The devil is in the detail: discrepancy between SOC stocks estimated from regional and local data sources in Flanders, Belgium

Sam Ottoy^{a,b,*}, Laura Vanierschot^a, Stefaan Dondeyne^{a,c}, Karen Vancampenhout^d, Martin Hermy^a, Jos Van Orshoven^a

^a Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E box 2411, 3001 Leuven, Belgium

^b PXL Bio-Research, PXL University College, Agoralaan H, 3590 Diepenbeek, Belgium

^c Department of Geography, Ghent University, Krijgslaan 281, S8, 9000 Gent

^d Cluster for Bioengineering Technology, Department of Microbial and Molecular Systems, KU Leuven Campus Geel, Kleinhoefstraat 4, 2440 Geel, Belgium

*Corresponding author: e-mail: sam.ottoy@pxl.be

Abstract

Storage of soil organic carbon (SOC) is an essential function of ecosystems underpinning the delivery of multiple services to society. Regional SOC stock estimates often rely on data collected during land use specific inventory schemes with varying sampling depth and density. Using such data requires techniques than can deal with the associated heterogeneity. As the resulting SOC assessments are not calibrated for the local scale, they might suffer from oversimplification of landscape processes and heterogeneity. This might especially be the case for sandy regions where typical historical land-use practices and soil development processes determine SOC storage. The aims of this study were (i) to combine four land use specific SOC stock assessments to estimate the total stock in Flanders, Belgium and (ii) to evaluate the applicability of this regional-scale estimate at the local scale. We estimated the SOC stock in the upper 100 cm of the unsealed area in Flanders (887,745 ha) to be 111.67 Mt OC, or 12.58 ± 5.65 kg OC m⁻² on average. In general, soils under (semi-)natural land use types, e.g. forests, store on average more organic carbon than under agriculture land. However, agricultural soils store the largest amounts of SOC due to their vast extent. Zooming in on a sandy case study (13.55 km²) revealed the poor performance of the regional estimates, especially where Histosols occurred. Our findings show that a higher spatial sampling density is required when SOC stock estimates are needed to inform carbon-aware land management rather than to serve for regional reporting.

Keywords: Soil organic carbon; ecosystem service; regional assessment; local assessment; spatial scale; carbon hotspot

Introduction

The effective and potential level of services that ecosystems provide is increasingly inspiring policy makers (Broekx *et al.*, 2013). In assessments of soil fertility and climate regulation services, the soil organic carbon (SOC) stock is an important indicator (Schneiders & Müller, 2017). Policies target SOC storage at various scales and hence require information at different spatial levels (O'Rourke *et al.*, 2015; Vrebos *et al.*, 2017). Whereas carbon accounting and its reporting are typically performed at the national or regional scale (IPCC, 2006), the impacts of land management practices on SOC stocks need to be assessed at local levels (e.g. Parras-Alcántara *et al.* 2015). In many countries and regions, including the region of Flanders in Belgium (13,522 km²), heterogeneous or low spatial density data sources (e.g. Saby *et al.* 2008) limit their usability at both regional and local levels. For local assessments, e.g. focusing on the optimisation of land management in relation to SOC storage, it is essential to capture site-specific variation such as anthropogenic and hydrological gradients and hot and cold spots of SOC storage. As field surveys are costly and time-consuming, the question arises to what extent regional scale estimates are informative and relevant, and can be reused at the local scale. After all, regionally calibrated SOC models are often not calibrated for the local scale and might suffer from oversimplification of landscape processes and from landscape heterogeneity (Stevens *et al.*, 2014). This might especially be the case for sandy regions where typical historical land-use practices (e.g. plaggen management) and soil development processes (e.g. podzolisation) can result in the accumulation of SOC at different depths (Springob *et al.*, 2001).

Until now, estimates of the SOC stock in the upper 100 cm of the region of Flanders were based on legacy soil survey data, i.e. the Aardewerk dataset with data collected between 1947 and 1974, (Van Orshoven *et al.*, 1993)), so an estimate of the contemporary SOC stock is unavailable. To obtain total SOC stock estimates, data collected using different, often land use specific inventory schemes need to be combined. An example of the heterogeneous nature of SOC data is the difference in sampling depth, which can vary from one mixed sample of the topsoil (upper 15 – 30 cm) to horizon-based sampling up to 100 – 150 cm (e.g. see overviews in Lettens *et al.* (2005) and Minasny *et al.* (2013)). In addition, the limited and heterogeneous spatial density of soil inventories hamper regional assessments (Ottoy *et al.*, 2015). In earlier research, we have developed a depth extrapolation function (Ottoy *et al.*, 2016), and tested multiple digital soil mapping approaches (Ottoy *et al.*, 2017b) to deal with these shortcomings in the region of Flanders.

The overall aim of this study was to test the usability of regional-scale SOC estimates at the local scale. To this end, three specific objectives were formulated: (i) combine various regional-scale datasets to calculate the total SOC stock in the region of Flanders, Belgium, (ii) use these regional-scale assessments to estimate the total SOC stock of a particular study area and (iii) compare the latter estimate with stocks derived from local information with higher spatial density.

