Imperfect information management in GIS: A logical-mathematical approach

H. Hosni, Manchester University and ISTI-CNR http://www.maths.man.ac.uk/ hykel/
hykel@maths.man.ac.uk mailto:hykel@maths.man.ac.uk
M.V. Masserotti, ISTI-CNRhttp://www.isti.cnr.it
Maria.Vittoria.Masserotti@isti.cnr.it mailto:Maria.Vittoria.Masserotti@isti.cnr.it
C. Renso, ISTI-CNR http://www.isti.cnr.it
Chiara.Renso@isti.cnr.it mailto:Chiara.Renso@isti.cnr.it

March 9, 2004

Abstract

We survey the relevance of the maximum entropy inference process to the management of uncertainty in Geographic Information Systems. We show how the former constitutes a theoretically well-founded solution to problems that arise naturally in GIS facing imperfect information. We also put forward how, as a consequence of the encouraging developments on computational techniques for reasoning under maximum entropy, the latter must be considered as a most crucial approach to uncertainty management in various fields of GIS science.

1 Introduction

A relevant problem in approaching GIS (Geographic Information Systems) consists in giving a reasonably agreed definition of what GIS actually are. For present purposes it is reasonable to consider GIS as being characterized by a twofold nature: On the one hand GIS consist of a *technology* used for certain purposes. From this perspective, the crucial issues in GIS research amount to *computing* problems, both on the hardware and software level. On the other hand, however, GIS research is increasingly more focussed on *theoretical* issues concerning the representation of geographic information. According to the latter point of view GIS problems include, at the very least, issues of *knowledge representation and reasoning*. In this paper we investigate some of the consequences deriving from approaching GIS from the latter point of view. In particular, we will be insisting on the fact that its 'conceptual side', so to speak, commits GIS research to achieving *scientific goals* which happen to be closely related to some of those pursued in Artificial

Intelligence (AI) research.[1] In doing so, we adopt a perspective according to which GIS are essentially construed as artificial intelligent agents reasoning about a certain classes of natural environments.

The reason for our focussing on the issues of knowledge representation and reasoning in relation to GIS is that this perspective provides us with a promising methodological framework for tackling a critical issue in contemporary GIS science: the management of *imperfect information*. As we'll be insisting in what follows, by adopting this perspective we can provide and then exploit a strong link between GIS science and the logical-mathematical discipline of *uncertain reasoning*.

A remark is necessary here in order to avoid possible misunderstandings. Although the distinction between "the scientific" and "the technological" is by no means a sharp one –especially in the domain of information systems– we see plenty of methodological advantages in considering the scientific aspects of GIS as a rather separated issue: it is only by clearly stating where we want to go that we can attempt to tell whether we are following a reasonable path.

The paper is organized as follows. Section 2 introduces and motivates the agent-perspective on the main problem. Section 3 analysis the inevitability of imperfect information in a way that motivates the need for theoretical solutions to the problem. It turns out that our approach is relatively close to the *naïve geography* programme. Section 3.3 is devoted to the investigation of the main commonalities as well as differences between the two perspectives. Section 4 introduces the crucial aspects of what we consider to be an "intelligent GIS-agent". Equipped with this conceptual and methodological framework we move on to discuss the essential aspects of the corresponding formal setting. We do so by introducing a case study concerning terrestrial GIS in section 5 and illustrating in section 6.2 the applicability of *maximum entropy logic programming* to a representative class of GIS-queries extrapolated from the case study. The brief discussion on the implementation of maximum entropy reasoning of section 7 illustrates how the latter constitutes a methodologically principled as well as computationally viable solution to the management of uncertainty in GIS. Section 8 concludes the paper by pointing to the future directions of research.

2 From a tool-box to a multi-agent system: Putting GIS into the agents perspective.

We have mentioned above the fact that GIS are characterized by a twofold nature. In the *practice* of GIS it seems that this fact has been implicitly

[1] Notice that the 'obvious' connection between GIS and AI research is to be found on the computational essence of the two areas. As we will see, however, much more fine-grained similarities are worth investigating.

assumed over the past two decades, as witnessed by the following definitions.

Duecker 1979 "A geographic information system is a special case of information systems where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines, or areas. A geographic information system manipulates data about these points, lines, and areas to retrieve data for ad hoc queries and analysis."

Burrough 1986 [3] "A powerful set of tools for storing and retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes."

Clarke 1995 [6] "Automated systems for the capture, storage, retrieval, analysis, and display of spatial data."

Chrisman 1997 [5] "The organized activity by which people:

measure aspects of geographic phenomena and processes;

represent these measurements, usually in the form of a computer database, to emphasize spatial themes, entities, and relationships;

operate upon these representations to produce more measurements and to discover new relationships by integrating disparate sources;

transform these representations to conform to other frameworks of entities and relationships.

These activities reflect the larger context (institutions and cultures) in which these people carry out their work. In turn, the GIS may influence these structures."

What is the common path followed by these definitions? It is quite clear that the *fil rouge* among them all is to be found in the characterization of GIS goals: *devising computing systems capable of collecting, representing, storing, manipulating and retrieving geographic information.* On the basis of such a tentative definition, we can explore some relevant commonalities between GIS and AI research.

Put crudely, AI aims at the design of autonomous agents exhibiting intelligent behaviour in non-trivial environments[2]. It is commonly held that AI has, as we maintain for GIS research, two sorts of goals: in the first place accounting for such a thing as an "agent behaving intelligently in a non-trivial environment" and then, building it up. The former encompasses the "scientific goals" of AI whereas the latter is what gives AI its "engineering" flavour. Without going into the details, noteworthy at this point

[2]See e.g. [46, 32, 39]

is that the crucial issues of the former are essentially related to knowledge representation and reasoning.[3]

Among the above mentioned definitions, the sort of parallel we are drawing here is quite easily traced in Duecker's account where particular emphasis is put on the fact that knowledge representation and reasoning in GIS amount to a *particular instantiation* of a general AI problem: accounting for the "intelligent" management of information.

Here is our first methodological step: by regarding the geographic ones as a special case of (general) information management systems, an entirely natural link can be constructed between GIS and uncertain reasoning research. We address this problem in section 3.

In order to appreciate how logical theories of reasoning under imperfect information could be deployed in modelling "intelligent GIS", however, we have to make explicit the fact we are thinking of GIS as essentially *autonomous agents*. Among the above mentioned characterizations of what GIS actually are, only the one due to Chrisman considers explicitly the (human) agent acquiring and manipulating information as an essential part of the geographic information system.[4] Our interest, however, goes somehow beyond Chrisman's definition, for we think of GIS as being centered around autonomous *artificial* agents. For this reason, throughout this paper, we will be calling such broader systems **GIS-agents**.

2.1 GIS-agents and multi-agent systems

A major achievement of the past two decades in AI research has been the introduction of the 'agents metaphor' in the study of intelligent artificial systems.[5] Among its many consequences one of particular interest to us is the anchorage of the rather abstract notion of "reasoning" to several types of environments. In a nutshell, an agent is understood as an *embodied entity*, that is to say an entity whose reasoning is about an environment which is subject to its (and other agents') actions. Were we able to carefully explain how this *interaction* takes place, most of the fundamental questions of AI would be answered. It goes without saying that we set ourselves a much more modest aim here: accounting for the relation between GIS-agents and some aspects of the formalization of reasoning under imperfect information.

The main reason we argue in favour of the necessity of the agents perspective is easily explained by noticing that were GIS mere tool-boxes, no

[3] We are extremely sympathetic with the view proposed in [40] according to which "the problems of AI are not primarily computing problems (hardware or software), but problems concerning what it is for a device to reason."

