Ontologies Contribution to link thematic and remote sensing knowledge: preliminary discussions

Samuel Andrés [1]
Damien Arvor [1]
Marie-Angélique Laporte [1]
Laurent Durieux [1]
Thérèse Libourel [1]
Isabelle Mougenot [1]
Christelle Pierkot [1]

[1] UMR 228 ESPACE-DEV
500 rue Jean-François Breton, 34093 Montpellier, France
{firstname.lastname}@ird.fr

Abstract. The semantic gap between image data and expert knowledge has been well identified for several years. Knowledge representation techniques such as ontologies are expected to help reducing this gap. Such techniques have already been used in different fields of application, from medical images to landscape pictures. We think they also have a great potential in remote sensing applications due to the interdisciplinary expert knowledge necessary for image interpretation. In this paper, we introduce what are ontologies and we discuss a potential ontological architecture for remote sensing applications. This architecture involves top-domain ontologies (OBOE, SWEET) and domain ontologies (land cover, image, spatio-temporal relations, protocol, sensor). We illustrate our thoughts with an example of Landsat TM image interpretation for detecting beaches.

Keywords: **R**emote Sensing, Image Processing, Ontology, Expert knowledge

1. Introduction

Global change is a major issue to be faced by today's societies requiring interdisciplinary approaches involving thematic knowledge from diverse scientific domains such as geography, ecology, biology, etc. However, large part of this knowledge is focused on a specific scientific area and consequently is hardly shared with other domains. We argue that environmental sciences would be largely improved if heterogeneous expert domain-knowledge could be unified on the basis of shared core concepts.

Earth Observation (EO) data has a major role to play for supporting interdisciplinary environmental researches. The development of enhanced remote sensors with increasingly high spatial, temporal and spectral resolutions allows the access to more details in satellite images, which has broadened the spectrum of remote sensing applications (see the nine Societal Benefit Areas defined by the Group on Earth Observation - GEO). Many scientific domains are now concerned by the use of EO data. As a consequence there is a need for improving the ability of researchers from different areas to process and interpret remote sensing images according to their specific purposes. One step toward such goal has consisted in implementing new image processing techniques allowing users to include their expert knowledge in the interpretation process. Especially, Geographic Object-Based Image Analysis (GEOBIA) represents a paradigm shift for image interpretation. Its main interest lies in its capacity for considering semantics based on a descriptive assessment and knowledge, i.e. it incorporates the wisdom of the user (Blaschke and Strobl, 2001; Hay and Castilla, 2008).

However, although such approach has led to tremendous improvements in the interpretation of remote sensing images, it highlighted some important methodological issues: 1) each GEOBIA expert has its own conceptualization of the reality he intends to represent on the image; 2) the image processing is performed in a very laborious way at the end of repeated cycles of “trial-and-error” analysis; and 3) the output products and the implemented methods are difficult to communicate and share with other scientists.

Summarizing, large parts of the knowledge involved in any geographic object-based image analysis is coming from specific scientific areas and consequently is hardly shared with other domains. To overcome these issues, we argue that it is necessary to formalize the knowledge involved in any geographic object-based image analysis. For this purpose, ontologies could play a pivotal role due to their ability to represent the knowledge of any domain based on the identification of key concepts and their inter-relationships. Furthermore, ontologies would then allow linking various domain knowledge representations. In this paper, we discuss a theoretical ontological architecture that could be used in support of interdisciplinary Earth Observation studies. We then illustrate our thoughts with a concrete application.

2. Ontologies for conceptualization and knowledge representation

2.1 Definition

The definition of the term “ontology” has been discussed in many previous papers (the reader is invited to refer to Mark et al. (2005), Agarwal (2005), Couclelis (2010) for a complete review on the definition of ontologies for GIS applications or to Madin et al (2008) for ecological applications). A widely agreed definition is the one proposed by (Gruber, 1993) who defines an ontology as a *formal, explicit specification of a shared conceptualization*.