Materials and Methods

Study areas

Soil organic carbon stocks were estimated for the region of Flanders (13,522 km²), northern Belgium. This region is characterised by a maritime temperate climate, with a mean annual temperature of 9.8 – 10.5 °C and a mean annual precipitation of 733 – 832 mm (Peel *et al.*, 2007). In general, soil texture varies from sandy soils in the north to silt-loam in the south. The local study area was selected in the valley of the Kleine Nete, situated in the north of Flanders (Figure 1). This 13.55 km² area is characterised by a gradient in soil types and anthropogenic influences, resulting in a range of natural and historical controls of soil carbon stocks. The natural topographic gradient ranges from valley soils with high groundwater tables to dry land dunes. The soils in the valley are dominated by Histosols, Gleysols, Gleyic Arenosols and Umbrisols on mixed parent materials of Tertiary and Quaternary sand deposits and alluvial material. Further from the river, Podzols with different groundwater influence are present and the influence of poorer Pleistocene cover sands increases. In the North, the valley rim consists of Tertiary marine deposits locally capped by iron sandstone and Pleistocene cover sands. Some historical land dunes also occur and are classified as Protic Arenosols. Centuries of transferring organic matter from heathlands and swamps to arable fields around village centres resulted in the formation of Plaggic Anthrosols. Furthermore, part of the study area's peatlands were severely drained for agricultural purposes in the 1950's.

Regional data

Land use specific datasets of SOC stocks are available for arable land, grassland, forest and Low-Input High-Diversity (LIHD) systems covering the region of Flanders (Figure 1 and Table 1). The Soil Service of Belgium had collected 94,342 topsoil samples from agricultural land (388,572 ha of arable land and 315,074 ha of grassland) (Maes *et al.*, 2012). These measurements are limited to the topsoil (23 cm for arable land

and 6 cm for grassland) and are, for privacy reasons, made available as 3-year averages for each possible combination of municipality (n=308) and agricultural region (n=7). SOC contents were determined by a modified Walkley and Black procedure (Maes et al., 2012) and a factor 1.14 was used to correct for incomplete oxidation (Lettens et al., 2005). For forested land, the ForSite database is available, containing descriptive and analytical data of 303 soil profiles sampled by horizon up to a depth of 100 to 150 cm throughout the 153,544 ha of forests (De Vos, 2009). SOC contents of the ForSite database were determined either by loss-on-ignition, Walkley and Black wet oxidation or total carbon analysis and converted to total SOC contents using locally calibrated regression equations (De Vos et al., 2005b, 2007). Low-Input High-Diversity systems such as ecologically managed grasslands in nature reserves receive limited inputs (e.g. fertilisation) and house a high plant diversity. For these LIHD systems (30,556 ha), 139 topsoil measurements (upper 15 cm) are available (Van Meerbeek *et al.*, 2014). As for arable land and grasslands, SOC contents for LIHD systems were determined by a modified Walkley and Black procedure (Maes et al., 2012) and a factor 1.14 was used to correct for incomplete oxidation (Lettens et al., 2005).

Besides sampling depth, spatial density and SOC determination procedure, also the time of sampling differs: arable land and grassland has been sampled between 2008 and 2011, forests between 1997 and 2002 and LIHD systems between 2009 and 2011. Furthermore, only the forest dataset contains measured bulk density data. For the remaining datasets, bulk density was estimated using the pedotransfer function of Rawls (1983) with locally calibrated values of the bulk density of the mineral fraction (Boon, 1984).

As the local study area comprised parts of two communes and was situated within one agricultural region for both arable land and grassland two aggregated topsoil SOC values were available. None of the forest soil profile descriptions was situated within the study area. For LIHD systems, three topsoil measurements were taken in the southern part of the valley.

Local data

In 2013, 31 parcels in the local study area were sampled (Figure 5d): 4 under arable land, 12 under grassland, 9 under forest and 6 under LIHD systems (Vanierschot, 2014) with, on average, a spatial sampling density of 2.3 augerings per km². Soils were sampled to a depth of 100 cm by fixed depth intervals: 0 – 10, 10 – 20, 20 – 40, 40 – 70 and 70 – 100 cm. Soil organic matter was determined by loss-on-ignition (LOI) and converted

to SOC content using a fixed correction factor of 0.58 (De Leenheer *et al.*, 1957). Bulk density of each depth interval was measured using Kopecky rings of 100 cc.

Regional SOC stocks

The characteristics of the datasets listed in Table 1 show their heterogeneity in terms of sampling depth, spatial density and level of aggregation. To derive region-wide SOC stock estimates of the upper 100 cm, in earlier research we developed, tested and applied (i) an exponential change decline function to extrapolate topsoil SOC measurements under agricultural land downwards (Ottoy *et al.*, 2016), (ii) boosted regression tree models to spatially upscale forest SOC stocks (Ottoy *et al.*, 2017b) and (iii) a combination of these techniques to estimate stocks under LIHD systems (Ottoy *et al.*, 2017a). The exponential change decline function has been developed as an alternative to the more commonly applied exponential decline function (Ottoy *et al.*, 2016). By assuming that not the SOC content, but rather its change over time declines exponentially with depth, this function allows is able to accounting for subsurface horizons with a larger SOC content. Boosted regression trees (BRT) was selected as the most appropriate digital soil mapping technique to estimate regional SOC stocks under forest and LIHD systems after comparison with multiple linear regression, artificial neural networks and least-squares support vector machines. The BRT-technique aims to improve the performance of a single model tree by fitting many model trees and combining (boosting) them for prediction (Elith *et al.*, 2008). Not only did this non-linear modelling technique result in the best fit, it also provided insights into the soil system by showing average predictor effects in partial dependence plots. In each of these three assessments (for agricultural land, forest and LIHD-systems), land units were defined by reclassification of the digitised Belgian soil map units (scale of publication 1:20,000) according to the third edition World Reference Base (WRB) classification system (Dondeyne *et al.*, 2014) and its subsequent topological overlay with the ECOPLAN land cover dataset with a resolution of 5 m × 5 m (Vrebos, 2015). The classes of the latter were aggregated into five land use types: arable land, grassland, forest, LIHD system and other land use. In this study, we spatially merged the resulting per land use type assessments to obtain the total regional SOC stock estimate.

