Can soil properties and land use explain glomalin-related soil protein (GRSP) accumulation? A nationwide survey in France

Siobhan Staunton, Nicolas P.A. Saby, Dominique D. Arrouays, Herve Quiquampoix

Can soil properties and land use explain glomalin-related soil protein (GRSP) accumulation? A nationwide survey in France

Siobhan Staunton[1], Nicolas P.A. Saby[2], Dominique Arrouays[2], & Hervé Quiquampoix[1]

[1]Eco&Sols, INRAE, IRD, Cirad, SupAgro, University of Montpellier, place Viala, 34060 Montpellier, France

[2]INRAE, Unité Infosol US1106, 45075 Orléans 2, France

Corresponding author: Siobhan Staunton

Phone +33 (0) 4 99 61 23 31. E-mail: siobhan.staunton@inrae.fr

Abstract

Organic matter plays essential roles in soil, including physical stabilization, nutrient storage and carbon sequestration. An operationally defined fraction of soil organic matter known as glomalin or glomalin-related soil protein, GRSP, is obtained by autoclaving soil in citrate solution. It is reputed to be of fungal origin, very stable and responsible for enhanced soil physical stability. This is the first nationwide survey of GRSP content in soil and was conducted to test the hypotheses on the origin and identity of GRSP. Nearly 200 archived soils were selected on the basis of organic matter content from the French National Soil Inventory, representing mostly cropland, grassland and woodland land uses. Two fractions of GRSP were measured, easily extractable ($GRSP_{EE}$) and total ($GRSP_T$), extraction. The median values of $GRSP_{EE}$ and $GRSP_T$ were 0.7 and 2 g kg^{-1} respectively, with a strong correlation between the two fractions. Scatter plots and cubist modelling were used to explore the relationships between the contents of the two GRSP fractions and both soil properties and land use. Land use effects were almost entirely attributable to soil characteristics. No evidence was found to support the hypotheses that GRSP is solely of fungal origin, nor that easily extracted GRSP is more recent that the total fraction, although this does not disprove either hypothesis. Soil organic matter was enriched in GRSP C-depleted cropland soils and lower in C-rich woodland, this may result from inherent stability or differences in plant-related composition of carbon input.

Keywords : soil, glomalin, protein, autoclaved citrate-extractable protein, carbon sequestration, carbon dynamics, fungal tracer

Highlights

- Land use effect of GRSP content was indirect and due to soil characteristics
- Fungal origin of GRSP could not be confirmed
- Correlation between $GRSP_{EE}$ and $GRSP_T$ did not support claim of contrasting ages
- Organic matter was enriched in GRSP as soils were depleted in C

1. Introduction

Soil organic matter, SOM, has many important functions including the supply of nutrients, stabilizing soil to prevent erosion and sequestering carbon to limit the build-up of green-house gases. Much attention has recently been given to the possible role of soil to sequester carbon, which requires a knowledge of the storage of soil carbon (Hernandez-Soriano et al., 2018; Houghton et al., 2012; Janzen, 2006; Janzen, 2015; King, 2011; Lal, 2004; Schmidt et al., 2011; Stockmann et al., 2013; Yu et al., 2017). A component of SOM, glomalin or Glomalin Related Soil Protein, GRSP, is claimed to be particularly stable and to convey great physical stability to soils, and has been a focus of research since its first report in 1996 by Wright and co-workers (Wright and Upadhyaya, 1996a). It is an operationally defined fraction, extracted by autoclaving in neutral or alkaline citrate solution to yield Easily Extractable (EE) or Total (T) fractions. The term autoclave citrate-extractable protein has recently been proposed to describe this fraction (Hurriso et al, 2018). The protein is quantified using either an antibody directed against fresh fungal hyphae (Wright et al., 1996; Wright and Upadhyaya, 1996a) or more commonly by the non-specific colorimetric assay developed by Bradford (Bradford, 1976). Other substances, operationally classed as humic substances, are co-extracted particularly at alkaline pH (Bolliger et al., 2008; Gillespie et al., 2011a; Gillespie et al., 2011b; Halvorson and Gonzalez, 2006; Schindler et al., 2007; Whiffen et al., 2007).

GRSP is proposed as an indicator of soil physical stability and of fungal activity and turnover (Bedini et al., 2009; Fokom et al., 2012; Gispert et al., 2013; Šarapatka et al., 2019; Young et al., 2012). A strong correlation is often observed between GRSP and organic carbon content, this along with the realization that the composition of GRSP is not as simple as was originally claimed, has led to doubt as to the identity of GRSP as a distinct component of soil organic matter. Nevertheless the seasonal variation of GRSP, its sensitivity to land use change, exhibiting better resistance to depletion than average organic matter, accredit the hypothesis of

GRSP as a distinct fraction (Emran et al., 2012; Gispert et al., 2013; Oliveira et al., 2016; Preger et al., 2007; Rillig et al., 2001). It is generally stated that the ease of extraction of the GRSP fractions reflects the period of contact between soil and protein (Koide and Peoples, 2013; Lovelock et al., 2004b) because the strength of association of protein with soil organo-mineral surfaces was expected to increase with time. The easily extractable fraction, obtained at neutral pH, is thus considered to be more recent that the more strongly fixed, total fraction, extracted at pH 8. GRSP is reputed to be a glycoprotein originating from the hyphal walls of arbuscular mycorrhizal fungi (Driver et al., 2005; Wright and Upadhyaya, 1996b).

The aim of this study was to assess the evidence to support some of the various claims made concerning the origin and identity of GRSP and the factors that determine its accumulation. An underlying question is whether or not GRSP is a distinct fraction of organic matter. We also consider the validity of the premise that GRSP is largely of fungal origin, and hence is a useful probe of fungal activity. If this were true then there should be a link between soil composition and land-use favourable towards fungal activity and the amount of GRSP. Finally, by comparing the ratio of the two operationally defined fractions of GRSP, $GRSP_{EE}$ and $GRSP_T$, we aim to determine if there is any reason to presume that they probe soil protein pools of different ages. To achieve these aims we have surveyed a large number (197) of top soils from metropolitan France and compared the GRSP contents to soil properties and land-use. The study was undertaken in the context of the suggestion that carbon sequestration in soil could be engineered, including phytoengineering, to mitigate climate change due to increasing levels of atmospheric greenhouse gases (Janzen, 2006; Janzen, 2015; Lal, 2004; Lorenz et al., 2007).

2 Materials & Methods

Soils had been sampled throughout mainland France as part of the National Soil Quality Monitoring Programme (RMQS) using a 16x16 km grid. Sampling strategy and soil analysis and conservation have been described in detail elsewhere (Arrouays et al., 2002; Meersmans et al., 2012). Soils were air-dried and sieved to < 200 µm. Ten percent of the top-soils (0-30 cm) in the collection were selected to be representative of the full set on the basis of soil organic carbon content (median, average and standard deviation of about 20 g kg^{-1}). Several physico-chemical parameters had been measured on each soil during the soil monitoring survey performed by the Soil Analysis Laboratory of INRA (Arras, France, http://www.lille.inra.fr/las). Data are available upon request, and a synthesis is given https://doi.org/10.15454/BNCXYB). Soil class was not included in statistical analysis because more than half the samples were Cambisols. Figure SI_1 shows the boxplots of relevant soil properties for the complete sample set and for each of the major land-use categories (cropland, grassland and woodland). We calculated the fractions of soil organic carbon and of clay that were complexed or uncomplexed using the relationship proposed by Dexter (Dexter et al., 2008b). This model assumes that organic matter is preferentially complexed with clay minerals in soil and so there is no non complexed organic matter until the saturation limit of clay minerals is reached. This limit is reported to be 10 g clay g^{-1} organic carbon organic matter in European soils.

