Reproducing and exploring past events using agent-based geo-historical models

Nasser Gasmi, Arnaud Grignard, Alexis Drogoul, Benoit Gaudou, Patrick Taillandier, Olivier Tessier, Duc An Vo

Reproducing and exploring past events using agent-based geo-historical models

Nasser Gasmi, Arnaud Grignard, Alexis Drogoul, Benoit Gaudou, Patrick Taillandier, Olivier Tessier, and Vo Duc An

UMR 5505 IRIT, CNRS, University of Toulouse, France
UMI 209 UMMISCO, IRD, UPMC, Hà Nội, Việt Nam
UMR 6266 IDEES, CNRS, University of Rouen, Rouen, France
EFEO, French School of Asian Studies, Hà Nội, Việt Nam
Contact author: `benoit.gaudou@ut-capitole.fr`

Abstract The field of "digital humanities" is about using the latest digital methodologies in order to tackle humanities disciplines and social sciences questions. The ARCHIVES project belongs to this new research area. It proposes a methodology to build agent-based models of historical events, in particular crisis events, in order to answer new questions about them or explore them in new ways. In this paper, we present the first implementation of ARCHIVES on the case study of the management of floods in Hà Nội (Việt Nam) in 1926. We show how we collected, digitized and indexed numerous historical documents from various sources, built a historical geographic information system to represent the environment and flooding events and finally designed an agent-based model of human activities in this reconstructed environment. We then show how this model helped us understanding the decisions made by the different actors during this event, testing multiple scenarios and answering several questions concerning the management of the flooding events.

Keywords: Agent-based model, historical geographic information systems, historical event simulation, crisis management.

1 Introduction

It is now widely accepted that the adaptation of human communities to natural hazards is partly based on a better understanding of similar past events and of the measures undertaken by impacted groups to adapt to them. This "living memory" has the potential to improve their perception of the risks associated to these hazards and, hopefully, to increase their resilience to them. However, it requires that: (1) data related to these hazards are accessible; (2) relevant information can be extracted from it; (3) "narratives" can be reconstructed from these information; (4) they can be easily shared and transmitted. This is classically the task of archivists and historians to make sure that these conditions are fulfilled. However, these last years, the new area of research of Digital Humanities that consists in integrating computer technology into the activities of

humanities scholars, has brought innovative approaches and methods to fulfill these tasks.

The ARCHIVES project is part of this new research area. It has for goal to propose a methodology that would enable to support the work of the historians, in a systematic and automated way, from the analysis of documents to the design of realistic geo-historical computer models. Our aim is that, using these models, users can both visualize *what happened* and explore *what could have happened* in alternative "what-if" scenarios. Our claim is that this tangible, albeit virtual, approach to historical "fictions" will provide researchers with a novel methodology for synthesizing large corpuses of documents and, at the same time, become a vector for transmitting lessons from past disasters to a contemporary audience. In addition, it is an attempt to give an experimental approach to History [6]. Previous works have used agent-based modeling and simulation to tackle historical and archeological questions, e.g.[1] tested possible causes of the disappearance of the Kayenti Anasazi and [5] investigated the resilience of the ancient Maya civilization. But such projects simulate a whole civilization over hundreds years, whereas we focus on a short event (a crisis) at a very low level (individual human beings), with (often) a large amount of detailed documents.

In this paper, we present a first implementation of ARCHIVES concerning the study of the floods in Hà Nội (Việt Nam) at the beginning of the 20th century. The aim of this study is to better understand decisions that have been taken during those exceptional events and especially to understand the role of the political actors implied in the decision process. This case study is particularly interesting as the problem of floods in Hà Nội is still topical. Moreover, a lot of period documents have been written and stored concerning these events.

This paper is organized as follow: Section 2 presents the historical context of the case study. Section 3 is dedicated to the presentation of the methodology that has been carried out. Section 4 presents the final model that has been developed. Section 5 exposes the results and proposes a discussion about them. At last Section 6 concludes and proposes some perspectives.

