

Adaptation of Compositional Data Analysis in Deep Learning to Predict Pasture Biomass Proportions

Badri Narayanan^{1,3,4}, Mohamed Saadeldin^{1,3,4}, Paul Albert^{2,3,4}, Kevin McGuinness^{2,3,4}, Noel E. O'Connor^{2,3,4}, and Brian Mac Namee^{1,3,4}

¹ School of Computer Science, University College Dublin

² School of Electronic Engineering, Dublin City University

³ Insight SFI Centre for Data Analytics

⁴ VistaMilk

badri.narayanan@insight-centre.org

Abstract. Dry biomass weight measurements from a quadrat in a paddock for grass, clover and weeds when expressed as percentages of total dry herbage mass are compositional in nature. Unlike real valued regression problems, prediction of compositional data is handled differently in statistics because of its closure property where the components of the composition are positive data adding up to a constant sum and is therefore constrained in the simplex space, in our case 100%. Our motivation in this paper was to study whether the adaptation of compositional data analysis (CoDa) techniques in deep learning improves the prediction results over the best performing deep learning model we used in our earlier paper [Narayanan et al., 2021]. Although the log ratio transformation of targets is an appropriate adaptation of CoDa and is interesting for Biomass prediction, our study indicates that the CoDa adaptation does not improve the prediction errors over our earlier method.

Keywords: Deep Learning · Compositional Data Analysis · Isometric Log Ratio · Simplex · Softmax

1 Introduction

The dairy industry uses clover and grass as fodder for cows. Grass and clover are grown together in fields to improve the consistency of high-quality biomass yield and to reduce the need for external fertilizers. Accurate estimation of the dry biomass percentages of grass and clover species (as well as weeds) in fields is very important for determining optimal seeding density, fertilizer application and elimination of weeds. The dry biomass weights of the individual components, when expressed as percentages of overall weight of the harvested and dried biomass, are compositional in nature.

Compositional data are positive data summing to a constant value, and measure relative changes in the components. They are constrained in the simplex space. Standard multivariate statistical analysis and regression techniques assume

the sample space to be \mathbb{R} . The sample space for compositional data, however, is restricted in the simplex due to the sum constraint. Compositional data analysis (CoDa) [Aitchison \[2005\]](#) is a set of mathematical techniques that helps in analysing the relative proportions of individual components.

In this paper, we examine the applicability of the principles of CoDa to the problem of predicting biomass composition from farm imagery using deep learning. We present an adaptation of the approach used in statistical CoDa where the compositional data is transformed from the simplex space to the real space using the isometric log ratio (ILR) transformation [[Egozcue et al., 2003](#)]. These transformed values are used as targets with a model that predicts the dry mass fractions of grass, clover and weeds from images of a section of grass marked with a square frame (known as a *quadrat*). We compare the prediction results to the best performing model from our previous paper [[Narayanan et al., 2021](#)] that uses the composition data directly as a target. This comparison shows that the addition of the approaches from CoDa do not improve the performance of the model.

In the rest of this paper, Section 2 outlines related work for CoDa techniques in applied statistics and highlights the few available adaptations by the machine learning community. This is followed by a description of our experimental design in Section 3 and a discussion of the results in Section 4. Finally, Section 5 summarises the findings from this paper and suggests directions for future work.

2 Related Work

This section introduces compositional data analysis, establishes the relationship between the softmax transformation and the simplex, and reviews the limited applications of CoDa to machine learning and biomass composition prediction.

2.1 Compositional Data Analysis

Compositional data of D -parts is constrained in the simplex space \mathcal{S}^D by a constant sum of the components. The sum constraint induces negative correlations between variables [[Chayes, 1960](#)], which violates the independence assumptions and the central limit theorem [[Aitchison, 1982, 2005](#)]. Therefore, the sum constraint needs to be broken before standard statistical methods can be applied for analysis. This is often achieved by transforming the compositional data from the simplex space into real space using log ratio transformations [[Pawlowsky-Glahn and Egozcue, 2006](#)].

