Modelling, interpreting and visualizing uncertainties for the North Wyke Farm Platform baseline field surveys

Paul Harris*[1], Chris Brunsdon[2], Lex Comber[3], Hadewij Sint[1], Robert Orr[1],

Michael Lee[1], Phil Murray[1]

[1] Rothamsted Research, North Wyke, Devon, EX20 2SB, UK
[2] National Centre for Geocomputation, Maynooth Unuveristy, Ireland
[3] University of Leeds, Leeds, West Yorkshire, LS2 9JT, UK

*Corresponding author: paul.harris@rothamsted.ac.uk

Abstract
This study demonstrates a new approach for the visualization of spatial uncertainty using data from three agricultural field surveys.

Keywords
caricRture; kriging; geographically weighted models; grasslands research

I INTRODUCTION

The North Wyke Farm Platform (NWFP) is a systems-based, farm-scale experiment with the aim of addressing grassland agricultural productivity and ecosystem responses to different management practices. The 63 ha site captures the data necessary to develop a better understanding of the dynamic processes and underlying mechanisms that can be used to model how agricultural grassland systems respond to different management inputs.

Via cattle beef and sheep production, the underlying principle is to manage each of three farmlets in three contrasting ways: (i) improvement of permanent pasture (i.e. 'business as usual' *green* farmlet); (ii) improvement through the use of legumes (*blue* farmlet); and (iii) improvement through innovation (*red* farmlet). The connectivity between the timing and intensity of the different management operations, together with the transport of nutrients and potential pollutants from the NWFP is evaluated using numerous inter-linked data collection exercises, operating at various spatial and/or temporal scales. Fig. 1a maps the NWFP experiment, where its 15 hydrologically-isolated sub-catchments are shown; some of which consist of multiple fields. In this study, we introduce some of the modelling and visualization opportunities that are possible with this rich data resource, with respect to baseline field survey data only.

II METHODS

We study three surveys that were conducted in the summers of 2012 and 2013. Baseline data entails that the *blue* and *red* treatments of NWFP farmlets are not as yet in place (thus the map in Fig. 1a is effectively all *green* during these surveys). Surveys sampled over a mixture of 25m and 50m grids for: (a) plant nutrients in 2013 (n = 544); (b) plant species in 2013 (n = 294); and (c) soil nutrients in 2012 (n = 250). These data sets (and many more) are freely available at www.rothamsted.ac.uk/farmplatform.

For plant nutrients, bulk stable isotopes of Carbon (δ^{13}C) and Nitrogen (δ^{15}N) along with Total Carbon (C), Total Nitrogen (N) and sward height were measured. For plant species, 18 species were observed, where two species, Agrostis stolonifera and Loium perenne, clearly dominated. For soils nutrients, $\delta^{13}C$, $\delta^{15}N$, C and N, together with Bulk Density, Soil Organic Matter and pH were measured (Fig. 1b). Surveys can be analysed as a self-contained entity or related to each other.

Our study implements geostatistical and geographically weighted (GW) models to these surveys. Outputs are visualized to reflect different aspects of uncertainty, showcasing functions provided in the caricRture R package (Brunsdon 2016). Functions extend the 'sketchy' rendering techniques of Wood et al. (2012) to use with spatial data.

Using caricRture, we map a model output of interest (predicted/estimated to a finer 15m grid) via short lines that are coloured to demark the output range. To reflect uncertainty, the lines are drawn with varying degrees of 'sketchiness', where straight lines suggest relatively low levels of uncertainty, whilst highly sketchy ('hand-drawn') lines suggest relatively high levels of uncertainty. The technique provides a useful way of communicating uncertainty, where information that is commonly presented as two maps, are presented using only one. We present four case studies, each with a different visualization theme.