2 Historical context

2.1 Việt Nam and floods: a thousand years story

Due to its climatic and geographical location, and in particular to the fact it is structured around two huge deltas (the Red River delta in the north and the Mekong River in the south), Việt Nam has faced devastating floods all along its history [7,8]. Opposite strategies have been used in the two deltas that structure the country: while the north has emphasized the construction of dykes to stem the Red River, the south has adapted by digging a dense network of canals in the Mekong River delta. And, despite the political upheavals undergone by the country in the last centuries, and more recently during the Nguyễn dynasty (1802-1945), the French colonization (1865-1954), the independence (1955) or the reform policy (đổi mới, 1986), these strategies have remained virtually unchanged.

In this study we focus on the Red River delta. In this area both in the past and today, effects of floods are often devastating both for the population and for economy and agriculture, because the area is very populous and produces a huge amount of rice. Even in 1930, 6,5 millions inhabitants lived in the delta (with a density very high of 430 people par km2) and cultivated 1,1 millions ha of rice fields.

Works to build dykes on the Red River began in 1099. All over history, speed and progress of works have followed the strength and wealth of the government. In particular in the XIIIth century, with the reinforcement of the power the Trần dynasty, huge works have been done from the source of the river to the sea. Gourou in [3] considered that the embanking of the whole river has been achieved at the eve of the French colonial period (1865), with about 4000 km of dykes.

2.2 Political and social context in Việt Nam in 1926

In 1926, the north (Tonkin) and the center (Annam) of the Việt Nam were under a French protectorate, whereas the south (Cochinchina) was a French colony. They were all part of the French Indochina, a federation that also included Cambodia and Laos protectorates.

The fact that the Tonkin was at that time a protectorate (and not a colony) had several consequences in terms of governance. In particular indigenous institutions,from the emperor and his ministries to local authorities, were maintained. The country was divided into provinces, provinces into districts and districts into commune. Each of these levels was managed by local authorities hierarchically organized. The French authority was limited to the Resident Superior of Tonkin that managed all the area and one Resident of France per province. There was no French authority in sub-levels. In addition to authorities (Residents), the French organization also included technical services (public works, hydraulic...).

It is important to note that officially, there is no hierarchical relationship between the French organization and the Vietnamese one. As illustrated in the Figure 1, the Resident of France in Bắc Ninh (the province on which we focus for the floods, on the east of Hà Nội on the other side of the Red River) could only request intervention to his Vietnamese homolog, but cannot order anything.

2.3 Floods in Hà Nội in 1926

This study focuses on the particular case of floods in Hà Nội during the summer 1926 (from the 25th of July until November when the last breach have been plugged). We can distinguish two main parts in the disaster management. In the sequel we will focus on the first one.

The first period, from the 25th of July to the 30th of July, corresponds to the increase of the water level and the breaches in dykes. The first priority was to protect the dykes and in particular to avoid water submersion, which is the main reason of dyke breaches. The best (and the only) way to do it is to increase their height by building small dykes on top of them using anything available (e.g.
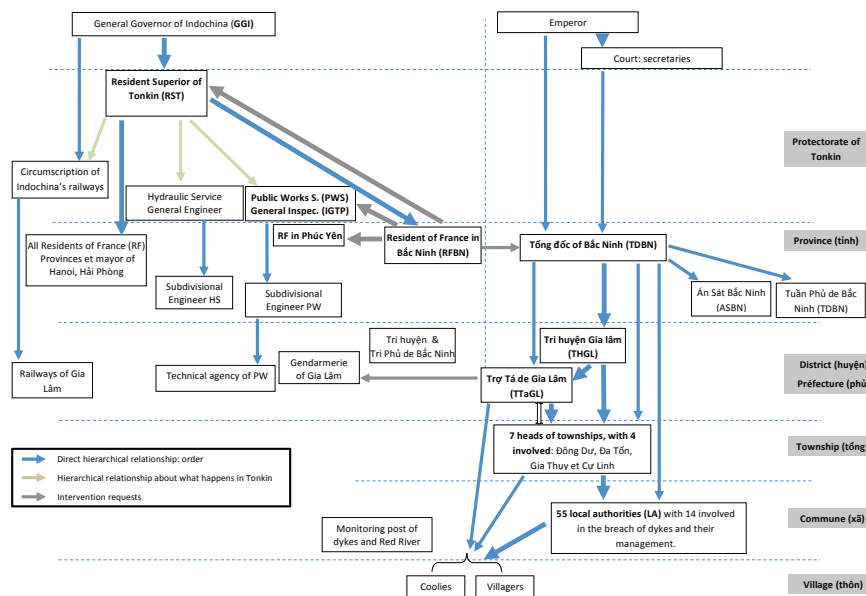
Figure 1. Hierarchical organization of the Tonkin and the Bắc Ninh province

ground of the dyke itself). So the main task of the authorities was to go from commune to commune to order the construction of small dykes and to supervise works. Even with these works, in the 28th and 29th of July, dykes were broken in three main locations: Gia Quất, Ái Mộ and Lâm Du.

