JNCC Report 737

Towards Indicators of Soil Health

Annex 3:

Developing conceptual ecological models and Bayesian belief networks for a soil health indicator

Hannam, J., Harris, M., Hutchison, J., Withers, A., Harris, J.,
Hoskins, H., Way, L. & Rickson, J.

June 2023

For further information please contact:
Joint Nature Conservation Committee
Quay House
2 East Station Road
Fletton Quays
Peterborough PE2 8YY
https://jncc.gov.uk/
Communications@jncc.gov.uk

This document should be cited as:
Hannam, J.,[1] Harris, M.,[2] Hutchison, J.,[2] Withers, A.,[2] Harris, J.,[1] Hoskins, H.,[2] Way, L.[2] & Rickson, J.[1] 2023. *Developing conceptual ecological models and Bayesian belief networks for a soil health indicator. JNCC Report 737: Annex 3.* JNCC, Peterborough. https://hub.jncc.gov.uk/assets/71cece04-eef3-4d34-b118-33ddad50912c#jncc-report-737-annex-3-developing-models.pdf

Author affiliations:
[1] Cranfield University
[2] JNCC, Peterborough/Aberdeen

Contents

1 Policy context

There have been policy statements and parliamentary debates (see Main Report), which pave the way towards an increased interest in a framework through which soil health can be concluded upon. Subsequently, JNCC and Cranfield University have been commissioned by Defra to develop an indicator of soil health.

For this project, soil health is defined as follows:

> "Soil health is the ability of the soil to perform its functions and to deliver ecosystem goods and services. The range of functions and ecosystem services provided should reflect the different capabilities of different soils – a 'healthy' soil is therefore one in which ecosystem services are provided at an acceptable level given inherent underlying constraints and the purpose of the land" (see Glossary of the Main Report).

From this definition, the indicator therefore needs to include an estimate of the level of soils' contribution to delivery of relevant ecosystem services, as well as an assessment of whether those delivery levels are "acceptable" given the soil type, environmental context, and land use.

This appendix describes the analysis and modelling used to develop an initial proof-of-concept for both the indicator and a dashboard to make the outputs accessible and interpretable for users.

2 Scope

With the limited time available for this pilot phase of the project, not all factors (e.g. all land use types, all soil types and all pertinent ecosystem services) possible for inclusion in a model of soil health have been included. Decisions behind factors included are explained below.

3 Ecosystem services included

The 25 YEP takes an ecosystem services approach to contextualise the plan's goal to "help the natural world regain and retain good health" and the Natural Capital Ecosystem Assessment conceptual ecological model (In prog.) aims to provide a mechanism by which ecosystems can be understood in this way. It is therefore taken into consideration how a soil health indicator could:

a. be used to understand whether the above goal is being met, in a manner that fits with the approach being taken by government, and
b. use the mechanism outlined by the NCEA model to "describe the system that is being monitored, identify the ecosystem components, processes and functions that compose the natural environment, and how they interact.", taking this to the latter phase which "consider[s] the drivers, pressures and responses acting on the system".

It is believed that this is the first project to adopt the NCEA CEM to understand a component of the system.

Soils contribute to the delivery of many ecosystem services, with healthy soils having properties that enable them to do so to a greater extent than unhealthy soils. Four such services were identified as priorities for the project (climate regulation through soil carbon storage; water regulation through soils' contribution to runoff reduction; maintenance of biodiversity through soil biodiversity; and food and fibre production potential through soils' contribution to land capability for agriculture). Two of these have been completed through to the final output at the time of writing (climate regulation and water regulation). Other services that were considered include water quality, cultural heritage, human disease regulation and genetic heritage. The four selected were chosen through discussion in project and steering group meetings as those considered to have the most significant benefits, be most affected by the soil system and be most feasible to model based on current data and expertise notwithstanding the fact that ecosystem services are important across the board and so these four should be recognised at a starting set to which others can be added.

Each of the four services selected represents an overarching, high level function. Intermediate functions (such as infiltration rather than water regulation) are included within the networks and can be reported on alongside the 'headline' services.

For the purposes of this project, ecosystem services are defined as "The contributions that ecosystems make to human well-being, and arise from the interaction of biotic and abiotic processes" (Haines-Young 2010), and each of the four services selected for modelling are defined as:

- **Climate regulation:** The ability to act as a net sink for GHG, for example through sequestration and storage of carbon, and prevention of carbon losses to the atmosphere. In the context of this project, focus is only on soil carbon storage, not other regulation via other means or of other GHGs.
- **Water regulation:** Control of excess water (flooding) and deficit of water (drought) in the environment. In the context of this project, focus is only on soils' contribution to this through runoff reduction.
- **Food and fibre production potential:** Soils' contribution to land capability for agriculture and ability to support food and fibre production.
- **Maintenance of soil biodiversity:** Maintenance of "The variation in soil life, from genes to communities, and the ecological complexes of which they are part, that is from soil micro-habitats to landscapes" (European Commission 2010, in line with the CBD's definition of biodiversity).

4 Ecological models

4.1 Rationale

A conceptual model is a graphical representation of a system, identifying components of the system and the relationships between them, they offer a means to represent complex environmental systems such as soils. Conceptual ecological models were developed to represent soils' contribution to the delivery of each of the four ecosystem services described above. They represent links and interactions between soil properties and environmental variables and intermediate ecological processes that ultimately contribute to soils' delivery of the ecosystem service. They simplify the complexity of the soil system by identifying key components, and their interactions that can affect the most change in the delivery of the ecosystem service. The conceptual model is intended to be used to demonstrate how one can arrive at an outcome indicator for a specific ecosystem service (i.e. how well the soil is delivering against this service).

Soils' contribution to the delivery of the ecosystem service in some cases will be constrained by the inherent characteristics of the environment such as the soil type and land use (e.g. sandy soils have lower capacity to store organic carbon than clay soils). These parameters are included in the conceptual model so that these boundary conditions can be taken into account when reporting on the indicator for the ecosystem service.

The variables within the conceptual models are expected to be represented by available data sources, that either directly measure the property or variable, or can be used as a reasonable proxy. These data sources can be spatial representations of the variable (e.g. rainfall) or soil property (e.g. soil organic carbon) or can be input directly by an observer or user where a dataset or information is not readily available (e.g. tillage).

Outputs from the Annex 1 literature review reported key soil properties that can be used to indicate soil health (Annex 1, Table 9 and 10) and their contributions to the delivery of ecosystem services (Annex 1, Table 5 to 8). This was used as a starting point to determine key properties that together can represent or inform the delivery of the ecosystem services identified.

For climate regulation, soil carbon storage was used as a contributor to climate regulation. This is a soil property that can be directly measured and was identified as a key property reflecting soil health and delivering a climate regulation ecosystem service in Annex 1. In the absence of a measurement of soil carbon at the required scale of reporting (e.g. land parcel) a conceptual model was developed so that this could be reported. The conceptual model is based on known controls of soil organic carbon content. A value for soil organic carbon arises from the balance between inputs from plant and animal biomass, roots, and organic amendments, and turnover or loss via microbial respiration, disturbance (tillage or drainage) and soil erosion (e.g. Paustain *et al.* 2019). The capacity for carbon storage in soils is also determined by the soil type, with soils with higher clay content offering greater protection from mechanisms of soil carbon losses (Prout *et al.* 2021). Land use also affects the capability of soil carbon storage. It has been shown empirically that for the same soil, it has lower organic carbon contents in cropland compared to grassland or semi-natural habitats.

For water regulation, runoff generation from soil was used as an indicator of soils' contribution to water regulation through flood mitigation. The conceptual ecological model was developed from the concepts of the Hydrology of Soil Types (Boorman *et al.* 1995) where run-off generation from soil is determined by the amount of water transmitted across the surface or laterally within the soil profile (sub-surface flows). Soil properties from the literature review in Annex 1 (water holding capacity, infiltration rate organic matter and soil structure) also informed the conceptual model. Surface runoff occurs when the infiltration rate of water into soil is low, for example when the soil is completely sealed (built environments) or compacted by heavy machinery or livestock. Sub-surface flows within the soil are characterised by how quickly water moves within the soil profile due to infiltration rate and the presence of any slowly permeable layers, determined by soil structure, organic matter, and soil texture. Together surface and lateral runoff determine the amount of water entering the stream network from soils, and soils with lower runoff potential have greater capacity for water regulation.

4.2 Methods

Outcomes from the logical sieve (Table 10, Annex 1) highlighted ranked soil properties that were most pertinent to soil health and were relevant to the delivery of ecosystem services. These soil properties provided the basis for the initial development of the conceptual ecological models, although not all of these highlighted soil properties would be relevant for the delivery in all the ecosystem services considered here.

The conceptual models were constructed based on expert knowledge of mechanistic and empirical processes that link environmental variables to ecological processes to the delivery of an ecosystem service. A draft model was developed in an interactive environment (stormboard.com) and reviewed by the project team, eventually providing consensus on the variables included and links to ecological processes within the model.

