte thématique sur les zones à risque d'érosion et de ruissellement au Grand-Duché du Luxembourg à partir de la carte des sols.*

The spatial variation of OC was modeled through Generalized Additive Models (GAM) [42] as implemented in the mgcv package [43]. This regression technique can account for non-linear relationships in the data using a set of smooth functions. A general equation for a GAM regression model can be written as (Eq. 2.3):

$$E\left(Y \mid X_1, X_2, \ldots, X_p\right) = \alpha + f_1\left(X_1\right) + f_2\left(X_2\right) + \ldots + f_p\left(X_p\right) \tag{2.3}$$

where Y is the dependent variable, $X_1 X_2, \ldots, X_p$ represent the covariates and the f_i's are the smooth (non-parametric) functions. Since GAM is in essence an extension of generalized linear models, the conditional mean $\mu\left(X\right)$ of the dependent variable Y is related to an additive function of the covariates via a link function g [44](Eq. 2.4):

$$g\left[\mu\left(X\right)\right] = \alpha + f_1\left(X_1\right) + f_2\left(X_2\right) + \ldots + f_p\left(X_p\right) \tag{2.4}$$

Due to the logarithmic distribution of OC in topsoils, we used the log-linear link function $g\left(\mu\right) = \log\left(\mu\right)$. The GAM model is build using penalized regression splines and the smoothing parameters were estimated by Maximum Likelihood. An extra penalty is added to each term so that it can each term can potentially be set to zero during the fitting process. The model for each land cover was developed on a training set containing two-thirds of the total number of samples, selected by random sampling within subgroups of the data defined by the percentiles of the response. The remaining samples were assigned to a test set to assess model performance. A first model with all the covariates was developed on the training set and a backward selection of the terms was followed using their approximate p-values. This was done by sequentially dropping the single term with the highest non-significant p-value from the model and re-fitting until all terms are significant as indicated in [43]. The level of significance was set at $p < 0.05$. The mgcv package provides a Bayesian approach to compute standard errors and confidence interval for the predictions. Model accuracy was evaluated with the Root Mean Square Error, as given by (RMSE, Eq. 2.5):

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}\left(\hat{y}_i - y_i\right)^2}{n}} \tag{2.5}$$

where $\hat{y}_i$ is the predicted value of observation i in the test set, y_i is the observed value and n the total number of observations in the test set. We also computed the Ratio of Performance to Deviation (RPD), which allows to normalize RMSE values (Eq. 2.6):

$$RDP = \frac{SD}{RMSE} \tag{2.6}$$

where SD is the standard deviation of the test set. After this calibration phase, a final model was built with all the samples (i.e. in both the training and test sets) using the covariates selected by the stepwise procedure, in order to improve model accuracy and representativity over te GDL territory. The fitted models for each land cover was then applied to the stack of spatial layers (covariates) to map topsoil OC content and associated model uncertainties (confidence intervals). We should note that RMSE computed in Eq. 2.5 does not give a correct measure of the true *map* accuracy, which should preferably assessed using a set of samples taken from randomly selected locations [45].

Chapter 3

Results

3.1 Exploratory analysis

3.1.1 Summary statistics and indicative OC ranges

To assist soil management in GDL, it is important to define likely or indicative ranges of observed OC in different soils, which could serve as reference levels on which a decline or increase in OC could be monitored[1] [13]. Here, we propose to classify OC observations based on soil associations (which reflects somewhat also climatic conditions) and land cover and use the median and inter-quartiles range (IQR) of the OC distributions (Table 3.1-3.4) to define indicative OC range in each soil association-land cover class. In each category, any value close to the median or within the boundaries defined by the IQR can be qualified as normal or standard values prevailing under current environmental conditions. Values outside the IQR indicate soils either depleted or enriched in OC in relation to the majority of observations in the same category.

Cropland soils in ASTA-SOC (n = 984) have a median OC content as low as 16 g C kg^{-1}but a large variation exists between the different soil associations (Table 3.1; Figure 3.1-3.2). The Oesling region reach indeed a median of 28.3 g C kg^{-1}while soil associations from the south of Luxembourg (Gutland) have median values between 10 g C kg^{-1} (*Grès du Luxembourg*) and 21 g C kg^{-1}(*Argiles lourdes des Schistes Bitumineux*). Large OC variations are also apparent when samples are grouped according to the 4 texture classes defined by ASTA (Figure B.1), giving the following sequence in terms of OC content: L (*leicht* = light texture) < M (*mittel* = medium texture) < S (*shwer* = heavy texture) < MO (*mittel Oesling* = medium texture of the Oesling region, stony soils). In cropland soils, the Inter-Quartiles Ranges (IQR) computed for each association are generally less that 10 g C kg^{-1} indicating a relatively small OC variation within each association. Some soil associations in the Gutland show a significant portion of observations with very low OC content (under 10 g C kg^{-1}), which make these soils probably sensitive to various degradation processes such as erosion. These OC observations in croplands are similar to published values in Belgium for similar environmental conditions as observed in the REQUASUD database [47]. In the Belgian Ardennes, corresponding to the Oesling region, the median is 31 g C kg^{-1} (q1 = 27 g C kg^{-1}, q3 = 35 g C kg^{-1}) and in the Belgian Jurassic region, corresponding roughly to the Gutland, the median is 16 g C kg^{-1} (q1 = 13 g C kg^{-1}, q3 = 20 g C kg^{-1}) .

The OC content in grassland (n = 323) is sensibly similar in the Oesling and Gutland, with a median around 34 g C kg^{-1}(Figure 3.1). There are however large variations between soil associations in Gutland (Table 3.2; Figure 3.2). Median OC values reach 40 to 50 g C kg^{-1} for clay-rich soils of the *Argiles du Lias Inf. et Moyen*, *Argiles lourdes du Keuper* and *Argiles lourdes des schistes bitumineux* while loamy and sandy soils of Gutland (*Buntsandstein*, *Dolomies du Muschelkalk*, *Calcaires du Bajocien*, *Grès de Luxembourg*, *Dépôts limoneux sur Grès*) have median OC generally less than 26 g C kg^{-1}. The IQR of grassland is on average approximately 18 g C kg^{-1} indicating larger within-class OC variations than in cropland soils.

The median OC content of forest soils is 33 g C kg^{-1}(Table 3.3), which is remarkably similar to the

[1] Defining critical threshold of OC content under which the fertility of soils is impaired due to degradation of the soil physical properties or disruption of nutrient cycle would be also of great practical interest but is difficult [46]. There are however evidences that such a treshold is ca. at 10 g C kg^{-1}under which, irrespective of the soil type, it may not be possible to reach potential crop yields. At this point, the decomposition of organic matter is in equilibrium with inputs of organic matter from plant residues and yields may be affected due to the low amount of mineralizable-N, even with the addition of N-P fertilisers [46].

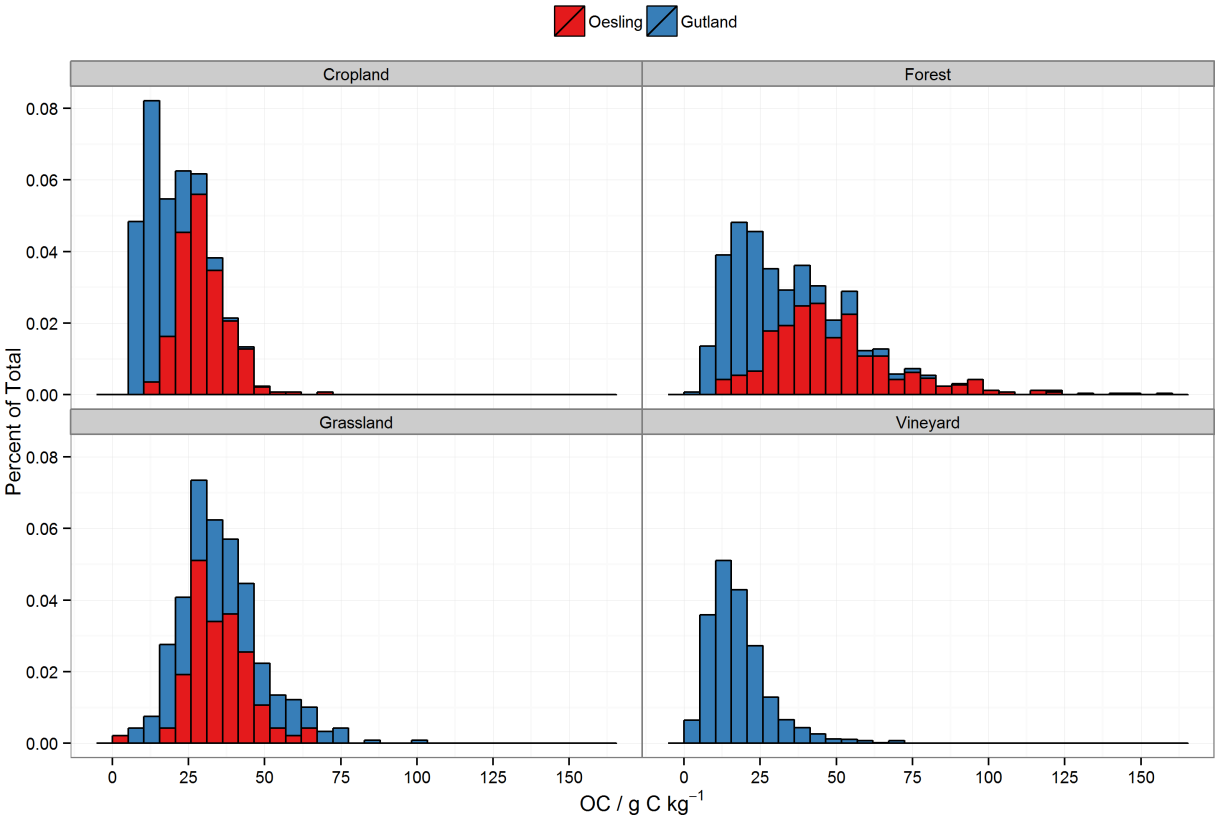

Figure 3.1: Histogram of OC (/g C kg^{-1}) in the ASTA-SOC database under cropland, grassland, forest and vineyard land cover for the Northern (Oesling) and Southern (Gutland) areas of GDL.

median OC content of grassland soils. This observation masks however great discrepancies between the two land cover classes. First, in the Oesling, the OC content is distinctly higher in forest soils than in grassland, with a median OC content of almost 45 g C kg^{-1}. Secondly, in Gutland, the quantity of OC stored in clayey soils is higher in grassland than in forests (Figure 3.2). Thirdly, forest seems to have a larger dispersion of OC values in their soils (Figure 3.2) with a IQR reaching almost 30 g C kg^{-1}(Table 3.3). This variation is mainly due to a large OC variation in the Oesling region, *Buntsandstein* and *Dolomies du Muschelkalk* associations (Table 3.3; Figure 3.2) because other soil associations show generally lower IQR.

Soil associations	**n**	**Min**	**q_1**	**$\bar{x}$**	**$\tilde{x}$**	**q_3**	**Max**	**IQR**
Oesling	266	12.8	24.3	29.8	28.3	34.6	68.7	10.3
Buntsandstein	59	7.1	13.5	16.6	15.1	18.0	43.9	4.4
Dolomies du Muschelkalk	27	11.3	16.4	25.8	27.8	32.0	48.4	15.6
Calcaires du Bajocien	1	15.7	15.7	15.7	15.7	15.7	15.7	0.0
Grès de Luxembourg	215	6.5	8.9	10.9	10.2	11.7	32.5	2.8
Dépôts limoneux sur Grès	146	6.3	10.3	12.6	11.9	14.3	26.7	4.0
Argiles du Lias inf. et moyen	129	8.0	15.6	19.2	19.0	22.8	33.2	7.3
Argiles lourdes du Keuper	85	7.4	12.5	15.0	14.3	17.3	25.5	4.8
Argiles lourdes des schistes bitumineux	7	11.3	16.7	22.6	21.2	25.0	42.6	8.3
Autres	42	7.9	11.8	17.1	15.3	21.2	40.4	9.4
	7	12.8	13.5	17.2	16.5	19.9	23.9	6.4
all	984	6.3	11.3	18.9	16.0	24.7	68.7	13.4

Table 3.1: Summary statistics of OC in Cropland per soil associations

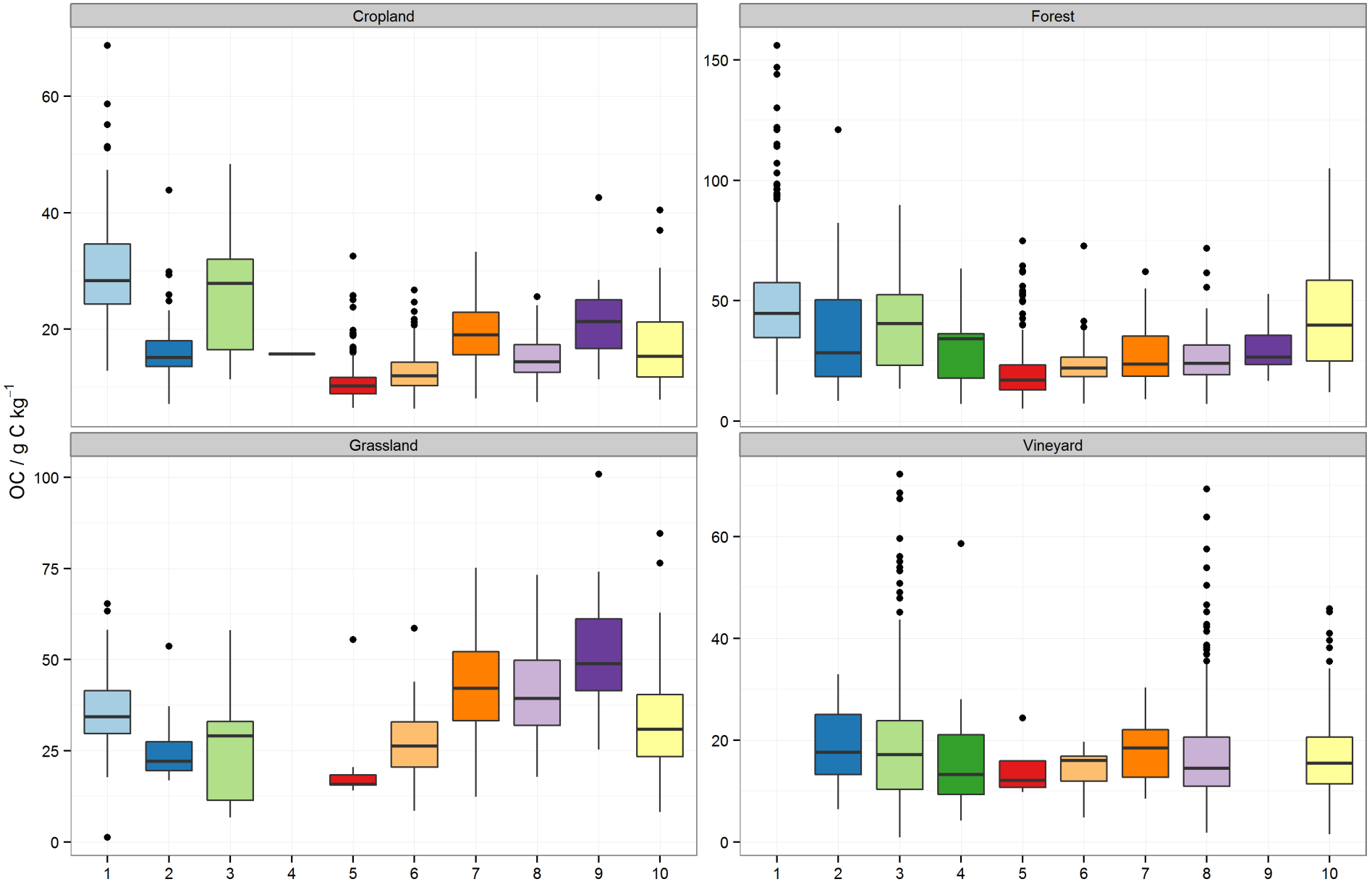

Figure 3.2: Boxplot[2] of OC (/g C kg^{-1}) in the ASTA-SOC database under cropland, grassland, forest and vineyard land cover as a function of the soil associations (1 = Oesling, 2 = Buntsandstein, 3 = Dolomies du Muschelkalk, 4 = Calcaires du Bajocien, 5 = Grès de Luxembourg, 6 = Dépôts limoneux sur Grès, 7 = Argiles du lias inf. et moyen, 8 = Argiles lourdes du Keuper, 9 = Argiles lourdes des schistes bitumineux, 10 = Others).

Soil associations	n	Min	q_1	$\bar{x}$	$\tilde{x}$	q_3	Max	IQR
Oesling	77	1.2	29.7	36.0	34.3	41.5	65.4	11.9
Buntsandstein	11	16.9	19.5	26.0	22.1	27.4	53.7	7.9
Dolomies du Muschelkalk	11	6.8	11.5	25.8	29.1	33.1	58.1	21.6
Grès de Luxembourg	11	14.1	15.6	20.1	15.8	18.4	55.5	2.8
Dépôts limoneux sur Grès	30	8.5	20.5	27.4	26.3	33.0	58.6	12.4
Argiles du Lias inf. et moyen	49	12.4	33.3	42.4	42.2	52.2	75.2	18.9
Argiles lourdes du Keuper	43	17.8	32.0	41.7	39.4	49.8	73.2	17.9
Argiles lourdes des schistes bitumineux	18	25.3	41.5	52.1	48.9	61.1	100.8	19.6
Autres	71	8.2	23.4	33.6	30.9	40.5	84.6	17.1
	2	27.6	29.6	31.7	31.7	33.7	35.7	4.0
all	323	1.2	26.0	36.1	33.7	43.9	100.8	17.8

Table 3.2: Summary statistics of OC in Grassland per soil associations

Soil associations	n	Min	q_1	$\bar{x}$	$\tilde{x}$	q_3	Max	IQR
Oesling	455	11.0	34.7	49.1	44.6	57.5	156.0	22.9
Buntsandstein	50	8.3	18.5	36.8	28.3	50.3	121.0	31.8
Dolomies du Muschelkalk	23	13.4	23.1	41.5	40.5	52.4	89.8	29.3
Calcaires du Bajocien	5	7.0	17.8	31.7	34.1	36.3	63.3	18.5
Grès de Luxembourg	217	5.1	12.9	20.2	17.0	23.3	74.8	10.4
Dépôts limoneux sur Grès	63	7.2	18.4	23.2	22.0	26.4	72.7	8.0
Argiles du Lias inf. et moyen	65	9.0	18.5	27.4	23.6	35.2	62.0	16.7
Argiles lourdes du Keuper	63	7.0	19.1	26.4	24.0	31.6	71.8	12.4
Argiles lourdes des schistes bitumineux	8	16.6	23.4	30.3	26.6	35.6	52.7	12.2
Autres	60	11.9	24.8	42.2	39.9	58.4	105.0	33.6
	2	19.5	27.8	36.0	36.0	44.3	52.6	16.5
all	1011	5.1	20.0	37.0	32.9	49.2	156.0	29.2

Table 3.3: Summary statistics of OC in Forest per soil associations

Soil associations	n	Min	q_1	$\bar{x}$	$\tilde{x}$	q_3	Max	IQR
Buntsandstein	20	6.4	13.2	18.6	17.6	25.0	32.9	11.8
Dolomies du Muschelkalk	349	0.9	10.3	19.1	17.1	23.8	72.2	13.5
Calcaires du Bajocien	18	4.2	9.3	16.9	13.2	21.1	58.6	11.7
Grès de Luxembourg	4	9.8	10.7	14.6	12.1	15.9	24.3	5.2
Dépôts limoneux sur Grès	9	4.8	12.0	13.9	16.0	16.8	19.7	4.8
Argiles du Lias inf. et moyen	22	8.5	12.7	18.1	18.4	22.0	30.3	9.3
Argiles lourdes du Keuper	403	1.8	10.9	16.7	14.5	20.6	69.3	9.7
Autres	238	1.5	11.4	16.7	15.4	20.6	45.8	9.2
	82	3.4	13.0	20.9	18.5	25.3	61.5	12.3
all	1145	0.9	11.1	17.8	16.0	21.8	72.2	10.8

Table 3.4: Summary statistics of OC in Vineyard per soil associations

Soils under vineyard have about the same median OC content than in cropland soils of GDL (Table 3.4). In vineyard soils, soil associations show a median OC content oscillating between 12 (*Grès du Luxembourg*) and 18.4 g C kg^{-1}(*Argiles du Lias Inf et Moyen*; Table 3.4) with an IQR lower than 13.5 g C kg^{-1}, indicating low inter- and intra-class OC variation. However, it can be observed that samples from the *Dolomies du Muschelkalk* and *Argiles lourdes du Keuper* have a relatively large number of outlying values with high OC content (Figure 3.2) and maximum OC content reaching around 72 g C kg^{-1}.