[Aitchison \[1982\]](#) formalises three key principles of CoDa: scale invariance, permutation invariance, and subcompositional coherence. Any statistical analysis of compositional data must conform to these principles. Scale invariance is characterised by the relative information that the compositional data carry, rather than the individual size of the components. Permutation invariance mandates that any statistical inference should be independent of the ordering of the components within the composition. Finally, subcompositional coherence stipulates that results

from the analysis of components in a full composition should not contradict the results from a subcomposition, i.e., the distances between two compositions should decrease when subcompositions of the original ones are considered, and that scale invariance is preserved within arbitrary subcompositions.

Log Ratio Transformations The centered log ratio (CLR) and isometric log ratio (ILR) are the prevalent log ratios used in modern CoDa applications. In a composition \mathbf{x} of $D \geq 2$ components, the sum constraint of compositional data implies that there is at least one component that is negatively correlated with another in the composition, and that there are at most $D - 1$ independent components. The composition is therefore constrained in a $D - 1$ dimension vector space of \mathcal{S}^D , defined as a D -part simplex on \mathbb{R} . The values of these components, when scaled by their geometric mean and then log transformed, are mapped to a hyperplane in \mathbb{R}^D and referred to as the Centred Log Ratio (CLR):

$$\text{clr}(\mathbf{x}) = \mathbf{z} = \left[\ln \left(\frac{x_1}{g(\mathbf{x})} \right), \ln \left(\frac{x_2}{g(\mathbf{x})} \right), \dots, \ln \left(\frac{x_D}{g(\mathbf{x})} \right) \right],$$

where $g(\mathbf{x})$ is the geometric mean of the k components of \mathbf{x} :

$$g(\mathbf{x}) = \sqrt[D]{x_1 x_2 \dots x_D}.$$

Pawlowsky-Glahn et al. [2007] highlight that the CLR introduces a mathematical complexity in the form of a singular covariance matrix where the determinant is zero. Additionally, the CLR transformation is not subcompositionally coherent as the geometric mean of a subcomposition will differ from that of the whole composition.

These drawbacks led to the introduction of the isometric log ratio (ILR) by Egozcue et al. [2003] where an isometry from \mathcal{S}^D to \mathbb{R}^{D-1} is achieved from an orthonormal basis derived from Gram-Schmidt orthogonalization. In addition to being an isometry, the ILR is an isomorphism too, and conforms to the three CoDa principles outlined above. Following Egozcue et al. ILR can be defined as follows:

$$\text{ilr}(\mathbf{x}) = [\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \langle \mathbf{x}, \mathbf{e}_2 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a],$$

where $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}]$ is an orthonormal basis in the simplex, the default one being the orthonormal basis built by Egozcue et al. using Gram-Schmidt orthogonalization. $\langle \mathbf{x}, \mathbf{e}_i \rangle_a$ represents the Aitchison inner product between \mathbf{x} and \mathbf{e}_i . The inverse of ILR transformation is given by

$$\mathbf{x} = \text{ilr}^{-1}(\mathbf{y}) = \bigoplus_{i=1}^{D-1} (\langle \mathbf{y}, \vec{\mathbf{e}}_i \rangle_a \odot \mathbf{e}_i),$$

where $\vec{\mathbf{e}}_i = \text{ilr}(\mathbf{e}_i)$ for all i . \bigoplus and \odot denote the compositional operations of perturbation and power transformation described in the Aitchison geometry of the simplex [Pawlowsky-Glahn and Egozcue, 2006].

Although both the CLR and the ILR are isometric and allow for statistical operations in the Euclidean space, the ILR is the most prevalent in modern applications of CoDa, simply because of its representation of the composition in an orthogonal coordinate system. Unlike CLR, the ILR allows for the association of angles and distances in the simplex with those in the real space, and adheres to the 3 key principles of CoDa, thereby making it a better choice. For interested readers, Tolosana-Delgado [2008] provide a short and comprehensive mathematical representation of these log ratios and other foundational aspects of CoDa.

Handling zero values Zero values in compositional data, if not handled, can be problematic. Essential zeros refer to the absence of a component in the observation, whereas rounded zeros indicate approximate recording of a component below detection limit [Martín-Fernández et al., 2003] and need to be addressed. Rounded zeros are replaced with a threshold value using a multiplicative replacement method that maintains the constant sum of the composition.