Local SOC stocks

The SOC stock derived from each of the 31 profiles was first compared with the regionally estimated SOC stock for the land-unit in which the profile is located. This procedure is also referred to as geomatching (Lettens *et al.*, 2004). For the 31 data pairs, goodness-of-fit indicators including the coefficient of determination (R^2), root mean squared error (RMSE) and relative RMSE values were calculated.

Next, the derived SOC stocks of the 31 profiles were spatially up-scaled through a multi-level spatial generalisation approach similar to the ones developed by Ottoy *et al.* (2015) and Ottoy *et al.* (2017a). To this end, land units were defined in analogy with the regional assessment and a set of five hierarchical levels of generalisation was defined (Table 2). The definition of a land unit was increasingly generalised according to these levels, up to the point where a predefined number of profiles could be matched. The highest possible level of generalisation is obtained when the profile (i) is associated with the same land use type, and (ii) has the same morphogenetic soil classification as the land unit. Lower level matches are obtained by semantic generalisation of the soil classification or by not considering the land use type. In line with the results of Ottoy *et al.* (2015), for each land unit the highest possible level with minimum one matching profile was selected. Finally, the estimated local stocks were compared with the stocks derived by clipping the study area from the regional geodataset.

Results

Regional SOC stocks

Combining the different datasets, the total SOC stock in the upper 100 cm of unsealed soil in the region of Flanders (887,745 ha) is estimated at 111.67 Mt OC, or $12.58 \pm 5.65 \text{ kg m}^{-2}$ on average (Table 3, Figure 2). A rasterised version (10 x 10 m) of this assessment is available on the ECOPLAN repository (http://www.ecosysteemdiensten.be/cms/nl/indicator/09_3_2-0). Areas with larger SOC stocks tend to correspond to soils under forest and in alluvial valleys. Soils under more natural land use types, i.e. forest ($17.58 \pm 5.71 \text{ kg OC m}^{-2}$) and LIHD systems ($13.72 \pm 6.92 \text{ kg OC m}^{-2}$), store on average more organic carbon than soils under arable land ($10.40 \pm 4.06 \text{ kg OC m}^{-2}$) or under grassland ($12.60 \pm 5.54 \text{ kg OC m}^{-2}$). In absolute terms, arable land holds the largest SOC pool (36.21 % of the total stock) given its large area fraction (43.77 % or 388,572 ha out of 887,745 ha).

Local SOC stocks

Comparing the SOC stock data of the 31 profiles with the SOC stocks of the geomatched land units from the regional estimation (Table 4 and Figure 3) shows large differences between the two methods for SOC stock under grassland and under LIHD systems, resulting in small R^2 and large (r)RMSE values. Especially for two Histosols under LIHD vegetation, the local measurements (79 resp. 125.8 kg OC m⁻²) were largely underestimated in the regional assessment (14.6 resp. 51.1 kg OC m⁻²). On the other hand, the stock measured in three Histic Gleysols under LIHD systems (26.1 kg OC m⁻² on average), was smaller than the regional estimate (44.0 kg OC m⁻² on average). For the four profiles under arable land, a small R^2 value (0.01) but also small (r)RMSE values (5.27 kg m⁻², 29.26%) were found. The SOC stock of the 9 forested land units fitted the local measurements fairly well, resulting in a R^2 value of 0.73 and an rRMSE of 37.99 %. Given the small sample sizes (e.g. n=4 for arable land) caution is required when interpreting these goodness-of-fit measures.

To obtain the total SOC stock of the study area from the local measurements, 188 land units (encompassing 1539 polygons) were defined: 53 under arable land, 47 under forest, 69 under grassland and 19 under LIHD systems. As a result, 24 land units were matched with a soil profile at the reference level of generalisation: 4 under arable, 6 under forest, 12 under grassland and 2 under LIHD systems (Figure 4). The majority of land units (86) were assigned a SOC value at the third level of generalisation, i.e. after simplification of the soil classification to the soil texture class. For 8 land units classified as built-up soil, no matching profiles were found.

Both the regional and local methods of estimations resulted in similar estimates for arable land units: 45 kt OC using regional data and 43.54 kt OC using local information (Table 5). Evaluating stocks using local measurements led to larger estimates for grassland (+4.57 kt OC or +6.52 %), forests (+25.20 kt OC or +37.99 %) and LIHD systems (+2.94 kt OC or +26.34 %) compared to using regional data. This resulted in a larger total stock estimate using local data (223.84 kt OC) than when regional data was used (192.60 kt OC). The spatial distribution of SOC stocks estimated by local data, Figure 5b, reflects an anthropogenic gradient in the northern part, expressed by larger stocks surrounding the village centres and smaller stocks in the northwest. The north-south hydrological gradient with smaller stocks in dry (forested with mainly *Pinus* spp.) historical land dunes in the northwest and larger stocks in peatlands in the southeast is visible from both estimates (Figure 5a and Figure 5b), albeit more pronounced when local data were used.

Discussion

Regional SOC stocks

Accurate carbon accounting requires the availability of detailed national or regional SOC stock estimates, which are in most cases derived from different sources of routinely available soil datasets. These datasets are typically heterogeneous in terms of sampling depth and spatial density (Lettens *et al.*, 2005) requiring standardisation prior to combining them. With the exponential change decline function it was possible to extrapolate topsoil SOC of agricultural soils without overlooking SOC-rich subsoil horizons such as Plaggic, Spodic and Histic horizons (Ottoy *et al.*, 2016). Taking into account the subsoil stock is, given its magnitude and stability, especially valuable in an ecosystem service context. The digital soil mapping approach demonstrated an overall good performance of boosted regression trees in modelling SOC stocks of forests and LIHD-systems as this technique resulted in the best fit and provided insights in the importance of the different predictors (Ottoy *et al.*, 2017a; b).

