Analysis of Multi-Interpretable Ecological Monitoring Information

Frances Brazier, Joeri Engelfriet, Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam
Email: `{frances, joeri, treur}@cs.vu.nl`
URL: `http://www.cs.vu.nl/~{frances,joeri,treur}`

Abstract

In this paper logical techniques developed to formalise the analysis of multi-interpretable information, in particular belief set operators and selection operators, are applied to an ecological domain. A knowledge-based decision support system is described that determines the abiotic (chemical and physical) characteristics of a site on the basis of samples of plant species that are observed. The logical foundation of this system is described in terms of a belief set operator and a selection operator.

1. Introduction

In most real-life situations humans receive information that can be interpreted in many different ways. The context often determines the view with which this information is interpreted, but also other factors may be of influence. One domain in which multi-interpretable observations can be analysed using a technique based on the distinction of different views, is the domain of ecology.

Plants only grow in areas where conditions are appropriate. Knowledge of which set of factors is necessary for species to germinate and complete their life-cycle, has been acquired by experts over a large number of years. This knowledge of environmental preferences of plant species makes it possible to derive information about a terrain's abiotic (physical and chemical) characteristics on the basis of the plant species found. More specifically, experts are able to derive the abiotic conditions of the site studied in terms of acidity, nutrient value and moisture from the abiotic preferences of the species comprising the vegetation.

If knowledge on abiotic preferences of plant species is available, nature conservationists can use their knowledge of the plant species found in a specific terrain to determine the abiotic conditions. Often, however, nature conservationists responsible for terrains do not possess this detailed knowledge. An Environmental Knowledge-based System, EKS, has been designed to support them in this decision making process. Once the abiotic conditions of a terrain have been determined, nature conservationists can then use this knowledge to manage the terrain; e.g., new measures can be derived to improve the quality of the site.

The specific domain of application in the current implementation is grasslands. The knowledge-based system, the development of which was funded by the organisations International Plant Technology Services (IPTS) and the State Forestry Department of the Dutch Ministry of Agriculture (Staatsbosbeheer), is based on knowledge acquired from experts in the fields of Plant Ecology, Eco-hydrology, and Soil Sciences. Acquiring consensus between experts on the meaning of individual plant species with respect to their specific abiotic conditions is one of the main aims of this project. The observations made in the field, a sample, can often be interpreted in different ways. To model this expert reasoning task, an approach based on belief set operators (introduced in [8]) is applied.

In this paper, the application domain is introduced in Section 2. In Section 3 the knowledge-based system EKS is described. Section 4 introduces belief set operators and shows how the expert reasoning task can be formalised using these operators. In Section 5 the correspondence between the formalisation and the system design is shown. Finally, in Section 6 the reported results are discussed.

2. Domain of Application

Experts identify the current abiotic conditions of a terrain on the basis of plant species they encounter. The process of identification of abiotic conditions was analysed in cooperation with experts, resulting in the distinction of three tasks: (1) grouping the plant species that "belong together", (2) selecting the set of plant species experts consider most "defining", and (3) identifying the related abiotic conditions. These conditions are expressed as values for each of the abiotic factors: *acidity* (basic, neutral, slightly acid, fairly acid, acid), *nutrient value* (nutrient poor, fairly nutrient rich, nutrient rich, very nutrient rich) and *moisture* (very dry, fairly dry, fairly moist, very moist, fairly wet, very wet).

In a sample of plant species taken from an abiotic homogeneous site, a common set of abiotic conditions can be found that are shared by the plant species. A technique to determine the abiotic conditions in this case is described in Section 2.1. In practice, however, the samples often include groups of plant species that, according to the knowledge available, could not possibly grow under the same abiotic conditions. One cause could be that the knowledge about the abiotic conditions in which species can live is incomplete. Another cause could be that the sample has been taken from a heterogeneous site: a site where the abiotic conditions vary over space and time (for instance, on a site in transition between dry and wet soil). An expert needs to analyse and interpret the available information and can, for example, determine that a sub-set of the sample is most dominant. A method to determine which compatible groups of plant species can be distinguished within a sample is described in Section 2.2.

2.1. Homogeneous Sample: Greatest Common Denominator

In a sample of plant species taken from a homogeneous site, at least one set of abiotic conditions can be found that is shared by all species on the site. An example of a sample of species that can all grow in a homogeneous site is used to illustrate a technique to find this set of common abiotic conditions. Examination of the plant species, depicted in Table 1, shows all possible values for each of the three abiotic factors, for each of the plant species. For example, the abiotic requirements of *Caltha palustris* L., are:

- very moist or fairly wet,
- basic, neutral or slightly acid,
- nutrient poor, fairly nutrient rich or nutrient rich terrain.

For the species *Poa trivialis* L. a terrain needs to be

- fairly moist, very moist or fairly wet,
- basic or neutral,
- nutrient rich or very nutrient rich.

If both species occur in a terrain, this implies that the terrain can only be:

- very moist or fairly wet,
- basic or neutral,
- nutrient rich.

	Moisture						Acidity					Nutrient Value			
Species	vd	fd	fm	vm	fw	vw	bas	neu	sac	fac	ac	np	fnr	nr	vnr
Angelica sylvestris				x	x		x	x					x	x	
Caltha palustris ssp palustris				x	x		x	x	x			x	x	x	
Carex acutiformis				x	x		x	x					x	x	
Carex acuta				x	x	x	x	x	x				x	x	x
Deschampsia caespitosa			x	x	x		x	x	x				x	x	x
Epilobium parviflorum			x	x			x	x	x				x	x	
Equisetum palustre			x	x	x	x	x	x	x			x	x	x	
Galium palustre				x	x		x	x	x			x	x	x	x
Glyceria fluitans				x	x	x	x	x	x	x			x	x	x
Juncus articulatus				x	x		x	x	x			x	x	x	x
Lathyrus pratensis			x	x			x	x	x				x	x	
Myosotis palustris				x	x		x	x	x				x	x	x
Phalaris arundinacea			x	x	x	x	x	x						x	x
Phleum pratense ssp pratense			x	x			x	x						x	x
Poa trivialis			x	x	x		x	x						x	x
Scirpus sylvaticus				x	x	x	x	x	x				x	x	

Moisture (vd: very dry, fd: fairly dry, fm: fairly moist, vm: very moist, fw: fairly wet, vw: very wet)
Acidity (bas: basis, neu: neutral, sac: slightly acid, fac: fairly acid, ac: acid)
Nutrient value (np: nutrient poor, fnr: fairly nutrient rich, nr: nutrient rich, vnr: very nutrient rich)

Table 1. A homogeneous sample.

Note that not only can the occurrence of a *single* species restrict the possible abiotic conditions of the terrain, but the occurrence of species *in combination* can restrict the possible abiotic conditions even further.

Analysis of the abiotic conditions for all plant species presented in Table 1 shows that only a restricted number of possibilities (but more than one) for the abiotic conditions can be found in which all of these plant species can abide. This *greatest common denominator* for the given plant species is defined by the following set of abiotic conditions:

- very moist
- basic or neutral
- nutrient rich

The combination of these plant species indicates that a terrain on which these plant species are found has to fulfill these conditions.

