Project no: 043268

Project acronym: PATRES

PATTERN RESILIENCE

Instrument: STREP

Thematic Priority: New and Emerging Science and Technology

Deliverable 1.1: Report on case studies

Start date of project: 01/02/2007

Duration: 36 months

Project coordinator: Guillaume DEFFUANT
Project coordinator organisation: Cemagref

Revision: Version 2.1
March 2010

Content

1 Introduction

This report summarises the main steps of the work on the case studies carried out in the PATRES project. The aim of these case studies was initially twofold:

- To test the developed methods and tools for exploring and managing pattern resilience in complex systems, and to identify how these tools should evolve according to the needs identified in the case studies.
- To develop new knowledge about the case studies from the applications of these new methods.

Several of these case studies are described with more details in PATRES book (and also in published papers), and we shall only mention the main aspects of the work, referring to the corresponding chapters of the book and papers for the details. We also mention some exploratory work which is not presented in the book.

The case studies implement the global approach for pattern resilience, developed in the project. In this approach, patterns are regularities in the systems' structure and dynamics which are generated by their internal organization. For example, in savannas trees and grass coexist for a very wide range of environmental conditions, with the tree cover usually not exceeding 20% of the areas. Savannas are thus characterized by a spatial pattern (more or less scattered distribution of trees in a continuous grass layer), a pattern in organization (not more than 20% tree cover), and a pattern in time (long-term coexistence of trees and grass). In the case studies, we try to identify these regularities using methods coming from physics, which for instance summarise a spatial distribution using pair distributions (see deliverable 2.1 and the book for more details about the methodological aspects).

Resilience is generally defined as the property of systems to maintain their internal organization and, in turn, pattern dynamics, despite disturbances and changing environmental conditions. Accordingly, "pattern resilience" is understood as the capacity of a system to maintain or to recover some desired pattern dynamics (which are related to useful functions) in a changing environment. The key word in this context is "desired", because PATRES aims at methods and tools that support complex system management.

The problem of maintaining a system within a desired subset of its state space can be seen as a viability problem, and viability theory can be used to solve it. In particular, this theory led to develop tools for computing action policies maintaining the system in the desired set. This idea is the starting point of a viability based definition of resilience: a state of the system is resilient if it is possible to find an action policy which drives back the system to its viability kernel (from where it is possible to maintain it indefinitely in the desired states). Deliverable 3.1 and the book provide more details about this approach.

Moreover, a main idea of PATRES was to develop and test new methods and tools across different disciplines so that that the final results would be independent of the idiosyncrasies of specific fields and systems. Therefore, PATRES focused on case studies from different domains (ecology, microbiology, linguistics, social sciences). In the following, we describe these case studies: first, we present the system in question, then the models used to explore the systems and their pattern dynamics; finally, achievements regarding their viability and resilience are summarized.

2 Case studies with full application of the pattern resilience approach

We first report the case studies on which we managed to apply the whole PATRES approach. For each of these case studies, the viability - resilience part (derived from the simplified dynamics) is implemented in the KAVIAR software tool.

2.1 Savanna dynamics

More than 10% of the Earth land is covered by savannas. Due to the increase in human population, they are increasingly used for livestock grazing and fire wood production. Overused savannas tend to loose their patterns and, accordingly, their ability to provide ecosystem services because they develop into steppes dominated by shrub species that cannot be used for livestock or fire wood production. To explore pattern resilience of savannas, we built on an existing computational micro-level model which simulates the relevant patterns, the Jeltsch model (Jeltsch et al., 1996, 1997, 1998).

The Jeltsch model is widely cited and well-tested. It has been used to explore tree-grass coexistence, the risk of shrub encroachment because of overgrazing, and to explain vegetation patterns around waterholes. It has, however, also a number of limitations: it is specific to sites in the Kalahari Desert in Namibia, and it is relatively complex so that it never has been systematically been explored regarding pattern dynamics.