Two operationally defined fractions of glomalin related soil protein, easily extractable and total protein were obtained. The easily extractable fraction was obtained by a single cycle of autoclaving of soil in 20 mM sodium citrate at pH 7. The total GRSP was obtained by a single cycle of autoclaving of soil in 50 mM sodium citrate at pH 8. Many studies employ a variable number of extractions in alkaline citrate which are pooled to obtain the total fraction, using the criterion of colour to decide the number of extractions. This was originally proposed by Wright and co-workers who assumed that the colour was due to the protein, glomalin, and not to humic substances (Wright and Upadhyaya, 1996b). Unpublished data have shown that extracted protein decreases exponentially with extraction rank, and so we chose to use a single extraction. Soil:solution ratio was always 1:8 (0.25 g + 2 ml). After cooling, soil and solution were separated by centrifugation at 15 000 *g* for 20 minutes. Bradford reactive protein was assayed in microplates, after dilution if necessary, using the Bradford Quick Start kit purchased from BioRad (20µl sample + 230 µl reagent). The absorbance was read at 590 nm and bovine serum albumin (0 – 200 mg dm^{-3}) used for calibration. Variation between triplicate extractions was about 7 % and average variation between triplicate protein assays was 5 %. Data are presented either as the Bradford reactive protein with respect to mass of air-dry soil, or as the ratio of GRSP to soil organic carbon content, OC. We prefer to report the ratio of GRSP:OC, rather than assuming a C content for GRSP, since the reported C contents are not typical of those of most proteins (53% (Rouwenhorst et al., 1991)). In previous studies a factor of 32% has been used to convert GRSP to GRSP-carbon, and other studies report that the C content of GRSP vary between samples (Lovelock et al., 2004b; Seguel et al., 2008; Woignier et al., 2014).

In addition to simple linear regression, we used a Cubist model using the dedicated R package. The advantage of the cubist approach, as with other machine learning approaches, is that no regression type is imposed on data. Regression using Cubist works with condition-based rules where the output is a set of rules, and each rule has a specific multivariate linear model. The algorithm was first described by Quinlan (Quinlan, 1992). A decision tree is constructed and a model is used to calculate the predicted value for each leaf (Minasny and McBratney, 2008). Unlike regression trees, which predict a rigid value for each 'leaf,' regression rules build a multivariate linear function. The Cubist model can also use a boosting–like scheme called committees where iterative model trees are created in sequence. To select the best iteration numbers, we used a k-fold cross validation on the dataset with 300 replications. We tested a set of values for the number of committees (1, 2, 5, 10, 15, 50 and 100). The following prediction quality criteria were calculated during the cross validation: Coefficient of determination (R^2) and Root mean square error (RMSE). The soil properties considered for these analyses were: organic carbon (OC), total nitrogen (N_T), texture (clay, silt and sand contents), pH (in water), total iron, exchangeable phosphorus, total carbonate, base saturation, exchangeable aluminium, extractable DNA and land use. The influence of different soil properties in predicting the ratio

$GRSP_{EE}$:$GRSP_T$ was also checked. The utility of each variable during the model building was derived. This is known as variable importance (VI) and is computed as the percentage of times where each variable was used in a condition and/or a linear model.

3. Results and Discussion

3.1 Relations between GRSP and soil properties

The median value of $GRSP_T$ was 2 g kg^{-1}, with the median value of $GRSP_{EE}$ about three times less at 0.7 g kg^{-1}. Figure 1 shows the box plots of $GRSP_{EE}$ and $GRSP_T$ for the full data set and classed by major land-use categories: cropland, grassland and woodland which accounted for more than 90% of the soils studied. The range of values of both $GRSP_{EE}$ and $GRSP_T$ are similar to previously reported values. For example Nichols & Wright (2005) found values of $GRSP_T$ in the range 1.6-4 g kg^{-1} (with up to 4 successive extractions pooled) for eight US soils with textures ranging from sandy loam to silty clay loam. With some seasonal variation, annual average values of $GRSP_T$ in the range 1-5 g kg^{-1} have been reported for sandy to sandy loam soils with different land uses in Spain, with $GRSP_{EE}$ being about a third less in the range 0.5-1.3 g kg^{-1} (Emran et al., 2012). Treseder & Turner (2007b) report much lower average values of immuno-reactive $GRSP_T$ for relevant biomes, agricultural and temperate grassland (0.5±0.1 and 0.7±0.3 g kg^{-1}), and greater values for temperate forests (2.5 g kg^{-1}) and even greater for tropical forests (7 g kg^{-1}). When immuno-reactive and Bradford reactive protein can be compared, the former is about three-fold smaller than the latter (Franzluebbers et al., 2000; Nichols and Wright, 2005), bringing the relevant data of Treseder and Turner in line with those presented in this study. The values reported in this study for woodland soils are somewhat lower than those previously reported for ten French mainland forest soils (3.5 – 10.5 g kg^{-1} for $GRSP_T$ and 0.7-2.1 g kg^{-1} for $GRSP_{EE}$) (Jorge-Araújo et al., 2015). The main reason for this difference is that the former study investigated forest soils with organic carbon contents of about 100 g kg^{-1} whereas the median value for forest soils in this study was considerably less, 28 g kg^{-1}.

Previous studies have pointed to a link between GRSP with some soil properties, especially soil organic carbon content, although with much smaller data sets (Rillig et al., 2003; Treseder and Turner, 2007a). Figure 2 shows that there are positive but non-linear relationships between each of the GRSP fractions and both OC and total N. There are approximately linear relations at lower values of OC and N_T, with a threshold value reached at about the 95th percentile. The relation with N_T is consistent with a large proportion of soil N being organic and that proteins typically contain about 16% N. There is no relationship with clay content, therefore there is no evidence that GRSP is stabilised by association with clay minerals, or that the measured GRSP content depended on extraction efficiency.

Simple regression analysis cannot take into account the complex interrelaton between land use and soil properties and so is of limited value for studies of a large number of soils with contrasting properties. In order to avoid bias in the choice of parameters we adopted the Cubist approach to identify the determinants of soil GRSP content. Correlations between observed and predicted values were calculated during the cross validation. They were significant for $GRSP_{EE}$ (r^2=0.54) and $GRSP_T$ (r^2=0.75) with root mean square prediction errors of 0.287 and 0.383 respectively (Table SI_1). The agreement between predicted and observed values showed little bias for $GRSP_T$, whereas lower values were overestimated for $GRSP_{EE}$.

Figure 3 shows the relative importance of the parameters identified by this approach to explain the measured values of $GRSP_{EE}$ and $GRSP_T$. The most obvious feature is that soil organic carbon and total nitrogen contents are the most important soil parameters to explain both GRSP fractions, which is consistant with N-containing proteins being important components of GRSP. Sand is also important, particularly for $GRSP_{EE}$, which may be related with the ease of extraction of protein. However clay and iron oxide contents are of less importance, lending little support to the widely held assumption that minerals contribute to the protection and hence accumulation of GRSP, like organic matter in general (Six et al., 2002). A striking feature of these rankings is that land use does not appear to be important, as is discussed later. The precise order of importance of the less important soil properties must be subject to caution. It is recognised that the selection of the most relevant features when variables are correlated is challenging (Kursa and Rudnicki, 2011) . We choose to use the VI criteria of the Cubist model in this study.

The curvilinear relation between both $GRSP_{EE}$ and $GRSP_T$ and both OC and N_T (Figure 2) suggests a difference between organic matter complexed with soil minerals and free, uncomplexed organic matter. Separate analyses were therefore performed in which either organic carbon was classed as complexed or non-complexed with clay (c. or n.c.) or clay classified as saturated or non-saturated with organic matter (s. or n.s.), or both This classification was obtained using the clay-organic matter relationship proposed by Dexter (Dexter et al., 2008a). Neither r^2 nor RMSPE were markedly changed by these changes to the models (Table SI_1), the relative importances of all parameters for $GRSP_{EE}$ and $GRSP_T$ and each of the variants of the model are shown in Figure SI_2. For both GRSP fractions the sum of relative importances of complexed and non complexed clay or OC fractions was very similar to that of the simpler model. Non complexed OC was always more important than complexed OC when the distinction was made. This may seem surprising, since both fractions of GRSP levelled off at larger values of OC, when more organic matter would be uncomplexed. However the explanation for this is that samples with no uncomplexed organic matter were excluded to avoid a strong weighting of zero, and so only the residual regression contributed.

3.2 Relations between GRSP and land use

In their review, Treseder & Turner (2007b) stated that average immuno-reactive $GRSP_T$ increased in the order desert<agricultural≈temperate grassland<boreal forest<temperate forest<tropical forest. Emran (Emran et al., 2012) also reports a strong land use effect on $GRSP_T$, with lowest values for vineyard soil and greatest values for pasture soil. Figure 1 shows a marked land use effect for both GRSP fractions, with median, first and third quartile values being larger for grassland and woodland soils than for cropland soils. However, Figure 2 shows

that the relations between GRSP and other soil parameters exhibit little distinction between land use classes, suggesting that the apparent land use effect may arise from other soil properties, and notably OC. Furthermore, the cubist modelling approach showed that land use was of very minor importance, since land use effects have been largely explained by soil properties. It seems likely that for this data set and most previously studied soils, the observed land use effect might be largely explained by soil composition, and in particular organic matter content.

