A Decision Analytic Framework and Exploratory Statistical Case Study Analysis of Grass Growth in Northern Ireland

Orla McHugh, Fiona Browne, Jun Liu, and Philip Jordan
Ulster University, Northern Ireland, UK
Email: {mchugh-o3, f.browne, j.liu, p.jordan}@ulster.ac.uk

***Abstract*—Sustainable land management, with pressures from climate change, is a highly multidisciplinary research field. There are challenges to exploit an abundance of data and apply data-processing technologies to integrate environmental, economic, and social considerations and manage uncertainties originating from imperfect data quality. Motivated by these challenges, the present work proposes a multi-layered mapping methodological framework to bridge or reduce the problems identified by developing a transparent and explainable decision support system as a precision agriculture tool. This should be designed both for farmers and agriculture decision makers, and integrate soft (e.g., legislation, policy, regulation and experience) and hard data (measured data), along with geographical information that presents key information in the form of spatial mapping, information mapping, and causal structure mapping. Presented is a preliminary exploratory statistical case study analysis on grass growth data in order to examine patterns and determine which factors have the greatest influence on grass growth in Northern Ireland.**

***Index Terms*—sustainable land management, climate change, decision support system, uncertainty, exploratory statistical analysis**

I. Introduction

Climate change is one of the most urgent risks to our environment at present and the United Kingdom (UK) is under increasing pressure to reduce the amount of greenhouse gases produced [1]. The relationship between climate change and agriculture is cyclical as climate change contributes to agricultural activities, and agriculture is affected by climate change [2]. The Northern Ireland (NI) agricultural sector is one of the most significant industries in the region, as there is an estimated 2.5% of the civic workforce in employment [3]. Some of the biggest earning sectors within the agricultural industry are the production of beef, milk and milk products, and sheep meat [3]. The importance of these industries highlights the reliance of farmers on grassland management, as this is the main feed source for livestock.

Global climate change and a growing realisation that natural resources are finite have made sustainability a top scientific, industrial, societal, and governmental priority. More specifically, sustainable development policies and plans of the dairy farm sector of NI to reduce nutrient losses, causing water pollution and air pollution, while ensuring the economic and social values associated, are also becoming more crucial. Integrated decision making and a long-term approach to planning are defining characteristics and prerequisites of sustainable development and represent key management challenges. Data within this domain includes hard data, e.g., geochemical data, including soil pH and moisture, bedrock geology, elemental distributions across NI, and meteorological data, including solar radiation and rainfall, as well as soft data, e.g., legislation, regulation, law, policies, and farmer experience/advices. All these data are held by public/private agencies in NI. These are often siloed and as a result are not being used to their full potential due to a lack of effective and holistic information management and decision-support techniques and systems [4]. Agricultural yields and environmental health in NI are currently suboptimal [5] and there are a number of scenarios following BREXIT that require urgent planning.

Decision Support Systems (DSS) have been used to aid in decision making and have been applied to the sustainable land management domain, to exploit and handle the abundance of data available using different data-processing technologies [6], [7]. These include, e.g., model based DSS [8], [9], knowledge based DSS [10], [11], data driven DSS [12], Geographic information System (GIS) based DSS, or can be a combination of multiple types [13]. However, there are still some limitations for these techniques to be widely and effectively applied to land management decision making. This is due to the problems/challenges such as:

- Low uptake of DSSs by farmers for multiple reasons;
- Gap between research and practice – no real world uses;
- The modelling approaches have been widely used, but often fall short in their ability to permit the incorporation of subjective and/or vague data, which is important to ensure its viability and relevance to the problem;

Manuscript received August 13, 2019; revised December 24, 2019.

doi: 10.12720/jait.11.1.15-20

- Collecting sufficient data to base a statistical probability of prediction is costly, and in many situations, such data is limited or unavailable due to a lack of research or the complexity of the system/process considered;
- It is extremely difficult to quantify the effects and consequences of some events as they involve too many factors with a high level of uncertainty, even in those cases where the physical processes are clearly understood;
- It may be extremely difficult to construct an accurate and complete mathematical model to correlate the sustainability factors and sustainability levels;
- Transparent reasoning for the decision solution must be present to allow accountability
- DSS must be highly visual to allow for ease of use, and to be accessible by a wider audience.
- Crucially, handling uncertainty is one of the vital issues in land management in complex systems with diverse environments.