2.2. Inhomogeneous Sample: Maximal Indicative Subsets

In a sample taken from an inhomogeneous site, the sample does not have a common denominator of abiotic conditions. A real example sample is shown in Table 2, together with the possible values for the three abiotic factors for each plant species. Focusing on the acidity of a terrain shows that the plant species *Angelica sylvestris* L., for example, only grows on a basic or neutral terrain, whereas the species *Carex panicea* L., also found in the same sample, only grows on a slightly or fairly acid terrain. These two species, however, are in the same sample. One common set of possible values of the abiotic factors for all plant species can not be derived.

Further analysis of the abiotic factors of the plant species in the sample is required. Groups of plant species for which a set of shared abiotic conditions can be found are grouped together. These groups of plant species are homogeneous groups of plants as defined above in Section 2.1. The largest possible homogeneous groups of plant species are called *maximal indicative subsets*.

These subsets are maximal with respect to compatibility of the plant species in the subset. In other words, all plant species in the sample that are compatible with the group of plant species in a maximal indicative subset (those plant species that can grow on a site with the same abiotic conditions), are in the subset. As shown in Table 3, in the example sample two maximal indicative sets of plant species can be distinguished. The *first maximal indicative subset* contains all plant species that can grow in

- very moist
- basic or neutral
- nutrient rich

environments. The *second maximal indicative subset* contains all plant species that can grow in

- very moist
- slightly acid
- fairly nutrient rich

environments.

	Moisture						Acidity					Nutrient Value			
Species	vd	fd	fm	vm	fw	vw	bas	neu	sac	fac	ac	np	fnr	nr	vnr
Angelica sylvestris				x	x		x	x					x	x	
Anthoxanthum odoratum		x	x	x					x	x		x	x		
Caltha palustris ssp palustris				x	x		x	x	x			x	x	x	
Carex acutiformis				x	x		x	x					x	x	
Carex acuta				x	x	x	x	x	x				x	x	x
Carex nigra			x	x	x				x	x	x	x	x		
Carex panicea			x	x	x				x	x		x	x		
Carex riparia				x	x	x	x	x						x	x
Cirsium oleraceum				x	x		x	x					x	x	
Cirsium palustre				x			x	x	x			x	x	x	
Crepis paludosa			x	x	x		x	x	x				x	x	
Deschampsia caespitosa			x	x	x		x	x	x				x	x	x
Epilobium palustre			x	x	x				x			x	x		
Epilobium parviflorum			x	x			x	x	x				x	x	
Equisetum palustre			x	x	x	x	x	x	x			x	x	x	
Filipendula ulmaria				x			x	x	x			x	x	x	
Galium palustre				x	x		x	x	x			x	x	x	x
Glyceria fluitans				x	x	x	x	x	x	x			x	x	x
Juncus articulatus				x	x		x	x	x			x	x	x	x
Juncus conglomeratus		x	x	x					x	x		x	x		
Lathyrus pratensis			x	x			x	x	x				x	x	
Lotus uliginosus			x	x	x		x	x	x			x	x	x	
Lychnis flos cuculi				x	x		x	x	x				x	x	
Lysimachia vulgaris			x	x	x		x	x	x			x	x	x	
Myosotis palustris				x	x		x	x	x				x	x	x
Phalaris arundinacea			x	x	x	x	x	x						x	x
Phleum pratense ssp pratense			x	x			x	x						x	x
Poa trivialis			x	x	x		x	x						x	x
Scirpus sylvaticus				x	x	x	x	x	x				x	x	

Moisture (vd: very dry, fd: fairly dry, fm: fairly moist, vm: very moist, fw: fairly wet, vw: very wet),
Acidity (bas: basis, neu: neutral, sac: slightly acid, fac: fairly acid, ac: acid),
Nutrient value (np: nutrient poor, fnr: fairly nutrient rich, nr: nutrient rich, vnr: very nutrient rich)

Table 2. An inhomogeneous sample.

Note that the two maximal indicative subsets share a number of plants (the intersection of the two subsets). These plants have a relatively broad spectrum of environmental preferences.

	Moisture						Acidity					Nutrient Value			
Species	vd	fd	fm	vm	fw	vw	bas	neu	sac	fac	ac	np	fnr	nr	vnr
Angelica sylvestris				x	x		x	x					x	x	
Carex acutiformis				x	x		x	x					x	x	
Carex riparia				x	x	x	x	x						x	x
Cirsium oleraceum				x	x		x	x					x	x	
Phalaris arundinacea			x	x	x	x	x	x						x	x
Phleum pratense ssp pratense			x	x			x	x						x	x
Poa trivialis			x	x	x		x	x						x	x
Caltha palustris ssp palustris				x	x		x	x	x			x	x	x	
Carex acuta				x	x	x	x	x	x				x	x	x
Cirsium palustre				x			x	x	x			x	x	x	
Crepis paludosa			x	x	x		x	x	x				x	x	
Deschampsia caespitosa			x	x	x		x	x	x				x	x	x
Epilobium parviflorum			x	x			x	x	x				x	x	
Equisetum palustre			x	x	x	x	x	x	x			x	x	x	
Filipendula ulmaria				x			x	x	x			x	x	x	
Galium palustre				x	x		x	x	x			x	x	x	x
Glyceria fluitans				x	x	x	x	x	x	x			x	x	x
Juncus articulatus				x	x		x	x	x			x	x	x	x
Lathyrus pratensis			x	x			x	x	x				x	x	
Lotus uliginosus			x	x	x		x	x	x			x	x	x	
Lychnis flos cuculi				x	x		x	x	x				x	x	
Lysimachia vulgaris			x	x	x		x	x	x			x	x	x	
Myosotis palustris				x	x		x	x	x				x	x	x
Scirpus sylvaticus				x	x	x	x	x	x				x	x	
Anthoxanthum odoratum		x	x	x					x	x		x	x		
Carex nigra			x	x	x				x	x	x	x	x		
Carex panicea			x	x	x				x	x		x	x		
Epilobium palustre			x	x	x				x			x	x		
Juncus conglomeratus		x	x	x					x	x		x	x		

Moisture (vd: very dry, fd: fairly dry, fm: fairly moist, vm: very moist, fw: fairly wet, vw: very wet)
Acidity (bas: basis, neu: neutral, sac: slightly acid, fac: fairly acid, ac: acid)
Nutrient value (np: nutrient poor, fnr: fairly nutrient rich, nr: nutrient rich, vnr: very nutrient rich)

Table 3. Maximal indicative subsets within an inhomogeneous sample.

Note also that the conditions for the plant species that these two groups do not have in common are mutually exclusive with respect to acidity and (partially) nutrient value.

To decide which maximal indicative set is the most appropriate for a given site, additional knowledge is required. For example, in this case, the expert knows that the sample has been taken from a site that has a particular type of stratification (so-called rainwater lenses): two different layers of soil can be found on the same site. This explains the presence of the two abiotic indicative sets of plant species. Additional detailed knowledge on abiotic conditions for plant species can also be taken into account; e.g., knowledge on the optimal conditions for specific plant species.