From the 31st of July to November, French authorities began large works to plug the three breaches. Technical services called up thousands of Vietnamese workers and provided a large amount of materials (e.g. sandbags or bamboos) to plug the three main breaches, which took months for the Lâm Du's one. The authorities had also to deal with thousands of victims: they provided foods and raw materials to build shelters. They also tried to determine responsibilities: an inquiry has concluded to the guilt of the vice-chief of the district where the three breaches occurred (Gia Lâm). Thanks to this inquiry, a lot of written documents and reports have survived until now in the Vietnamese archives.
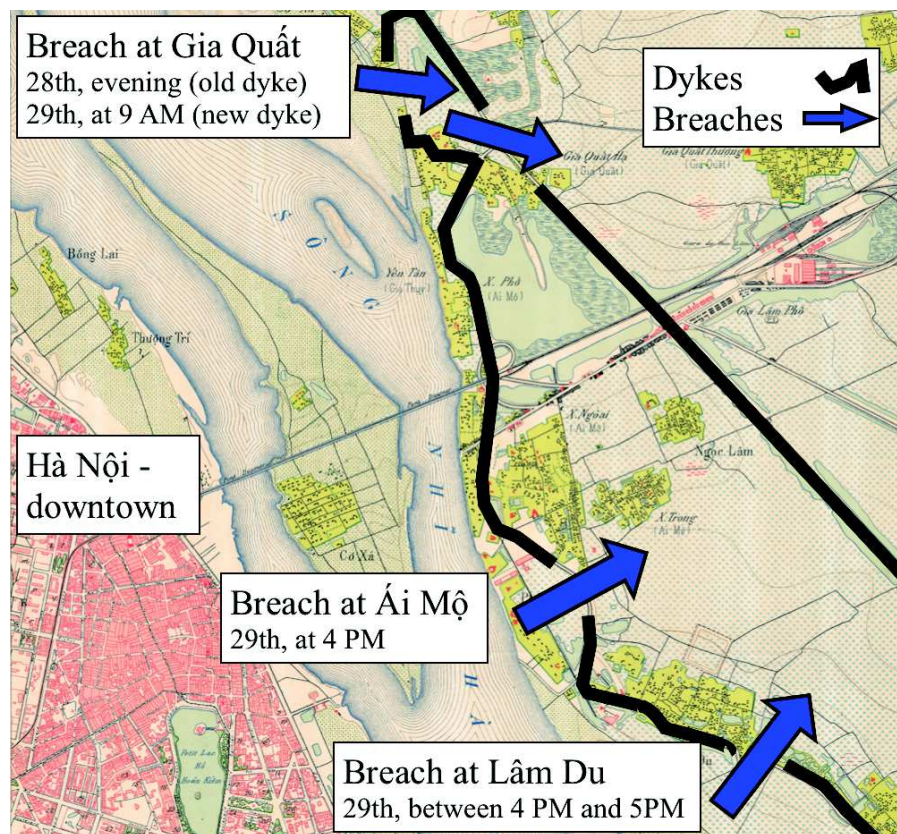

Figure 2. Summary of the three main breaches during the floods of 1926

3 Methodology and tools

3.1 General methodology

In order to achieve our goal, we propose a methodology composed of 3 steps:

1. Building of a comprehensive and navigable corpus from the heterogeneous set of archives.
2. Representation of this corpus in its multiple (spatial, temporal, social) dimensions so that it can be manipulated and explored freely by users.
3. Building of models of the reality depicted by the corpus so that they can offer an account of what happened but also become the objects of experiments.

