REVERSE ENGINEERING OF BIOCHAR

Verónica L. Morales[a,1,*], Francisco J. Pérez-Reche[b], Simona M. Hapca[a], Kelly L. Hanley[c], Johannes Lehmann[c], Wei Zhang[d]

[a] *SIMBIOS Centre, Abertay University, Dundee DD1 1HG, UK*
[b] *Institute for Complex Systems and Mathematical Biology, SUPA, Department of Physics, University of Aberdeen, Old Aberdeen AB24 3UE, UK*
[c] *Department of Crop and Soil Sciences, Cornell University, Ithaca, NY 1453, USA*
[d] *Department of Plant, Soil and Microbial Sciences; Environmental Science and Policy Program, Michigan State University, East Lansing, MI 48824, USA*

Abstract

This study underpins quantitative relationships that account for the combined effects that starting biomass and peak pyrolysis temperature have on physico-chemical properties of biochar. Meta-data was assembled from published data of diverse biochar samples (n=102) to (i) obtain networks of intercorrelated properties and (ii) derive models that predict biochar properties. Assembled correlation networks provide a qualitative overview of the combinations of biochar properties likely to occur in a sample. Generalized Linear Models are constructed to account for situations of varying complexity, including: dependence of biochar properties on single or multiple predictor variables, where dependence on multiple variables can have additive and/or interactive effects; non-linear relation between the response and predictors; and non-Gaussian data distributions. The web-tool *Biochar Engineering* implements the derived models to maximize their utility and distribution. Provided examples illustrate the practical use of the networks, models and web-tool to engineer biochars with prescribed properties desirable for hypothetical scenarios.

Keywords: physico-chemical properties, slow-pyrolysis, correlation networks, Generalized Linear Models, web-tool

*Corresponding author.
Email address: morales@ifu.baug.ethz.ch (Verónica L. Morales)
[1] Present address: Inst. of Environmental Engineering, ETH Zürich, Zürich 8093, CH

1. INTRODUCTION

Biochar, the product of biomass thermochemical conversion in an oxygen depleted environment, has gained increasing recognition as a modernized version of an ancient Amerindian soil management practice, with at times wide-ranging agronomic and environmental gains (Lehmann et al., 2003; Atkinson et al., 2010; Novak and Busscher, 2013). Some of the most commonly acclaimed benefits of biochar application to soils include: increased long-term C storage in soils (Atkinson et al., 2010; Joseph et al., 2010; Cross and Sohi, 2011; Ennis et al., 2011; Karhu et al., 2011; Novak and Busscher, 2013), restored soil fertility (Glaser et al., 2002; Lehmann et al., 2003; Gaskin et al., 2008; Novak et al., 2009; Atkinson et al., 2010; Laird et al., 2010; Beesley et al., 2011; Lehmann et al., 2011; Enders et al., 2012; Spokas et al., 2012b; Novak and Busscher, 2013), improved soil physical properties (Novak et al., 2009; Joseph et al., 2010; Ennis et al., 2011; Karhu et al., 2011; Lehmann et al., 2011; Novak and Busscher, 2013), boosted crop yield and nutrition (Novak et al., 2009; Major et al., 2010; Lehmann et al., 2011; Rajkovich et al., 2012; Spokas et al., 2012a; Novak and Busscher, 2013), enhanced retention of environmental contaminants (Cornelissen et al., 2005; Loganathan et al., 2009; Cao and Harris, 2010; Beesley et al., 2011), and reduced N-emission and leaching (Spokas et al., 2012b; Novak and Busscher, 2013). Examples of the specific biochar properties responsible for these benefits are summarized in Table 1.

Biochar quality can be highly variable, and its performance as an amendment – whether beneficial or detrimental– is often found to depend heavily on its intrinsic properties and the particular soil it is added to (Lehmann et al., 2003; Novak et al., 2009; Atkinson et al., 2010; Major et al., 2010; Lehmann et al., 2011; Spokas et al., 2012a). As has been previously concluded, biochar application to soil is not a "one size fits all" paradigm (Spokas et al., 2012a; Novak and Busscher, 2013). Consequently, detailed knowledge of the biochar properties and the specific soil deficiencies to be remediated is critical to maximize the possible benefits and minimize undesired effects of its use as a soil amendment. While soil deficiencies must be identified on a site-by-site basis, it is conceivable that biochar properties can be engineered through the manipulation of pyrolysis production parameters and proper selection of parent biomass type (Zhao et al., 2013). The capacity to produce biochars with consistent and predictable properties will, first, enable efficient matching of biochars to soils, and second, facilitate the deployment of this soil management strategy at large and commercial scales. Although the properties and effects of biochar samples produced from a variety of methods and starting biomasses have been intensively studied, as yet, the analytical techniques for characterization and effect quantification are not standardized. This creates a challenge when comparing biochar properties and effects across studies. At the same time, making such comparisons is imperative to gain a comprehensive understanding of alterable biochar properties.

The prevailing hypothesis in the literature is that the selection of peak pyrolysis temperature and parent biomass –as two key production variables– fundamentally affects resulting biochar properties. Identification of relationships between production variables and biochar properties has been pursued by many investigators, but has been limited to the small number of samples produced and analyzed for each study (e.g., Karaosmanoğlu et al., 2000; Zhu et al., 2005; Gaskin et al., 2008; Nguyen and Lehmann, 2009; Cao and Harris, 2010; Joseph et al., 2010; Keiluweit et al., 2010; Cao et al., 2011; Cross and Sohi, 2011; Hossain et al., 2011; Mukherjee et al., 2011; Enders et al., 2012; Rajkovich et al., 2012; Zhao et al., 2013), with few reports combining measurements from more than one source (Cordero et al., 2001; Glaser et al., 2002; Atkinson et al., 2010; Ennis et al., 2011; Spokas et al., 2012a). The knowledge gained from the above studies does not provide a quantitative understanding of the relationships between production variables and biochar properties. The shortcomings responsible for such lack of systematic insight include: (i) reported trends that are primarily qualitative with respect to the independent effect of parent biomass or temperature (e.g., decrease in labile carbon with increasing pyrolysis temperature for selected samples (Cross and Sohi, 2011)), (ii) trends that are often in conflict with similar samples of other studies (e.g., positive effect (Rajkovich et al., 2012) vs. negligible effect (Nguyen and Lehmann,

2009) of temperature on pH for oak biochar), and (iii) correlations that are not convincing (e.g., correlation $r = 0.5$ between volatile matter content and microporous surface area (Mukherjee et al., 2011)). A recent study by Zhao et al. (2013) reports, for the first time, a quantitative evaluation of the individual influence of feedstock source and production temperature on various biochar properties. The authors classified a variety of physical and chemical biochar properties as predominantly controlled by either feedstock or temperature. While this initial knowledge is critical to guide the production of designed biochar, it falls short when the influence of both parameters is significant, as is the case with most properties of interest.

The present study advances the quantitative approach one step further by constructing relationships that capture the combined influence that starting biomass and temperature has on various biochar physico-chemical properties of agronomic and environmental interest. The first objective was to gather comparable data from various sources to create an unbiased meta-data set on which to perform statistical analyses. The second objective was to identify groups of inter-correlated properties to gain an insight into how individual properties may be affected when others are manipulated. The third objective was to underpin quantitative relationships between production variables and the measured properties of biochar in the meta-data, as listed in Table 1. The fourth objective was to implement the identified relationships in a simple-to-use web application, which provides an estimate of the expected properties of biochar when produced under a user-defined set of production variables. The overarching goal is to improve the efficiency in production of biochar with engineered properties so that it can best match the needs of a particular soil or crop system.

2. MATERIALS AND METHODS

2.1. Assembly of meta-data library

A library of meta-data (summarized in Table A.1) was created using information from 102 different biochar samples measured for 22 unique physical and chemical characteristics. To build the library, data were gathered from published studies that: (i) used slow-pyrolysis biochar, (ii) reported the production details, and (iii) extensively characterized the physical and chemical properties of biochar materials (Karaosmanoğlu et al., 2000; Cordero et al., 2001; Gaskin et al., 2008; Keiluweit et al., 2010; Mukherjee et al., 2011; Enders et al., 2012; Rajkovich et al., 2012). Production variable details for each study are summarized in Table 2. These studies were chosen because the analytical methods for characterization were similar, thus permitting the comparison of data across studies. Based on these selection criteria, we focused our efforts to test the effects of starting biomass and peak pyrolysis temperature on each of the 22 biochar characteristics. It is important to note that although additional pyrolysis production parameters varied among the samples in our meta-data, the distribution of these variables was too skewed or not documented in a sufficient number of studies to adequately test their effect.