3.1.2 Relations with the soil covariates

Since the process to create OC maps is essentially empirical, it is important, as a first step, to understand the relationships between OC content and the covariates. The level of OC content in soils is controlled by the balance between the input of Organic Matter (OM) into the soil and the decomposition (or mineralization) of OM by micro-organisms, and these two processes are to some extent captured by the covariates. Soils under cropland and vineyard have on average ca. 50 % less OC than those under grassland and forest (Table 3.1-3.4), showing the strong effect of land cover on OC levels. Cropland

[2]The two hinges of the boxes are the first and third quartile, the middle bar indicates the median of the distribution. The whiskers (i.e. lines extending away from the boxes) extend to the most extreme data point which is no more than 1.5 times the length of the boxes.

and vineyard soils show generally depleted OC levels due to ploughing which breaks up soil aggregate structures and increase thereby the exposition of physically-protected OC to micro-organisms [48, 49].

In croplands, high OC values can be observed in the Oesling but also, to some extent, in clay-rich soils of Gutland, especially in the south-east corner of GDL near the French-German border (Figure 3.3). This pattern also appears in Figure 3.4 showing an U-shaped relationship between OC and the longitude. This difference in OC content between soils of the Oesling and Gutland can be linked to the climatic gradient between the south and the north of the GDL territory (Figure 2.8), which itself is highly related to altitude, as indicated by a relatively strong relationship between OC and temperature ($\rho = -0.51$), precipitation ($\rho = 0.54$) and altitude ($\rho = 0.58$; Figure C.1). The Oesling region experiences indeed a colder climate and longer frost periods than Gutland. Higher OC content in areas with higher precipitation and lower temperature is often observed due the effects of precipitation on Net Primary Productivity, lower levels of oxygen concentration in wetter soils (anaerobic conditions), and decreased microbial activity or decomposition of organic matter in colder climate [50, 51, 52]. In general, OC levels increase with increasing precipitation and for any given level of precipitation, decrease with increasing temperature [51]. A clear, but somewhat less strong relationship can be observed between OC and C factor ($\rho = -0.35$). This negative correlation is explained by the fact that the C factor increase with a decrease in crop cover, which in turn can be related to the rate of OM incorporation into the soils trough plant residues and OM decomposition through variations in runoff and erosion rates. Livestock intensity show a positive correlation with OC in croplands reflecting the positive influence of manure application on OC levels but the relationship seems only limited ($\rho = 0.17$). The aggregation of the data at the municipality level might explain this low correlation. While amendments such as manure, lime and fertilizer are well-known to induce changes in OC content in soils (see [53]), these variables are very difficult to include in the environmental regression approach due to the scarcity of detailed spatio-temporal crop management data at large scales.

Morphometric variables, including slope and aspects, have very small correlations with OC content ($|\rho| < 0.15$; Figure C.1), suggesting that position in the landscape and its control on the water supply and erosion rates plays a little role in the patterns of OC content in croplands of GDL. It is however also possible that the scale at which samples have been collected (i.e. field composites) and/or the spatial resolution of the morphometric covariates are too coarse to capture variations linked to local OC transfer processes. Soil texture seems to be a major driver of OC distribution in croplands, although the positive correlation is only visible for soils of Gutland ($\rho = 0.56$) since the texture of Oesling soils varies only little (Figure 3.5). This is in agreement with the results of previous studies in various environments[54, 55, 18, 56] and empirically confirms the positive impact of the clay fraction on OC through chemical stabilization [48] and higher soil moisture content (due to poor drainage status) leading to lower OC mineralization rates [57, 58]. Fine-textured soils tend also to contain more OC due to their greater nutrient and water-holding capacity, favoring plant production and thereby the amount of fresh OM returning into the soil.

The spatial distribution of OC seems more difficult to interpret in grasslands than in croplands, partly because of the low number of observations (Figure 3.3). Climatic, management and morphometric variables have only a weak correlation with OC in grassland soils (Figure 3.6, C.2). Clay content is the only covariate showing a strong and positive influence on OC levels in Gutand ($\rho = 0.5$; Figure 3.5). As noted earlier (section 3.1.1), clay deposits from Gutland tend indeed to store more OC than other soil associations (Figure 3.2). In forest, a clear distinction between the Oesling and Gutland can be observed (Figure 3.3). Most of observations with high OC ($>$ 100 g C kg^{-1}) occur in the Oesling region. The correlations with climatic variables are higher than in grasslands (Figure 3.7) and the correlation with clay content in Gutland is less pronounced ($\rho = 0.24$; Figure 3.5). Similarly to croplands and grasslands, the relations between OC and morphometric variables are weak in forest soils (Figure C.3). Concentrated along the Moselle river valley, OC observations for vineyard soils do not show evident spatial patterns (Figure 3.3), nor any significant correlations with the covariates (Figure 3.8, C.4). This contrasts with observations in vineyards of France [56], where higher correlations were found with texture ($\rho = 0.32$), temperature ($\rho = -0.31$) and slurry and famyard manure-related OC ($\rho = 0.41$). Vineyards in GDL are concentrated in only one river valley characterized by rather homogeneous environmental conditions which can explain these poor correlations with the covariates. Another explanation could be the high percentage of land reallocation done in the past.

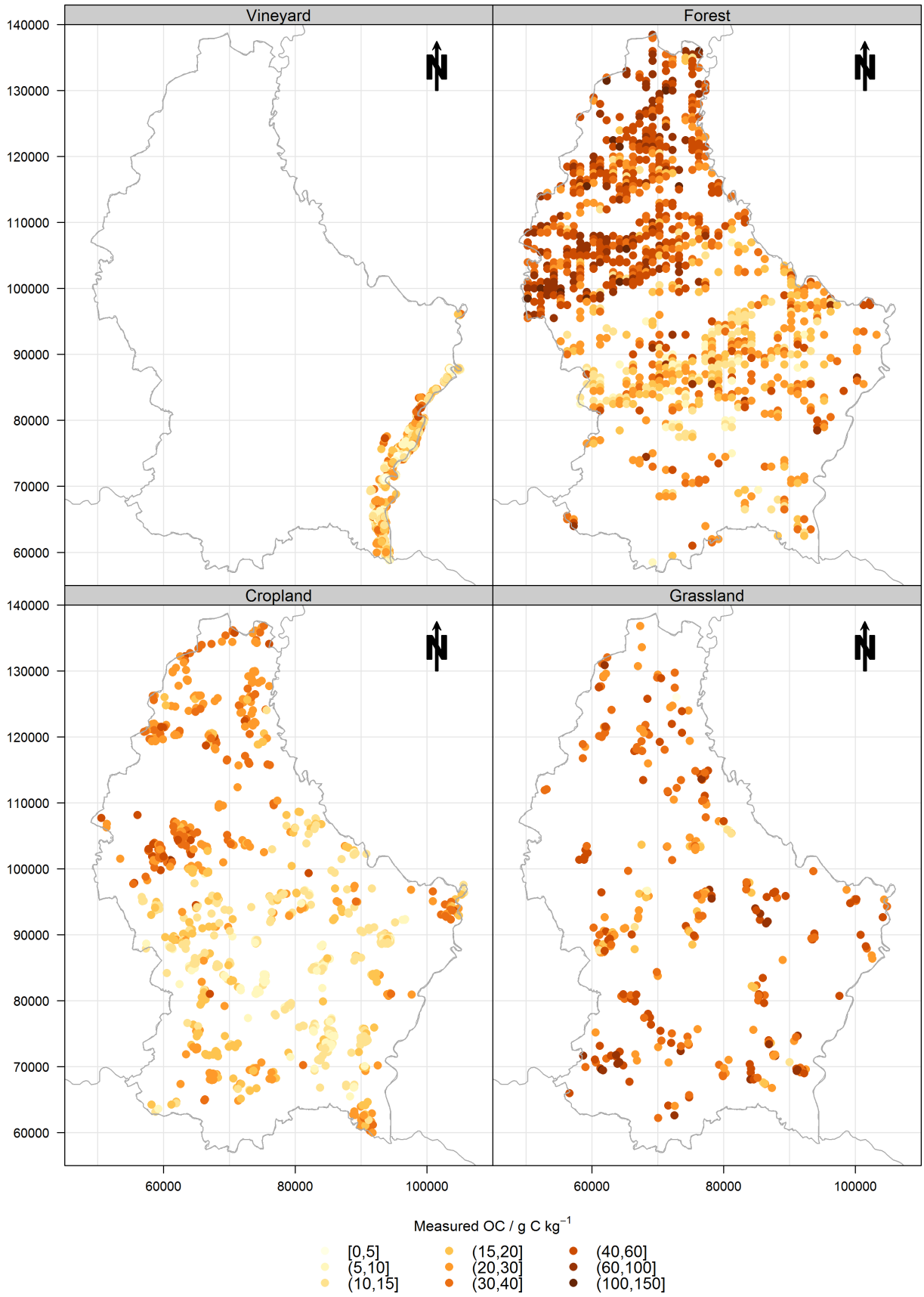

Figure 3.3: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under cropland, grassland, forest and vineyard land cover. For clarity, these four maps are available in larger size in Annex B.2 to C.4

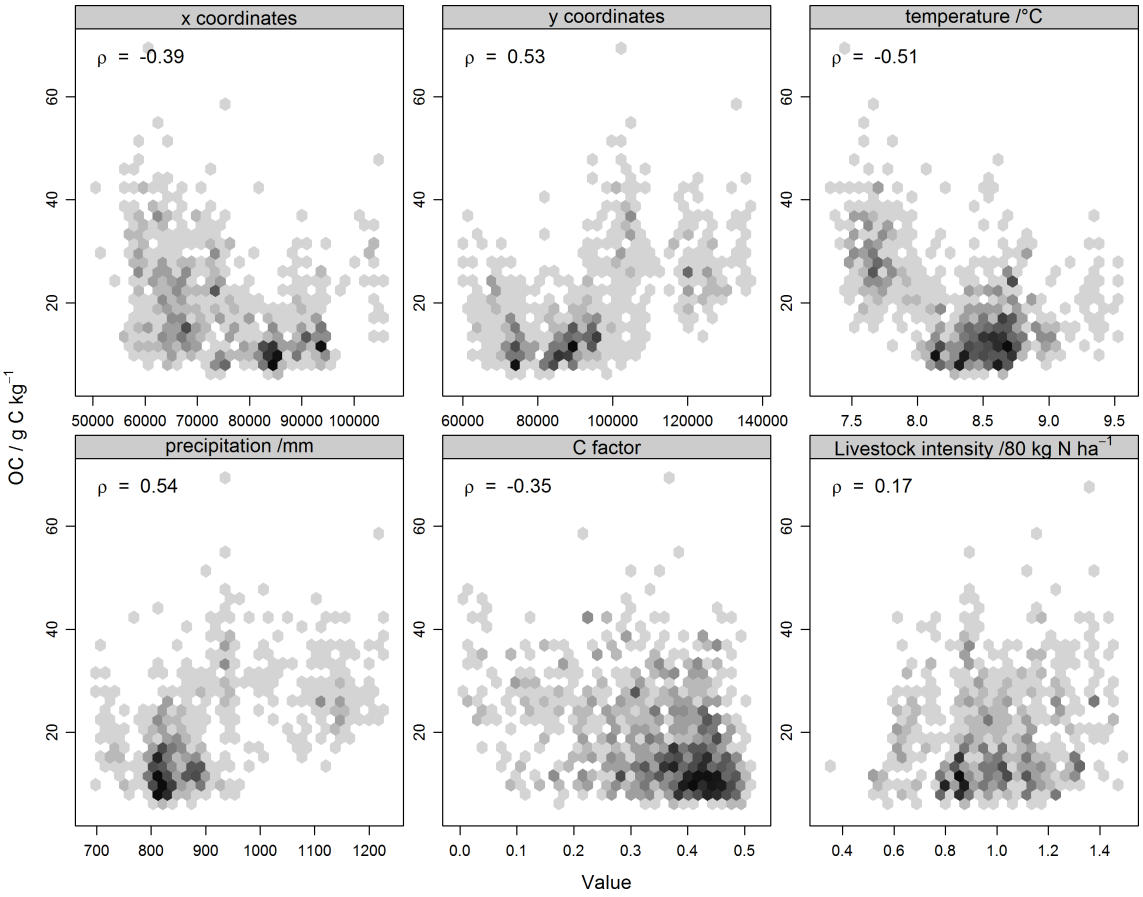

Figure 3.4: Scatter plot of OC (/ g C kg^{-1}) as function of geographical coordinates (x,y, in LUREF), mean annual temperature (/°C), mean annual precipitation (/mm), C factor and livestock intensity (/ 80 kg N ha^{-1}) for cropland soils. The gray color scale represents the density of points in each hexagonal bin of the plot.

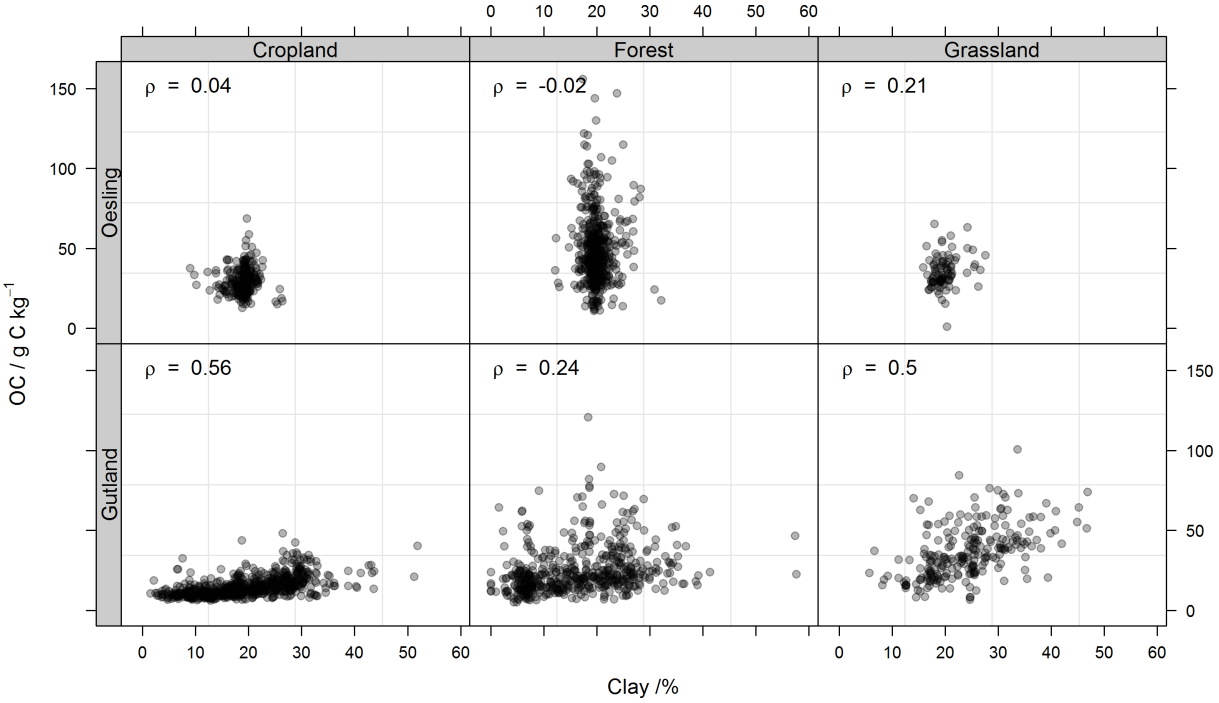

Figure 3.5: Scatter plot of OC (/ g C kg^{-1}) as function of clay content (/ %), land cover type and region (Oesling, Gutland).

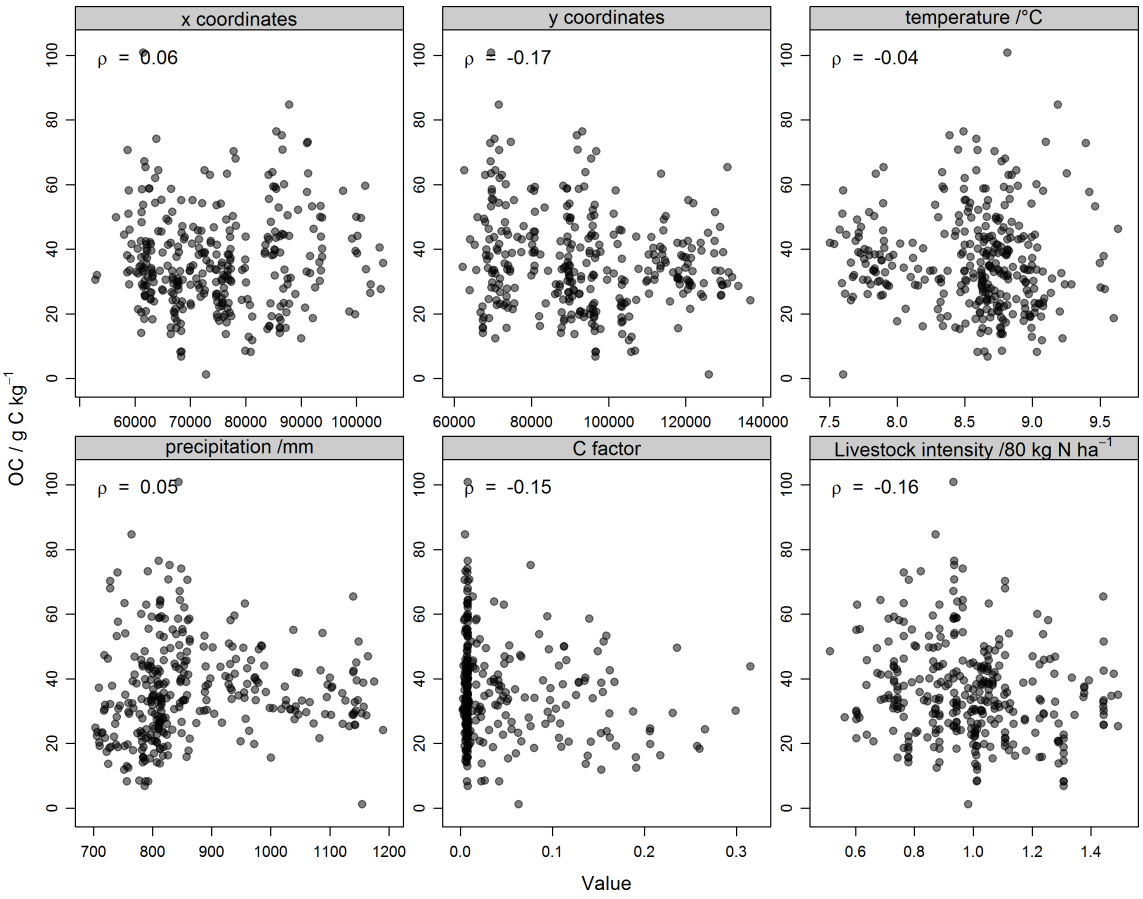

Figure 3.6: Scatter plot of OC (/ g C kg^{-1}) as function of geographical coordinates (x,y, in LUREF), mean annual temperature (/°C), mean annual precipitation (/mm), C factor and livestock intensity (/ 80 kg N ha^{-1}) for grassland soils.