Novel decision methods are therefore required to bridge domain knowledge and data analytics in an acceptable way in various environments where mature tools cannot be effectively or efficiently applied. In this paper we focus on grass-land management within NI as it represents the largest agricultural land use. Various types of information exist within the domain of land management, from geographical data, to farmer experience, to policy. This data is often interpreted within their silos. This paper will firstly outline our proposed layered mapping methodological framework which aims to integrate diverse heterogenous data types to aid with decision making. Mapping techniques will include geographical mapping, informative mapping, and causal structure mapping which will provide information on key attributes of the land which coupled with intelligent computations, should enable farmers and other individuals to make decisions on the use of the land. Models may be used including geographical models, grass growth models etc. to simulate aspects of the decision-making process to create the most appropriate outcome as a solution. Secondly, the solution will integrate the knowledge in land management of grassland; this knowledge (legislations, policy, regulations, advices, etc.) must be converted into a machine-readable language so it can be used in the decision-making process carried out by the DSS.

The initial step in this methodology is the application of a case study which will firstly analyse the patterns on historical grazing event data in order to determine the most influential factors affecting grass growth in NI. A statistical analysis of these factors is also presented, using approaches such as correlation between the different variables and linear regression.

The paper is organised as follows; in Section 2 a summary of the proposed integrated methodology framework is presented. Section 3 presents the case study with an exploratory analysis of grass growth in NI along with results. Section 4 presents the discussion and conclusions of the paper.

II. A Proposed 3-Layer Mapping Methodological Framework

This section outlines a smart land management decision methodological framework consisting of three-levels of virtual mapping (Fig. 1). These will include geographical mapping, causal structure mapping and information mapping. This is aimed at capturing, measuring, formalising, evaluating and visualising the various inputs and their integration. Application of this approach will enable the anticipation of future agri-developments and crises and for these to be included in a strategic decision-making process.

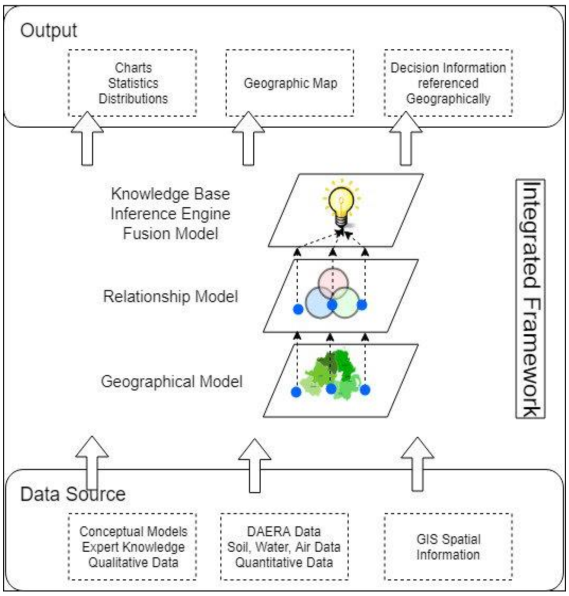

Figure 1. The methodological framework

The key inputs into the system is the expert knowledge of key players in the dairy industry, regarding legislation, law, policies, farmer experience, and government advice. This framework will be developed through the combination of rule-based algorithms and networking structures to convert knowledge from the farmers and policy makers into if-then rules, allowing the knowledge to be machine readable. The network structure of the system allows complex analysis of intricate links between diverse data. This framework will be developed with three layers in the structure: geographical model, relational model, and decision model.

The geographical layer of the structure is to be a GIS map of NI with the physical location of the various farms (approx.) and the surrounding geology of the area. It will be composed of the spatial and visual information gathered in the system and thus, the visible, physical factors leading to a decision will be considered, for example, the soil type and properties. The GIS map, as well as an explanation, will result in a wider range of audience being able to use the model, as opposed to just a textual explanation.

The second layer of the framework is the relational model, in which the data analysis of the three dimensions of sustainability factors is carried out. The relationships between the three sustainability factors are formed. This involves the relationships between the factors, as well as within the factors. The relationships can be described as a target for the model to achieve; for instance, a regulation policy may be a target. The relationships between the variables ultimately form the rules which need to be adhered to when providing a solution. The data analysis techniques used in the system will ultimately depend on the data types of the inputs in the system. These data types include quantitative, qualitative, and spatial information which provides the necessary information to produce a viable solution.