2.2. Correlation matrix and networks

For the first statistical analysis, a correlation matrix was built to identify the links among the physical and chemical properties of biochar in this study (see Fig. 1). To construct the correlation matrix, the Pearson product-moment correlation coefficient between each pair of variables was determined using all complete pairs of observations on those variables. Significance of the relationships was simultaneously determined with a confidence interval of 0.95. Absolute value of correlation and its significance (p-values denoted by star symbols) are reported in the matrix. A threshold for the absolute value of correlation coefficient, $|r|$, of 0.75 was arbitrarily chosen to resolve sufficiently strong relationships. The correlation matrix gives a great deal of information that is not always easy to interpret. In order to visualize the most relevant details, we identified the significant and strong enough correlated pairs of properties, and made a network graph representation (see Fig. 2). The nodes of the graph represent the biochar properties and edges are drawn between pairs of nodes if the properties are strongly correlated and the relationship is significiant ($|r| \geq 0.75$ and p-value < 0.001). Edge thickness in the network graph is proportional to the correlation strength between node pairs. From the correlation networks it is further possible to classify biochar properties into interdependent groups or as independent properties. Alternative network graph representations built with different correlation coefficient thresholds can be obtained from the web-tool, as described in subsequent sections. The authors note that the only difference between network representations of different correlation coefficient thresholds is the number of connections which are displayed, meaning that weak correlations are filtered out in order to ease analysis of network properties that are generally obscured by the complexity of the complete (i.e., unfiltered) network.

2.3. Generalized Linear Model analyses

To accommodate for the different relationships between biochar properties and production variables, a Generalized Linear Models (GLMs) approach was used. GLMs are an extension of ordinary linear regression analysis that account for non-Gaussian distributions of the response as well as non-linear dependencies between explanatory and response variables (the interested readers are referred to Myers et al. (2010) for greater details). When there is a non-linear relation between the response and predictor, GLMs can be used by applying a transformation to the response variable before fitting the model. The other possibility consists in modelling the non-linear dependence by means of a non-linear link function.

2.3.1. GLM candidates

The following steps have been used to build GLMs for the biochar system:

(a) In this study, the response variables are the biochar properties listed in Table 1. The predictors correspond to the production variables which are parameterized by the pyrolysis peak temperature (T: 250-650 °C) and details about the starting biomass, which can be introduced in the model by two categorical variables. A first variable denoted as biomass (B) contains the categories: bull manure, corn, dairy manure, digested dairy manure, food waste, grass, hazelnut, oak, paper waste, pine, poultry litter, and rapeseed. The second variable corresponds to a nested category for B referred to as feedstock class (F), and contains the categories: animal waste, plant material, or combination. Variable T was introduced as covariate in the model, while B and F were introduced as factors.

(b) Under GLMs, the response is assumed to follow a probability density function $p(Resp|X)$ belonging to the exponential family (Myers et al., 2010). In this study the Gaussian and Gamma distributions were initially investigated. However, the Gamma distribution did not show a good fit for any of the response variables and therefore it will not be presented here. Instead, where the response variables did not meet the criteria for a Gaussian distribution, transformation of the response using the Log transform and the Box-Cox transform was applied. As a result, the data distributions we have investigated include (untransformed) Gaussian and two power-transformations for non-Gaussian data (Log transformed and Box-Cox transformed) to describe the biochar system.

(c) A linear relation between the response (biochar property) and the predictors (production variables) of the form

$$g(E(y_i)) = \beta_{i0} + \sum_{j=1}^{N_c} \beta_{i,j} x_{i,j} + \sum_{j=1}^{N_c} \sum_{k=1}^{N_c} \beta_{i,jk} x_{i,j} x_{i,k} \ , \tag{1}$$

is assumed, where $E(y_i)$ signifies the expected values of the i-th response, N_c is the number of predictors, $x_{i,j}$ are the values of the predictor variables (dummy values are used for categorical predictors), and $g(\cdot)$ is the link function. In particular, the link functions *identity* and *log* were explored for all models. The β quantities are unknown parameters to be estimated by maximum-likelihood. The first contribution, β_{i0}, is referred to as the intercept. The parameters $\beta_{i,j}$ quantify the effects of individual variables, while the parameters $\beta_{i,jk}$ account for combined effects associated with interacting pairs of variables. The predictor variables were assessed in all possible individual (B, T, F) and interacting ($B{:}T$, $F{:}T$) combinations. That is, possible formulas relating biochar property (*Resp*) to temperature (T). starting biomass (B) and feedstock class (F) include: Resp $\sim T$, Resp $\sim B$, Resp $\sim B+T$, Resp $\sim B:T$, Resp $\sim B+B:T$, Resp $\sim F$, Resp $\sim F+T$, Resp $\sim F:T$, Resp $\sim F+F:T$.

With all the available options, 54 iterations of GLM models (covering 9 formula possibilities, 3 data transformations, and 2 link functions) were tested to describe each biochar property. These options provide the extra flexibility in the model to describe the biochar system with alternative data transformations and link functions that are not included in ordinary linear regression models, which are limited to Gaussian $p(Resp|X)$ and *identity* $g(\cdot)$.

2.3.2. "Best" model selection and goodness-of-fit tests

The process of "best" model selection requires, first, grouping the GLMs by initial data transformation type: untransformed, Log transformed, and Box-Cox transformed. Quantitative diagnostics were determined for each model, including Akaike Information Criterion (AIC) as an estimate of the quality of a model relative to the collection of candidate models for the data, Shapiro-Wilk (SW) test to determine whether the sample came from a Normally distributed population, and Durbin-Watson (DW) test to detect autocorrelation in the residuals. Within each transformation group, the different model formulations and the different link functions were ranked by the individual model's AIC score. The model with the lowest AIC was then selected as the top candidate model in its group. This step reduces the list of candidate models from 54 to 3, one for each transformation type.

In the second step, the three candidates belonging to each data transformation group were compared against each other. To do this, diagnostic plots were generated for each candidate model, including: (i) residual plots to illustrate the distance of the data points from the fitted regression, (ii) Normal Quantile-Quantile plots to graphically compare the probability distribution of the data against a theoretical Normal distribution, (iii) square root of standardized residual plots to check for heterogeneity of the variance, and (iv) leverage with Cook's distance to identify outliers and points with disproportionate influence on regression estimates. Outlier points were removed from a data set only when the Cook's distance of a datum exceeded 0.5 and re-evaluation of the model did not result in new points with large Cook's distance. Performance of the candidate models for *SW* and *DW* tests, together with the diagnostic plots were used as goodness-of-fit tests to evaluate the assumptions of the models.

The following criteria were used to assess model adequacy. The residual plot was checked for a random scatter of points producing a flat-shapped trend to verify that the appropriate type of model was fitted. The Normal Quantile-Quantile plot was assessed for deviation from the theoretical distribution to confirm Normality in the residuals. The standardized residual plot was examined for a symmetric scatter and flat-shapped trend to test the homogeneity of the variance. The leverage plot was inspected for influential outliers when points fell far from the centroid or were isolated. *SW* quantitatively tested for assumptions of Normality (p-value ≥ 0.05), while *DW* evaluated the level of uncorrelation of the residuals (p-value ≥ 0.05). The "best" model was finally selected as that which satisfied the most criteria, preferring the simpler data transformation if diagnostics were comparable. All computations were performed using RStudio, version 0.96.331.

2.4. Interactive web-tool

The interactive web application *Biochar Engineering* (available at: http://spark.rstudio.com/veromora/BiocharEng/) was built to implement the GLMs constructed in this study into a user-friendly tool, which requires no prior knowledge of advanced statistics or programming language. It is accessible free of charge through a web browser as a stand-alone application hosted by Shiny-RStudio. The primary intention of the tool is to maximize the utility of the models herein developed so that anyone can use them to obtain a statistical outlook for expected physical and chemical properties of biochar from user-defined production values. As is demonstrated in examples to follow, the tool can be used to make informed decisions of the optimum selection of parent biomass type and peak pyrolysis temperature that is required to produce biochars with tailored physical and chemical properties.

3. RESULTS AND DISCUSSION

3.1. Correlation matrix and networks

Related biochar properties identified from the correlation matrix (Fig. 1) were used to build a network representation of the 22 responses included in this study (Fig. 2). From the generated networks, three groups of interdependent biochar properties were distinguished and five individual properties found to be independent (i.e., the correlation coefficient between any pair of properties was $|r| < 0.75$). As illustrated in Fig. 2, the first correlated group includes Fe, Yield, Ash, Ca, C, FixedC, and SSA(CO_2), which contains a mixture of positively and negatively correlated pairs. The second group includes EC, Na, P, K, Mg, Mn, Zn, and S, which contains all positive correlations (linked by solid edges). The third group includes C:N and pH_w, which are negatively correlated (linked by dashed edges). The five independent properties are represented as edge-free nodes and include BulkD, SSA(N_2), N, MatVol, and CEC. Interestingly, SSA(N_2) and CEC were found to have mostly very weak and insignificant relationships with all other biochar properties ($|r| \leq 0.53$ with p-value ≥ 0.01 and $|r| \leq 0.44$ with p-value ≥ 0.001, respectively). The exception for CEC is its relationship with BulkD, which is significant albeit still weak ($|r| = 0.58$ with p-value < 0.001). As a result, SSA(N_2) and CEC could be considered the two most independent biochar properties, which are the least likely to be affected when other properties are modified. It is noted that Principal Component Analysis (analyzed with SPSS v.21) was initially explored to find clusters of biochar properties. However, the meta-data contained too many samples that were not characterized in full, thus producing an incomplete matrix that required the omission of a vast number of samples or of entire response variables from the analysis. As these omissions were considered to affect the results excessively, a correlation matrix and network approach was adopted being considered less biased by missing data.