An example ecological model for the climate regulation ecosystem service is shown in Figure 1. This model links various environmental properties (rainfall, temperature, land management) and soil properties (soil organic matter, fertilisers, biological activity – all three identified in Annex 1 as pertinent to soil health) resulting in soil carbon inputs and turnover to the final delivery of climate regulation through soil carbon storage. An ecological model for water regulation is shown in Figure 2. It links inherent properties of soil that determine infiltration and water movement with rainfall to determine surface and subsurface runoff. This ends with total runoff from soil, where lower runoff amounts would contribute to the delivery of water regulation.

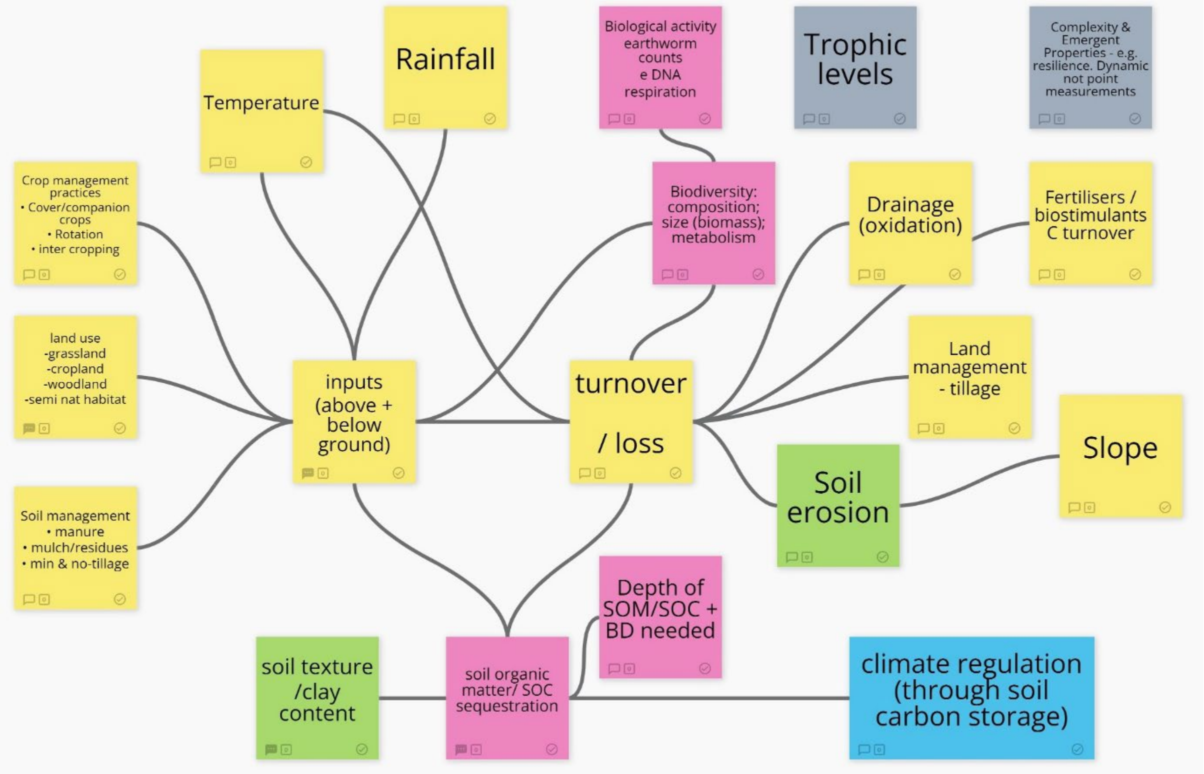


Figure 1. A conceptual ecological model showing soils' contribution to the delivery of climate regulation as an ecosystem service through soil carbon storage. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (pink); Ecosystem Service outcome indicator (blue); other consideration (grey).

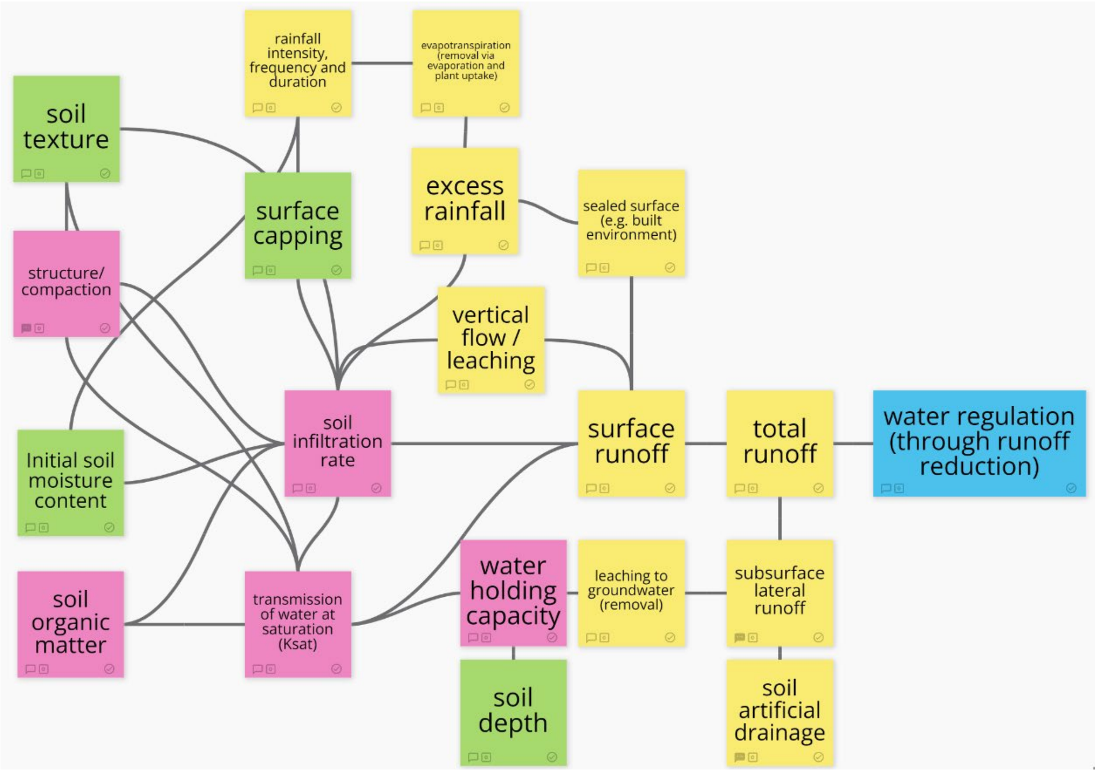


Figure 2. A conceptual ecological model showing soils' contribution to the delivery of flood mitigation as an ecosystem service through soil water regulation. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (pink); Ecosystem Service outcome indicator (blue); other considerations (grey).

4.3 Conceptual Ecological Model (CEM) uncertainties

The ecological models were initially constructed from the extensive outputs of the logical sieve (Annex 1) that highlighted key soil properties that have relevance for measuring soil health and that can link to the delivery of ecosystem services. This process instils reasonable certainty that the soil properties selected from the literature review outputs (Annex 1) can deliver indicators or outcomes for soils' contribution to the selected ecosystem services.

The inclusion of additional environmental properties and processes is based on mechanistic and empirical understanding of the relationships between the variables driving ecological processes. For example, some components of the ecosystem models are represented by well-established mechanistic models (e.g. RothC models soil organic carbon turnover). The inclusion of these relationships is conceptually simpler in the ecosystem model to avoid excessive parameterisation and demand for data. Additionally, well-established empirical relationships are also built into the model (e.g. soil drainage and soil erosion reduce soil carbon content).

For some services such as maintenance of soil biodiversity, the links to ecosystem service delivery is less well understood and therefore the model is simpler in the representation of the ecosystem service delivery. There may be greater uncertainty in the indicator for this service.

Further refinement of the links and direction of change between one or more variables and ecosystem process can be achieved through extensive literature and meta-data analyses. Models will also need to be refined as our understanding of more complex relationships in the soil system are developed.

5 Land use and soil types included

There are likely to be different responses and contributions that different soil types make to the outcomes of the ecosystem services selected. There are different types of classification systems for soil, and all have different relationships to representing soil ecosystem services and functions (Table 1). Taxonomic classification systems such as the Soil Survey of England of Wales or the World Reference Base use hierarchical systems based on key soil processes that give rise to morphological differences (diagnostic features) between soils. However, these processes do not necessarily map onto or completely represent a specific ecosystem service. Other systems use 'functional' classifications, such as the Hydrology of Soil Types or Forest Soil Type, these often only map onto one ecosystem service, or are bias towards certain habitats. Some ranges in ecosystem service contribution for each soil type could be determined should sufficient data exist but it is likely that some soil types would be difficult to associate with a suite of ecosystem services. Finally, "attribute-based" classification such as soil texture classes, are not soil types per se, but group soils by a specific soil property. These properties do not necessarily link to a soil function or ecosystem service but can provide useful information on the boundary conditions for the mechanisms that contribute to a particular ecosystem service or can be used as a parameter that feeds directly into a model. For example, it is known that clayey soils have a higher affinity to store carbon and drain water more slowly than sandy soils.

