ARIMA

épisciences

Combining Scrum and Model Driven Architecture for the development of an epidemiological surveillance software

Azanzi Jiomekong[*1], **Hippolyte Tapamo**[1], **Gaoussou Camara**[2]

[1]UMMISCO, Faculty of Sciences, University of Yaounde I, Cameroon
[2]EIR-IMTICE, University Alioune Diop de Bambey, Sénégal

*E-mail : fidel.jiomekong@facsciences-uy1.cm

DOI : 10.46298/arima.9873
Submitted on July 31, 2022 - Published on June 21, 2023
Volume : **Volume 39** - Year : **2023**

Editors : *Nabil Gmati, Mathieu Roche, Maurice Tchuenté, Bruce Watson*

Abstract

Epidemiological surveillance systems evolve with time, depending on the context and the data already collected. Then, the software used must evolve in order to meet requirements. However, introducing new requirements in order to update the software takes time, is expensive and may lead to the problem of software regression. The problem of failed software developed for epidemiological surveillance are often the result of an unsystematic transfer of business requirements to the implementation. This problem can be avoided if the system is established using a well-defined framework/architecture permitting the rapid development/update of the surveillance software. Empirical research shows on the one hand that Model Driven Techniques such as Model Driven Architecture (MDA) are more effective than code-centric approaches for the development and the maintenance of software. On the other hand, Agile Processes such as Scrum are more effective than Structured Processes when requirements are subject to frequent change. Researchers demonstrated that developers of medical software such as epidemiological surveillance software are experiencing difficulties when following Structured Processes and code-centric approaches. The main goal of this empirical study was to apply the combination of Scrum and Model Driven Architecture for the development of epidemiological surveillance of tuberculosis. During this research, we found the approach ease of use and very useful when the MDA tool can generate the complete source code. It has had positive effects on programmer productivity and satisfaction, cost-effectiveness, timelines and customer satisfaction. In addition, we learned that to involve non-informatic experts in the development/update, the modeling user interface must be as simple as possible.

Keywords

Tuberculosis; Epidemiological surveillance systems; Agile; Scrum; Model Driven Engineering; Model Driven Architecture; EPICAM

I INTRODUCTION

Over the centuries, communicable diseases have always been one of the main problems for human health because they spread rapidly and result in high mortalities. Currently, respiratory infectious diseases such as Influenza, Coronavirus disease, Tuberculosis (TB) are in the top of the most killers globally. Effective management of these diseases requires to put in place epidemiological surveillance systems. Epidemiological surveillance systems enable the collection, analysis, and interpretation of data, closely integrated with the dissemination of these data to the public health practitioners, clinicians, policy makers and the general population for preventing and controlling disease and injury. To be usable, epidemiological surveillance systems use surveillance software to provide timely, accurate, comprehensive and update information to stakeholders [28, 49, 54–56]. Depending on the data gathered on the field, the epidemiologists may need to collect a new parameter in order to explain a phenomenon. For example, collect the data on the height of the patients in order to calculate their body mass index and understand why some of them are dying the most. This task may be done by using supplementary materials such as paper forms or spreadsheet software and lead to the problem of data integration [6]. In fact, at the end of data collection, the data obtained using supplementary materials must be put together with the other data. To collect new data, another solution consists of updating the software used for the surveillance. This solution is expensive and may lead to the problem of software regression [61] because bugs can be introduced during the source code update. These problems occur because of the problem of failed software developed for epidemiological surveillance.

The problem of failed software in the health domain is often the result of an unsystematic transfer of business requirements to the implementation [44, 47]. In the case of epidemiological surveillance, this problem can be avoided if the system is established using a well-defined framework/architecture permitting the rapid development/update of the surveillance software by non-informatics experts such as health workers.

Model Driven Techniques in general relies on models and model transformations for the development, maintenance and evolution of software [40, 48, 64]. Through the architectural separation of concerns, the Model Driven Architecture (MDA) in particular assumes the evolutivity, portability, interoperability and reusability [10, 15, 68]. Graphical modeling permits non-informatics experts to be able to build graphical models and automatically generate their software. Empirical research shows that the main motivations for companies to use Model Driven techniques are cost saving, rapid development / update / maintenance, improvement of productivity, software quality, and the maintaining the architecture consistency from analysis to implementation [32, 33, 39, 59]. Model Driven Techniques experience incredible success in many domains such as automotive, business process engineering, and embedded systems [64]. In the health domain, many authors has proven that Model Driven techniques is an effective approach for the development of health software [11, 14, 20, 47, 53, 57, 62].

Translating requirements to a solution using either code-centric or MDA approach requires the use of a software development process. Software development process defines a scheme to structure and manage the various aspects of the development which are requirements elicitation, design, implementation, verification, maintenance. Software development processes are grouped into Structured Processes such as the waterfall or V-model and Agile Processes such as Scrum [46]. With Structured Processes, specifications are fixed in advance and the software is developed and delivered thereafter. In the health domain, empirical research demonstrated that software developers experience difficulties when following Structured Processes [31, 38,

44, 65]. In the particular case of epidemiological surveillance of diseases like Covid-19, it can be very difficult to capture all the requirements at such an early stage of a project. In addition, any change introduced once a project is underway can create cost and budget overruns. However, Agile software development processes such as Scrum [9] experience great success in software development organizations [29, 42, 45]. Many empirical studies report significant benefits being gained from utilising Agile practices such as reduce costs, reduce time to market and increase quality [31, 34, 38, 65].

Many authors have written about the benefits of deploying Agile [21, 27, 30, 31, 34] on the one hand and Model Driven Techniques [11, 14, 20, 24, 53, 57, 62] on the other hand to develop health software. But, unfortunately, the percentage of empirical studies be them surveys, experiments, case studies or postmortem analyses that provide data to illustrate the effects of the combination of Agile and MDA approaches over different quality characteristics such as productivity of the development team and satisfaction of stakeholders is still very low, which hampers the generalizability of the results. Thus, our main contributions are:

- to provide empirical data that contributes to giving a scientific answer to the following research question: how to combine Scrum and MDA to develop and maintain epidemiological surveillance software?
- provides a framework based on MDA that can be used for the rapid development of any epidemiological surveillance system.

The remainder of the paper is organized as follows: Section II presents the research context; Section III presents the MDA approach and Scrum Process; Sections IV and V present respectively the research design and the research results; Sections VI and VII present respectively the threats to validity and related works. We end with the conclusion in Section VIII.

II RESEARCH CONTEXT

The EPICAM[1] (Epidemiology in Cameroon) project aims to improve Cameroon's health system by strengthening the Health Information System. Given the problem of failed software developed in the health domain [44, 47], the project of academic (University of Yaounde I in Cameroon[2]), clinical (fifty hospitals in Cameroon), epidemiological (Centre Pasteur du Cameroun[3] - Epidemiology and Public Health department and NTCP), and industrial (MEDES[4] in France) partners were set-up. It aims to provide methodologies, methods, models and tools for improving the development of epidemiological surveillance of infectious diseases. Firstly, a pilot project consisting to set-up the epidemiological surveillance of Tuberculosis (TB) was put in place. In effect, given the burden of TB in Cameroon, the government has put in place the National Tuberculosis Control Program (NTCP). Thereafter, the NTCP has set-up epidemiological surveillance. Even if some tools like Microsoft Excel are used, this system is mainly manually managed. Given the limit of the manual management [13, 62], the NTCP expressed the need for an electronic system. Thus, the purpose of the pilot project was to describe the methodologies, methods and tools for the development of epidemiological surveillance of TB and to investigate factors considered as important for its adoption.

Given the limits of the code-centric approach [39] and the structured process [31, 38, 44, 65] for software development, Scrum as the process which aims to describe step by step the devel-

[1] http://www.medes.fr/fr/nos-metiers/la-e-sante-et-l-epidemiologie/la-tele-epidemiologie/projet-epicam.html
[2] https://uy1.uninet.cm
[3] https://www.pasteur-yaounde.org
[4] http://www.medes.fr

opment of epidemiological surveillance software and the MDA approach which aims to define models and automatically generate the software were adopted.

III MDA AND SCRUM

Epidemiological surveillance software can be developed on the one hand using a code-centric or a MDA approach. With the code-centric approach, generally, the developers use languages like Unified Modelling Language (UML) and methods like Unified Process (UP) or Merise in the analysis and design steps. The conceptual model such as class diagram/Entity Relational model is usually designed at once and used to develop the software [10]. With the MDA approach, the conceptual model such as Platform Independent Model (PIM) and Platform Specific Model (PSM) are used to automatically generate the executable source code [8, 15]. Then, we choose MDA in this research because related research [39] shows that it improves programmers productivity and satisfaction, shortens development time, improves software quality, and promotes reuse of tools and models and saving costs.

On the other hand, whether code-centric or MDA approaches are used during the development of a software, there is a need for a software development process [31, 38, 65]. Given that epidemiological surveillance systems evolve faster, the recommended process is Agile. While several agile development processes have existed for some years, a large number of software organizations have implemented agile processes such as Scrum [65]. Then, in this research we used Scrum.

In the next subsections, Model Driven Architecture (Subsection 3.1) and Scrum (subsections 3.2) are presented.

3.1 Model Driven Architecture

Model Driven Engineering (MDE) is a software development approach based on the use of models to develop software artifacts. Its goal is to improve the productivity and maintainability of software by raising the level of abstraction from source code to high-level domain-specific models such that developers can concentrate on application logic rather than implementation details [8, 33, 48, 64]. It is identified by several paradigms whose Model Driven Architecture (MDA). MDA [8, 10, 15, 40] is promoted by the the Object Management Group (OMG)[5] as an architecture and a framework for software development based on models.