[4] Notice however that the other definitions do not appear to be inconsistent with such an idea.

[5] Due to obvious space constraints we plainly assume here the agents perspective in AI. Complete details can be found in [20] and in the more recent [51].

sensible relevant issue concerning GIS *reasoning* would even arise, let alone reasoning under imperfect information.

But there are further methodological reasons for turning our attention to the relation between GIS and (multi-)agent systems research. We have stressed in the previous section that the *practice* of GIS requires a variety of agents (both human and not) to interact.[6] Therefore it is extremely natural to think of GIS as multi-agent systems in which the various tasks characterizing the diverse nature of GIS are distributed, say among agents responsible for acquiring, storing, manipulating relevant information. As a consequence, approaching theoretical GIS research from the agents perspective opens up to a variety of computing issues. We limit ourselves to list the mayor ones:

communication : i.e. the sharing of information among agents;

cooperation : i.e. the joint effort of several agents aimed at achieving some goal which would go beyond their individual capabilities;

negotiation : i.e. the activity by means of which several agents (who normally have competing goals) achieve agreement.

Such issues are being extensively studied in the broader domain of multi-agent systems (see e.g. [51] for a recent reference work). Thus it seems quite clear that when talking about the role of multi-agent systems in theoretical GIS research, the 'direction of fit' is understood to be from the general to the particular, that is to say, the general framework developed in the context of multi-agent systems will have to be adapted to the particular case of GIS-agent.[7] Being this study essentially devoted to the foundational issues, we postpone further details on this until subsequent stages of the research underway.

Rather, our focus now is on the desirable *abstract properties* to be imposed on GIS-agents modelling. In this context 'abstract' points both to the fact that such properties are largely independent of the implementing details and on the fact that they are investigated in the formal framework of mathematical logic.

Abstraction plays nowadays a critical role in multi-agent system design. As a consequence, a variety of *architectures* have been proposed during the past decade, the best known of which is perhaps the so-called BDI architecture [41, 20, 50]. The present paper, as part of a broader study, is just

[6]In the perspective of AI, we clearly aim at delegating as much workload as possible to artificial agents, putting ourselves in the comfortable position of 'users'. As discussed in section 3.3 this is one of the motivations underlying the 'naïve geography' approach.

[7]There are surely different sorts of constraints characterizing the ontology of, say, marine and terrestrial GIS. The abstract framework that we set out to develop, however, aims at capturing the problem of uncertainty management in GIS in its full generality.

oriented along the lines of abstraction. Hence our present aim consists in a first investigation on the crucial issues concerning imperfect information representation and management in GIS-agents.

Summing up, the relation between research in (multi)agent systems and GIS appears to be as natural as it is necessary. We have appreciated how natural it is simply by recalling the definitions of geographic information systems which arise from the practice of actually deploying GIS. Thus, it turns out to be extremely natural to think of a GIS-agents as a distributed system. Among distributed/concurrent systems, however, multi-agent systems provide a compelling framework for the development of GIS. This happens at two levels: on the computing level, advances in multi-agent systems just have to be contextualized according to the problems arising in GIS; whereas on the abstract level, the agents perspective, by anchoring GIS to some environment, provides a necessary key for tackling our main concern: imperfect information management, to which we now turn our attention.

3 Uncertainty and GIS

Like any other agent operating in *realistic environments* –that is to say environments complicated enough so as to be seen as portions of the "real world"– GIS *ought* to face the inevitability of imperfect information.[8] But *why* is it the case? This is best explained by briefly recalling what sort of features characterize realistic agents.

As we have just seen, agents make full sense only in terms of embodied entities. Then, whether they have a physical (robots) or software (softbot) nature, agents are necessarily *limited* entities. This has crucial consequences that we now briefly consider by looking at the sort of things we expect a realistic agent like a GIS-agent to do.

First of all, an agent acquires information by means of perceiving its environment. Then perceived information (or data) is suitably represented and eventually stored in the system's memory. Let's call the set of stored representations of information, the agent's *knowledge base*.[9] We take, for present purposes, agents' *reasoning* as a function of their knowledge base whose output consists of *new* pieces of information (or equivalently, old ones made more specific). From this point of view, as it will be exemplified in section 6.2 below, the modelling of GIS-agents reasoning arises from the specification of suitable constraints on their *inference process*.

Agents' reasoning process might be triggered by a variety of different circumstances which might span from performing particular actions to an-

[8]This is even more true in the case of marine GIS; more about this below.

[9]Within GIS research, the term *database* is often preferred. Databases, usually, correspond to particularly structured pieces of information. At this stage, however, it seems more appropriate to deal with a more general notion.

swering certain queries. Whatever the goal reasoning is to fulfill, we expect intelligent agents to make the best possible use of the information available to them, given their essential limitations. As we are about to discuss, it is just because of such limitations that agents cannot undertake the sort of reasoning formally captured by classical logic.

Having drawn this rough picture of the GIS as an agent we now tackle the crucial question: Where does all the uncertainty come from?

As far as realistic agents are concerned, there are essentially two sources of imperfection which GIS research has to care about and which we call "agent-independent" and "agent-dependent" respectively. Those sources are simultaneously active at the perception, representation and reasoning levels and can purposefully be distinguished. The reason for so doing is that of distinguishing between *technological* and *scientific* problems, that is to say distinguishing among problems which should be tackled at the hardware (or software) level and those for which an abstract, theoretical solution is needed.[10] Thus whereas agent-dependent sources of imperfection are generally to be tackled technologically, the agent-independent ones require geographic information systems to satisfy (some of) the rational/commonsensical[11] constraints on agents' reasoning as studied in the context of *uncertain reasoning* research.

3.1 Agent dependent vs agent independent limitations

To the purpose of briefly illustrating the contrast we are making here, let's take the case of marine GIS. As put forward in [29] two crucial activities in collecting appropriate information (equivalently: data acquisition) are *sensing* and *measuring*. Both activities inevitably generate imperfection in the resulting representations *independently* of the agents' representational systems but *dependently* on the nature of agents and their environment.

Sensing is clearly subject to the actual physical constraints of the particular agents at hand. Although there will always be a limited number of objects that the agent can simultaneously perceive from a finite (and predetermined) number of distinct perspectives, the number of such objects and perspectives can clearly be maximized by means of appropriate technological solutions.[12]

Nonetheless, as just remarked, it is evident that –even in principle– the information agents are able to collect cannot be exhaustive if they operate in realistic environments. This leads to what we will be calling **incomplete**

[10] Unlike "abstract", in the context of our study, "theoretical" will essentially mean "logical-mathematical".

[11] We do not distinguish here between the (slightly different) notions of "rationality" and "commonsense".

[12] A point closely related to the latter has been investigated within the context of *naïve geography* to which our approach is compared in section 3.3.

knowledge-bases. For the present we take the latter to mean that there is some aspect of the agent's environment which is not represented in the agent's knowledge base. Such an issue cannot be solved by means of improving, say the perceptual processes of the agent itself. Rather, what is needed, in order for the GIS to make the most of the available information, is a clear understanding of what sort of constraints that should be imposed on the agent's reasoning function so as to allow it to draw commonsensical conclusions given its incomplete knowledge base.

Having said that, still one might be inclined to think that such imperfect knowledge-bases might amount to incomplete yet accurate (i.e. crisp) representations of the agents' environment(s). Due to the intrinsic *granularity of measuring* however, this does not turn out to be the case.[13] In a nutshell, the problem of granularity consists in the fact that when measuring say, in the case of marine GIS through water devices, agents transform essentially continuous entities into essentially discrete representations. This, in general, affects the representation of *spatio-temporal* relations, leading to the phenomenon of **vagueness**. The implications of incomplete and vague knowledge bases are described in more details in section 4, whereas in section 6.2 we focus on one particularly important case of incomplete knowledge bases leading to *background ignorance*.