Table 1. Soil classification system examples, alongside considerations for their use in representing ecosystem services and whether ranges of ecosystem service delivery can be detected.

Soil classification system [examples]	Considerations for representing ecosystem services (ES)	Can ranges of ES delivery be detected?
Taxonomic soil classification systems [Soil Survey of England and Wales; Brown Earth] [World Reference Base; Cambisol]	• Don't always map well onto soil functions or ES	• Some ranges could be determined from existing categories/soil types • Ranges would need to be developed
"Functional" classification systems [Hydrology of Soil Types; HOST classes] [Forest Research Soil Types]	• Often focus on one ES or function • Habitat specific	• Some ranges could be determined from existing categories/soil types • Ranges would need to be developed
"Attribute-based" classification systems [Soil texture class; Sandy loam]	• Key attributes not necessarily linked to function or ES but provide some boundary condition • Often feeds directly into model as a parameter	• Different attributes drive different ES • Some attributes drive multiple ES

6 Bayesian Belief Network modelling

6.1 Rationale

Bayesian Belief Network (BBN) modelling can be used to build on an ecological model, by adding numerical and probabilistic values. This allows for the prediction of the likely consequences of changes to inputs of the system, which can be particularly useful for applications such as predicting the likely effects on ecosystem services of management actions and environmental factors. With the aim of this project being to model soils' contribution to delivery of ecosystem services as a representation of soil health, BBNs were therefore selected as a useful tool.

Use of BBNs mean it is possible to integrate different types of knowledge, including spatial data and stakeholder expertise. Whilst data on many soil factors and the relationships between them are available, this is not the case for everything. The ability to make use of expert knowledge where required, without losing out on the value of the data that are available, is therefore key to developing a complete indicator. The ability to use either or both types of input will allow for improvement of the model in future if new data sources become available.

They are particularly useful in cases where there are lots of different interacting factors that lack data, but that are reasonably well understood from an ecological perspective in terms of direction of change. This is the case for soils.

The fact that BBNs are based on an underlying conceptual model provides a good visual representation of what the BBN involves, whilst the more complicated mathematics is only introduced at a later stage. This is useful for stakeholder engagement and communication of outputs.

The probabilistic nature of BBNs also allows them to report on uncertainty. The most likely outcome of a particular combination of factors is often not the only possible outcome. Understanding the probability associated with this outcome gives users a more complete understanding of risk.

Overall, BBN modelling was selected as the approach to take forward due to its ability to make predictions from complex networks, its flexibility in terms of data inputs and its communicability.

6.2 Methods

BBN modelling took place in RStudio, using the R package 'bnlearn' (Scutari 2010). The ecological model described in the previous section of this report was used as a basis to form a directed acyclic graph (the graphical depiction of the network of spatial data and management decisions through to contribution to ecosystem service delivery) for each service. Climate regulation is used here as an example to illustrate this process (Figure 3). Some small adaptations (such as the removal of the fertiliser and soil management nodes, and the splitting of crop management into several separate nodes for each management option within the climate regulation network) were made to the ecological models, based on data constraints and practicalities of the format that inputs are required to be in to feed into the model. Where possible, spatial data on each input were compiled, making use of a random sample of 100,000 points to assess node states and probabilities (Table 2). Expert input was then used to define the probabilities for nodes which did not have data, and the relationships between inputs, in order to develop a model that could predict soils' contribution to likely ecosystem service delivery based on a given set of inputs. This allowed it to be queried based on specific inputs from different land parcels via the dashboard (Section 7).

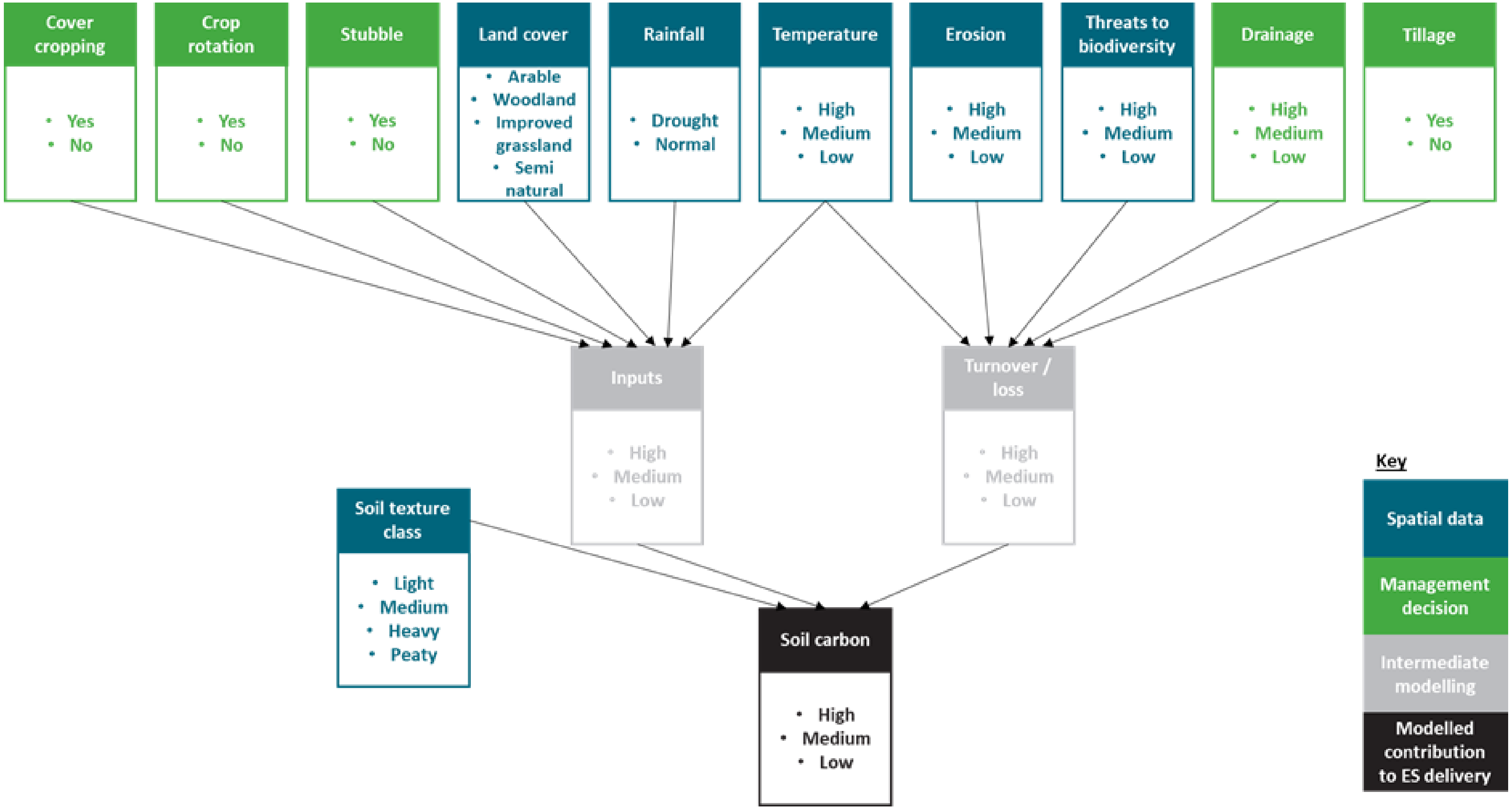


Figure 3. A directed acyclic graph (DAG) showing the nodes and node states used in the BBN modelling for the climate regulation through soil carbon storage model. Inputs for the nodes in blue are based on spatial data, whilst nodes coloured green will rely on users to input a land management decision (e.g. cover cropping) via the dashboard, in order to model soils' contribution to the final ecosystem service delivery.

Table 2. Node data sources for the climate regulation through soil carbon storage model and the water regulation through run off reduction model.