3. The Decision Support System EKS

The above described expert knowledge on the determination of abiotic conditions on the basis of a terrain's vegetation, has been used to design a knowledge-based system to support ecologists in the upkeep of nature reserves. This knowledge-based system, the EKS system, has been modelled, specified and implemented within the compositional development method DESIRE (see e.g., [1], [4]).

3.1. The Compositional Development Method DESIRE

DESIRE is a compositional development method for the design and implementation of knowledge-based and multi-agent systems. A knowledge engineer is supported during all (iterative) phases of design: from initial conceptualisation to implementation, by the DESIRE development method supported by the dedicated software environment.

The development method focuses on the identification and specification of the following types of knowledge, the types of knowledge used to define a model:

(1) *process composition*

- identification of the *processes* or *tasks* involved at different levels of *process abstraction;*
- knowledge of task and role *delegation* between systems (human and/or automated): *task and role delegation;*
- knowledge of the information exchanged between processes: *information exchange;*
- knowledge of when and how processes are activated (in parallel or sequential, under which conditions): *task control*;

(2) *knowledge composition*

- identification of the types of *information* and *knowledge* used at different levels of knowledge abstraction;

- specification of the knowledge structures and the way in which they are composed;

(3) *relations between process composition and knowledge composition*

- Knowledge on which knowledge structures are used in which processes.

Initial knowledge analysis focuses on the acquisition of a shared task model: an intermediary agreed model shared by both the expert and the knowledge engineer, in which these types of knowledge are made explicit (see [2], [3]). This knowledge is first identified at an abstract level, and refined during the further design process.

Tasks distinguished during conceptual design are modelled as components. Components can be primitive or complex: a component may encompass a number of other (either primitive or complex) components, or it may not. If not, the component is either a reasoning component with a knowledge base or a component with a so-called *alternative specification* (meaning that only its input and output are explicitly specified in the DESIRE modelling language, e.g., databases, OR-algorithms, neural networks, etc.). A knowledge-based system's behaviour as a whole is defined by the interaction between components, and between the system and its users. The DESIRE software environment consists of:

- a graphical editor to support conceptual and detailed design;
- an implementation generator that translates DESIRE specifications into executable
  code;
- an execution environment in which the translated code can be executed.

3.2. Design of EKS

In Section 2, three tasks are distinguished: (1) grouping of plant species that "belong together", (2) selecting the set of plant species experts consider most "defining", and (3) identifying the related abiotic conditions.

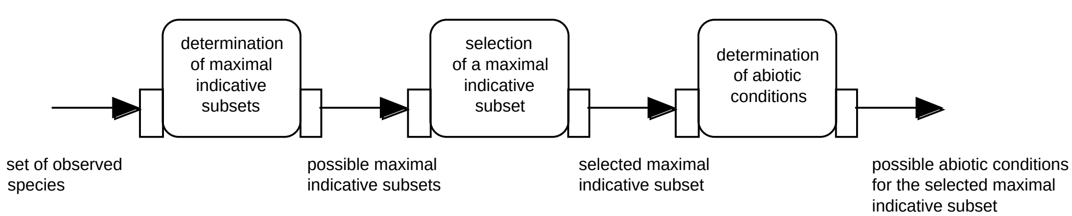

Figure 1. The global design of EKS.

Task Composition

These three tasks are modelled by three components as shown in Figure 1. The first task, the *determination of maximal indicative subsets*, entails analysis of the plant species in the sample and the corresponding abiotic conditions to determine maximal indicative subsets of plant species. The choice of the most defining subset is performed by the component *selection of a maximal indicative subset*. The third task, *determination of abiotic conditions*, is relatively simple, and includes the presentation of the abiotic conditions of a maximal indicative subset.

Information Exchange

The initial information needed by the system to determine the abiotic conditions of a terrain is a list of observed plant species. This is the input for the first component. The maximal indicative sets of plant species derived in the first task are the input for the second task. The result of the selection process (the second task), one of the maximal indicative subsets, in turn, is input for the third task (determination of abiotic conditions). The final output consists of the possible abiotic conditions for the selected maximal indicative subset.

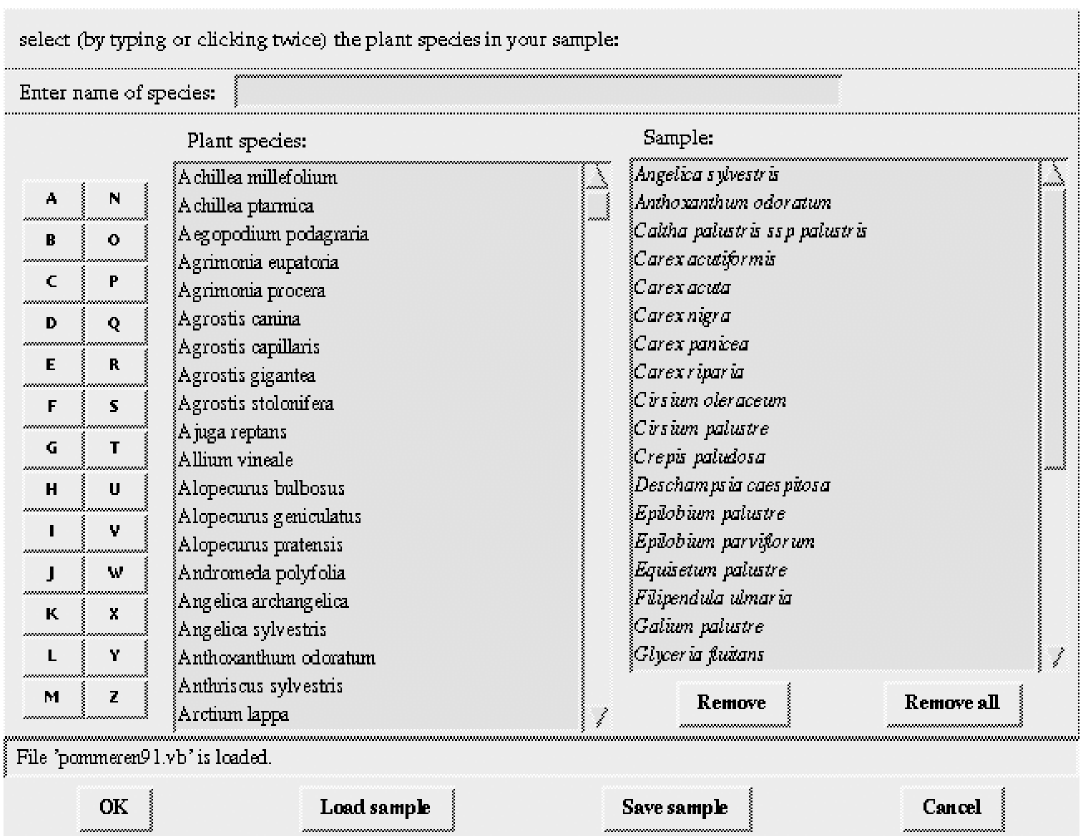

Figure 2. Input window of EKS.

Task Activation

Task activation is straightforward. Completion of the first task results in activation of the second. Completion of the second task results in activation of the third. Completion of the third task results in completion of the entire task.