The imperfection which originates as a byproduct of data acquisition clearly interferes with the subsequent manipulation of the stored representations. Therefore the imperfection which is so 'inherited' affects the capability that agents have of drawing conclusions from their knowledge bases, like e.g. planning some appropriate actions or answering some particular queries.

We have briefly pointed out above that although they might enable better performances on the whole, technological solutions cannot annihilate such interferences: Due to the presence of agent-independent constraints, this cannot be the case, even in principle. Therefore, we claim, in order for GIS-agents to overcome the difficulties in handling such imperfect knowledge bases, they have to satisfy some commonsense constraints. Acknowledging the inevitability of imperfect information, our proposal suggests understanding how imperfect information can be *managed* by means of commonsensical reasoning.

3.2 The need for theoretical solutions

Consider again the problems related to sensing and measuring. Their negative effects –i.e. the amount of imperfection generated– can be *reduced* by means of suitable hardware/software solutions. For example the range of data and the amount of details acquired crucially depends on the nature of

[13] Notice also that in normal circumstances sensing will also lead to inaccurate representations due to various forms of noise.

sensors: the more efficient the sensors, the less the imperfection generated by them.[14]

To see why technological improvements cannot guarantee the required solution to the problem of imperfect information management we need to focus on the nature of realistic environments. Let's focus, for the time being, on marine environments (a proper case study concerning terrestrial GIS is discussed in section 5). Of particular interest for our purposes is to notice that *any portion of marine environment corresponds to a inhomogeneous and proactive environment.* Marine environments are **inhomogeneous** to the extent that the behaviour on the shoreline is considerably different from the one on open sea, with respect to, say, marine currents. Of a similar nature is the problem constituted by *temperature fronts*, that is to say areas where a significant change of temperature is recorded within tiny distances. Thus the inhomogeneity of marine environments gives rise to the phenomenon of **context dependent information**, that is to say, information pertinent to a specific marine area may not be adequate for other –even 'similar', in the case of fronts– areas. Of course, the fact that marine information is essentially layered adds in further variables.[15]

On the other hand, **proactiveness** captures the fact that marine environments change over time independently of agents' actions in a way that agents cannot, in general, predict. Thus, proactiveness is much stronger than dynamism. Proactiveness clearly makes any knowledge base whatsoever imperfect to the extent it puts the agent in a condition that we'll be calling **background ignorance**: at any point in time the agent cannot be taken to have complete information about its environment. As discussed later on, the phenomenon of background ignorance is to be found in any (marine or terrestrial) applications of GIS.

Let us put to work what has been said so far. By means of briefly analyzing the GIS tasks it emerged that there are essentially two possibilities in order to cope with imperfect information in GIS: either by improving the mechanisms of data acquisition/representation –and this, we have suggested, would essentially remain within the domain of GIS technology– or by isolating constraints that would enable commonsensical reasoning in GIS. The existence of agent-independent sources of imperfection, however, rules out the former as a satisfactory –in terms of generality– solution to our main problem. This clearly leaves open to us only the latter option.

[14] Similar remarks apply to the representation of geographical information. In this context, an early typical example is given by the *raster* vs. *vector* debate with respect to models for continuous surface-data.

[15] Normally such further variables correspond to : satellite data, marine current data, bathymetric data and sea floor data. Such an essentially layered structure of data acquisition on marine GIS, that is to say the fact that information is distributed through space, very naturally suggests a multi-agent approach to GIS-agents.

3.3 GIS and common sense: the *Naïve Geography* approach

Those pointed out above are not the only reasons for imposing commonsense constraints on GIS-agents overall behaviour. With respect to its main goal –enabling commonsense reasoning in GIS– the *naïve geography* approach [10] is indeed notably close to the perspective we are suggesting in the present study, though the motivations underlying the naïve geography programme are somehow different.

Naïve geography is defined as the "the body of knowledge that people have about the surrounding geographical world" [10]. The main motivation for the formalization (and hence the implementation) of "naïve geographic knowledge" consists in allowing non specifically trained people to make the most of their interaction with geographic information systems. We now focus on the most important (in our view) commonalities while postponing further discussion on the differences until the end of this section.

Spatial and temporal reasoning A cornerstone of the framework we are suggesting consists in taking GIS to be autonomous agents, that is to say, spatially and temporally embodied entities. Spatial and temporal reasoning, on the other hand, are "central to naïve geography"[10]. Notice, however, that we have not directly argued for specific spatio/temporal constraints for commonsensical GIS. Rather we have limited ourselves to notice that –already at a purely abstract level– taking GIS as (multi)agent systems is sufficient to bring spatially and temporally embedded reasoning into the picture. Recall, from section 3.2, that without agents' reasoning being temporal no notion of proactiveness could make sense, pretty much in the same way no issue of context-dependence can arise in non-spatial reasoning. It must be stressed that spatio/temporal reasoning enters *explicitly* into the picture only by means of appropriate formal constraints.

Realistic environments The fact that GIS-agents are embedded in realistic environments constitutes a key assumption underpinning our framework. We have seen before how this, combined with their necessary physical limitations, results in agents being able to perceive a limited number of objects from a limited number of different perspectives. Those are precisely the features that characterize "geographic space" in the naïve geography approach.

Qualitative reasoning We have paid particular attention to the phenomenon of background ignorance stressing that it corresponds, by and large, to the usual epistemic or informational state of the agent. Although this is a topic for the next sections, it is appropriate to recall here that a very similar concern is put forward in the naïve geography approach through the notion of "qualitative reasoning":

> In qualitative reasoning a situation is characterized by variables that can only take a small, predetermined number of values and inference rules that use these values *in lieu* of numerical quantities approximating them. Qualitative reasoning enables one to deal with partial information, which is particularly important for spatial applications when only incomplete data sets are available [10] .

We conclude this section by considering some significant differences between our approach and the one developed in the context of naïve geography. In the formalization of geographical commonsense reasoning as proposed by the latter, geographical space is two-dimensional. In the general context of GIS (including marine ones) however, this cannot be the case. Indeed –as we have stressed above– marine space is extremely inhomogeneous requiring *a fortiori* GIS-agents to make sense of both partial and contextual information. Related to the original motivations is the second main point of divergence. By aiming at making easier the deployment of GIS technologies by non-experts, naïve geography is focussed on human-machine interaction to an extent that approach is not.

4 Characterizing imperfect information management

Unfortunately, acknowledging a tight interaction between GIS and uncertain reasoning research does not provide the former with a dried and cut account of what it *is* to manage imperfect information. To-date, it is by and large impossible to satisfactorily represent *all* the interesting features of reasoning under imperfect information within a unique mathematical framework. This in not to say, however, that *no* feature at all can be captured. It is precisely in this spirit that we focus on rational/commonsensical constraints on agents' reasoning.

We have seen above that reasoning about (and hence acting upon) geographic domains already requires the ability to face background ignorance, context-dependency and vagueness. The purpose of this section is to investigate the sort of formal constraints that GIS-agents need satisfy if they are to exhibit commonsense. We propose the following *slogan*:

> **Commonsensical GIS-agents must be able to implement nonmonotonic and fuzzy reasoning.**

We now briefly discuss why this *has* to be the case and then move on to focussing on the management of background ignorance.