Model	Node	Levels	Data source
Climate and water regulation	Land use	Woodland, arable, improved grassland, semi natural	CEH Land Cover Map 2019 (Morton *et al.* 2020), see table 4
Climate regulation	Drought	Yes / no	HadUK-Grid gridded and regional average climate observations for the UK (Met Office 2018), with the lowest 20% of rainfall data points considered as drought
Water regulation	Rainfall	High / medium / low	HadUK-Grid gridded and regional average climate observations for the UK (Met Office 2018)
Climate regulation	Temperature	High / medium / low	HadUK-Grid gridded and regional average climate observations for the UK (Met Office 2018)
Climate regulation	Erosion	High / medium / low	Pan European Soil Erosion Risk Assessment – PESERA (Kirkby *et al.* 2008, 2004)
Climate regulation	Threats to soil biological function	High / medium / low	Potential threats to soil biodiversity in Europe (Orgiazzi *et al.* 2016, 2015)
Climate (topsoil only) and water regulation	Texture class	Light / medium / heavy / peaty	LandIS NATMAP Topsoil Texture and Subsoil Texture (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022]), see Table 5

Model	Node	Levels	Data source
Water regulation	Evapotranspiration	High / medium / low	EA_PETI Environment Agency Potential Evapotranspiration Datasets January to March 2022 (Environment agency © 2021)
Water regulation	Organic matter	High / medium / low	LandIS NATMAP Carbon (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022])
Water regulation	Initial soil moisture	High / medium / low	Countryside Survey (CS) topsoil moisture data (Henrys *et al.* 2014)
Climate regulation	Cover cropping	Yes / no	Management choice – will be input by the user
Climate regulation	Crop rotation	Yes / no	Management choice – will be input by the user
Climate regulation	Stubble	Yes / no	Management choice – will be input by the user
Climate and water regulation	Drainage	High / medium / low	Management choice – will be input by the user
Water regulation	Tillage direction	Up/down slope / across slope	Management choice – will be input by the user
Climate regulation	Reduced tillage	Yes / no	Management choice – will be input by the user
Water regulation	Compaction	High / medium / low	Management choice – will be input by the user

Model	Node	Levels	Data source
Water regulation	Capping/Sealing extent	High / medium / low	Management choice – will be input by the user
Climate regulation	Inputs	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Climate regulation	Turnover / loss	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Carbon regulation	Climate regulation via soil organic carbon storage	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Water regulation	Excess rainfall	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Water regulation	Soil infiltration rate	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Water regulation	Surface runoff	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Water regulation	Subsurface lateral runoff	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes
Water regulation	Water regulation via runoff reduction	High / medium / low	Calculated by the model, based on expert defined relationships between input nodes

For the land-use node, land-use mapping was taken from CEH Land Cover Map 2019 (Morton *et al.* 2020), but land-use categories were condensed into just four classes (arable, improved grassland, woodland and semi natural, Table 3). Land-use categories that did not fall into one of these four classes were considered out of scope for the proof-of-concept. This includes coastal and urban land cover types, which were excluded from this initial stage due to the very different factors influencing service provision compared to non-urban, terrestrial land cover types. It is recognised that urban soils will be important to include in future to ensure representation of all soils. The decision to condense the remaining land cover types into four categories was taken to simplify the modelling approach. A similar approach was used in the water regulation model; however, three categories (arable, urban, or other) were used in this model (Table 3).

Table 3. Land cover classes from the original data source (Morton *et al.* 2020) aggregated into the classes used within this project.

Land cover classes from Morton *et al.* 2020	**Land use classes used in the current project for the carbon storage model**	**Land use classes used in the current project for the water regulation model**
• Deciduous woodland • Coniferous woodland	Woodland	Other
• Arable	Arable	Arable
• Improved grassland	Improved grassland	Other
• Neutral grassland • Calcareous grassland • Acid grassland • Fen • Heather • Heather grassland • Bog • Inland rock	Semi natural	Other
• Urban • Suburban	Na	Urban
• All other land use classes	Na	Other

For the model of climate regulation through soil carbon storage, soil texture mapping for the relevant node was taken from the LandIS NATMAP Topsoil Texture dataset (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022], n.d.), but soil texture types were condensed into four simple texture classes (light, medium, heavy, and peaty, Table 4). Again, the decision to condense these into just four categories was taken based on the need to simplify the modelling approach. For the model of water regulation through water run off reduction, soil texture mapping was taken from the LandIS NATMAP Topsoil and Subsoil Texture datasets (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022], n.d.), and as for the climate regulation model, these soil texture types were also condensed into four simple texture classes (light, medium, heavy, and peaty, Table 4).

Table 4. Soil texture classes from the original data source (LandIS NATMAP Topsoil Texture and LandIS NATMAP Subsoil Texture (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022])) aggregated into classes used within this project.

Soil texture classes used in LandIS NATMAP Topsoil Texture (Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022])	Top and subsoil texture classes used in the current project
Clay	Heavy
Silty clay	
Clay loam	Medium
Silty clay loam	
Sandy clay loam	
Fine sandy loam	Light
Medium sandy loam	
Coarse sandy loam	
Loamy fine sand	
Loamy medium sand	
Loamy coarse sand	
Fine sand	
Medium sand	
Silt loam	
Fine sandy silt loam	
Medium sandy silt loam	
Course sandy silt loam	
Peaty sand	Peaty
Loamy peat	
Peaty loam	
Peat	

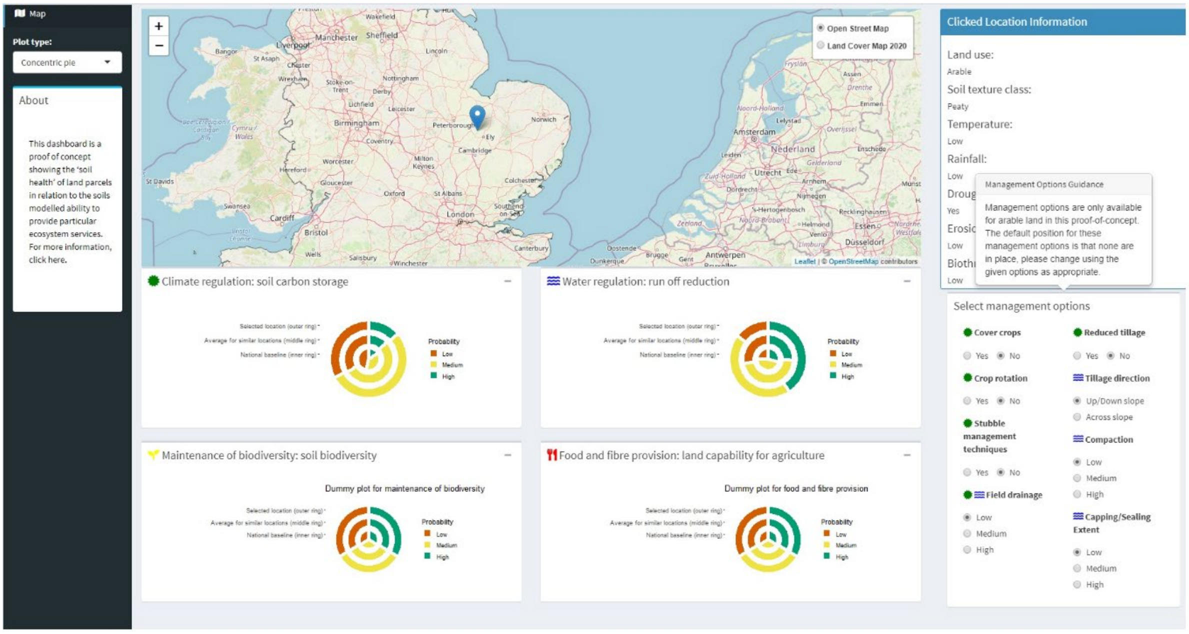
7 Dashboard visualisation

In order to present results in a visual and interactive manner, results from the modelling were connected to an R Shiny application (Figure 4). This allows users to select land parcels, input current or potential management decisions or local data (where available) and view the predicted level of ecosystem service provision. In the current version, data is presented as a series of concentric pie charts. There are three rings in each chart. The inner ring represents the average across all land parcels within the UK. The middle ring represents the average for all land parcels of the same land use and soil type as the land parcel selected. And the outer ring represents the land parcel selected on the map, reacting to input from the drop-down boxes on management decisions taken. Each ring is composed of three segments, which represent the probability that soils' contribution to ecosystem service delivery given the particular combination of inputs selected is 1) low, 2) medium and 3) high.

The ecosystem services selected for inclusion are made up of many contributing functions. For example, water regulation through soils' contribution to runoff reduction is made up of functions, including water storage. Such contributing functions are not available to view as part of the headline results via the dashboard at this stage, but a more detailed analysis of this type can be taken from the model itself.

It is hoped that the information provided by the dashboard will be helpful for a wide range of use cases, including:

- individual landowners/landholding organisations for informing land management
- individual farmers who are participating in agri-environment schemes
- governments, country conservation bodies and the Office for National Statistics, to provide information about national status, environmental policy, and national land use decisions
- local councils, for informing planning decisions and mitigation options



Plot type: Concentric pie

About

This dashboard is a proof of concept showing the 'soil health' of land parcels in relation to the soils modelled ability to provide particular ecosystem services. For more information, click here.

Clicked Location Information

Land use: Arable
Soil texture class: Peaty
Temperature: Low
Rainfall: Low
Droug… Yes
Erosi… Low
Biothi… Low

Management Options Guidance

Management options are only available for arable land in this proof-of-concept. The default position for these management options is that none are in place, please change using the given options as appropriate.

Select management options

Cover crops: ○ Yes ● No
Crop rotation: ○ Yes ● No
Stubble management techniques: ○ Yes ● No
Field drainage: ● Low ○ Medium ○ High
Reduced tillage: ○ Yes ● No
Tillage direction: ● Up/Down slope ○ Across slope
Compaction: ● Low ○ Medium ○ High
Capping/Sealing Extent: ● Low ○ Medium ○ High

Climate regulation: soil carbon storage

Water regulation: run off reduction

Maintenance of biodiversity: soil biodiversity

Dummy plot for maintenance of biodiversity

Food and fibre provision: land capability for agriculture

Dummy plot for food and fibre provision

Figure 4. A screenshot of the dashboard produced to present the results of the proof-of-concept. The explanatory pop-up activated by clicking in the 'About' section gives an overview of the app, its uses, and potential future developments.