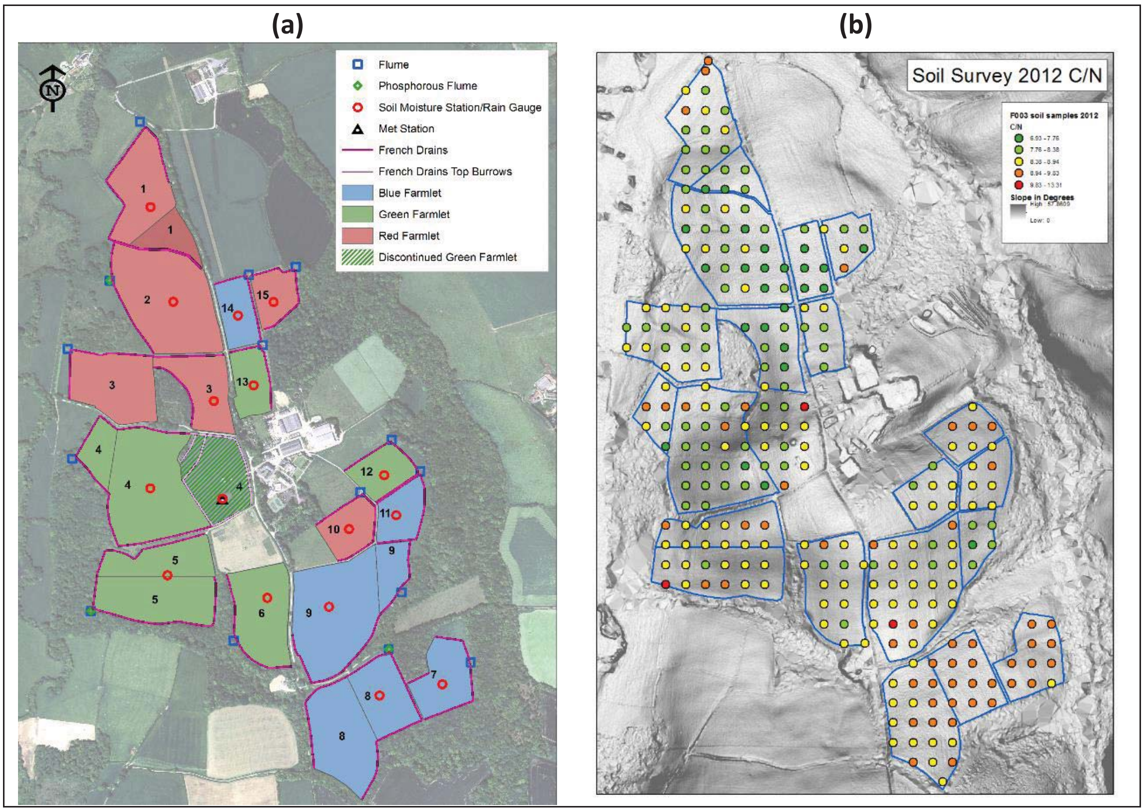

Figure 1:(a) The North Wyke Farm Platform and (b) Map depicting the C/N ratio for soils.

III CASE STUDIES

Visualization of spatial prediction uncertainty

In our first visualization, we predict C from the plant nutrients survey using universal kriging (UK) with a linear drift. Instead of using the UK variance as a measure of prediction uncertainty, we use the UK interpolation variance of Yamamoto (2000) which, unlike the UK variance, accounts for local changes in sample variance. The method is sensitive to the kriging neighbourhood size, and we choose N = 20. Fig.2a provides the UK prediction map, Fig. 2b provides the corresponding UK interpolation variance map, and Fig. 6a. provides the resultant 'sketchy' map - combining both UK predictions and their uncertainty. Thus the prediction levels remain the same for maps in Fig 2a and Fig. 6a, whereas the high interpolation variances coloured dark green in Fig. 2b correspond to areas of high 'sketchiness' in Fig. 6a. Thus plant C predictions should be viewed cautiously in sub-catchments 2 and 7 (see also Fig. 1a).

Visualization of species metric uncertainty

Next, we suggest how to visualize species richness together with species diversity (for the plant species survey). Species diversity is a function the number of species present (i.e. species richness) and the evenness of how the individuals are spread (Hurlbert 1971). We choose to view species richness as our main output (Fig. 3a) and species diversity (Shannon's index) as its uncertainty or variance (Fig. 3b). To aid these visualizations we have smoothed both indices to the 15m grid. Fig. 6b provides the resultant 'sketchy' map - combining both species metrics. Species richness levels remain the same for maps in Fig 3a and Fig. 6b, whereas high levels of species diversity coloured dark green in Fig. 3b correspond to areas of high 'sketchiness' in Fig. 6b. Thus plant species are both rich and diverse, in for example, sub-catchments 14 and 15.