The first step consists in building the corpus. Indeed, the memory of a past event is preserved by society and state through testimony or records, materialized by various documents from heterogeneous sources that offer a fragmented vision, sometimes contradictory, of reality. The goal of this step is to collect all the documents that are relevant for the studied event and to digitize them.

The second step consists in enriching and contextualizing this textual corpus. It requires linking it to a geospatial representation. Maps are essential in disaster preparedness as they provide information of paramount importance: communities know where impacts have been the greatest in past events or where infrastructures are likely to be weaker which gives a chance to adapt more quickly and thoroughly to a new disaster. Spatializing this corpus also represents a great opportunity for historians to establish relationships between events distant in time.

The last steps consists in deriving *geo-historical models* from *geo-historical representations*. Indeed, the corpus and its associated multidimensional representation carry a linear causality, which can prevent users to gain an empirical understanding of the decisions taken by the actors in their context. The course of the event and the decisions are all *inputs* of this representation, while an experimental approach based on hypothetical reasoning ("what would have happened if ...", "what effect this decision could have had on ...") would require some of them to become *outputs* as well, so as to enable the exploration of alternative paths or the assessment of hypotheses.

In the next sections, we detail the three steps and their implementation for the Hà Nội floods of 1926.

3.2 Step 1: Building of the corpus

The first step consists in building the corpus. This step requires to:

- define the event to study and the questions that the model should answer,
- identify the possible sources of data,
- delimit the study area,
- delimit the calendar.

Note that these delimitations could be modified at any moment during the complete process (our methodology follows the Agile trend). Thus, the goal here is to define area and calendar delimitations big enough in order to be sure not to miss documents that could be important later. It is the same for the questions that the model would have to answer. During this step, it is not mandatory to have precise questions: some questions could be deleted and modified throughout the complete process if necessary.

Concerning our case study, the main question that we wanted to answer about floods in Hà Nội in 1926 is: *what were the impacts of the decisions that have been taken during this event?*

For the sources of data, we identified 3 main sources:

– The National Archives Center #1 (Hà Nội, Việt Nam),
– The French mapping agency - IGN (Paris, France),
– The French School of Asian Studies - EFEO (Hà Nội, Việt Nam).

After a quick analysis of the documents, we proposed a first area delimitation: Hà Nội and its surrounding. Concerning the calendar delimitation, we choose to consider months around July 1926.

Once these required elements defined, the next sub-step consists in collecting, digitizing and indexing all the available data in the identified sources that have a link with the study question and that are consistent with the area and calendar delimitations. In our case study, we collected topographic maps of Hà Nội (and its surroundings) of 1925. In addition, we collected documents about the management of these floods. We got qualitative data about the water level and qualitative data about exchanges and communication between actors. More precisely, we used:

– The report of the Resident of France in Bắc Ninh (RFBN) with the chronology of events and the causes and responsibilities of each breach.
– The report of the Vietnamese Head of the district of Gia Lâm to the RFBN with the events, decision chains and dysfunctioning in order execution
– The report of the inquiry of the province judge about responsibilities of local actors in dyke breaches (in particular the district Vice-Head's responsibility).

Note that in our context, the collection process was complicated by the linguistic diversity of the documents, written in hán nôm (Sino-Vietnamese), French or quốc ngữ (romanized script). And, despite progresses in digitization, analyzing and indexing documents remains a largely manual and tedious task.

3.3 Step 2: Representation of the corpus

This step concerns the representation of the corpus. It is composed of two sub-steps: the first one is to create accurate GIS data from the disparate maps available in the corpus; resorting to recent sources (Digital Elevation Models, satellite imagery, etc.) if necessary. The second is to integrate temporal information and to offer the same querying and navigation facilities in time than the ones existing for 2D and 3D spatial data.

In the context of our case study, we build a GIS (see Figure 3) composed of following layers:

- buildings, source: vectorisation of the 1925 Hà Nội map (IGN),
- Red River, source: vectorisation of the 1925 Hà Nội map (IGN),
- lakes, vectorisation of the 1925 Hà Nội map (IGN),
- dykes, source: vectorisation of a map from National Archives Center #1,
- DEM (resolution of 10 meters), source: vectorisation of the 1925 Hà Nội map
  - contour lines (IGN).