The networks of correlated properties provide an overview of which combinations of biochar properties are more likely to occur in a given sample. The correlation networks prove very useful as a tool for qualitative design of biochar samples with desired properties. For example, a hypothetically desirable biochar might be needed to neutralize soil acidity (high pH_w), return lost macronutrients P and S that were removed during harvest (high P and S), prevent excess atrazine from leaching into the groundwater (high SSA(CO_2) and/or high Ash), and maximize the amount of biochar produced by pyrolysis (high Yield). Using the network diagram of Fig. 2, it is possible for example to infer the following. A biochar sample engineered for high pH_w will not affect the other desired properties, given that pH_w is in a separate network to all other properties of interest. The addition of macronutrient P will concomitantly supply S, as these properties belong to the same positively correlated network. The remaining three properties belong to the same network from which we extrapolate that a single sample of biochar has a negative tradeoff between high SSA(CO_2) and high Ash[2], meaning that it is less probable that a sample will have both high SSA(CO_2) and high Ash. Yield will be reduced if the sample is prioritized for high SSA(CO_2) and (indirectly) maximized when high Ash content is favored. Networks obtained from different correlation coefficient thresholds can be created in the web-tool as displayed in the *Networks* tab and interpreted in the fashion described above. Increasing the correlation coefficient threshold will simply result in the removal of weak connections from the final graphic, while decreasing it will result in the display of more connections.

3.2. Generalized Linear Models

In this section the versatility of GLMs as an extended linear regression approach is leveraged to model the biochar system. The candidate GLMs are compared against one another and the most appropriate models for each biochar property selected. Lastly, the "best" models are evaluated for goodness-of-fit.

[2]While SSA(CO_2) is not directly linked to Ash, high SSA(CO_2) implies high C and FixedC which, in turn, are negatively correlated with Ash. In other words, SSA(CO_2) and Ash are indirectly anti-correlated.

3.2.1. GLM candidates

As indicated in the methods section, selection of the "best" model is a two-step process. First, the list of candidates is reduced to three. To do so, candidate models belonging to each of the three data transformation groups (untransformed, Log transformed and Box-Cox transformed) are ranked according to their AIC score. Top scoring models for each group are those with the lowest AIC value, and are reported in tables for each biochar property in section II of the supplementary data. The tables summarize the top candidate model for each data transformation group, where details of the model are reported concerning: formula, type of data transformation used, link function, AIC, p-value for the SW test, as well as d and p-value for the DW test. Second, diagnostic plots are generated for the reduced candidate list, and the overall "best" model is selected according to their relative performance in SW and DW tests and diagnostic plot criteria. Diagnostic plots of the overall "best" model are included in the same section of the supplementary data, and noted by a star in the table.

Model selection required a certain level of flexibility, as very few candidate models met all evaluating criteria. This is a common feature of real data sets of a limited size. Model performance in the SW test was relatively poor, since candidate GLMs of 15 of the biochar properties failed SW for all types of data transformation. Nevertheless, candidate GLMs of the remaining biochar properties consistently satisfied this criterion for the overall "best" model. Performance in DW was useful in quantitatively evaluating the assumption for uncorrelated residuals, but not to differentiate the candidate GLMs against each other because often all candidates satisfied or failed this criterion. Diagnostic plots, on the other hand, were much more insightful in illustrating the suitability and relative performance of the models, and were given more consideration during "best" model selection.

In general, all four diagnostic plots corresponding to one candidate model performed well above the other two, and demonstrated that the goodness-of-fit (GOF) assumptions were satisfactorily met. For certain biochar properties two candidate models produced diagnostic plots of similar performance, in which case the model corresponding to the simpler data transformation was given preference; that is, untransformed is simpler than Log transformed, which is simpler than Box-Cox transformed. In the case of Na, for example, diagnostic plots for Log and Box-Cox transformation GLMs showed a nearly identical model improvement (see Figs. A.15 and A.16), and all three candidate models performed the same for *SW* and *DW* (see Table A.16). Consequently, the Log transformed model was selected as the "best" model. The models for Fe, N, and SSA(N_2) were difficult to select given the pronounced heterogeneity in variance and heavy deviation from the theoretical Normal Quantile-Quantile distribution across all candidate models (see Fig. A.8, A.14 and A.21). These three models were therefore considered to violate too many GOF criteria to be recommended for use with confidence; the situation would improve with additional data. Irrespective of that, the large proportion of properties found to be properly described by the corresponding "best" model clearly demonstrates the feasibility of reverse engineering multiple biochar properties simultaneously. We note that initial analysis with fewer samples comprising the meta-data resulted in the selection of "best" models with satisfactory GOF criteria that were very similar to those chosen from the larger data set (presented in Table 3). This indicates that replication of suitable results (i.e., those that comply with GOF standards) from different studies are consistent.

Table 3 summarizes the "best" models chosen for all biochar properties, where the last column indicates whether the model complies with GOF standards. The Maximum Likelihood Estimates (MLEs) of the "best" model coefficients for each biochar property are reported in section III of the supplementary data and can be requested from the web-tool in the *Stats* tab.

3.2.2. "Best" GLMs

The formulas of the "best" models (column 2 in Table 3) indicate that for the vast majority of cases it is imperative to have information about both starting biomass and peak pyrolysis temperature to properly define the relationship between biochar properties and production variables. In the simplest case a single predictor variable statistically dominates. We find that this only occurs for S, which depends entirely on B, while T is not statistically significant (as shown in Fig. 3A). No response variable was found to depend exclusively on T. The next level of complexity is that in which the response depends on both B and T, but the two factors do not interact ($B + T$). This occurs for pH_w, Ash, C:N, and most micronutrients. In this type of relationship, B affects the response, but the rate at which T has an influence is the same across all types of B (illustrated in Fig. 3B). The following level of complexity is that in which there is a significant interaction between B and T, but no main effect of B ($B : T$), as in the case for SSA(CO_2) and FixedC. A general trend in this type of relationship is that the rate of change in the response with the increase in T is different for the different B, whereas the intercept is the same (as shown in Fig. 3C). Finally, the most complex relationship is given by the full model (B+B:T or F+F:T). In this model, both intercept and temperature regression slope are significantly different for the different B (or F). The relationships for BulkD, SSA(N_2), Yield, EC, CEC, MatVol, C, N, P, Ca, and K fall into this category. In this case, changes in B (or F) and T are not trivial, as the relationship permits the greatest level of flexibility and rules out any general trends (as in Fig. 3D).

For the three simplest relationships (B, B+T, and B:T), a change in B does not affect the response order relative to the other types of B. Conversely, for the most complex relationship (B+B:T or F+F:T), a change in biomass affects the response in such a way that it crosses over responses from other biomass types as T changes; thereby not necessarily maintaining the relative order among the different types of biomass. This assessment of multiple predictor variable influence corroborates the perception that biochar properties are deeply shaped by the collective effect of both production variables, whether additive and/or interactive. Furthermore, it warrants against statistical bias that is introduced when biochar production decisions are based on the dominance of a single variable on a biochar property of interest. Interestingly, only the "best" model for MatVol favored the nested starting biomass, F. All other "best" models performed better when this information was entered in its more detailed form, B.

The frequency in response variable transformation for the selected "best" models (column 3 in Table 3) indicates that a minority of the data are Normally distributed and meet the constant variance assumption. Most responses require power-transformation to stabilize their variance. Specifically, 7 response variables were satisfactorily modeled without transformation of the response values, while 9 others needed Log transformation and the remaining 6 required the more advanced Box-Cox transformation. This observation draws attention to the fact that non-constant variance is ubiquitous in the characteristics of biochar, which requires transformation of the response variable to comply with Normality assumptions. Depictions of different functional shapes are presented in Fig. 4 for models sharing the same formula ($B+T$) and *identity* link. In this figure, (A) is the reference for the untransformed response for pH_w, (B) is the Log transformed response for Mn, and (C) is the Box-Cox transformed response for Ash. In these plots, it is evident that the untransformed data have a perfectly linear relationship. In contrast, Log and Box-Cox transformations are suitable to describe non-linear behavior associated with a more cumbersome relationship between biochar properties and production variables.