Task Delegation

The first task and the third task are performed by the system. The second task is performed by the user.

Knowledge Structures

The knowledge includes knowledge of plant species and the abiotic conditions in which they can abide, part of which is presented above in table format (see Tables 1 and 2). Each plant species has related values for each of the three abiotic factors. For reasons of efficiency, the first component is specified by an alternative specification.

The system EKS has been developed using the DESIRE method and software environment. In addition, a graphical user interface has been designed specifically for EKS.

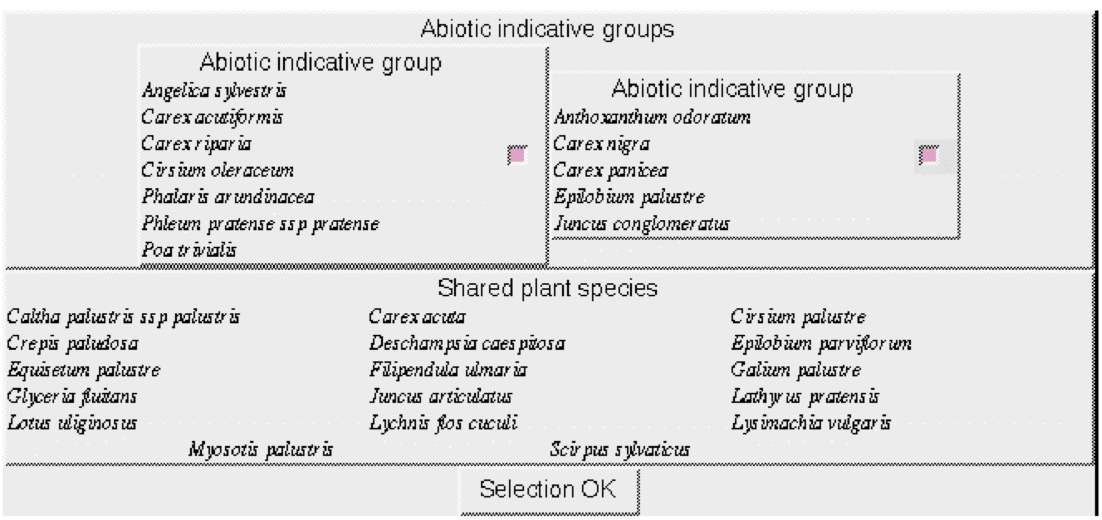

Figure 3. Presentation of the maximal indicative subsets.

3.3. User-System Interaction

Initially a user is presented with a screen with which he/she can enter the plant species found on a terrain, as shown in Figure 2. The system analyses this information, resulting in the two maximal indicative subsets of plant species. This information is presented to the user as shown in Figure 3. The overlap between the two maximal indicative subsets of plant species is presented on the screen as the list of *shared plant species*. The remaining plants are listed separately for each of the maximal indicative subsets as *abiotic indicative groups*. The user chooses which maximal subset is most appropriate.

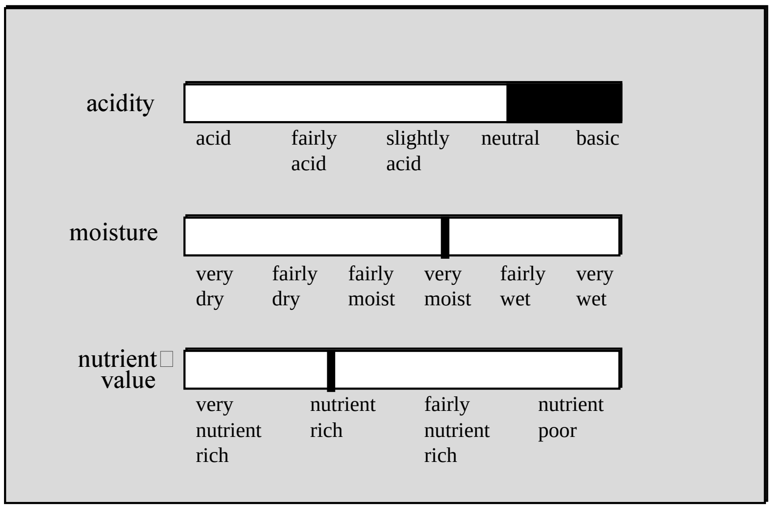

Figure 4 Abiotic conditions for the first maximal indicative set.

The final output of the system is a graphical presentation of the abiotic conditions for the terrain in question. In Figure 4 and 5 the two possible outputs are shown: for the first maximal indicative subset and for the second maximal indicative subset, respectively. Figure 4 shows that the first subset indicates that the terrain is nutrient rich and very moist, and the acidity is somewhere in the interval from basic to neutral. The second subset indicates that the terrain is slightly acid, very moist and fairly nutrient rich.

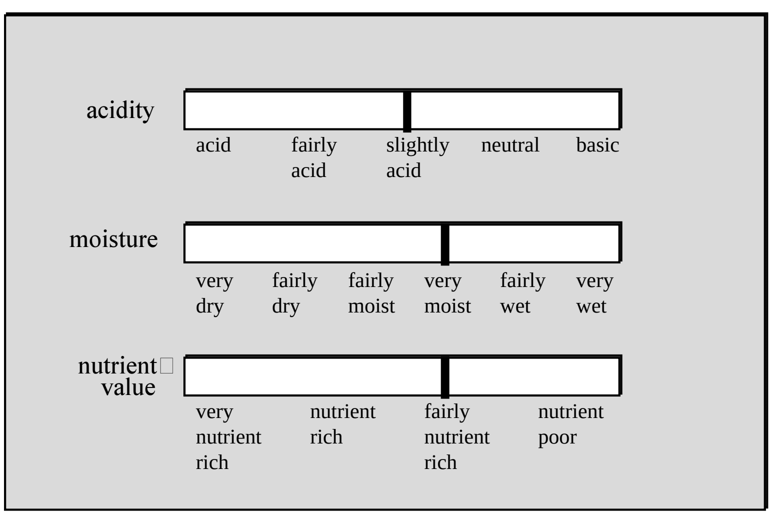

Figure 5 Abiotic conditions for the second maximal indicative set.

4. Formalisation by a Belief Set Operator and a Selection Operator

In Section 2, three tasks are distinguished: determination of groups of plants that belong together, selection of one of these groups, and then identification of its related abiotic conditions. The second task is performed by the user, and the last task is rather straightforward. As mentioned before, the first task is the most complex; it is formalised in this section. From an abstract point of view, this task can be seen as follows. The given observations together provide a partial view of the world (these observations are the plant species on the terrain). It is partial in the sense that it is not yet known which (group of) plants are most defining for the terrain. The first task consists of finding possible extensions of this partial view, by adding additional beliefs about which plants are defining (and, as said before, this can, in general, be done in multiple ways). Such forms of reasoning, in which a partial view on the world is extended to multiple (more informed) views, after which a selection from these can be made, have been formalised using belief set operators and selection operators in [8]. In this section these formalisations are related to the application at hand. To this purpose, a brief overview of the main ideas and concepts of [8] is presented.