3.3 Is GRSP a distinct fraction of organic matter?

The average values of $GRSP_{EE}$/OC and $GRSP_T$/OC were 0.038 and 0.097 respectively, with ranges of 0.006-0.096 and 0.02-0.19. Thus the average value for $GRSP_T$ would account for 3 or 5% of OC taking the C content of GRSP to be 32% or 53% (Rouwenhorst et al., 1991; Seguel et al., 2008). The strong correlation of both fractions of GRSP with OC content in this and other studies might suggest that GRSP is not a distinct fraction of organic matter with different patterns of production and accumulation. However the positive correlation of GRSP with N_T supports the assumption that GRSP is largely composed of proteins, that contain on average 16% N (Rouwenhorst et al., 1991). Another indication that GRSP is not a uniform fraction of OC is that the positive correlation of GRSP with clay content (Figure 2) was weaker than that of OC with clay (not shown, r^2=0.155, P<0.001) for these soils. Most importantly, the nonlinear relations between both fractions of GRSP and OC (Figure 2) indicate that GRSP does not simply follow OC. There is a smaller proportion of GRSP in organic matter in soils with large OC contents. This is in line with observations from chronosequences (Preger et al., 2007) that indicate that GRSP resists mineralisation following land-use change better than organic matter. As previously discussed, model predictions of GRSP were not greatly improved when complexed and non complexed OC and clay were considered separately. Another way of considering the distinction between complexed and free organic matter is to compare GRSP with the ratio of organic carbon to clay content. Values of this ratio superior to 0.1 correspond to the presence of free, uncomplexed organic matter, and values less than 0.1 correspond to under-saturated clay, according to the classification of Dexter (Dexter et al., 2008a). Significant positive linear correlations ($P<0.001$, $r^2= 0.416$ and 0.345 for $GRSP_{EE}$ and $GRSP_T$ respectively) were observed between both fractions of GRSP and the ratio OC:clay (Figure 4), with no inflection in the trend around the value of 0.1. However these correlations were led by extreme points, correlations were less significant for the first three quartiles and not significant when data from the interquartile range of OC:clay were considered.

Another way of exploring the non linear relationship between GRSP and OC is to study the trends of the ratio GRSP:OC as a function of soil composition and land-use. Figures 5 a) and b) show that GRSP as a fraction of OC decreases non linearly as OC increases. Although Preger (Preger et al., 2007) attributed a similar trend following land use change to the greater stability of GRSP, the trend could arise if GRSP represented a component of organic matter continuously produced by plant growth. OC content is an integrated measurement that takes no account of changes in the quality and composition of organic matter. Similar trends (not shown) were observed for both GRSP fractions and the soil nitrogen content. No land use effect could be separated from the overall trends.

Figure 5 c) and d) show that there was a strong trend for the ratio GRSP:OC to decrease with increasing clay content. Both OC and the GRSP fractions are positively correlated with clay content ($r^2 = 0.155$ for OC, and 0.0083 and 0.0076 for $GRSP_{EE}$ and $GRSP_T$ respectively). The slope of the correlation of OC is greater than that of GRSP (when normalised to the average value of each). The inverse relation is thus a mathematical consequence of the different slopes of the correlation of OC and GRSP *vs* clay content. This illustrates that, contrary to previous assumptions, GRSP is not more strongly protected by association with clay minerals than is organic matter in soil. No land use effect was observed for $GRSP_{EE}$ within this trend. For $GRSP_T$, correlations for soils from each of the three major land use categories differed slightly, with cropland soils having larger fractions of GRSP/OC at lowest clay content and a steeper decline with increasing clay content. This observation has important implications for the understanding of the chemical nature and biological origin of GRSP and for the hypothesis that this fraction could be exploited as a sink to mitigate rising levels of atmospheric carbon. It is possible that land use could indirectly affect the extent of mineral protection of organic matter. For example, it has been suggested that the solubilization of SOM by chelating ligands contributes to the priming effect (Keiluweit et al., 2015). The strong adsorption of proteins on organo-mineral surfaces would not be influenced by chelating ligands (Cheshire et al., 2000; Dümig et al., 2012; Quiquampoix and Burns, 2007). These trends thus converge to suggest that the greater the pressure on soil organic matter, due to various processes leading to organic matter depletion, including land management and absence of the stabilizing potential of clay minerals, the more favoured will be GRSP accumulation.

3.4 Is GRSP of fungal origin?

If GRSP were largely of arbuscular mycorrhizal fungal (AMF) origin, then strong relations would be expected with land use and management that influence fungal activity. Fungal activity is favoured by low pH (Joergensen and Wichern, 2008; Rousk et al., 2010). In this study pH was not identified as being very important for either fraction of GRSP. The importance of pH was greater for $GRSP_{EE}$ than for $GRSP_{T}$, leading to pH being the most important factor for the ratio $GRSP_{EE}/GRSP_{T}$. Another influence on mycorrhizal abundance and activity is the nutrient status, in particular the phosphorus status of soils (Abbott et al., 1984). Neither the cubist model, nor simple regression analysis found a strong effect of available P on either GRSP fraction. AMF are reported to be reduced by cropping, mainly due to tillage effects, biocidal treatments, monoculture and enrichment in P due to fertilization, and this could be invoked to explain the lower GRSP contents of cropland soils (Galvez et al., 2001; Gosling et al., 2006; Oehl et al., 2003; Pereira et al., 2018; Treseder and Allen, 2002). However, although both the OC and GRSP contents of cropland soils were lower than for other soils, their organic matter appears to be enriched in GRSP. For example, the median values of the ratio $GRSP_{T}/OC$ decrease in the order cropland (0.116)>grassland (0.094)>woodland (0.075) soils. This is in contradiction with the hypothesis that GRSP is predominantly of AMF origin (Driver et al., 2005; Wright and Upadhyaya, 1996b). Furthermore, in forest soils where fungal activity is strong, even if the dominant mycorrhizal fungi associated with temperate climate tree species are ectomycorrhizal (Joergensen and Wichern, 2008), GRSP is under-represented in organic matter. These data therefore provide no evidence to support the hypothesis that GRSP is predominantly of AMF fungal origin, and so further investigation would be necessary to confirm this assumption.

3.5 Do EE and Total fractions represent different ages?

On average the ratio between $GRSP_{EE}$ and $GRSP_T$ was 0.40±0.15 (Figure 1) with no significant difference between land use classes although the median value was larger in woodland soils than in crop or grassland soils. There is a significant linear relation between the two fractions of GRSP ($r=0.757$, $P<0.01$, Figure 6) but with no difference between land use categories. Previous studies have also reported linear relations between these fractions, albeit never for such a large sample size nor such a wide range of soil properties (Jorge-Araújo et al., 2015; Koide and Peoples, 2013; Lovelock et al., 2004a). An important consequence of the correlation between both operationally defined fractions of GRSP is that both are equally good probes of soil quality or stable soil organic matter content.

The cubist model identifies a simple situation for the ratio $GRSP_{EE}/GRSP_T$ with pH being the most important parameter, followed by sand and extractable P contents (Figure 3). There is no obvious explanation for relation between the ratio $GRSP_{EE}/GRSP_T$ and pH, which was also observed with simple linear regression (Figure SI_3). One explanation is the positive relation between pH and clay content, particularly for non calcareous soils (not shown). Since extraction of proteins, and indeed many organic compounds, is more difficult from clays than from sand due to strong binding processes, then relative extraction yield of the less harsh extraction would be expected to decrease with increasing clay, as observed with simple linear regression.

The cubist analysis was also performed for the difference of the two GRSP fractions $GRSP_T$-$GRSP_{EE}$. This difference has been proposed to represent the older fraction of GRSP, assuming the more easily extractable fraction to be recent, it should thus be favoured by soil components that protect the protein (Koide and Peoples, 2013). In each case (Figure 3) the most important variable was N_T, twice as important as the next most important variable. The order of relative importance of the other parameters varied only slightly between models (organic matter or clay or both separated onto complexed and non complexed categories). The importance of the clay content on the ratio $GRSP_{EE}/GRSP_T$ was greater than for either GRSP fraction. This results from the greater difficulty in extracting protein from clay than sand due to the strong interaction of proteins with organo-mineral surfaces (Helassa et al., 2011; Quiquampoix and Burns, 2007).