Similarly, the prevalence of non-linear link functions in the "best" model population (column 4 in Table 3) exposes the common violation of the linearity assumption. It is interesting that all 7 responses that demonstrated constant variance (i.e., not requiring data transformation) also met the linearity assumption (favoring *identity* link function). This was also the case for 8 of the responses with unequal variances that required data transformation. The remaining 7 responses required transformation to address variance instability and the *log* link function to further correct for non-linearity. The *log* link function contributes to the non-linear function shape of the response in a way that resembles that of Log and Box-Cox data transformation. Fig. 4 illustrates this effect for responses that have been Log transformed. The data in (B) satisfies the linearity assumption and is adequately modeled with the *identity* link function. In contrast, the property in (D) needs a *log* link function to adjust for non-linearity. In short, both non-Gaussian and non-linear features were found to be ubiquitous in the biochar system.

3.3. Biochar Engineering*: the web-tool*

The *Biochar Engineering* tool is an integrated calculator for the biochar models in Table 3. The web-tool can be navigated through the various tabs on display at the top of the page. The *About* tab introduces the tool, the *Graphic* and *Table* tabs contain the model results, the *Stats* tab summarizes individual model parameters, and the *Networks* tab displays networks of correlated biochar properties. The side bar panel is always visible and can be modified at any time to re-run the model with new input variable values for biomass, peak temperature, and confidence coefficient, request the statistical summary of a specific response model, set a correlation coefficient cutoff for the networks, and download the output of any tab. The model output for the user-defined production variables is automatically generated and updated in the *Graphic* and *Table* tabs. Correlation networks are similarly updated in the *Networks* tab for newly defined correlation coefficients. Ultimately, this information can be used to select production variable values that yield biochar with the most desirable set of properties for the user, thereby facilitating the possibility to efficiently engineer biochar resources to meet multiple agricultural demands.

3.4. Using GLMs and web-tool to engineer a biochar

Recommendations for the use of the GLMs in Table 3 cannot be generalized because they depend on the particular set of properties needed from biochar to mitigate deficiencies in a specific soil or crop, as well as on the type of biomass available and limitations of the pyrolysis unit. Rather than attempting to examine all possible scenarios, this section presents two examples that demonstrate how the GLMs and the web-tool can be used to engineer the hypothetical biochar described in section 3.1 (requiring high pH_w, high P and S, high $SSA(CO_2)$ and/or high Ash, and high Yield). In the first example we assume a situation where all production variables can be modified, and identify the optimum combination of starting biomass and temperature that return the desired qualities. In the second example we assume a situation where the type of starting biomass is fixed (e.g., to concurrently dispose of a byproduct from another process), and determine the temperature that is most suitable to obtain the desired qualities.

3.4.1. A worked example for total optimization of production variables

In the case where all production variables can be modified, we propose to refer to the prediction plots corresponding to the properties of interest. Prediction plots for all properties analyzed in this study are included in Fig. A.24-A.45 of the supplementary data; see the particular case for pH_w in Fig. 5. To facilitate interpretation of the model results, the predictive plots are presented as composite figures where each subfigure corresponds to a unique type of starting biomass and the property of interest is plotted as a function of pyrolysis temperature. The predicted (mean) values are presented as a solid line, while regions corresponding to 75, 85, and 95% confidence intervals are indicated by the shaded regions (dark gray, gray, light gray, respectively). For reference, the data points from the meta-data are overlaid as solid circles.

We begin by analyzing Fig. 5 to identify the variables that can deliver biochar with high pH_w. This figure shows that as T increases pH_w increases, and this rate is constant across all B. Among the different types of B included in the pH_w model, biochars made from Poultry litter would typically result in the highest achievable pH_w at any T, followed by Digested dairy manure, Corn, Food waste, and Paper waste. Next, we analyze the predictive plot for P (Fig. A.38). From this figure it is apparent that most Bs result in biochars with low P concentrations that are minimally variable with T; crossovers associated with the $B{:}T$ coupling are mainly observed on the low T range. Notably, samples made from Poultry litter contain the highest concentration of P (by orders of magnitude greater than samples of lowest P), with Food waste and Digested dairy manure following significantly behind in P concentration. Then, we examine the predictive plot for S (Fig. A.40), which is exclusively dependent on B (in agreement with the "best" model formula for S in Table 3). It is easy to distinguish that Poultry litter has the highest S content, followed by Digested dairy manure and Dairy manure. Next, we consider predictions for SSA(CO_2) (Fig. A.41), which also show a general increase in response with T at rates that depend on B (cf. formula $B{:}T$ for the "best" SSA(CO_2) model). From these predictions we identify that Hazelnut, Pine and Oak produce the highest possible SSA(CO_2), which is enhanced as T is increased. Conversely, the predictive plot for Ash (Fig. A.24) indicates that this property is typically around 30% and generally increases with T. Paper waste, Poultry litter and Food waste are ranked highest among the B types to show high ash at all T levels. Lastly, the predictive plot for Yield (Fig. A.44) demonstrates a pronouncedly decreasing trend with increasing T for all B types, with crossovers throughout, as expected from the "best" model formula $B{+}B{:}T$ given in Table 3 for Yield. It is evident that biochars from Paper waste and Poultry litter produce the highest yield for the range of T investigated.

Based on the above observations, we conclude that Poultry litter pyrolysed at T above 500°C will return a biochar that meets most of the needed hypothetical properties. More concrete recommendations of T will depend on the producer's choice to compromise between Ash and Yield, which have opposing trends with T. One way to facilitate this decision is to refer to the predictions made by the *Biochar Engineering* web-tool at various temperatures. By specifying in the side bar panel the Biomass (Poultry), Peak Temperature (a value in the range 500-600°C), and a satisfactory Confidence Coefficient (e.g., 0.8), the web-tool automatically generates a table (located in the *Table* tab) that summarizes the expected biochar properties for the input variables. For discrete temperatures at 500, 550, and 600°C, the biochar would be expected to have an Ash content of 56.60, 61.31, and 66.4%, and Yield of 65.76, 64.38, and 63.03%, respectively. Considering that Ash is increased by 10% and Yield is only reduced by 2% when T is increased from 500 to 600°C, one might accept the small penalty in yield for gaining more ash. Assuming all other considerations are satisfactory in this hypothetical scenario, one could conclude that the customized biochar with the above listed characteristics is best produced by pyrolysing Poultry litter at 600°C. For a comprehensive outlook on the expected range of all 22 physico-chemical properties, the user may refer to the output generated in the *Graphic* or *Table* tabs of the web-tool, and save the results with the download buttons for future reference.

3.4.2. A worked example for restrictions in starting biomass

A similar approach to that followed in the first example can be used to engineer a biochar for cases in which the type of biomass is fixed. Take for instance a corn farm, which is interested in selling its corn stover resources as high quality biochar because livestock feed and bioenergy prices are low. The properties required from the biochar, as specified by the client, are assumed to be the same as those for the hypothetical biochar considered above. In this case, the farmer or pyrolysis contractor would be referred to the web-tool directly. In the side bar panel, the Biomass should be set to Corn and a suitable Confidence Coefficient selected (e.g., 0.8). The Peak Temperature slider can then be used to study the changes in biochar properties with temperature, as the only production variable that can be adjusted. The model output results can be monitored in either the *Graphic* tab (bar plots indicate predicted values with error bars marking the confidence interval range) or in the *Table* tab (table summary of predicted values with their corresponding standard error and confidence interval). By shifting the Peak Temperature slider from low to high temperatures it is evident that Yield is diminished, SSA(CO_2), pH_w, Ash, and P are intensified, and S remains constant. Assuming in addition to the required biochar properties that in order to make a profit, the Yield should be at least 30%, we can conclude that the corn stover should be pyrolysed at 467°C, so the lower end of the expected yield range is above 30%. The *Table* tab of the web-tool (see screenshot in Fig. 6) summarizes the expected value and confidence interval for each biochar property, according to the production variables specified. For corn pyrolysed at 467°C, the estimated range (with 80% confidence level) for the desired properties is: 8.6-9.9 pH_w, 1647-2214 Total (mg/kg) P, and 633.1-869.9 Total (mg/kg) S, 330.6-450.6 m^2/g SSA(CO_2), 11.8-16.2% Ash, and 30.0-33.1% Yield.

4. CONCLUSION

Statistical results demonstrate that arbitrary choices of starting biomass or peak pyrolysis temperature are unlikely to produce biochar with prescribed physico-chemical properties. Generalized Linear Models were used to quantify the combined effect that starting biomass and peak temperature has on different biochar properties. These properties are typically non-Gaussian and exhibit non-linear dependence on the two predictor variables. Proper description of most biochar properties by GLMs demonstrates the feasibility to engineer biochar. A web-application of the GLMs together with correlation networks are offered as tools to guide biochar engineering.