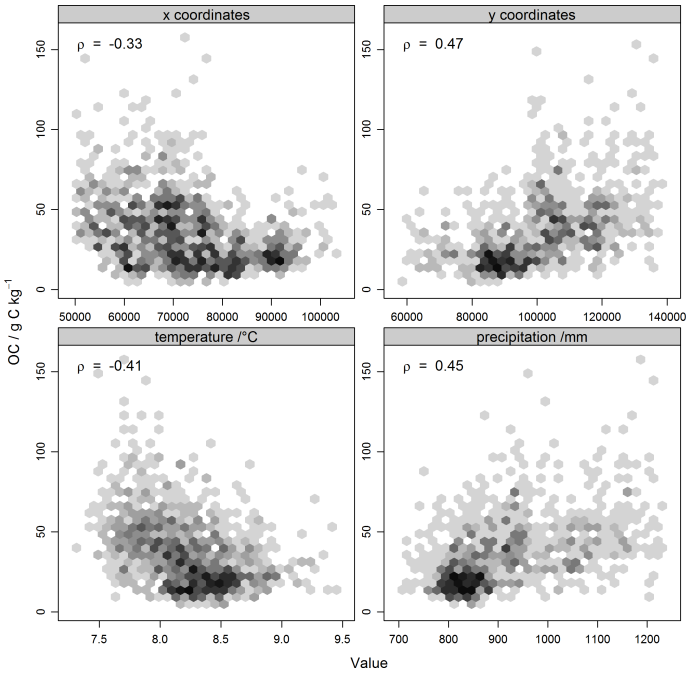

Figure 3.7: Scatter plot of OC (/ g C kg^{-1}) as function of geographical coordinates (x,y, in LUREF), mean annual temperature (/°C) and mean annual precipitation (/mm) for forest soils. The grey color scale represents the density of points in each hexagonal bin.

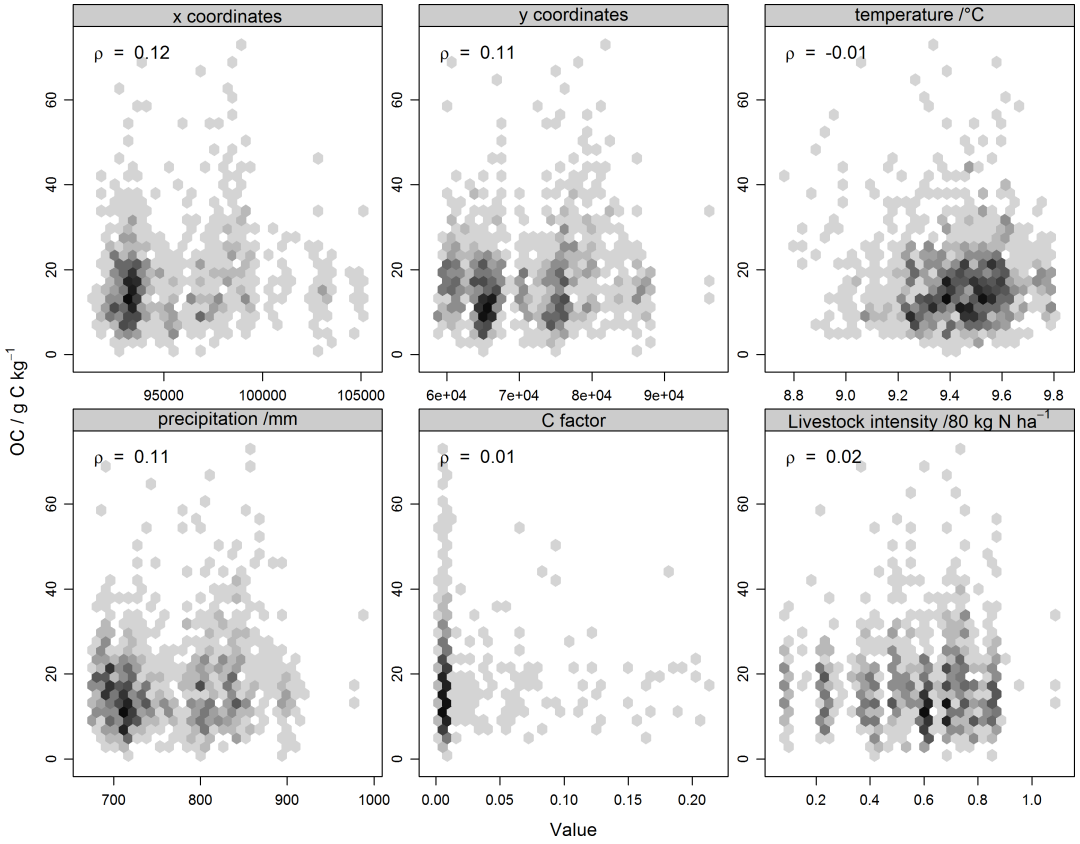

Figure 3.8: Scatter plot of OC (/ g C kg^{-1}) as function of geographical coordinates (x,y, in LUREF), mean annual temperature (/°C), mean annual precipitation (/mm), C factor and livestock intensity (/ 80 kg N ha^{-1}) for vineyard soils.

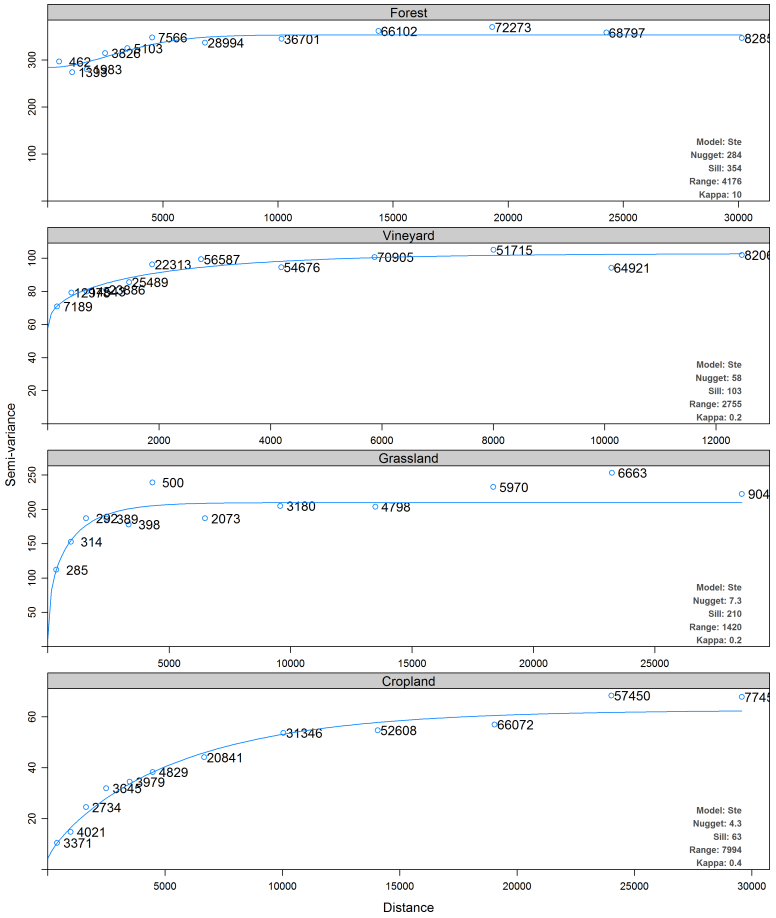

Figure 3.9: Semi-semivariograms of OC under cropland, grassland, forest and vineyard land cover.

The experimental semivariograms of the OC observations show some very different spatial structures between the different land cover (Figure 3.9). Cropland soils have a small nugget-to-sill ratio, indicating a high degree of spatial dependence. The range of the semivariogram is large which can express the fact that OC in croplands is mostly determined by long-range factors such as temperature and precipitation. The semivariogram is possibly unbounded, which can be related to the presence of a trend in the data (probably due to the differences between the Oesling and Gutland regions). In grasslands, OC content show also a low nugget-to-sill ratio but with a much smaller range (ca. 2 km). The spatial dependence of OC in grasslands occurs at a much lower distance than croplands and can be due to the above-mentioned role of the clay content (which can vary on short distances) on OC spatial variation. Vineyard and forest soils show very little spatial structures in their semivariograms. This suggests that OC content vary as it was almost randomly distributed in space and hence the spatial variation of OC in these environments will be very difficult to model.

3.2 Model results, performance & OC maps

The exploratory analysis showed that there are conspicuous relationships between OC content of soils and the natural and anthropic variables. The main drivers of OC spatial distribution in GDL are land cover, climatic conditions and soil types. The same combination of driving factors has been identified in similar environments, for instance in England and Wales [13] and in France [20]. These relations set the basis for the extrapolation of the OC observations of the ASTA-SOC database in space that we present thereafter. GAM model results for each land cover are presented in Tables 3.5-3.8.

3.2.1 Croplands

The backward stepwise procedure selected elevation, precipitation, temperature, clay and projected co-ordinates (x,y) to model the spatial variation of OC in cropland soils, achieving an explained deviance of 0.78 (Table 3.5). The remaining deviance should be related to other factors not included in the analysis, such as fertilizer application or rotation patterns. The predicted-observed graph (Figure 3.10) shows that the GAM model fit well to not only the training set but also to the test set, with a $R^2$of 0.66 and RMSE of 5.5 g C kg^{-1}. There are indications, however, that observations with high OC content (> 40 g C kg^{-1}) are under-predicted (Figure 3.10), due to the low number of samples with such high OC content in croplands. The GAM model was applied to the set of covariates, resulting in a continuous map of OC content (Figure 3.11) and associated standard error of the predictions (Figure 3.12). The spatial distribution of OC in croplands appears to be mainly driven by elevation. The map depicts a north-South gradient in OC in GDL, with high OC content generally found in high plateaus of the Oesling. In Gutland, OC variations are related to the soil associations (Figure 2.2) and the spatial distribution of clay content (Figure 2.9), with higher OC content found in the south-west (Minette area). The map shows however unexpected high OC concentrations in the Harlange area (x: 57 500; y: 111 000) and along the Moselle river, which might be unrealistic and probably due to the absence of sampling points in this area (Figure B.2). Generally, it is probable that pixels that are predicted above 40 g C kg^{-1} are in the extrapolation domain and likely to be inaccurate because almost no observations in the ASTA-SOC database were predicted above this value (Figure 3.10). The map of the prediction standard error confirms indeed that the predictions in these areas should be viewed with caution. We can expect that such artifacts related to the mapping process will be fixed when new samples from these areas will be incorporated in the database.

3.2.2 Grasslands

The GAM model for grasslands, explaining 35 % of the deviance in the calibration set, included three terms: clay, hydrological classes (as defined by the ERRUISSOL project, only significant in the Gutland) and the coordinates (only significant in the Gutland, Table 3.6). This confirms the important role of texture and drainage on OC concentrations in grasslands. This is true only in Gutland, because in the Oesling these factors are more constant.The GAM model shows poor results when applied to test set, with an R^2 of 0.36 and a RMSE of 11.4 g C kg^{-1}(Figure 3.10). The predicted-observed plot show, as for croplands, that the model tends to under-estimate high OC values (> 60 g C kg^{-1}). Compared to croplands, the OC map show more contrasted patterns with large variations within a few kilometers (Figure 3.13). In the Oesling, there is overall more OC at high altitude and in the west, although the map of standard errors suggest that predictions in the west of the Oesling should be taken with care. In Gutland, sandy soils in the center and west have less OC (± 10-20 g C kg^{-1}) than elsewhere.

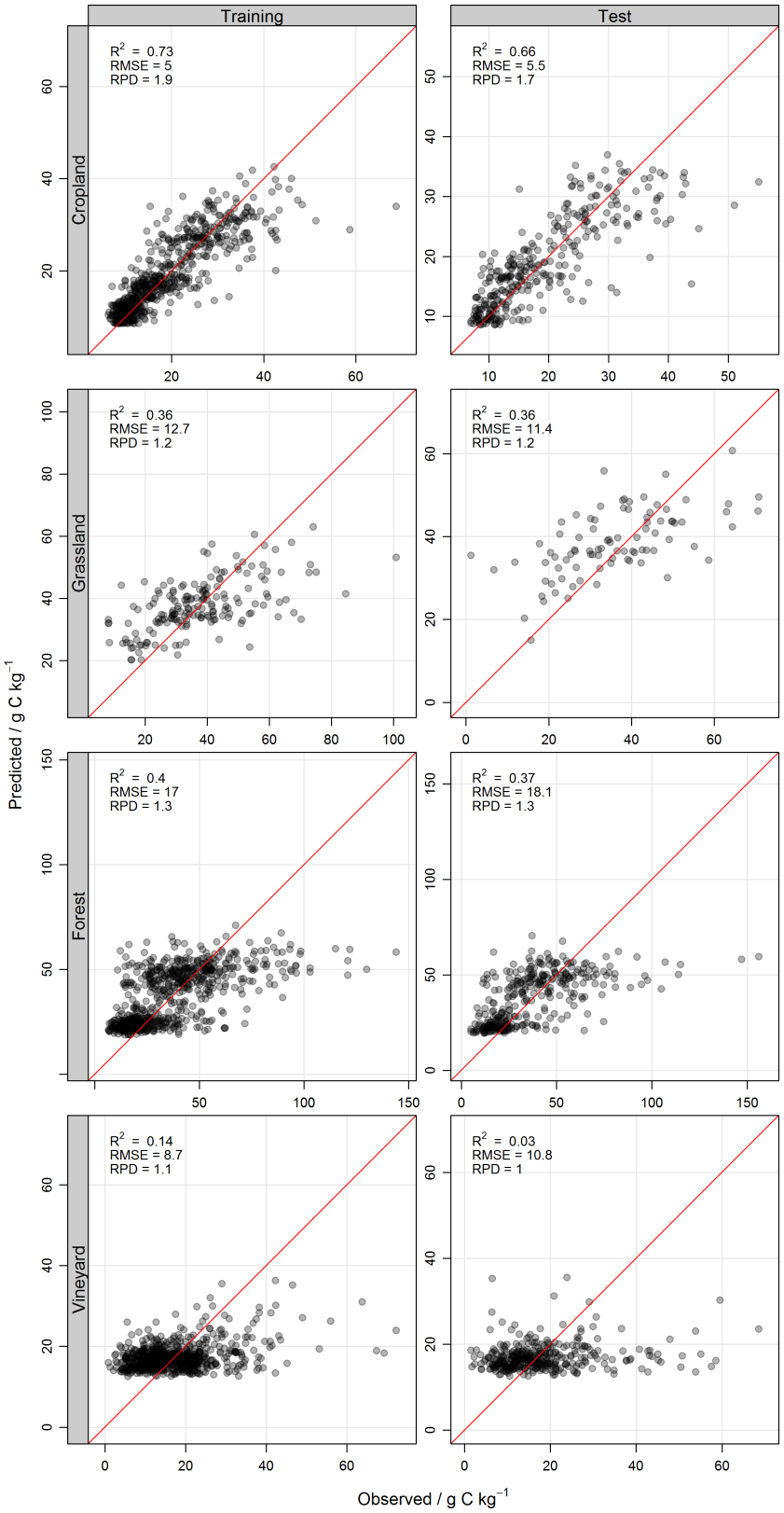

Figure 3.10: Predicted *vs* observed OC (/ g C kg^{-1}) in the training and test sets as obtained by the GAM model for cropland, grassland, forest and vineyard soils.

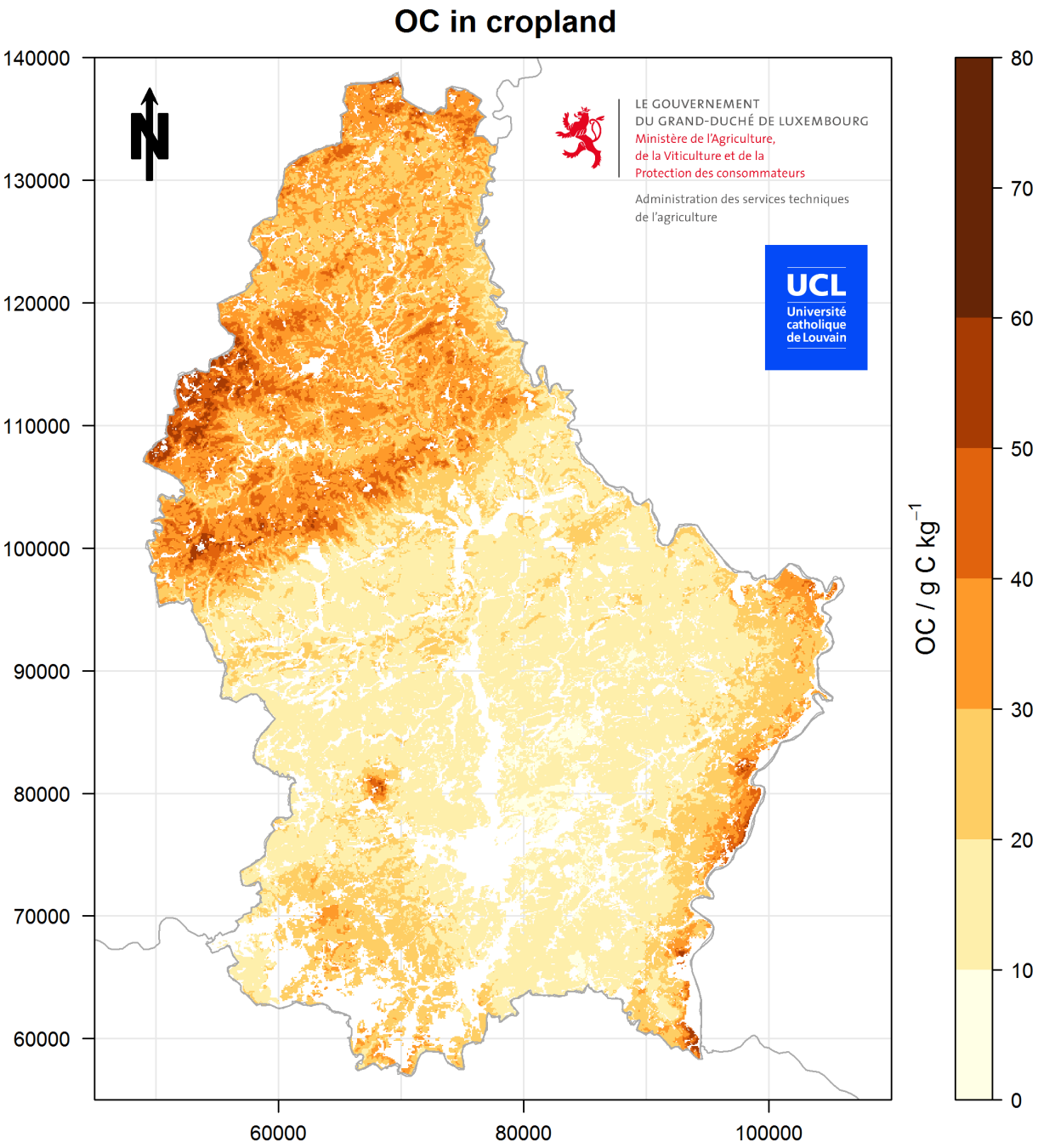

Figure 3.11: Map of predicted OC content (/ g C kg^{-1}) in cropland soils, assuming the whole territory is covered by cropland. Areas in white corresponds to environments where no predictions could be made.