Applications of CoDa Applied statistics has seen many applications of Compositional Data Analysis (CoDa) in geostatistics [Tolosana-Delgado et al., 2019], bioinformatics, environmental science and chemistry [Filzmoser et al., 2010] where many problems are compositional in nature. Liu et al. [2016] trace the underlying factors that influence rock weathering and mineralisation in stream sediments of the Nanling tectono-magmatic belt using robust factor analysis and compositional data analysis.

In the biomass composition problem studied in this paper the targets are compositional. Aitchison [2005] presents an example similar to this that quantifies the extent of dependence of sediment composition on water depth in arctic lakes. Three mutually exclusive and exhaustive constituents (sand, silt and clay) are recorded in their proportions by weight for 39 samples at different water depths. The objective is to quantify the extent of dependence of sediment composition on water depth and hence, identify the nature of sedimentation process. In another similar example the relative mass of water, fat and protein in a meat sample is predicted from its NIR spectrum [Verwaeren, 2014].

2.2 Softmax and the Simplex

In the context of the deep learning experiments in this work, it is necessary to understand the relationship between the softmax activation function used in neural networks and the simplex space. The softmax is a mathematical function most commonly used in the output layers of neural networks for multi-class classification, and provides a generalisation of the sigmoid function in logistic regression. The softmax function has been extensively used in state-of-the-art deep neural network models and has been used very successfully in classification and regression problems.

In a typical multi-class classification setting, the softmax function converts a vector of k real values into a vector of k probabilities that sum to 1—a probability

distribution for the predicted classes in the target. Each of these probabilities is a proportion of the relative scale of the corresponding individual component of the input vector:

$$\sigma(\mathbf{z}_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \text{ and } \mathbf{z} = (z_1, \dots, z_K) \in \mathbb{R}^K.$$

[Amos, 2019, Theorem 4, pg 13] provides a theorem and proof that establishes the relationship between the softmax activation function and the simplex, where the softmax acts as the projection of a point $x \in \mathbb{R}^n$ onto the interior of the $(n - 1)$ -simplex.

2.3 Applications of CoDa in Machine Learning & Biomass Prediction

There are limited examples in the literature of the adaptation of the CoDa principles to machine learning. The use of random forest models trained on data pre-processed with log ratio transformations [Harris and Grunsky, 2015; Talebi et al., 2018] illustrate the few attempts in the use of machine learning; however, this area remains largely unexplored and there are no specific instances of literature of experiments / benefits of adapting CoDa techniques with deep learning.

A body of recent research [Skovsen et al., 2018; Larsen et al., 2018; Sindic and Riday, 2020; Castro et al., 2020; Sun et al., 2021] employs state-of-the-art deep learning techniques to predict dry matter yield from proximal and UAV images of grass paddocks. These works typically rely on transfer learning [Pan and Yang, 2009] to transfer latent representations that were learnt from large corpus of images by deep networks like VGG16 [Simonyan and Zisserman, 2014] and Resnet [He et al., 2016]. To our knowledge, there are no references to the adaptation of CoDa techniques in these deep learning approaches. Therefore it is interesting to explore the integrated approach of CoDa principles and deep learning to solve the biomass composition prediction problem. The next section describes the design of an experiment to assess the effectiveness of adapting the concept of isometric log ratio transformation (ILR) introduced by Egozcue et al. [2003] to the deep learning architecture used in Narayanan et al. [2021].

3 Experimental Design

This section describes the design of a set of experiments undertaken to assess the effectiveness of adopting CoDa techniques in deep learning models for biomass composition prediction. The section describes the dataset used, the architecture of the models built and the experimental method used.

3.1 Data Description

The Grass Clover Image Dataset for the Biomass Prediction Challenge [Skovsen et al., 2019] provides us with 261 images of quadrats of grass with corresponding

dry biomass composition of grass, white clover, red clover and weeds. These are expressed in terms of their weights proportional to the total biomass, and sum to 1. Five example images and their target values are presented in Table 1.