By standardisation of three different types of source datasets through depth extrapolation and digital soil mapping, this study is the first to come up with a region-wide assessment of the contemporary (1997 – 2011) rather than legacy SOC stock (1947 – 1974) in the upper 100 cm of the region of Flanders. Our estimate (111.67 Mt OC) is larger than the legacy estimates by Lettens *et al.* (2004), i.e. 99.78 Mt OC, Meersmans *et al.* (2008), i.e. 97.61 ± 1.05 Mt OC and Meersmans *et al.* (2009b), i.e. 103.19 ± 1.27 Mt OC. For the complete unsealed territory of Belgium (24,042 km²) encompassing not only the region of Flanders but also Wallonia and the Brussels Capital region, Lettens *et al.* (2005) estimated the total SOC stock in 2000 at 264 Mt, or 10.98 kg OC m⁻² on average. To obtain estimates of the upper 100 cm, Lettens *et al.* (2005) derived depth extrapolation factors from the legacy data. The larger overall average found in this study (12.58 ± 5.65 kg OC m⁻²) is likely to be the result of the explicit consideration of SOC-related soil characteristics, i.e. the ability of the exponential change decline function to model SOC-rich subsurface horizons and of the digital soil mapping approaches to take the soil type into account.

Local SOC stocks

Site-specific factors The particular nature of the study area is reflected in the local SOC stocks. Larger estimates (Table 5) compared to the regional averages (Table 3) can partly be explained by the presence of site-specific particularities. For arable land, the local average was 18.02 ± 3.15 kg OC m⁻² (using local data)

and 18.62 ± 3.13 kg OC m⁻² (using regional data), which is much larger than the overall regional average of
 Flanders found in this study (10.40 ± 4.06 kg OC m⁻²) and in earlier regional assessments by Sleutel et al.
 (2003) (7.8 kg OC m⁻²) and Meersmans et al. (2009a) (8.2 kg OC m⁻²). The observation of larger SOC
 contents in arable sandy soils is in line with the results of Springob et al. (2001) and can be explained by the
 accumulation of OC through illuviation (podzolisation) and historical land-use practices (plaggen
 management). For grassland, the stocks (22.17 ± 10.83 kg OC m⁻² using regional data and 23.62 ± 7.95 kg
 OC m⁻² using local data) are larger than the regional estimate (12.60 ± 5.54 kg OC m⁻²) and the averages for
 Flanders of 11.1 kg OC m⁻² found by Meersmans et al. (2009a) and 14.3 kg OC m⁻² found by Mestdagh et
 al. (2009). This difference can be explained by the occurrence of grasslands in the valley, mainly on Gleyic
 Podzols (Figure 5d). The presence of SOC-rich peat layers in soils under forest (19.06 ± 8.91 kg OC m⁻²
 using regional data and 26.29 ± 25.42 kg OC m⁻² using local data) and LIHD systems (23.01 ± 12.15 kg OC
 m⁻² using regional data and 29.06 ± 15.93 kg OC m⁻² using local data) in the valley goes together with larger
 average stocks than the regional estimate (17.58 ± 5.71 kg OC m⁻² resp. 13.72 ± 6.92 kg OC m⁻²).
 The weak performance of the regional estimate to estimate SOC hotspot characteristics such as the presence
 of peat (Figure 3) resulted in smaller estimates under forest and LIHD systems using regional data ($19.06 \pm$
 8.91 kg OC m⁻² resp. 23.01 ± 12.15 kg OC m⁻²) compared to the stock obtained using local data ($26.29 \pm$
 25.42 kg OC m⁻² resp. 29.06 ± 15.93 kg OC m⁻²). For forest soil stocks, regional estimates of SOC in
 Histosols were based on only one soil profile description (Ottoy et al., 2017a) which is a weak basis for both
 regional and local SOC assessments. As this soil group is commonly underrepresented in profile collections
 and soil maps (e.g. Baritz et al., 2010), additional surveys are needed to enable more accurate estimations.
 Similarly, reference profile descriptions of Histosols under LIHD systems are lacking at the regional level.
 Furthermore, the regional assessment pointed to drawbacks from the use of potentially outdated information
 on the soil classification derived from the soil map (1947 – 1974), especially for Histosols (Ottoy *et al.*,
 2017a). Associated map impurities can impede comparison of geomatching results because profiles may be
 assigned to land units with a different soil and land use type (see e.g. Lettens et al., 2004).
 When upscaling, however, regional data can offer solutions at the local level. This is especially the case in
 data-scarce situations in which land units need to be highly generalised in order to find a matching profile.
 The soil classification of land units characterised by a level of generalisation larger than three was simplified
 by only taking into account soil textural classes. In some cases, this simplification can lead to over or

underestimations of the SOC stock (Ottoy *et al.*, 2015). In this study, the ‘Sfp(v)’ soil type (Histic Umbrisol) under LIHD vegetation could only be matched with a local profile at the third level of generalisation. As this simplification did not impose the presence of peat, the resulting SOC stock using local data (26.13 kg OC m⁻²) was lower than the estimate found using regional data (33.31 kg OC m⁻²).