A propositional language, $\mathbf{L}$, is assumed, together with its corresponding set of models, $\mathbf{Mod}$, and the standard (semantic) consequence relation $\models\ \subseteq \mathbf{Mod} \times \mathbf{L}$. A set of formulas which is closed under propositional consequence is called a *belief set*. A belief set can be seen as a possible set of beliefs of an agent with perfect (propositional) reasoning capabilities.

Definition 4.1 (Belief set operator)

a) A *belief set operator* $\mathbf{B}$ is a function $\mathbf{B} : \wp(\mathbf{L}) \rightarrow \wp(\wp(\mathbf{L}))$ that assigns a set of belief sets to each set of initial facts.

b) A belief set operator $\mathbf{B}$ satisfies *inclusion* if for all $\mathbf{X} \subseteq \mathbf{L}$ and all $\mathbf{T} \in \mathbf{B(X)}$ it holds $\mathbf{X} \subseteq \mathbf{T}$. A belief set operator $\mathbf{B}$ satisfies *non-inclusiveness* if for all $\mathbf{X} \subseteq \mathbf{L}$ and all $\mathbf{S}, \mathbf{T} \in \mathbf{B(X)}$, if $\mathbf{S} \subseteq \mathbf{T}$ then $\mathbf{S} = \mathbf{T}$.

The *kernel* $\mathbf{K_B} : \wp(\mathbf{L}) \rightarrow \wp(\mathbf{L})$ of $\mathbf{B}$ is defined by $\mathbf{K_B(X)} = \cap\, \mathbf{B(X)}$.

The first condition expresses conservativity: it means that a possible view on the world at least satisfies the given facts; the belief set operator defines a method of extending partial information (instead of, for instance, revising it). The condition of non-inclusiveness guarantees a relative maximality of the possible views. The kernel of a belief set operator yields the most certain conclusions given a set of initial facts, namely those which are in every possible view of the world. To give an example of a belief set operator, consider a set of default rules (the reader is referred to the next section for a definition of default logic). A set of initial facts, together with the default rules, gives rise to a number of extensions (which can be considered belief sets). An operator that assigns the corresponding set of extensions to each set of initial facts is a belief set operator. The kernel of this operator yields the sceptical (see e.g., [11]) conclusions.

Often, as is the case in the application, after a number of belief sets have been generated, the process will focus on (or make a commitment to) one (or possibly more) of the belief sets, because it seems the most promising, or interesting, possible view on the world. This selection process can be formalized by selection operators.

Definition 4.2 (Selection operator and selective inference operation)

a) A *selection operator* **s** is a function $\mathbf{s} : \wp(\wp(\mathbf{L})) \rightarrow \wp(\wp(\mathbf{L}))$ that assigns to each set of belief sets a subset (for all $\mathbf{A} \subseteq \wp(\mathbf{L})$ it holds $\mathbf{s(A)} \subseteq \mathbf{A}$) such that whenever $\mathbf{A} \subseteq \wp(\mathbf{L})$ is non-empty, $\mathbf{s(A)}$ is non-empty. A selection operator **s** is *single-valued* if for all non-empty **A** the set **s(A)** contains exactly one element.

b) A *selective inference operation* for the belief set operator **B** is a function $\mathbf{C} : \wp(\mathbf{L}) \rightarrow \wp(\mathbf{L})$ that assigns a belief set to each set of facts, such that for all $\mathbf{X} \subseteq \mathbf{L}$ it holds $\mathbf{C(X)} \in \mathbf{B(X)}$

A formalisation of (the first task of) the application described in this paper can be made using the notions defined above. The language **L** is the propositional language of which the atoms are the ground atoms defined by the following signature:

plant species names (**P**):	**achillea_millefolium, achillea_ptarmica, ...**
abiotic factors (**A**):	**moisture, acidity, nutrient_value**
values for each of the abiotic factors (**V**):	**very_dry, fairly_dry, ...,**
	basic, neutral, ...,
	nutrient_poor, fairly_nutrient_rich,... .

Predicates:

occurs(P)
is_negative_indication_for(P, A, V)
has_value(A, V)
is_indicative(P)

The constants **achillea_millefolium, achillea_ptarmica, ...** represent the names of the plant species (see Figure 2). The abiotic factors are the three factors introduced in Section 2. The predicate **occurs(P)** refers to the presence of plant species **P** in the sample of the terrain (this is input to the reasoning process). The predicate **is_negative_indication_for(P, A, V)** expresses the fact that abiotic factor **A** does not have value **V**. The predicate **has_value(A, V)** expresses the fact that factor **A** has value **V**, and **is_indicative(P)** the fact that **P** is regarded as an indicative species (giving evidence to the terrain having certain abiotic factors).

There is a set, **KB**, that consists of propositional formulae expressing knowledge (about the domain of determination of abiotic factors), which is of the following form:

- a (large) number of ground instances of:

  is_negative_indication_for(P, A, V)

These instances represent the experts' knowledge of which species may occur in terrains with certain abiotic factors.

- all ground instances of the generic rule

$$\textbf{is\_indicative(P)} \wedge \textbf{is\_negative\_indication\_for(P, A, V)} \rightarrow \neg\, \textbf{has\_value(A, V)}$$

This rule makes it possible to conclude that certain abiotic factors do not have a certain value. This derivation can be made if an indicative species has been found that does not (generally) occur in terrains for which the factor **A** has value **V**.

- statements expressing that for each abiotic attribute at least one value should apply

$$\textbf{has\_value(moisture, very\_dry)} \vee \textbf{has\_value(moisture, fairly\_dry)} \vee \ \textbf{...}$$
$$\textbf{has\_value(acidity, basic)} \vee \textbf{has\_value(acidity, neutral)} \vee \ \textbf{...}$$
$$\textbf{has\_value(nutrient\_value, nutrient\_poor)} \vee$$
$$\qquad \textbf{has\_value(nutrient\_value, fairly\_nutrient\_rich)} \vee \ \textbf{...}$$

For a given set of observed species **OBS**, i.e., input of the form

$$\{\ \textbf{occurs(p)} \mid \textbf{p} \in \textbf{OBS}\ \}$$

the set

$$\textbf{X} = \textbf{KB} \cup \{\ \textbf{is\_indicative(p)} \mid \textbf{p} \in \textbf{OBS}\ \}$$

may be inconsistent. That is, it may be inconsistent to assume that all observed species are indicative for the terrain. This may occur if there is an abiotic factor $\mathbf{A_0}$ such that for all of its possible values **V**, a species **P** is observed that negatively indicates this value (which means we have both **is_indicative(P)** and **is_negative_indication_for(P, $\mathbf{A_0}$, V)**). With the generic rule, the conclusion $\neg$ **has_value($\mathbf{A_0}$, V)** is drawn for all possible values **V** of $\mathbf{A_0}$. But this is inconsistent with the statement **has_value($\mathbf{A_0}$, $\mathbf{V_0}$)** $\vee$ **has_value($\mathbf{A_0}$, $\mathbf{V_1}$)** $\vee$ **...** which is in **KB**. However, as explained earlier, the set of maximal indicative subsets containing **KB** may be considered. This is defined as follows:

Definition 4.3 (Maximal indicative subset)

Let **OBS** $\subseteq$ **P** be a given set of species

a) The set of species S $\subseteq$ **P** is an *indicative set of species* if the theory

KB $\cup$ **{is_indicative(p) | p** $\in$ **S}**

is consistent.

b) The set **S** $\subseteq$ **OBS** is a *maximal indicative subset of* **OBS** if it is an indicative set of species and for each indicative set of species **T** with **S** $\subseteq$ **T** $\subseteq$ **OBS** it holds **S = T**.