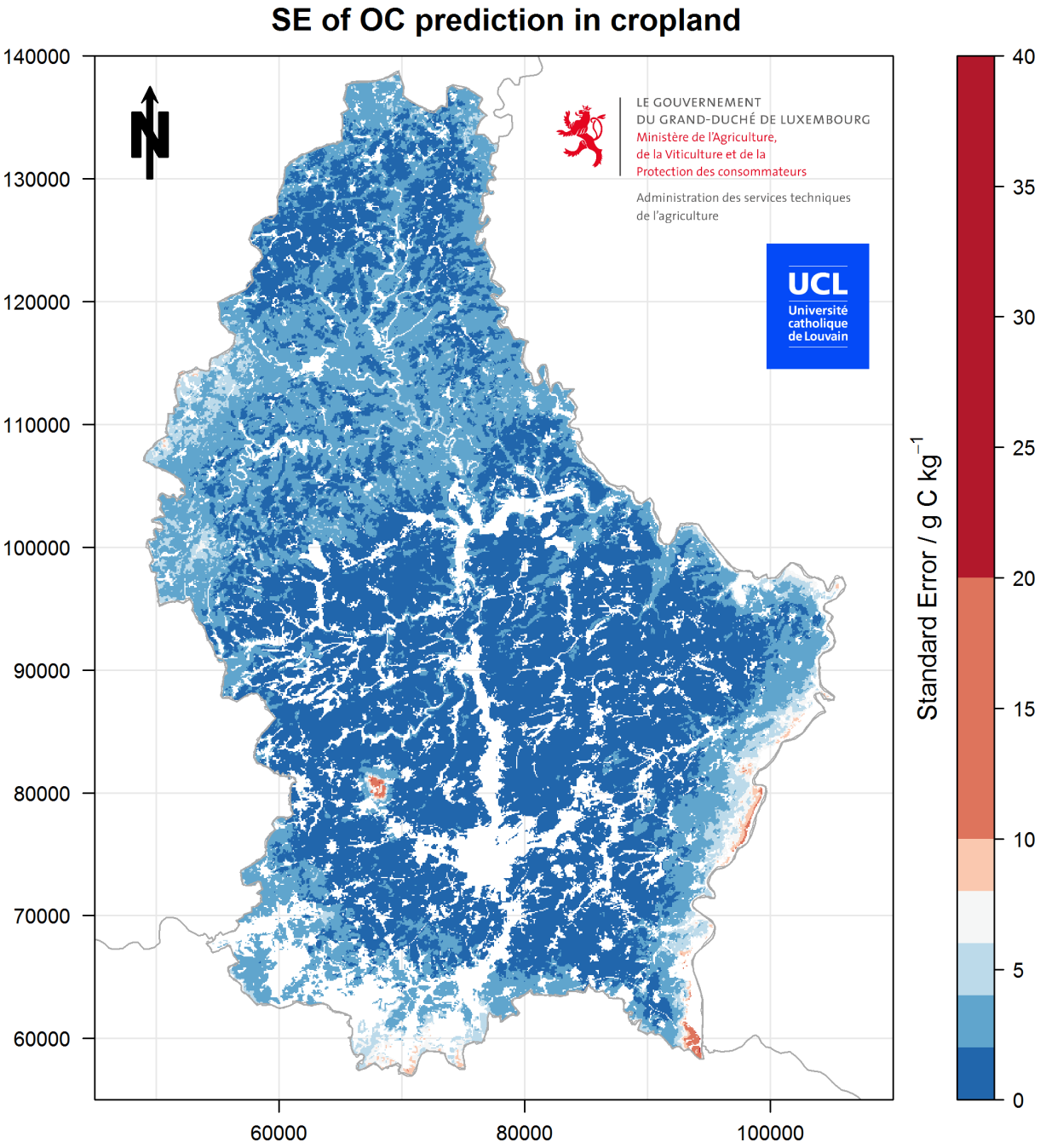

Figure 3.12: Map of the standard error of predicted OC content (/ g C kg^{-1}) in cropland soils.

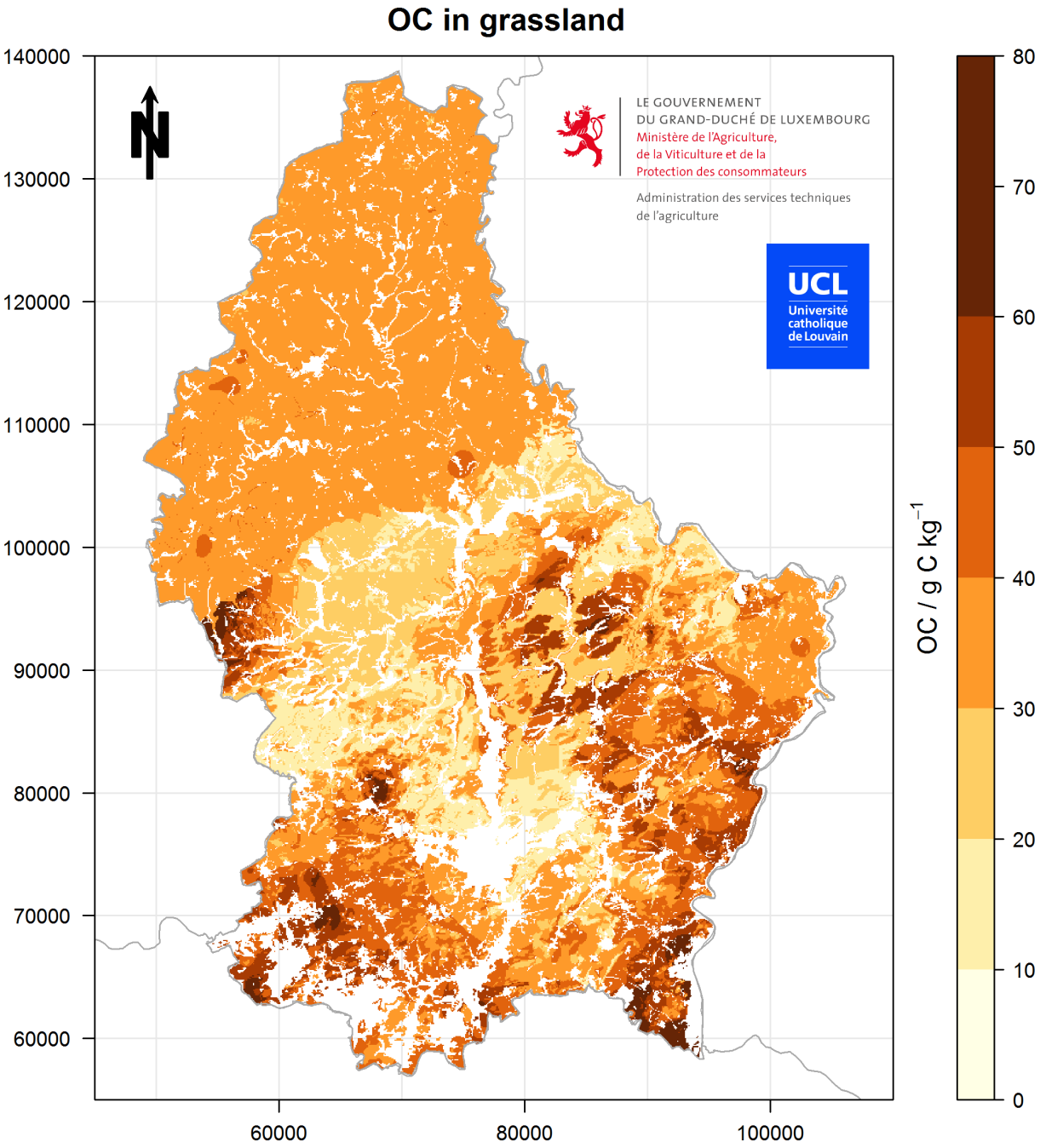

Figure 3.13: Map of predicted OC content (/ g C kg^{-1}) in grassland soils, assuming the whole territory is covered by grassland. Areas in white corresponds to environments where no predictions could be made.

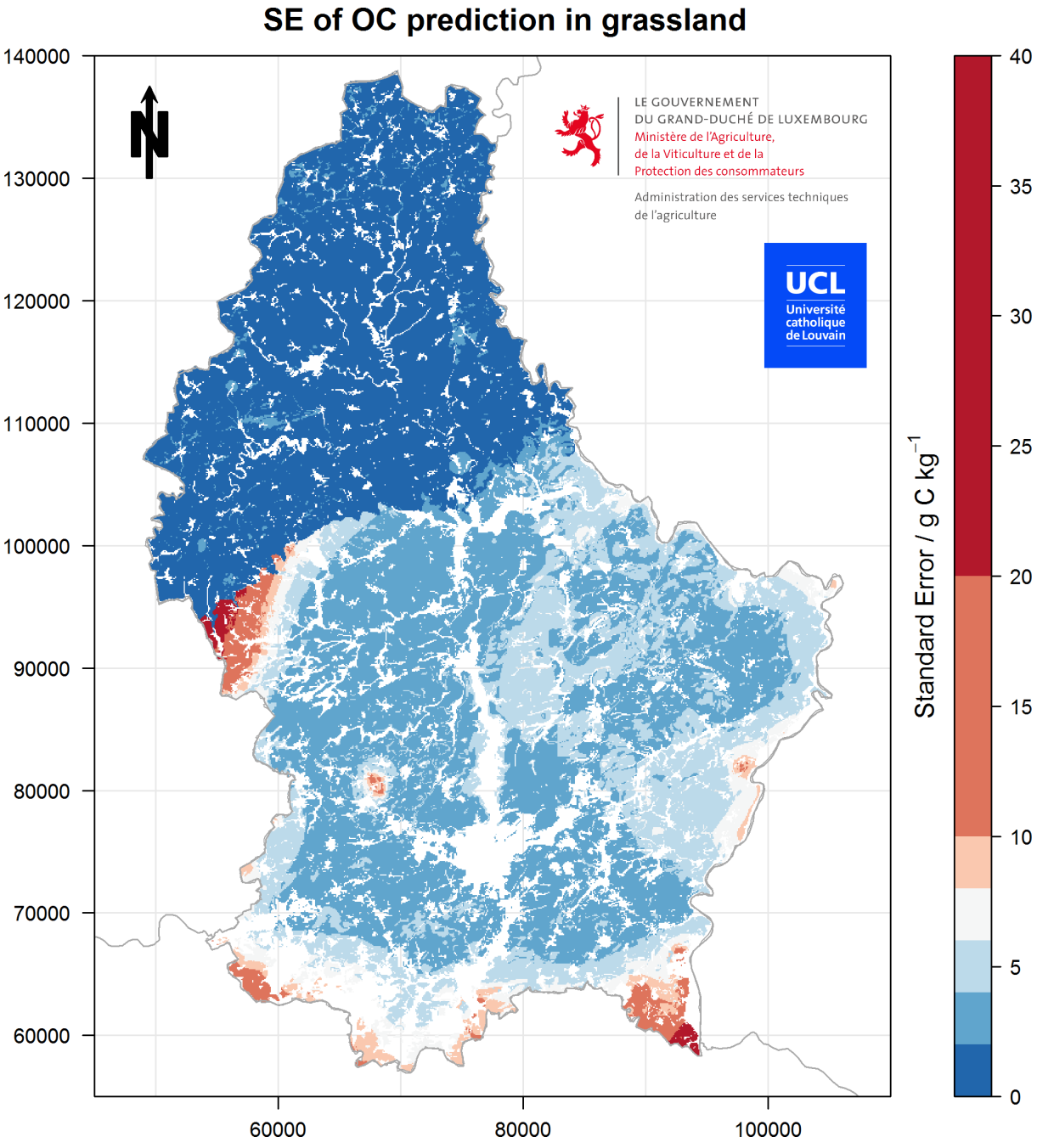

Figure 3.14: Map of the standard error of predicted OC content (/ g C kg^{-1}) in grassland soils.

Table 3.5: Summary of the full and reduced GAM models for croplands

	Full model	Reduced model
(Intercept)	2.85***	2.86***
elevation	5.99***	5.91***
eastness	1.88***	
northness	1.21	
slope	0.76*	
precipitation	2.55***	3.22***
temperature	6.38***	6.25***
C factor	3.08***	3.06***
clay	2.97***	2.95***
Livestock intensity	1.93*	
LS factor	1.96**	
path length	3.43**	
x,y	12.17***	12.25***
AIC	4003.92	4035.78
BIC	4211.47	4195.50
Log Likelihood	-1955.65	-1982.25
Deviance	15266.92	16562.65
Deviance explained	0.74	0.72
Dispersion	25.12	26.78
R^2	0.72	0.71
GCV score	2049.63	2057.01
Num. obs.	653	653
Num. smooth terms	12	6

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.6: Summary of the full and reduced GAM models for grasslands

	Full model	Reduced model
(Intercept)	3.53***	3.52***
elevation	0.55*	
eastness	0.81*	
northness	0.97	
slope	0.00	
precipitation	0.00	
temperature	0.56	
C factor	0.88**	
clay	0.96***	0.94***
Livestock intensity	0.00	
x,y:Oesling	0.00	0.49
x,y:Gutland	4.61**	5.32**
hydrological class†:Oesling	0.49	0.82*
hydrological class†:Gutland	0.89**	0.87**
AIC	1418.50	1430.68
BIC	1459.00	1463.93
Log Likelihood	-696.52	-704.89
Deviance	26107.42	28681.34
Deviance explained	0.41	0.35
Dispersion	157.02	170.17
R^2	0.37	0.32
GCV score	714.13	718.98
Num. obs.	178	178
Num. smooth terms	13	5

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The influence of texture and drainage on OC content is clearly visible, although uncertainties are very high for pixels that are predicted to contain high OC content (Figure 3.13-3.14) because of high clay content (Figure 2.9). There are indeed no observations with high clay content (> 60 %; Figure 3.5) in grasslands, which make these predictions pure extrapolations. In the north-west of the Gutland, an area is predicted to have high OC content but is an artifact due to the low number of sampling points (Figure 2.3 and B.3). Due to the relative short-range spatial variation of OC under grasslands (Figure 3.9), it is critical to have a large number of observations located such as to cover appropriately the spatial domain and the variation in the covariates in order to produce reliable predictions over the entire territory.

3.2.3 Forest and vineyards

For forest soils, the GAM model included precipitation, clay and coordinates and explained 39 % of the deviance (Table 3.7). Better results were obtained by using two separate smooth terms of the coordinates for the Oesling and the Gutland. Despite a R^2 of 0.4 in the training set and 0.37 in the test set, the GAM fit should be considered to be poor. The R^2 is indeed inflated by the presence of two clusters of prediction points, at relatively low and high OC content (Figure 3.10), which corresponds to the two different regions. Looking at the two regions separately, the correlation between predicted and observed OC is close to zero. This means that the model is able to catch the general shape of OC variation in GDL (low OC in the south *vs* high OC in the north) but the spatial distribution of OC within each region is very poorly taken into account. So, despite the high number of sampling points (Figure 2.3, B.4), modeling OC variation in forest soils seems difficult, suggesting that important factors controlling the OC content of forest soils were not taken into account. Analysis of the IFL database indicated that tree species and age did not influence OC content (results not shown). There are however other factors that were not included in the modeling, such as forest management practices (e.g. fertilization, harvesting) or land use change history which are known to have an effect on OC content in forest [59, 60]. Furthermore, the space between observation on the grid used for collecting IFL data (500x1000 m) were collected is too coarse to capture variations in OC that can occur at very short range [61]. It is also possible that variations between sampling units may derive from slight differences in sampling depth and imperfect removal of the ectorganic horizons. Another reason could be the long period of time between sampling and analysis (approximately 11-12 years) and the question of OC stability in samples. The map of OC content displays relatively smooth spatial variations over GDL (Figure 3.15), bearing in mind that predictions holds limited quantitative information, and that the map represents only the average OC content prevailing under given precipitation-clay content conditions. Standard errors of predictions are generally low due to the high number of samples (Figure 3.16).

With a percentage of explained deviance close to 0.14, the GAM model is is not good for soils under vineyard (Table 3.8). The RPD in the test set is equal to 1 (Figure 3.10), meaning that predictions are as good as if we used the mean OC content in vineyards to predict all the observations. This poor fit was expected since the exploratory analysis demonstrated that the covariates were very poorly correlated with OC observations. As a matter of fact, this emphasizes that vineyards are intensively managed so that soil and environmental conditions does not play a great role in the dynamic and patterns of OC. Large variation on very small distance can also be related to land reallocation operations that are very common in vineyards. Finally, this can also be due to bad assignation of the OC observations to specific fields because of wrong FLIK values as the numbering system changed recently and may have caused confusion. Due to the very low accuracy achieved by the model, we decided not to produce a map of OC for this land cover class.

3.3 Map of OC content in GDL

The maps of OC for each land cover were combined into one final map displaying the OC distribution of topsoils in GDL. For vineyards, we assigned the mean value of all observations under vineyard to each field, due to poor prediction accuracy of the GAM model. The resulting map is shown in Figure 3.17 and the histogram of the predictions in each land cover in the Oesling and Gutland in Figure 3.18. Generally, the map shows a north to south increase in OC due to the climatic gradient in GDL and OC patterns linked to the spatial distribution of the different land cover types and clay content. In the Oesling, models predicted more OC in forest than in grassland and more in grassland than in croplands (Figure 3.18). In Gutland, however, the map shows more OC in grassland than in forest soils and also in grasslands of the Oesling (Figure 3.18). Compared to the OC distribution of observed values (Figure 3.1), it is evident

that predicted OC ranges are lower, indicating that models failed to predict high OC values, a problem that is especially acute in forest soils.