The final layer in the architecture is the knowledge base inference engine fusion model, in which the final decision is made. At this level, decision nodes are compared with each other in order to produce a solution that weighs the possibilities against each other, to ensure the best possible solution. This structure is a bottom-up hierarchy in which the outcome is only known at the final decision, after the rules are adhered to, and comparisons made.

The proposed framework emphasises the importance of work that ensures that decision models are not merely efficient or accurate about predictions, but also offer increased transparency and enable explanation. In this framework, an existing knowledge centralised DSS methodology, called Belief Rule-Base (BRB) methodology [14], [15], will be used and expanded along with multilayer network analysis [16], [17] to integrate varying data types and model the various relationships found within the data for land management policy making [18]. BRB takes advantage of Bayesian probability theory, Dempster-Shafer theory, and fuzzy logic. By integrating these frameworks together, the system can offset the negative qualities of each individual one and provide a solution that has high confidence and support.

III. Exploratory Statistics: Analysis of Grass Growth

This section summarises the preliminary case study of the exploratory statistical analysis of the most influential grazing and meteorological factors that affect grass growth in NI based on grass-land data provided by the Agri-Food Biosciences Institute (AFBI).

A. Grass-Land Management in NI

One of the specific areas of land management that this research aims to address is the grass growth prediction of farmland across NI. This means that the utilisation of grass needs to be optimised to maintain one of the farms most valuable resources and improve the sustainability of Northern Irish agriculture, which needs to remain competitive in a global market. AFBI carries out research across all livestock type farms in NI, in the GrassCheck project [19], which originally began in 1999 [19]. It involves farmers carrying out monitoring of their grass growth in order to provide high quality grass information to aid them in making decisions about their grassland management [19].

Grass growth in NI is seasonal with the highest growth occurring between early spring and late autumn, and little to no growth over winter [20]. This is due to the seasonality of sunlight, rainfall, and high temperatures that is optimal for grass growth. NI can be described as having a maritime impacted climate which results in a temperate and humid climate meaning mild summers and cool winters [21]. Grass growth prediction is challenging due to the variability of weather, thus resulting in difficulties budgeting additional fertiliser application via concentrate feedstuffs and fertiliser.

B. Grass Related Data

The data received from AFBI to date includes weather information, grass quality information, grazing event information, and grass growth information across a sample of 20 farms in their GrassCheck research program. This information is gathered by the farmers involved in the project, collected at farms spread across NI, with differences in features across the region. The data available can be classified as three types: 1) Grazing events data: includes the paddock identifier, county, date, pre-grazing cover, post-grazing cover, available, offtake, utilisation, month, total rainfall, air temperature, radiation, and soil moisture; 2) Daily grass growth data: includes the date, month, week, county, and grass growth measurement; 3) Grass quality data: includes the date, month, week, field, farmer opinion of weather conditions, dry matter %, crude protein %, Acid Detergent Factor % (ADF), Water Soluble Content % (WSC), and Metabolisable Energy MJ/kg DM (ME).

C. Exploratory Statistical Analysis and Results

In this section we provide an analysis on the grass growth data using different events such as rainfall, land utilisation, and grazing cover across different months in 2018 for a farm located in County Down. The objective of this study is to examine the patterns in data gathered in one year and determine which factors have the greatest influence on grass growth. All statistical analysis has been performed in R using packages ggPlot2 for the graphs and the correlation function when performing the correlation matrix analysis. Correlation analysis was performed to identify variables which were highly correlated with each other and with the grass growth values. A summary of the variables along with the descriptive statistics including mean, median, standard deviation, minimum value, maximum value, and kurtosis skew value. These are mapped across the months. Measurements were only collected in the months March to October summarised in Table I below. There is a total of 11 variables and 357 instances.

1) Grass growth data

The next stage of the analysis is reviewing the variability in grass growth measured on the County Down farm between March and October. Box plots have been used to illustrate the grass growth measured over the time period presented monthly in Fig. 2, in 2018. Grass growth is measured in terms of kg DM/ha/day. In Fig. 2 the boxplots provide details on the median line, minimum and maximum, outliers, and extremes for each box. The largest grass growth is observed in May, June, and August 2018.