8 BBN uncertainties

Whilst the models are useful representations of a complex system that would otherwise be difficult to analyse and visualise, as with all models, they are based on a number of assumptions with inherent uncertainties, which are necessary to understand for correct interpretation of results. These include the fact that they are an expert driven simplification, the fact that this stage is simply the first iteration so it is expected that further improvements will be made, and the fact that the results will be qualitative not quantitative.

Being based on expert judgement means that the models are not primarily driven by data. Whilst some data are used where available at an appropriate scale, they are simply not available for many of the nodes of the ecological models. This means that some nodes would be missed out if one were to follow a method that relied only on data. It also means that the interactions between nodes must be based on expert judgement instead of modelled from the data. In cases where it is known or suspected that a property has a significant impact on the final contribution of soils to the ecosystem service, its inclusion based on expert judgement is therefore likely to be a more accurate representation of reality than its exclusion. However, it does open the model up to the risk of misrepresentation or bias in the results based on the experts' perceptions. Whilst the model is a visual representation of the experts' understanding of how the model works, it could be biased by how the experts interpret the world. Results should therefore be viewed as an estimate or 'best guess' of how the system works, rather than an absolute. Whilst this is the case for all models (even those based on data), it should be especially recognised when interpreting results of BBNs due to the added uncertainty brought about by the inclusion of expert judgement.

Similarly, this pilot phase of the project means that the current models should only be viewed as a first iteration. Subsequent iterations would be needed to test the model results against real scenarios, in order to sense check and better understand the accuracy of the results. This would also help to edit the model inputs to best represent real outputs. It will also be important in the next stages of the project to get input from a wider variety of experts and stakeholders, as the current phase has utilised only a small number of experts from within the project group to create an initial set of models that can form the basis for further improvements. There are also a number of manual quality assurance steps that could take place as part of further project development, such as investigating the extent of variation within each land class.

As well as this, the outputs of the model will be qualitative, not quantitative. This means that they will be able to show relative variation in provision of a service but will not give results in terms of real numbers. There is again potential for future improvements of the model (such as with additional data that may become available in the future, through sources such as the Environmental Land Management schemes and Soil Survey England) to be able to provide quantitative results, but the current, expert-judgement-driven version cannot. In the current version, results simply report on whether a particular set of inputs are most likely to lead to high, medium, or low provision of that ecosystem service. The probability of that combination of inputs leading to each category can also be viewed. As it is based on modelled probabilities, results showing that there is a high probability of a situation being linked to high provision of an ecosystem service does not mean that this will definitely be the case; once again, results should be viewed as estimates, not absolutes.

It should also be noted that soils are part of a very complex system of interacting biological, chemical, and physical factors. Whilst the ecological models and BBNs capture a set of factors thought to be most significant to the ecosystem service in question in each case have been selected, it would never be possible to capture the complexity in its entirety. There are therefore undoubtedly factors that do affect soils' contribution to ecosystem service delivery that are not included in the model.

Overall, whilst the models are associated with uncertainties, if users interpret the results correctly (i.e. as estimates) and understand how and why these uncertainties arise, they should not affect the usefulness of the results, for example in estimating the direction of change associated with differing management practices in a given context.

9 Challenges and solutions

The main challenge faced by this project was the complexity of the soil system. Each ecosystem service to which soil contributes delivery is the result of a huge range of different interacting factors. These factors include both inherent properties of the soil such as texture class, properties that change geographically such as average rainfall or temperature, and properties that vary depending on management practices. Each factor may have a different level of impact on final service delivery and interact with each of the other factors in the network whereby the same value for a particular node may have differing levels of impact depending on the value of another node. For many of these factors, data are not available. This was particularly the case for those based on management practices, which is significant as a challenge when considering that a likely use case for the indicator is comparing the estimate for a land parcel's current soil health with the estimate for the same land parcel under different management scenarios.

However, experts do have a reasonably good ecological understanding of directional change in this context. The BBN approach presented a solution to this challenge, allowing for this understanding to be incorporated into the modelling rather than relying entirely on data as traditional modelling would. The use of conceptual ecological models to simplify the network also meant that it was possible to use a simplified version of reality, which allows for initial estimates to be made, whilst also leaving open the option for adding further complexity in further iterations of the model as understanding develops.

There are challenges around how changes seen on the dashboard are understood. This will require appropriate communication of likely timescales for change that a user can expect to see on the ground which reflect the changes to the dials/pie charts (Figure 4) provoked by changing management practise. Additionally, there are other factors than may lead to the outcome of an intervention not being as predicted by the dashboard. It must be ensured, therefore, that users understand that changes to management interventions will likely affect a soils ability to contribute to an ecosystem in the positive or negative manner depicted but that overall if this service is being delivered to a better or worse degree will include other contributing factors.

10 Future research

As previously mentioned, the project reported on in this document is currently at a pilot stage. It is expected that further work, including consultations and iterative improvements to the basic models presented here, will be required before results are used to inform real decisions in the field. Future research should include a ground truthing process, a wider expert consultation process, consideration of additional nodes and addition of extra ecosystem services.

Ground truthing could be used to help compare results from the dashboard with field data or sampled values. This could help to give an idea of model accuracy. It also presents potential to train the model with data and reduce the reliance on expert input and related risk of bias.

Filling in current data gaps and allowing for less of the model to be based on expert judgement than is currently the case should therefore be a priority for future research. While some of this could come from field data, which may become available in the future through sources such as the Environmental Land Management schemes and Soil Survey England, other data could come from consulting papers to obtain experimental values on interactions between different factors where these already exist.

A wider expert consultation process could also reduce the risk of bias from expert judgement, by averaging across many experts, rather than only utilising input from those involved within the project. This will be a crucial aspect to improving the model beyond the pilot phase. In addition, the consultation should form part of an iterative process, where experts also provide input on model outputs (not just inputs and their interactions), in order to identify whether results match up with those that would be expected based on their expertise. This would again help to improve and calibrate the models to make them more useful for end users.

The current BBNs make use of relatively simple conceptual ecological models, with fewer nodes than would affect ecosystem service delivery in reality. Another avenue for future work could involve expanding the networks to include more factors. The inclusion of more data will improve the nodes and the number that can be included. However, this must be balanced with a need to retain simplicity for communication purposes and for practicality of developing and running the models, which may become computationally intensive and have reduced sensitivity to changes in input factors if there are too many factors feeding in.

The four ecosystem services selected in this pilot study are not the only ecosystem services to which soil contributes. It may also be useful for future work to develop similar models for other ecosystem services of interest, such as water quality, cultural heritage, human disease regulation and genetic heritage. Some of these are likely unfeasible at the current time due to a lack of understanding but may be possible in future as the field develops.

This project aspires to incorporate site-based soil data in the future; such incorporation will require careful consideration as to how it is presented within a dashboard and how it interrelates with any management option scenarios selected.

The components of the proof-of-concept indicator can all be upgraded as knowledge improves. The sift of properties with proven relationships to ecosystem services can be re-run to pick up new research. The models of soil function can be improved with wider expert and science input. The modelling technique (currently Bayesian) can adapt and include more empirical elements as more soil data becomes available. And the way the indicator is presented on the dashboard can improve as data allows range rather than probability information to be used. But crucially, despite these areas being open to improvement, it is still functionally useable presently.

References

Boorman, D.B., Hollis, J.M. & Lilly, A. 1995. Hydrology of Soil Types: A Hydrologically-Based Classification of the Soils of the United Kingdom. Institute of Hydrology Report No. 126. Institute of Hydrology: Wallingford.

European Commission. 2010. Soil biodiversity: functions, threats and tools for policy makers.

Environment Agency. 2022. EA_PETI Data, Environment Agency Potential Evapotranspiration Dataset.

Haines-Young, R. 2010. Proposal for a Common International Classification of Ecosystem Goods and Services (CICES) for Integrated Environmental and Economic Accounting. Report to the European Environment Agency 30.

Henrys, P.A., Keith, A.M., Robinson, D.A., Emmett, B A. 2014. Model estimates of topsoil moisture [Countryside Survey]. NERC Environmental Information Data Centre. DOI: https://doi.org/10.5285/8db84900-5fdb-43be-a607-e56c843d9b87

Kirkby, M., Irvine, B., Jones, R. & Govers, G. 2008. The PESERA coarse scale erosion model for Europe. Model rationale and implementation. *European Journal of Soil Science* 59, 1293–1306.