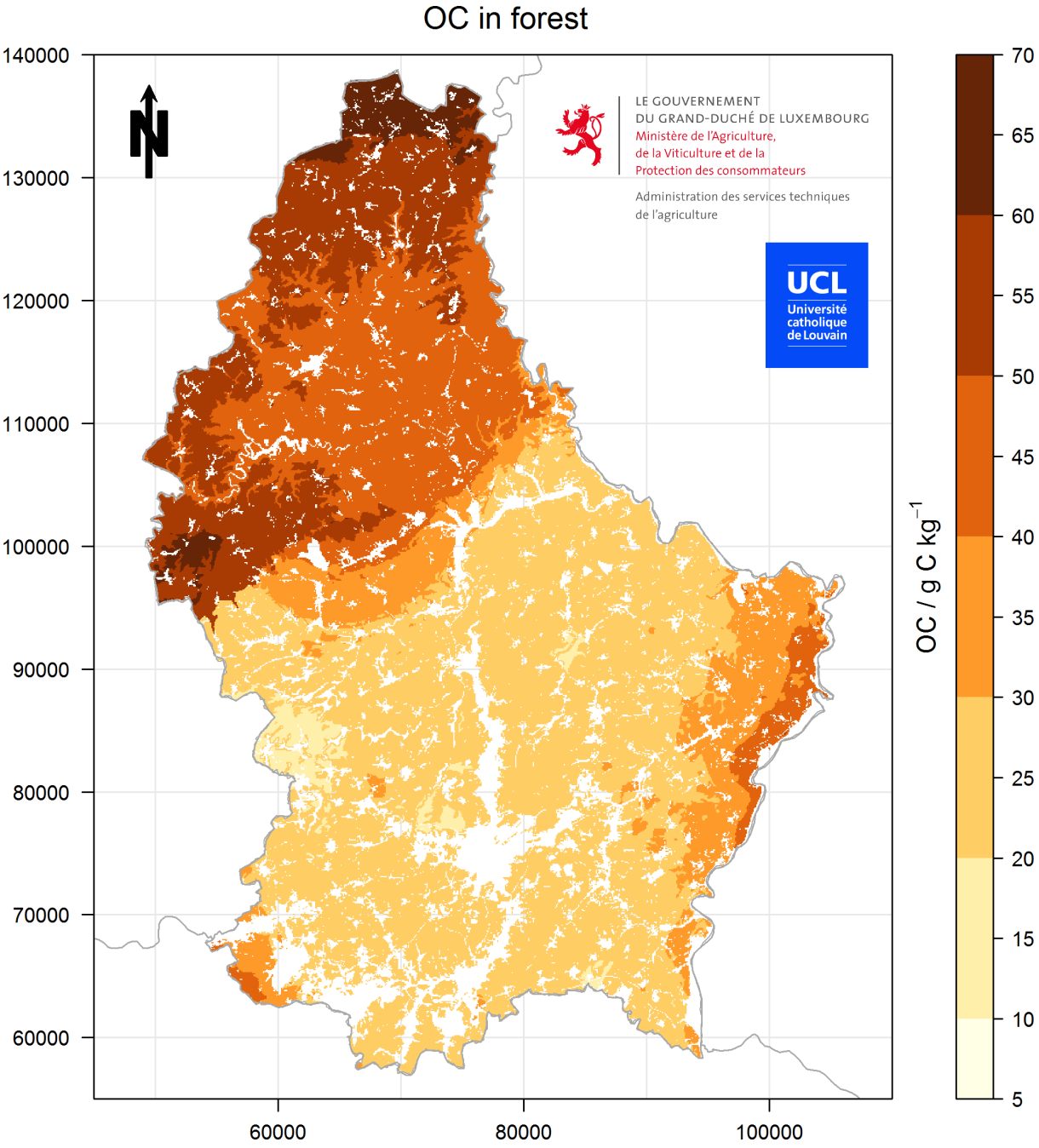

Figure 3.15: Map of predicted OC content (/ g C kg^{-1}) in forest soils, assuming the whole territory is covered by forest. Areas in white corresponds to environments where no predictions could be made.

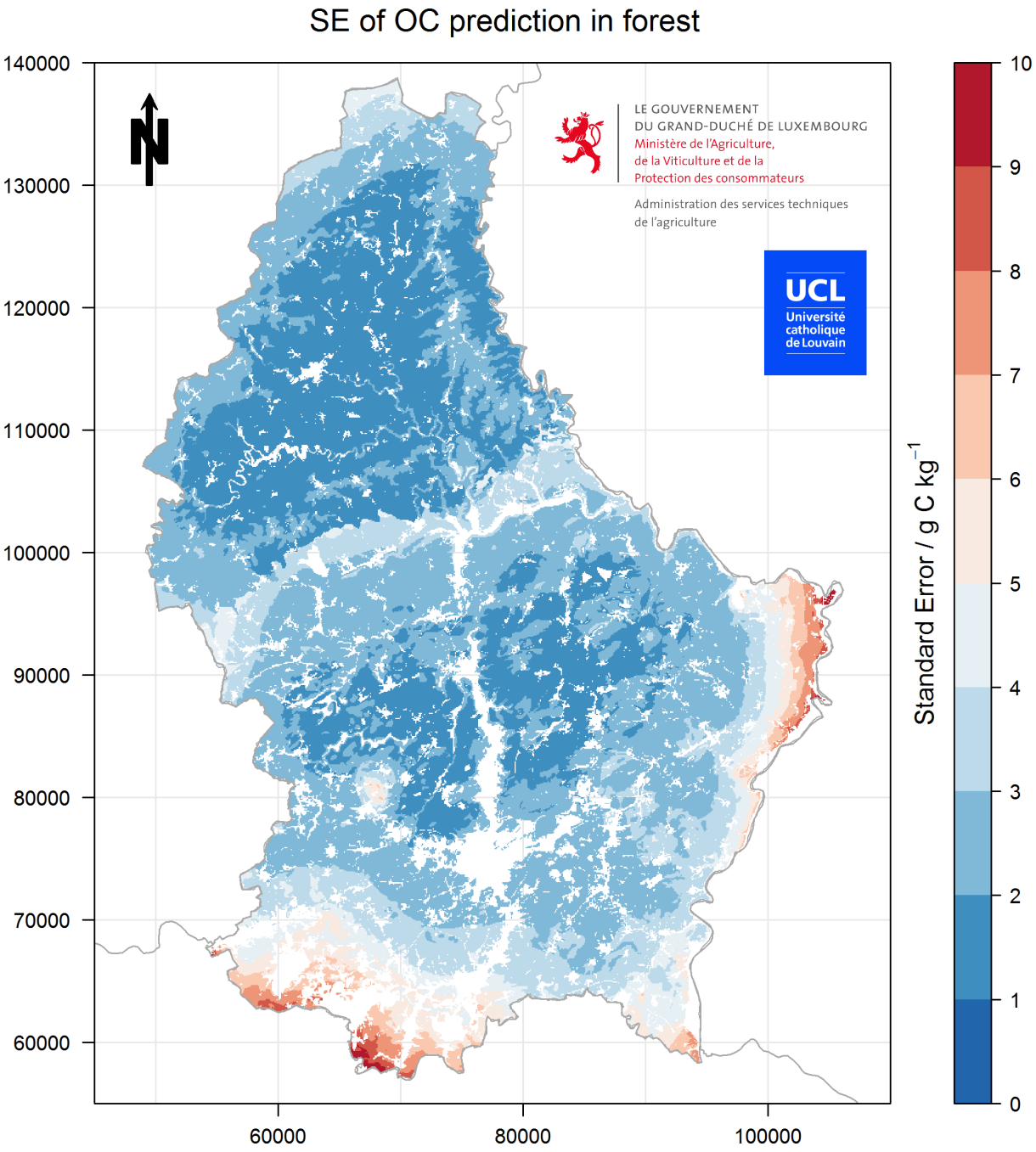

Figure 3.16: Map of the standard error of predicted OC content (/ g C kg^{-1}) in forest soils.

Table 3.7: Summary of the full and reduced GAM models for forests

	Full model	Reduced model
(Intercept)	3.75***	3.72***
elevation	2.26***	2.03***
eastness	0.00	
northness	0.00	
slope	1.94***	
precipitation	0.00	
temperature	0.00	
clay	0.89**	0.87**
LS factor	0.00	
path length	0.00	
x,y:Oesling	2.86***	4.93***
x,y:Gutland	8.64***	8.69***
AIC	5767.35	5775.60
BIC	5851.26	5859.19
Log Likelihood	-2865.09	-2869.29
Deviance	192143.43	194546.53
Deviance explained	0.40	0.39
Dispersion	292.27	295.89
R^2	0.39	0.38
GCV score	2896.61	2900.91
Num. obs.	675	675
Num. smooth terms	11	4

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.8: Summary of the full and reduced GAM models for vineyards

	Full model	Reduced model
(Intercept)	2.83***	2.84***
elevation	1.46	3.50***
eastness	1.38**	
northness	0.00	
slope	0.00	
precipitation	0.00	
temperature	5.47***	3.05***
C factor	0.00	
clay	5.30***	5.71***
Livestock intensity	0.00	
LS factor	2.04**	
path length	1.85***	
x,y	8.12***	9.14
AIC	5080.12	5149.63
BIC	5206.30	5256.54
Log Likelihood	-2512.44	-2551.41
Deviance	48416.84	54019.13
Deviance explained	0.23	0.14
Dispersion	70.64	78.33
R^2	0.20	0.11
GCV score	2563.01	2581.90
Num. obs.	712	712
Num. smooth terms	12	4

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

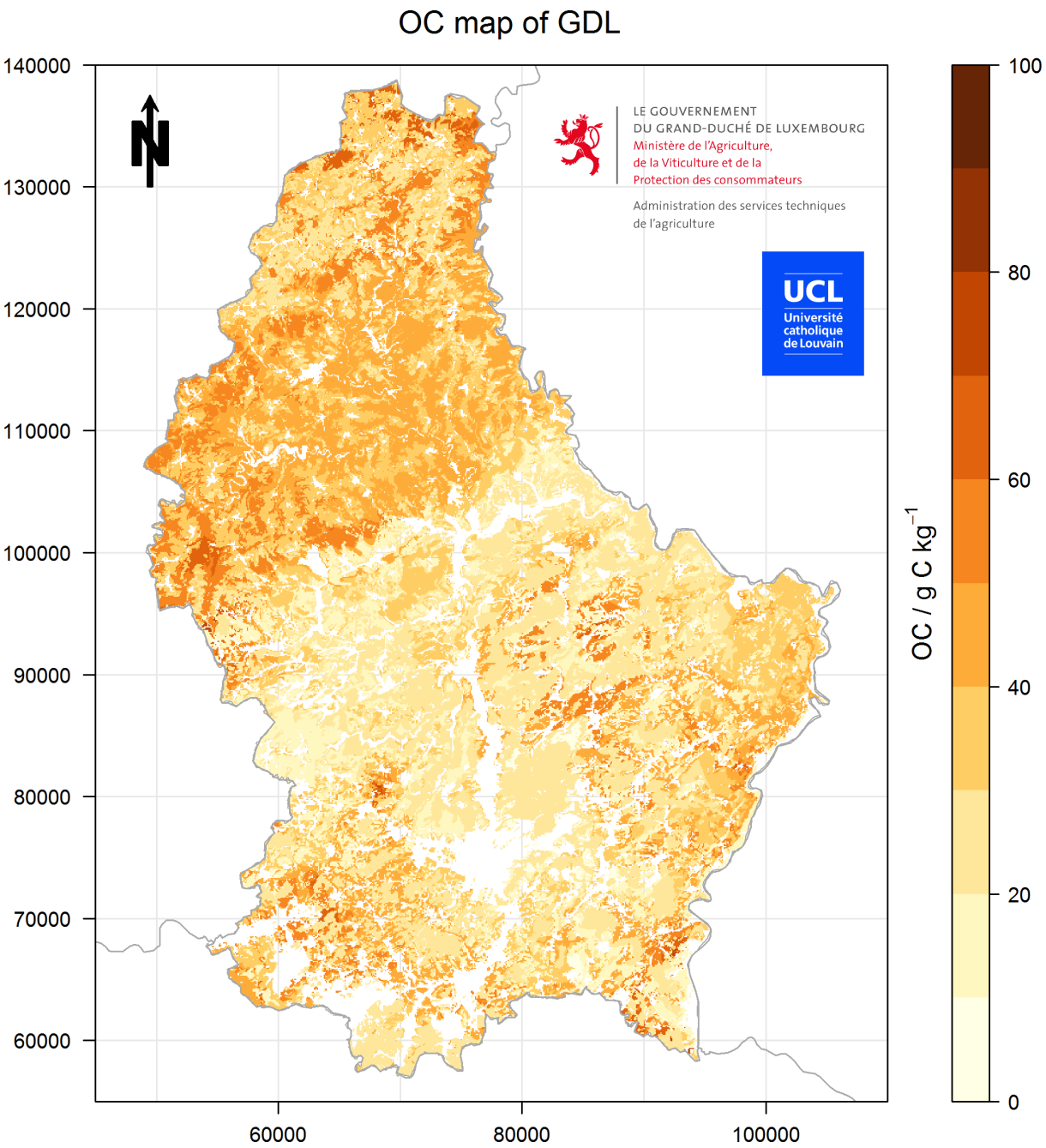

Figure 3.17: Map of predicted OC content (/ g C kg^{-1}) in GDL. Areas in white corresponds to environments where no predictions could be made.

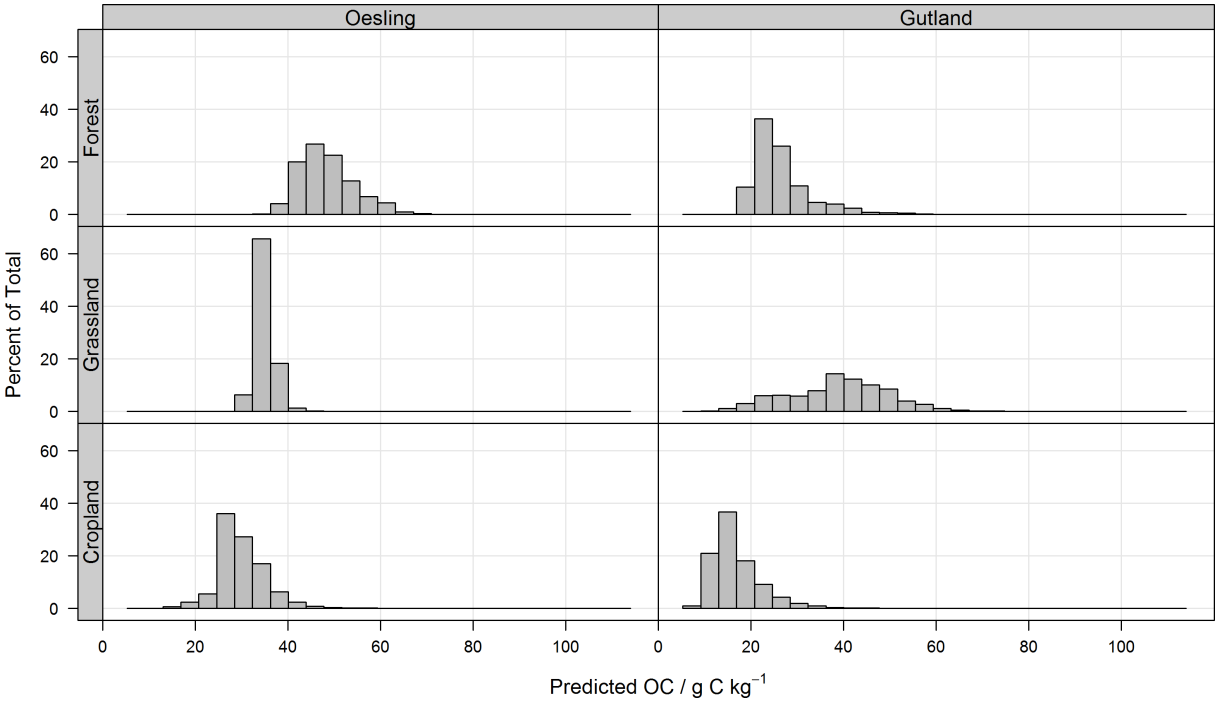

Figure 3.18: Histogram of values shown in the map of OC (/ g C kg^{-1}) in GDL soils.

Chapter 4

Conclusion

Soil is a non-renewable resource providing a range of services to our societies, from plant production to water filtering, which human activities are putting under pressure. The soil organic carbon is the main constituent of the soil organic matter, a key parameter of soil quality through its influence on soil aggregate stability, erosion sensitivity, nutrient and water holding capacity, biological activity and *in fine* its fertility. Soil carbon represents also the main terrestrial carbon pool, and, to this regard, is at the heart of many contemporary environmental and economical stakes. In Europe, the decline in soil organic matter due to agriculture intensification has been identified as one of the major threats to the soil. Hence, the knowledge of OC variation in space and the understanding of its dynamic and controlling variables is of crucial importance to support and target protection measures.

Here, we presented the results of a study analyzing the spatial distribution of OC in the Grand-Duchy of Luxembourg in four main land cover classes (cropland, grassland, forest and vineyard), based on an extensive database compiled from soil tests conducted by the *Service de Pédologie* at ASTA. Firstly, indicative or likely OC ranges were defined for combination of land cover classes and soil associations, to provide benchmark levels to which new OC observations can be compared. Then, the spatial variation of OC in the four land cover classes was modeled through Generalized Additive Models (GAM) using a set of available covariates representing climatic, pedologic, topographic and anthropic factors that are known to control accumulation and decomposition of soil organic matter. The GAM models were then applied to spatially-continuous layers of the covariates to produce four maps OC content for each land cover that were combined into one final map covering as much as 87% of the GDL territory. The resulting OC maps depict spatial patterns related to covariates selected by the backward stepwise procedure: climatic gradients and texture for cropland soils, texture and drainage for grasslands, elevation for forest soils. In vineyards, due to the poor model results, no map was produced. We evaluated the model prediction accuracy on a random portion of the observations and obtained R^2's of 0.66 for cropland, 0.33 for grassland, 0.37 for forest and 0.03 for vineyard soils. The maps of OC content are based on empirical models and their accurary depends strongly on the number of sampling points and spatial coverage over the territory. The mapping procedure that we used can be reproduced and is given as R codes in digital Annex attached to the report (see Annex E). This will allow to rapidly update and improve the OC maps when new samples will be analyzed by ASTA. One can especially foresee future improvements for the maps of croplands and grasslands because they both show a relatively low nugget-to-sill ratio, indicating a strong spatial dependence of the observations which could be leveraged to extrapolate in space. We should note however that many factors influencing soil OC operates at very small scales (e.g. erosion) and gathering spatial data for some of them is difficullt (e.g. local management), so that high uncertainties in regional/national OC assessments is innevitable [62], as observed for instance in forest and vineyard soils.

Bibliography

[1] Johan Röckstrom, Will Steffen, Kevin Noone, Asa Persson, F. Stuart Chapin, Eric F. Lambin, Timothy M. Lenton, Marten Scheffer, Carl Folke, Hans Joachim Schellnhuber, Björn Nykvist, Cynthia A. de Wit, Terry Hughes, Sander van der Leeuw, Henning Rodhe, Sverker Sörlin, Peter K. Snyder, Robert Costanza, Uno Svedin, Malin Falkenmark, Louise Karlberg, Robert W. Corell, Victoria J. Fabry, James Hansen, Brian Walker, Diana Liverman, Katherine Richardson, Paul Crutzen, and Jonathan A. Foley. A safe operating space for humanity. *Nature*, 461(7263):472–475, September 2009.

[2] Rattan Lal. Soil carbon sequestration impacts on global climate change and food security. *science*, 304(5677):1623–1627, 2004.

[3] Kristof Van Oost, T. A. Quine, G. Govers, S. De Gryze, J. Six, J. W. Harden, J. C. Ritchie, G. W. McCarty, G. Heckrath, C. Kosmas, J. V. Giraldez, J. R. Marques da Silva, and R. Merckx. The impact of agricultural soil erosion on the global carbon cycle. *Science*, 318(5850):626–629, October 2007.