The trends observed do not lend support to the hypothesis that the easily extractable fraction is more recent than the total fraction. It should be noted that there is no independent verification of this hypothesis either for GRSP or for proteins in general in soil on the time scales of seasons or years. It is known that proteins may become increasingly strongly adsorbed on mineral surfaces with increasing contact period. This phenomenon has been studied on clean surfaces for short periods, never for complex surfaces such as soils over longer periods of months or years (Nakanishi et al., 2001). For complex surfaces such as soils, it is difficult to separate the effects of time dependent fixation that renders extraction less efficient, protein breakdown and loss of biological or biochemical activity (Helassa et al., 2011; Hung et al., 2016; Kedi et al., 2013). Hung and co-workers (Hung et al. 2016) attributed the loss of insecticidal activity of a Cry protein in contact with soil to fixation, but this appeared complete within a period of about 2 weeks. Koide & Peoples cite the poor correlation between easily extracted GRSP and the difference between total and easily extractable as confirmation of this hypothesis (Koide and Peoples, 2013). However, in the absence of marked land use change, it seems likely that GRSP accumulated from previous years would be a function of annual production, and hence strongly related to the most recent season.

In conclusion, the previously reported land-use effect on GRSP content appears to be explained by soil organic matter content. Differing trends in OC and GRSP contents lend support to the existence of GRSP as a distinct fraction of SOM. However none of the observed relations between GRSP and either land-use or soil composition could support the hypothesis that GRSP is solely of arbuscular mycorrhizal fungal origin.

Acknowledgements

RMQS soil sampling and physico-chemical analyses were supported by the GIS Sol, which is a scientific group of interest on soils involving the French Ministry for ecology and sustainable development and Ministry of agriculture, the French national forest inventory (IFN), ADEME (Agence de l'environnement et de la maîtrise de l'énergie), IRD (Institut de recherche et développement) and INRAE (Institut national de la recherche pour l'agriculture, l'alimentation et l'environnement). None of the funders played a role in study design, in the collection, analysis and interpretation of data, in the writing of the report; and in the decision to submit the article for publication. D.A. is supported by LE STUDIUM Loire Valley Institute for Advanced Studies through its LE STUDIUM Research Consortium Programme. Claudy Jolivet is thanked for his strong involvement in the RMQS monitoring network. The project was funded by an INRA "Projet Innovant" grant. The authors thank Josiane Abadie, Priscila Jorge-Araújo and Yvain Denat for their contribution to the experimental analyses.

List of Figure captions

Figure 1

Box plots of values of $GRSP_T$, $GRSP_{EE}$ and $GRSP_{EE}/GRSP_T$ for the full data set and for each of the major land use categories.

Figure 2

Various simple regressions of $GRSP_{EE}$ and $GRSP_T$ as a function of (from top to bottom) organic carbon content (OC), pH, total nitrogen content, N_T, and clay content with different symbols chosen to represent each of the major land uses; ; ○ cropland, □ grassland, Δ woodland and ♦ other land uses. Left column for $GRSP_{EE}$ and Right column for $GRSP_T$.

Figure 3

Relative importance parameters identified by Cubist modelling. a) $GRSP_{EE}$, b) $GRSP_T$, c) $GRSP_{EE}$ /$GRSP_T$ and d) $GRSP_T$ – $GRSP_{EE}$.

Figure 4

GRSP as a function of the ratio of contents of OC and clay, OC:Clay with different symbols chosen to represent each of the major land uses a) $GRSP_{EE}$ and b) $GRSP_T$; ○ cropland, □ grassland, Δ woodland and ♦ other land uses.

Figure 5

GRSP as a fraction of soil organic carbon, OC as a function of OC; a) $GRSP_T$ and b) $GRSP_{EE}$ and as a function of clay content; c) $GRSP_{EE}$ and d) $GRSP_T$ with different symbols chosen to represent each of the major land uses; ○ cropland, □ grassland, Δ woodland and ♦ other land uses.

Figure 6

Relation between $GRSP_{EE}$ and $GRSP_T$ with different symbols chosen to represent each of the major land uses; ○ cropland, □ grassland, Δ woodland and ♦ other land uses. The straight line shows the best fit linear regression through all the data.

Supplementary Information

Table caption

Table SI 1

Cross validation results for the different models, showing the coefficients of determination and root mean square prediction errors obtained for each of the four models for $GRSP_{EE}$ and $GRSP_T$ the ratio of the two fractions and their difference. Model 1 (OC, Clay); Model 2 (cOC, ncOC, Clay); Model 3 (sClay, nsClay, OC); Model 4 (cOC, ncOC, sClay, nsClay) where s/ns are abbreviations for saturated and non saturated and c/nc for complexed and non complexed.

Figure captions

Figure SI_1

Box plots of values of various soil properties for the full data set and for each of the major land use categories.

Figure SI_2

Effect on relative importance parameters identified by Cubist modelling of separating organic matter into complexed (c) and non-complexed (nc) classes and/or clay as saturated (s) in organic matter or non-saturated OC/clay (ns) for $GRSP_{EE}$ (left-hand columns) and $GRSP_T$ (right-hand columns). For ease of comparison, data of Model 1 (OC and Clay not subdivided) already presented in Figure 4 are shown.

Figure SI_3

Ratio of the two GRSP fractions, $GRSP_{EE}$ /$GRSP_T$ as a function of a) pH and b) Clay content with different symbols to indicate each of the major land uses; ○ cropland, □ grassland, Δ woodland and ♦ other land uses. There is a significant ($P< 0.001$, $r^2 = 0.31$) inverse linear relationship between the ratio and pH.

Graphical highlight

Geological map of France showing $GRSP_{EE}$ contents of soils

References cited

Abbott, L.K., Robson, A.D., De Boer, G., 1984. The effect of phosphorus on the formation of hyphae in soil by the vesicular-arbuscular mycorrhizal fungus, Glomus fasciculatum. New Phytologist, 97, 437-446.

Arrouays, D., Deslais, W., Badeau , V., 2001. The carbon content of topsoil and its geographical distribution in France. Soil Use and Management, 17, 7-11.

Arrouays, D. et al., 2002. Une initiative nouvelle en France : la mise en place d'un réseau institutionnel de mesure de la qualité des sols (RMQS). Comptes Rendus de l'Académie d'Agriculture de France, 88, 93-103.

Bedini, S. et al., 2009. Changes in soil aggregation and glomalin-related soil protein content as affected by the arbuscular mycorrhizal fungal species Glomus mosseae and Glomus intraradices. Soil Biology and Biochemistry, 41, 1491-1496.

Bolliger, A. et al., 2008. Re-examining the glomalin-purity of glomalin-related soil protein fractions through immunochemical, lectin-affinity and soil labelling experiments. Soil Biology and Biochemistry, 40, 887-893.

Bradford, M.M., 1976. A rapid and senstive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72, 248-254.

Cheshire, M.V., Dumat, C., Fraser, A.R., Hillier, S., Staunton, S., 2000. The interaction between soil organic matter and soil clay minerals by selective removal and controlled addition of organic matter. European Journal of Soil Science, 51, 497-509.

Dexter, A.R. et al., 2008a. Complexed organic matter controls soil physical properties. Geoderma, 144, 620-627.

Dexter, A.R. et al., 2008b. Complexed organic matter controls soil physical properties. Geoderma, 144, 620-627.

Driver, J.D., Holben, W.E., Rillig, M.C., 2005. Characterization of glomalin as a hyphal wall component of arbuscular mycorrhizal fungi. Soil Biology & Biochemistry, 37, 101-106.

Dümig, A., Häusler, W., Steffens, M., Kögel-Knabner, I., 2012. Clay fractions from a soil chronosequence after glacier retreat reveal the initial evolution of organo–mineral associations. Geochimica et Cosmochimica Acta, 85, 1-18.

Emran, M., Gispert, M., Pardini, G., 2012. Patterns of soil organic carbon, glomalin and structural stability in abandoned Mediterranean terraced lands. European Journal of Soil Science, 63, 637-649.

Fokom, R. et al., 2012. Glomalin related soil protein, carbon, nitrogen and soil aggregate stability as affected by land use variation in the humid forest zone of south Cameroon. Soil and Tillage Research, 120, 69-75.