4.1 Managing background ignorance: nonmonotonic reasoning

The discussion of section 3.2 lead us to the conclusion that if we are to model commonsensical GIS-agents, then we ought to account for the phenomenon of *background ignorance*. This must be the case since the proactiveness of marine environments implies that at no point in time, will the agent's knowledge base ever be complete.[16]

Classical logic is *monotonic* to the extent that the conclusions that can be drawn from a given set of premisses have to be valid at any time and in any place where *at least* such premisses hold. This means that nothing the agent might be in a position to learn afterwards can affect the (logical) status of such conclusions. Put very loosely (and somehow circularly) monotonic reasoning can never lead to the *revision* of an agent's knowledge base. Straightforwardly, this cannot be a desirable pattern of reasoning for GIS-agents.[17]

Therefore it is easily seen that abstract models of GIS with commonsense cannot be provided by means of monotonic logical systems. Different ways of getting rid of this undesirable form of reasoning have given rise to different formal approaches to *nonmonotonic reasoning*.

The first formal account of a type of nonmonotonic reasoning was given in Reiter's [42] where the so-called "close world assumption" is introduced.[18] According to it, the agent is allowed to draw conclusions *as if* it had complete knowledge of the world with the proviso that the agent must withdraw those very conclusions in the case it acquires new information contradicting it. This idea culminated in the formalization of *Default Logic* [43].

Since the first steps a tremendous number of logical systems rejecting (unconstrained) monotonicity has arisen (see [2] for a recent comprehensive survey and [7] for a computationally-oriented account of the main 'survivors'). Therefore we cannot properly speak of a single nonmonotonic logic, but we should consider it as a family of logics. The approach based on consequence relations, pursued by [13, 27, 24] makes rather transparent the reciprocal relations between such distinct systems, hence qualifying it as a suitable framework for investigating commonsensical constraints on GIS-agents.[19]

[16]Similarly, we have seen that the essentially layered nature of marine information, together with the particular phenomenon of temperature fronts, imply the context-dependency of agents' knowledge bases.

[17]Immunity to new information is only compatible with omniscient agents or absolutely stable environments, neither of which is compatible with the underlying notion of GIS-agents.

[18]It is perhaps more than a curious historical accident the fact that after Aristotle, the interest in qualitative nonmonotonic forms of reasoning was revived by issues in the theory of databases.

[19]Recent mathematical developments, like [11, 22, 18], show the inevitability of non-

Nonmonotonicity is also required for dealing with *property-inheritance* in inhomogeneous environments. As discussed above, context-depending information, as the result of say, temperature fronts, is likely to make the inheritance of some properties undesirable. The problem of blocking undesirable inheritance has been extensively treated in the early literature on nonmonotonic reasoning (see, e.g.[14]).

4.2 Managing vagueness: fuzzy reasoning

Space and time are obviously critical issues to be accounted for when modelling GIS-agents' reasoning. One of the difficulties they generate in modelling geographic reasoning is that they correspond to non-discrete entities.

Representing information about non-discrete (or equivalently, continuous) entities typically gives rise to *vague* statements. One example which has been studied in the GIS perspective is the *nearness relation* [52, 9]. It clearly follows that GIS-agents with commonsense must be able to deal with relations such as the nearness one. This is even more compelling if we agree with the programme of naïve geography, and we seek for an easier non-expert- human/GIS interaction.

In order to better appreciate the requirement of "fuzziness" for commonsensical GIS we need to recall that the main consequence of vagueness is the impossibility of defining truth values without generating paradoxes like for example, the most famous Sorites one. Put the other way round, the only way for determining the truth value of a vague statement (i.e. a statement involving a vague predicate or relation) is that of allowing (possibly infinitely many) *degrees of truth* in place of the standard binary values. Back in the 20's Lukasiewicz provided the first suggestions for devising logical systems capable of making sense of degrees of truth, or equivalently, *many-valued logics*.[20]

The topic of many-valued logics had a tremendous revival in the mid-Sixties essentially due to Zadeh's introduction of the notion of *fuzzy set*, in his seminal [54][21]. Zadeh's work has given rise to a fairly big number of formal approaches for modelling fuzziness the most well-studied among them being the domain of *possibility measures* (see, e.g. [8] for a survey).

monotonic reasoning in the characterization of rational reasoning under imperfect information.

[20]Notice that, pretty much for the same reasons given above for nonmonotonic logics, we speak of a family of many-valued logics, rather than a single many-valued logic. This essentially depends on the fact that there are slightly distinct ways in which the notion of degree of truth can be captured. See [1] for a discussion on that.

[21]The enthusiasm of the early days contributed to the rather rhapsodic development of the subject. See [17] for a brief yet compelling conceptual review, and [16, 1] for the advanced mathematical details with respect to the formalization of reasoning under vagueness.

4.3 Qualitative vs. quantitative approaches

When it comes to formal theories of reasoning under imperfect information, the debate arises about which side of the great divide better serves the purposes of the theory. We now briefly deal with the issue by addressing the following question: What sort of conceptual difference mirrors the distinction between qualitative and quantitative methods?

We first distinguish between problems which are essentially quantitative and those which are not. It is clear that if we are dealing with an *essentially* quantitative problem, no debate can reasonably arise: only quantitative solutions can be adequate. A case in point is given by *many-valued semantics*. It is clearly quantitative to the extent its intended goal –as remarked just above– is that of capturing the notion of "degrees of truth", which is clearly a genuinely quantitative issue.

A case in which the problem is not necessarily quantitative is that of characterizing agents' belief-formation, that is to say –in our present context– characterizing the kind of flexible and sensible reasoning we expect from commonsensical GIS-agents. The point is better illustrated by means of an example.

Let us consider Pearl's at al. $\varepsilon-$semantics. In a nutshell, ε-semantics uses infinitesimal probabilities to give meaning to default rules.[22] In other words, defaults like "θ typically implies ϕ" are taken to hold just if the probability of the corresponding conditional ("ϕ given θ") is infinitesimally close to 1. Now, despite using numbers[23], $\varepsilon-$semantics clearly captures a *qualitative* aspect of reasoning under imperfect information. This follows from the fact that the expressive power of $\varepsilon-$semantics does not go beyond what the agent takes to be "extremely probable". Put in other words, if we interpret $\varepsilon-$semantics as giving a formal meaning to agents' beliefs, then *no degrees of belief can be captured by $\varepsilon-$semantics.* This clearly contrasts with the case of standard Bayesian probability accounts, where there are infinitely (actually uncountably) many values a belief function can range over.[24]

Thus, from this point of view, there is a purely extensional difference between nonmonotonic reasoning and classical probabilistic reasoning and it lies in the expressive power of their characterization of belief. Again, this is easily seen by considering the fact that nonmonotonic reasoning can be given a probabilistic semantics provided that no degrees of probability values are allowed. This is nothing but an easy example showing how, in general, qualitative reasoning can be considered to be a very special case of

[22]See [37] for an introductory discussion of ε-semantics in the context of Default Logic.

[23]A particularly unfruitful aspect of the debate consists in taking the distinction qualitative vs. quantitative to be mirrored by the symbolic vs. numeric one.

[24]See [33] for the mathematical details on the Bayesian approach and [19] for the related epistemological issues.

the quantitative one.

Another remark on the expressive power of formal models is appropriate. In standard Bayesian models of reasoning under imperfect information, probability values are taken to be pointed. In other words the value of a probability (belief) function is an exact –precise– numeric value in the interval $[0, 1]$. Giving such a numerical value, however, would not be possible for realistic agents in a number of cases of interest. Those are the cases in which indeed information is *imprecise*.[25] This requirement of precision has being relaxed in the *imprecise probability model* fully described in [48] (but see also the more manageable [49]).[26] The generalization to imprecise probabilities allows the agent to manage –obviously enough– *imprecise information*; something neither fuzzy nor standard probabilistic quantitative models can do.