[4] Z. G. Bai, D. L. Dent, L. Olsson, and M. E. Schaepman. Proxy global assessment of land degradation. *Soil Use and Management*, 24(3):223–234, 2008.

[5] R. Lal. Challenges and opportunities in soil organic matter research. *European Journal of Soil Science*, 60(2):158–169, 2009.

[6] Pedro A. Sanchez, Sonya Ahamed, Florence Carré, Alfred E. Hartemink, Jonathan Hempel, Jeroen Huising, Philippe Lagacherie, Alex B. McBratney, Neil J. McKenzie, Maria de Lourdes Mendonça-Santos, Budiman Minasny, Luca Montanarella, Peter Okoth, Cheryl A. Palm, Jeffrey D. Sachs, Keith D. Shepherd, Tor-Gunnar Vågen, Bernard Vanlauwe, Markus G. Walsh, Leigh A. Winowiecki, and Gan-Lin Zhang. Digital soil map of the world. *Science*, 325(5941):680–681, August 2009.

[7] S. Grunwald, J. A. Thompson, and J. L. Boettinger. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Science Society of America Journal*, 75(4):1201, 2011.

[8] E. G. Gregorich, C. M. Monreal, M. R. Carter, D. A. Angers, and BH74 Ellert. Towards a minimum data set to assess soil organic matter quality in agricultural soils. *Canadian journal of soil science*, 74(4):367–385, 1994.

[9] David A. Robinson, Bridget A. Emmett, Brian Reynolds, Ed C. Rowe, Dave Spurgeon, Aidan M. Keith, Inma Lebron, and Neal Hockley. Soil natural capital and ecosystem service delivery in a world of global soil change. *Soils and food security*, 35:41–68, 2012.

[10] P. Capriel. Trends in organic carbon and nitrogen contents in agricultural soils in bavaria (south germany) between 1986 and 2007. *European Journal of Soil Science*, 64(4):445–454, 2013.

[11] Jaakko Heikkinen, Elise Ketoja, Visa Nuutinen, and Kristiina Regina. Declining trend of carbon in finnish cropland soils in 1974–2009. *Global Change Biology*, 19(5):1456–1469, 2013.

[12] European Commission. Proposal for a directive of the european parliament and of the council establishing a framework for the protection of soil and amending directive 2004/35/EC. Technical Report COM(2006) 232, Brussels, 2006.

[13] F.g.a. Verheijen, P.h. Bellamy, M.g. Kibblewhite, and J.l. Gaunt. Organic carbon ranges in arable soils of england and wales. *Soil Use and Management*, 21(1):2–9, 2005.

[14] Dominique Arrouays, B. P. Marchant, N. Saby, Jeroen Meersmans, T. G. Orton, M. P. Martin, Patricia H Bellamy, R. M. Lark, and Mark G. Kibblewhite. Generic issues on broad-scale soil monitoring schemes: A review. *Pedosphere*, 22(4):456–469, August 2012.

[15] Bas van Wesemael, Keith Paustian, Olof Andrén, Carlos E. P. Cerri, Mike Dodd, Jorge Etchevers, Esther Goidts, Peter Grace, Thomas Kätterer, Brian G. McConkey, Stephen Ogle, Genxing Pan, and Clemens Siebner. How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO2 fluxes in agricultural soils? *Plant and Soil*, 338(1-2):247–259, January 2011.

[16] Tomislav Hengl, European commission, Joint research centre, and Italie) Institute for environment and sustainability (Ispra. *A practical guide to geostatistical mapping of environmental variables*. Publications Office, Luxembourg, 2007.

[17] Suzanna Lettens, Jos Van Orshoven, Bas Van Wesemael, Bart Muys, and Dominique Perrin. Soil organic carbon changes in landscape units of belgium between 1960 and 2000 with reference to 1990. *Global Change Biology*, 11(12):2128–2140, 2005.

[18] J. Meersmans, F. De Ridder, F. Canters, S. De Baets, and M. Van Molle. A multiple regression approach to assess the spatial distribution of soil organic carbon (SOC) at the regional scale (flanders, belgium). *Geoderma*, 143(1–2):1–13, January 2008.

[19] A. B McBratney, M. L Mendonça Santos, and B Minasny. On digital soil mapping. *Geoderma*, 117(1–2):3–52, November 2003.

[20] M. P. Martin, M. Wattenbach, P. Smith, J. Meersmans, C. Jolivet, L. Boulonne, and D. Arrouays. Spatial distribution of soil organic carbon stocks in france. *Biogeosciences*, 8(5):1053–1065, May 2011.

[21] B. L. Henderson, E. N. Bui, C. J. Moran, and D. A. P. Simon. Australia-wide predictions of soil properties using decision trees. *Geoderma*, 124(3–4):383–398, February 2005.

[22] R. Grimm, T. Behrens, M. Märker, and H. Elsenbeer. Soil organic carbon concentrations and stocks on barro colorado island - digital soil mapping using random forests analysis. *Geoderma*, 146(1–2):102–113, July 2008.

[23] Tomislav Hengl, Gerard B.M. Heuvelink, and Alfred Stein. A generic framework for spatial prediction of soil variables based on regression-kriging. 120(1-2):75–93, 2004.

[24] Rainer Baritz, Dietmar Zirlewagen, Robert Jones, Dominique Arrouays, Roland Hiederer, Marion Schrumpf, and Winfried Riek. Carbon in european soils. In Robert Jandl, Mirco Rodeghiero, and ts Olsson, editors, *Soil Carbon in Sensitive European Ecosystems*, page 49–84. John Wiley & Sons, Ltd, 2011.

[25] Gustavo M. Vasques, S. Grunwald, and D. Brenton Myers. Influence of the spatial extent and resolution of input data on soil carbon models in florida, USA. *Journal of Geophysical Research*, 117(G4), October 2012.

[26] Karen W. Holmes, Dar A. Roberts, Stuart Sweeney, Izaya Numata, Eraldo Matricardi, Trent W. Biggs, Getulio Batista, and Oliver A. Chadwick. Soil databases and the problem of establishing regional biogeochemical trends. *Global Change Biology*, 10(5):796–814, 2004.

[27] Dan-Dan Wang, Xue-Zheng Shi, Hong-Jie Wang, D. C. Weindorf, Dong-Sheng Yu, Wei-Xia Sun, Hong-Yan Ren, and Yong-Cun Zhao. Scale effect of climate and soil texture on soil organic carbon in the uplands of northeast china. *Pedosphere*, 20(4):525–535, August 2010.

[28] S. Lettens, J. Van Orshoven, B. van Wesemael, and B. Muys. Soil organic and inorganic carbon contents of landscape units in belgium derived using data from 1950 to 1970. *Soil Use and Management*, 20(1):40–47, 2004.

[29] J. Meersmans, B. van Wesemael, F. De Ridder, and M. Van Molle. Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (flanders, belgium). *Geoderma*, 152(1–2):43–52, August 2009.

[30] Jeroen Meersmans, Manuel Pascal Martin, Eva Lacarce, Sarah De Baets, Claudy Jolivet, Line Boulonne, Sébastien Lehmann, Nicolas Philippe Anthony Saby, Antonio Bispo, and Dominique Arrouays. A high resolution map of french soil organic carbon. *Agronomy for Sustainable Development*, 32(4):841–851, October 2012.

[31] Panos Panagos, Roland Hiederer, Marc Van Liedekerke, and Francesca Bampa. Estimating soil organic carbon in europe based on data collected through an european network. *Ecological Indicators*, 24:439–450, January 2013.

[32] IUSS Working Group WRB. World reference base for soil resources 2014. international soil classification system for naming soils and creating legends for soil maps. Technical Report 106, FAO, Rome, 2014.

[33] Ruth Kerry, Pierre Goovaerts, Barry G. Rawlins, and Ben P. Marchant. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170:347–358, January 2012.

[34] Jacques Rondeux, Vincent Colson, André Thibaut, Nils Bourland, Gaetan Cuartero Diaz, and Marc Wagner. L'inventaire forestier national permanent du grand-duché de luxembourg et ses aspects méthodologiques. *Revue forestière française*, 57(1):51–62, 2005.

[35] Robert J. Hijmans and Jacob van Etten. raster: Geographic analysis and modeling with raster data. *R package version*, 1:9–92, 2012.

[36] V. Olaya. A gentle introduction to SAGA GIS. *The SAGA User Group eV, Gottingen, Germany*, 208, 2004.

[37] Jeff Jenness. *Topographic Position Index (tpi_jen. avx) extension for ArcView 3. x, v. 1.3 a. Jenness Enterprises.* 2006.

[38] J. H. Zar. *Biostatistical Analysis.* Prentice Hall, New Jersey, 1999.

[39] J. Meersmans, Bas Van Wesemael, E. Goidts, M. Van Molle, S. De Baets, and F. De Ridder. Spatial analysis of soil organic carbon evolution in belgian croplands and grasslands, 1960–2006. *Global Change Biology*, 17(1):466–479, 2011.

[40] Eric P. J. Boer, Kirsten M de Beurs, and A Dewi Hartkamp. Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation*, 3(2):146–154, 2001.

[41] Andy Jarvis, Hannes Isaak Reuter, Andrew Nelson, and Edward Guevara. Hole-filled SRTM for the globe version 4. *available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org)*, 2008.

[42] Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical science*, 1(3):297–310, 1986.

[43] Simon N. Wood. mgcv: GAMs and generalized ridge regression for r. *R news*, 1(2):20–25, 2001.

[44] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Additive models, trees, and related methods. In *The Elements of Statistical Learning*, Springer Series in Statistics, pages 295–336. Springer New York, January 2009.

[45] D. J. Brus, B. Kempen, and G. B. M. Heuvelink. Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62(3):394–407, 2011.

[46] P Loveland and J Webb. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. *Soil and Tillage Research*, 70(1):1–18, March 2003.

[47] V. Genot, M. Renneson, G. Colinet, M. J. Goffaux, T. Cugnon, B. Toussaint, D. Buffet, and R. Oger. Base de données sols de REQUASUD. 3ème synthèse. Technical report, Requasud, Gembloux, Belgium, 2011.

[48] J. Six, R. T. Conant, E. A. Paul, and K. Paustian. Stabilization mechanisms of soil organic matter: Implications for c-saturation of soils. *Plant and Soil*, 241(2):155–176, April 2002.

[49] J Balesdent, C Chenu, and M Balabane. Relationship of soil organic matter dynamics to physical protection and tillage. *Soil and Tillage Research*, 53(3–4):215–230, February 2000.

[50] Miko U. F. Kirschbaum. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic c storage. *Soil Biology and Biochemistry*, 27(6):753–760, June 1995.

[51] Wilfred M. Post, William R. Emanuel, Paul J. Zinke, and Alan G. Stangenberger. Soil carbon pools and world life zones. *Nature*, 298(5870):156–159, July 1982.

[52] Susan E. Trumbore, Oliver A. Chadwick, and Ronald Amundson. Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change. *Science*, 272(5260):393–396, April 1996.

[53] R. J. Haynes and R. Naidu. Influence of lime, fertilizer and manure applications on soil organic matter content and soil physical conditions: a review. *Nutrient Cycling in Agroecosystems*, 51(2):123–137, June 1998.

[54] Eric A. Davidson. Spatial covariation of soil organic carbon, clay content, and drainage class at a regional scale. *Landscape Ecology*, 10(6):349–362, December 1995.

[55] I. C. Burke, C. M. Yonker, W. J. Parton, C. V. Cole, D. S. Schimel, and K. Flach. Texture, climate, and cultivation effects on soil organic matter content in u.s. grassland soils. *Soil Science Society of America Journal*, 53(3):800, 1989.

[56] Jeroen Meersmans, Manuel Pascal Martin, Fjo De Ridder, Eva Lacarce, Johanna Wetterlind, Sarah De Baets, Christine Le Bas, Benjamin P. Louis, Thomas G. Orton, Antonio Bispo, and Dominique Arrouays. A novel soil organic c model using climate, soil type and management data at the national scale in france. *Agronomy for Sustainable Development*, 32(4):873–888, October 2012.

[57] J. Skopp, M. D. Jawson, and J. W. Doran. Steady-state aerobic microbial activity as a function of soil water content. *Soil Science Society of America Journal*, 54(6):1619, 1990.

[58] Eric A. Davidson and Ivan A. Janssens. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, 440(7081):165–173, March 2006.

[59] Dale W Johnson and Peter S Curtis. Effects of forest management on soil c and n storage: meta analysis. *Forest Ecology and Management*, 140(2–3):227–238, January 2001.

[60] A. Stevens and B. Van Wesemael. Soil organic carbon dynamics at the regional scale as influenced by land use history: a case study in forest soils from southern belgium. *Soil Use and Management*, 24(1):69–79, March 2008.

[61] Ingo Schöning, Kai Uwe Totsche, and Ingrid Kögel-Knabner. Small scale spatial variability of organic carbon stocks in litter and solum of a forested luvisol. *Geoderma*, 136(3–4):631–642, December 2006.

[62] Esther Goidts and Bas van Wesemael. Regional assessment of soil organic carbon changes under agriculture in southern belgium (1955–2005). *Geoderma*, 141(3–4):341–354, October 2007.

[63] David J. Brown, Keith D. Shepherd, Markus G. Walsh, M. Dewayne Mays, and Thomas G. Reinsch. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, 132(3–4):273–290, 2006.

Appendix A

Replicate analysis

To assess temporal discrepancies between OC analyzes due to the use of different instruments and protocols at ASTA, we conducted a replicate analysis of the samples analyzed with the TruSpec (LECO) in 2010-2011 and measured them with the Multi EA 4000 (Analytik Jena; Table 2.2). We first grouped samples in two classes with either low pH_{CaCl_2} ($pH_CaCl_2 < 7$) or high pH_{CaCl_2} ($pH_CaCl_2 \geq 7$). Then, 50 samples in each class were selected by stratified random sampling on the basis of the deciles of the observed OC distribution (i.e. 5 samples in each decile). Samples were re-analyzed and two outliers with replicate difference ≥ 2 were removed. As indicated in Figure A.1, the error is almost twice as large for samples with a high pH_{CaCl_2} (coming mostly from vineyards), strongly suggesting that the main source of error is related to the way Total Inorganic Carbon (TIC) is measured. The Root Mean Square Error is evaluated at 3.7 g C kg^{-1} for samples with $pH_CaCl_2 < 7$ and 7.7 g C kg^{-1} $pH_CaCl_2 \geq 7$. Although such errors might seem relatively large, one should note that this is far from unusual when comparing OC analyzes retrieved by different reference methods. For instance, Brown et al. [63] conducted an inter-laboratory comparison of OC analyzes and observed a RMSE of 4.4 g C $kg^{-1}$for TIC and 7.7 g C $kg^{-1}$for TC-TIC (=TOC).

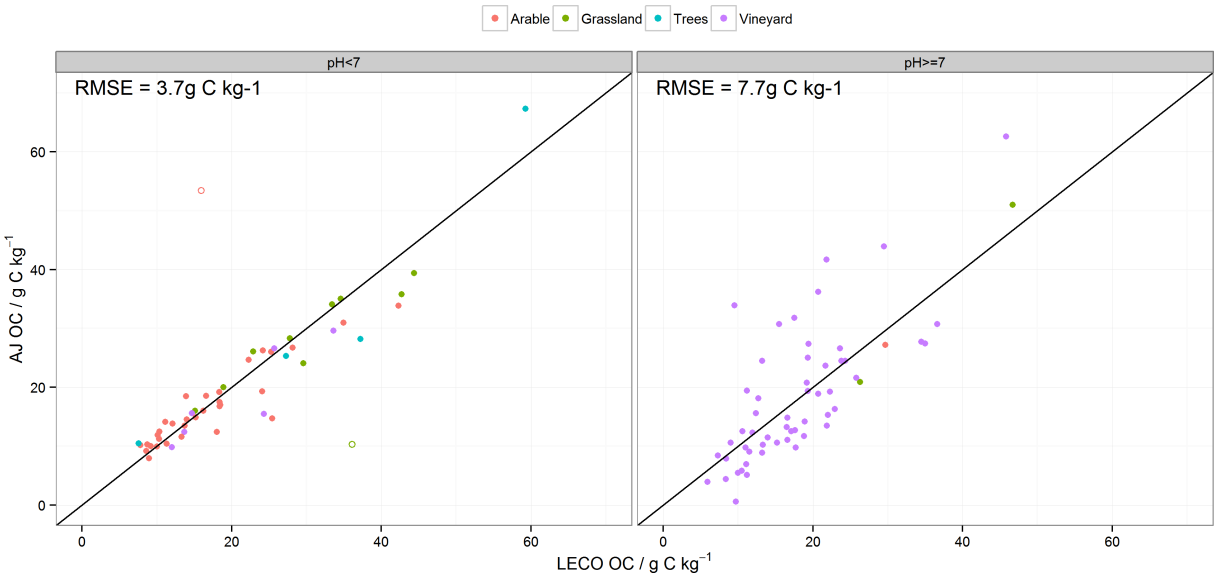

Figure A.1: Replicate analysis of OC measured with a LECO TruSpec and a Multi EA 4000 Analytic Jena at ASTA. Two outliers with a replicate difference ≥ 2 are indicated with circle symbols (o).

Appendix B

Observed OC values in the ASTA-SOC database

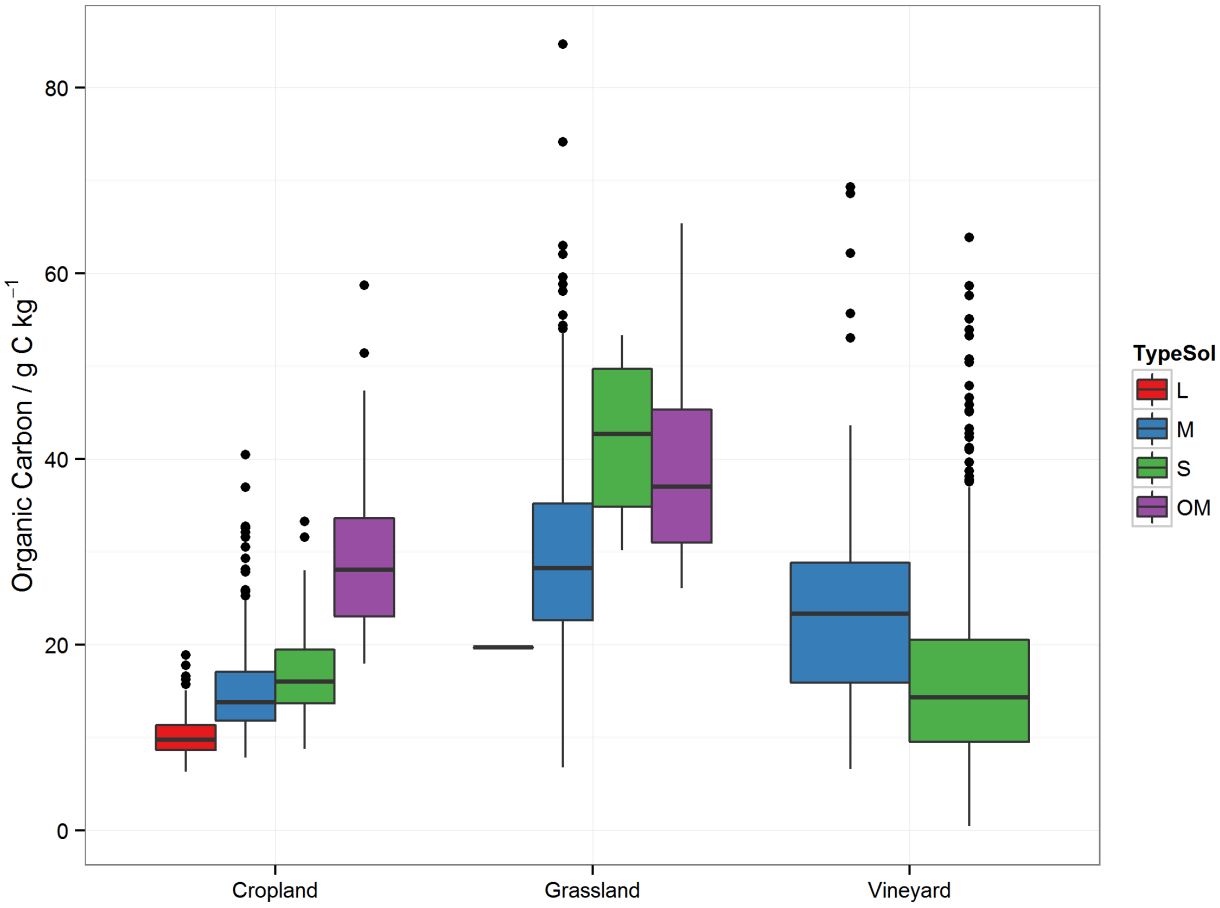

Figure B.1: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under cropland, grassland and vineyard land cover as a function of soil texture classes defined by ASTA (L = {S+Z}; M = {L+P+E+A}; OM = G; S = U). Note: texture class was not available for forests.