Franzluebbers, A.J., Wright, S.F., Stuedemann, J.A., 2000. Soil Aggregation and Glomalin under Pastures in the Southern Piedmont USA. Soil Science Society of America Journal, 64, 1018-1026.

Galvez, L., Douds, D.D., Drinkwater, L.E., Wagoner, P., 2001. Effect of tillage and farming system upon VAM fungus populations and mycorrhizas and nutrient uptake of maize. Plant and Soil, **228**, 299–308.

Gillespie, A.W. et al., 2011a. Glomalin-related soil protein contains non-mycorrhizal-related heat-stable proteins, lipids and humic materials. Soil Biology and Biochemistry, 43, 766-777.

Gillespie, A.W. et al., 2011b. XANES and Pyrolysis-FIMS Evidence of Organic Matter Composition in a Hummocky Landscape. Soil Science Society of America Journal, 75, 1741.

Gispert, M., Emran, M., Pardini, G., Doni, S., Ceccanti, B., 2013. The impact of land management and abandonment on soil enzymatic activity, glomalin content and aggregate stability. Geoderma, 202-203, 51-61.

Gosling, P., Hodge, A., Goodlass, G., Bending, G.D., 2006. Arbuscular mycorrhizal fungi and organic farming. Agriculture, Ecosystems & Environment, 113, 17-35.

Halvorson, J.J., Gonzalez, J.M., 2006. Bradford reactive soil protein in Appalachian soils: distribution and response to incubation, extraction reagent and tannins. Plant and Soil, 286, 339-356.

Helassa, N. et al., 2011. Effects of physicochemical interactions and microbial activity on the persistence of Cry1Aa Bt (Bacillus thuringiensis) toxin in soil. Soil Biology & Biochemistry, 43, 1089-1097.

Hernandez-Soriano, M.C. et al., 2018. Soil Organic Carbon Stabilization: Mapping Carbon Speciation from Intact Microaggregates. Environmental Science & Technology, 52, 12275-12284.

Houghton, R.A. et al., 2012. Carbon emissions from land use and land-cover change. Biogeosciences, 9, 5125-5142.

Hung, T.P. et al., 2016. Persistence of detectable insecticidal proteins from Bacillus thuringiensis (Cry) and toxicity after adsorption on contrasting soils. Environmental Pollution, 208, 318-325.

Hurisso, T.T., Moebius-Clune, D.J., Culman, S.W., Moebius-Clune, B.N., Thies, J.E. & van Es, H.M. 2018. Soil Protein as a Rapid Soil Health Indicator of Potentially Available Organic Nitrogen. Agricultural and Environmental Letters, 3, 10.2134/ael2018.02.0006.

Janzen, H.H., 2006. The soil carbon dilemma: Shall we hoard it or use it? Soil Biology and Biochemistry, 38, 419-424.

Janzen, H.H., 2015. Beyond carbon sequestration: soil as conduit of solar energy. European Journal of Soil Science, 66, 19-32.

Joergensen, R., Wichern, F., 2008. Quantitative assessment of the fungal contribution to microbial tissue in soil. Soil Biology and Biochemistry, 40, 2977-2991.

Jorge-Araújo, P., Quiquampoix, H., Matumoto-Pintro, P.T., Staunton, S., 2015. Glomalin-related soil protein in French temperate forest soils: interference in the Bradford assay caused by co-extracted humic substances. European Journal of Soil Science, 66, 311-319.

Kedi, B., Sei, J., Quiquampoix, H., Staunton, S., 2013. Persistence of catalytic activity of fungal phosphatases incubated in tropical soils. Soil Biology and Biochemistry, 56, 69-74.

Keiluweit, M. et al., 2015. Mineral protection of soil carbon counteracted by root exudates. Nature Climate Change, 5, 588-595.

King, G.M., 2011. Enhancing soil carbon storage for carbon remediation: potential contributions and constraints by microbes. Trends in Microbiology, 19, 75-84.

Koide, R.T., Peoples, M.S., 2013. Behavior of Bradford-reactive substances is consistent with predictions for glomalin. Applied Soil Ecology, 63, 8-14.

Kursa, M.B., Rudnicki, W.R., 2011. The All Relevant Feature Selection using Random Forest, https://arxiv.org/abs/1106.5112.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma, 123, 1-22.

Lorenz, K., Lal, R., Preston, C.M., Nierop, K.G.J., 2007. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. Geoderma, 142, 1-10.

Lovelock, C.E., Wright, S.F., Clark, D.A., Ruess, R.W., 2004a. Soil stocks of glomalin produced by arbuscular mycorrhizal fungi across a tropical rain forest landscape. Journal of Ecology, 92, 278-287.

Lovelock, C.E., Wright, S.F., Clark, D.A., Ruess, R.W., 2004b. Soil stocks of glomalin produced by arbuscular mycorrhizal fungi across a tropical rain forest landscape. Journal of Ecology, 92, 278-287.

Meersmans, J. et al., 2012. A high resolution map of French soil organic carbon. Agronomy for Sustainable Development, 32, 841-851.

Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems, 94, 72-79.

Nakanishi, K., Sakiyama, T., Imamura, K., 2001. On the adsorption of proteins on solid surfaces, a common but very complicated phenomenon. Journal of Bioscience and Bioengineering, 91, 233-244.

Nichols, K.A., Wright, S.F., 2005. Comparison of glomalin and humic acid in eight native U.S. soils. Soil Science, 170, 985-997.

Oehl, F. et al., 2003. Impact of Land Use Intensity on the Species Diversity of Arbuscular Mycorrhizal Fungi in Agroecosystems of Central Europe. Applied and Environmental Microbiology, 69, 2816-2824.

Oliveira, S.P. et al., 2016. Conversion of forest into irrigated pasture II. Changes in the physical properties of the soil. Catena, 143, 70-77.

Pereira, C.M.R., da Silva, D.K.A., Goto, B.T., Rosendahl, S., Maia, L.C., 2018. Management practices may lead to loss of arbuscular mycorrhizal fungal diversity in protected areas of the Brazilian Atlantic Forest. Fungal Ecology, 34, 50-58.

Preger, A.C. et al., 2007. Losses of glomalin-related soil protein under prolonged arable cropping: A chronosequence study in sandy soils of the South African Highveld. Soil Biology and Biochemistry, 39, 445-453.

Quinlan, J.R., 1992. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart 16-18 November 1992, 343-348.

Quiquampoix, H., Burns, R.G., 2007. Interactions between Proteins and Soil Mineral Surfaces: Environmental and Health Consequences. Elements, 3, 401-406.

Rillig, M.C., Maestre, F.T., Lamit, L.J., 2003. Microsite differences in fungal hyphal length, glomalin, and soil aggregate stability in semiarid Mediterranean steppes. Soil Biology & Biochemistry, 35, 1257-1260.

Rillig, M.C., Wright, S.F., Nichols, K.A., Schmidt, W.F., Torn, M.S., 2001. Large contribution of arbuscular mycorrhizal fungi to soil carbon pools in tropical forest soils. Plant and Soil, 233, 167-177.

Rousk, J., Brookes, P.C., Baath, E., 2010. Investigating the mechanisms for the opposing pH relationships of fungal and bacterial growth in soil. Soil Biology & Biochemistry, 42, 926-934.

Rouwenhorst, R.J., Jzn, J.F., Scheffers, W.A., van Dijken, J.P., 1991. Determination of protein concentration by total organic carbon analysis. Journal of Biochemical and Biophysical Methods, 22, 119-128.

Šarapatka, B., Alvarado-Solano, D.P., Čižmár, D., 2019. Can glomalin content be used as an indicator for erosion damage to soil and related changes in organic matter characteristics and nutrients? Catena, 181, 104078.

Schindler, F.V., Mercer, E.J., Rice, J.A., 2007. Chemical characteristics of glomalin-related soil protein (GRSP) extracted from soils of varying organic matter content. Soil Biology and Biochemistry, 39, 320-329.

Schmidt, M.W. et al., 2011. Persistence of soil organic matter as an ecosystem property. Nature, 478, 49-56.

Seguel, A., Rubio, R., Carrillo, R., Espinosa, A., .Borie, F., 2008. Levels of glomalin and their relation with soil chemical and biological soil (andisol) characteristics in a relic of native forest of southern Chile. Bosque, 29, 11–22.

Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils Plant and Soil, 241, 155-176.