5 Methodology in action: Maximum entropy reasoning for GIS

Let's now put to work the methodological precepts discussed so far. In order to illustrate some crucial features related to the management of uncertainty in GIS, we'll be drawing on a case-study on animal behaviour concerning the *Hystrix Cristata* (crested porcupine) in the Natural Park of Maremma, Tuscany [4].

The latter provides a particularly representative example, especially with respect to the various aspects of imperfect information that feature in such a scenario. It goes without saying, however, that our main goal here is that of discerning, among the uncertainty-related phenomena of this case study, those general patterns of "uncertainty" that GIS in general will almost inevitably have to face.[27]

Our goal for the remainder of this paper is to show how *maximum entropy reasoning* qualifies as a methodologically sound and computationally viable solution for the management of background ignorance. The other crucial ingredient of imperfect information management in GIS –fuzzy reasoning– will be addressed in subsequent work.

[25]The difference between imprecision and vagueness might be recalled intuitively by pointing out that whereas the former is an attribute of degrees of belief, the latter is an attribute of degrees of truth.

[26]Within the REV!GIS project, the imprecise probability model is discussed in relation with GIS in the Deliverable N. R211 covering 1 June 2000–15 May 2002.

[27]As hinted above, *marine* GIS pose, as far as uncertainty management is concerned, more complicated and hence even more interesting theoretical problems. This, however, must be a topic for subsequent investigations.

5.1 Thorny queries

So let's briefly recall the scenario from the case study (precise details can be found in [4]). Animals are studied from collected data about them. Data collection is performed by means of a radio tracking techniques, where animals are provided with a radio collar and their localization is *fixed* at a given time interval by means of a technique called bearing, where radio signal emitted from fixed stations match the signal coming from the radio collar. Each localization is called a *fix* and is of the form fix(Id, X,Y, T), where Id is the identifier of the tracked animal, X,Y are the spatial coordinates and T is the time of the bearing.

What behavioural ecologists expect to gain from the analysis of tracked animals is essentially an accurate account of certain observed behaviours of the porcupine *in relation* to its environment. As an example of such kind of analysis, the study of the behaviour of the animal with respect to its den play a crucial role in the overall understanding of the porcupine characteristics. Typically, scientists will be interested in understanding what (if any) are the preferences individuals have concerning the location of the den. Or determine how much time is being spent in the den by the animal. Other questions concerning the mating system can benefit from the knowledge of individuals that share the same den. The census of the den can be done by scientists by a technique called *homing-in*, that is to say they build a database of the dens after their physical location has been ascertained (this is possible since individuals are equipped with appropriate radio-collars). Homing-in technique provides a localization of den almost free of errors (just the minimal error caused by instruments or possible human mistakes) but it is a very expensive procedure in terms of human effort. Therefore it cannot be done more often than bi-monthly. Since animals typically change theirs dens over time, there are some kind of gaps where the location of the den of animals is unknown. It can easily be seen that, by the above discussion on the various sources of imperfection in geographical information, many of the relevant parameters will not have –in principle– crisp or otherwise "perfect" geo-referencing. Moreover, and this is the crucial element in this example, GIS designed for this sort of tasks need to be capable of handling what is called *background ignorance*. Considering the problem of dens localization, every inference relating certain particular individuals with their own dens is necessarily being performed under ignorance of the possible changes that might have taken place since the last physical measurement. The inevitability of this type of imperfection in any reasonably complicated environment justifies the terminology *background ignorance*. On a more general level, this is perhaps the most fundamental issue in the uncertainty management for GIS. Every "dynamic system" (whatever the actual realization, be it a society of porcupines in their environment or a fragment of coastal territory) in fact, put *agents reasoning about the system* itself in a condition of background ignorance. GIS seen as agents, hence, must be capable of inferring as much information as possible from the available (geo-referenced) observation *taking into account* the fact that they are referred to a dynamic system. This facet of imperfect information, which can be seen as a special case of incompleteness, clearly adds to the ones relating to the granularity and possible inconsistency of the geo-referenced data discussed above. Let us take a short step back and point out that the formalization of geographic uncertain reasoning requires us to distinguish between two levels which, by and large, correspond to the familiar ones of "object-" and "meta-" level. On the object-level, geographic information must be formalized by means of a language capable of representing the spatio-temporal attributes of the observed data. The product of this formalization leads to what is usually called geo-referenced data. On the meta-level geo-referenced data is used to yield geographical reasoning. The intrinsic imperfection of geo-referenced information discussed above, however, requires the formal reasoning at the meta-level to be constrained by adequate principles of reasoning under uncertainty, and in particular under background ignorance. Since it is by far less understood, the main focus of this paper is on formalization of the meta-level. We are now in a position to see how the phenomenon of background ignorance comes into the foreground in the porcupine case study and related GIS. One non-trivial problem that behavioural ecologists ask GIS science to help solving consists in the localization of the dens. The authors of [4] translate this into the following

> **Query** Given a number of known dens positions (collected by homing-in) and animal fixes, infer the position of the dens in periods of time when no information is available.

If we abstract from the nature of the parameters occurring in a successful dens localization, we can see that the above query embeds a very general pattern of what is normally called "inductive inference".[28] Put crudely, inductive inference relies on the data possessed by the agent in order to draw principled conclusions about as yet unobserved phenomena. And this pattern of reasoning is where the agent perspective on GIS introduced above becomes essential. In such a framework we can in fact understand the above query as the following

> **Main Query** Given that the GIS-agent possesses information of a number of dens positions and the animal fixes, determine the probability that the dens $d_1, \ldots, d_n$ will be in the locations $\delta_i, \ldots, \delta_n$[29] in periods when no information is available.

[28]Notice that by no means we are opposing "deduction" to "induction", here.

[29]Here, clearly each δ_i stands for the ordered pair $\langle x_i, y_i \rangle$ on a bi-dimensional of space.

Notice that the probabilistic nature of the query is extremely natural both as a direct representation of the intrinsic uncertainty of the GIS-agent answer and in relation to the general information provided by domain experts. Solving queries of this sort in fact, among other things, involves deploying domain experts specific knowledge (in this case concerning the porcupine's "den behaviour") to construct a reliable hypothesis that can accommodate the possible changes occurring within the dynamic system. But domain specific knowledge cannot be taken to be "perfect" or otherwise "certain" knowledge. Rather it is best represented in probabilistic form. So, for example, the conjecture that

"the main activity of porcupines during the time spent outside the day is feeding" [4]

can be informatively represented by

$$\omega(feeding \mid outside\_den) = 0.8, \tag{1}$$

where ω is a probability function (the precise details of the representation are discussed below). Within this framework it is reasonable to require that an adequate solution to our query should provide "the most accurate" probability evaluation *consistent* both with the information possessed by the GIS-agent and with the domain expert general hypotheses. The crux of this requirement is clearly to characterize formally what we mean by "the most accurate" evaluation. A natural way of doing so consists in requiring that the construction of the probability assignment (i.e. the solution to the probabilistic query) should, while satisfying the consistency requirements, *introduce as few arbitrary information as possible*, that is to say information that goes beyond the collected data and the domain-expert evaluations. From a mathematical (information-theoretical) point of view this requirement amounts to *maximizing the entropy* of the solution. The remaining part of this work is aimed at introducing and discussing the relevance of maximum entropy inference with respect to the management of background inference as well as the computational and implementational aspects of the resulting inferential system.