The “conceptualization” is an abstract, simplified view of the world that we wish to represent for some purpose. It intends to identify the concepts and their relationships within a scientific domain. Here, “explicit” refers to the fact that concepts and constraints are explicitly defined. “Shared” specifies that the ontology captures a consensual knowledge, but this is not a consensual view. Finally, “formal” means that the ontology should be machine understandable. Thus, the ontology was implemented in dedicated software (e.g Protégé: protege.stanford.edu). Concepts were then defined through axioms constructed based on Description Logics that allow inferring new knowledge thanks to inference engines called *reasoners*.

2.2 Methodology to build ontologies

Since ontologies are used for creating, aggregating and sharing knowledge, any potential user should be able to extend existing ontologies in order to include its own knowledge. Pinto et al. (2004) reviewed several methodologies (TOVE, Methontologie) for building ontologies from terminologies. Indeed, a terminology is very useful to identify key-concepts of a given domain. These methods also recommend reusing already existing ontologies formalizing the relations between concepts. Such existing ontologies are often structured in various more or less abstract ontological levels, including top-domain and domain ontologies. Describing these ontological levels allows us identifying first requirements for the future construction of an GEOBIA ontology.

3. Ontological foundations for GEOBIA: an overview

3.1 Top-domain ontologies

As a sub-discipline of remote sensing science, an ontology representing GEOBIA concepts should be supported by concepts referring to Earth Observation. We then propose to use already existing top-domain ontologies about Earth and Observation.

3.1.1 Observation ontology: OBOE

OBOE (Extensible Observation Ontology) (Madin et al., 2007; Madin et al. 2008) is a top-domain ontology allowing to represent scientific observation and measurements (Figure 1). An *observation* is about an *entity* (abstract concept representing anything which can be observed) and can be used to aggregate different measurements in a single observation event. A *measurement* associates a *value* to the corresponding *characteristic* of the observed *entity* (for example, the height of a tree). A *value* itself is considered as an *entity* (*primitive values* are a kind of *entity* concept). Measurements are associated to measurements standards (units) and are potentially enriched by other information linked to the measurement process, method, protocol or accuracy.

An observation may have a *context* linked to other observations. Observations linked by a context are interdependent. For instance, let's consider an observation about the measure of the height of tree. This observation is carried out in a specific context, such as in a specific geographical place. Geographical place is important to interpret the measure of height of the given tree, but also to compare this measurement to other tree measurements around. All of this is made possible thanks to existence of context relation between the observations and a localization.

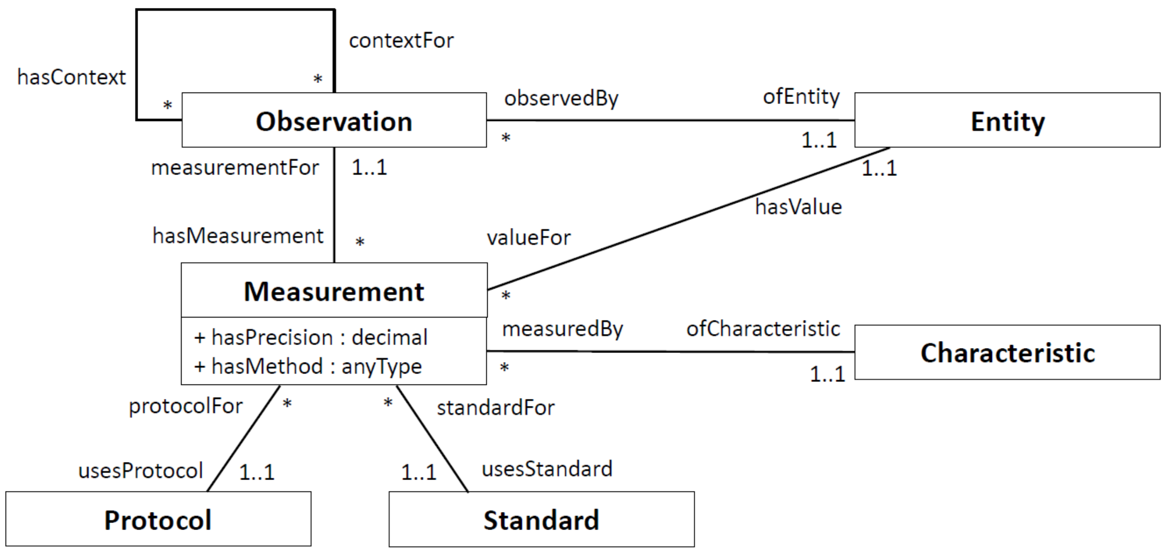

Figure 1: Core model describing the OBOE ontology.