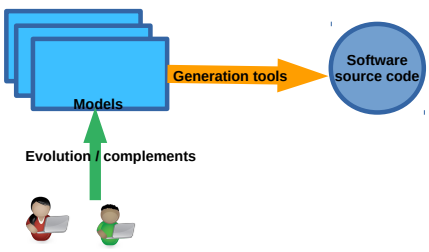

Figure 1: MDA Software development approach

Globally, the MDA approach (Figure 1) is an approach to software specification, design, and implementation which provides guidelines for structuring software specifications that are ex-

[5]https://www.omg.org/

pressed as models. It focuses on forward engineering in which the executable source code is (semi)automatically generated from abstract, human-elaborated modeling diagrams such as a class diagram [20, 47, 53, 57]. A class diagram describes the structure of a domain by identifying the domain classes (e.g., patient), their attributes (e.g., age, sex), their operations (e.g., calculate a body mass index) and the relationships amongst classes (e.g., the relation between a patient and his/her appointments at the hospital) [14, 18]. Models are used throughout the development life cycle of an application, that is, from the application specification to code generation. For the generation task, the MDA approach proposes to provide developers with automation tools [10, 23]. When adopting the MDA approach, a natural counterpart is the framework which is the body of code that implements the aspects that are common across an entire domain. This framework allows us to link the business level and the information technology [15, 47].

The two key concepts of MDA are models and transformations. All the models are well defined by a modeling language, itself defined by syntax and semantics in a meta-model. The transformation from one model (e.g., requirements) to another model (e.g., design) is a systematic process explicitly defined by transformation rules. During the development of applications using the MDA approach, the business models as a result of the discussion between domain experts and modeling experts are (partially) automatically transformed through the three models layers ($CIM \longrightarrow PIM \longrightarrow PSM$) [14, 15]:

- **Computational Independent Model (CIM):** presents exactly what the system is expected to do. It hides all the information technology related specifications to remain independent of how that the system will be (or currently is) implemented;
- **Platform Independent Model (PIM)**: obtained from the CIM, the PIM is exactly what its name denotes: a model that is free of specific hardware or computing platform requirements and constraints. An example of a PIM is the class diagram modeling the domain and presenting the entities and the relationships between these entities;
- **the Platform Specific Model (PSM):** The PIM constructed can be transformed (manually or automatically) into one or more PSM for different target platforms by integrating platform specific details to the PIM;
- **Executable source code:** The PSM is used to generate the software solutions which are application source code, configuration files, component descriptor files, deployment files, build scripts, documentation, etc. Depending on the maturity of the MDA tool-set, code generation varies from significant to substantial or, in some cases, all the source code is generated.

Model Driven Techniques make a greater impact in the software engineering domain. Researchers claim several advantages over code-centric approaches among which [32, 33, 39, 59, 64]:

- Improve productivity and shorten development time by increasing automation (generating code and other artifacts from models) and promoting reuse. For instance, Martínez et al. [39] demonstrate that maintaining Web applications using Model Driven Techniques improves the performance of developers.
- Improve software quality by: earlier detection of bugs, maintaining the architecture consistency from analysis to implementation, improving communication and information sharing among stakeholders, and enabling traceability throughout software development artifacts.

In the domain of health, the literature review shows that the MDA approach has been widely used [11, 14, 20, 47, 53, 57, 62]. To cite a few:

- Hajou et al. [44] showed how Model Driven techniques can be used to overcome the difficulties of adapting pharmaceutical software to the standard by proposing a rapid update of pharmaceutical products;
- Schlieter et al. [47] showed how an MDA approach can be used in order to build a method, fundamental for a systematic creation of an application system. Their use-case is a software for IT based workflow support for an interdisciplinary stroke care project;
- Curcin et al. [36] presented a work on the use of MDA to build tools which allow non-IT experts to model their requirements and to generate data collection tools for medical organizations;
- Blobel and Pharow [16] presented a work done in Germany consisting of the establishment of a health telematics platform combining card enabled communication with network based interoperability;
- Jones et al. [11] presented the use of MDA to develop m-health (Mobile-health) application providing an integrated set of personalized health-related services to users;
- Rayhupathi and Umar [24] proposed the use of MDA to develop a system capable of tracking patient information.

Concerning the domain of epidemiological surveillance, we found some tools such as Imogene[6], Magpi[7], Open Data Kit(ODK)[8] generally used for the modeling and the generation of data collection software. However, epidemiological surveillance software involved more features.

3.2 Scrum

Agile software development Processes started officially when the agile movement presented the agile manifesto in 2001. The agile manifesto contains the central values which proposes to place more emphasis on people, interaction, working software, customer collaboration, and change, rather than on processes, tools, contracts and plans during software development [4, 29, 42, 45].

Agile Processes rely on an iterative, incremental and adaptive development cycle which considers that the needs cannot be fixed and proposes to adapt the development to the changes. It's intended to support early and quick production of working code. In the particular case of health information, it allows fast deployment and adoption [21, 30, 31, 34, 44, 53, 65]. It consists firstly of setting a short-term goal and of getting into the project without delay. Once the first objective is attained, we take a short break and adapt the path according to the situation of the moment, and so on, until we reach the final destination. This helps us to give more visibility, involving the client from the beginning to the end of the project and adopting an iterative and incremental process. In a broader sense, an Agile Process can be seen as a process consisting of an initialization step, iterative steps and incremental steps [4, 7]. While several agile development processes have existed for some years, a large number of software organizations have implemented agile processes such as Scrum [65].

The first reference of the term "Scrum" originates from Japan and refers to an adaptive, quick, and self-organizing product development. Scrum describes how the team members should be organized in order to develop a system using Agile principles [9].

Scrum Process includes three steps: Pre-Development, Development and Post-Development.

[6] https://github.com/medes-imps/imogene
[7] www.magpi.org
[8] https://opendatakit.org/

- **Pre-Development:** the Pre-Development step includes the planning and the design. The planning is done based on the Product Backlog, in collaboration with the customer. The requirements are prioritized, the resources (team, tools, etc.) needed for their implementation are estimated, risks assessed, identification of training needs and verification management approvals made. The design phase is done by defining the architecture based on the Product Backlog List; preliminary plans for the contents of releases are proposed. The design meeting permits us to make decisions about the implementation and preliminary plans for the contents of releases are prepared;
- **Development:** during the development steps, the system is built in Sprints during which the functionalities are developed or enhanced to produce new increments;
- **Post-Development:** this phase is reached when the requirements are completed. The system is ready for the release and tasks such as system integration and documentation are made.

Systematic literature review papers demonstrate the relevance of Agile in industry [4, 29, 42, 45]. Many empirical studies report significant benefits being gained from utilizing Agile practices such as reduce costs, reduce time to market and increase quality [31, 38, 65]. In addition, in October 2012, the Association for the Advancement of Medical Instrumentation (AAMI) released a Technical Information Report (TIR) entitled "Guidance on the use of agile practices in the development of medical device software [27]". The AAMI recognised the shift in the non-regulated software development industry towards more agile practices and the evidence presented from successful adoption of agile practices in medical device software development organizations.

The use of Agile Processes in the health domain is an old practice. Many authors presented its adoption for the design and implementation of health software [21, 30, 31, 34, 44, 53, 65].

However, these studies use Agile to develop health software with the code-centric approach. In this study, we choose to use Agile to develop software using the MDA approach. The next section presents the research design.

IV RESEARCH DESIGN

The research design was inspired by papers on empirical research methods in software engineering [17, 66]. They describe a number of empirical methods available for software engineering research, examine the goals of each method and analyze the types of questions each best addresses. Theoretical stances behind methods, practical considerations in the application of the methods and data collection are also reviewed. Taken together, these information provide us with a suitable basis for both understanding and selecting a research method from the variety of methods applicable to empirical software engineering research. These works allowed us to design our study. Thus, we designed the research question (Section 4.1), stakeholders involved in the research (Section 4.2) the research methods (Section 4.3), data collection and analysis (Section 4.4), and how the models were constructed and transformed 4.5).

4.1 Research question

On the question on how to develop epidemiological surveillance software, evidence shows on the one hand that Agile Processes are more effective than Structured Processes when the specification cannot be fixed at the beginning of the project [31, 38, 65]. On the other hand, MDA is more effective than code-centric approach during the development and the update/maintenance of the software [32, 39, 59]. Then, given the problem of failed software, our intuition

is that: "meshing Scrum and MDA improves the development of epidemiological surveillance software". This led to the following research question: how to mesh Scrum and MDA to develop and maintain epidemiological surveillance software? We have used the term meshing as the common term for combining, mixing, and coexisting.

4.2 Stakeholders involved

The research team consisted of two persons playing the role of researchers and software developers. Both had a strong background and experience in Scrum and Model-Driven Engineering Techniques and tools. The target population of the research was a big stakeholder group including:

- **The product owner:** working at NTCP, he is the member of the Scrum Team responsible for the prioritization of the features to develop. He also plays the role of interchange belt between the software team and the others stakeholders;
- **The NTCP Staff:** this is the technical staff at the NTCP who are going to participate in the development and the maintenance of the surveillance software. They assists to the Scrum meeting and will be trained on the update of the system;
- **End users:** they are the health workers at the different levels of the health system. They are the final users of the platform developed. They used and validate the different versions developed before the release of the final product;
- **Two epidemiologists:** one epidemiologist from the NTCP and one epidemiologist from Centre Pasteur du Cameroun (CPC). It should be noted that the Service of Epidemiology of CPC monitored many diseases including influenza, HIV, rabies, etc.

The NTCP represented by the Product Owner led the recruitment of the NTCP staff and the end users.

4.3 Research methods

Our aim in this research as presented by the research context (Section II) was to do intensive research on a specific case of the development of epidemiological surveillance software in order to provide a deeper understanding. Then, we choose the case study method because it is suited for understanding the complex real-world setting [22, 25, 65, 66]. The selection of the case of epidemiological surveillance of TB was made because of the NTCP's interest to replace manual management with software management. However, to understand the current system, we thought it necessary to investigate the opinions (through survey) of the stakeholders about the current system and its replacement with a software approach. Stakeholders allowed us to find a real problem that health software generally faced: this is the problem of failed software mainly caused by the fact that requirements cannot be fixed. In this situation, we completed our methodology with action research. In effect, action research is used in real situations where major challenges cannot be studied without implementing them, and where implementation implies a long term commitment because effect may take time to emerge [17, 35, 63, 66]. In addition, we were having the NTCP willing to collaborate to both identify the problem, and engage in an effort to solve it. Combining empirical research methods is largely accepted in empirical software engineering research [3, 5, 17, 66]. The research methodology is presented by the Fig. 2.