Following the general approach of PATRES to devise simpler macro-level models, Calabrese et al. (2010a) used a reimplementation of the Jeltsch model, which was produced for PATRES, and focussed on identifying dynamic and spatial patterns that could be used as starting point for formulating a macro-level model. These patterns were a logistic curve of the number of trees vs. time if fire was excluded, and specific features of the trees spatial distribution if fire was included. This task is typically of the type for which SimExplorer can be used. This software tool was developed in the project and aims at facilitating the simulation experiments. More information about it is given in deliverable 2.1 and in chapter 9 of the book.

Then, we developed a simplified simulation model which on the one hand reproduced the key patterns identified in the Jeltsch model, on the other hand could be approximated by an analytical model. Two approximation methods were used: mean-field approximation and multi-scale pair approximation.

The resulting models were simple enough to explore systematically their pattern dynamics. Calabrese et al. (2010b) found that the interaction between tree-tree establishment competition and fire is key to explaining savanna dynamics. Interestingly, competition between adult and establishing trees was included in the Jeltsch model, but never identified as key generative mechanism.

Moreover, this model was simple enough to use the tools developed in the project for computing viable or resilient policies. Indeed, the mean-field approximation is two dimensional (it involves the density of trees and the level of grazing), and the pair approximation model involves 4 dimensions (tree density of and grazing, plus two variables: one representing the density of pairs at short distance, and the other and far distance). One difficulty in this case study has been to choose properly the two last variables to enter in the viability – resilience model, because the initial variables of the model have no constant bounds when tree density varies, and this causes practical difficulties (see chapter 6 for more details).

Our approach to resilience is related to the choice of the desired set of state in which we would like to maintain the system. It is interesting to make different hypotheses about this set, and to study their consequences. The first step we defined the desired set around the attractor tree density of 0.22, given by the mean field model, and supposed that the grazing cannot be changed. The corresponding viability kernel is in this case the same as the desired set, and the resilience set is very large, corresponding to the attraction basin of this attractor (this type of result is not surprising, and explained in details in chapter 3). Moreover, a second attractor for a tree density of 0 appears, with a much smaller attraction basin (for small values of tree density and grazing). This attraction basin includes the states which are not resilient for this problem setting.

In a second step, we supposed that it is possible to modify slightly the grazing at each time step, and that the desired set does not include the tree density attractor around 0.22. We defined the desired set of states with a tree density strictly above 0, and below 0.2. In this case, a management is necessary to prevent the "natural" dynamics of the system which is attracted either by tree density 0.22 or tree density 0. The result of the computation yields a viability kernel which is significantly smaller than the desired set (meaning that from many states of the desired set, whatever the management policy, the system crosses the constraints at some point), and more so when the minimum grazing level increases. The viability kernel lies around the boundary between the attraction basins, and the policy of action must pay constantly attention to the possibility to cross the limit of this kernel, and act preventively.

The resilience analysis of this setting shows that there is a limit from which it is not possible to drive back the system into the viability kernel (given the hypotheses on the possible changes of grazing). This limit corresponds to the possibility for the management action to reach the boundary between the attraction basins.

The study based on the pair approximation model shows an important difference between the two models: in the pair approximation model, there exists attractor points for a tree density around 0.18, which leads to significantly different viability kernels when considering the same desired set as previously (maximum tree density 0.2 and minimum grazing and tree density strictly superior to a threshold value). Indeed, the viability kernel includes the points from which it is possible to reach this new attraction basin. The other part of the viability kernel has a projection in the space (tree density, grazing) which is similar to the one found with the mean field model: it is located around the boundary between the two attraction basins, which has approximately the same location.

Globally, this case study illustrates well the whole approach of the project: from an existing complex model involving a lot of different processes spatially distributed, we derived an approximate individual-based model which is much simpler, from which it is possible to apply the pattern dynamics techniques (pair approximation). This provides a dynamics which is based on a small set of variables (2 to 4) for which it is possible to make a viability-based resilience study.

2.2 Bacteria biofilms

Biofilms are thin layers of bacteria attached to a surface, which can show a large variety of spatial structures. They emerge spontaneously from the interactions between bacteria and with their substrate. The spatio-temporal dynamics of biofilms is important both for the function of biofilms and for species and functional diversity. Understanding the generative mechanisms underlying the spatial structure is thus important both in terms of community ecology of bacteria and industrial applications, e.g. medicine, bioreactors, and water treatment.