Discrepancy in SOC stock determination procedures In addition to site-specific factors, the use of different SOC stock determination procedures can lead to differences between the assessments. The use of estimated rather than measured bulk densities (BD) can result in incorrect BD and SOC stock estimates when these estimation functions are applied outside their calibration range (e.g. De Vos *et al.*, 2005a). For example, Liebens and Van Molle (2003) found that the use of different bulk density estimation methods led to a variation of up to 7.5 % in BD estimates and 6% in the total SOC stock estimate in Flanders. In our case, only the forest soil dataset contained measured bulk densities and the pedotransfer function of Rawls (1983) was used in the other assessments (Table 1). As for each interval of the 31 profiles, bulk density was measured, these values were used to test the performance of the PTF of Rawls (1983). This resulted in an overall R² of 0.79, mean error (ME) of -0.11 t m⁻³ and root mean squared error (RMSE) of 0.23 t m⁻³. The negative mean error indicates the global underestimation of bulk density and hence of SOC stock by the PTF of Rawls. The values per land use type (Table 6) show that larger ME values were found for arable and grassland soil depth layers, whereas soil layers under forest and LIHD systems had larger RMSE values. Underestimation of the bulk density can thus partly explain the smaller, underestimated SOC stock estimates derived by using regional data not encompassing BD measurements.

Also the procedure used to derive the SOC content can have an impact on the resulting SOC stocks. For example, incorrect determination of SOC contents through incomplete oxidation of OC using the Walkley and Black (1934) procedure or additional losses of structural water from clay minerals using the loss-on-ignition procedure lead to under- or overestimations of the SOC stock (De Vos *et al.*, 2005b, 2007; Meersmans *et al.*, 2009a). To solve these issues, the use of correction factors or regression equations converting the resulting SOC content into actual SOC content is recommended. These factors are dependent upon land use and soil type (De Vos *et al.*, 2005b, 2007; Meersmans *et al.*, 2009a).

Implications for soil inventories and SOC assessments

For arable land, and to a lesser extent also for grassland, the regional SOC estimate was able to reflect local phenomena such as anthropogenic and hydrologic gradients, resulting in similar average and total SOC stocks. For forests and LIHD systems, however, the discrepancy between the regional and local assessment was relatively large. This was especially the case for forests and LIHD systems on Histosols which play an import role in SOC storage. In our study area, this soil group covered 3.41 % of the area, but contained 6.94 % (calculated using regional data) to 14.90 % (using local data) of the total SOC stock. Given the low availability of data on Histosols in regional assessments, inventories focusing on these carbon hotspots are needed. These surveys could benefit from recently developed remote sensing techniques, e.g. peat mapping (Aitkenhead, 2017). Furthermore, quantitative estimates of the various sources of uncertainty present at different scales (see e.g. Goidts et al. (2009) and Vanguelova et al. (2016)) would serve in-depth comparison between assessments. Several methods exist to estimate uncertainty, e.g. through stochastic Monte-Carlo simulations (Chartin *et al.*, 2017) or decision trees (Stumpf *et al.*, 2017). Such measures can be used to guide and optimize soil sampling (Stumpf *et al.*, 2017).

With the presented research a top-down approach was evaluated, i.e. how well is the local SOC stock estimated when data collected for regional purposes are used. Accuracy improvements of both local and regional estimates might be gained by reverse thinking: a bottom-up approach incorporating a diverse set of local field campaigns. The main requirement for such an approach is the availability of a central data platform on which local inventories can be gathered. The OPPLA web platform (www.oppla.eu), containing a database of case studies on ecosystem services, natural capital and nature-based solutions, can be used as an example. Collecting appropriate metadata in a standardised form is necessary for acknowledging the heterogeneity in data collection and selecting the appropriate harmonisation techniques. Furthermore, the availability of these datasets can support the identification of data-scarce land units, on which future soil inventories should focus.

Conclusion

By combining four land use type specific SOC datasets, the total stock in the upper 100 cm of soil in the 887,745 ha region of Flanders was estimated at 111.67 Mt OC: 40.44 Mt OC under arable, 39.70 Mt OC under grassland, 26.99 Mt OC under forest and 4.53 Mt OC under LIHD systems. Soils under more natural land use types including forests (17.58 ± 5.71 kg m⁻²) and LIHD systems (13.72 ± 6.92 kg m⁻²) on average stored more organic carbon than agricultural soils: arable (10.40 ± 4.06 kg m⁻²) and grassland (12.60 ± 5.54

kg m⁻²). Given their large share in area, however, the latter contain the largest absolute stock. Zooming in on a sandy study area marked the regional assessment's poor performance for identifying carbon hotspots, such as Histosols resulting in overall SOC stock underestimation. Regional assessments must be complemented by local surveys to capture site-specific patterns. Similarly, local assessments can benefit from regional data to accommodate missing data.

Acknowledgements

We thank the anonymous reviewers for their constructive comments and suggestions, Elias for his logistical support, and Bert Veris and his team for their support in the Histosol sampling. The research was supported by the IWT SBO Project 120014 "EcoPlan" funded by the Agency for Innovation by Science and Technology (IWT).