This methodology is composed of the Pre-Intervention Phase, the Intervention Phase and the Post-Intervention Phase. During the Pre-Intervention, the problems at the NTCP are identified, the method to solve these problems and the list of interventions to implement are defined. During the Intervention, the solution to the problem is built and during the Post-Intervention, this

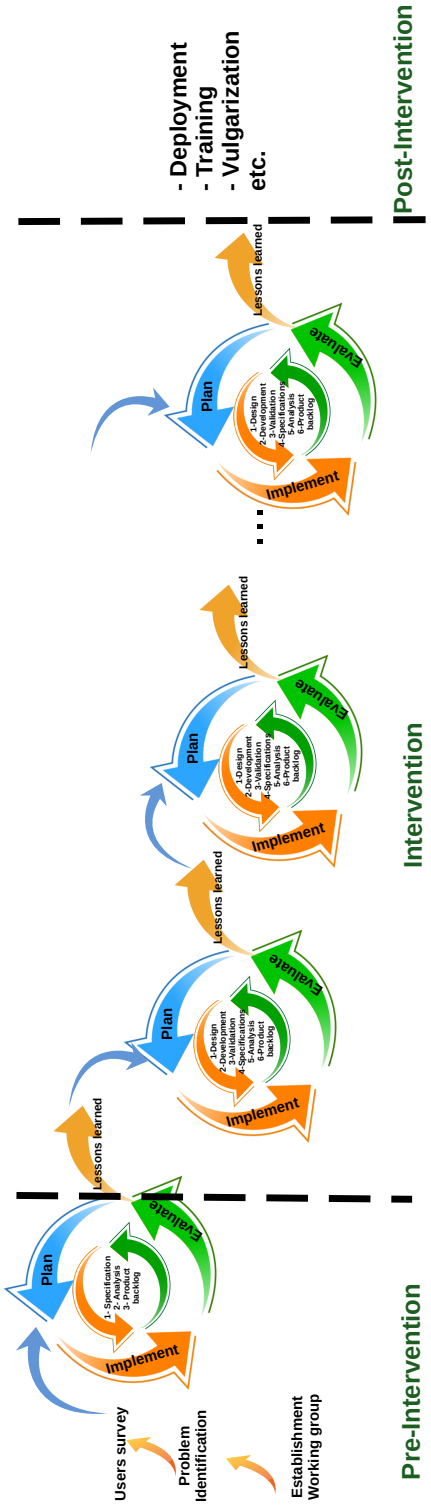

Figure 2: Research methodology

solution is deployed. In the following paragraphs, we present the different research methods and how they were used.

4.3.1 Case Study

The aim of studying the case of the development of epidemiological surveillance software was to portray this development so that it can be reproduced to other epidemiological surveillance. Then, we choose the case study method because it is suited for understanding the complex real-world setting [22, 25, 65, 66]. The selection of the case of epidemiological surveillance of TB was made because of the NTCP interest to improve epidemiological surveillance and the researcher's interest in understanding the effects of development approach and process during the development. Thus, our aims were to portray and provide a deeper understanding of the development of epidemiological surveillance of TB using a software development approach (either code-centric or MDA) and a software development process (either Structured or Agile). Section 4.5 provides the justification elements of the choice of Scrum and MDA.

4.3.2 User survey

Putting in place an electronic surveillance network requires the collection of user needs, a good analysis of the existing network, information and knowledge of the users in the use of Information Technology (IT) tools. Therefore, to put in place the electronic surveillance of TB in Cameroon, we collected the opinions of the final users at all levels of the NTCP. To this end, we conducted the survey to better document the problem so that the system deployed will be accepted.

The main goal of this survey was to find out what the stakeholders think of the replacement of manual management by software management. The specific objectives were:

- Understand the current network,
- Know how stakeholders collaborate,
- Collect the specific needs of users in the field to complete the specifications already gathered,
- Determine the computer skills of users of the current surveillance network,
- Know the opinion of final users about an electronic network.

It should be noted that survey research is accepted in empirical software engineering research [2, 17, 66]. In the next section, we present action research.

4.3.3 Action Research

Action research is an iterative process in which plans are created, implemented, revised, then implemented, lending itself to an ongoing process of reflection and revision. During this process, findings emerged as interventions are implemented. It is used where major challenges cannot be studied without implementing them, and where implementation implies a long term commitment because effect may take time to emerge [12, 17, 35, 66]. It was successfully used by Adam Alami et al. [63] to study how to implement changes in Free and Open Source Software (FOSS) communities. Thus, the researchers joined with the FOSS community and proposed solutions for improving and enhancing FOSS practices. The study included interviews with FOSS developers in order to seek solutions to their problems.

Like in the case of epidemiological surveillance, there was a clearly identified group of stakeholders that suffered problems. Thus, we collaborated with the NTCP which were using manual

tools for epidemiological surveillance in order to identify the problems they have and the challenges to effectively replace this manual management by the software approach. To this end, we needed a method and a process that fully engaged the stakeholders. That is why we set-up an action research method in which we as researchers and the NTCP as the problem owner were engaged collaboratively to identify the problems and develop effective solutions. This process is constituted by a set of consecutive actions and reflections forming a collaborative approach to research whereby the researcher partners with the subject being studied to achieve a particular outcome [12].

As presented by the Fig. 2, the action research we applied consisted of a set of aggregated interventions to develop epidemiological surveillance software of TB. These interventions involved a series of actions taken during the development of the epidemiological surveillance of TB in order to come up with an effective system. It is organized into the Pre-Intervention, the Intervention and the Post-Intervention phase.

Pre-intervention phase. During the Pre-Intervention, we identified the problems and proposed solutions. These solutions are divided to give a list of interventions. Thereafter, we conduct an empirical investigation of the research environment by interviewing health workers at all the levels of the health system. This empirical investigation allows us to identify problems and issues related to the current system. At the end, we identified a shared practical problem that was presented to the NTCP staff. Globally, the Pre-Intervention is composed of the following interventions (see Fig. 2):

1. Establishment of the working group: this is a group of people involved in the whole research;
2. Problem identification: this is the identification of the problems of the current system;
3. Users survey: this is the survey of users of the current system;
4. Specification: this is the description of users needs;
5. Analysis: after the description of users' needs the analysis allow us to well understand these needs;
6. Identification of the list of interventions: once the system is well understood, the list of tasks to implement in order to provide the solution is defined.

Intervention phase. Involves the development of the solution to the problem. It is cyclic, implemented through iterations of planning, implementing, validation, and reflecting [12, 35]. During the Intervention Phase, the list of interventions identified during the Pre-Intervention Phase are implemented. These interventions are executed in iterations (see Fig. 2). Each iteration is composed of the planning, the implementation, the validation, the reflection.

1. The planning consists of selecting a list of intervention and planning their execution. The planning is designed and enhanced iteratively. It is obtained from requirements, precedents implementations, problems encountered, etc. Once these elements are identified, the list of interventions is updated and the planning is updated.
2. The implementation consists to implement all the interventions planned during the planning.
3. The validation consists of the confirmation that the result of the implementation corresponds to the outcome. During the implementation and the validation of the solution, new learning may emerge.
4. The reflection consists of drawn lessons, revise and prioritize the list of interventions.

The iteration nature of the process provides us opportunities for improvements to the knowledge, methods, and better understanding of the approaches and methods.

	Research method	Data collection	Data analysis
Pre-Intervention	Survey and Action research	Interviews	Open-coding
Intervention	Case study and Action research	Observation and Interview	Thematic coding
Post-Intervention	Case study and Action research	Interview	Thematic coding

Table 1: Data collection and analysis methods

Post-Intervention Phase. Once the Intervention Phase is completed, Post-Intervention takes place. It includes all the tasks to implement (deployment, training end users, vulgarization, etc.) in order to make the solution available to the end users.

4.4 Data collection

To collect data, we used two data collection techniques: semi-structured interviews and observations. The data were recorded by taking notes in structured form in a notebook. These notes were supplemented with pictures (for instance, the photos of data collection forms). After the data collection, the analysis consisted of labeling and organizing the data in order to identify codes, themes and relationships between them (the summary is given in table 1). It should be noted that in case study research, many authors acknowledge the use of interviews for data collection [19, 25, 65]. Interviews and observations were used for data collection during the action research conducted by Adam Alami et al. [63]. Survey data are collected either using quantitative data or qualitative data such as interviews [3, 5, 17]. In the next paragraphs, we present the data collection during the different phases of the research.

Pre-intervention. The data collection during the Survey was made using semi-structured interviews. The participants were health workers involved in the epidemiological surveillance of TB at the hospital, regional and NTCP levels. These interviews were conducted by the two researchers and one epidemiologist. Each interview lasted between 30 to 45 min and was recorded in a notebook. The text and conclusions were presented to the NTCP staff.

The following open-ended questions were asked to participants:

- How do you manage your data?
- What are the main problems you generally have? What are the solutions you generally use to solve these problems?
- What is your knowledge on the use of IT tools?
- What do you expect from the use of a software for epidemiological surveillance?

Alongside the data collected from the health workers, other sources of data used in the research were the supports that are used at NTCP for TB surveillance.

To analyze this data, we started from scratch and created codes based on the data. Thus, we go through the transcript of every interview and highlight everything that jumps out as relevant or potentially interesting.

Intervention. During the Intervention phase, the data were gathered by watching how the combination of Agile and MDA is being used during the development of epidemiological surveillance of TB. Thus, data from daily observation of the implementation of the interventions was collected in order to gain an understanding of the software development phases and decisions. These observations were documented by note taking in notebook. In some cases, we used semi-structured interviews to investigate a phenomenon. For instance, we observed that only one health worker investigated himself in the use of the modeling tools. To understand this phenomenon, a semi-structured interview was conducted.