In this case study, we developed a first individual-based model of bacteria biofilms (Mabrouk et al. 2010). This biofilm model is relatively complex and computationally demanding, mainly because bacteria are represented as spheres competing for space, and also because of the careful representation of the substrate diffusion. Moreover, this model includes particles representing polymer produced by bacteria, which are supposed to modify their motility. We studied the properties of this model through experimental designs of simulation (using SimExplorer), where we tried to identify extreme behaviours of the model, depending on the parameter value ruling the motility of bacteria as a function of the density of polymer. We show that for the extreme values, we get well known spatial patterns, either a single big colony, or randomly distributed micro-colonies of different sizes. When considering intermediate values of the variable, we get more complex patterns, with interconnected colonies. Such patterns are observed in real experiments.

We developed a simpler model representing bacteria cells as points with only a position but no spatial extension, and no substrate dynamics. The question addressed with the new model was to explore how the processes represented in the model affect the spatial pattern and dynamics of the biofilm. Only one hypothetical species was included, and the model processes were: reproduction and mortality; detachment of bacteria from the surface; migration of bacteria to new locations; production of extracellular polymers; detachment of polymers. These processes are expressed through simple kernel functions to express their locality, and they are stochastic (taking place with defined probabilities).

The basic assumption of the model is that bacteria produce extracellular polymers, which are released to and remain in the immediate neighbourhood of a cell. Aggregations of bacteria cells produce enough polymers to reduce local mobility of the bacteria. The interplay of the model process then leads to different spatial patterns of bacteria distribution. To identify its pattern dynamics, the model has been first systematically analyzed using the software tool Simexplorer. The metric used for analysing the spatial distribution were, as for the savanna model, spatial statistics, for example pair correlation functions. This systematic exploration showed that the dynamics produce three main patterns, depending on the values of the parameters: uniform distribution, regularly spaced micro-colonies, and labyrinth like patterns. The intermediate patterns between these can also be found.

These analyses also revealed that for the simplified model it was sufficient to include only the bacteria, not polymers, because for biologically meaningful parameter values, the spatial output of this simpler model showed the same patterns as the model with polymer. Hence, for applying the viability- resilience approach, we used this model without polymer.

The state variables of the analytical approximation the biofilm IBM are based on the first and second moment of the bacteria's spatial distribution, i.e. average density and the pair correlation function. This pair correlation function is expressed in two dimensions and is represented by a lot of variables (all distances from the considered bacterium). This is more complex than the analytical approximation used for savannas (which includes only two variables), but still could be entered directly in the tools for computing viability and resilience.

We needed a description of the pattern, based on a small number of variables. After some investigations, we noted that, at least for a restricted set of dynamics where the size of the kernel functions determining the dynamics of the bacteria are kept fixed, the integral of the pair approximation function above 1 was a particularly interesting indicator to characterise the pattern. Indeed, the emergence of new colonies tends to increase systematically this indicator. We verified that this indicator coupled with the squared density of bacteria can be associated with a single pattern of correlation function.

In an idealised viability-resilience problem, the control variable is the density of bacteria that can be added to the system. We constituted a data base of the possible patterns that can be reached, by a systematic exploration of different values of this control, and we associated the patterns of this database to its indicators in two dimensions (the integral over 1, and the squared bacteria density). The actual dynamical evolution that is implemented in the software tool Kaviar, is the pair approximation evolution which is in a large state space (100 x 100 dimensions), and once this evolution is performed, we come back to the indictor space by computing the corresponding indicators for the reached pattern (see chapter 7 of the book for more details).

2.3 Language competition

Currently, thousands of languages are at risk of getting extinct in this century (Crystal, 2000). This loss of diversity in languages would be linked to a major, and irreversible, loss in our historical and cultural heritage. It has therefore been tried to use model to understand the reasons why languages go extinct and which measure can be taken to stop the decline of a certain language. One focus in this context has been on communities with bilingual agents which can speak two languages and have the choice to use only one of them and thereby contribute to the decline of the other language, both because of their own use and because their children might grow up more or less monolingual. Examples of bilingual societies include Welsh and English in Wales and Catalonian and Spanish in Catalonia, Spain.