6 Maximum entropy inference

We start by recalling briefly some of the main motivations[30] underlying maximum entropy inference, with particular reference to our underlying problem.

[30] There is an extremely rich literature on maximum entropy in uncertain reasoning. An accessible introduction for the non-specialist is [34] whereas for more presentations [33, 36, 22]. The latter works provide also extensive references to the relevant literature.

We have already remarked that a feature of imperfection in geo-referenced information that we should constantly expect, is related to the periodicity of relevant measurements. Although this feature arises essentially from pragmatic concerns like, e.g. the cost of obtaining and recording relevant data, it must be stressed that there is virtually no way around this source of imperfection of information, no matter how efficient our measurement and recording methods are. For this reason we have suggested the terminology "background ignorance".

The most peculiar feature of background ignorance consists simply in the fact that independently of the accuracy of our previous information, the available data cannot be taken to be either complete or certain. Hence, statements which are conditional on such data cannot be expected to hold with no exception. Maximum entropy inference not only allows us to define meaningful notions of logical consequence for dealing with background ignorance; it also admits of extremely robust methodological justifications which are tied-up by principles of rational reasoning.

The latter is clearly a fundamental requirement in a framework in which GIS as taken as "reasoners", that is to say, agents. We have seen, in fact, that talk about the management of uncertainty implies considering Geographic Information Systems as agents which, among others, have the capability of manipulating geo-referenced and then of solving queries *solely* on the basis of such information.[31] Construing GISs in this way allows us to discuss them from an abstract, logical perspective. So we consider a GIS as an agent which, as far as its reasoning capabilities are concerned, is formally characterized by means of an appropriate *inference process*. The particular instance of inference process, discussed below in section 6.4, is given by probabilistic logic programs under maximum entropy.

6.1 Knowledge representation

Knowledge[32] is represented probabilistically in the framework under discussion. More precisely we consider agents' knowledge to be represented by means of a *finite set of (linear) constraints on agents' subjective degrees of probability*. It is important to stress the subjective component to the extent that we are discussing here of GIS as agents whose task consists in the manipulation of geo-referenced data. Knowledge being encoded by subjective degrees of probability means just that agents' answers are expected to be based *only* the information available to the agent itself, with the explicit

[31] In doing so we implicitly separate the problem of modelling *reasoning* from the problem of *acquiring* information. Although the latter is of fundamental importance in GIS science, it goes beyond the scope of this work.

[32] Since our main focus is on GIS-agents we presume no distinction here between knowledge, information, data, the preference of one expression over the others being justified only on the grounds of intuitive understanding.

assumption that is *all* the knowledge possessed by the agent.[33]

In order to see how this could be captured and hence applied for the solution of our Main Query, we need a bit of formal setting-up which we take from [33].

6.2 The ME inference process

Let $\mathcal{S}L$ be the set of sentences (denoted by θ, ϕ, etc.) of a finite propositional language $\mathcal{L} = \{p_1, \ldots, p_n\}$ and let G be our GIS-agent. Then we let K_G, or simply K (if we are interested in just one GIS-agent) denote the probabilistic knowledge possessed by G at a fixed point in time. We can conveniently think of K as a finite set of conditional constraints of the form:

$$\omega(\theta \mid \phi) = x, \tag{2}$$

where ω is a probability function[34] on $\mathcal{S}L$, and $x \in [0,1]$. Note that unconditional constraints are obtained by taking the conditioning sentence (ϕ) to be any tautology. Moreover we assume, mainly for the sake of simplicity, that $\omega(\phi) > 0$.

An *Inference Process* on $\mathcal{L}$ then, is defined as a function N such that, for K a consistent finite set of linear constraints of the form $\omega(\theta_j \mid \phi_j) = x_i$, with $x_i \in [0,1]$, $N(K)$ is a probability function ω on $\mathcal{S}L$ satisfying K.

We can see how the definition of an inference process translates formally the consistency requirements pointed put above for the solution of the Main Query. It should be noted that, from the representational point of view, there is no qualitative difference between the data possessed by the agent and the domain-specific knowledge supplied by the expert (in the porcupine case the behavioural ecologist), but only the quantitative difference that arises from the (subjective) degrees of belief. This is why it makes perfect sense to consider both sources of knowledge as defining the constraints against which the consistency of the solution should be checked.

Given a finite consistent set of constraints K however, there will be many (indeed infinitely-many) *formally consistent* ways of extending the information possessed by a GIS-agent to the query at hand. What is needed then, is a principled way of discarding those solutions that though logically consistent, would nonetheless turn out to be "non-optimal". This is obtained by

[33]In the mathematical literature this entirely straightforward –yet sometimes overlooked– principle is so fundamental that it has been given a name of its own: the *Watts Assumption* [33].

[34]Recall that a function $\omega : \mathcal{S}L \to [0,1]$ is a *probability function* if $\omega(\theta) \geq 0 \ \forall \theta \in \mathcal{S}L$ and the following are satisfied:

$$\text{If } \models \theta \text{ then } \omega(\theta) = 1 \tag{\textbf{P1}}$$

$$\text{If } \not\models (\theta \wedge \phi) \text{ then } \omega(\theta \vee \phi) = \omega(\theta) + \omega(\phi). \tag{\textbf{P2}}$$

identifying the solution to our query with the *Maximum Entropy Inference Process* which, for K as above, is that *unique* probability function $N^{(ME)}(K)$ consistent with K and for which the Shannon-entropy

$$-\sum_{i=1}^{J} x_i \log x_i \tag{3}$$

is maximal.[35]

Shannon-entropy [47] is a measure of the "uncertainty of information" so that it maximization leads to an inference process that introduces as few arbitrary information as possible, where arbitrary loosely means "departing from what is given in K".[36] We can now see how the maximum entropy inference process captures formally the two intuitive constraints on the solution of our Main Query above.

There are many logical-mathematical realizations of maximum entropy inference processes. A particularly interesting one being, as far as our Main Query is concerned, the framework of probabilistic logic programming under maximum entropy recently investigated by G. Kern-Discerner and T. Lukasiewicz [23]. Although we are only interested here in laying down the main conceptual issues concerning the relevance of the maximum entropy inference process to the management of imperfect information in GIS science, we will say something about the consequence relations defined by such a logic programming-based approach below. In light of the logic programming-based formulation of the maximum entropy inference process, we will sometimes refer to the finite set of consistent linear constraints K as a consistent probabilistic logic program.[37]

Before doing so, however, let's focus on the key properties satisfied by $N^{(ME)}(K)$.

6.3 Properties of ME inference processes

The crucial result in this area is that $N^{(ME)}(K)$ is the only inference process which satisfies (along with the consistency requirements introduced above) a number of rationality principles on the determination of subjective degrees of probability.[38] We discuss here two of them which happen to be directly related to our main problem, namely *Irrelevance* and *Obstinacy*.

[35]Notice that we identifying here the probabilistic constraints $i \in J$ with their values. See, e.g., chapter 2 of [33] for precise details on the representation of probability functions. Notice also that by the usual convention, $\log 0 = 0$.

[36]We refer the interested reader to [33, 21] for the mathematical analysis of Shannon's measure of uncertainty

[37]Precise details can be found in [26] and [23].

[38]See chapter 7 of [33] and [34] for the case in which K is a finite set of *linear* constraints. [36] extends the discussion to the case of *non-linear* constraints.