3.1.2 Earth ontology: SWEET

As regards to Earth Environment framework ontology, the Semantic Web for Earth and Environmental Terminology (SWEET) proposed by NASA seems appropriate (Raskin and Pan, 2005) (http://sweet.jpl.nasa.gov/). This ontology is freely released and is composed of nine top-level ontologies (fig. 2) written in the OWL language and containing around 6000 concepts. The nine ontologies are organized in "faceted" and "integrative" ontologies, including orthogonal concepts and integrative science knowledge concepts, respectively (Madin et al., 2008). The "faceted" ontologies refer to **Representation**, **Process**, **Matter**, **Realm**, **Property**, **State** and **Relation** while "integrative" ontologies refer to **Phenomena** and **Human Activities**.

These ontologies are highly modular since, potentially, any compound concept related to environment can be described based on concepts introduced in orthogonal ontologies. As an example, the compound concept *air temperature* can be described by the physical property *temperature* (coming from the ontology on Properties), which applies to the substance *air* (coming from the ontology on Matter) (Tripathi and Babaie, 2008). Then, it is expected that complex compound concepts of interest for domain applications (e.g. land cover mapping) can be described by faceted concepts selected in the SWEET "faceted" ontologies.

The strength of OBOE and SWEET ontologies lies in their extension capabilities that can be used to connect various ontologies. Thus, OBOE and SWEET can be interconnected and can be used to develop relevant domain and task ontologies.

3.2 Domain ontologies

Domain ontologies describe the vocabulary related to a specific domain (e.g. ecology, geography, remote sensing) by specializing the terms introduced in the top-domain ontologies. Based on the OBOE and SWEET ontologies, it is possible to identify main domain ontologies that have to be built for remote sensing applications. We identified five domain ontologies.

3.2.1 Land cover ontology

Important efforts for building land cover ontologies have already been carried out. The HarmonIsa project (http://harmonisa.uni-klu.ac.at/content/land-use-land-cover-ontologies) listed main land cover ontologies including a Corine ontology. Currently, the Land Cover Meta Language (LCML), also named Land Cover Classification System (LCCS), produced by the Food and Agriculture Organization (FAO) is another attempt to build a land cover ontology. In this ontology, groups of characteristics are arranged in different ways that act as building blocks to describe more complex land cover classes (http://www.glcn.org/ont_2_en.jsp). For the moment such ontologies are constructed independently, without any link to higher level ontologies. We propose to extend the OBOE ontology for describing land cover classes expressed in the LCCS. The Entity concept should be specified as a Geographic Entity, which corresponds to any entity that occupies a position in space (Mark, 1993), i.e. Lake Victoria, Amazon River, Amazonian Forest, Paris, etc.

3.2.2 Satellite image ontology

The representation of a geographic entity in a digital image is called a geographic object. Whereas the geographic entity can be described by its real characteristics (vegetation height, leaf type, etc), the geographic object is described by its characteristics in the image.

In a GEOBIA approach, the image characteristics are organized in three levels:

1) features referring to single object intrinsic characteristics, including its spectral response, texture, geometry;

2) features referring to topological spatial relations between two objects;

3) features referring to the way two or more objects are arranged together. This refers in photo-interpretation to what is called "structure" and is linked to the context.

We propose a satellite image ontology containing some useful framework concepts for remote sensing image analysis in figure 2, which Falomier et al. (2011) call a *reference conceptualization*.

Concepts and properties from image ontology have two main purposes:

1) they are used to explicitly describe remote sensing images relying on analysis products (pixel or segment description values) to their corresponding concepts;

2) they are also recombined via description logics in order to define more complex and abstract concepts resulting from remote sensing expert knowledge (For instance, a vegetated segment is defined as a segment whose NDVI mean value is between 0,3 and 0,8.).