The creation of GIS has required an important work of digitization and calibration of the data (in particular, for the dykes).
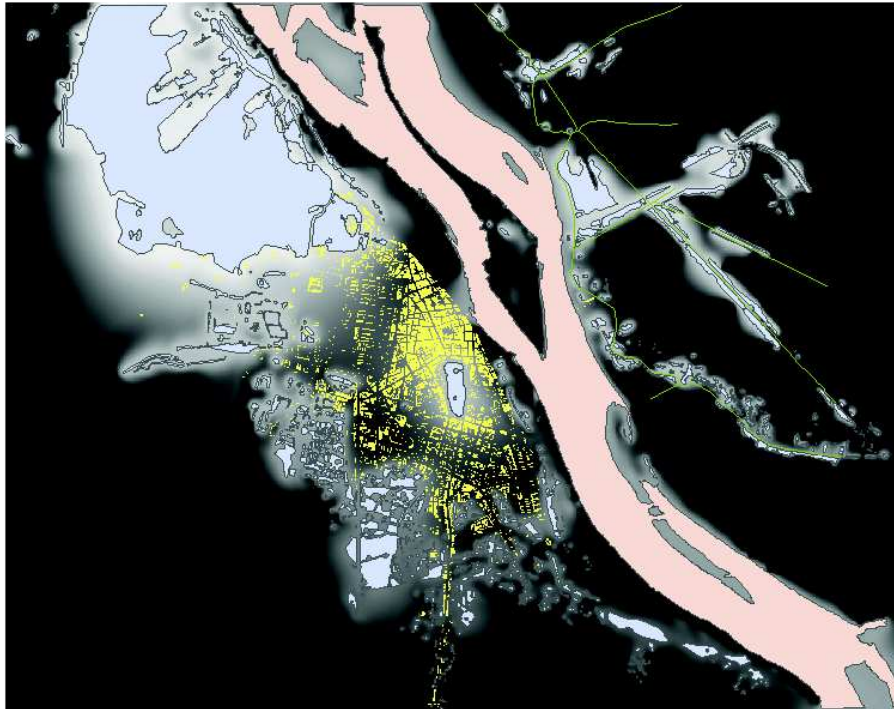

Figure 3. GIS built from historical data and maps

Concerning the second sub-step, we have developed a web-mapping application (Figure 4), based on Geoserver, allowing navigation through a web interface in time and space. This work required georeferencing all the documents in time and space and transforming these information into a GIS layer.
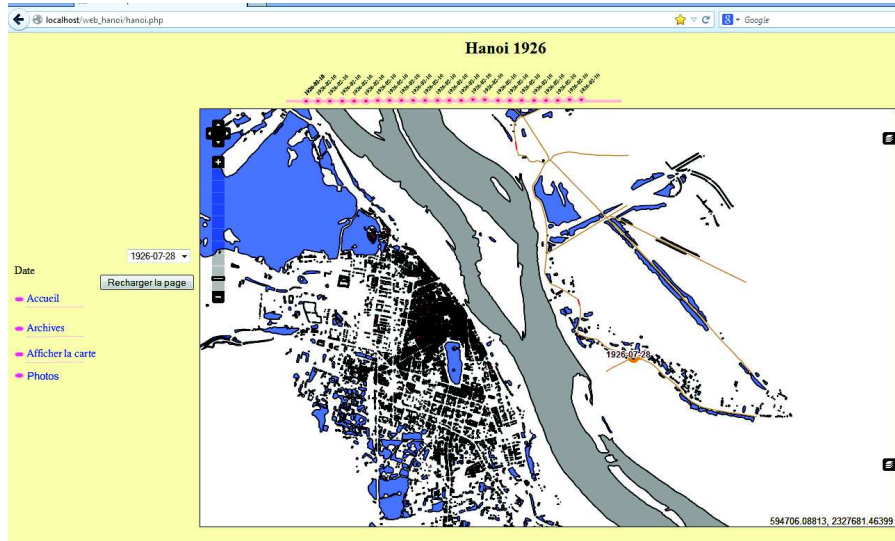
Figure 4. Snapshot of the web-mapping application

3.4   Step 3: Development of models

This step consists in developing a model according to the corpus that will answer the questions identified in Step 1. This model will rely on Agent-Based Modeling (ABM) as its core technology. ABM has been used over the last twenty years in a growing number of disciplines, including social sciences. This approach has now become the paradigm of choice to couple models from different domains into a same integrated model: in the case of flooding, for instance, hydrodynamic models representing the dynamics of the event itself need to be coupled with the models used to simulate human decision-making, taking into account the fact that each model can change the boundary conditions under which other models operate. A key issue of this step is to couple the GIS data with historical data and then to integrate them into the model to produce new knowledge. It is also very important that historians participate to this step. Indeed, our claim is that their implication and their understanding of the models produced is mandatory for the success of the modeling process. In this context the use of graphical formalisms like UML is very interesting as they are easily understandable by all.