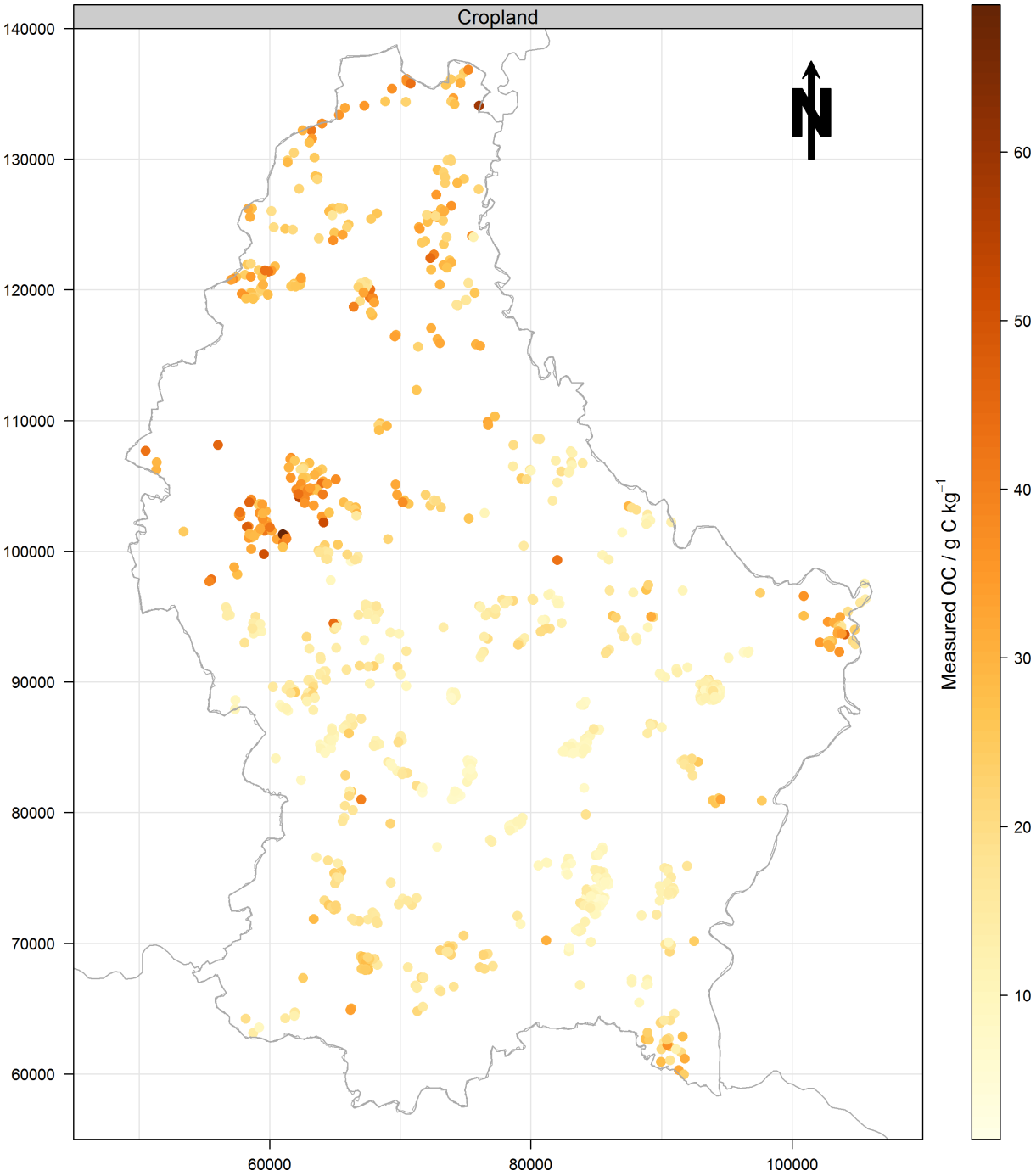

Figure B.2: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under cropland land cover.

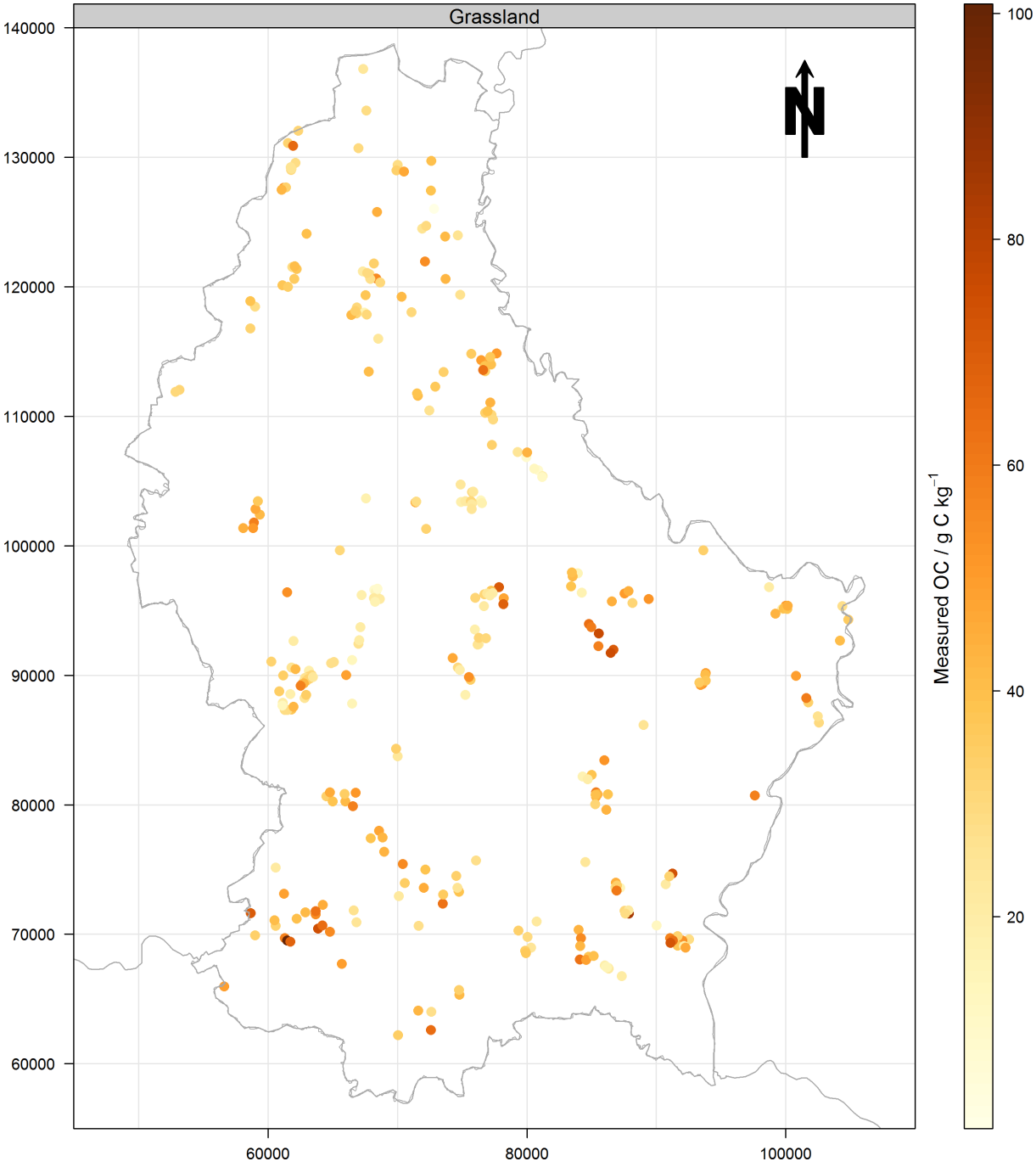

Figure B.3: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under grassland land cover.

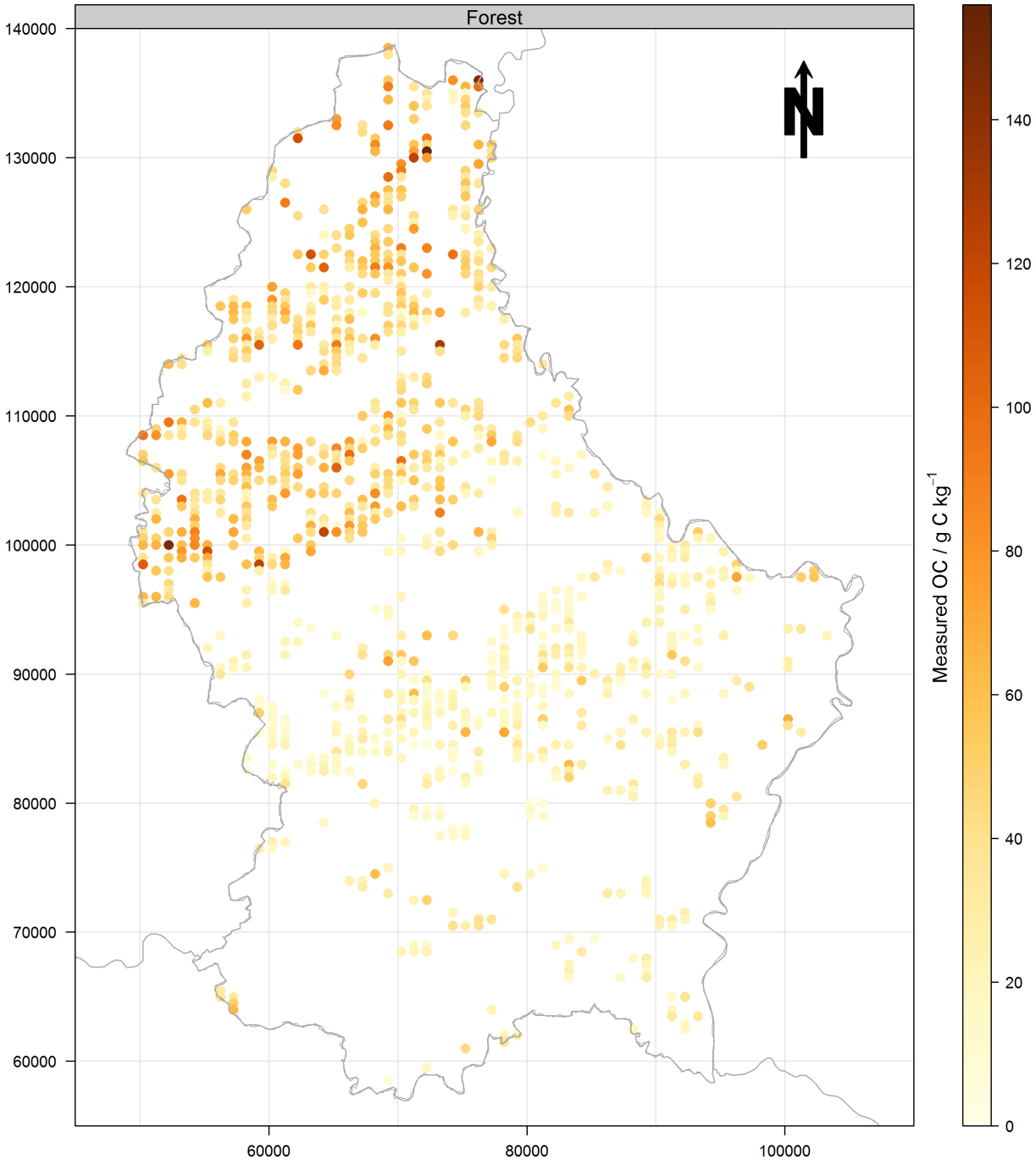

Figure B.4: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under forest land cover.

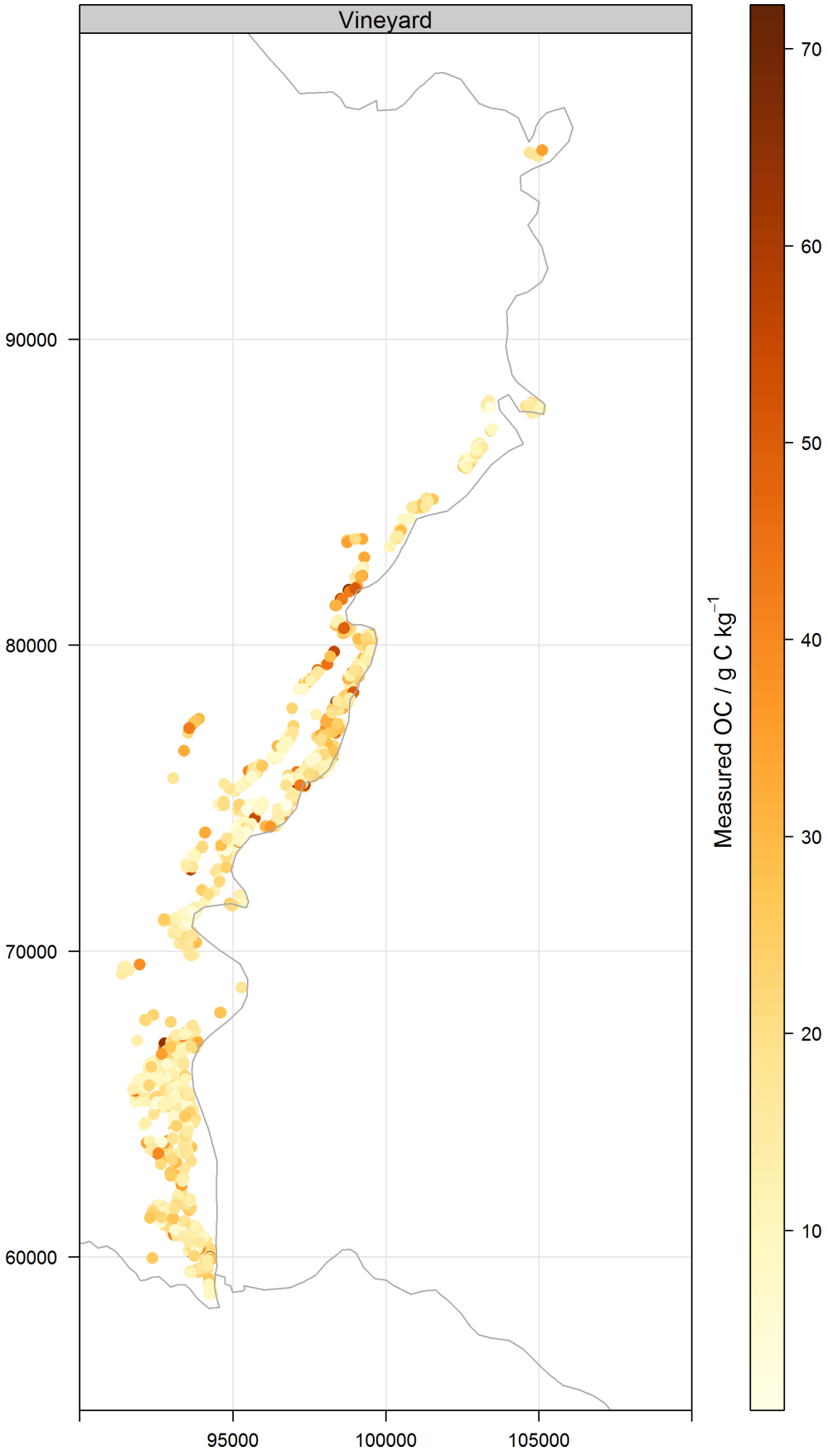

Figure B.5: Observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database under vineyard land cover.

Appendix C

Relation between observed OC values and morphometric variables

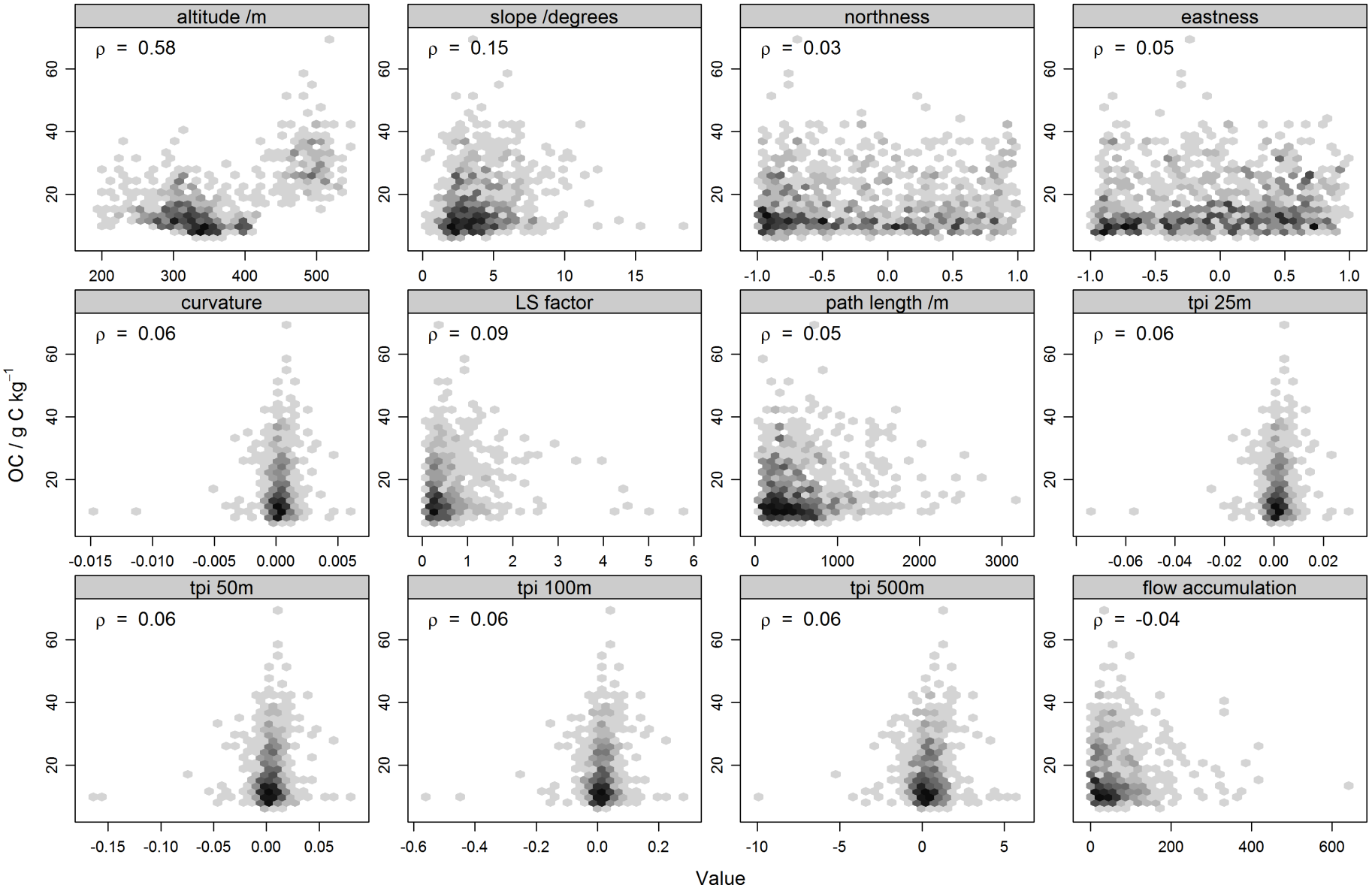

Figure C.1: Scatter plots of observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database against morphometric variables in cropland soils.