TABLE I. SUMMARY STATISTICS OF THE VARIOUS VARIABLES ACROSS THE 8 MONTHS OF MEASUREMENTS

	Mean	**Std**	**Median**	**Min**	**Max**	**Kurtosis**
G.G	34.62	12.86	37.6	17.49	18.9	30.5
Pre.G	2938.84	398.4	2948	357.31	1940	1864
Post.G	1630.86	151.08	1580	118.61	1406	528
A	1438.9	398.44	1448	357.8	440	1864.44
O	1308.04	393.59	1326.67	325.94	213.53	1940.92
U	0.98	0.14	1.01	0.1	0.63	0.6
M	10	0	10	0	10	0
T.R	1.79	4.73	0	0	0	17.2
A.T	10.8	2.9	11.26	3.13	4.26	10.32
S.R	64.57	28.14	63.9	34.78	16.04	101.44
S.M	34.2	5.26	34.23	4.29	21.1	23.15

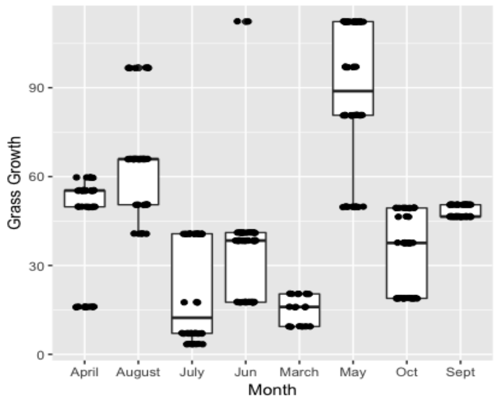

Figure 2. Boxplot of the grass growth in terms of kg DM/ha/day measured against months

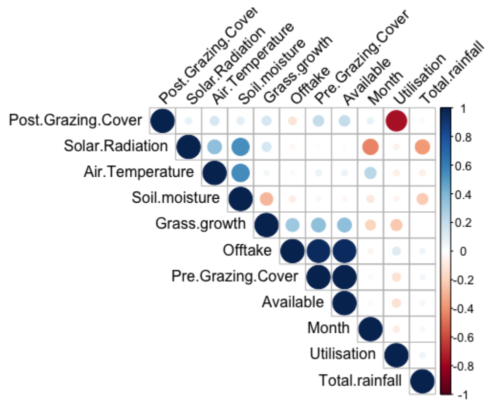

Figure 3. Correlation analysis of the various variables

2) Correlation analysis

Using the corrplot plot library, correlation analysis was performed with the results illustrated in Fig. 3. In this plot, correlation coefficients are coloured according to the value with red and blue high positive and negative correlation, and white for mid correlation. By calculating the correlation matrix, we identified variables that are highly corrected using a cut-off value of >0.50. (the closer to 1 the more the features are considered correlated).

Fig. 3 and Table II show that positive correlations were observed between grass growth and the variables offtake, pre-grazing cover and available. Positive correlations were also observed between air temperature and soil moisture along with solar radiation. A strong negative correlation was observed between utilisation and post grazing cover. Other negative correlations include total rainfall, solar radiation, soil moisture, and grass growth.

TABLE II. CORRELATIONS MATRIX OBSERVED BETWEEN GRASS GROWTH AND THE VARIABLES

	G.G	Pre.G.C	Post.G.C	A	O	U	M	T.R	A.T	S.R	S.M
G.G	1.000	0.353	0.149	0.353	0.311	-0.223	-0.181	0.009	0.043	0.150	-0.285
Pre.G.C	0.353	1.000	0.204	1.000	0.950	-0.125	-0.022	0.041	0.066	-0.014	-0.038
Post.G.C	0.149	0.204	1.000	0.204	-0.111	-0.763	0.072	-0.030	0.143	0.076	0.099
A	0.353	1.000	0.204	1.000	0.950	-0.125	-0.022	0.041	0.066	-0.014	-0.038
O	0.311	0.950	-0.111	0.950	1.000	0.115	-0.045	0.051	0.022	-0.038	-0.070
U	-0.223	-0.125	-0.763	-0.125	0.115	1.000	-0.071	0.057	-0.064	-0.056	-0.042
M	-0.181	-0.022	0.072	-0.022	-0.045	-0.071	1.000	0.032	0.230	-0.434	-0.091
TR	0.009	0.041	-0.030	0.041	0.051	0.057	0.032	1.000	-0.063	-0.375	-0.218
A.T	0.043	0.066	0.143	0.066	0.022	-0.064	0.230	-0.063	1.000	0.353	0.555
S.R	0.150	-0.014	0.076	-0.014	-0.038	-0.056	-0.434	-0.375	0.353	1.000	0.541
S.M	-0.285	-0.038	0.099	-0.038	-0.070	-0.042	-0.091	-0.218	0.555	0.541	1.000