The *set of maximal indicative subsets* of **OBS** is denoted by **maxind(OBS)** .

Note that if **OBS** is an indicative set of species itself, there is only one maximal indicative subset of **OBS**, namely **OBS** itself.

Based on these notions the following belief set operator can be defined.

Definition 4.4 (Belief set operator for the application domain)

For a set **X** $\subseteq$ **L**, define the *set of observations implied by* **X** by

OBS(X) = {p | occurs(p) $\in$ **Cn(X)}**.

The *belief set operator* $\mathbf{B_{maxind}}$ is defined by

$\mathbf{B_{maxind}}$ **(X) = { Cn(X φ KB φ {is_indicative(p) | p** $\in$ **S}) | S** $\in$ **maxind(OBS(X)) }**

for each **X** $\subseteq$ **L**.

Actually, here the interesting sets **X** are the sets of the form **{p | occurs(p)** $\in$ **OBS}** for some set of species **OBS** $\subseteq$ **P**. The operator $\mathbf{B_{maxind}}$ satisfies a number of properties of well-behavedness as defined in [8].

The fact that in the case of an observed set of species **OBS** a unique interpretation occurs is expressed as: for each subset of species **OBS** $\subseteq$ **P** the following are equivalent:

(i) $\mathbf{B}_{\text{maxind}}$ **({p | occurs(p) ∈ OBS})** contains just one element.

(ii) the set **OBS** is an indicative set of species.

If these (equivalent) conditions are satisfied, all observed species are indicative, and the user does not need to do selection. The possible values of the abiotic factors are contained in $\mathbf{B}_{\text{maxind}}$ **({p | occurs(p) ∈ OBS})**.

If $\mathbf{B}_{\text{maxind}}$ **({p | occurs(p) ∈ OBS})** contains more than one element, the user must select one. But even before this selection process, conclusions can be drawn: the kernel of the $\mathbf{B}_{\text{maxind}}$ operator contains the most certain conclusions, so $\mathbf{K}_{\mathbf{B}_{\text{maxind}}}$ **({p | occurs(p) ∈ OBS})** may be inspected. For instance, there may be two possible views in $\mathbf{B}_{\text{maxind}}$ **({p | occurs(p) ∈ OBS})** as species have been observed which only grow in dry terrains, and other species have been observed which only grow in moist terrains. However, all of these species may indicate that the terrain is not acid, and this conclusion will be in $\mathbf{K}_{\mathbf{B}_{\text{maxind}}}$ **({p | occurs(p) ∈ OBS})**. If acidity is all one is interested in, there is no need for selection. If one is interested also in the moistness, this selection has to take place. If one is interested in the species which are in both maximal indicative sets, one can either examine $\mathbf{K}_{\mathbf{B}_{\text{maxind}}}$ **({p | occurs(p) ∈ OBS})**, or the intersection of the maximal indicative sets:

$$\mathbf{K}_{\mathbf{B}_{\text{maxind}}}(\mathbf{X}) \cap \{\, \textbf{is\_indicative(p)} \mid \mathbf{p} \in \mathbf{P} \,\} =$$

$$\{\, \textbf{is\_indicative(p)} \mid \mathbf{p} \in \bigcap \textbf{maxind(OBS(X))} \,\}.$$

So, the kernel contains the atoms **is_indicative(p)** precisely for **p** in the set of shared plant species (see Figure 3, lower part), which is the intersection of the maximal indicative subsets (the two rectangles in Table 3).

The formalisation in terms of a belief set operator is semantical of nature. However, a syntactical representation can be found as well, in terms of a normal default theory based on (in addition to the world theory **KB**) the following set of defaults **D**:

(occurs(p) : is_indicative(p)) / is_indicative(p) for all species **p** in **P**.

To see the equivalence, the following is needed. Let $\Theta = \langle \mathbf{W}, \mathbf{D} \rangle$ be a default theory. A set of sentences **E** is called a *Reiter extension* of Θ if the following condition is satisfied:

$$\mathbf{E} = \bigcup \mathbf{E}_i$$

where

$$\mathbf{E}_0 = \mathbf{Cn(W)},$$

and for all $\mathbf{i = 0}$

$$\mathbf{E_{i+1} = Cn(E_i \cup \{ \omega\_ \mid (\alpha : \beta_1,...,\beta_n) / \omega\_ \in D, \alpha \in E_i \text{ and } \neg \beta_1 \notin E, ... , \neg \beta_n \notin E \})}$$

The *set of Reiter extensions* of Θ is denoted by $\mathbf{Ext(\Theta)}$.

Theorem 4.5

The belief set operator $\mathbf{B_{maxind}}$ is representable by the normal default theory $\mathbf{< KB, D >}$, i.e., for all $\mathbf{X}$ it holds $\mathbf{B_{maxind}(X) = Ext(<KB \cup X, D>)}$.

The proof is as follows. Let $\mathbf{X}$ be a set of formulas in $\mathbf{L}$. Let $\mathbf{X \cup KB}$ be consistent (if it is not, verification is straightforward and omitted). The extensions of $\mathbf{< KB \cup X, D >}$ are sets of the form $\mathbf{Cn(KB \cup X \cup S)}$, where $\mathbf{S}$ is a subset of $\mathbf{\{ is\_indicative(p) \mid occurs(p) \in Cn(X) \}}$, which is maximal such that $\mathbf{Cn(KB \cup X \cup S)}$ is consistent. This is proved below. The sets $\mathbf{Cn(KB \cup X \cup S)}$ with $\mathbf{S}$ as above together comprise $\mathbf{B_{maxind}(X)}$. First of all, let $\mathbf{S}$ be such a maximal set, and let $\mathbf{E = Cn(KB \cup X \cup S)}$. Then if the $\mathbf{E_i}$ are defined in the definition of a default extension as above, the following holds:

$$\mathbf{E_0 = Cn(KB \cup X)},$$

$$\mathbf{E_1 = Cn(E_0 \cup \{ is\_indicative(p) \mid occurs(p) \in E_0 , \neg is\_indicative(p) \notin E \})}$$

As $\mathbf{E_1}$ does not contain more instances of the **occurs** predicate than $\mathbf{E_0}$ (this follows from the fact that $\mathbf{X}$ contains only the **occurs** predicate, whereas $\mathbf{KB}$ does not), $\mathbf{E_i = E_1}$ for all $\mathbf{i > 1}$. The claim is that

$$\mathbf{\{ is\_indicative(p) \mid occurs(p) \in E_0 , \neg is\_indicative(p) \bullet E \} = S}.$$