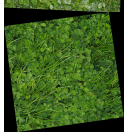
| Image | Proportions of weight | | | | ILR coefficients | | |
|---|-----------------------|--------------|------------|--------|------------------|---------|---------|
| | Grass | White clover | Red clover | Weeds | ilr_1 | ilr_2 | ilr_3 |
|  | 0.7648 | 0.2058 | 0.0000 | 0.0294 | -3.7659 | -3.2462 | 0.5262 |
|  | 0.9085 | 0.0655 | 0.0260 | 0.0000 | -0.6518 | -2.5238 | 4.1137 |
|  | 0.9532 | 0.0073 | 0.0000 | 0.0395 | -1.4092 | -4.7866 | -0.6271 |
|  | 0.2065 | 0.5590 | 0.0314 | 0.2031 | -2.0354 | -0.3619 | -0.2418 |
|  | 0.2747 | 0.5189 | 0.2064 | 0.0000 | -0.6518 | 0.1432 | 4.9636 |

Table 1: Five example images from the Grass Clover Image Dataset along with their biomass compositions and ILR transformed targets. The coefficients `ilr_1`, `ilr_2` and `ilr_3` are obtained from the transformation of the proportions of grass, white clover, red clover and weeds after replacement of zero values.

The dataset was divided into 209 training examples and 52 validation examples. All the images were standardized to 500×500 pixels and, to ensure adequate training examples for the network to learn efficiently, the training set images were subject to $10\times$ expansion through runtime augmentations [Krizhevsky et al., 2012]. The transformations in the augmentation included a rotation (up to 15°), zoom ($\pm 15\%$), height and width shift (20%), shearing ($\pm 15\%$), horizontal reflections, channel shift (± 50), and image wrapping to minimize loss of information.

We use the Compositional Statistics package⁵ in Python for the ILR and inverse ILR transformations in this work. Given the non-zero minimum values of each component in the dataset presented in Table 2, a threshold value of 0.001

⁵ Compositional Statistics: <https://composition-stats.readthedocs.io/>

| Grass | White Clover | Red Clover | Weeds |
|----------|--------------|------------|----------|
| 0.051104 | 0.001333 | 0.003056 | 0.001025 |

Table 2: Non-zero minimum values of the components

was selected for undetectable measurements. Zero values in the data are then replaced with this minimum threshold value using the multiplicative replacement method, while ensuring the sum closure of 1. The transformed targets from the five examples presented in Table 1 are also shown.

3.2 Model Architecture

The model architecture used in this study is the same convolutional neural network (CNN) architecture from our previous work [Narayanan et al., 2021] where we used weak supervision [Zhou, 2018], transfer learning from a VGG-16 model pre-trained on the ImageNet dataset, and a multi-target output layer with softmax regression trained to minimise root mean squared error (RMSE) loss. The weak supervision is necessary as the dataset has missing values in 104 examples for the red and white clover subcomposition while the overall clover values were available, and therefore had to be imputed with their corresponding mean values and readjusted to match the overall clover proportion in the total biomass. In doing so, the approximated examples were given a lesser weighting during the loss calculation in a ratio of 1:1.5 with respect to the examples with recorded values. Latent feature representations through transfer of weights from the final convolutional layer of the pre-trained VGG-16 network enabled faster and better optimization of the two trainable dense layers with 4,096 and 256 neurons. The dense layers were equipped with ReLU activations and uniform random kernel initialization, and each of the dense layers was followed by a layer of batch normalization to help prevent overfitting. The network was compiled with the Adam optimizer with an initial learning rate 0.001 and decay factor $10^{-3}/200$. The output layer had 4 neurons with softmax activation, each neuron corresponding to regression output for grass, white clover, red clover and weeds. As a direct interpretation of Amos’s theorem [Amos, 2019], the softmax probabilities from the output layer can be construed as equivalent to the simplicial proportions of the predicted values of the individual components. This provides a framework for interpreting the results of the model with softmax outputs in the context of the simplex. In our current work, we transform the 4 target variables of grass, white clover, red clover and weeds into 3 ILR coefficients. We modify the output layer of the network in this experiment to 3 neurons with linear activations for real valued outputs, and use RMSE as the loss function.