In our case study, a difficulty comes from the representation of the various decision-making processes, as they rely on information on management and adaptation measures undertaken by the social actors before, during and after the event considered. This has implied the ability to identify these actors, to understand their individual roles and inter-individual relationships and to build models of their decision-making and communication processes. To this purpose, we relied on existing Artificial Intelligence techniques (such as the FIPA com-

munication language [2]). The challenge was to extract from the transcript of decisions and communications, the information related to the management and adaptation measures undertaken by the social actors.

4 Model

4.1 Purpose of the model

The model developed has for main purpose to study the decisions made by the different actors during the Hà Nội floods of 1926 and their impacts in terms of flooded area and causality.

4.2 Entities

To this purpose we have defined two kinds of entities: physical entities and actors. All these entities (with their attributes) are summarized in a UML diagram presented in Figure 5.

We have defined **physical entities** in our model that can be divided in 3 groups:

- Space entities: cells. A cell represents a partition of the space that is homogenous in terms of altitude and water level.
- Hydrological entities: rivers and lakes.
- Obstacle entities: buildings and dykes.

From data collected in the previous step, we have been able to identify all the individual **actors** involved in the floods management, i.e. all the actors mentioned in the (French) archives (cf. Figure 1). Among all these actors, we identify three main groups, with different roles in the flood management:

- Decision-makers, who include the French administration, Resident of France in Tonkin and in Bắc Ninh province, and central technical services. These actors are the senders of about 86% of all the messages.
- Order transmitters, i.e. actors involved in the Vietnamese hierarchy from the province to the commune level.
- Order executors, i.e. the manpower that practically execute the orders.

This distinction has driven our modeling of actors. As we focus on the management of the event during the crisis, i.e. the reaction to it, we do not need to model high-level decision-makers, as they are mostly involved in planning phases. Among order transmitters, we simplify by considering only one actor at each level (province, district and township, denoted as local authorities). Finally, we choose to introduce a kind of agent "commune" that will gather the local authority, its manpower and the dykes the local authority is responsible of. In addition this kind of agent will be an intermediate between actors and the hydraulic part of the model. This modeling choice is based of several hypotheses

that we made on this model: we do not represent each coolie and villager individually; in particular we consider they do not have an autonomous behavior (they cannot disobey). In addition, we consider that manpower is not a blocking resource, so we do not represent it explicitly. These hypotheses are of course strong, they will be discussed in conclusion.
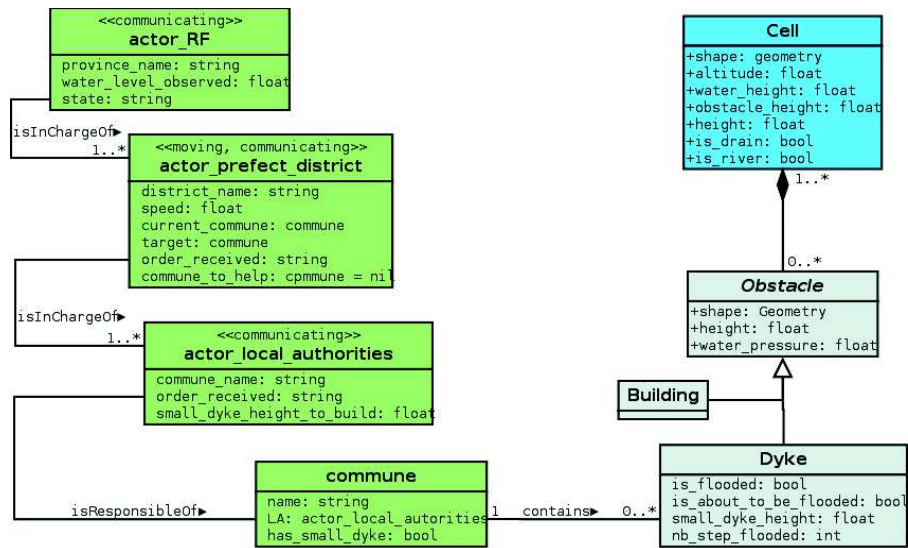