330 References

- 331 Aitkenhead, M.J. 2017. Mapping peat in Scotland with remote sensing and site characteristics. *European Journal*
332 *of Soil Science*, **68**, 28–38.
- 333 Baritz, R., Seufert, G., Montanarella, L. & Van Ranst, E. 2010. Carbon concentrations and stocks in forest soils of
334 Europe. *Forest Ecology and Management*, **260**, 262–277.
- 335 Boon, W. 1984. *Onderzoek naar het verband tussen het koolstofpercentage en de volumedichtheid van de grond.*
336 *Intern Rapport Bodemkundige Dienst België*, Heverlee.
- 337 Broekx, S., De Nocker, L., Liekens, I., Poelmans, L., Staes, J., Van Der Biest, K. & Verheyen, K. 2013. *Raming*
338 *van de baten geleverd door het Vlaamse NATURA 2000-netwerk*.
- 339 Chartin, C., Stevens, A., Goidts, E., Krüger, I., Carnol, M. & van Wesemael, B. 2017. Mapping Soil Organic
340 Carbon stocks and estimating uncertainties at the regional scale following a legacy sampling strategy
341 (Southern Belgium, Wallonia). *Geoderma Regional*, **9**, 73–86.
- 342 De Leenheer, L., Van Hove, J. & Van Ruymbeke, M. 1957. Détermination quantitative de la matière organique du
343 sol. *Pédologie*, **7**, 324–347.
- 344 De Vos, B. 2009. *Uncertainties of forest soil carbon stock assessment in Flanders (Unpublished doctoral thesis).*
345 *KU Leuven, Leuven, Belgium*.
- 346 De Vos, B., Lettens, S., Muys, B. & Deckers, J.A. 2007. Walkley–Black analysis of forest soil organic carbon:
347 recovery, limitations and uncertainty. *Soil Use and Management*, **23**, 221–229.
- 348 De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J. & Muys, B. 2005a. Predictive quality of pedotransfer
349 functions for estimating bulk density of forest soils. *Soil Science Society of America Journal*, **69**, 500–510.
- 350 De Vos, B., Vandecasteele, B., Deckers, J. & Muys, B. 2005b. Capability of Loss-on-Ignition as a Predictor of
351 Total Organic Carbon in Non-Calcareous Forest Soils. *Communications in Soil Science and Plant Analysis*,
352 **36**, 2899–2921.
- 353 Dondeyne, S., Vanierschot, L., Langohr, R., Van Ranst, E. & Deckers, J. 2014. *The soil map of the Flemish region*
354 *converted to the 3rd edition of the World Reference Base for soil resources*. KU Leuven, Universiteit Gent
355 & Departement Leefmilieu, Natuur & Energie van de Vlaamse Overheid, Brussels.
- 356 Elith, J., Leathwick, J.R. & Hastie, T. 2008. A working guide to boosted regression trees. *The Journal of animal*
357 *ecology*, **77**, 802–813.
- 358 Goidts, E., van Wesemael, B. & Crucifix, M. 2009. Magnitude and sources of uncertainties in soil organic carbon
359 (SOC) stock assessments at various scales. *European Journal of Soil Science*, **60**, 723–739.
- 360 IPCC. 2006. *IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse*
361 *Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds)* (HS
362 Eggleston, L Buendia, K Miwa, T Ngara, and K Tanabe, Eds.). IGES, Japan.
- 363 Lettens, S., Van Orshoven, J., van Wesemael, B. & Muys, B. 2004. Soil organic and inorganic carbon contents of
364 landscape units in Belgium derived using data from 1950 to 1970. *Soil Use and Management*, **20**, 40–47.
- 365 Lettens, S., Van Orshoven, J., van Wesemael, B., De Vos, B. & Muys, B. 2005. Stocks and fluxes of soil organic
366 carbon for landscape units in Belgium derived from heterogeneous data sets for 1990 and 2000. *Geoderma*,
367 **127**, 11–23.
- 368 Liebens, J. & Van Molle, M. 2003. Influence of estimation procedure on soil organic carbon stock assessment in
369 Flanders, Belgium. *Soil Use and Management*, **19**, 364–371.
- 370 Maes, S., Elsen, A., Tits, M., Boon, W., Deckers, S., Bries, J., Vogels, N. & Vandendriessche, H. 2012. *Wegwijs*
371 *in de bodemvruchtbaarheid van de Belgische akkerbouw- en weilandpercelen (2008-2011)*. Bodemkundige
372 Dienst van België, Heverlee, Belgium.
- 373 Meersmans, J., De Ridder, F., Canters, F., De Baets, S. & Van Molle, M. 2008. A multiple regression approach to
374 assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium).
375 *Geoderma*, **143**, 1–13.
- 376 Meersmans, J., van Wesemael, B. & Van Molle, M. 2009a. Determining soil organic carbon for agricultural soils:
377 a comparison between the Walkley & Black and the dry combustion methods (north Belgium). *Soil Use and*
378 *Management*, **25**, 346–353.
- 379 Meersmans, J., van Wesemael, B., De Ridder, F., Dotti, M.F., De Baets, S. & Van Molle, M. 2009b. Changes in
380 organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. *Global Change*
381 *Biology*, **15**, 2739–2750.
- 382 Meersmans, J., van Wesemael, B., De Ridder, F. & Van Molle, M. 2009c. Modelling the three-dimensional spatial
383 distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, **152**, 43–52.
- 384 Mestdagh, I., Sleutel, S., Lootens, P., Van Cleemput, O., Beheydt, D., Boeckx, P., De Neve, S., Hofman, G., Van
385 Camp, N., Vande Walle, I., Samson, R., Verheyen, K., Lemeur, R. & Carlier, L. 2009. Soil organic carbon-
386 stock changes in Flemish grassland soils from 1990 to 2000. *Journal of Plant Nutrition and Soil Science*,
387 **172**, 24–31.
- 388 Minasny, B., McBratney, A.B., Malone, B.P. & Wheeler, I. 2013. Digital Mapping of Soil Carbon. *Advances in*