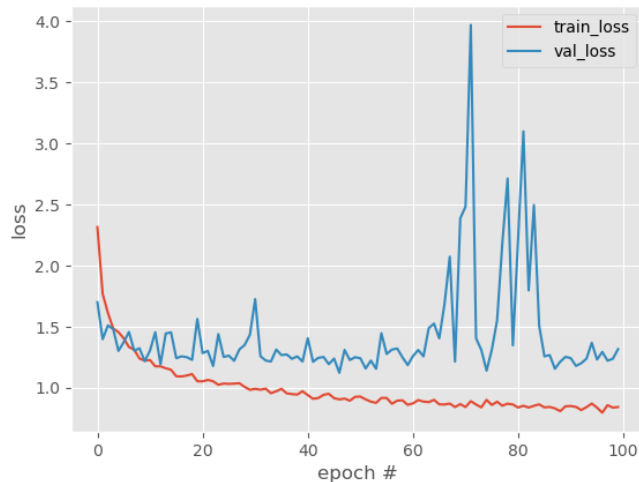


Fig. 1: Training and validation losses (RMSE) for the ILR transformed data.

4 Results and Discussion

An examination of the training and validation losses during model training, presented in Figure 1, confirms the ability of the model to learn from the ILR transformed targets in the training data. The top row in Figure 2 shows scatter plots from the baseline model results from our previous work, of the actual values versus predicted values for each component of the biomass composition for data in the validation set. Similarly, the bottom row shows scatter plots of actual vs predicted values for the ILR transformed targets. It is interesting to note that the ILR model predicts the proportions of grass and white clover reasonably well. In the case of red clover, however, the predictions are generally erroneous. Prediction of weeds is reasonably accurate when the actual weed percentage is less than 20% of the composition, but erroneous results can be observed above this range.

The results from our previous work is the baseline for comparison with the results of the CoDa adaptation experiment. Table 3 compares the performance of this baseline model against the model trained to predict ILR transformed targets. It is clear that the baseline model outperforms the CoDa-inspired model.

This experiment shows convincing evidence of the ability to learn from ILR transformed compositional data. Nevertheless, the CoDa adaptation results do not improve upon the performance of this baseline model. We surmise that there are two reasons for this. First, it is interesting to note that the softmax function projects the real valued output vector of the network onto the simplex (as explained in Section 2.2). In our case the training targets too are in the same simplicial dimensions, and therefore, it is effective within the simplex. Second,