Then, through self-assessment, we determined how things went and how they can be improved. To this end, after each Sprint, the following questions were answered:

- What went well? - we determine if what has been done can be replicated in another context.
- What didn't go well? - we acknowledge mistakes and pledge improvements.
- What can we do to improve? - We look at what was done during the Sprint and if there is a better way to do it.

The responses to these questions were used to measure the acceptance of the approach using Technology Acceptance Model (TAM) as proposed by Fred Davis [1]:

- Perceived ease of use: measure the extent to which the developers and the non-informatic experts involved in the development think that using the combination of agile and MDA to develop their system is easy.
- Perceived usefulness: measure the extent to which the user (the developer and the non informatic-user) thinks that using the methodology enhances his job performance. This is measured using the following indicators: programmers productivity and satisfaction, development cost-effectiveness, timelines, customer satisfaction, absence of problems.

Other sources of data during the Intervention Phase were:

- Email: this was the online channel of information exchange, decision discussion and sometimes, decision making;
- Meetings: notes were taken during meetings.

At the end of each step, a report was prepared and shared to all the team by emails and during meetings.

Post-Intervention. The post-intervention phase consists of many activities such as the deployment of the solution and its vulgarization. In the case of the EPICAM project, after the intervention phase, a meeting was organized by the NTCP. During this meeting, we have presented the approach in order to have the opinion of the users. After the presentation, the questions, remarks about the approach were noted.

4.5 Model construction and transformation

In this research, the PIM is constructed using UML class diagram. To construct UML class diagram, many tools are proposed. We choose to use Dia Diagram Editor[9] because we master its use to build UML models. On the other hand, Dia is used to build large diagrams spanning multiple pages that can be saved in XML format and be printed.

The assessment of tools used for complete code generation in epidemiological surveillance (presented in Section 3.1) reveals that Imogene were more close to the features expected by the NTCP. Our proximity with the MEDES, the company which developed Imogene motivated us to adopt it.

Imogene is a platform to model and generate integrated data collection information systems. Thus, the class diagram (PIM) designed using Dia is transformed into the imogene model (Imogene). Thereafter, the model is used to generate the application. Imogene is developed using the Eclipse framework, the Eclipse Modeling Framework including Ecore meta model (to describe model) and the plugins furnished by Eclipse for code generation from the model.

[9] www.dia-installer.de

Class diagram	Imogene model	Source code	Comments
Class	Card Entity	Form for data collection/-consultation and table in the database	The source code generated by the imogene generator involved the form containing the class attributes as field and a table containing class attributes as column names in the database.
Attributes	form fields	Attributes of Java classes	Each card entity contains many fields and will correspond to attributes of Java classes. These fields may have different types.
Class relations	Relation fields between two card entities	One-to-one, one-to-many, many-to-one and one-to-many relations between classes	The source code corresponds to the relation between classes as prescribed by the Object relational mapping[10]

Table 2: Some mapping elements between PIM, PSM and source code

As the Eclipse EMF, the Imogene platform provides graphical tools to model business entities and generation tools to generate a set of applications based on the model. The modeling phase consists of defining the forms (with their fields) which will be used to collect/manage data. These forms contain the fields, their types and specific constraints; and the relations with other forms.

Additional information such as the application description, the description of each form (card entity) and each field (field entity) may be added using the imogene editor.

To manage access to the data, Imogene editor provides a particular card entity named actor. It's used to generate the application for user management. In our case, the user rights to the data is defined in the class diagram and converted in the imogene model using actor card entity.

Once the PSM is constructed, the corresponding applications are generated. In our research, given that there was not a tool for direct transformation of PIM to PSM, this task was done manually. The table 2 presents an overview of the transformation from the PIM (class diagram) into the PSM (Imogene model) and from the PSM into the source code.

V RESULTS

Globally, this research allowed us to demonstrate that the combination of Scrum and MDA allows effective development of epidemiological surveillance software when the MDA tool is mature enough to generate the entire source code. In effect, Scrum practices allow users to participate in the entire process of modeling and generation, which make the solution acceptable by end users. The details of the research results are given in the following subsections: Section 5.1 presents the results of the survey beside the health workers, Section 5.2 presents the results obtained during the implementation of the approach, and Section 5.3 presents the results of the evaluation of the approach by people who participated to this research. The results obtained, the finding and the lessons learned allowed us to provide an answer to the research question in Section 5.4.

5.1 Users survey

The survey of health workers involved 8 (53%) health professionals working in four hospitals where TB patients are treated, 1 (40%) health regional delegate, 6 (4%) managers at the National TB Control Program. All the managers have worked in hospitals before being managers. During the survey, a number of data collection tools (such as registers, statistics templates, data collection sheets - these elements are added as additional materials to this paper) were collected at the hospital, regional and central levels. Given the small sample size, we did not produced the statistics based on the questionnaire, but have summarized the findings per questions asked:

- On the question: "How do you manage your data?" We noted that at the hospital level that only manual management was used. In effect, at the central level, the data collection tools used are constructed using Microsoft Word, printed and distributed to health personnel. At the regional and the central level, manual management was combined with software management. Excel was generally used to put all the data collected in the hospital together and produce statistics. For instance, at the regional level, all the data from the hospitals of the region are compiled in an Excel file to obtain the overall situation of the region. At the NTCP level, all the data at the national level are compiled in an Excel file in order to have the overall situation in the country. Paper data is generally transmitted using the land transport from the hospitals to the Region. From the Region to the NTCP, they used USB keys and email.
- On the question: "What are the main problems you generally have?" Many problems were reported, the main are: (1) The problem of manual collection and management of data. In effect, everything at the hospital level is manual, which causes many difficulties to the health workers: difficulties to verify and correct the errors in the data collection tools, difficulties to guide the patients to the health center near to his home, difficulties to recall patients who have not come to their appointments, difficulties to update data collection (which is modify each years), difficulties to integrate data (for instance, children data) with the entire data. In addition, the use of many data collection tools (patient card, treatment register and laboratory register) to collect the same data (information on patient) is a tedious task; archiving paper data takes too many places and makes it difficult to find information. (2) The problem of manual analysis of data. This problem causes many other problems such as stock-outs of medicine, completeness, promptitude and erroneous statistics. (3) The problem of transmission of data using land transport. The main consequences of this problem noted at the Regional and the NTCP were the late arrival of data collection tools, data collected and epidemiological reports at their destination. At the Regional level, the promptness was estimated less than 70%. Generally, it is advised to transmit $\frac{2}{3}$ reports and to send the rest soon it is available. The rapid exchange of data between doctors is done using phone calls (for instance, during a transfer/referral). The awareness of patients and the general population is done using SMS. (4) Another problem not directly linked with the system such as power and Internet outage were reported.
  Only one health personnel reports that he did not have any problem with data management. We noted however that he treated less TB patients per week (generally, less than thirteen), compared to the other hospitals.
- On the question: "What is your knowledge on the use of IT tools?" Six health personnel affirm that they have basic knowledge on IT tools like Word and Excel. One of the six affirms that she already used an electronic health record. All the people from the Regional and the NTCP affirm that they have good knowledge on the use of Word and Excel, but they prefer to hire IT experts to build the Excel files they need for their statistics.
- On the question "What do you expect from epidemiological surveillance software?" Only one health worker estimated that he did not see what would be the benefit of the electronic system to his work. From the others, if we want to improve their work with an electronic tool, this tool must: (1) Solve the previous problems and be close as possible to the current tools (data collection sheets and reports) to facilitate the easy use; (2) Data management such as patient registration and following, sharing patient information to entitled users, automatic generation of patient cards, easy identification of lost sigh patients, easy identification of all health centers where TB patients are treated; (3) Allow the production and distribution of basic statistics automatically. They report to be aware that the

power and Internet outage will be the weakness of this system compared to the existing system and that the training of health personnel on the use of the electronic tool is the key to its success.

The table 3 presents the summary of the results.

Question	Participants	Code	Comments
How do you manage your data?	Health center	Manually	Data are collected using data collection forms and analyzed by counting by hand and filling the reporting form.
	Regions	Manually and with software	The paper data arrive from the health center. These data are aggregated in an Excel form. Some data is collected at our level manually.
	Central	Manually and using software	The data comes from the region in Excel form or paper form. We aggregate all these data in an Excel sheet. Some statistics are used in the data form.
What are the main problems you generally have?	Health center	Errors in data and tedious tasks	The fact that the data is in the paper form make it easy to make errors. In addition, collecting and analyzing data using paper is a tedious task.
	Regional	Errors and tedious task	Some data comes from the health centers (paper forms) with errors and we have to contact the health personnel to correct these errors
	Central	Data integration	When the data comes from the Region in an Excel sheet, we should aggregate these data in order to have a global and local view of the burden of the disease.
What are your knowledge on the use of IT tools?	Health center	Basic knowledge	Almost all the health workers in the health centers have basic skills (use computers only for typing or searching information on Internet) in the use of IT tools.
	Regional	Basic and Advanced	Some health personnel from the region have basic knowledge on the use of IT tools. However, others master the use of spreadsheet for form design and data analysis
	Central	Basic and Advanced	At the central level, we found some health personnel having intermediate and advanced level in the use of IT tools such as Word, Excel and the navigation on Internet
What do you expect from the use of a software for epidemiological surveillance?	Health center	Friendly UI	They proposed to develop the UI so that it's as close as possible to the tools they currently use for data management
	Regional	Automatic production of reports	The main concern at this level was the automatic production of reports
	Central	Automatic production of reports	As at the Regional level, the main concern was the automatic production of reports

Table 3: Summary of the survey results

At the end of this survey, the users confirmed that the electronic system will enhance their work as long as they are trained on the use of computers and the problem of power/Internet outage is considered.

5.2 Implementation of the approach

In this section, we present the outputs of the interventions implementation. This constitutes the results of the action research applied in this study. In effect, the combination of Scrum and MDA is deemed successful when its effectiveness can be demonstrated. The successful implementation of interventions is a testimony to the success of the approach. In the next sections, we present these results organized into Pre-Intervention, Intervention and Post-Intervention.