Acknowledgements

The authors thank Dr D.R. Fuka and Dr M.P. Allan for advise on programing and statistical methods, and Dr S. Sohi and Dr O. Mašek for valuable discussions. This study was financed in part by the Teresa Heinz Foundation for Environmental Research and Projekt Unicorn. V.L. Morales acknowledges support from Marie Curie International Incoming Fellowships (FP7-PEOPLE-2012-SoilArchnAg).

References

[1] Abe, F., 1988. The thermochemical study of forest biomass. Bulletin of the Forestry and Forest Products Research Institute, Japan 352, 1–95.

[2] Atkinson, C. J., Fitzgerald, J. D., Hipps, N. A., 2010. Potential mechanisms for achieving agricultural benefits from biochar application to temperate soils: A review. Plant Soil 337, 1–18.

[3] Beesley, L., Moreno-Jimenez, E., Gomez-Eyles, J., Harris, E., Robinson, B., Sizmur, T., 2011. A review of biochars' potential role in the remediation, revegetation and restoration of contaminated soils. Environ. Pollut. 159, 3269–3282.

[4] Cao, X., Harris, W., 2010. Properties of dairy-manure-derived biochar pertinent to its potential use in remediation. Bioresour. Technol. 101, 5222–5228.

[5] Cao, X., Ma, L., Liang, Y., Gao, B., Harris, W., 2011. Simultaneous immobilization of lead and atrazine in contaminated soils using dairy-manure biochar. Environ. Sci. Technol. 45, 4884–4889.

[6] Chapman, H., 1965. Cation-exchange capacity. In: Norman, A. (Ed.), Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties. Agronomy monograph, pp. 891–901.

[7] Cordero, T., Marquez, F., Rodriguez-Mirasol, J., Rodriguez, J., 2001. Predicting heating values of lignocellulosics and carbonaceous materials from proximate analysis. Fuel 80, 1567–1571.

[8] Cornelissen, G., Gustafsson, ., Bucheli, T. D., Jonker, M. T. O., Koelmans, A. A., van, Noort, P. C. M., 2005. Extensive sorption of organic compounds to black carbon, coal, and kerogen in sediments and soils: mechanisms and consequences for distribution, bioaccumulation, and biodegradation. Environ. Sci. Technol. 39, 6881–6895.

[9] Cross, A., Sohi, S. P., 2011. The priming potential of biochar products in relation to labile carbon contents and soil organic matter status. Soil Biol. Biochem. 43, 2127–2134.

[10] Enders, A., Hanley, K., Whitman, T., Joseph, S., Lehmann, J., 2012. Characterization of biochars to evaluate recalcitrance and agronomic performance. Bioresour. Technol. 114, 644–653.

[11] Ennis, C. J., Evans, A. G., Islam, M., Ralebitso-Senior, T. K., Senior, E., 2011. Biochar: Carbon sequestration, land remediation and impacts on soil microbiology. Crit. Rev. Env. Sci. Tec. 42, 2311–2364.

[12] Gaskin, J., Steiner, C., Harris, K., Das, K., Bibens, B., 2008. Effect of low-temperature pyrolysis conditions on biochar for agricultural use. T. ASABE 51, 2061–2069.

[13] Glaser, B., Lehmann, J., Zech, W., 2002. Ameliorating physical and chemical properties of highly weathered soils in the tropics with charcoal - a review. Biol. Fertil. Solis 35, 219–230.

[14] Hossain, M., Strezov, V., Chan, K., Ziolkowski, A., Nelson, P., 2011. Influence of prolysis temperature on production and nutrient properties of wastewater sludge biochar. J. Environ. Manage. 92, 223–228.

[15] Joseph, S., Peacocke, C., Lehmann, J., Munroe, P., 2009. Developing a biochar classification and test methods. In: Lehmann, J., Joseph, S. (Eds.), Biochar for Environmental Management: Science and Technology. Earthscan, London, pp. 107–126.

[16] Joseph, S. D., Camps-Arbestain, M., Lin, Y., Munroe, P., Chia, C. H., Hook, J., van, Zwieten, L., Kimber, S., Cowie, A., Singh, B. P., Lehmann, J., Foidl, N., Smernik, R. J., Amonette, J. E., 2010. An investigation into the reactions of biochar in soil. Aust. J. Soil Res. 48, 501.

[17] Karaosmanoğlu, F., Işıgıgür-Ergüdenler, A., Sever, A., 2000. Biochar from the straw-stalk of rapeseed plant. Energ. Fuel. 14, 336–339.

[18] Karhu, K., Mattila, T., Bergstrom, I., Regina, K., 2011. Biochar addition to agricultural soil increased ch4 uptake and water holding capacity - results from a short-term pilot field study. Agr. Ecosys. Environ. 140, 309–313.

[19] Keiluweit, M., Nico, P., Johnson, M., Kleber, M., 2010. Dynamic molecular structure of plant biomass-derived black carbon (biochar). Environ. Sci. Technol. 44, 1247–1253.

[20] Laird, D. A., Fleming, P., Davis, D. D., Horton, R., Wang, B., Karlen, D. L., 2010. Impact of biochar amendments on the quality of a typical midwestern agricultural soil. Geoderma 158, 443–449.

[21] Lehmann, J., Pereira da Silva, J. J., Steiner, C., Nehls, T., Zech, W., Glaser, B., 2003. Nutrient availability and leaching in an archeological anthrosol and a ferralsol of the central amazon basin: Fertilizer, manure and charcoal amendments. Plant Soil 249, 343–357.

[22] Lehmann, J., Rillig, M. C., Thies, J. E., Masiello, C. A., Hockaday, W. C., Crowley, D., 2011. Biochar effects on soil biota - a review. Soil Biol. Biochem. 43, 1812–1836.

[23] Loganathan, V. A., Feng, Y., Sheng, G. D., Clement, T. P., 2009. Crop-residue-derived char influences sorption, desorption and bioavailability of atrazine in soils. Soil Sci. Soc. Am. J. 73, 967.

[24] Major, J., Rondon, M., Molina, D., Riha, S. J., Lehmann, J., 2010. Maize yield and nutrition during 4 years after biochar application to a colombian savanna oxisol. Plant Soil 333, 117–128.

[25] Mukherjee, A., Zimmerman, A., Harris, W., 2011. Surface chemistry variations among a series of laboratory-produced biochars. Geoderma 163, 247–255.

[26] Myers, R. H., Montgomery, D. C., Vining, G. G., Robinson, T. J., 2010. Geleralized Linear Models With Applications in Engineering and the Sciences. John Wiley & Sons, Inc., Hoboken, New Jersey.

[27] Nguyen, B. T., Lehmann, J., 2009. Black carbon decomposition under varying water regimes. Org. Geochem. 40, 846–853.

[28] Novak, J., Busscher, W., 2013. Selection and use of designer biochars to improve characteristics of southeastern usa coastal plain degraded soil. In: Lee, J. W. (Ed.), Advanced Biofuels and Bioproducts. Springer, New York, pp. 69–96.

[29] Novak, J. M., Busscher, W. J., Laird, D. A., Ahmedna, M., Watts, D. W., Niandow, M. A., 2009. Impact of biochar amendment on fertility of a southeastern coastal plain soil. Soil Sci. 174, 105–112.

[30] Rajkovich, S., Enders, A., Hanley, K., Hyland, C., Zimmerman, A. R., Lehmann, J., 2012. Corn growth and nitrogen nutrition after additions of biochars with varying properties to a temperate soil. Biol. Fertil. Soils 48, 271–284.

[31] Sander, M., Pignatello, J. J., 2005. Characterization of charcoal adsorption sites for aromatic compounds: insights drawn from single-solute and bi-solute competitive experiments. Environ. Sci. Technol. 39, 1606–1615.

[32] Spokas, K., Cantrell, K., Novak, J., Archer, D., Ippolito, J., Collins, H., Boateng, A., Lima, I., Lamb, M., McAloon, A., Lentz, R., Nichols, K., 2012a. Biochar: A synthesis of its agronomic impact beyond carbon sequestration. J. Environ. Qual. 41, 973–989.

[33] Spokas, K. A., Novak, J. M., Venterea, R. T., 2012b. Biochar's role as an alternative n-fertilizer: Ammonia capture. Plant Soil 350, 35–42.

[34] Zhao, L., Cao, X., Mašek, O., Zimmerman, A., 2013. Heterogeneity of biochar properties as a function of feedstock sources and production temperatures. J. Hazard. Mater. 256-257, 1–9.