Stockmann, U. et al., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems & Environment, 164, 80-99.

Treseder, K.K., Allen, M.F., 2002. Direct nitrogen and phosphorus limitation of arbuscular mycorrhizal fungi: a model and field test. New Phytologist, **155**, 507-515.

Treseder, K.K., Turner, K.M., 2007a. Glomalin in Ecosystems. Soil Science Society of America Journal, 71, 1257.

Treseder, K.K., Turner, K.M., 2007b. Glomalin in ecosystems. Soil Science Society of America Journal, 71, 1257-1266.

Whiffen, L.K., Midgley, D.J., McGee, P.A., 2007. Polyphenolic compounds interfere with quantification of protein in soil extracts using the Bradford method. Soil Biology and Biochemistry, 39, 691-694.

Woignier, T., Etcheverria, P., Borie, F., Quiquampoix, H., Staunton, S., 2014. Role of allophanes in the accumulation of glomalin-related soil protein in tropical soils (Martinique, French West Indies). European Journal of Soil Science, 65, 531-538.

Wright, S.F., Franke-Snyde, M., Morton, J.B., Upadhyaya, A., 1996. Time-course study and partial characterization of a protein on hyphae of arbuscular mycorrhizal fungi during active colonization of roots. Plant and Soil, 181, 193-203.

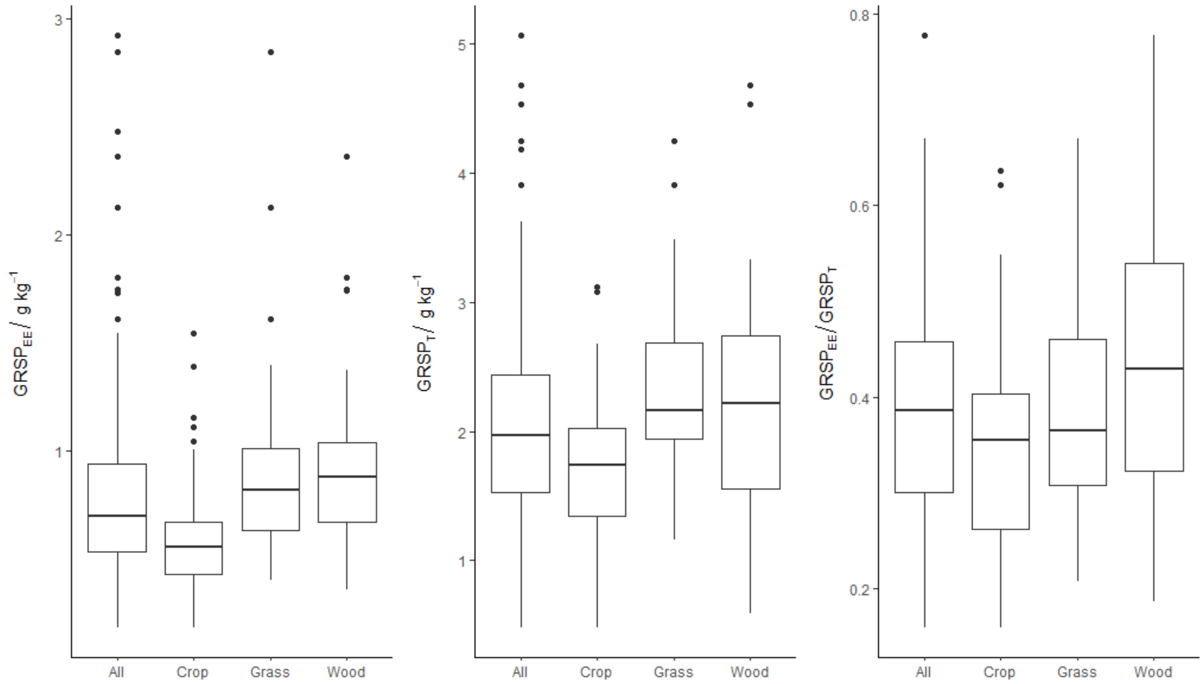
Wright, S.F., Upadhyaya, A., 1996a. Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. Soil Science, 161, 575-586.

Wright, S.F., Upadhyaya, A., 1996b. Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. Soil Science, 161, 575-586.

Young, I.M., Feeney, D.S., O'Donnell, A.G., Goulding, K.W.T., 2012. Fungi in century old managed soils could hold key to the development of soil water repellency. Soil Biology & Biochemistry, 45, 125-127.

Yu, G. et al., 2017. Mineral Availability as a Key Regulator of Soil Carbon Storage. Environmental Science & Technology, 51, 4960-4969.

Figure 1



3
2
1
$GRSP_{EE}$/ g kg^{-1}
All
Crop
Grass
Wood
5
4
3
2
1
$GRSP_T$/ g kg^{-1}
All
Crop
Grass
Wood
0.8
0.6
0.4
0.2
$GRSP_{EE}$/ $GRSP_T$
All
Crop
Grass
Wood