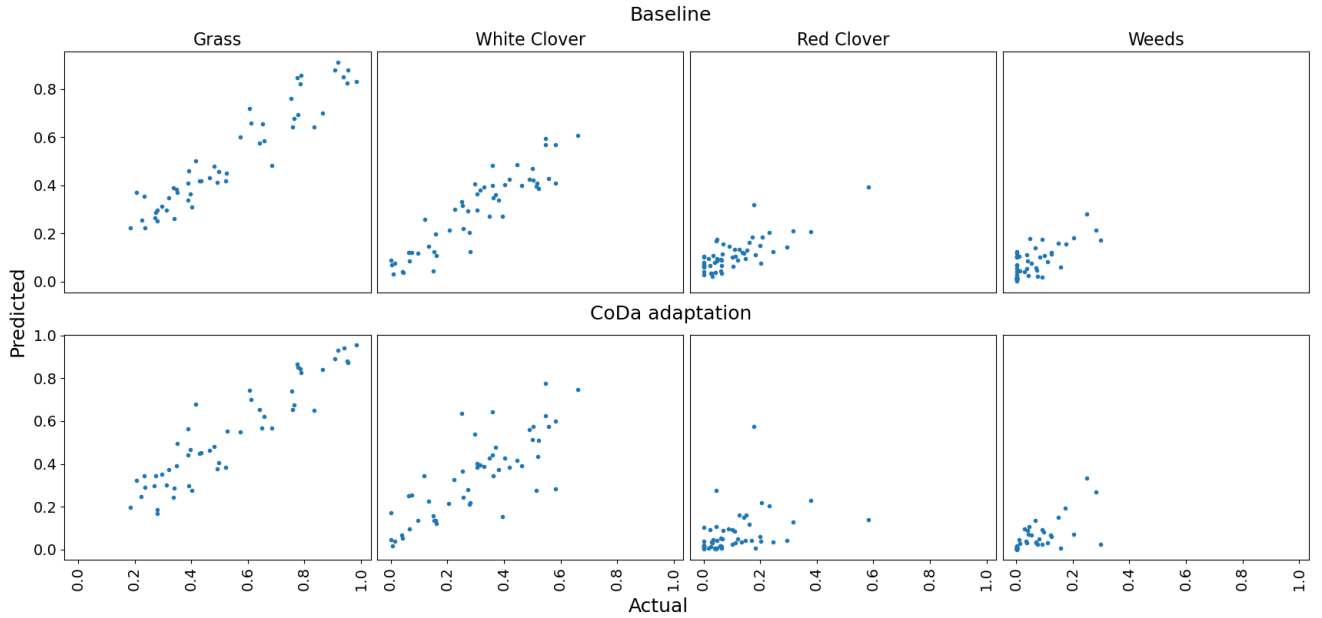


Fig. 2: Scatter plots of actual vs predicted values for the components.

the premise of CoDa is to ensure data transformations that will satisfy the requirements of standard statistical analyses, like the central limit theorem and conformance to the rules of linear independence. On the contrary, deep neural networks do not require such assumptions and have the ability to effectively approximate a non-linear estimation function to fit an unknown distribution of the target data. The problem that CoDa is designed to solve using standard statistical methods does not exist in deep neural networks. Therefore, the CoDa adaptation for deep neural networks using the ILR transformation of the targets is an additional step over a network with an intrinsic ability to learn these non-linear functions. We believe that these two reasons provide a plausible explanation for

| Model | Validation metrics RMSE, MAE | | | | | | | | | |
|----------|------------------------------|------|--------------|------|------------|------|-------|------|---------|------|
| | Grass | | White clover | | Red clover | | Weeds | | Overall | |
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| Baseline | 8.00 | 6.21 | 7.44 | 5.99 | 7.33 | 5.63 | 5.68 | 4.20 | 7.11 | 5.51 |
| ILR | 8.87 | 7.04 | 12.64 | 8.94 | 11.98 | 7.39 | 5.63 | 2.95 | 9.78 | 6.58 |

Table 3: Comparison of validation metrics RMSE and MAE with the baseline from our earlier work.

the better performance of the network with softmax activation over the ILR transformed approach.

5 Conclusion

In this paper we explored the usefulness of techniques from statistical compositional data analysis in a deep learning context. In particular, we tested this with a pasture biomass prediction problem, which is compositional in nature. We presented an approach that transformed the biomass composition data using the isometric log ratio (ILR) from the simplex space onto the real space and used these transformed targets for training a deep network. This paper demonstrates that it is possible to train a reasonably accurate prediction model using this approach. Nevertheless, based on the evidence of the results, we conclude that the softmax works better in the deep learning context than a model trained to predict targets transformed using ILR. This suggests that it is not useful to adapt techniques from statistics CoDa to deep learning models. Our further work will focus on improving the prediction for red clover and weeds.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number [16/RC/3835]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Our sincere thanks to Prof. Claire Gormley, School of Mathematics and Statistics, University College Dublin, for her suggestion to explore Compositional Data Analysis.

Bibliography

- Aitchison, J.: The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)* 44(2), 139–160 (1982)
- Aitchison, J.: *A Concise Guide to Compositional Data Analysis* p. 134 (2005)
- Amos, B.: Differentiable optimization-based modeling for machine learning. Ph.D. thesis, PhD thesis. Carnegie Mellon University (2019)
- Castro, W., Marcato Junior, J., Polidoro, C., Osco, L.P., Gonçalves, W., Rodrigues, L., Santos, M., Jank, L., Barrios, S., Valle, C., et al.: Deep learning applied to phenotyping of biomass in forages with uav-based rgb imagery. *Sensors* 20(17), 4802 (2020)
- Chayes, F.: On correlation between variables of constant sum. *Journal of Geophysical research* 65(12), 4185–4193 (1960)
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C.: Isometric logratio transformations for compositional data analysis. *Mathematical geology* 35(3), 279–300 (2003)