[35] Zhu, D., Kwon, S., Pignatello, J. J., 2005. Adsorption of single-ring organic compounds to wood charcoals prepared under different thermochemical conditions. Environ. Sci. Technol. 39, 3990–3998.

Captions, Figures and Tables

Figure 1. Correlation matrix of biochar properties. The diagonal indicates the biochar properties. The upper triangular sector shows the absolute value of correlation between pairs of properties and significance symbol (defined in the legend). Highly correlated pairs (with $|r| \geq 0.75$) are highlighted in bold font. The lower triangular sector displays the respective bivariate scatterplots with a trend line.

Figure 2. Correlation networks of inter-correlated biochar properties ($|r| \geq 0.75$). Nodes represent individual biochar properties, and edges indicate whether the correlation is positive (solid line) or negative (dashed line). Line thickness is proportional to the correlation strength.

Figure 3. Formula interpretation for GLMs of link *identity*. (A) Resp ~ B. (B) Resp ~ B + T. (C) Resp ~ B:T. (D) Resp ~ B + B:T.

Figure 4. Data transformation interpretation for GLMs of link *identity* and Formula $B{+}T$. (A) Untransformed. (B) Log transformed. (C) Box-Cox transformed. (D) Log transformed of link *log*.

Figure 5. Model predictions for pH_w content (solid line) with confidence intervals for 75, 85, and 95% (dark gray, gray, light gray shading, respectively). Data points from meta-data are overlain (solid circles).

Figure 6. Interface of the *Biochar Engineering* tool. Model output compiled in the *Table* tab.

Table 1. Benefits from specific biochar properties.

Table 2. Production details of meta-data.

Table 3. Summary of "best" models selected for each biochar characteristic.

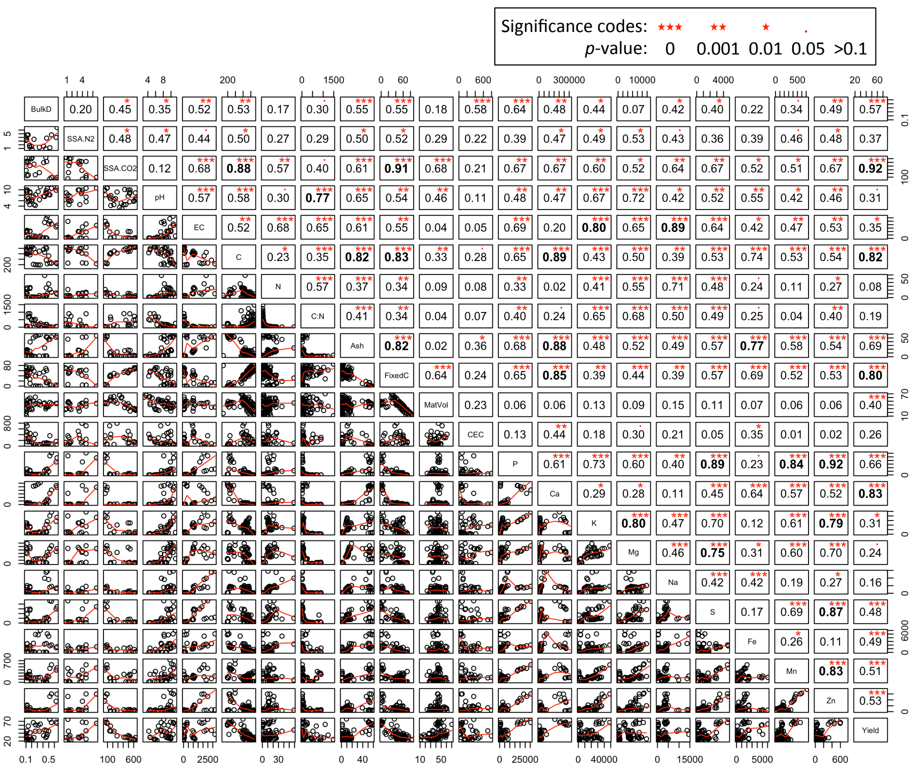

Significance codes: *** ** * .
p-value: 0 0.001 0.01 0.05 >0.1
BulkD
SSA.N2
SSA.CO2
pH
EC
C
N
C:N
Ash
FixedC
MatVol
CEC
P
Ca
K
Mg
Na
S
Fe
Mn
Zn
Yield

Figure 1:

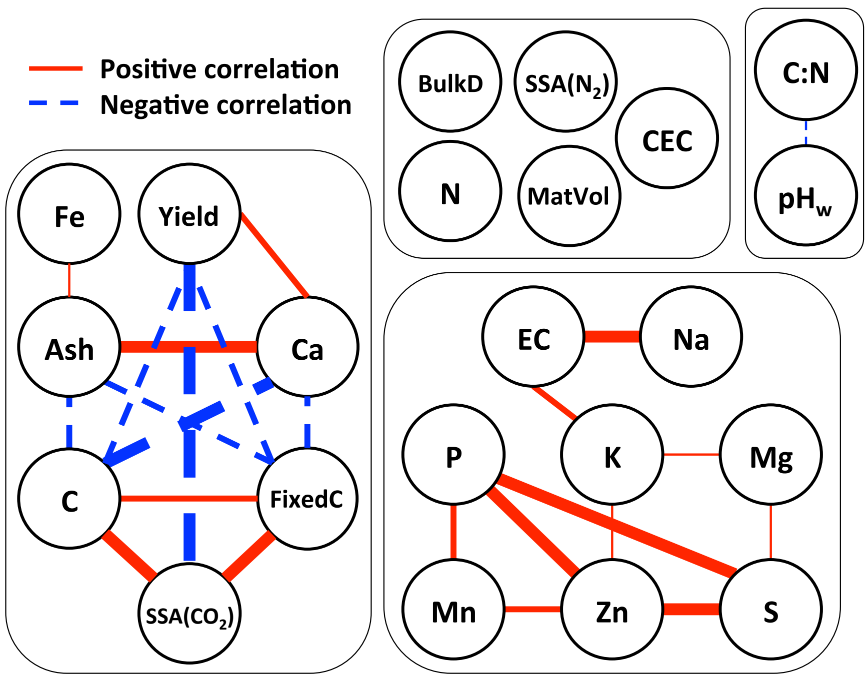
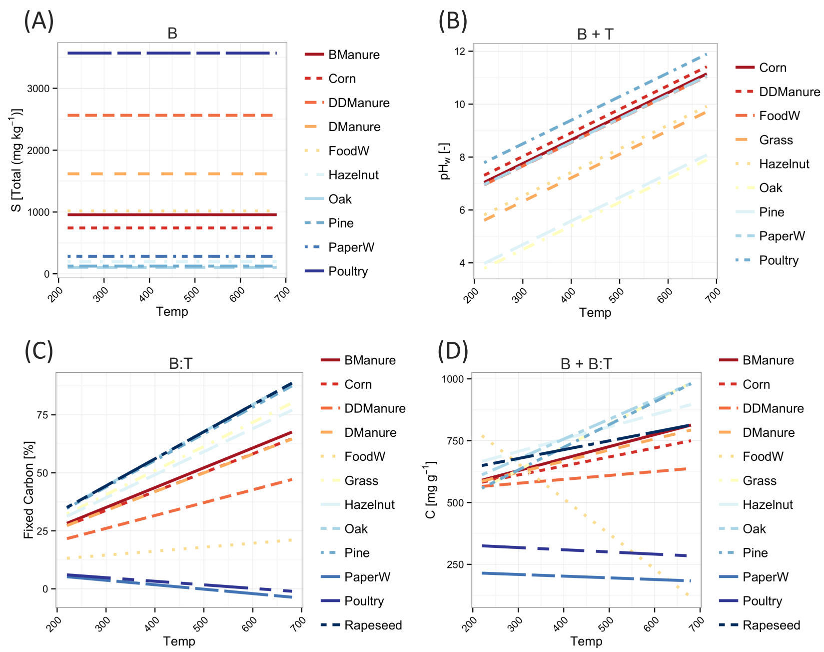
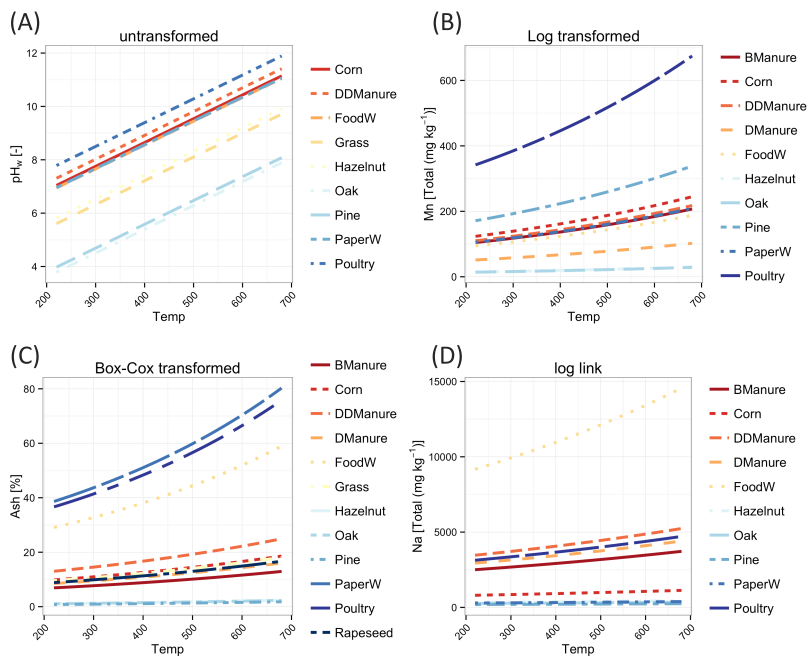