5.2.1 Pre-Intervention

The aim of the Pre-Intervention was to produce a preliminary list of interventions for the NTCP to assess, enhance and approve. Then, the following activities were planned: Establishment of a working group, identification of the problem faced by the current system, stakeholders survey to gain more insights in the problem, reflection on the solution, software specification and analysis. Specification, analysis, product backlog definition and validation were implemented by following the Scrum iteration practice. Then, three iterations of one week each were conducted. The

Interventions	People involved	Scrum practice	Status	Comments
Establishment of the working group	Researchers and product owner	Face-to-face	Complete	The researchers and the product owner select health workers to form the NTCP staff of the project.
Problem identification	Researchers and NTCP staff	Face-to-face and Scrum meetings	Complete	The NTCP staff meet regularly to discuss the problems NTCP have during the epidemiological surveillance of TB.
Users survey	Researchers and Epidemiologist	Scrum meeting	Complete	The survey were conducted by the researchers and the epidemiologist (product owner). At its end, a Scrum meeting was organized to share the results at NTCP.
Specifications and Analysis and Product Backlog	Researchers and NTCP staff	Product Backlog, Sprint, Scrum meetings, Scrum iteration	Complete	After the definition of the product backlog, the definition of the specifications, analysis and product backlog goes in Scrum iteration. During this process, Scrum meetings allowed the NTCP to make remarks and improve the product.

Table 4: Summary of the list of interventions planned and implemented during the Pre-Intervention

work was divided into tasks (for instance, writing the specifications of data collection, identification of actors) and the Scrum meetings make it easy to define common goals and commit to achieving the goal with the product owner. The face-to-face communication during these meetings allowed us to discuss, define and validate intervention outputs. At the end of each week, the Scrum Weekly Meeting was held at the NTCP to evaluate the comprehension of the needs by the researchers. The table 4 presents a summary of the Pre-Intervention.

Establishment of the working group. The aim of this intervention was to establish the group of people involved in the development of the system. This was successfully implemented and a team composed of two persons acting as researchers and developers, a product owner, a staff at NTCP, an epidemiologist from Centre Pasteur du Cameroun and end users were put in place. A mailing list group was put in place as a means to facilitate the collaboration and a channel to disseminate the decisions.

Problem identification. We observed that epidemiological surveillance at NTCP is managed manually. According to the staff at NTCP, this manual management caused many problems such as low rates of promptitude and completeness, poor statistics, etc. Because of these problems, information may arrive to decision makers late.

Users survey. The aim of this intervention was to have the opinion of the end users at all the levels (health centers, district, region central levels). This was successfully implemented and the results are presented in section 5.1.

Specification. The aim of this intervention was to gather and document the most requirements of the epidemiological surveillance system of TB. A study of existing software used for epidemiological surveillance and the survey of the end users made previously allowed us to identify the main functionalities of epidemiological surveillance systems. These are: data collection, data analysis, and information dissemination to stakeholders.

This intervention was particularly difficult because in addition to the data that the NTCP demands to collect, many health centers collect additional data. Then, these features have to be integrated in the specifications. On the other hand, the data to be collected may vary every year. The presence of the product owner impacted positively on the validation of the specifications. Then, the first version of the specification went through an approval process with the working group team and was validated.

Scrum iterations were very useful during the specifications and analysis. In effect, in order to understand very well what was really needed, interviews were made, transcribed by the re-

searchers and validated during the Scrum meetings with the working group. This was a cycle composed by the software specification (or update), analysis (or update), validation of specification and analysis (or update) and product backlog definition (or update). The positive effect that we observed was the openness that the Scrum meetings give to us in order to communicate with the working group.

The involvement of the NTCP staff was very useful to avoid poorly written requirements specifications. This allowed us to avoid severe problems presented by [65] which are: starting over with the product specification and restructuring it, difficulties to understand the requirements, and rewriting requirements.

The presence of the product owner, who was one of the users of the system at the NTCP level, has had positive effects on the Scrum practice. In effect, customer participation in the definition of product backlog prioritization of the backlog items, planning meetings and sprint review meetings make it easy to define common goals and commit to achieve this goal.

Analysis. This intervention consisted of determining the actors, use cases and to describe most important use cases. It was implemented successfully. The main actors identified were at the central level, all the staff of the NTCP; at the regional level, the regional delegate and the personnel in charge of TB surveillance; at the district level, the head district and the personnel in charge of TB surveillance; at the health center level, the nurses, the doctors, the laboratory technicians and the pharmacists in charge of TB management. Each user of the system has a particular role and collects data corresponding to their activities. After the identification of the actors, we grouped use cases by actors and at the end, we described the most important use cases.

Product backlog definition. Once the needs of the NTCP were well understood, the developer team transformed requirements into a product backlog. This product backlog was prioritized with the help of the product owner. Globally, the product backlog comprises the following backlog items:

- Platform design: this consists to design the software architecture, determine the MDA tool to use and design the user interface;
- Development of user management module: this module is used to grant data access to users according to their rights;
- Development of data management module (patient and medical data): data management module involves the forms used for data registration, data searching;
- Development of the reporting module: the reporting module allows users to have a synthesize version of their data (in form of statistics);
- Development of the mapping module: the mapping module allows health workers to have a map containing the different hospitals involved to the system;
- Development of the sensitization module: this module allows the NTCP to sensitize the population by SMS.

The Sprint Review Meetings permitted us to adopt the software specification, analysis and design with the whole staff at the NTCP. We designated the user management module and the data management module as our initial list of interventions which we presented to the NTCP for discussion and deliberation.

Lesson learned. The first lesson we learned was that epidemiological surveillance systems requirements cannot be fixed and change generally. Then, we should be careful of the approach and tools to use for the development.

This lesson led to a deep reflection by the software team on the choice of the development approach to use. Given the limits of the code-centric approach presented in the introduction, insight was then to model and generate the software instead of coding it. Then, the combination of Scrum that were already chosen with MDA were adopted. Thus, software development was divided into Sprints and with the first Sprint composed of the following interventions: software design, development of user management module and development of data management module.

5.2.2 Intervention

Intervention phase were organized in six Sprints. During the development, Scrum boards were used to display the status and progress of each sprint. Ad hoc meetings were also used when some issues were identified. During the Scrum meetings, the developers found it easy to define a common goal and commit to achieving the goal with the product owner on the one hand, and the NTCP staff on the other hand. The following paragraphs give the details of the execution of each Sprint.

Sprint 1: development of a first prototype. The goal of the first Sprint was to design the platform and to develop the first prototype. This Sprint took 4 iterations of one week each. All the interventions involved in this Sprint were not completely implemented successfully. In effect, the user management feature generated by the MDA tool was not corresponding to how the NTCP managed their users. The table 5 presents a summary of the interventions implemented in this Sprint. The details about these interventions are given below.

Intervention	People involved	Scrum practices	MDA practice	Status	Comments
System design	Researchers and NTCP staff	Sprint, Iterations, Scrum meetings	Modeling PIM	Completed	During this intervention, beside the other tasks, the product researchers and the two health workers (including the product owner) were involved in the modeling of the PIM with Dia. Their role was to validate the different elements and relations between them.
Prototype implementation	Researchers and NTCP staff	Sprint, Iterations, Scrum meetings	Transforming the PIM to the PSM and the PSM to the source code	Not complete	This intervention involved the researchers and two health workers. The health workers participated in the transformation of the PIM to the PSM using Imogene. The production of this version was presented to the NTCP staff during Scrum meetings. However, the user management feature proposed by Imogene did not correspond to the features expected by the users. We observed that the two health workers involved have had great difficulties in modeling with UML.
Product backlog update	Researchers and Product owner	Scrum meetings	None	Complete	Given to the remarks on the users' management features, the product backlog is updated with new features to develop and integrate in the Imogene generator.

Table 5: List of interventions implemented during the first Sprint

System design. Given the system specification and analysis, the architecture, the user interfaces and the Platform-Independent model were made.

1. **Architecture design.** The architecture design intervention consisted of designing an architecture (given in appendix A) which suits an environment like Cameroon where access to the Internet is not always available.
2. **Choice of MDA tool:** to choose the MDA tool that is more suitable to our problem, an overview of existing tools used in epidemiological surveillance were made. Either these tools were develop using code-centric approach (DHIS[11], OpenRMS[12]), either they were specialized in the generation of mobile-based data collection tools (ODK[13], Magpi[14]). Only Imogene[15] were found to cover most of our requirements. Then, a closed collaboration between the University of Yaounde I and MEDES allowed us to build good skills in the modeling and the generation of data collection applications with Imogene. A deep analysis of Imogene by the researchers shows that it permits to model and generate only two modules of the TB surveillance software: the data management module and the user management module.
3. **User Interfaces (UI) Design:** during this intervention, a deep discussion with the working group and the constraints of the Imogene platform on the design of UI allowed us to decide how the elements on the UI will be organized to resemble the manual management tools.
4. **Platform Independent Model design:** this intervention consisted of determining the entities, their properties, their relationships. Once these elements were determined, we set-up a UML class diagram using Dia. The working group were involved in this intervention. In fact, given that our main goal was to allow them to be in charge of the update of the system, they were invited to install Dia and use it. However, only 2/11 (18%) installed the tool. 3/11 (27%) estimated that the tools were out of their domain of expertise and 6/11 (55%) were not available. The final version of the class diagram validated by the staff is presented in appendix B.

Prototype implementation. Imogene was used to build the Platform Specific Model in only one week. This intervention involved the researchers and two NTCP personnel. The whole activities were executed by the researchers and followed by the health workers. In fact, the passage from the PIM to the PSM was manual following the rules given in the table 2 of the Section 4.5. Additional information such as field description, field name in French and English were added to the PSM using the Imogene editor. The screenshot of the epicam model obtained during this intervention is given in appendix C. This model was used to generate the first prototype of the application. This first prototype consisted of:

- **An administration application:** Used to manage health workers information, their roles, and their access rights on data;
- **A Web application:** Containing the main user interfaces needed. It is used for the registration of patients and their follow-up;
- **Offline application:** The offline application has the same functionalities as the Web application, but is used in an asynchronous mode. When Internet connection is not available, the system uses the local database as storage system and the local database is synchronized with the server when the Internet is available;
- **Synchronization application:** Used to synchronize the client with the server (updating local or remote database).