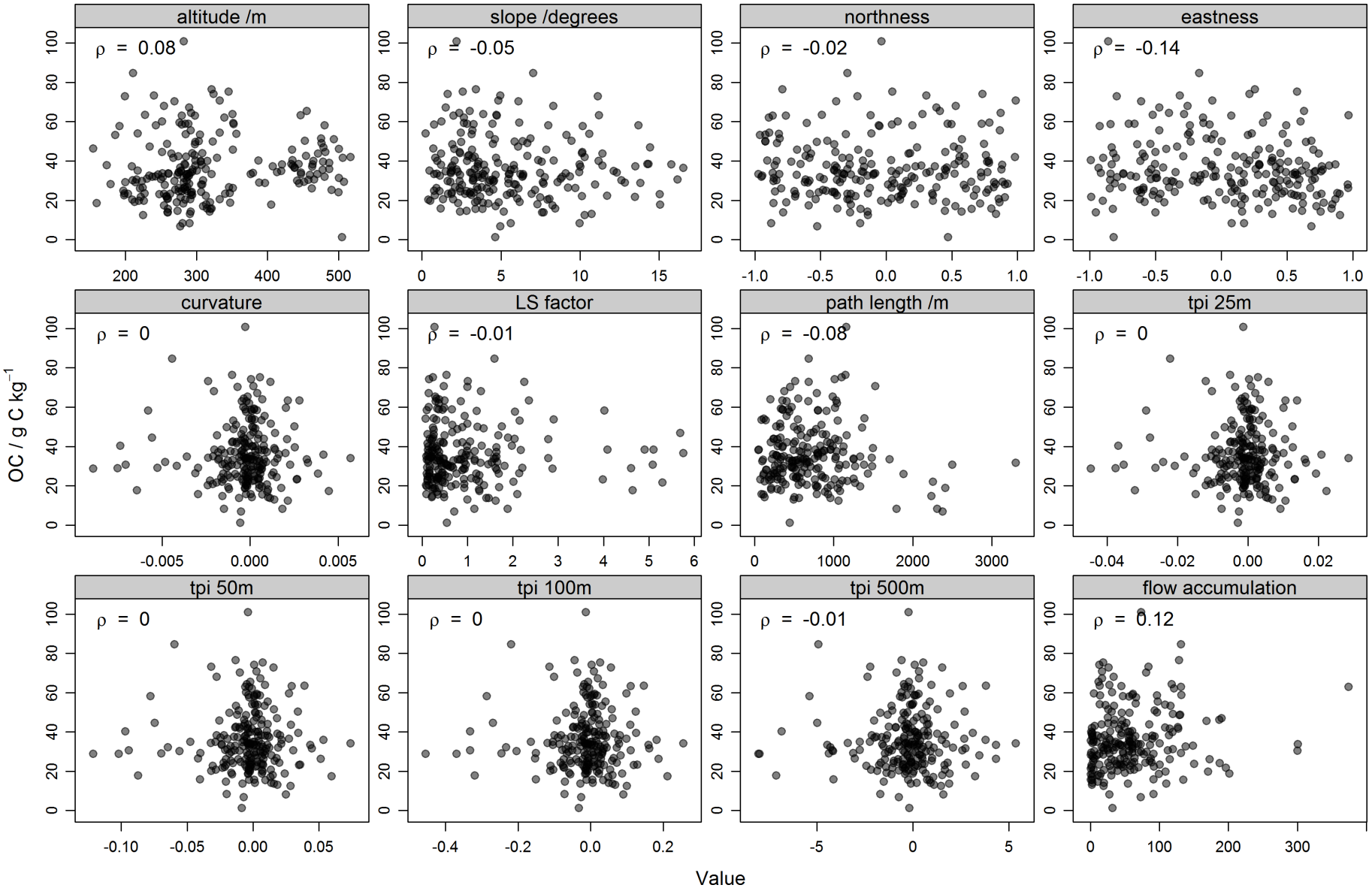

Figure C.2: Scatter plots of observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database against morphometric variables in grassland soils.

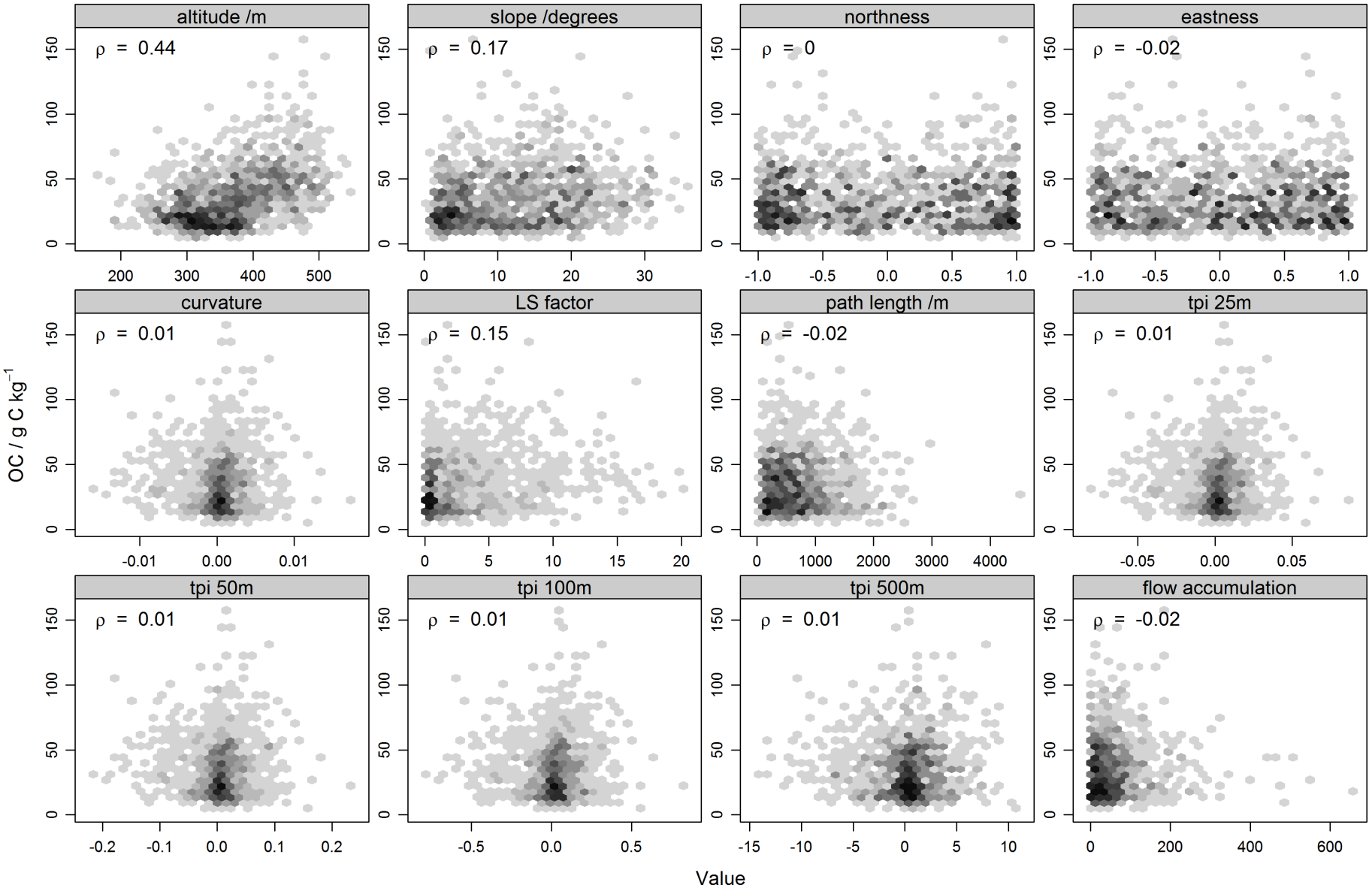

Figure C.3: Scatter plots of observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database against morphometric variables in forest soils.

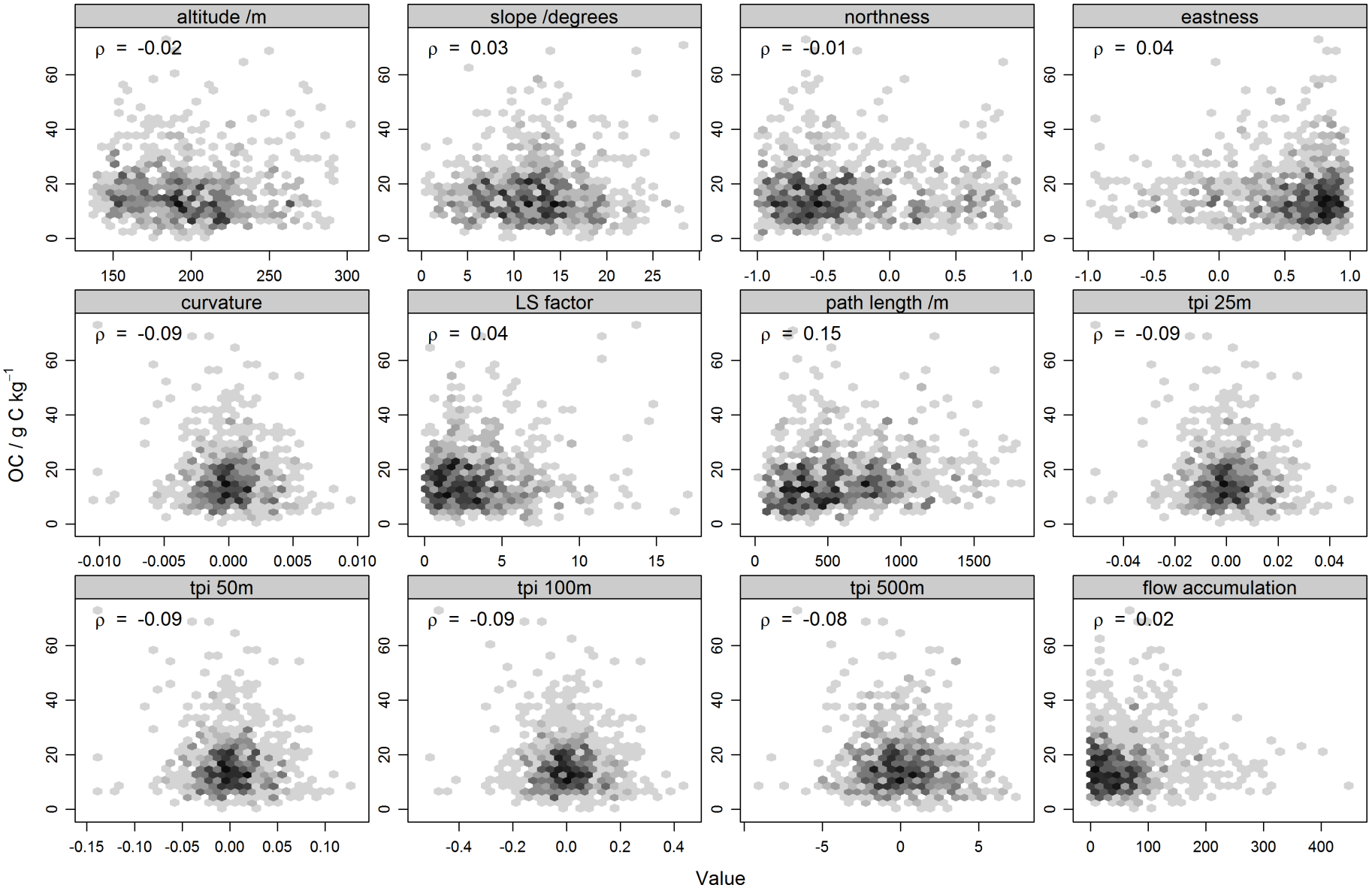

Figure C.4: Scatter plots of observed OC values (/ g C kg^{-1}) of samples of the ASTA-SOC database against morphometric variables in vineyard soils.

Appendix D

Summary of predicted OC values

Soil associations	n	Min	q_1	$\bar{x}$	$\widetilde{x}$	q_3	Max	IQR
Oesling	26748	14.0	26.4	29.9	29.2	32.7	59.8	6.3
Buntsandstein	4373	10.5	15.3	17.4	16.5	17.9	41.5	2.5
Dolomies du Muschelkalk	3746	12.4	19.7	24.3	23.3	28.1	45.7	8.5
Calcaires du Bajocien	310	12.7	15.8	21.5	23.4	26.3	35.9	10.4
Grès de Luxembourg	6038	7.5	10.7	12.3	12.1	13.5	23.0	2.8
Dépôts limoneux sur Grès	7581	7.8	11.9	14.3	13.9	15.4	44.6	3.6
Argiles du Lias inf. et moyen	10333	9.5	14.7	17.5	16.8	20.0	47.4	5.3
Argiles lourdes du Keuper	4629	9.2	13.5	16.8	16.2	19.0	45.8	5.5
Argiles lourdes des schistes bitumineux	1535	11.3	15.0	21.0	21.2	25.7	34.4	10.7
Autres	2312	8.7	15.1	20.6	17.7	26.8	59.5	11.7
all	67605	7.5	14.9	22.1	21.0	28.4	59.8	13.5

Table D.1: Summary statistics of OC predictions (g C kg-1) in Cropland

Soil associations	n	Min	q_1	$\bar{x}$	$\widetilde{x}$	q_3	Max	IQR
Oesling	18775	14.6	34.3	35.1	35.2	36.0	49.9	1.7
Buntsandstein	3619	11.7	22.1	31.8	26.8	37.5	79.9	15.4
Dolomies du Muschelkalk	5299	16.9	29.1	37.5	38.4	44.9	68.8	15.8
Calcaires du Bajocien	250	24.8	37.8	41.9	39.2	47.1	80.0	9.3
Grès de Luxembourg	3574	12.0	18.6	24.5	23.9	28.3	51.7	9.7
Dépôts limoneux sur Grès	8133	12.8	24.6	31.8	32.0	37.7	75.9	13.1
Argiles du Lias inf. et moyen	14451	15.3	37.6	42.4	42.3	47.0	76.4	9.4
Argiles lourdes du Keuper	13664	15.6	37.4	42.0	41.7	47.6	79.9	10.2
Argiles lourdes des schistes bitumineux	3443	29.4	48.6	52.0	53.2	56.8	79.5	8.2
Autres	6889	14.0	33.0	36.9	36.0	41.5	75.2	8.5
all	78097	11.7	33.5	37.8	36.5	43.3	80.0	9.7

Table D.2: Summary statistics of OC predictions (g C kg-1) in Grassland

Soil associations	n	Min	q_1	$\bar{x}$	$\widetilde{x}$	q_3	Max	IQR
Oesling	42055	31.4	44.1	48.4	47.7	51.8	69.2	7.7
Buntsandstein	4901	21.1	26.2	35.2	31.9	42.7	60.5	16.6
Dolomies du Muschelkalk	4673	22.7	30.3	35.4	35.9	40.4	50.1	10.1
Calcaires du Bajocien	3030	23.7	25.6	29.2	26.4	32.5	45.2	6.9
Grès de Luxembourg	21308	17.3	20.4	21.4	21.2	21.9	42.7	1.5
Dépôts limoneux sur Grès	11903	18.7	23.3	26.4	24.7	28.8	47.3	5.5
Argiles du Lias inf. et moyen	9351	20.4	24.1	25.8	25.2	26.4	46.4	2.3
Argiles lourdes du Keuper	8160	19.8	25.0	28.3	26.6	30.4	46.0	5.5
Argiles lourdes des schistes bitumineux	3263	21.3	26.4	28.4	28.1	29.8	40.8	3.4
Autres	6296	19.4	23.9	36.2	37.9	45.7	67.5	21.7
all	114940	17.3	23.8	35.0	30.8	46.0	69.2	22.2

Table D.3: Summary statistics of OC predictions (g C kg-1) in Forest

Appendix E

R scripts used to generate OC maps

R scripts used to (i) import and format the ASTA database and spatial data, (ii) explore data and (iii) model OC content and create maps of topsoil OC content can be found in digital annex to this report. These scripts suppose that you have the necessary data in folders named soil_data and spatial_data. A short description of their purpose is given in the following list (the order of the files in the list is important and corresponds to the flow of procedures to create maps):

Prepare soil data

1. intro.r: Script run at beginning of each script - Load R packages, import shapefiles of the GDL, create generic functions
2. format_data.r: import and join ASTA xls files - clean data and carry out multiple checks - save data for later use
3. format_data_IFL.r: import ASTA and IFL xls files - join data, clean and save for later use
4. check_analytical_methods: replicate analysis between the different analytical instruments (see Appendix A)

Prepare spatial (covariate) data

1. aggregate_UF_data.r: import data about livestock intensity by field - aggregate it by GDL municipalities
2. create_clim_map.r: import climatological data from Lux and neighbouring countries - load a raster of altitude from SRTM mission - create maps of mean annual temperature and precipitation in GDL
3. create_texture_map.r: import soil profile data from ASTA - create maps of clay, silt and sand content in topsoils by regression kriging
4. get_spatial_data.r import spatial data from various sources - combine data into one raster stack with the same resolution and extent - save the covariate stack for later use

Join soil and spatial data

1. attach_spatial_data.r : using a mean or majority rule, extract data from the covariate for each polygon
2. attach_spatial_data_IFL.r : idem but for IFL data (points)

Explore, model and map OC in Lux

1. exploratory_analysis.r: combine data from FLIK's (cropland, grassland, vineyard) and from IFL (forest) - explore the relationships between OC and the covariates - compute variograms - compute summary statistics - map OC observation for each land cover

2. gam_model.r: calibrate a GAM model for each land cover - apply the model to the covariate stack to predict OC in space - plot the resulting maps
3. gam_model_IFL.r: idem than previously but for IFL data
4. create_OC_map.r: combine maps of each land cover into one - plot the resulting map

Miscellaneous / not used

1. test_kriging.r: test for residual kriging
2. textural_triangle.r: create textural triangle of GDL and plot texture of BDSOL
3. autocrop.r: auto-crop images in the Rapport/img directory