- Filzmoser, P., Hron, K., Reimann, C.: The bivariate statistical analysis of environmental (compositional) data. *Science of The Total Environment* 408(19), 4230–4238 (Sep 2010), <https://linkinghub.elsevier.com/retrieve/pii/S0048969710004845>
- Harris, J., Grunsky, E.: Predictive lithological mapping of Canada’s North using Random Forest classification applied to geophysical and geochemical data. *Computers & Geosciences* 80, 9–25 (Jul 2015), <https://linkinghub.elsevier.com/retrieve/pii/S0098300415000709>
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
- Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105 (2012)
- Larsen, D., Steen, K.A., Grooters, K., Green, O., Nyholm, R., et al.: Autonomous mapping of grass-clover ratio based on unmanned aerial vehicles and convolutional neural networks. In: *International Conference on Precision Agriculture*. International Society of Precision Agriculture (2018)
- Liu, Y., Cheng, Q., Zhou, K., Xia, Q., Wang, X.: Multivariate analysis for geochemical process identification using stream sediment geochemical data: A perspective from compositional data. *Geochemical Journal* 50(4), 293–314 (2016)
- Martín-Fernández, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V.: Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology* 35(3), 253–278 (2003)
- Narayanan, B., Saadeldin, M., Albert, P., McGuinness, K., Mac Namee, B.: Extracting pasture phenotype and biomass percentages using weakly supervised multi-target deep learning on a small dataset. *arXiv preprint arXiv:2101.03198* (2021)
- Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22(10), 1345–1359 (2009)
- Pawlowsky-Glahn, V., Egozcue, J.J.: *Compositional data and their analysis: an introduction*. Geological Society, London, Special Publications 264(1), 1–10 (2006)
- Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana Delgado, R.: *Lecture notes on compositional data analysis* (2007)
- Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
- Sindic, C.M., Riday, H.: Using image object recognition to increase biomass in red clover (*trifolium pratense* l.) breeding. *Crop Science* 60(4), 1770–1781 (2020)
- Skovsen, S., Dyrmann, M., Eriksen, J., Gislum, R., Karstoft, H., Jørgensen, R.N.: Predicting dry matter composition of grass clover leys using data simulation and camera-based segmentation of field canopies into white clover, red clover, grass and weeds. In: *Proceedings of the 14th International Conference on Precision Agriculture*. Montreal, CA: International Society of Precision Agriculture (2018)

- Skovsen, S., Dyrmann, M., Mortensen, A.K., Laursen, M.S., Gislum, R., Eriksen, J., Farkhani, S., Karstoft, H., Jorgensen, R.N.: The grassclover image dataset for semantic and hierarchical species understanding in agriculture. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
- Sun, S., Liang, N., Zuo, Z., Parsons, D., Morel, J., Shi, J., Wang, Z., Luo, L., Zhao, L., Fang, H., et al.: Estimation of botanical composition in mixed clover–grass fields using machine learning-based image analysis. *Frontiers in Plant Science* 12, 87 (2021)
- Talebi, H., Mueller, U., Tolosana-Delgado, R., Grunsky, E., McKinley, J., Caritat, P.d.: Surficial and Deep Earth Material Prediction from Geochemical Compositions. *Natural Resources Research* 28 (Oct 2018)
- Tolosana-Delgado, R.: Compositional data analysis in a nutshell. University of Gottingen on-line reference (2008)
- Tolosana-Delgado, R., Mueller, U., van den Boogaart, K.G.: Geostatistics for compositional data: an overview. *Mathematical geosciences* 51(4), 485–526 (2019)
- Verwaeren, J.: Mathematical optimization methods for the analysis of compositional data: subset selection, unmixing and prediction. Ph.D. thesis, Ghent University (2014)
- Zhou, Z.H.: A brief introduction to weakly supervised learning. *National Science Review* 5(1), 44–53 (2018)