Visualization of local multivariate data structure uncertainty

Here, we apply a localised PCA (GWPCA, Harris et al. 2011) to all seven variables of the soils nutrients survey. We then map the local percentage of the total variance (PTV) accounted for by the first two components (PC1 and PC2) in Fig. 4a, where it appears that the seven soils variables tend to be relatively uncorrelated in the central sub-catchments of the NWFP, but relatively correlated elsewhere. From the usual PCA, the single global PTV value is 64.6% for the first two components, and in this instance, we map the absolute difference between this global value and each local PTV value in Fig. 4b. In Fig. 6c, we use that depicted in Fig. 4b, to reflect local structural uncertainty in our soils data with respect to how different the local PTV data is to that found globally. Thus data structure in sub-catchment 4 is most like that found, if we were to naively apply a non-spatial PCA to this data.

Visualization of local regression coefficient uncertainty

For our final visualisation, we apply a bootstrap methodology (Harris et al. 2015) to test for non-stationarity in the coefficients of a GW regression (GWR) as an alternative to a multiple linear regression (MLR), the null hypothesis. Our response is plant species diversity and our predictor is the C/N ratio for soils. GWR reduces AIC by 22 units from the MLR fit, suggesting clear value in a localised relationship. Fig. 5a maps the local regression coefficients for the C/N ratio for soils, which compares to a global coefficient of 0.015. Fig 5b maps p-values that have been formatted such that low p-values indicate coefficients which are significantly different to that found globally. Fig. 6d displays the corresponding 'sketchy' map, where high levels of 'sketchiness' indicate areas of significant non-stationarity in the coefficient estimates.

IV CONCLUDING COMMENTS

We have presented four case studies on the value of 'sketchy' maps to communicate uncertainty. Further 'sketchy' visualizations are possible when it comes to relaying uncertainty in the boundaries of spatial data, such as those of each NWFP sub-catchment, which we are currently working on.

References

Brunsdon (2016). Representing Uncertain Geographical Information with Algorithmic Map Caricatures. *Geophysical Research Abstracts Vol 18 EGU2016-2609,* EGU General Assembly 2016.

Harris P, Brunsdon C, Charlton M (2011) Geographically weighted principal components analysis. *International Journal of Geographical Information Science* 25(10), 1717-1736.

Harris P, Brunsdon C, Gollini I, Nakaya T, Charlton M (2015) Using Bootstrap Methods to Investigate Coefficient Non-stationarity in Regression Models: An Empirical Case Study. *Procedia Environmental Sciences* 27:112-115.

Hurlbert SH (1971). The Nonconcept of Species Diversity. *Ecology* 52(4), 577-586.

Wood J, Isenberg P, Isenberg T, Dykes J, Boukhelifa N, Slingsby A (2012). Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 18(12), 2749-2758.

Yamamoto JK (2000). An alternative measure of the reliability of ordinary kriging estimates. *Mathematical Geology* 32, 489-509.