First models addressing bilingual systems where adopted from population ecology and classical Lotka-Volterra type models of competition. However, such models often cannot take into account local interactions and individual heterogeneity. Therefore, micro-level simulation models (cellular automata, individual-based) are increasingly used (Abrams and Strogatz 2003, Stauffer 2007).

In this case study, an individual-based version of a simple model of Abrams and Strogatz (2003) was taken as the point of departure. In this model, agents are distributed on a square grid and speak one of two languages. In each time step, they can choose to switch their preferred language. The factors affecting the choice of language are: local densities of the two languages spoken in the neighbourhood (four neighbours on the grid); prestige, i.e. the agent's social benefits of speaking a language, and volatility, a parameter which determines how stochastic the agent's decisions are. Here, stochasticity represents factors other than local density and prestige that might influence the agents' decisions. In the case study, also an extension of the Abrams and Strogatz model was used, where agents can also can chose to speak both languages simultaneously.

Simulations of this model showed that, depending on volatility and the difference in prestige of the two languages, both languages either coexist for long times, or one of the languages takes over, driving the other language to extinction. High volatility favours coexistence if prestige differences are high, whereas for low prestige differences, low volatility increases time to extinction considerably.

To obtain macro-level equations for the language competition model, first mean-field approximations were formulated, assuming that no spatial relationships exist, i.e. a well-mixed system where all agents interact with each other. Then, first different complex networks of interaction were included, based again on mean field and pair approximations. Finally, interactions on a square grid, as in the simulation model, were studied, which required a more sophisticated approximation techniques, based on a continuous field describing the local dominance of a language, and a time dependent Ginzburg-Landau equation describing the dynamics of this field.

We applied the viability – resilience approach to the mean-field model, taking two state variables: the percentage of usage of the languages, and the prestige. We considered as control the possibility to modify slightly the prestige by policy actions. We defined the desired set with the constraint that both languages should be used at least at 20%.

The study shows two different situations, depending on the volatility a, the main parameter of the model. Indeed, when this value is lower than 1, there are attractors of the dynamics for intermediate percentages of usage of the language, whereas if this value is higher than 1, then the only attractors are when one language is extinguished. In both cases, the viability kernel is not void, but it is much larger for $a < 1$, than for $a > 1$. Moreover, for $a > 1$, the action policy should be very careful, in alternating the actions in favour of one language and then of the other, whereas for $a < 1$, generally the system stabilises within the constraints without action.

The resilience study shows an even sharper difference. For $a < 1$, almost the whole state space is resilient. This is only when one language is almost not used at all and its prestige is very small that the situation becomes very long to revert. But for $a > 1$, a significant part of the state space is not resilient, meaning that whatever the policy of action, from these states one of the language will become extinct.

This case study has been very complete in terms of pattern dynamics, with the use of the Langevin approach to model the stochastic variations around the mean-field. However, the study of viability – resilience was limited to the mean-field case. The details are provided in chapter 3 of the book and in (Chapel et al. 2010) and in (Vasquez et al. 2010).

2.4 Social dilemma

In this case study, we explore the possibility to apply the approach of the project to the social dilemma and emergence of cooperation dynamics. The social dilemma is modeled as a game having different outcomes depending on the attitudes of the players. If both players cooperate, they get a good reward, if both players defect, they get a lower reward, if one player cooperates and the other defects, the cooperative player gets no reward, and the defective player gets a big one. The model includes a population of agents with different behaviours (altruist, reciprocator, defector) in varying proportions. The agents tend to imitate the behaviour which is the most rewarding.

The main idea is that there are several ways for an institution to act on the strength of a social dilemma (tax, subsidies, etc) but this action is often costly. Thus the different measures that an institution can undertake can thus be viewed as control parameters that help the institution to maintain the social system within its domain of normal functioning. Since these measures are costly, the institution faces a viability problem with resilience episodes: how to maintain the social systems in the viability kernel and what are the optimal controls to apply to restore the system when it leaves this viability kernel.

In the first phase of PATRES, we began to study this model in the viability framework. In the first setting, we considered only as state variable the strength of the social dilemma, the proportion of altruists, and the proportion of reciprocators. The global dynamics were given with a mean field approximation, derived from the agent model. We checked the accuracy of this approximation.