Figure 5. UML class diagram of the model

4.3   Processes

Hydrological process. Due to the lack of precise data concerning the Red River at the study period (flow, cross section...), we chose to use a simple flowing model based on a grid (the cell entities). Another advantage of this type of models is to be more easily understood by historians.

Each cell represents an area which has homogenous altitude and water level. In addition, it has a height variable that corresponds to the sum of its altitude, water level and max height of its obstacles. At each simulation step that represents 2 hours, all the cells are sorted by height. Then, from the lowest to the highest, each cell that contains water (water level > 0) diffuses its water to the lower cells (in terms of height) in its (Moore) neighborhood. This diffusion works as follow: the cell sorts by height all its neighbor cells that are lower than it. Then, from the lowest to the highest, it diffuses to this cell a quantity of water that corresponds to the difference of height between the two cells.

We also implemented a dyke destruction process: if the water level on the cell containing the dyke is not null (meaning that there is water above the dyke) for at least 12 consecutive simulation steps, the dyke breaks (the dyke agent dies).

Crisis management. Given the hydraulic model presented above, we introduce a model dealing with the crisis management including the actors introduced above. This model will mainly represent the way actors fight against floods (in particular against the water level increase) and the way orders and information are transmitted through the actor network. We thus aim at investigating how orders and information transmissions induce the building of small dykes.

To this purpose, the main process we consider is the top-down order chain. Basically, when the Red River water level is higher than a given threshold, the Head of the province will order to Heads of district to build small dykes. Heads of districts will thus move to meet local authorities actors, order them to build small dykes (with a given height) and supervise works that will be performed by villagers or coolies. Then the Head of districts will move to another commune and so on. This pattern has been (manually for the moment) identified in several exchanges in archives.

4.4   Implementation of the model

The model was implemented with the GAMA platform[1] [4]. This generic modeling and simulation platform is particularly well-suited to our application context as it allows to simply integrate GIS data and offers the possibility to use FIPA-ACL [2] for agent communication. In addition, its integrated agent-based modeling language (the GAML), which is easy to understand, has allowed to ease the exchange between the different participants of the project.

The implemented model offers a realistic rendering of the city of Hà Nội in 1926 (Figure 6). It allows users to interact with the simulation at runtime to see the impact of a dyke breaking. In addition, it provides a multi scale online analysis of the flow of messages occurring during the flood event. Messages coming from different parts of the administrative hierarchy is represented as a new layer on the already existing representation of the city as shown in Figure 6 where aggregated and instantaneous graph are displayed on top of Hà Nội map.

5   Results and discussion

First results are very encouraging: we are able to reproduce in a realistic way from historical documents both the physical phenomenon (the floods) and its management by authorities (in a given time scale). They tend to show that, even in an ideal situation from a social point of view, without disobeying villagers or local authorities who do not execute orders, submersion of dykes could not have been avoided. This reinforces conclusions of some of the people involved in the
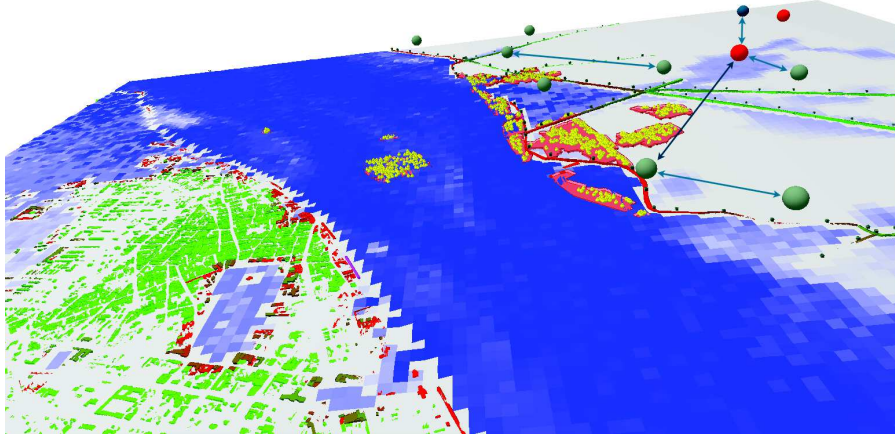
---