[11] https://dhis2.org
[12] https://openmrs.org
[13] https://opendatakit.org
[14] https://www.magpi.com
[15] https://github.com/medes-imps/imogene

Testing and updating the Product Backlog. During this Sprint, six review meetings took place with the working group. The goal of these meetings was to evaluate the users and data management modules. During these meetings, the missing fields and some field names errors were noted. The missing fields information were used to update the imogene model in order to make the source code correspond to the requirements. Error fields were corrected. It was also noted that some users from the central level can access personal patient information. The NTCP staff warned us that only the health workers in charge of the patient must have access to personal information of patients. Users' remarks permitted us to integrate the specification, analysis and design of user rights to the data in the Product Backlog.

Next iterations. After the development of the data collection module, five iterations were conducted. These iterations permitted us to integrate all the information that was not present in the PIM and correct some typology errors. This involved the design, the development, and the testing. The design consisted of the updating of the PSM, the development consisted of the generation of a new version of the platform.

Lessons learned many lessons were learned during this Sprint:

- We should be careful when using a tool for the complete generation of the software: we remarked that it is difficult to find a tool that fits the requirements of a specific context.
- Domain experts are not generally available: NTCP staff were involved in the modeling and the generation of the software. However, we noted during the Scrum meetings that they were not generally attending. It should be noted that most health workers in the NTCP are nurses, physicians, lab technicians, epidemiologists. They have full time employment and their commitment to the project was during their spare time. This constrains the implementation of the interventions.
- Modeling can be a difficult task for domain experts: among the domain experts that were involved in the modeling, only 2/11 (18%) installed the modeling environment and 3/11 (27%) estimated that installing and using the modeling environment is out of their competencies.

After a deep reflection during ad hoc meetings between the researchers, the MEDES engineer, the NTCP staff and CPC expert it was decided to develop and update the Imogene platform with the user management module that fit to the NTCP needs. This new intervention was put as a main priority. Given that medical data is very sensitive, it was decided to prioritize the development of this module. Then, the second Sprint was dedicated to the development of the generator of user management modules that fulfill epidemiological user management.

Sprint 2: implementing user management module generator. The interventions involved in this Sprint comprised on the one hand the specification, analysis, design, development and integration of the user management module in the Imogene platform and on the other hand, the modeling and generation of the first prototype. These interventions were done by the researchers supervised by the MEDES engineers. It lasted 4 months with 8 Sprints. The Product Backlog contains the following interventions: training on the development technologies, specifications, analysis, design, development, validation. The iterations processes involved the specification, analysis, development, validation. The table 6 presents the summary of the list of interventions implemented during this Sprint.

Only the developers (supervised by the MEDES engineers) and the product owner were involved in the development of this module. The main role of the product owner was to verify that the user management module allowed us to generate the user management system of TB surveillance at the NTCP.

Intervention	People involved	Scrum practice	MDA practice	Status	Comments
Training	Developers and MEDES engineers	Scrum meetings	None	Completed	The developers were trained to the software to be used to develop new modules.
Specifications	Researcher and NTCP staff	Scrum meetings	None	Completed	The NTCP staff explained to the developers how users should be managed in their system.
Development	Developers and MEDES engineers and Product owner	Scrum meetings, Scrum iteration, Sprint	Constructing the User Model	Completed	This intervention consisted to constructed the user model, use it to complete the imogene model so that the user management feature generated will fit to the NTCP needs. During this intervention, the product owner verified during Scrum meetings that we are developing what they expected.
Validation	Researchers and NTCP staff	Scrum meetings	None	Completed	After the development, we present the interface that is generated for user management to the NTCP staff.
Prototyping	Researchers and NTCP staff	Scrum meetings	Modifying the PIM and the PSM	Completed	The PIM and the PSM is modify in collaboration with the two health workers involved in the development
Prototype validation	Researchers and NTCP staff and end users	Scrum meetings	None	Completed	Once validated by the NTCP staff, the product is presented to end users in the field for validation.

Table 6: List of interventions implemented during the second Sprint

It was noted during the specifications and analysis that users have different roles. In some hospitals where there is not enough staff, a user may be the person who consults the patient and gives its medication. With the user management module, when the administrator creates a user, she/he allows her/him to read/modify the data. When a user login, she/he can access only the data she/he is entitled to. The appendix E presents a screenshot of the UML model and the user interface allowing the creation of a role and the attribution of rights to users.

We used the same tools used to develop Imogene to develop this module: the programming language was Java, the Integrated Development Environment was Eclipse Kepler for Java EE, the Google Web Toolkit to develop User Interface, the Spring 3 framework, the Spring security framework, code management with Maven and git, the modeling environment was Eclipse Modeling Framework and the code generation environment was composed of oAW (involved Xtext, Xpand, MWE components). Before the development of this module, the developers were trained on these technologies. After the development, this module was integrated in the source code generator of Imogene.

Once the generator of the user management module was developed and integrated into Imogene, the epicam model (appendix C) was updated manually and used to generate a new version of the prototype. It should be noted that this version of the prototype allowed us to manage users and TB data.

Tests and experimentation. The second Sprint permitted us to obtain a first working version of the system that the end users in the field can use in a real environment. After the tests and the Scrum review meeting with the working group, we decided to validate the developed platform in the field. Then, we deployed it in six hospitals in Douala and Yaounde (which are the largest cities in Cameroon). Eight health personnel were involved in this validation which lasted one month. The goal were to evaluate the perceived usefulness and the perceived ease of use of the end users:

- **Perceived usefulness:** All the users found the tool useful for data management. They estimated that this tool will allow them to obtain data of good quality and improve their work. They also noted that the tool may not be suitable in certain hospitals. In effect, the

scenario presented by the prototype considers only the registration of patients when they are tested positive to TB. However, in certain scenarios, the patient is registered in the system when they are suspected to have TB. We noted also that amongst the TB exams that the NTCP generally uses, some hospitals also use radiology exams. The system was not considering the registration of this exam. New requirements were noted as a new intervention and the product backlog was updated.

- **Perceived ease of use:** the users estimated that the use of the system was intuitive. All the eight personnel involved found that the forms resemble the paper forms they are using in the hospital.

Lesson learned globally, two lessons were learned at the end of this Sprint:

- We learned during this Sprint that beyond the Scrum team, the end users must be integrated in the validation of the software. After a deep reflection with the working group, it was decided to invite at least 50 users around the country for the validation before the deployment;
- We learned that the maturity of the MDA tool can slow down the development of the software system. Then, it can be a serious barrier to adoption. In effect, the development of the user management module has decelerated the execution of interventions and have had a negative impact on the timeline of the development.

Initially, 11 health personnel were involved in the project. However, after the first and the second Sprint, this number dropped to 5 (0.45%) with 3 (27%) completely engaged. Most participants justified their absence by their job. In addition, the product owner got another job and left the project. Unfortunately, he was among the personnel accustomed with the modeling. He was replaced by one person among the NTCP staff.

Sprint 3: Updating the application. The third Sprint consisted of the updating of the application by the integration of the end users remarks and the generation of another prototype. This was done in only one week. Then, a Scrum planning meeting was done, followed by the updating of the model and the generation of the new version of the system. The Scrum weekly meeting allowed us to validate the new version.

It should be noted that involving the NTCP staff in the development allowed 3 (27%) to grow more accustomed to the Scrum practices such as Product Backlog definition, organization of Scrum Sprint, and Scrum meetings. They found Scrum advantageous in the development of their system.

Sprint 4: Developing the reporting and mapping modules. During the fourth Sprint, the reporting and the mapping modules were developed (by the developers supervised by MEDES engineers) in two months, organized in 5 Sprints. The reporting module was developed using Business Intelligence and Reporting Tool (BIRT[16]). The mapping module was developed using Open Street Map (OSM[17]). During the Scrum meetings with the working group, the users validated these modules. Then, they were tested, validated at the NTCP and integrated in the code generator. A new version of the platform was validated by the users in the field.

[16] https://www.eclipse.org/birt/
[17] https://www.openstreetmap.org/

Intervention	Status	Comments
Product Backlog definition	Complete	The health worker define the list of tasks to execute
Sprint organization	Complete	The list of tasks were organized into 2 Sprints of one week each
Organizing Scrum meetings	Complete	The simulation of the organization of Scrum meetings were done by the health personnel
Modeling the PIM	Complete	The health worker used Dia to model a small application
Installing MDA tool	Complete	This intervention was the most arduous. The health worker found the installation of Eclipse and the different plugins out of his competence. However, with the help of the developers, he finally installed it
Construction of the PSM from the PIM	Not complete	The installation difficulties discourage the health worker to continue the experience.

Table 7: List of interventions implemented during the sixth Sprint

Sprint 5: Developing the SMS module. The fifth Sprint consisted of the development of the SMS module in one month with 3 Sprints. To this end, we used Spring Rest API. Actually, to send SMS using the platform, we decided to use a service provided by an SMS provider. The SMS module was tested, validated at the NTCP and integrated in the code generator. The new version of the epidemiological surveillance software generator containing a new version of user management module, reporting module, mapping module and SMS module were named EPICAM[18] and hosted on github.

We found the development of additional modules arduous than the development of code-centric application. In effect, it was questioned to reverse engineer Imogene in order to add new elements to generate in the code generator and to link these elements to the graphical interface for the modeling. During this development, we observed that Scrum practices, particularly face-to-face, Product Backlog, Scrum iterations and Scrum meetings were very helpful to organize the work and communicate among the developers, with the developers and the MEDES engineers and the developers and the stakeholders. It should be noted that all the fields used to automatically generate the mapping module were already presented in the PIM, which was transformed into the PSM.