Kirkby, M., Jones, R., Irvine, B., Gobin, A., Govers, G., Cerdan, O., Van Rompaey, A., Le Bissonnais, Y., Daroussin, J., King, D., Montanarella, L., Grimm, M., Vieillefont, V., Puigdefabregas, J., Boer, M., Kosmas, C., Yassoglou, N., Tsara, M., Mantel, S., Van Lynden, G. & Hunting, J. 2004. European Soil Bureau Research Report No.16, EUR 21176, 18pp. and 1 map in ISO B1 format. Office for Official Publications of the European Communities, Luxembourg.

Met Office. 2018. HadUK-Grid gridded and regional average climate observations for the UK.

Morton, R.D., Marston, C.G., O'Neil, A.W. & Rowland, C.S. 2020. Land Cover Map 2019 (25m rasterised land parcels, GB).

Orgiazzi, A., Dunbar, M., Panagos, P., de Groot, G. & Lemanceau, P. 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biology and Biochemistry* 80, 244–250.

Orgiazzi, A., Panagos, P., Yigini, Y., Dunbar, M., Gardi, C., Montanarella, L. & Ballabio, C. 2016. A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. *Science of the Total Environment* 545–546, 11–20.

Paustian, K., Larson, E., Kent, J., Marx, E. & Swan, A. 2019. Soil C Sequestration as a Biological Negative Emission Strategy. *Frontiers in Climate*. https://doi.org/10.3389/fclim.2019.00008

Prout, J.M., Shepherd, K.D., McGrath, S.P., Kirk, G.J.D. & Haefele S.M. 2021. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. *Eur J Soil Sci.*,72, 2493-2503. https://doi.org/10.1111/ejss.13012

Soils Data © Cranfield University (NSRI) and for the Controller of HMSO [2022], n.d. LandIS NATMAP Topsoil Texture.

UN Committee on World Food Security. 2012. Global Strategic Framework for Food Security and Nutrition.

Appendix 1: Conceptual Ecological Models and Directed Acyclic Graphs (DAGs)

We provide here the four conceptual ecological models produced for the proof of concept (Figures 5 to 8) as well as the DAGs for the two Bayesian models (Figures 9 to 10).

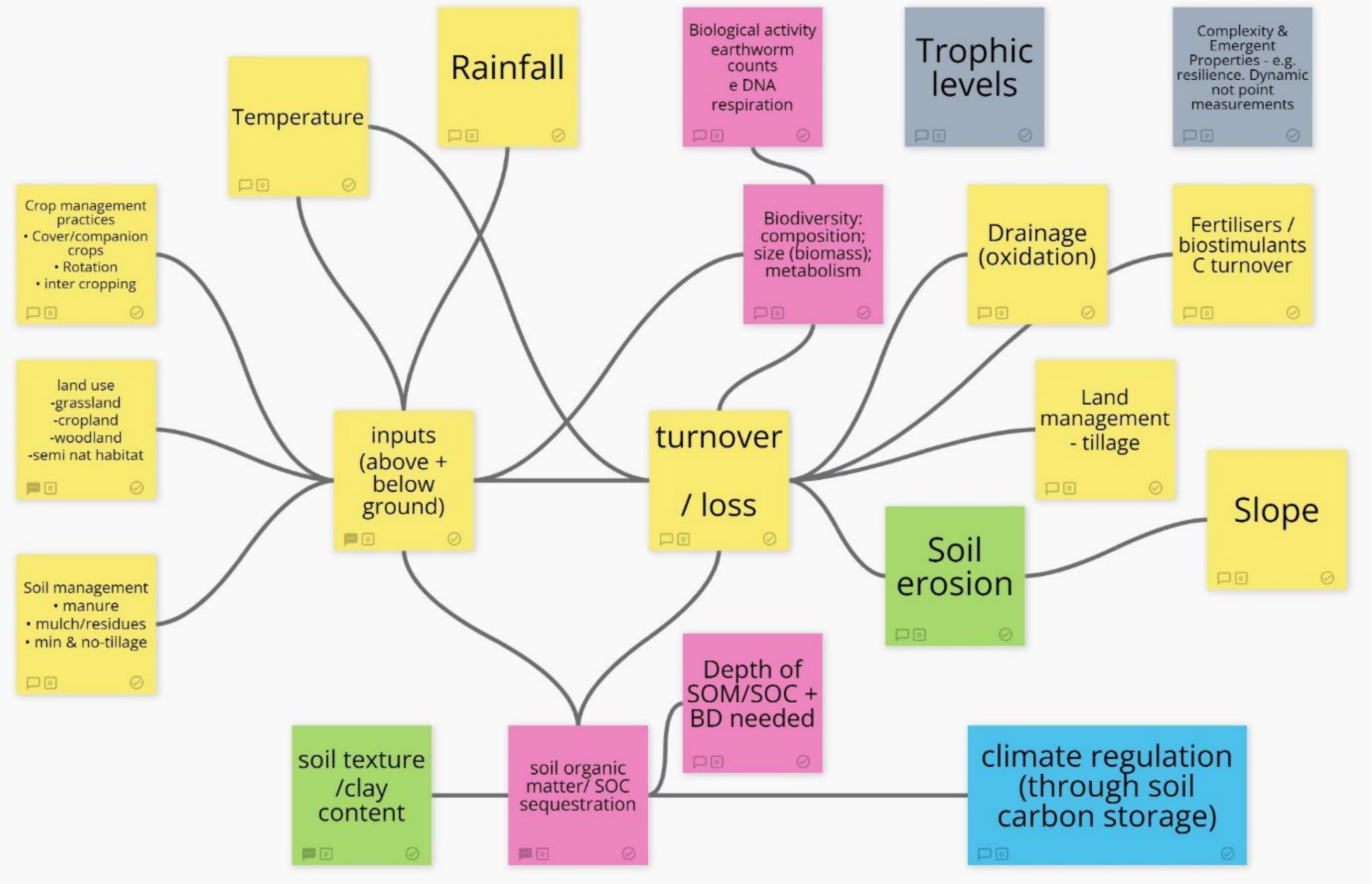


Figure 5. A conceptual ecological model showing soils' contribution to the delivery of climate regulation as an ecosystem service through soil carbon storage. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (Table1) (pink); Ecosystem Service outcome indicator (blue); other consideration (grey).

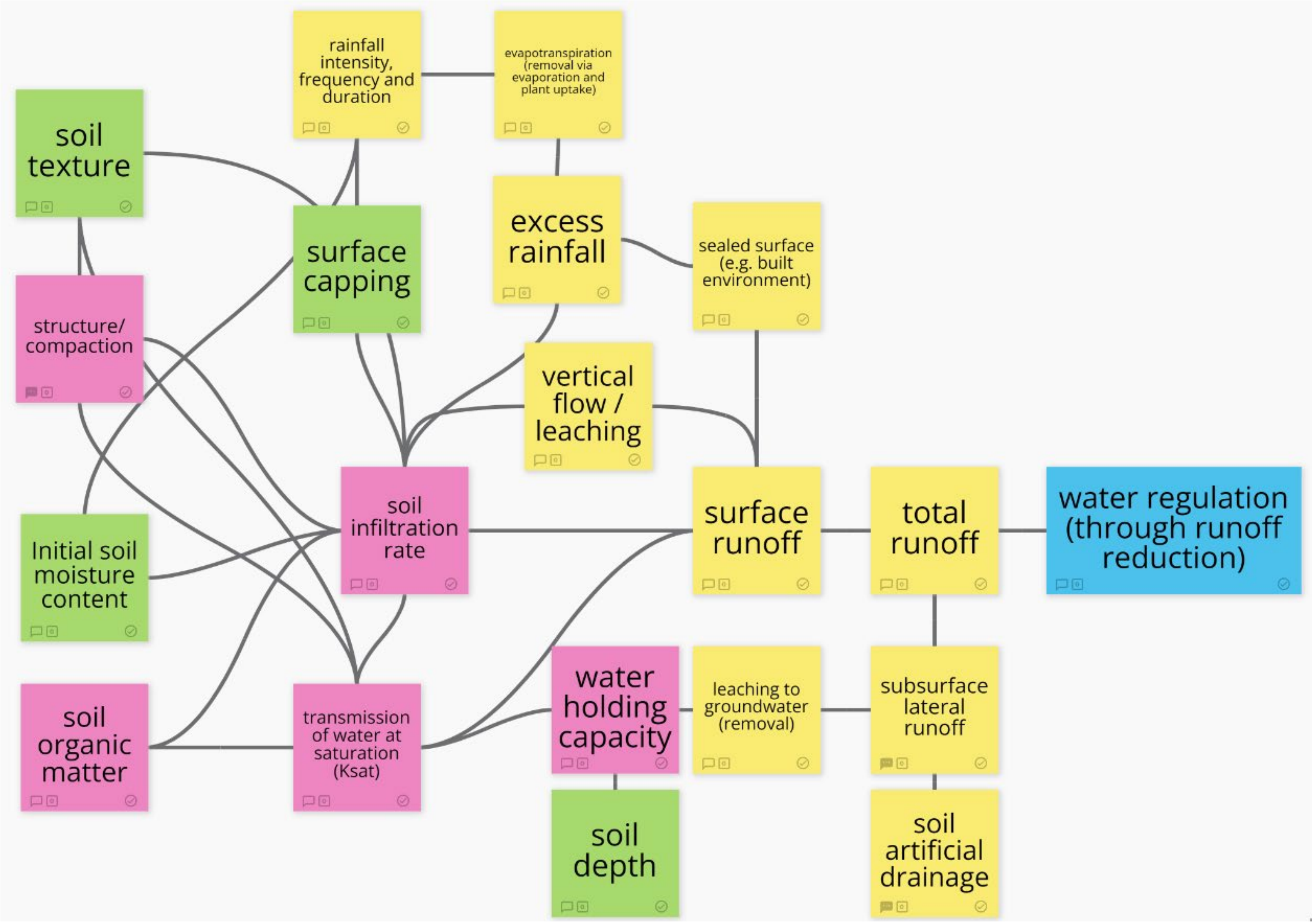


Figure 6. A conceptual ecological model showing soils’ contribution to the delivery of flood mitigation as an ecosystem service through soil water regulation. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (Table1) (pink); Ecosystem Service outcome indicator (blue); other consideration (grey).