Note: Grass.growth (G.G), Pre.Grazing.Cover (Pre.G.C), Post.Grazing.Cover (Post.G.C), Available (A), Offtake (O), Utilisation (U), Month (M), Total.rainfall (T.R), Air.Temperature (A.T), Solar.Radiation (S.R)

3) Linear regression

Linear regression analysis was performed to determine which variables had the greatest effect on grass growth. The variables from pre-grazing cover to solar radiation were considered the explanatory variables. Selection of statistically significant explanatory variables was based on the cut-off p<0.05. Table III provides a summary of the 12 variables and their measured effect on the dependent variable which is grass growth.

Measurements include the estimate, standard deviation, t-value, adjusted R-squared, and F-statistic presented as a p-value. Fig. 4 presents the graphical representation of the logistic regression plots with a b line for each variable modelled against grass growth.

The best fit models are the models built using the explanatory variables pre-grazing cover and available (adjusted R^2=0.1216). Other statistically significant variables include offtake, utilisation, month, solar radiation, and soil moisture.

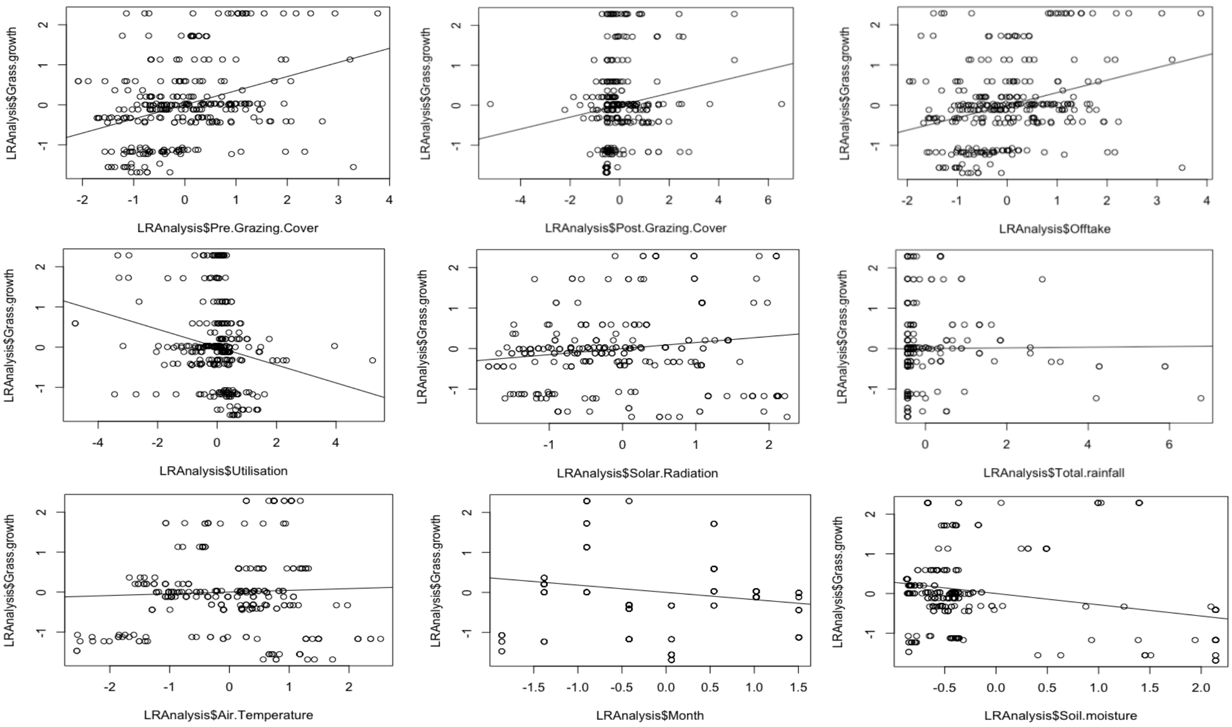

Figure 4. Relationship between grass growth and variable indicators (including grazing data through to soil moisture data) measured across 10 months in 2018