The intuitive idea underlying the former can be described by saying that each query determines a "range of relevance" in such a way that information outside this range should not take part to the solution of the query itself. Suppose, for instance, that the GIS-agent is facing a query concerning the localization of one particular den. It would be *irrational* for the agent to build the solution to such a query on top of its information regarding say the proportion of male researchers in the Italian Research Council. In other words, a solution based on the latter would be considered to be fallacious (i.e. unjustified) on the grounds of its irrelevance. Moreover, if we allowed GIS-agents to work out solutions to queries irrespectively of the relevance of the constraints being used, the computational task of solving the query will almost inevitably suffer an unmotivated increase in complexity.

Characterizing what (ir)relevance is in all its subtleties corresponds, to a large extent, to resolving one of the hardest problems in the formalization of intelligent behaviour.[39] The maximum entropy inference processes however, satisfies a very natural formalization of relevance for finite sets of consistent linear constraints. It is formulated as follows:

> If K_1 and K_2 are finite consistent sets of linear constraints in $\mathcal{L}$, $\theta \in \mathcal{S}L$ but no propositional variable appearing in θ or in K_1 appears also in K_2, then
>
> $$N(K_1 + K_2)(\theta) = N(K_1)(\theta).$$

Data acquisition is, under many respects, an expensive business. An important consequence of this, as we have recalled above, is the fact that the actual physical dens localization cannot be performed more frequently than once every two months. This is clearly to be generalized to the whole domain of GIS. It is therefore of the greatest importance that the process of *revising* probability assignments should be constrained in such a way as to preserve as much information as possible, that is to say by *avoiding* unnecessary revisions. This is the idea underlying the principle of Obstinacy which is captured formally as follows:

> If K_1 and K_2 are finite consistent sets of linear constraints in $\mathcal{L}$, $N(K_1)$ satisfies K_2, then
>
> $$N(K_1 + K_2) = N(K_1).$$

Obstinacy is a particular case of *constrained monotonicity*, that is to say is a principle that constraints the enlargement of the set of premises from which a conclusion has been previously drawn. So, for example, suppose that for a particular den, the GIS-agent provides a solution to the Main Query.

[39] The issue is clearly most debated in logic-based AI. See, for instance [15, 38, 30].

Such a solution will have the form of a probability assignment (consistent with K) to the (relevant) sentence of $\mathcal{S}L$. Now, if we expand K with a new constraint that is *already satisfied* by the solution to our Main Query, then the solution for the enlarged set of constraints should not be a revision of the one previously obtained.

The intuition captured by Obstinacy is clearly deeply related with *non-monotonic* reasoning. Indeed, as shown by [23], all the key properties of non-monotonic consequence relations are satisfied in their framework for probabilistic logic programming under maximum entropy.[40]

6.4 ME logic programming

The powerful combination of ME reasoning with logic programming results in a theoretically very well-behaved framework for reasoning under (various forms of) uncertainty and in particular for reasoning under background ignorance. Probabilistic logic programs under maximum entropy then, bridge between the theoretical results concerning maximum entropy inference processes and the corresponding computational logic (the full formal details can be found in [26, 23]).

We just stress here that section 5 of [23] discusses in depth the relation between probabilistic logic programming under maximum entropy and the normative properties of nonmonotonic consequence relations. It is shown that, among other properties the former satisfy the rules/conditions for the theory of Rational Consequence Relations which (possibly in its restriction to the theory of Preferential Consequence Relations) is regarded as capturing the "core aspects" of nonmonotonic reasoning (see, e.g. [24, 25, 22]).[41]

Notice that in framework investigated by Kern-Isberner and Lukasiewicz, probabilistic logic programs are (finite) collections of clauses of the form

$$\omega(\theta \mid \phi) = [l, u],$$

where $l, u \in [0, 1]$ stand for the *l*ower and *u*pper bounds of the probability assignment. A *practically* important consequence of this for GIS science is that domain experts are allowed to provide interval-valued rather than pointed-valued probability assessments.

7 Implementing maximum entropy inference

Despite the numerous arguments in favour of reasoning under maximum entropy, there is still a theoretical issue that might discourage AI practi-

[40] For another aspect of the connection between non-monotonic reasoning and maximum entropy see [18].

[41] It is worth recalling that the theory of Rational Consequence Relations is both formally and conceptually closely related to the standard AGM paradigm for belief revision. See [45] for an extensive discussion on this.

tioners.[42] Unsurprisingly, this is computational complexity. As proven in chapter 10 of [33], in fact, the following problems are in general infeasible:

i. Checking the satisfiability of a set of linear constraints K;

ii. Given a satisfiable set of constraints K and $\theta \in \mathcal{S}L$ computing an approximation for $\omega(\theta)$ consistent with K.

Recent developments on the computational techniques for maximum entropy reasoning, however, provide good evidence for the claim that –in many *practical* circumstances– ME reasoning is indeed feasible.

The expert system shell SPIRIT[43], for instance, provides an efficient computational engine for the solution of such problems. The key factor on which SPIRIT relies in order to optimize the complexity of ME reasoning is the construction of a *dependency-graph* for the problem at hand. Roughly,[44] this involves firstly, introducing a (distinct) vertex for each (distinct) propositional variable occurring in the probabilistic logic program under consideration and secondly, connecting any two vertices such that the corresponding propositional variables appear in the same constraint in K. The dependency-graph is then used to the effect that the actual probability evaluation is performed on it rather than on the set of all possible atoms.[45] Once the constraints have been met, the maximum entropy distribution is immediately propagated to the queries, providing the required solution.[46]

8 Conclusions and future research

This paper suggests a new framework for GIS research. The key methodological precept consists in viewing GIS as agents who possess their own knowledge (independently of how information is acquired) and are expected (say by humans or other agents) to make the most out of it. This is the only scenario in which talk of *uncertainty* in GIS can make full sense.

Managing imperfect information, we have suggested, requires a robust methodological framework. We propose here coupling GIS research with some prominent logical-mathematical theories of reasoning under uncertainty. In our framework this is the key part of a *divide et impera* plan: We have distinguished between agent-dependent and agent-independent sources of imperfect information and then illustrated how the latter can only be accounted for by a logical-mathematical framework. The vastity of the domain of imperfect information, however, requires us to refine our framework

[42]See [35] for a survey of the common criticisms raised against maximum entropy reasoning.

[43]Available at `http://www.fernuni-hagen.de/BWLOR/spirithome.html`

[44]See [31] and the references therein for precise details.

[45]If the constraints are all logically independent, then the probability evaluation must be performed on all the 2^n possible atoms, where n is the cardinality of $\mathcal{L}$.

[46]See [31, 44] for precise details on SPIRIT.

further. To this effect we propose distinguishing between the phenomena of background ignorance and granularity of information.

After setting-up the conceptual framework we've tackled the problem of constructing a logical-mathematical bridge between the methodological requirements on GIS-agents and a computational realization of them. We have argued for the appropriateness of maximum entropy logic programming to provides such a link. Maximum entropy inference processes lie at the heart of uncertain reasoning. Computational logic connects it very naturally with Belief Revision. It is deeply connected with the formalization of Common-sense by means of methodological normative principles for belief formation and more generally "inductive inference". This clearly has a bearing on the *naïve geography* programme (especially with respect to providing "commonsensical interaction" with GISs). Inconsistency handling is another important feature that is addressed by means of the computational treatment of maximum entropy reasoning, as provided by the shell SPIRIT. A traditionally hard problem for information systems and databases in general, in fact, is given by the spectre of inconsistency. Under classical reasoning even a single inconsistency is enough to cause the trivialization of the *entire* information possessed. Potential inconsistency can be measured by the maxent computational shell SPIRIT.