- Agronomy*, **118**, 1–47.
- O'Rourke, S.M., Angers, D.A., Holden, N.M. & Mcbratney, A.B. 2015. Soil organic carbon across scales. *Global Change Biology*, **21**, 3561–3574.
- Ottoy, S., Beckers, V., Jacxsens, P., Hermy, M. & Van Orshoven, J. 2015. Multi-level statistical soil profiles for assessing regional soil organic carbon stocks. *Geoderma*, **253–254**, 12–20.
- Ottoy, S., Elsen, A., Van De Vreken, P., Gobin, A., Merckx, R., Hermy, M. & Van Orshoven, J. 2016. An exponential change decline function to estimate soil organic carbon stocks and their changes from topsoil measurements. *European Journal of Soil Science*, **67**, 816–826.
- Ottoy, S., Van Meerbeek, K., Sindayihebura, A., Hermy, M. & Van Orshoven, J. 2017a. Assessing top- and subsoil organic carbon stocks of Low-Input High-Diversity systems using soil and vegetation characteristics. *Science of The Total Environment*, **589**, 153–164.
- Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M. & Van Orshoven, J. 2017b. Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological Indicators*, **77**, 139–150.
- Parras-Alcántara, L., Díaz-Jaimes, L. & Lozano-García, B. 2015. Management effects on soil organic carbon stock in Mediterranean open rangelands-treeless grasslands. *Land Degradation & Development*, **26**, 22–34.
- Peel, M.C., Finlayson, B.L. & McMahon, T.A. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, **11**, 1633–1644.
- Rawls, W.J. 1983. Estimating soil bulk density from particle size analysis and organic matter content. *Soil Science*, **135**, 123–125.
- Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A., Üveges, J.B., Freudenschuß, A. & Simota, C. 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology*, **14**, 2432–2442.
- Schneiders, A. & Müller, F. 2017. A natural base for ecosystem services. In: *Mapping ecosystem services* (eds. Burkhard, B. & Maes, J.), pp. 35–40. Pensoft Publishers, Sofia, Bulgaria.
- Sleutel, S., de Neve, S. & Hofman, G. 2003. Estimates of carbon stock changes in Belgian cropland. *Soil Use and Management*, **19**, 166–171.
- Springob, G., Brinkmann, S., Engel, N., Kirchmann, H. & Böttcher, J. 2001. Organic C levels of Ap horizons in North German Pleistocene sands as influenced by climate, texture, and history of land-use. *Journal of Plant Nutrition and Soil Science*, **164**, 681–690.
- Stevens, F., Bogaert, P., Van Oost, K., Doetterl, S. & van Wesemael, B. 2014. Regional-scale characterization of the geomorphic control of the spatial distribution of soil organic carbon in cropland. *European Journal of Soil Science*, **65**, 539–552.
- Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., Xiang, W. & Scholten, T. 2017. Uncertainty-guided sampling to improve digital soil maps. *Catena*, **153**, 30–38, (At: <http://dx.doi.org/10.1016/j.catena.2017.01.033>).
- Vanguelova, E.I., Bonifacio, E., De Vos, B., Hoosbeek, M.R., Berger, T.W., Vesterdal, L., Armolaitis, K., Celi, L., Dinca, L., Kjønaas, O.J., Pavlenda, P., Pumpanen, J., Püttsepp, Ü., Reidy, B., Simončič, P., Tobin, B. & Zhiyanski, M. 2016. Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales - review and recommendations. *Environmental Monitoring and Assessment*, **188**, 630.
- Vanierschot, L. 2014. *Bodemorganische koolstof in het landschap: relatie met natuurlijke en antropogene gradiënten in de Kempen (Unpublished Master's thesis)*. KU Leuven, Leuven, Belgium.
- Van Meerbeek, K., Van Beek, J., Bellings, L., Aertsen, W., Muys, B. & Hermy, M. 2014. Quantification and Prediction of Biomass Yield of Temperate Low-Input High-Diversity Ecosystems. *Bioenergy Research*, **7**, 1120–1130.
- Van Orshoven, J., Deckers, J.A., Vandenbroucke, D. & Feyen, J. 1993. The completed database of Belgian soil profile data and its applicability in the planning and management of rural land. *Bulletin des Recherches Agronomiques de Gembloux*, **28**, 197–222.
- Vrebos, D. 2015. Bodembedekking. (At: http://www.ecosysteemdiensten.be/cms/indicator/bl_lc).
- Vrebos, D., Bampa, F., Creamer, R., Gardi, C., Ghaley, B., Jones, A., Rutgers, M., Sandén, T., Staes, J. & Meire, P. 2017. The Impact of Policy Instruments on Soil Multifunctionality in the European Union. *Sustainability*, **9**, 407.
- Walkley, A. & Black, I.A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, **37**, 29–38.

445
 446 Figure 1: Geographical distribution of the regional datasets. For forests and LIHD systems, point data were
 447 available. For arable land and grasslands, data were made available as aggregates for each possible
 448 combination of municipality and agricultural region. The boundary of the local study area is marked by a red
 449 rectangle.

450

451 Figure 2: Estimated distribution of the SOC stock (kg m^{-2}) in the upper 100 cm of soil in Flanders, Belgium.

452 The boundary of the local study area is marked by a yellow rectangle.

Figure 3: Comparison of the SOC stock of the 31 local profiles (X-axis) with the SOC stock of the land unit from the regional assessment in which the profile is situated (Y-axis).

Figure 4: Distribution of the land units over the different levels of generalisation in the multi-level spatial generalisation approach using local data.

(a) SOC stock estimated using regional data (kg OC m⁻²) (b) SOC stock estimated using local data (kg OC m⁻²)

(c) Difference in SOC stock (kg OC m⁻²): (b) - (a)

(d) Land use type and augerings

Figure 5: Geographical distribution of the estimated SOC stock (kg OC m⁻²) using (a) local data, and (b) regional data. The difference in stock between (a) and (b) is given in (c). (d) shows the distribution of land use types and plots. The background map is the OpenStreetMap (www.openstreetmap.org).