TABLE II. SUMMARY STATISTICS OF THE LINEAR REGRESSION MODELS BUILT USING THE VARIABLES WITH GRASS GROWTH THE DEPENDENT VARIABLE ACROSS THE 8 MONTHS OF MEASUREMENTS

	Estimate	Std. Error	t-value	Adj R-sqr	F
Pre.G.C	3.53E-01	5.53E-02	6.381	0.1216	7.08E-10*
Post.G.C	1.49E-01	5.85E-02	2.551	0.01883	0.01126
A	3.53E-01	5.53E-02	6.381	0.1216	7.08E-10*
O	3.11E-01	5.62E-02	5.533	0.09354	7.11E-08*
U	-2.23E-01	5.76E-02	-3.867	0.04636	0.0001366*
M	-1.81E-01	5.82E-02	-3.119	0.02952	0.001998*
T.R	8.52E-03	5.91E-02	0.144	-0.003424	0.8856
A.T	4.34E-02	5.91E-02	0.735	-0.001606	0.4631
S.R	1.50E-01	5.85E-02	2.557	0.01893	0.01107*
S.M	-2.85E-01	5.67E-02	-5.019	0.07773	9.15E-07*

IV. DISCUSSION AND CONCLUSIONS

NI is in need of sustainable land management as its land is used for agriculture, agri-food, livestock, wildlife, and tourism. However, a balance needs to be found so that the land will not suffer from constant use and the environment can be preserved. The current measures of sustainable land management are not performing as well as they are needed to ensure the environmental protection in terms of its health and agricultural yields. With the UK leaving the European Union, the results will have an impact on the way the environment is controlled. These are just a few of the challenges to sustainable land management that will need to be addressed in the decision support tool that is to be created.

Future work aims to investigate this further with more work done using the advanced data analytics, as well as the proposed 3-layer decision model to conduct advanced investigation on grass growth prediction and sustainability analysis.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Fiona Browne carried out the experiment and analysed the data. Orla McHugh conducted the research and wrote the paper, with support from Jun Liu, Fiona Browne, and Philip Jordan. All authors approved the final version of the paper.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Deborah McConnell of the Agri-Food and Biosciences Institute, for her insight and expertise that greatly assisted the research, and for the provision of data used in this paper.

REFERENCES

[1] The Home Office, "The National Emission Ceilings Regulations 2018," 2018.

[2] M. A. Liebig, *et al*., "Agriculture and climate change," *Manag. Agric. Greenh. Gases*, pp. 3-11, 2012.

[3] Department of Agriculture, Environment, and Rural Affairs, "Statistical review of Northern Ireland agriculture 2017," 2018.

[4] T. Hujer, *Design and Development of a Compound DSS for Laboratory Research*, INTECH Open Access Publisher, 2011.

[5] Department of Agriculture, Environment, and Rural Affairs, "Northern Ireland greenhouse gas emissions 2017," 2019.

[6] H. V. Delden and J. Hurkens, "A generic integrated spatial decision support system for urban and regional planning," *Int. Congr. Model. Simul.*, pp. 12-16, 2011.

[7] K. O. Wenkel, *et al.*, "LandCaRe DSS – An interactive decision support system for climate change impact assessment and the