[1] http://code.google.com/p/gama-platform/

Figure 6. Snapshot of the model implemented in GAMA

crisis management that there is no human responsibility in the set of breaches in the dykes during these floods in contrary to the French inquiry conclusions.

A very important outcome of this work were the reflections it generates about the role and the interest of simulation for historical research and Digital Humanities in general. It is important to notice that we have not exactly reproduced the events occurred in 1926 (in particular in terms of location of dyke breaches). We built possible and realistic histories that have been produced by the processes that occurred during the actual History. Reproducing exactly the film of the actual events can have an interest to visualize in a 3D realistic environment what happened, but this would be possible only by scripting the simulation. From our point of view, this holds very little interest because this precludes us from testing various hypotheses on the model and the effects they could have had.

6   Conclusion

In this paper, we presented the ARCHIVES project that has for goal to propose a methodology to model historical events to better understand them. ARCHIVES aims at providing historians with new tools allowing them to not only replay historical events, but also to explore *what could have happened* in alternative "what-if" scenarios.

We already carried out a first implementation of ARCHIVES for the floods of Hà Nội (Việt Nam) in 1926 that were particularly destructive.The model built has allowed us to test several scenarios concerning the chain of commands (ideal situation vs actual one) and to extract knowledge from it. The next step of ARCHIVES will be to generalize this case-study in order to provide historians with generic tools allowing them to carry out this modeling work by themselves.

In this context, we propose first to work on digitization tools to support the historical document collection and indexation. We propose then to work on the integration of graphical modeling features inside GAMA. The goal is to allow historians - that have most of the time no programming skills - to develop the models by themselves (or at least the main part of the models).

7   Acknowledgements

References

1. J. S. Dean, G. J. Gumerman, J. M. Epstein, R. L. Axtell, A. C. Swedlund, M. T. Parker, and S. McCarroll. Understanding Anasazi culture change through agent-based modeling. In T. A. Kohler and G. J. Gummerman, editors, *Dynamics in Human and Primate Societies*, pages 179–205. Oxford University Press, UK, 2000.

2. Foundation for Intelligent Physical Agents. FIPA Communicative Act Library Specification, 2002. URL: http://www.fipa.org/repository/aclspecs.html.

3. P. Gourou. *Les paysans du delta tonkinois.* Publications de l'Ecole Française d'Extrême Orient, Les Editions d'art et d'histoire, Paris, 1936.

4. A. Grignard, P. Taillandier, B. Gaudou, D. A. Vo, N. Q. Hyunh, and A. Drogoul. GAMA 1.6: Advancing the art of complex agent-based modeling and simulation. In G. Boella, E. Elkind, B. T. R. Savarimuthu, F. Dignum, and M. K. Purvis, editors, *PRIMA 2013*, volume 8291 of *Lecture Notes in Computer Science*, pages 117–131, Dunedin, New Zealand, 2013. Springer.

5. S. Heckbert. MayaSim: An agent-based model of the ancient maya social-ecological system. *Journal of Artificial Societies and Social Simulation*, 16(4), 2013.

6. S. Moss and B. Edmonds. Towards good social science. *Journal of Artificial Societies and Social Simulation*, 8(4), 2005.

7. O. Tessier. Outline of the process of red river hydraulics development during the nguyen dynasty. *Environmental Change, Agricultural Sustainability and Economic Development in the Mekong Delta*, 45:45–68, 2011.

8. O. Tessier. Hydrological Development of the Red River Delta: A Historical Perspective of the Role of the Imperial then Colonial State (From the XIIth Century to the First Half of the XXth Century). In S. Lagrée, editor, *Water and its Many Issues. Methods and Cross-cutting Analysis*, Regional Social Sciences Summer University, chapter 1.6, pages 130–154. Journées de Tam Dao, 2013.