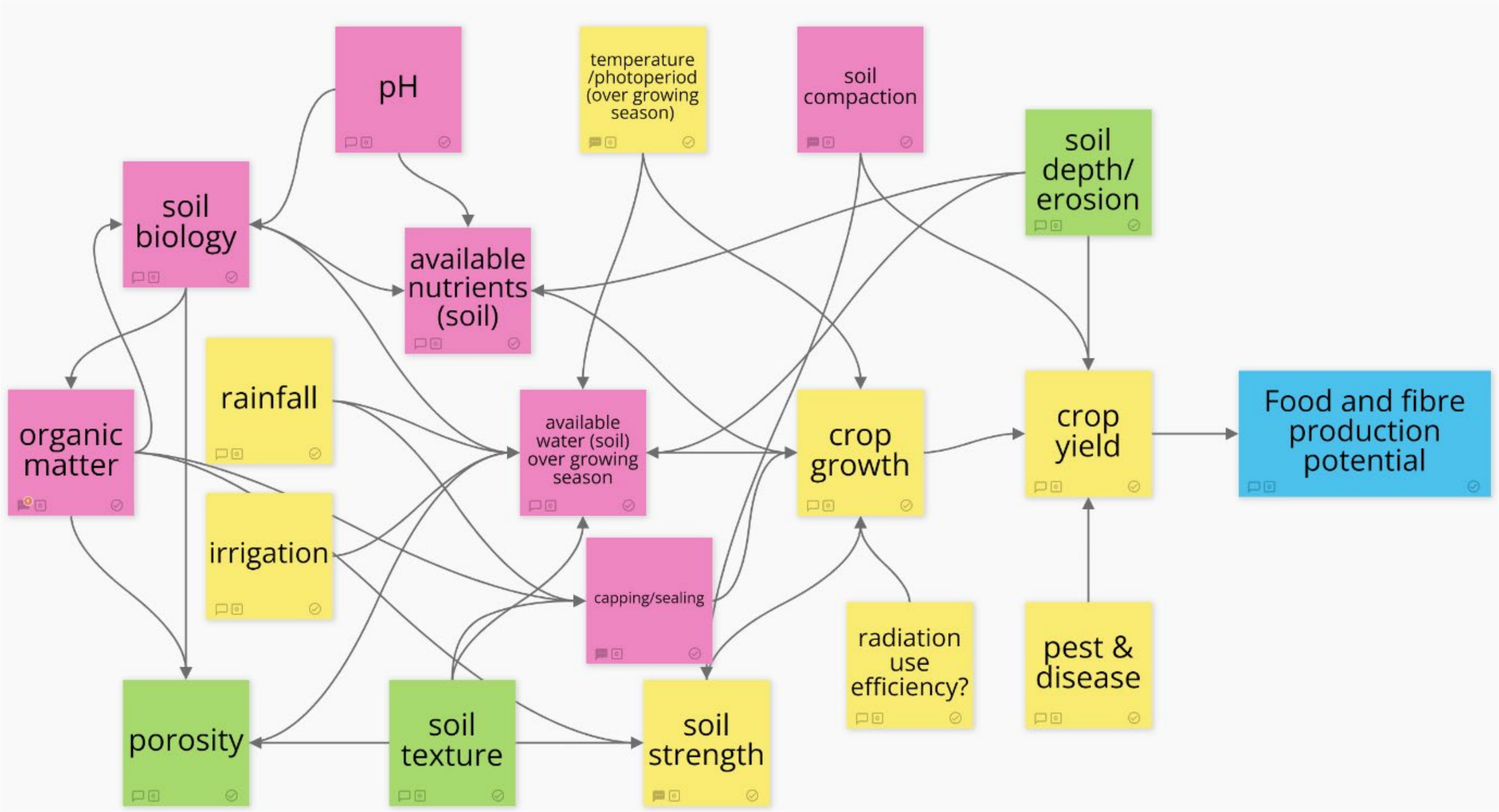


Figure 7. Conceptual ecological model for food and fibre production potential through soils' contribution to land capability for agriculture. Yellow: environmental variable; Green: soil property; Pink: soil property identified from the logical sieve (Annex 1, Table 10); Blue: Ecosystem Service indicator outcome. A conceptual ecological model showing soils' contribution to the delivery of agriculture as an ecosystem service through food and fibre production. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (See Annex 1) (pink); Ecosystem Service outcome indicator (blue); other consideration (grey).

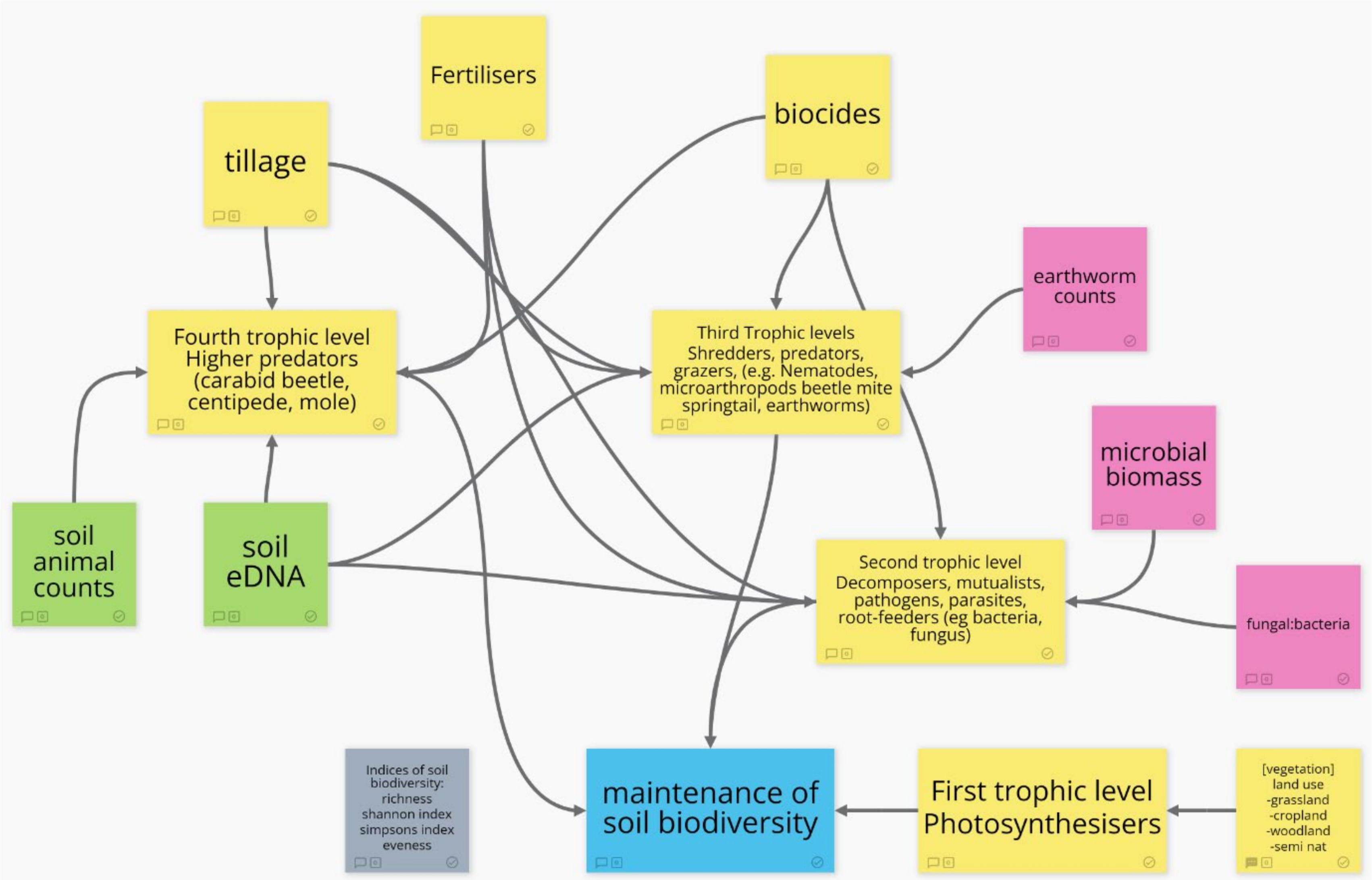


Figure 8. A conceptual ecological model showing soils' contribution to the delivery of maintenance of biodiversity as an ecosystem service through soil biodiversity. The elements within the model are categorised as: environmental variable (yellow); soil property (green); soil property identified from the logical sieve (See Annex 1) (pink); Ecosystem Service outcome indicator (blue); other consideration (grey).