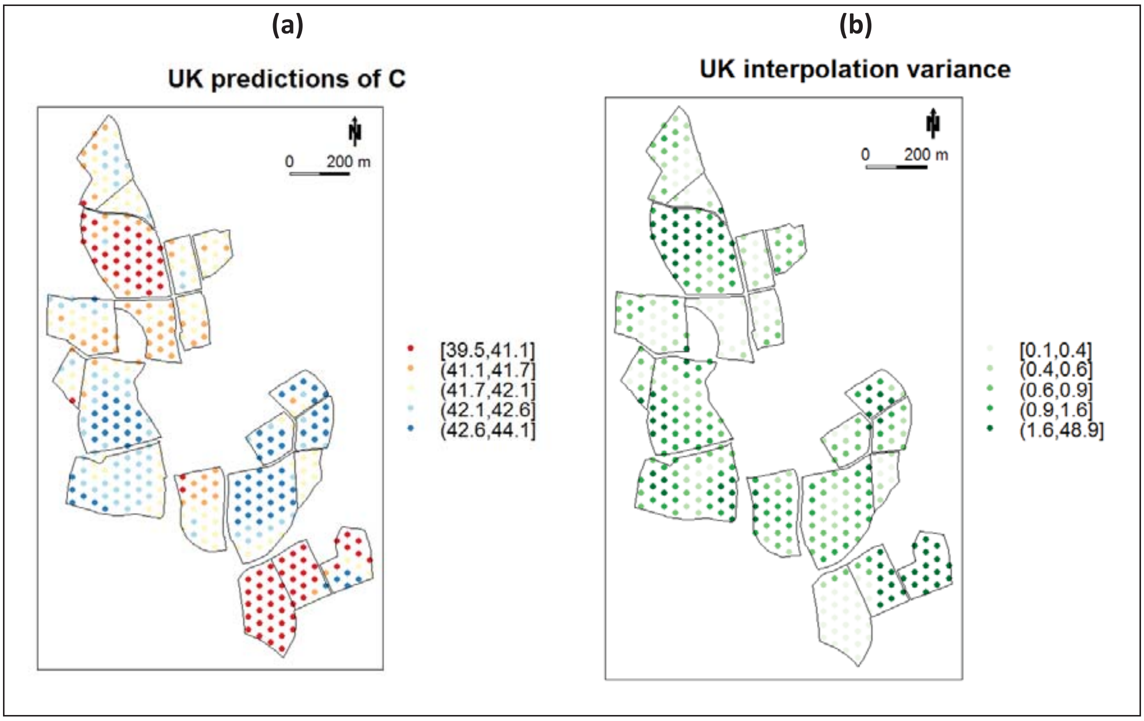

Figure 2: (a) UK predictions of C (plant nutrients) and (b) corresponding UK interpolation variances.

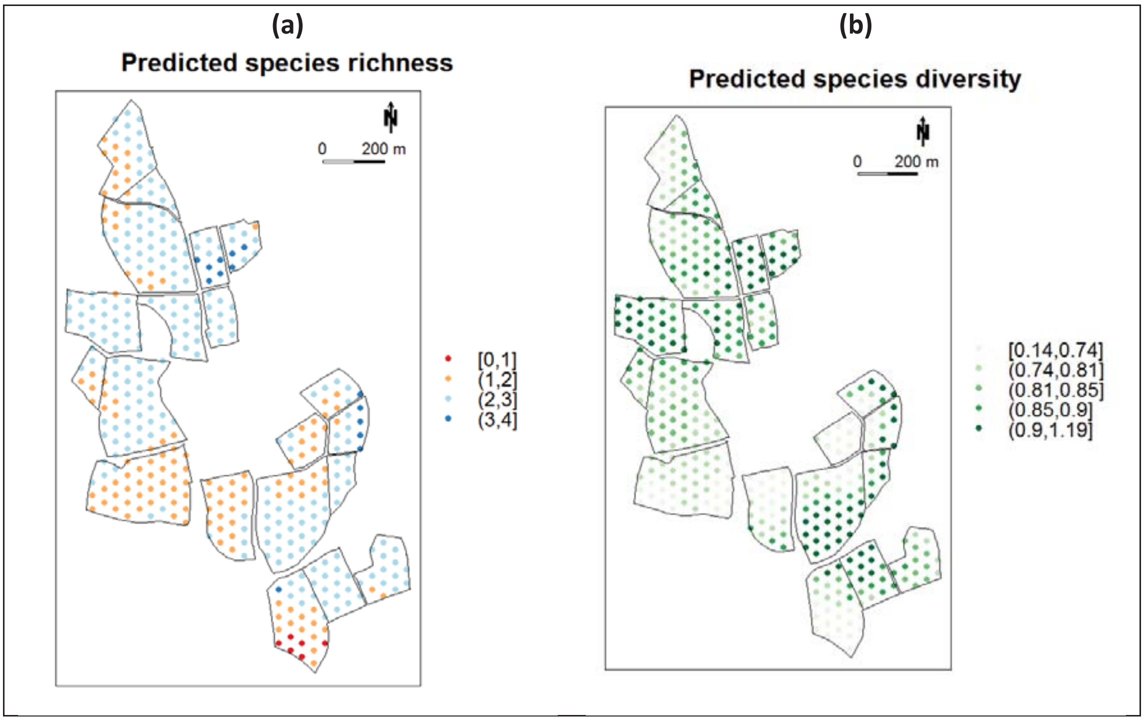

Figure 3: (a) Predicted species richness and (b) predicted species diversity.