463 **Tables**

464 Table 1: Overview of the different regional data sources and their characteristics.

	Arable land	Grassland	Forest	LIHD systems
Database	Soil fertility data		ForSite	-
Data owner	Soil service of Belgium		Research institute for Nature and Forest	KU Leuven
Inventory years	2008 – 2011		1997 – 2002	2009 – 2011
Number of samples	80,404	13,938	303	139
Sampling by horizon	No		Yes	No
Sampling depth (cm)	23	6	100/150	15
Aggregation level	3-year averages per municipality and agricultural region		Individual profiles	Individual measurements
Bulk density	Estimated using the PTF of Rawls (1983) with locally calibrated values of the bulk density of the mineral fraction (Boon, 1984)		Measured	Estimated using the PTF of Rawls (1983) with locally calibrated values of the bulk density of the mineral fraction (Boon, 1984)
OC determination method	Modified Walkley and Black (Maes et al., 2012)		Loss-on-ignition, Walkley and Black wet oxidation or Total Carbon analysis	Modified Walkley and Black (Maes et al., 2012)
Conversion of measured to total OC	Correction factor 1.14 (Lettens et al., 2005)		Locally calibrated regression equations (De Vos et al., 2005b, 2007)	Correction factor 1.14 (Lettens et al., 2005)
Regional assessment	Ottoy et al. (2016) using depth extrapolation		Ottoy et al. (2017a) using digital soil mapping	Ottoy et al. (2017b) using depth extrapolation and digital soil mapping

465

Table 2: Levels of generalisation and their corresponding assumptions used in the multi-level spatial generalisation approach to estimate local SOC stocks.

Level	Assumptions to assign a profile to a land unit	Example
Reference	Profile is associated with the same land use type AND Profile is of exactly the same morphogenetic soil classification as the land unit	s-Pfp(v) & Forest
1	Profile is associated with the same land use type AND Profile has the same substratum-, texture-, drainage- and profile development class as the land unit	s-Pfp & Forest
2	Profile is associated with the same land use type AND Profile has the same texture-, drainage- and profile development class as the land unit	Pfp & Forest
3	Profile is associated with the same land use type AND Profile has the same texture class as the land unit	P & Forest
4	Profile has the same texture class as the land unit	P

s: substratum – sand; P: texture – light sandy loam; f: drainage class – very wet with reduction horizon, p: profile development class – without profile development; (v): variant of profile development – presence of peat.

471 Table 3: Overview of the different land use types, their area and SOC stock estimate at the regional scale.

Land use type	Area (ha)	Total regional SOC stock (Mt OC)	Average regional SOC stock $\pm$ standard deviation (kg m ⁻²)
Arable land	388,572	40.45	10.40 $\pm$ 4.06
Grassland	315,074	39.70	12.60 $\pm$ 5.54
Forest	153,544	26.99	17.58 $\pm$ 5.71
LIHD systems	30,556	4.53	13.72 $\pm$ 6.92
Total	887,745	111.67	12.58 $\pm$ 5.65

472

473 Table 4: Goodness-of-fit measurements of the regional maps to estimate the SOC stock of the 31 profiles.

Land use type	n	R ²	RMSE (kg m ⁻²)	rRMSE (%)
Arable land	4	0.01	5.27	29.26
Grassland	12	0.05	12.61	69.35
Forest	9	0.73	8.22	37.99
LIHD systems	6	0.12	49.41	96.97
Total	31	0.05	23.60	92.55

474

475 Table 5: Total (kt OC) and average (kg m⁻²) SOC stocks in the valley of the Kleine Nete estimated by data
 476 derived at the regional and local scale.

Land use type	Area (ha)	SOC stock using regional data		SOC stock using local data	
		Total (kt OC)	Average ± standard deviation (kg m ⁻²)	Total (kt OC)	Average ± standard deviation (kg m ⁻²)
Arable land	241.59	45.00	18.62 ± 3.13	43.54	18.02 ± 3.15
Grassland	316.16	70.09	22.17 ± 10.83	74.66	23.62 ± 7.95
Forest	348.16	66.34	19.06 ± 8.91	91.54	26.29 ± 25.42
LIHD systems	48.52	11.16	23.01 ± 12.15	14.10	29.06 ± 15.93
Total	954.44	192.60	20.18 ± 8.98	223.84	23.45 ± 16.86

477

478 Table 6: Measured and estimated bulk densities of depth intervals per land use type, together with mean error
479 (ME) and root mean squared error (RMSE) values.

	<i>n</i>	Measured BD (t m ⁻³)	Estimated BD (t m ⁻³)	ME (t m ⁻³)	RMSE (t m ⁻³)
Arable					
0–10 cm	4	1.39	1.24	-0.15	0.17
10–20cm	4	1.41	1.22	-0.18	0.21
20–40 cm	4	1.50	1.27	-0.23	0.24
40–70 cm	4	1.55	1.42	-0.13	0.15
70–100 cm	4	1.56	1.45	-0.11	0.12
Forest					
0–10 cm	9	1.05	0.96	-0.09	0.21
10–20 cm	9	1.13	1.12	0.00	0.19
20–40 cm	9	1.33	1.25	-0.07	0.22
40–70 cm	9	1.42	1.42	0.00	0.28
70–100 cm	9	1.54	1.45	-0.09	0.28
Grassland					
0–10 cm	13	1.34	1.12	-0.22	0.24
10–20 cm	13	1.36	1.21	-0.15	0.20
20–40 cm	13	1.48	1.30	-0.18	0.23
40–70 cm	13	1.55	1.42	-0.13	0.15
70–100 cm	13	1.56	1.41	-0.15	0.18
LIHD system					
0–10 cm	5	0.33	0.33	0.01	0.26
10–20 cm	5	0.45	0.47	0.02	0.27
20–40 cm	5	0.88	0.82	-0.05	0.47
40–70 cm	5	1.09	0.99	-0.10	0.29
70–100 cm	5	1.08	1.02	-0.06	0.23
All	155	1.30	1.19	-0.11	0.23

480