The control was defined as a modification of the strength of the social dilemma, chosen between a minimum and a maximum. We found that below a critical value of this maximum, the viability kernel is restricted to a very tiny strip of the state space. Above the critical value, the viability kernel is a large part of the desired set (more than 80%).

In the second phase of the project, we included the cost of the action in the model. To do this, we considered that the institution has a budget for this type of action at each time step, and we compute the cumulated remaining budget. We fix a constraint that this cumulated remaining budget cannot be below a threshold value. This completes the desired set, and we could study the resilience of the system in this setting. The details are provided in chapter 5 of the book.

3 Other case studies

This section reports the work on the case studies on which we could not apply the whole approach. Nevertheless, the work of the project on these case studies led to valuable knowledge advances.

3.1 Pattern dynamics in Web communities

Web communities producing collaborative content are a new type of complex systems, e.g. encyclopaedias (Wikis), online photo repositories (Flickr), and many others. As with any other agent-based complex system, the properties of these communities, in particular the number of users and webpages, depend on the agents' decisions and on specific communities policies, for example whether the communities are completely open or have administrators. It is worthwhile studying such communities because they have an increasing impact on social networks and knowledge and data storage and retrieval (Taraborelli et al. 2008). Online communities like Facebook and Twitter are likely to strongly affect economies and politics in the future. Another reasons for studying web communities is that they sometimes provide unique data about the behaviour of a huge number of agents – a database that usually is not available for ecological, microbial, or language systems.

Therefore, PATRES started studying different sets of data in order to determine empirically the main dynamical patterns that are observable. State variables considered were mainly the number of users, pages, edits, and, on the governance side, administrator to user ratio, administrator density, and edition permission. As a first indicator of ‚viability', the growth rate and persistence of wikis in terms of users and pages was determined.

Out of data from more than 10,000 wikis, 360 were selected which had user number within certain bounds (i.e. from 400 to 20,000 users), which were not hosted by so-called ‚wiki farms', and which did not show too abrupt changes in user and page number (not more than 5% change within one day).

Patterns identified were: wikis with active users, who edit a lot, growth faster than wikis with less active users, in particular if the initial size of a wiki is larger; larger numbers of administrators tend to slow down the growth of wikis; wikis where users need to register grow significantly slower than wikis that do not require registration; overall, the growth of users and content of a wiki is correlated (Roth et al. 2008a, b).

In a similar analysis, Flickr, a photo and video sharing community, was analyzed. Here, the central structural elements are groups. Any user can start running a new group, which can be focussed on any topic, e.g. portrait, wildlife photos, etc. The main metrics used in this analysis were number of group members, number of photos in a group, and active member ship spread of group members (how many groups they join). Data were analyzed using a linear regression model, where the influence of various factors on absolute group growth. It was found that growth is affected by the size of a group and group cohesiveness (degree of social closure), whereas – strikingly – governance had little effect (Baldassari et al. 2008).

3.2 Pattern dynamics in scientific literature

The Internet provides, similarly to web communities, unprecedented data and means for their evaluation regarding pattern dynamics in the scientific literature. In this case study, more the 18 Million references of biomedical literature, taken from the database MedLine, were analysed. The aim was to identify different modes of science evolution and the viability constraints of different scientific fields.

We used different methods of co-word and co-occurrence analysis to identify scientific fields and their evolution in the course of time. Scientific fields are defined by sets of key-phrases, for example: *cancer, tumor, proliferation, apoptosis, suppressor, cell cycle*. The questions then are how scientific fields emerge, evolve, differentiate, and die.

To quantify the proximity between scientific fields, a new proximity measure was developed (Chavalarias & Cointet 2008), as well as a new measure of the consistency of scientific fields (Cointet & Chavalerias 2008). As a new tool for analyzing bibliometric datasets, the algorithm FieldBuilder was developed, that implement the new proximity and consistency measure. These approaches also allow to construct "phylogenies" of scientific fields, i.e. a field can give birth to one or more new branches, or children. A first application, research related to bias in biomedical research was analyzed (Chavalerias & Ioannidis, in press).