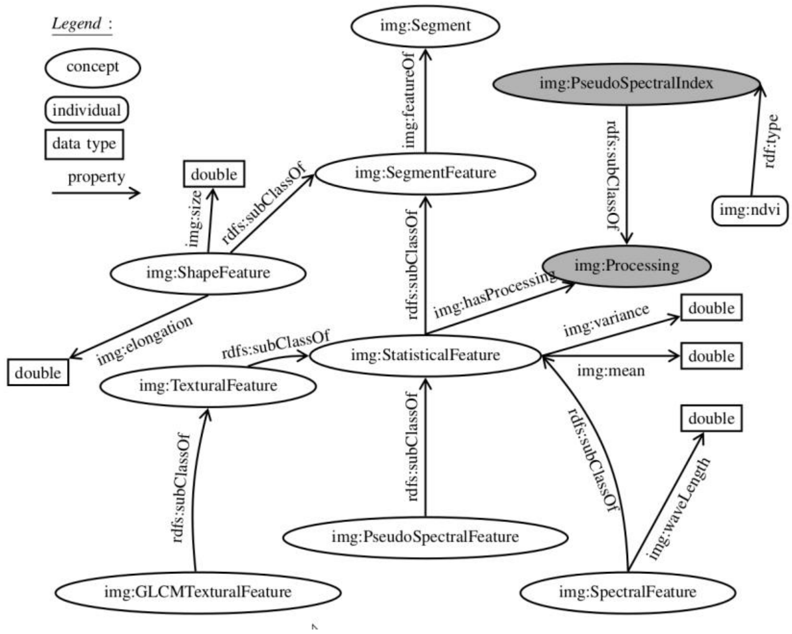

Figure 2: Extract of the ontology we propose to represent remote sensing image conceptualization.

3.2.3 Spatio-temporal ontologies

In OBOE, the context of the observation refers to the relationships the entity maintains with other entities at the time of the observation. In the case of remote sensing image interpretation, these relationships are especially spatial or temporal relationships. Important efforts have been made to model such relations. For example, spatial relations have been split in three categories, i.e. topological relations, cardinal direction relations and distance (or metric) relations (Shariff, Egenhofer, et Mark 1998). In naïve geography, it is usually considered that “topology matters, metric refines”. As a consequence, it has first been proposed to structure topological relations. Two main models are generally considered (the 9-intersection matrix and the RCC-8) to express spatial topological relations. Similar work has been carried out on temporal topological relations also by Allen (1983).

Such knowledge on spatial and temporal relations has then been formalized in ontologies. The SPAN/SNAP ontology is an ontology dedicated to such relations. The SWEET ontology includes a complete “facet” dedicated to relations between concepts, including spatial and temporal relations. The OBOE ontology can be extended to include spatial and temporal relations. It is worth mentioning that in OBOE, such relations can be expressed either as object properties (i.e. relations between concepts) or reified as concepts. This last case might be essential to allow refining topological relations with metrics.

3.2.4 Protocol ontology

While working with scientific observations, it is necessary to get traceability of the protocols used to measure the characteristics of the observed entities. Such a Protocol ontology should focus on (i) the methods used for extracting information from the remote sensing images, and on (ii) the tools used to carry out the method. It has thus been proposed to create method ontologies and software tool ontologies (Benjamin et al. 2005), where the method ontology describes the generic tasks involved in GEOBIA (data selection, image interpretation, final presentation) and the software tools ontology includes specific software operators that can be used to perform a defined task. The method and tools ontologies should then be connected to upper-level ontologies such as OBOE which already includes a Protocol concept defined as a specific procedure used for generating or processing data. Moreover, the applied Protocol often depends on the available data, and consequently on the sensors used to produce the data. A sensor ontology is then required.

3.2.5 Sensor ontology

In OBOE, the *Recorder* of a measurement can be human or non-human (J. Madin et al. 2007). In remote sensing, non-human recorders especially refer to *Sensors* onboard satellites (e.g. the Thematic Mapper onboard Landsat 5 that measures reflectance values in different spectral bands). It would then be necessary to develop a specific domain ontology for remote sensors. Efforts for describing sensors have already been carried out in the frame of SSN (Semantic Sensor Network). An ontology based around concepts of systems, processes, and observations has been proposed and is available at: http://www.w3.org/2005/Incubator/ssn/wiki/images/4/42/SensorOntology20090320.owl.xml.