Figure 2

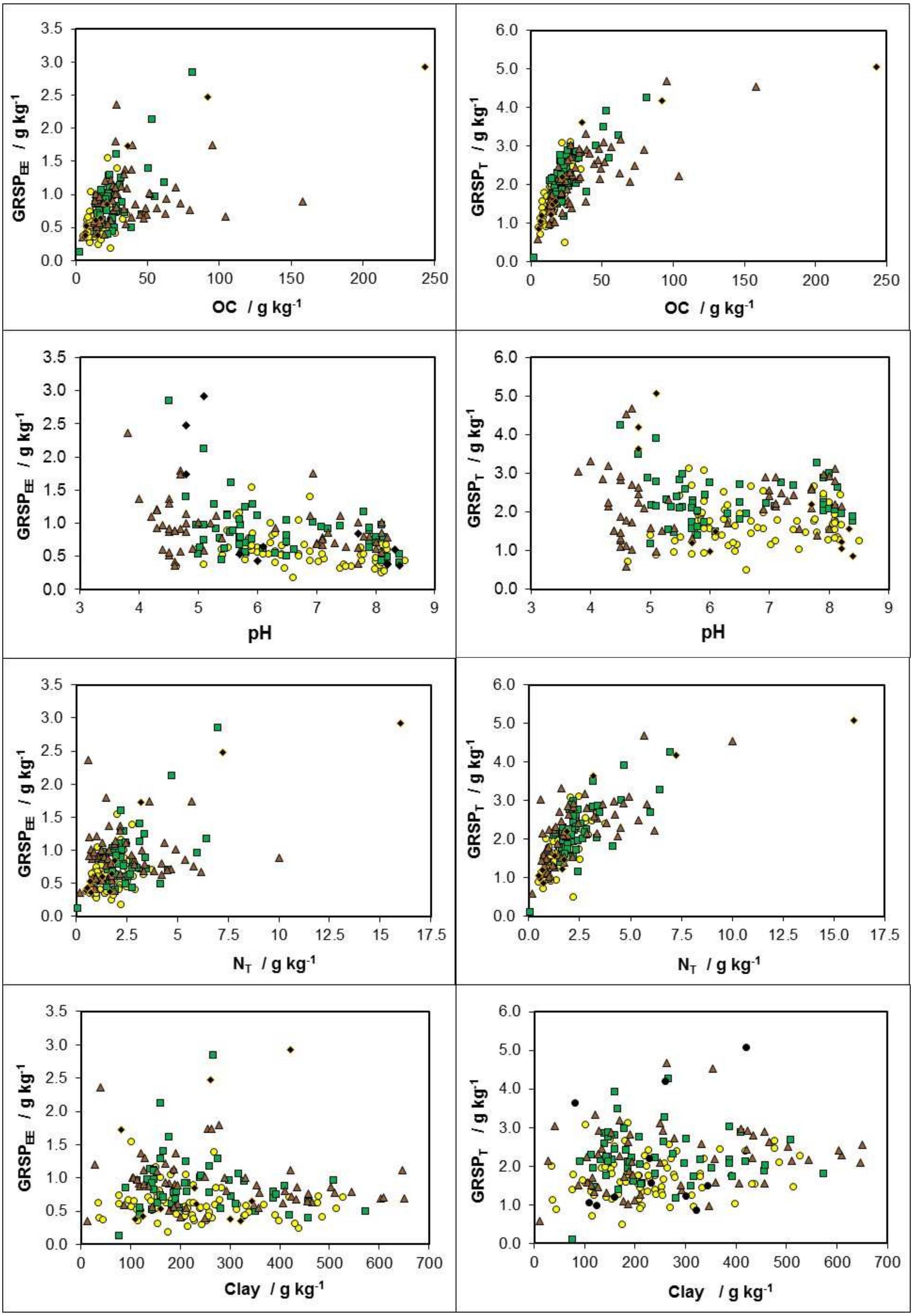
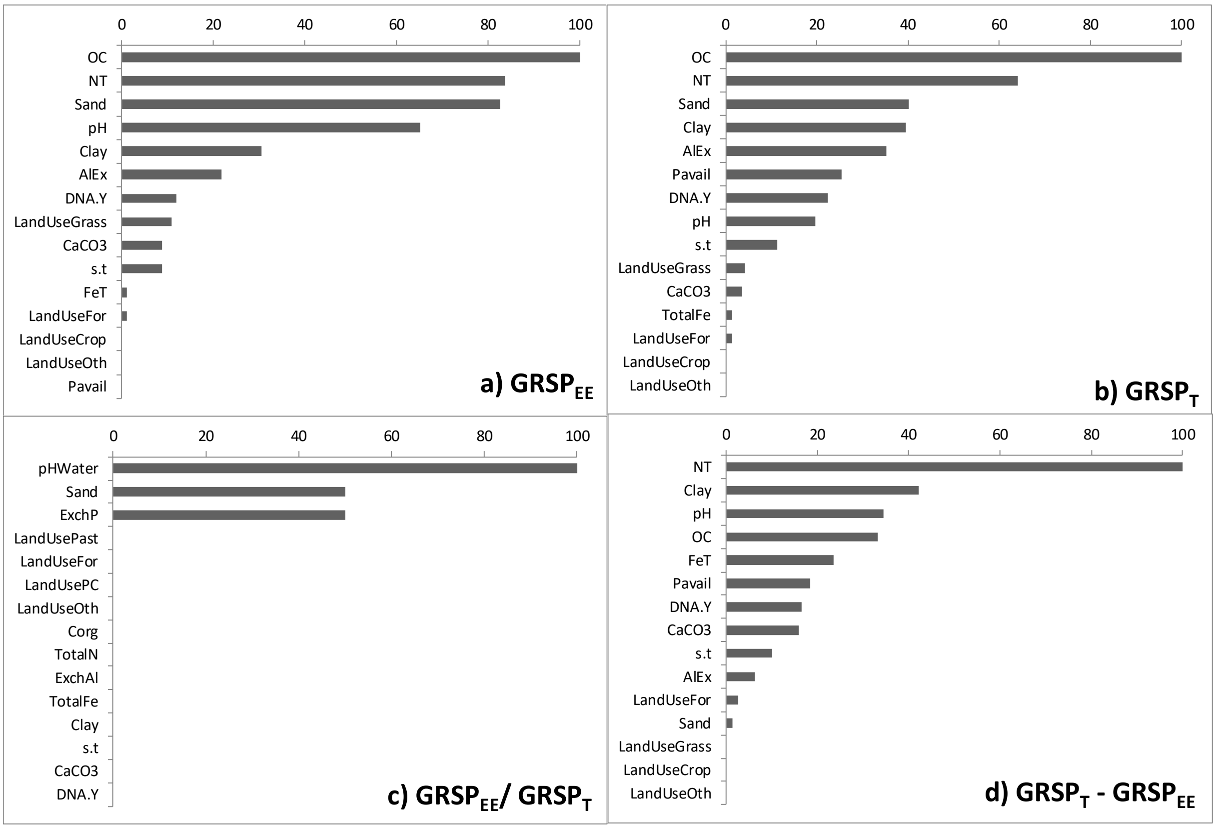
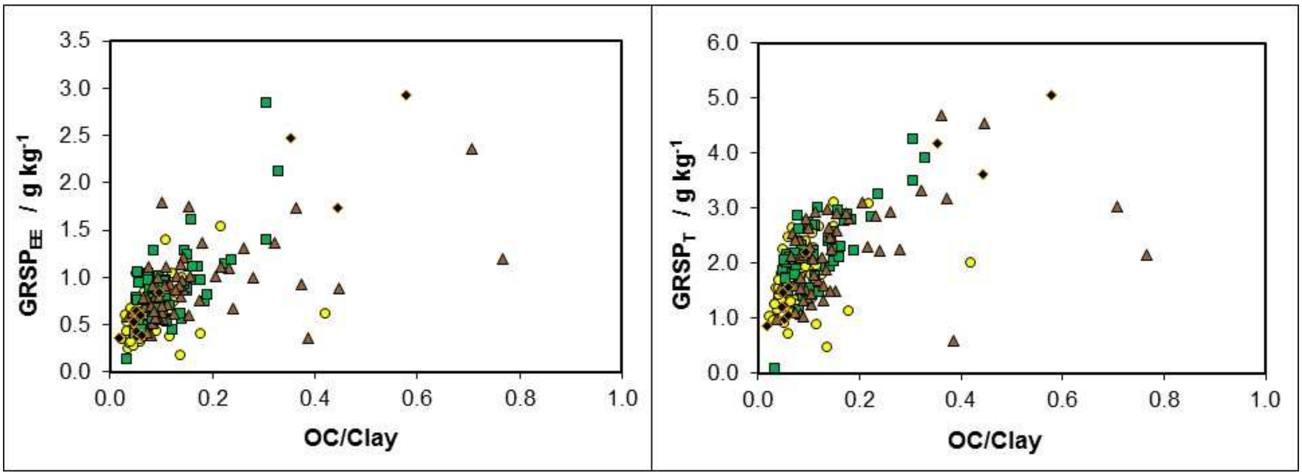

GRSP$_{EE}$ / g kg^{-1}
OC / g kg^{-1}
GRSP$_T$ / g kg^{-1}
OC / g kg^{-1}
GRSP$_{EE}$ / g kg^{-1}
pH
GRSP$_T$ / g kg^{-1}
pH
GRSP$_{EE}$ / g kg^{-1}
N$_T$ / g kg^{-1}
GRSP$_T$ / g kg^{-1}
N$_T$ / g kg^{-1}
GRSP$_{EE}$ / g kg^{-1}
Clay / g kg^{-1}
GRSP$_T$ / g kg^{-1}
Clay / g kg^{-1}

Figure 3



0 20 40 60 80 100
OC
NT
Sand
pH
Clay
AlEx
DNA.Y
LandUseGrass
CaCO3
s.t
FeT
LandUseFor
LandUseCrop
LandUseOth
Pavail
a) GRSP$_{EE}$
0 20 40 60 80 100
OC
NT
Sand
Clay
AlEx
Pavail
DNA.Y
pH
s.t
LandUseGrass
CaCO3
TotalFe
LandUseFor
LandUseCrop
LandUseOth
b) GRSP$_{T}$
0 20 40 60 80 100
pHWater
Sand
ExchP
LandUsePast
LandUseFor
LandUsePC
LandUseOth
Corg
TotalN
ExchAl
TotalFe
Clay
s.t
CaCO3
DNA.Y
c) GRSP$_{EE}$/ GRSP$_{T}$
0 20 40 60 80 100
NT
Clay
pH
OC
FeT
Pavail
DNA.Y
CaCO3
s.t
AlEx
LandUseFor
Sand
LandUseGrass
LandUseCrop
LandUseOth
d) GRSP$_{T}$ - GRSP$_{EE}$

Figure 4



3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
GRSP$_{EE}$ / g kg^{-1}
OC/Clay
6.0
5.0
4.0
3.0
2.0
1.0
0.0
GRSP$_{T}$ / g kg^{-1}
OC/Clay