Sprint 6: system update by non-informatic experts. In this research, we wanted a method that allows non-informatic experts to be engaged in the development/updating of their system. Thus, after the development of the platform, the goal of the current Sprint was to prepare the health workers to take charge of the updating. As presented before, only three NTCP staff followed the development to the end, only two were involved in the modeling task and one of them moved to another job. The main goal of this Sprint was to allow the health personnel to update the system using Scrum and MDA. Then it involved the simulation of the following interventions: building the product backlog and organization into Sprints, validation of the product backlog by the working group, organizing and conducting Scrum meetings (daily and review), modeling the PIM with Dia, installing the MDA tool, manual transformation of the PIM into PSM, generation of the new version of the software, deployment on a local machine and validation by the NTCP. These interventions were planned to be executed by the health personnel, supervised by the researchers. The summary of their implementations is presented by the table 7.

The team conducted two-weeks of iterations, trial, errors and learning. During the first week, the modeling environment was installed and the product backlog was defined. This product backlog contained the following interventions: adding new forms with new fields for data collection, constructing the PIM, transforming the PIM into PSM. Since the EPICAM platform (used for the modeling and the generation) is based on the Eclipse platform, we observed that installing

[18] https://github.com/UMMISCO/EPICAM

these tools by non-informatics experts is not an easy task. This did non encourage the latter to continue with the experience and he stopped after the second week.

The interview of this participant gave more insights on his motivation to participate in the development and update of the surveillance platform. Actually, he characterized himself as an advanced computer user that works with Microsoft Excel, and Epi Info [19]. He used Epi Info for modeling (with drag and drop) and generating data collection forms. He also had experience with the use of Microsoft Excel to create forms. We learned from him that the only barrier for the adoption of the approach by non-informatic experts was the MDA tool because Scrum was very intuitive. He stated that if the UI of the modeling software reassembles to the drag and drop UI of Epi Info, the modeling tool and the approach can be more accepted by the health personnel. This view was strongly supported by two other participants. However, developing this new feature was beyond the mandate of the developers. The problem was the time that this drag and drop UI can take to develop and the technologies that the developers must master in order to develop it.

We approached the two other staff members at NTCP to have their opinion about the approach. They characterized themselves as basic users of computers. They used computers mainly to type text and to navigate on the Internet. They had no experience with modeling and Scrum before joining the project. They found that Scrum is a good tool for work organization and for communication with developers. Some of them affirmed that they have developed good skills in the Scrum process and are using it in their own work. We asked the question "Why are they reluctant to participate in the experience". The most common reasons were the lack of expertise. They stated that it took a long time to feel like they understood the graphical models we were presenting during the Scrum meetings. They found the tools too much oriented towards the engineer. They stated that the complexity is high, because of the installation, the interface of modeling and also, they found UML very complex.

Lessons learned: we learned that counting on health workers to update the system is not advisable. In effect, generally, they have a full time employment and their commitment to contribute to implement interventions is during their sparse time. In addition, the update by non-informatic experts declined due to the lack of knowledge on modeling and the difficulty to install the MDA tool.

This lesson has had some implications at the NTCP level. They decided to hire a person with strong skills in Scrum and MDA to accompany them in the update of their system.

Although these can be classified as failed interventions, the experience has some merits. It gave insights on what should be done to make non-informatic experts update or even develop their system: develop a drag and drop modeling tool.

5.2.3 Post-intervention

The Post-Intervention consisted of training end users in the use of the software developed for epidemiological surveillance of TB and to disseminate the outcomes of the project.

To disseminate the outcome of the project, a demo session of the approach was presented to the stakeholders coming from the whole country (around 60) during a meeting organized by the NTCP. During this meeting, some participants said that the NTCP played a passive role in the

[19] www.cdc.gov/epiinfo

software development before, this is a new experience for them and they have to rethink their participation model to software development.

Fifty health personnel involved in the epidemiological surveillance of TB at the hospital, regional and NTCP levels were invited for a two days experience at the university of Yaounde I. Our main goal was to assess the tool before the deployment in the whole country. Conscious of the level of the entire health workers on the use of IT tools (found during the deployment of the prototype), we decided to train them before on the use of IT tools. Then, the first day consisted of getting all the health workers familiar with the use of the computer. For the perceived usefulness, the users found the tool useful to improve their job. However, they noted some typology errors at the User Interface level - these errors were corrected and a new platform generated. The users also reported that the system was easy to use. They particularly appreciate that all the data on the patient treatment is in one form (the patient form is given in appendix D).

The final version[20] were experimented in 25 pilot sites across the country. Around 3900 patients which represents 15.6% of the annual number of TB cases in Cameroon were registered and followed. Given the success of this pilot phase, the NTCP has adopted the software as its electronic epidemiological surveillance software and extended it in twenty new health centers during the years 2016 and 2017.

5.3 Perceived ease of use and perceived usefulness

The developers and the NTCP staff involved in the development were invited to give their opinion on the approach. The analysis of the case data using the Technology Acceptance Model presented in section 4.4 showed that the developers have found it easy and useful to combine agile and MDA for the development of epidemiological surveillance software. However, even if non-informatic experts easily handled and adopted Scrum practices, they found the approach difficult to use because of the lack of skills in modeling and the use of the modeling tool.

5.3.1 Perceived ease of Use

To measure perceived ease of use, we considered the developers evaluation and the stakeholders evaluation. We measure the degree to which these participants believe that using the combination of Scrum and MDA is free of effort.

As developers, globally, we can say that the application of Scrum and MDA for the development of epidemiological surveillance software is easy to use. In effect, once the features were validated and integrated in a Sprint Backlog, it generally took us a maximum of one week to update the model and generate a new version of the software.

The integration of Scrum in the development was very easy. In effect, once the different roles were introduced, the NTCP designed the product owner. The researchers explained his tasks and he conducted them very well without any difficulties.

After the production of the final version, we involved two health workers in the simulation of the update of the system using the combination of Scrum and MDA. Non-informatic experts constituted by the NTCP staff found Scrum simple, easy to learn, easy to handle and use. For them, it does not require a lot of mental effort. Thus, it was rapidly adopted. They found it easy to learn through experiences, the self organization when some members leave the team were very appreciated. Thus, during the project, they grew more accustomed to the Scrum practice.

[20]Source code generated: https://github.com/ummiscolirima/epicam_tb

However, they did not get accustomed to the MDA. For them, the main reason was the lack of knowledge in modeling and the use of the MDA tool.

5.3.2 Perceived Usefulness

As we have done above, in the following paragraphs, we present the developers and non-informatic experts views of perceived usefulness.

Developers found the combination of Scrum and MDA very useful in our work of software development. In effect, combined with MDA, the Scrum Process permitted us to gradually build models and generate the executable source code which are updated with the gradual expansion of the epidemiological surveillance. This has a significant impact on qualities such as: programming productivity and satisfaction, development cost-effectiveness, timelines and customer satisfaction.

- **Programming productivity and satisfaction:** we found the approach useful because it allowed us to develop/update as fast as possible the software. The use of models allowed us to reduce complexity in the development and the Scrum practices allowed us to better organize the work between us and the communication with the customers.
  In effect, to develop the system, all the tasks to perform are prioritized and consigned in the product backlog. Depending on the size of the product backlog, one or many Sprints can be put in place. The Scrum meetings and the face-to-face communications were very useful to clear up any misunderstanding and facilitate the discussion and the decisions during the development. During the Sprints, the developers tasks involved the modeling and generation of the software. The use of Scrum iterations allowed us to repeat during any Sprint until all the tasks planned are completed. For instance, once the first model was built and the software generated, during the users evaluation, modifications to apply were identified and the cycle resumed. The repetition of iterations helped us to learn from the errors and reinforce what went well, achieving continuous improvement. The use of Scrum increment and scrum meetings with the NTCP in the requirements specifications and analysis have had a great positive effect on the understanding of the system by the developers.
  To update the system: Once the product backlog is defined, the system is updated in one Sprint of maximum one week. We found the one week iteration very useful because it helps us to review the modification of the model and generate the new working version as fast as possible.
  In conclusion, the combination of Scrum and MDA have had a positive impact on programming productivity and satisfaction. We found a positive effect on:
  - long-term productivity improvement because of automating repetitive tasks especially adding/modifying new forms, report, fields, etc.;
  - satisfaction because the organization of the tasks and communication using Scrum practices allowed us to agree with the NTCP with the continuous improvement of the solution. Whatever the new requirements were added, the Scrum team refined the product backlog and the new software was generated. Consequently, all the feature sets were completed during the project.

  We found that to have a great positive effect on programmer productivity and satisfaction, the MDA tool must be mature enough to generate the code covering all the features demanded. In our case, the immaturity of the MDA tool has also had a negative effect in the development because the development of additional modules was hard (we have to be trained in new technologies) and took time.