Suppose $\mathbf{occurs(p) \in E_0}$ and $\mathbf{\neg is\_indicative(p) \notin E}$. Then $\mathbf{occurs(p)}$ is in $\mathbf{Cn(X)}$ and $\mathbf{Cn(KB \cup X \cup S \cup \{ is\_indicative(p) \})}$ is consistent. But as $\mathbf{S}$ was maximal with respect to these properties, $\mathbf{is\_indicative(p) \in S}$. On the other hand, if $\mathbf{is\_indicative(p) \in S}$, then $\mathbf{occurs(p) \in E_0}$ and $\mathbf{\neg is\_indicative(p) \notin E}$ (as $\mathbf{E = Cn(KB \cup X \cup S)}$ is consistent). Now let $\mathbf{E}$ be an extension of $\mathbf{< KB \cup X, D >}$, then it is of the form $\mathbf{Cn(KB \cup X \cup S)}$, where $\mathbf{S}$ contains (only) formulas of the form $\mathbf{is\_indicative(p)}$. Examination of $\mathbf{KB}$ (and the restriction on the language of $\mathbf{X}$), shows that only if $\mathbf{occurs(p) \in Cn(X)}$ is $\mathbf{is\_indicative(p) \in E}$. As extensions are always consistent (if each rule has a justification and the axioms are consistent), $\mathbf{Cn(KB \cup X \cup S)}$ must be consistent. Suppose there exists a $\mathbf{T \subset S}$ (strict inclusion) respecting the conditions, then there must be a default rule $\mathbf{occurs(p) : is\_indicative(p) / is\_indicative(p)}$, with $\mathbf{occurs(p) \in Cn(X) \subseteq E}$ and $\mathbf{Cn(KB \cup X \cup S \cup \{ is\_indicative(p) \})}$ consistent, implying that $\neg$ **is_indicative(p)** $\notin$ **E**. But that means there is an applicable default rule for which the conclusion is not in **E**, contradicting the assumption that **E** is an extension. Therefore **S** must be maximal.

At this point the reader may wonder what the benefit is of the (syntactical) representation in default logic. The belief set operator $\mathbf{B}_{\mathbf{maxind}}$ arose during the analysis and formalization of an application to be described in the next section. A system, EKS, was implemented based on this operator $\mathbf{B}_{\mathbf{maxind}}$. The implementation in fact follows the definition (Definition 4.1) rather closely. A representation in terms of default logic may provide some more familiarity for readers. Besides, the results of the current section indicate that alternatively a theorem prover for default logic (or, rather, a program computing extensions of default theories) could be used.

5. Correspondence Between the Formalisation and the System

The correspondence between the formalisation of the expert reasoning task and the interactive knowledge-based system EKS that models the task is shown in Figure 6. The first component of the system, determination_of_maximal_indicative_subsets, is formalised by the belief set operator $\mathbf{B}_{\mathbf{maxind}}$ defined in Section 4 (depicted by the grey arrow at the left hand side in Figure 6). The component selection_of_a_maximal_indicative_subset (which models the selection process by the user) is formalised by a single-valued selection function $\mathbf{s}_{\mathbf{user}}$ (depicted by the grey arrow at the right hand side in Figure 6). The composition $\mathbf{C}_{\mathbf{EKS}}$ of $\mathbf{B}_{\mathbf{maxind}}$ and $\mathbf{s}_{\mathbf{user}}$ defined by

$$\mathbf{C}_{\mathbf{EKS}}(\mathbf{X}) = \mathbf{s}_{\mathbf{user}}(\mathbf{B}_{\mathbf{maxind}}(\mathbf{X})) \qquad \text{for } \mathbf{X} \subseteq \mathbf{L}$$

is a non-monotonic inference operation, which is selective for $\mathbf{B}_{\mathbf{maxind}}$ (as described in Definition 4.2b). This inference operation formalises the reasoning of the system in interaction with the user as a whole (depicted by the grey arrow at the bottom of Figure 6). Note that one of the two functions of which this overall function is composed, is fixed and defined by the system itself (i.e., $\mathbf{B}_{\mathbf{maxind}}$), and that the other function can be changed dynamically, depending on the user (i.e., $\mathbf{s}_{\mathbf{user}}$).

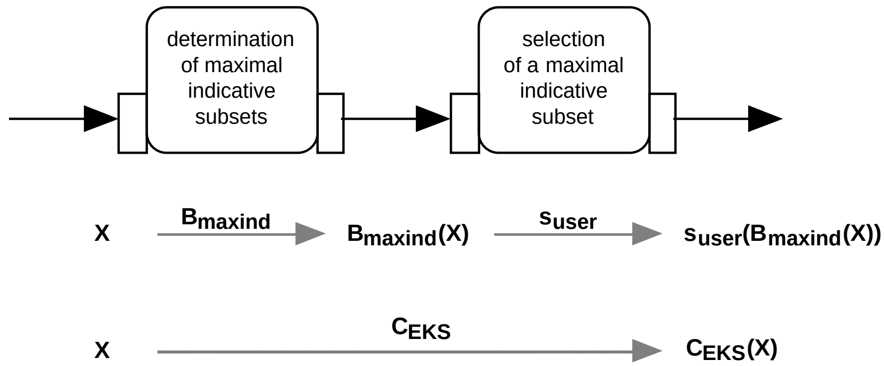

Figure 6. Correspondence between the formalisation and the system.

6. Discussion

The outcomes of the work reported in this paper can be discussed at two levels: the level of the specific application domain and system, and the more generic level of the logical techniques used.

6.1. Domain of Application and EKS

The multi-interpretability of samples of plant species has proven to be a central issue in this domain of application. Given the assumption that samples are always correct (the plant species named are indeed the plant species encountered), and that samples are only taken from sites which are homogeneous, the only reason for conflicting indicative information is that the specific domain knowledge on which conclusions are based is incorrect or incomplete. During the design of EKS this specific domain knowledge was continual subject of discussion between experts. The knowledge currently implemented in EKS is the result of consensus between experts, and is no longer a likely reason for conflicting indicative information.

The lack of homogeneity of a terrain is the cause of most conflicts, requiring additional expert knowledge to understand the nature of the inhomogeneity. The reason for the lack of homogeneity can, for example, be vertical stratification, as in the inhomogeneous example discussed earlier. Another possibility is the development of a terrain over time: what has and has not been done to a terrain can influence its vegetation and transitions in vegetation. Inhomogeneous terrains are more common than initially supposed: multi-interpretable samples are not the exception, but the rule. The way in which experts analyse samples from inhomogeneous terrains was at first unclear. A first model determined the, in some sense ‘average’, conditions for the species in the sample. This model, although originally agreed by the experts involved to be an acceptable model, did not work: experts found it difficult to interpret the outcome of the analyses. The second model displayed the ranges of conditions encountered as a kind of summary of results. However, this model was problematic for two reasons: (1) It was unclear to the user whether the ranges of conditions displayed were meant to be possible for all species or only for subsets (the difference between an *or* interpretation of a range and an *and* interpretation), (2) The different values in the ranges could not be traced back to the species on which they are based. As a result of experiencing these first two experimental models, the experts agreed that the different views of a sample were essential to the analysis of the plant species observed. EKS identifies these views and presents these views to the user. Which view is (or which views are) most appropriate requires additional heuristic (strategic) knowledge. The selection of a view is currently performed by the user of the system. Future research will focus on the acquisition of this knowledge to be able to support users in the selection process.