This study is part of a much larger project. Although maximum entropy logic programming can be extended to handle probability intervals, its relations with many-valued and fuzzy reasoning is still very much *terra icognita*. [47] Further research is then to be focussed on the formal investigation of the adaptation of maximum entropy logic programming to geo-referenced data and in particular on its relation with logic programming paradigms like MuTACLP [28].

References

[1] A.D.C. Bennett, J.B. Paris, and A. Vencovská. A new criterion for comparing fuzzy logics for uncertain reasoning. *Journal of Logic, Language, and Information*, 9(1):31–63, 2000.

[2] A. Bochman. *A Logical Theory of Nonmonotonic Inference and Belief Change*. Springer, 2001.

[3] P. A. Burrough. *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press, Oxford, 1986.

[4] T. Ceccarelli, D. Centeno, F. Giannotti, A. Massolo, C. Parent, A. Raffaetà, C. Renso, S. Spaccapietra, and F. Turini. The behaviour of the

[47] Some very recent studies seem, however, to point to this direction. See, eg.[53].

Crested Porcupine: the complete case study. Technical report, DeduGIS - European Working Group, 2001.

[5] N. Chrisman. *Exploring Geographic Information Systems.* John Wiley and Sons, 1997.

[6] K.C. Clarke. *Analytical and Computer Cartography.* Prentice-Hall, Englewood Cliffs, NJ., second edition, 1995.

[7] J Dix, U. Furbach, and I Niemelä. Nonmonotonic Reasoning: Towards Efficient Calculi and Implementations. In Andrei Voronkov and Alan Robinson, editors, *Handbook of Automated Reasoning, Volume 2, Chapter 18*, pages 1121–1234. Elsevier-Science-Press, 2001.

[8] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In Gabbay and Smets [12], pages 169–226.

[9] M. Duckham, K. Mason, J. Stell, and M. Worboys. A formal ontological approach to imperfection in geographic information. *GISRUK 2000*, 2000.

[10] M. J. Egenhofer and D. M. Mark. Naive geography. In *COSIT 95*, volume 988 of *Lecture Notes in Computer Science*, pages 1–15, 1995.

[11] N. Friedman and J. Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685, 2001.

[12] D. M. Gabbay and Ph. Smets, editors. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1. Kluwer, Doordrecht, The Netherlands, 1998.

[13] D.M. Gabbay. *Theoretical foundations for non-monotonic reasoning in expert systems*, pages 439–457. Proceedings NATO Advanced Institute on Logics and Models of Cuncurrent Systems. Springer, Berlin, 1985.

[14] D.M. Gabbay, C. Hogger, and J. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3. Oxford University Press, G.B, 1994.

[15] D.M. Gabbay and J. Woods. *Agenda Relevance: A Study in Formal Pragmatics.* North-Holland, 2003.

[16] P. Hájek. *Metamathematics of Fuzzy Logic.* Kluwer Academic Publishers, Dordrecht, 1998.

[17] P. Hájek. Ten questions and one problem on fuzzy logic. *ANNALSPAL: Annals of Pure and Applied Logic*, 96, 1999.

[18] L.C. Hill and J.B. Paris. When maximising entropy gives the rational closure. *Journal of Logic and Computation*, 12(1):1–19, 2002.

[19] C. Howson. The Baysian approach. In Gabbay and Smets [12], pages 112–134.

[20] M. N. Huhns and M. P. Singh. *Readings in Agents*. Morgan Kaufman Publishers Inc., 1998.

[21] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge, 2003.

[22] G. Kern-Isberner. *Conditionals in Nonmonotonic Reasoning and Belief Revision - Considering Conditionals as Agents*, volume 2087 of *Lecture Notes in Computer Science*. Springer, 2001.

[23] G. Kern-Isberner and T. Lukasiewicz. Combining probabilistic logic programming with the power of maximum entropy. *Technical Report Nr. 1843-02-12, Institut für Informationssysteme, Technische Universität Wien*, 2002. Available on line at.

[24] S. Kraus, D Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative inference. *Artificial Intelligence*, 44(1):167–207, 1990.

[25] D. Lehmann and M Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[26] T. Lukasiewicz and G. Kern-Isberner. Probabilistic logic programming under maximum entropy. In *Proceedings of the 5th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, London, UK*, volume 1638 of *Lecture Notes in Computer Science*, pages 279–292. Springer, 1999. Available on line at.

[27] D. Makinson. General patterns in nonmonotonic reasoning. In Gabbay et al. [14], pages 35–110.

[28] P. Mancarella, A. Raffaetá, C. Renso, and F. Turini. Integrating knowledge representation and reasoning in geographical information systems. *International Journal of GIS*. to appear.

[29] M.V. Masserotti, A. Montomoli, and E. Zucchelli. Trattamento dell'informazione geografica in ambiente marino. *Rapporto interno CNUCE, C95-43*, 1995.

[30] J. McCarthy. From here to human-level AI. In L. Carlucci Aiello, J. Doyle, and S Shapiro, editors, *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning*, pages 640–646, San Francisco, 1996. Morgan Kaufmann.

[31] C. Meyer and W. Rödder. Coherent knowledge processing at maximum entropy by SPIRIT. In *Proceedings of the 12th conference on Uncertainty in Artificial Intelligence, Portland, Oregon, USA*, pages 470–476, 1996.

[32] N. Nilsson. *Artificial Intelligence: A new syntesis*. Morgan Kaufmann Publishers, San Francisco, 1998.

[33] J.B. Paris. *The Uncertain Reasoner's Companion: A Mathematical Perspective*. Cambridge University Press, Cambridge, England, 1994.

[34] J.B Paris. Comon sense and maximum entropy. *Synthese*, 117(1):73–93, 1999.

[35] J.B. Paris and A. Vencovská. In defence of the maximum entropy inference process. *International Journal of Approximate Reasoning*, 17:77–103, 1997.

[36] J.B. Paris and A. Vencovská. Common sense and stochastic independence. In D. Corfield and J. Williamson, editors, *Foundations of Bayesianism*, pages 203–240. Kluwer Academic Press, 2001.

[37] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, Los Altos, California, 1988.

[38] D.A. Plaisted and A. Yahya. A relevance restriction strategy for automated deduction. *Artificial Intelligence*, 144:59, 2003.

[39] D. Poole, A. Mackworth, and R. Goebel. *Computational Intelligence*. Oxford University Press, Oxford, 1998.

[40] I. Pratt-Hartmann. *Artificial Intelligence*. Macmillan, London, 1994.

[41] A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, 1991.

[42] R. Reiter. On closed world databases. In H. Gaillaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, New York, 1978.

[43] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1–2):81–132, 1980.

[44] W. Rödder. On the measurability of knowledge acquisition and query processing. *International Journal of Approximate Reasoning*, 33(2):203–218, 2003.

[45] H. Rott. *Change, Choice and Inference : A Study of Belief Revision and Nonmonotonic Reasoning.* Oxford University Press, 2001.

[46] S. Russell and P. Norvig. *Artificial Intelligence: A modern approach.* Prentice-Hall, 1995.

[47] C.E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois, 1964.

[48] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[49] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[50] M. Wooldridge. *Reasoning about Rational Agents.* Intelligent Robots and Autonomous Agents. MIT Press, Cambridge, Massachusetts, 2000.

[51] M. Wooldridge. *An Introduction to Multiagent Systems.* John Wiley & Sons, Chichester, England, 2002.

[52] M.F Worboys. Nearness relations in environmental space. *International Journal of Geographic Information*, 2001.

[53] L. Zadeh. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *Journal of Statistical Planning and Inference*, 105:233–264, 2002.

[54] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.