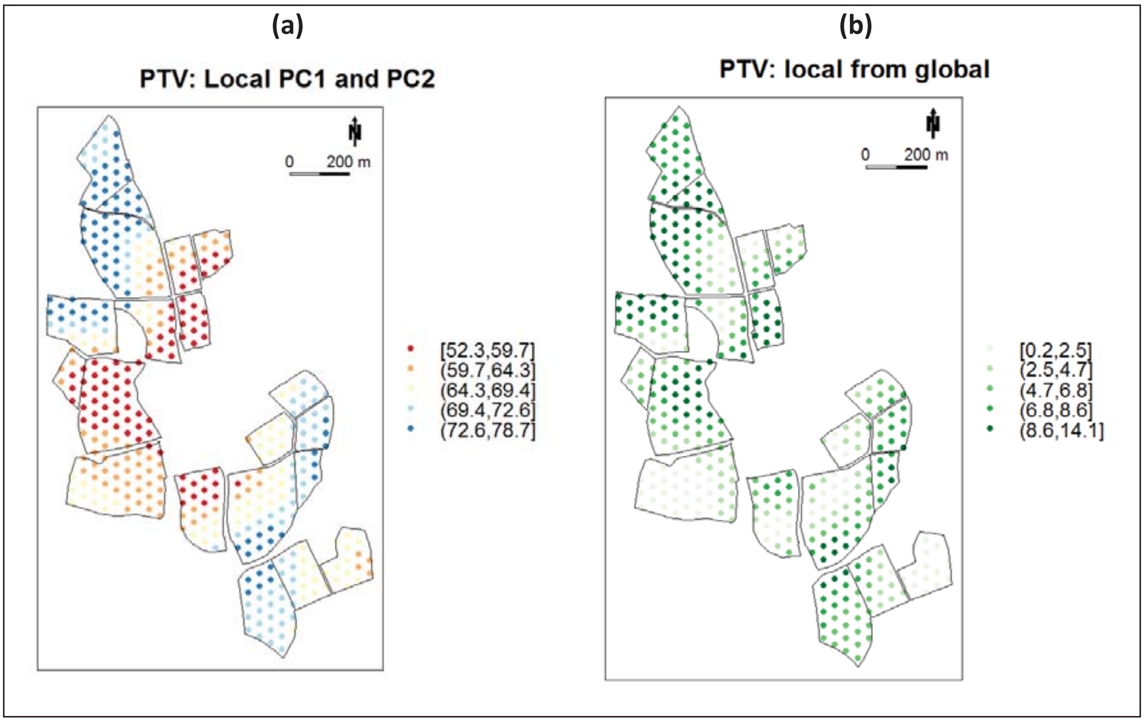

Figure 4: (a) Local PTV data from a GWPCA and (b) absolute difference in PTV from local to global.

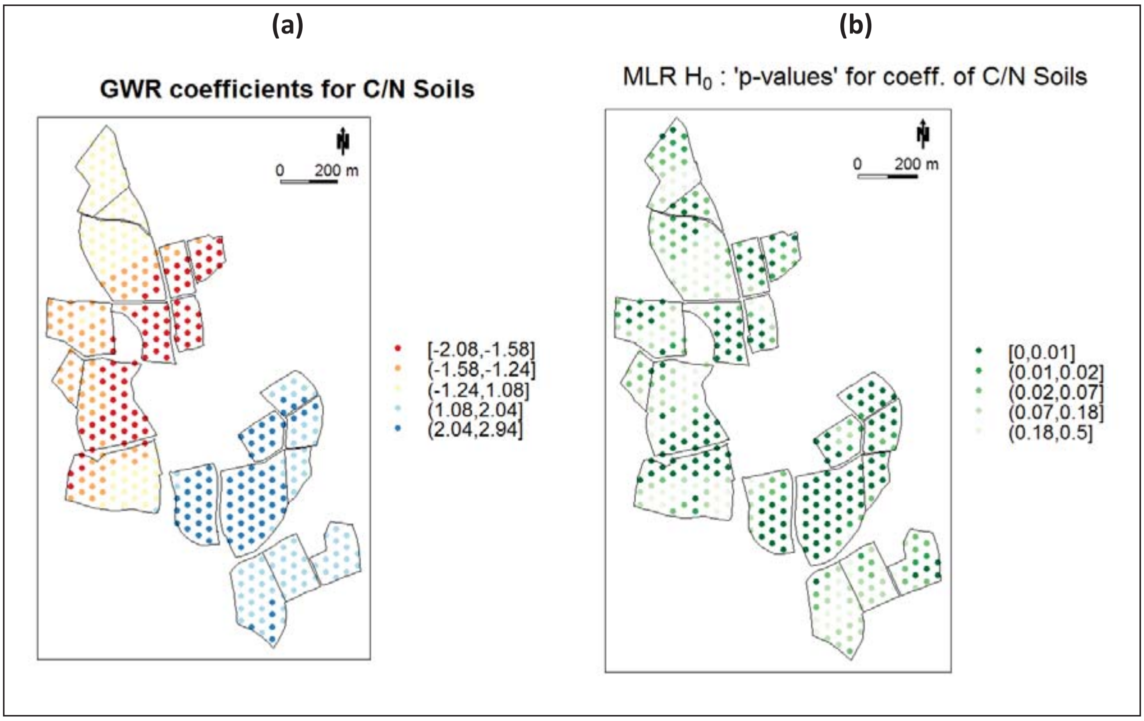

Figure 5: (a) GWR coefficients for C/N ratio for soils and (b) corresponding 'p-values'.