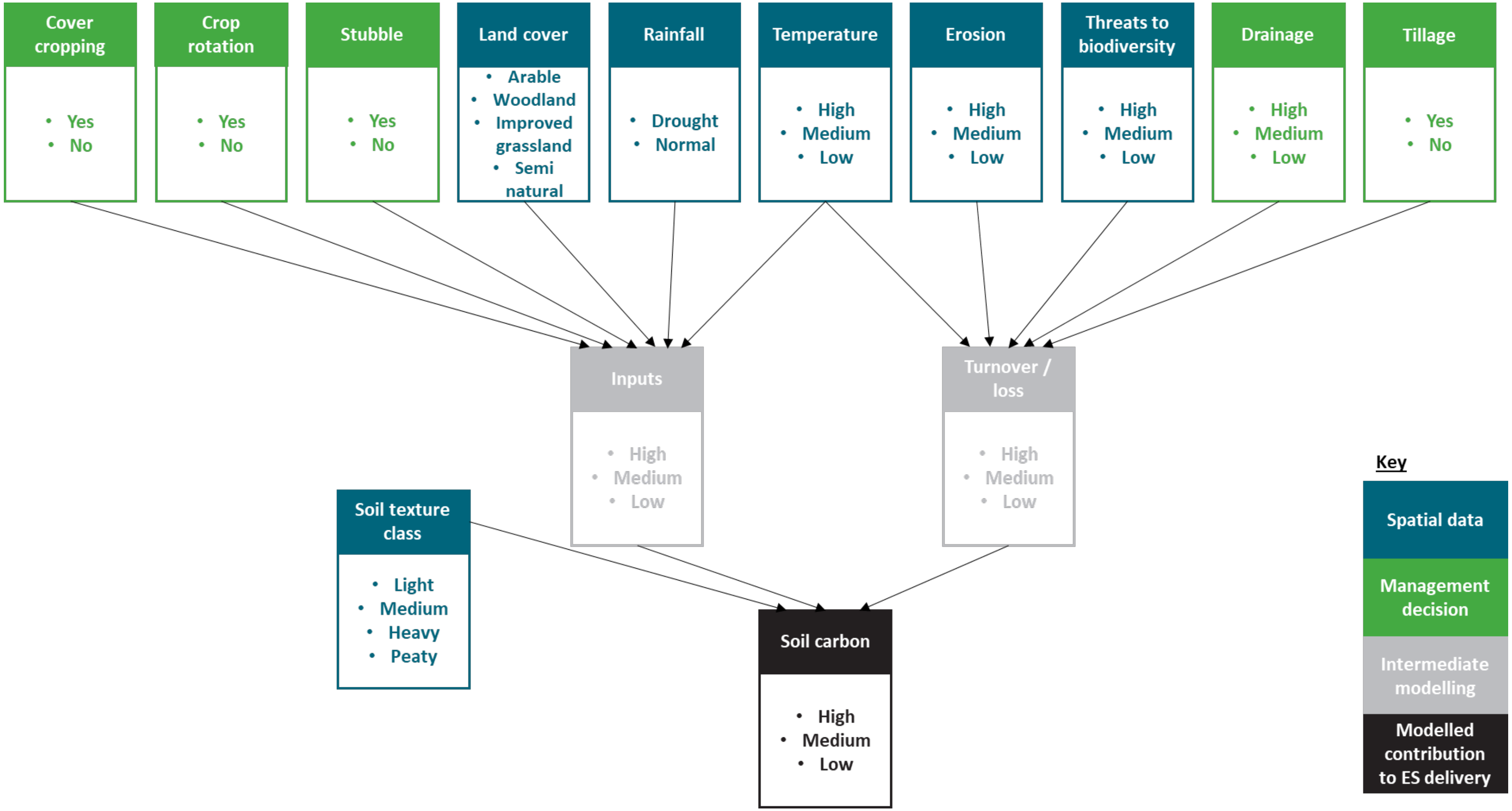


Figure 9. A directed acyclic graph (DAG) showing the nodes and node states used in the BBN modelling for the climate regulation through soil carbon storage model. Inputs for the nodes in blue are based on spatial data, whilst nodes coloured green will rely on users to input a management decision via the dashboard, in order to model soils' contribution to the final ecosystem service delivery (shown in black).

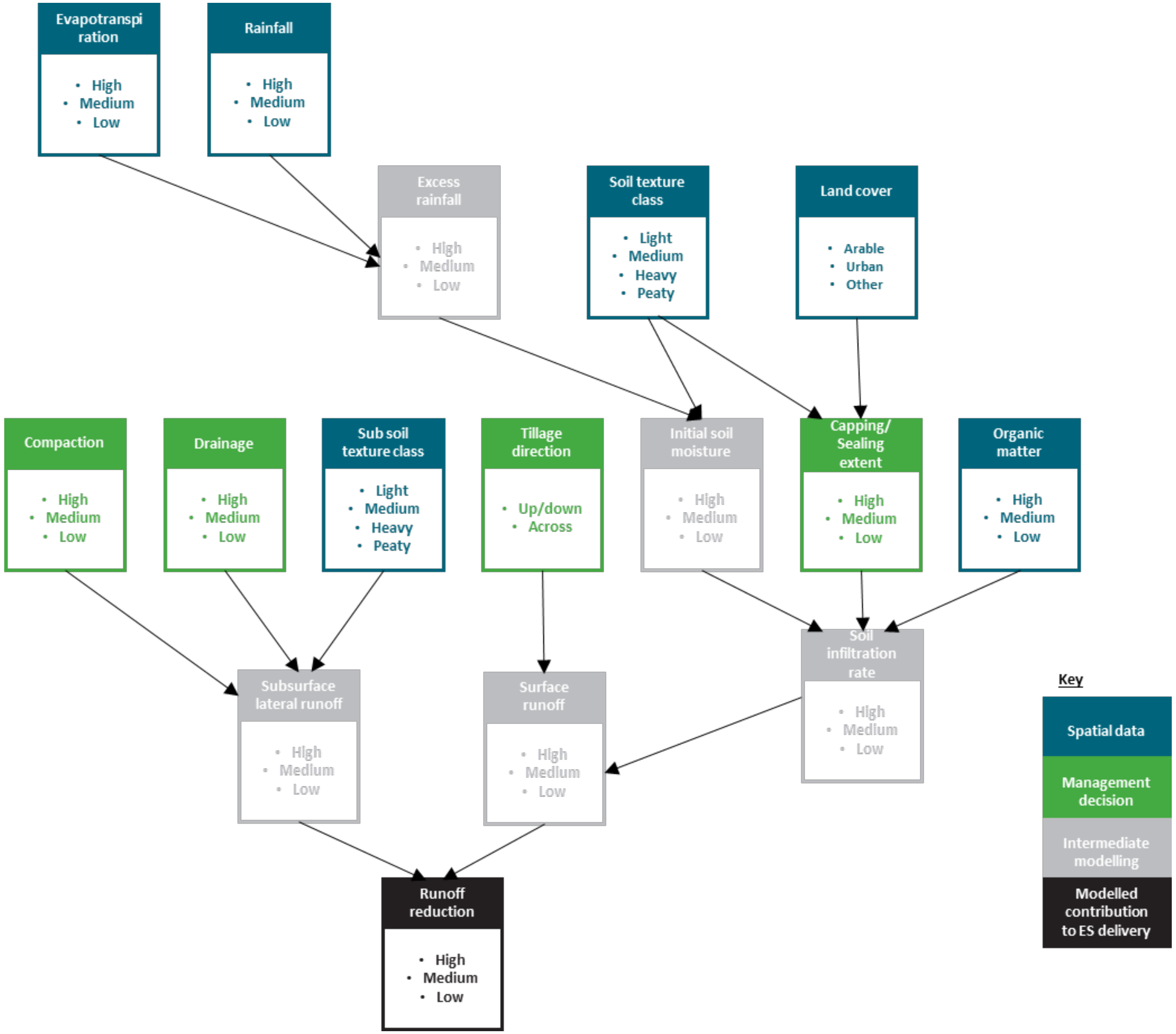


Figure 10. A directed acyclic graph (DAG) showing the nodes and node states used in the BBN modelling for the water regulation through runoff reduction model. Inputs for the nodes in blue are based on spatial data, whilst nodes coloured green will rely on users to input a management decision via the dashboard, in order to model soils' contribution to the final ecosystem service delivery (shown in black).