In general, it was found that younger scientific fields tend to be "bushy", i.e. have lots of "children", whereas older, more established fields will generally have a much more linear evolution with a lower rate of conceptual renewal. But when a phylogenetic branch begins to loose strength, it is quite rare that it stops just like that. Often it becomes bushy again, which corresponds to the introduction of new concepts, so that one of the offspring recovers, including higher structuration indexes.

This case study dealt with a complex system that so far received little attention in terms of pattern dynamics. The new methods and tools for analyzing electronic databases revealed interesting patterns, which provide the basis for developing dynamic models, both micro-level (agent-based) and macro-level. In order to apply the approach developed in the project, we should have developed a model of this dynamics, and approximate it in a small number of dimensions.

4 References

Abrams, D. M. and Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature*, 424:900.

Baldassarri, A., Barrat, A., Cappocci, A., Halpin, H., Lehner, U., Ramasco, J., Robu, V., Taraborelli, D. (2008) The Berners-Lee Hypothesis: Power laws and Group Structure in Flickr. *Proc. of Dagstuhl Seminar,* November 2008, http://drops.dagstuhl.de/opus/volltexte/2008/1789/pdf/08391.SWM.Paper.1789.pdf

Calabrese, J.M., Deffuant, G., Grimm, V. 2010a. Bridging the gap between computational models and viability theory in savanna ecosystems. In: Deffuant, G., Gilbert, N. (eds) *Pattern Resilience: Computing Resilient Action Policies for Social and Ecological Research*. Springer (to appear).

Calabrese, J.M., Vazquez, F., López, C. k San Miguel, M., Grimm, V.. 2010b. The individual and interactive effects of tree-tree establishment competition and fire on savanna structure and dynamics. *American Naturalist* 175: pp. E44–E65

Chapel, L, Castelló, X, Bernard, C, Deffuant, G, Eguíluz, V, Martin, S, San Miguel, M. 2010. Viability and Resilience of Languages in Competition. *Public Library of Science*. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2811194

Chavalarias D. & Cointet J-P. (2008) Bottom-up scientific field detection for dynamical and hierarchical science mapping - methodology and case study *Scientometrics* Vol. 75 No. 1 , (DOI): 10.1007/s11192-007-1825-6.

Cointet J-P., Chavalarias D. (2008) Multi-level Science mapping with asymmetric co-occurrence analysis: Methodology and case study, *Networks and Heterogeneous Media*, Vol 3 Number 2, june 2008, p267-276

Chavalarias & Ioannidis (in press) Science mapping analysis characterizes 235 biases in biomedical research. *Journal of Clinical Epidemiology.*

Crystal, D. (2000). *Language death*. Cambridge University Press, Cambridge.

Mabrouk N, Deffuant G, Lobry C. Tolker-Nielsen T. 2010: Bacteria can aggregate into interconnected microcolonies when a self-excreted product reduces their surface motility: evidence from individual-based model simulations, *Journal of Theoretical Biology* – In press.

Roth, C., Taraborelli, D., and Gilbert, N. (2008a) Measuring wiki viability. An empirical assessment of the social dynamics of a large sample of wikis. In WikiSym ’08: *Proceedings of the 4th International Symposium on Wikis* (New York, NY, USA, September 2008), ACM. http://nitens.org/docs/wikidyn.pdf

Roth, C., Taraborelli, D., and Gilbert, N. (2008b) Démographie des communautés en ligne: Le cas des wikis. *Réseaux* 26, 152 (2008), 205–240. http://dx.doi.org/10.3166/Reseaux.152.205-240

Stauffer, D., Castell´o, X., Egu´ıluz, V. M., and San Miguel, M. (2007). Microscopic abrams-strogatz model of language competition. *Physica A*, 374:835–842.

Taraborelli, D., Roth, C., and Gilbert, N. (2008) Measuring wiki viability (II). Towards a standard framework for tracking content-based online communities. Tech. rep., 2008. http://wikitracer.com/refs/wikitrack.pdf

F Vazquez , X Castelló and M San Miguel (2010) Agent based models of language competition: macroscopic descriptions and order–disorder transitions. Journal Statistical Mechanics P04007