One of the research questions to be addressed as well might be whether the basic assumption used, namely that the three main factors are considered independent variables could be replaced by some dependence relations as well. However, domain experts have strongly preferred to consider them as independent until now, and don’t see apparent dependence relations.

Other research questions concern the applicability of the approach in other domains. In further research it is aimed to find another suitable realistic domain and to apply the approach in this domain.

6.2. The Logical Techniques Used

The idea that information about the world can often be interpreted in different and conflicting manners was a central theme in the research reported in [14], [8]. Using techniques to formalise non-monotonic reasoning, such as default logic (e.g., [13], [5], [11]), often different (and often conflicting) possible outcomes of a reasoning process are obtained. In the area of research on non-monotonic reasoning, in general this is considered to be disturbing (e.g., it is called the multiple extension *problem*). To come to one set of conclusions, in the literature often the non-monotonic inference operation defined by the intersection of all possible outcomes is taken (sceptical approach), or sometimes the union of all possible outcomes (credulous approach). (The original paper on default logic, [13], however, proposed that a choice should be made for one outcome, using some mechanism outside default logic itself.)

For a particular domain such as the ecological domain addressed in this paper, both approaches are unsatisfactory: the sceptical approach often does not lead to any possible conclusions on the abiotic conditions, whereas the credulous approach often leads to inconsistent information. For reasons like these, in [14], [8], [9], [12] the multiple outcomes of a non-monotonic reasoning process are not considered to be a problem, but are instead exploited as a useful feature that can provide an adequate formalisation of the multi-interpretability often present in real-life information. In [14] this feature is expressed by adding as an extra parameter a selection function to a default theory. In [6] and [7] a similar approach is developed, based on priority orderings between defaults. In [8] the notion of belief set operator is introduced to formalise the multiple outcomes of a non-monotonic reasoning process, and a selection operator to make a choice between the different options.

For the application domain discussed in this paper the latter approach is more suitable, because in this approach first all alternative interpretations are generated, and the selection is made afterwards. In the approaches of [14], [6], and [7] the reasoning process itself is controlled by the selection knowledge in such a manner that only one outcome is generated, and other options remain invisible. Such strategic knowledge is not yet available. However, in the future of this project such strategic knowledge may be acquired so that not all possible options need to be generated. In that case approaches as described in [14], [6], or [7] might become useful. Another issue for future research is to characterize the domains in which the approach discussed in this paper for the ecological domain can be applied, thus making the method more general.

Acknowledgements

Within the EKS-project a number of persons have contributed their expertise: Frank Cornelissen, Edgar Vonk (scientific programmers); Christine Bel, Rineke Verbrugge (for parts of the analysis of the domain); Bert Hennipman and Frits van Beusekom (domain experts from IPTS); Piet Schipper, and Wim Zeeman (domain experts from Staatsbosbeheer).

References

[1] F.M.T. Brazier, B. Dunin-Keplicz, N.R. Jennings, and J. Treur, “Formal Specification of Multi-Agent Systems: a Real-World Case”, in: V. Lesser (ed.), *Proc. of the First International Conference on Multi-Agent Systems, ICMAS’95*, MIT Press, Cambridge, MA, pp. 25-32. Extended version in: *International Journal of Cooperative Information Systems*, M. Huhns, M. Singh, (eds.), special

issue on Formal Methods in Cooperative Information Systems: Multi-Agent Systems, vol. 6, 1997, pp. 67-94.

[2] F.M.T. Brazier, J. Treur, and N.J.E. Wijngaards, “The Acquisition of a Shared Task Model”, in: N. Shadbolt, K. O’Hara, G. Schreiber (eds.), *Advances in Knowledge Acquisition, Proc. 9th European Knowledge Acquisition Workshop, EKAW’96*, Lecture Notes in Artificial Intelligence, vol. 1076, Springer Verlag, pp. 278-289.

[3] F.M.T. Brazier, J. Treur, and N.J.E. Wijngaards, “Modelling Interaction with Experts: the Role of a Shared Task Model”, in: W. Wahlster (ed.), *Proc. European Conference on AI, ECAI’96*, John Wiley and Sons, 1996, pp. 241-245.

[4] F.M.T. Brazier, J. Treur, N.J.E. Wijngaards, and M. Willems, “Formal Specification of Hierarchically (De)Composed Tasks”, in: B.R. Gaines, M.A. Musen (eds.), *Proc. of the 9th Banff Knowledge Acquisition for Knowledge-based Systems workshop, KAW’95*, Calgary: SRDG Publications, Department of Computer Science, University of Calgary, 1995, pp. 25/1-15/20.

[5] P. Besnard, *An Introduction to Default Logic*, Springer-Verlag, 1989.

[6] G. Brewka, “Adding Priorities and Specificity to Default Logic”, in: C. MacNish, D. Pearce, L.M. Pereira (eds.), *Logics in Artificial Intelligence, Proceedings of the JELIA-94*, Lecture Notes in Artificial Intelligence, vol. 838, Springer-Verlag, 1994, pp. 247-260.

[7] G. Brewka, “Reasoning about Priorities in Default Logic”, in: *Proceedings of the AAAI-94*, 1994.

[8] J. Engelfriet, H. Herre and J. Treur, “Nonmonotonic Reasoning with Multiple Belief Sets”, in: D.M. Gabbay, H.J. Ohlbach (eds.), *Practical Reasoning, Proceedings FAPR’96*, Lecture Notes in Artificial Intelligence, vol. 1085, Springer-Verlag, 1996, pp. 331-344.

[9] J. Engelfriet, V.W. Marek, J. Treur and M. Truszczynski, “Infinitary Default Logic for Specification of Nonmonotonic Reasoning”, in: J.J. Alferes, L.M. Pereira, E. Orlowska (eds.), *Logics in Artificial Intelligence, Proceedings of the Fourth European Workshop on Logics in AI, JELIA’96*, Lecture Notes in Artificial Intelligence, vol. 1126, Springer-Verlag, 1996, pp. 224-236.

[10] D. Makinson, "General Patterns in Nonmonotonic Reasoning", in: D.M. Gabbay, C.J. Hogger, J.A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3,* Oxford Science Publications, 1994, pp. 35-110.

[11] V.W. Marek and M. Truszczynski, *Nonmonotonic logics; context-dependent reasoning*, Springer-Verlag, 1993.

[12] V.W. Marek, J. Treur and M. Truszczynski, "Representation Theory for Default Logic", *Annals of Mathematics and Artificial Intelligence*, vol. 21, 1997, pp. 343-358.

[13] R. Reiter, "A Logic for Default Reasoning", *Artificial Intelligence* 13, 1980, pp. 81-132.

[14] Y.-H. Tan, J. Treur, "Constructive Default Logic and the Control of Defeasible Reasoning", in: B. Neumann (ed.), *Proc. of the European Conference on Artificial Intelligence, ECAI'92*, John Wiley and Sons, 1992, pp. 299-303.