4. Example: towards an ontological interpretation of remote sensing images

We propose to use ontologies for describing remote sensing expert knowledge and then guide the interpretation of remote sensing images. First, we identified a land cover class to be mapped: beach. We defined the "beach" concept as a “mineral" entity "adjacent to" "water". Once this concept is described, we linked to the remote sensing ontology. For this purpose, we defined the "beach segment" concept that is the representation of a "beach" entity in a remote sensing image. The "beach segment" concept is defined as "mineral or built-up segment" "externally connected to" "water segment". The "externally connected to" relation refers to a spatial topological relation defined in the RCC8 model. For the moment, this relation was implemented in our own local ontology but we expect to import it from a spatio-temporal ontology. Furthermore, remote sensing rules for describing the "mineral or built-up segment" and the "water segment" were defined according to the OBOE ontology. For instance the "water segment" is defined by a measure of the NDWI (characteristic) higher than 0. Similarly, the "mineral or built-up segment" was defined by a measure of the NDBI (characteristic) higher than 0. The NDBI and NDWI characteristics refer to spectral indices computed thanks to specific operators implemented in Orfeo Tool Box (OTB). Such operators were used in order to prepare the NDBI and NDWI images but they should soon be included in the Protocol ontology. Also, the "Externally Connected" spatial relation was checked thanks to an OTB operator.

This approach was applied on a Landsat5 TM image of the city of Santarem and its surroundings (Brazil) from 07/12/2009. All the approach is illustrated in figure 4 that shows the entire processing for mapping the "beach" land cover class, from the input calibrated image, to the computation of spectral indices (NDBI and NDWI) and the identification of beaches.

5. Conclusion

Environmental researches will be improved by interdisciplinary studies. Ontologies are expected to help scientists from various areas to share knowledge. In the specific case of remote sensing, we argue that ontologies can help in guiding the image interpretation process with thematic expert knowledge. For this purpose, we introduced the definition of ontologies and listed some requirements for building a relevant ontological architecture for GEOBIA. We identified two interesting top domain ontologies (OBOE and SWEET) and five domain ontologies (land cover, image, spatio-temporal relations, protocol and sensors). We illustrated how concepts from these diverse ontologies could be associated to interpret images with a simple example (classification of beaches in a Landsat TM image). Our work is still at a preliminary stage and our perspectives consist in reinforcing the connections with existing ontologies. Indeed, for the moment, our approach was implemented locally with our own ontology and we are now working for linking it to the OBOE top-domain ontology.