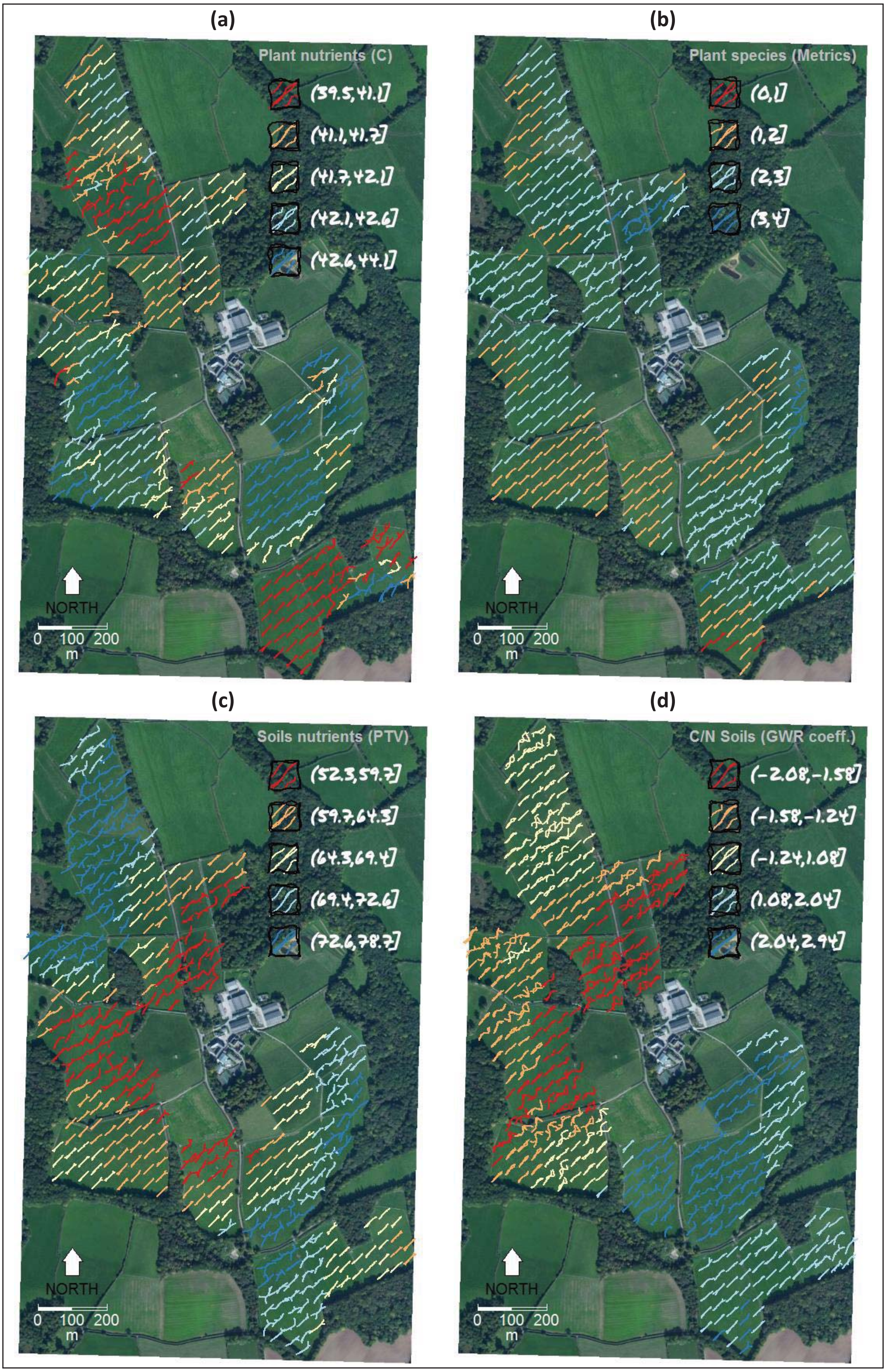

Figure 6: 'Sketchy' maps for (a) Fig. 2; (b) Fig. 3; (c) Fig. 4; and (d) Fig. 5.