Positive correlation
Negative correlation
Fe
Yield
Ash
Ca
C
FixedC
SSA(CO2)
BulkD
SSA(N2)
CEC
N
MatVol
C:N
pHw
EC
Na
P
K
Mg
Mn
Zn
S

Figure 2:



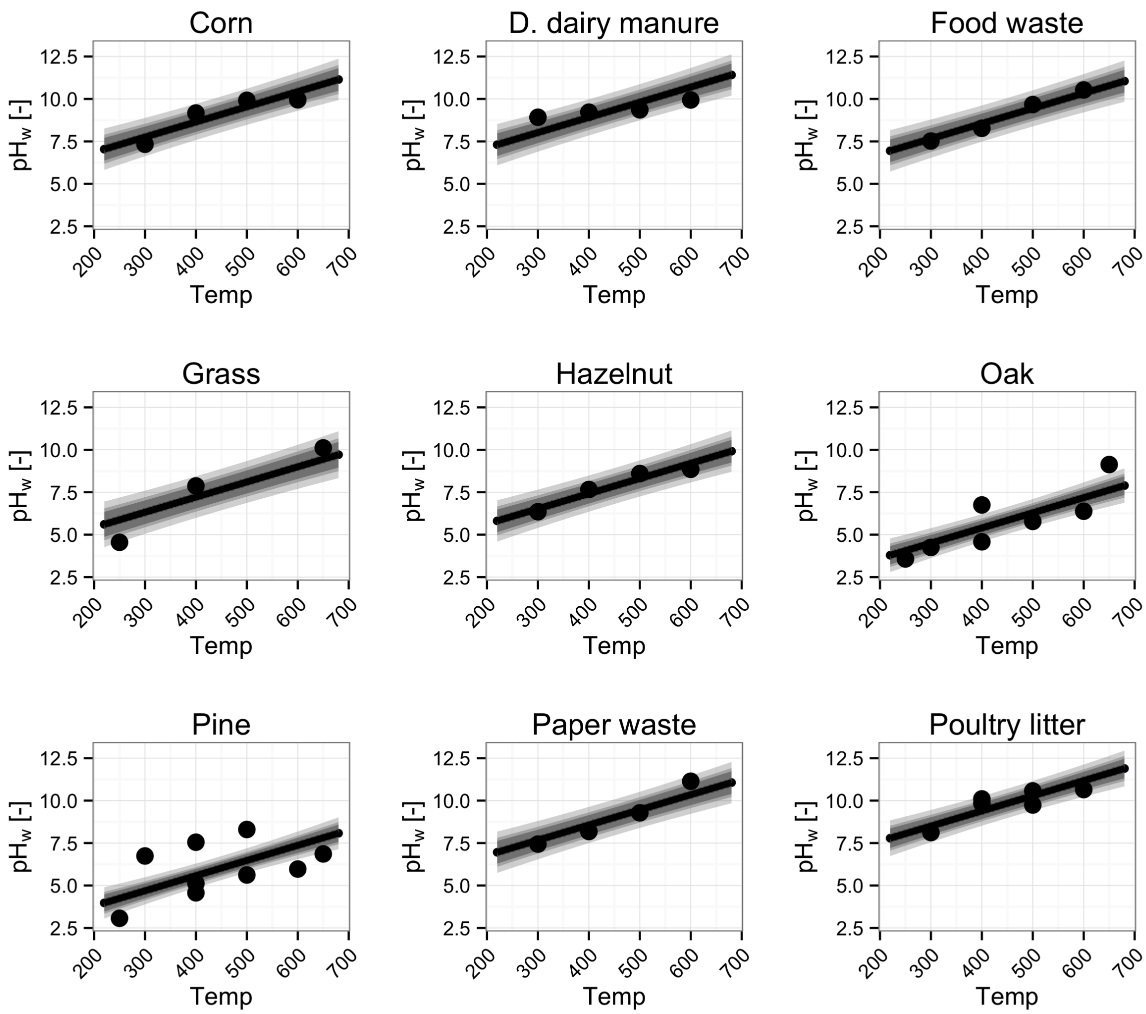
(A)
B
S [Total (mg kg^{-1})]
Temp
BManure
Corn
DDManure
DManure
FoodW
Hazelnut
Oak
Pine
PaperW
Poultry
(B)
B + T
pH$_w$ [-]
Temp
Corn
DDManure
FoodW
Grass
Hazelnut
Oak
Pine
PaperW
Poultry
(C)
B:T
Fixed Carbon [%]
Temp
BManure
Corn
DDManure
DManure
FoodW
Grass
Hazelnut
Oak
Pine
PaperW
Poultry
Rapeseed
(D)
B + B:T
C [mg g^{-1}]
Temp
BManure
Corn
DDManure
DManure
FoodW
Grass
Hazelnut
Oak
Pine
PaperW
Poultry
Rapeseed

Figure 3:



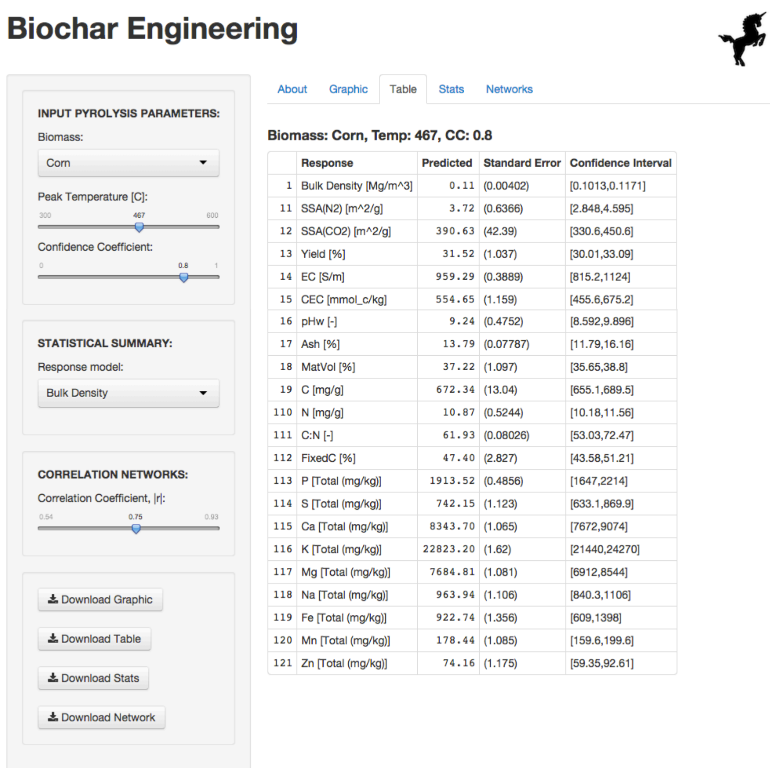
(A)
untransformed
pH$_w$ [-]
Temp
Corn
DDManure
FoodW
Grass
Hazelnut
Oak
Pine
PaperW
Poultry
(B)
Log transformed
Mn [Total (mg kg^{-1})]
Temp
BManure
Corn
DDManure
DManure
FoodW
Hazelnut
Oak
Pine
PaperW
Poultry
(C)
Box-Cox transformed
Ash [%]
Temp
BManure
Corn
DDManure
DManure
FoodW
Grass
Hazelnut
Oak
Pine
PaperW
Poultry
Rapeseed
(D)
log link
Na [Total (mg kg^{-1})]
Temp
BManure
Corn
DDManure
DManure
FoodW
Hazelnut
Oak
Pine
PaperW
Poultry

Figure 4:



Corn
D. dairy manure
Food waste
Grass
Hazelnut
Oak
Pine
Paper waste
Poultry litter
pH$_w$ [-]
Temp
12.5
10.0
7.5
5.0
2.5
200
300
400
500
600
700

Figure 5:



	Response	Predicted	Standard Error	Confidence Interval
1	Bulk Density [Mg/m^3]	0.11	(0.00402)	[0.1013,0.1171]
11	SSA(N2) [m^2/g]	3.72	(0.6366)	[2.848,4.595]
12	SSA(CO2) [m^2/g]	390.63	(42.39)	[330.6,450.6]
13	Yield [%]	31.52	(1.037)	[30.01,33.09]
14	EC [S/m]	959.29	(0.3889)	[815.2,1124]
15	CEC [mmol_c/kg]	554.65	(1.159)	[455.6,675.2]
16	pHw [-]	9.24	(0.4752)	[8.592,9.896]
17	Ash [%]	13.79	(0.07787)	[11.79,16.16]
18	MatVol [%]	37.22	(1.097)	[35.65,38.8]
19	C [mg/g]	672.34	(13.04)	[655.1,689.5]
110	N [mg/g]	10.87	(0.5244)	[10.18,11.56]
111	C:N [-]	61.93	(0.08026)	[53.03,72.47]
112	FixedC [%]	47.40	(2.827)	[43.58,51.21]
113	P [Total (mg/kg)]	1913.52	(0.4856)	[1647,2214]
114	S [Total (mg/kg)]	742.15	(1.123)	[633.1,869.9]
115	Ca [Total (mg/kg)]	8343.70	(1.065)	[7672,9074]
116	K [Total (mg/kg)]	22823.20	(1.62)	[21440,24270]
117	Mg [Total (mg/kg)]	7684.81	(1.081)	[6912,8544]
118	Na [Total (mg/kg)]	963.94	(1.106)	[840.3,1106]
119	Fe [Total (mg/kg)]	922.74	(1.356)	[609,1398]
120	Mn [Total (mg/kg)]	178.44	(1.085)	[159.6,199.6]
121	Zn [Total (mg/kg)]	74.16	(1.175)	[59.35,92.61]

Figure 6:

Table 1:

Biochar property	**Agronomic and environmental benefits**
BulkD [Mg m^{-3}]	Low bulk density biochar can reduce the density of compacted soils, thereby improving root penetration (Atkinson et al., 2010; Ennis et al., 2011; Novak and Busscher, 2013), water drainage and aeration (Joseph et al., 2009; Laird et al., 2010). The latter may mitigate green house gas emissions (Karhu et al., 2011).
SSA(N_2), SSA(CO_2) [m^2 g^{-1}]	High nanopore and micropore specific surface area, respectively, may increase the sorptive affinity of organic compounds to biochars (Cornelissen et al., 2005; Beesley et al., 2011), and improve water holding capacity (Karhu et al., 2011).
Yield [%]	Yield reflects the quantity of biochar material produced from the pyrolysis process.
EC [mS m^{-1}]	Electrical conductivity indicates the quantity of salt contained in the biochar. High EC can stabilize soil structure (Joseph et al., 2009; Hossain et al., 2011).
CEC [Av (mmol$_c$ kg^{-1})]	Increased cation exchange capacity can improve the soil's ability to hold and exchange cations (Chapman, 1965; Glaser et al., 2002).
pH$_w$ [-]	Soil solution pH directly affects soil surface charge, which determines the type of exchangeable nutrients and mineral ions it attracts (Mukherjee et al., 2011). Additionally, the buffering capacity of biochar can neutralize acidic soils, redude aluminum toxicity and change the soil microbial community structure (Abe, 1988; Lehmann et al., 2011).
Ash [%]	Ash may improve the sorption capacity of biochar for organic compounds and metals (Cao et al., 2011).
MatVol [%]	Volatile matter affects biochar longevity in soil (Lehmann et al., 2011; Enders et al., 2012). Residual volatiles can also impact organic substance sorption by blocking pores and changing surface chemical interactions (Sander and Pignatello, 2005; Zhu et al., 2005; Novak and Busscher, 2013).
C [mg g^{-1}]	Total carbon in organic matter benefits the soil.
N [mg g^{-1}]	Total nitrogen in the biochar supplies a macronutrient, but its availabiity is limited. Biochar may strongly sorb ammonia and act as a nitrogen-rich soil amendment (Spokas et al., 2012b).
C:N [-]	Carbon to nitrogen ratio influences the rate of decomposition of organic matter and release of soil nitrogen (Novak et al., 2009).
FixedC [%]	Fixed carbon is non-labile and therefore is a property attributed to biochar stability (Keiluweit et al., 2010; Enders et al., 2012; Rajkovich et al., 2012).
P, S [Total (mg kg^{-1})]	Macronutrients provided by biochar, which can improve soil fertility.
Ca, K, Mg, Na, Fe, Mn, Zn [Total (mg kg^{-1})]	Micronutrients provided by biochar, which can improve soil fertility.

Notes: BulkD = Bulk Density, SSA = Specific Surface Area, EC = Electrical Conductivity, CEC = Cation Exchange Capacity, MatVol = Volatile Matter, FixedC = Fixed Carbon

Table 2:

Biomass	**Feedstock**	**Milling size** [μ**m**]	**Moisture** [%]	**Reactor type**	**Feed capacity**	**Oxygen limitation**	**Heat rate**	**Holding time** [**min**]	**Peak temp.** [°**C**]	**Reference**
Bull manure	animal	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Enders et al., 2012)
Corn	plant	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Dairy manure	animal	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Enders et al., 2012)
Digested dairy manure	animal	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Food waste	combo	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,400,500,600	(Rajkovich et al., 2012)
Grass (*Tall fescue*)	plant	<1500	na	closed container muffle furnace	na	yes[a]	na	60	300,400,500,600	(Keiluweit et al., 2010)
Grass (*Tripsacum floridanum*)	plant	50,000	(5d drying at 60°C)	batch pyrolysis oven	4,749 cm^3	N_2	26°C	60	250,400,650	(Mukherjee et al., 2011)
Hazelnut	plant	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Oak (*Quercus rotundifolia*)	plant	177-250	na	horizontal tube furnace	na	N_2	continuous flow	120	300,350,400,450,500,550,600	(Cordero et al., 2001)
Oak (*Quercus lobata*)	plant	50,000	(5d drying at 60°C)	batch pyrolysis oven	4,749 cm^3	N_2	26°C	60	250,400,650	(Mukherjee et al., 2011)
Oak	plant	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Paper waste	plant	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,400,500,600	(Rajkovich et al., 2012)
Pine (*Pinus halepensis*)	plant	177-250	na	horizontal tube furnace	na	N_2	continuous flow	120	300,350,400,450,500,550,600	(Cordero et al., 2001)
Pine (*Pinus ponderosa*)	plant	<1500	na	closed container muffle furnace	na	yes[a]	na	60	300,400,500,600	(Keiluweit et al., 2010)
Pine (*Pinus taeda*)	plant	na	na	batch pyrolysis unit	na	N_2	na	na	400,500	(Gaskin et al., 2008)
Pine (*Pinus taeda*)	plant	50,000	(5d drying at 60°C)	batch pyrolysis oven	4,749 cm^3	N_2	26°C	60	250,400,650	(Mukherjee et al., 2011)
Pine	plant	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Poultry litter	animal	na	na	batch pyrolysis unit	na	N_2	na	na	400,500	(Gaskin et al., 2008)
Poultry litter	animal	149-850	10	kiln	3000 g	N_2	3°C 15-20 min^{-1}	80-90	300,350,400,450,500,550,600	(Rajkovich et al., 2012; Enders et al., 2012)
Rapeseed	plant	<1000	12.6	tubular reactor	30 g	N_2	5°C min^{-1}	30	400,500,600	(Karaosmanoğlu et al., 2000)

[a] Details not specified.

Table 3:

Response	**Formula**	**Transformation**	**Link**	**GOF**
BulkD	B + B:T	Box-Cox Transf	identity	✓
SSA(N_2)	B + B:T	-	identity	✗
SSA(CO_2)	B:T	-	identity	✓
Yield	B + B:T	Log Transf	log	✓
EC	B + B:T	Box-Cox Transf	log	✓
CEC	B + B:T	Log Transf	log	✓
pH_w	B + T	-	identity	✓
Ash	B + T	Box-Cox Transf	identity	✓
MatVol	F + F:T	-	identity	✓
C	B + B:T	-	indentity	✓
N	B + B:T	-	identity	✗
C:N	B + T	Box-Cox Transf	identity	✓
FixedC	B:T	-	identity	✓
P	B + B:T	Box-Cox Transf	log	✓
S	B	Log Transf	identity	✓
Ca	B + B:T	Log Transf	identity	✓
K	B + B:T	Box-Cox Transf	identity	✓
Mg	B + T	Log Transf	identity	✓
Na	B + T	Log Transf	log	✓
Fe	B + T	Log Transf	log	✗
Mn	B + T	Log Transf	identity	✓
Zn	B + T	Log Transf	log	✓