Figure 5

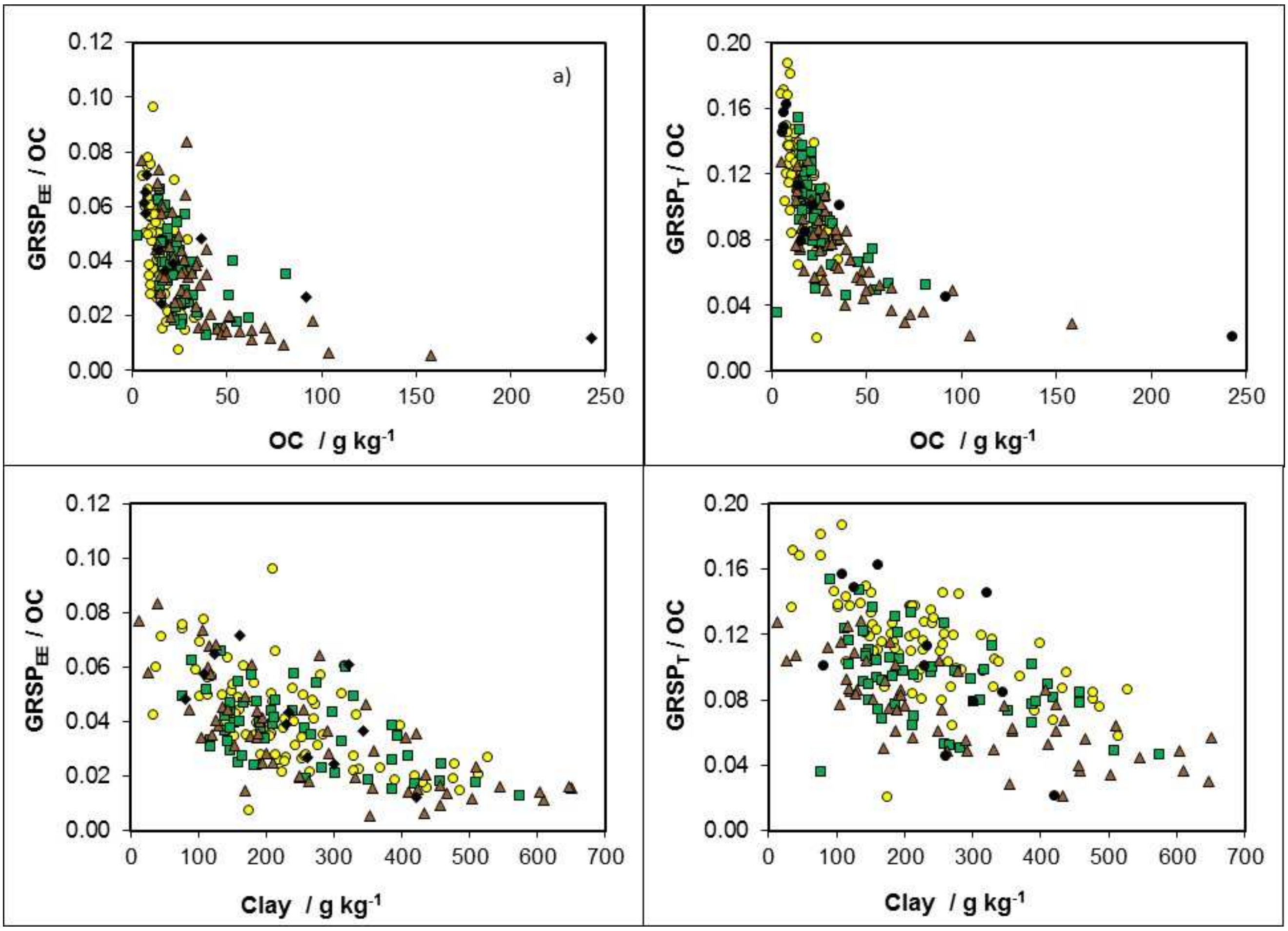
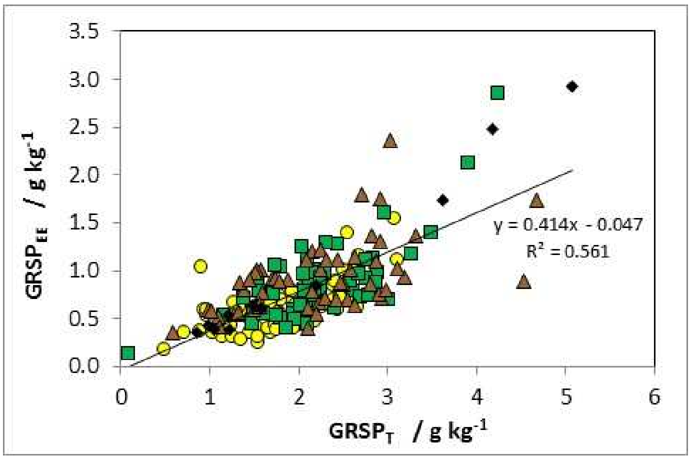
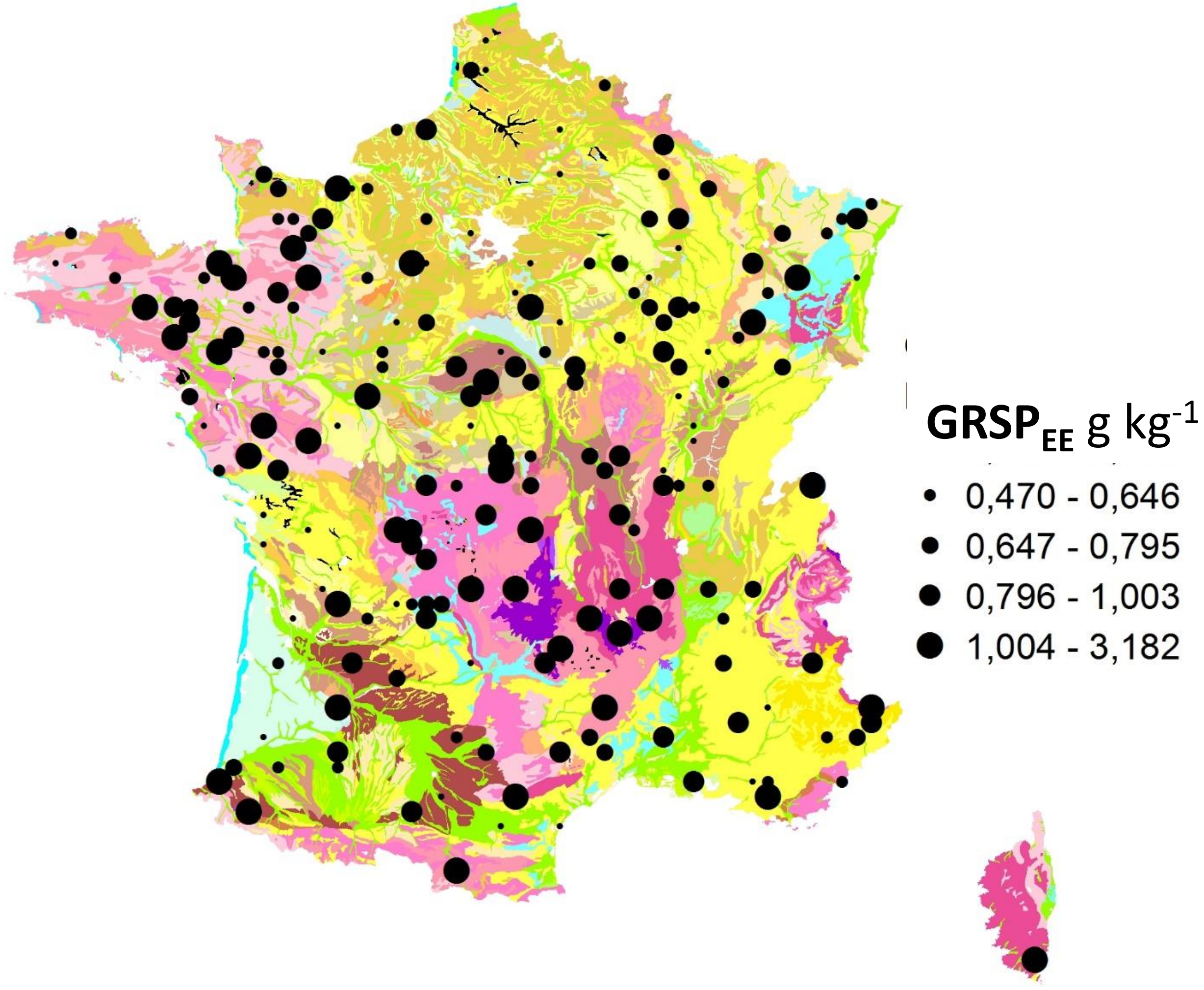


Figure 6



3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
GRSP$_{EE}$ / g kg^{-1}
0
1
2
3
4
5
6
GRSP$_T$ / g kg^{-1}
y = 0.414x - 0.047
R^2 = 0.561



Geological map of metropolitan France showing the measured values of GRSP$_{EE}$, the size of the spots indicating the range of GRSP$_{EE}$ content