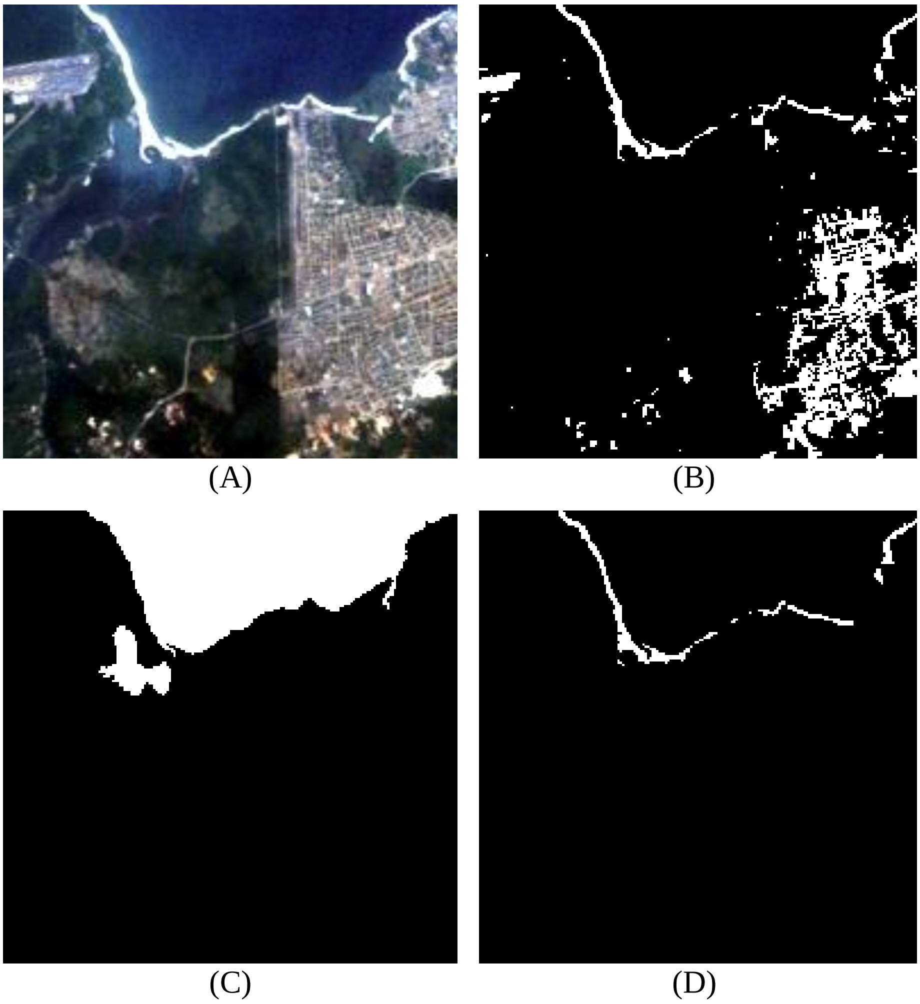

Figure 4: Overall procedure for classifying "beaches" on a Landsat image (A). Defining beach segments by mineral segments (colored in white on (B)) externally connected to water segments (C) gives the result (D).

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Program FP7/2007-2013 under grant agreement n° 263435 for the BIOSOS project and from the PO FEDER GUYANE 2007-2013 program under the frame of the CARTAM-SAT project.

References

Agarwal, P. Ontological considerations in GIScience, **International Journal of Geographical Information Science,** v. 19, n. 5, p. 501-536, 2005.

Allen, J. F. Maintaining knowledge about temporal intervals. **Commun. ACM *26***, v. 11, p. 832-843, 1983.

Benjamin et al. 2005

Blaschke, T.; Strobl, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS, **Interfacing Remote Sensing and GIS**, v. 6, p. 12-17, 2001.

Couclelis, H. Ontologies of geographic information, **International Journal of Geographical Information Science**, v. 24, n. 12, p. 1785-1809, 2010.

Falomir, Z.; Jiménez-Ruiz, E.; Escrig, M. T.; Museros, L. Describing images using qualitative models and description logics, **Spatial Cognition & Computation**, v. 11, n. 1, pp. 45–74, 2011.

Gruber, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, v. 5, n. 2., p. 199-220, 1993.

Hay G. J.; Castilla, G. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: Blaschke, T.; Lang, S.; Hay, G. J. (Org.) **Object-Based Image Analysis**, Berlin Heidelberg: Ed. Springer, p. 75-89, 2008,.

Madin, J. S.; Bowers, S.; Schildhauer, M. P.; Jones, M. B. Advancing ecological research with ontologies, **Trends in Ecology & Evolution**, v. 23, n. 3, p. 159-168, 2008.

Madin, J.; Bowers, S.; Schildhauer, M.; Krivov, S.; Pennington, D.; Villa, F. An ontology for describing and synthesizing ecological observation data, **Ecological Informatics**, v. 2, n. 3, p. 279-296, 2007.

Mark, D. Toward a theoretical framework for geographic entity types. **Computer Science**, **Spatial information theory: a theoretical basis for GIS**, v. 716, p. 270-283, 1993

Mark, D. M.; Smith, B.; Egenhofer, M. J.; Hirtle, S. C. Ontological Foundations for Geographic Information Science. In: McMaster, R. B.; Usery, E. L. (org.). **A research agenda for geographic information science**, Éd. Boca Raton, Fla.: CRC Press, 2005.

Pinto, H.S.; Martins, J. Ontologies: how can they be built? **Knowledge and Information Systems**, v. 6, p. 441–464, 2004.

Raskin, R. G.; Pan, M. J. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). **Computers & Geosciences**, v. 31, n. 9, p. 1119-1125, 2005.

Shariff, A.R.; Egenhofer, M.; Mark, D. Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English-Language Terms. **International Journal of Geographical information Science**, v. 12, p. 215-246, 1998.

Tripathi, A.; Babaie, H.A. Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies, **Computers & Geosciences**, v.34, n.9, p. 1022-1033, 2008.