analysis of potential agricultural land use adaptation strategies," *J. Environ. Manage.*, vol. 127, pp. S168-S183, Sep. 2013.
[8] Y. Yu, *et al.*, "Model based decision support system for land use changes and socio-economic assessments," *J. Arid Land*, vol. 10, no. 2, pp. 169-182, 2018.
[9] L. Hanrahan, *et al*., "PastureBase Ireland: A grassland decision support system and national database," *Comput. Electron. Agric.*, vol. 136, pp. 193-201, Apr. 2017.
[10] G. Anzaldi, *et al.*, "Towards an enhanced knowledge-based Decision Support System (DSS) for Integrated Water Resource Management (IWRM)," *Procedia Eng.*, vol. 89, pp. 1097-1104, 2014.
[11] A. Broome, *et al*., "Niches for species, a multi-species model to guide woodland management: An example based on Scotland's native woodlands," *Ecol. Indic*., vol. 103, pp. 410-424, Aug. 2019.
[12] E. Giusti and S. Marsili-Libelli, "A fuzzy decision support system for irrigation and water conservation in agriculture," *Environ. Model. Softw*., vol. 63, pp. 73-86, Jan. 2015.
[13] E. Bertone, *et al.*, "Extreme events, water quality and health: A participatory Bayesian risk assessment tool for managers of reservoirs," *J. Clean. Prod.*, vol. 135, pp. 657-667, Nov. 2016.
[14] J. B. Yang, *et al*., "Optimization models for training belief-rule-based systems," *IEEE Trans. Syst. Man, Cybern. Part A: Systems Humans*, vol. 37, no. 4, pp. 569-585, 2007.
[15] J. Liu, *et al*., "A novel belief rule base representation, generation and its inference methodology," *Knowledge-Based System*, vol. 53, pp. 129-141, 2013.
[16] F. Browne, *et al*. "Investigating the impact human protein-protein interaction networks have on disease-gene analysis," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 3, pp. 455-464, 2018.
[17] W. K. Chen, *Linear Networks and Systems*, Belmont, CA, USA: Wadsworth, 1993, pp. 123-135.
[18] J. C. Helton, *et al.*, "An exploration of alternative approaches to the representation of uncertainty in model predictions," *Reliab. Eng. Syst. Saf.*, vol. 85, no. 1-3, pp. 39-71, Jul. 2004.
[19] V. Brocard, *et al*., "Eurodairy: A bottom-up approach to transfer innovations on grass to European farmers," *Grassland Science in Europe - Sustainable Meat and Milk Production from Grasslands*, vol. 23, 2018.
[20] D. Läpple, *et al*., "Extended grazing: A detailed analysis of Irish dairy farms," *J. Dairy Sci*., vol. 95, no. 1, pp. 188-195, Jan. 2012.
[21] C. Hurtado-Uria, *et al*., "Relationships between meteorological data and grass growth over time in the south of Ireland," *Irish Geogr.*, vol. 46, no. 3, pp. 175-201, 2013.

Orla McHugh received the BSc degree from Ulster University, Jordanstown, in 2018, in Computing Science. She is currently pursuing an interdisciplinary PhD degree with the School of Computing and School of Geography and Environmental Sciences. Her research areas include data analytics, decision support systems, and knowledge integration.

Jun Liu received the BSc and MSc degrees in applied mathematics, and the PhD degree in information engineering from Southwest Jiaotong University, Chengdu, China. He is currently a Reader in Computer Science at Ulster University. He has been working in the field of Artificial Intelligence for many years. His current research is focused on: 1) knowledge-centralized data analytics under uncertainty for sensing decision making, with applications in management, engineering, and industry etc. (e.g., safety and risk analysis; policy decision making; security/disaster management; and heath care and smart home); 2) logic and automated reasoning methods for intelligent systems including software verification and automated theorem proving. He has published more than 80 papers in journals indexed by the SCI and more than 80 contributions in International Conferences.

Fiona Browne is Head of Artificial Intelligence at Datactics with over 15 years research and industrial experience. Her research interests focus on artificial intelligence, information fusion, reasoning under uncertainty, machine learning, pattern discovery and visualisation and their application in multi-disciplinary fields. She received the BSc (Hons.) degree in Computing Science from the Ulster University. She received a PhD on Artificial Intelligence in Bioinformatics from Ulster University for her work on developing integrative data analysis models and tools for the prediction of protein-protein interaction networks. She has published over 50 articles in peer reviewed international journals and conference proceedings.

Phil Jordan is Professor of Catchment Science in the School of Geography and Environmental Sciences, Ulster University. He has BSc Hons degree in Geography and a PhD degree in environmental science, from the University of Leeds and Ulster University, respectively. His research focuses on the dynamics and fate of nutrients and sediments in soil-water systems, approaches to high-resolution environmental data capture and analysis, and agri-environmental policy.