- **Development cost-effectiveness:** we also found that the approach has paid off by reducing the complexity of software development. In effect, the reusability of the generation tool and model allowed us to reduce the costs of the development and update. The use of Scrum practices such as Product Backlog and Scrum meetings allowed us to facilitate the development of the solution. The modification of the model and generation of the new software have made it more valuable. In addition, the approach reduced costs by saving the time that may be used to debug a system because of the problem of software regression.
  It should be noted that the EPICAM platform can be reused for the development of several epidemiological surveillance software.
  We concluded that the combination of Scrum and MDA has a positive effect on development cost effectiveness.
  On the other hand, the development of additional modules for the generation tool was time consuming and has had a negative effect on the cost-effectiveness.
- **Timelines:** the Scrum planning meetings, Scrum daily meetings facilitated the organization and prioritization of tasks. The Product Backlog allowed us to define the different tasks and prioritize them so as to deliver the first working version as soon as possible. The one week duration of iterations were useful to ensure that the development is advancing in a good way. The development of this first working version was greatly facilitated by the use of models. In effect, once the requirement and the analysis were defined, the transformation into product backlog with prioritized and estimated tasks allowed us to have insights on the main features of the software and develop the first working version of the platform as soon as possible.
  In conclusion, the approach presented in this research has a positive impact in the timelines of the development.
- **Customer satisfaction:** Scrum practices and MDA approach have had positive effects in customer satisfaction. In effect, the involvement of the stakeholders in the development allowed the improvement of short time visibility of common goals. For instance, the scrum meetings and the use of models to present the evolution of the work during these meetings have brought the customers closer to the development and revealed core requirements. Exchanges between the developers and the NTCP staff have had a positive effect in the development. The opinions of the stakeholders revealed that this was a positive way to keep people aware of the project. The Sprint planning meetings were very useful for the Scrum team. It allows the product owner to explain to the team the most valued requirements and the prioritized backlog items to be delivered to the customer at the end of the Sprint. Through these meetings, any misunderstanding between the developers and the customers were solved. Different Scrum Meetings with the NTCP staff and the end users allowed us to build a product which was as close to the users as possible. At the end, we observed that the software was accepted by the end user. Some of them found the user interface close to the tools they were using in their work.
  For the development phase, during the Scrum meetings, when the customer or the end users proposed to add or modify something, these tasks were added in the product backlog and the modifications were planned. These modifications and the testing took one Sprint of less than one week because in most time, it was question to modify the models and regenerate the updated solution. The stakeholders were very satisfied because during the Scrum weekly meetings, they found their remarks integrated in the solution. Sprint reviews were very positive, because we could also get a view of what other types of requirements the stakeholders had.

As prescribed by Scrum, the goal of producing a working version of the system for iteration reviews was a good way of noticing and communicating integration problems that would otherwise be communicated only after the overall implementation. This practice has a positive effect because the end users and the customer no longer wait until the end of the project before they have the results.
The fact that the customer commenced to use the software was useful to the development. The sprint review meeting shows the customer the new features and changes implemented in the Sprint. The Scrum team inspected the work done by the end users. Then, we found the approach very useful to reduce time-to-market. The end users started using the tool and got accustomed to it as soon as possible. This helps to realize if changes are necessary. In fact, sooner the changes are detected, the less its impact and cost.
After the success of the deployment in 25 pilot health centers, the NTCP extended the deployment in twenty new health centers.
In conclusion, the use of Scrum and MDA allows us to deliver a regularly working version of the software. This provided positive impacts in customer satisfaction.

Even if the approach has been found useful, we also found that the immaturity of the MDA tool can have negative effects in the development. It can slow down the programmer productivity and satisfaction, augment the development cost because new generators must be developed, and be costly in time.

Non-informatic experts also found the combination of Scrum and MDA useful. They were impressed by the rapidity with which new versions of the software were proposed every time where the changes were necessary. However, they found that they don't have good knowledge in the use of models and modeling tools to use the approach to update the system. The interview of the user and a deep observation on the use gave more insights on the non adoption of the method. The main bottleneck of the adoption was the installation of the tool and the UI. Developing tools based on Eclipse and Eclipse plugins such as EMF, etc. have had a negative effect on the acceptance of the approach by the stakeholders.

5.4 Conclusion

The lessons learned in this research and the main finding allow us to provide a scientific answer to the research question, independently to TB surveillance. This is required because the pilot project presented in Section II aims to study the epidemiological surveillance of TB in order to propose solutions that can be generalized to any other cases. We found that to mesh Scrum and MDA for the development and maintenance of epidemiological surveillance software, the following points should be considered:

- MDA framework (such as EPICAM for epidemiological surveillance) should be mature enough to generate the entire source code. In effect, the development of additional modules was not an easy task on the one hand because of the reverse engineering of the Imogene platform and on the other hand, because it is not easy to build a tool that should generate source code for all the surveillance systems. The additional development has had a negative impact on programmer productivity.
- Scrum practices such as involvement of stakeholders in the project, Product Backlog definition, Sprint definition, Scrum iteration, Scrum meetings must be used to build models, make these models validated by the domain experts and make the application developed validated by end users.
- Use the Scrum process involving the pre-development, the development, and the post-development:

1. Use the pre-development step to put in place the Scrum team, well document the problems so as to lift any ambiguity and misunderstanding between the developers and the users. In addition, write the application specifications, analysis and produce the first version of the product backlog. During the pre-development step, the Scrum iteration practice can be used and different iterations will be used to lift any misunderstanding.
2. During the development step, organize the development into Sprints. The first sprint consists of the design of the system by firstly justifying the adoption of the MDA approach and secondly identifying the most relevant tool to be used. If there is no adapted tool, the team should evaluate the cost of developing a new tool or updating an existing tool compared to the cost of developing a new system using the code-centric approach. It is during this sprint that the MDA approach is adopted. If there is not an adaptable tool, thus, the team should consider to develop the new ones during the sprints 2 to sprint n. Once the tool is adopted, the team should develop a first version of the system during the sprint n+1 by building the PIM, transforming into PSM, generating and deploying the software. In our case, it was not too difficult to pass from the PIM to the PSM manually. But, the automatic transformation can be a better solution to augment the programmer productivity. The first version of the system should be validated by the product owner who is part of the team and deployed to some end users. From the sprints n to the sprint m, the models are updated using the users feedback and the most complete solution is developed.
3. The post-development consists of training and making the software available to all the intended users.

VI THREATS TO VALIDITY

We found in this research that if the MDA tool is mature enough, the combination of Agile and MDA have positive effects on programmer productivity and satisfaction, cost-effectiveness, timelines and customer satisfaction. In this section, we evaluate under which conditions our research is applicable and offers benefits and under which circumstances it might fail. Section 6.1 presents threats to internal validity and section 6.2 presents threats to external validity.

6.1 Threats to internal validity

Threats to internal validity refers to whether the study supports the finding [17]. A number of threats to internal validity may surface at the Pre-Intervention Phase and the Intervention Phase.

At the Pre-Intervention phase, the main threat came from the survey of health personnel during the Pre-Intervention Phase. In fact, we choose the qualitative technique (data collection with interview) in order to lead to richer information on the subject of the study. The small number of subjects in a study does not allow any claim that the finding is representative of the organization in general [17, 66]. In effect, when the number of subjects is small, this may hamper the identification of all the problems of the system and give rise to the nonacceptance of the solution. This was remarked during the validation with health personnel where other patient management scenarios and additional features were proposed to be integrated.

During the Intervention Phase, the data were collected using direct observation by watching how the combination of Agile and MDA is being used. Collecting data with observation may lack clarity [17, 66]. On the other hand, we were the developers and the researchers who

developed and evaluated during the Intervention phase. It would have been better to have an external researcher who observes and evaluates everything.

In spite of the threats presented in this section, we have reasons to assume that they had only limited impact on the research. Firstly, the threats brought by the survey were corrected during the evaluation of the first prototype by end users. Secondly, interviews were used to solve the problem of lack of clarity of the observation. For instance, the health personnel involved in the updating of the platform using our approach were interviewed to know why they were reluctant. This allows us to discover the lack of competencies in modeling and in the use of modeling tools.

6.2 Threats to external validity

Threats to external validity refers to whether the finding can be generalized [17]. In this research, the generalization can be seen at two levels: at the level of epidemiological surveillance systems and at the level of the industrial systems.

The first threat to external validity came from the case study research. In fact, the selection of a case must be done so that if the research works for this case, it is likely to work for many other cases [17, 66]. However, we just chose the epidemiological surveillance of TB because the NTCP were available to collaborate with us in order to find a solution to their problems. This may hamper the generalizability at the level of epidemiological surveillance systems and industrial systems. It should be noted that the head of the epidemiological service at Centre Pasteur du Cameroon were involved in the research. This service is specialize in the monitoring of several infectious diseases whose rabies, HIV, Influenza, meningitis, etc.

The second threat came from the case study and action research. In fact, the generalizability of results is limited to the domain of health where the population shares similar characteristics with the sample involved. In Empirical Research in Software Engineering, to generalize an approach, one should replicate it to many other settings [17, 66]. Then, to generalize the approach we propose in Section 5.4, to the other industrial cases, it needs to be replicated with different languages, tools, applications and application size.

The fourth threat came from the developers. Only two developers were involved in the study. This can hamper the generalizability of the study to other developers.

VII RELATED WORK

The problem of failed software does not only concern the domain of epidemiological surveillance. In fact, software developers of today face the problem of increasingly complex software as clients demand richer functionalities in shorter timescales. This has given rise to the software development question and dilemma: how to provide the customers with the software and software artifacts that fulfill their expectations in time and within the budget [41]? To respond to this question, related work proposes approaches and tools amongst which Model Driven Techniques and Agile Processes (refer to Section III. To write this related work section on the combination of Agile and MDE, we used Scopus, Semantic Scholar and Google Search engine to search for papers related to the development of: (1) any software, (2) health software, and (3) epidemiological systems. As demonstrated by Hessa and Kevin [52] in a review article published in 2022, a lot of work was proposed on the combination of Agile methodologies and MDA for the development of software systems in general. However, we found only two works [50, 57] proposing the combination of Agile and MDA for the development of health software

in general and zero work for the development of epidemiological surveillance systems. In the following paragraphs, we present the two works and a table comparing these works with our work.

Kannan et al. [50] proposed the combination of Agile with Model Driven Development for the development of electronic health records (EHR) in the United States. On the one hand, the authors highlighted the problems when developing EHR which are: misunderstanding requirements, frequent product change requests, product defects, the developments are costly and time consuming. On the other hand, the authors have the challenge to develop EHR tools across 40 speciality academic medical center practices over one year. Thus, they thought it necessary to use a rapid yet high quality development approach such as the combination of Agile and MDD. Agile artifacts that were used involve User Stories, User Interface Storyboards, and Use Case Text. Concerning the MDD tools, the authors used the UML language to build diagrams to facilitate the communication amongst different stakeholders involved in the project. They used Class Diagram, Object Diagram, Use Case Diagrams, Decision Threes, Swimlane Workflow diagrams, and State Diagrams. They did not adopt any tool that is used to model and automatically generate the software source code. As results during the year of 2015, 58 EHR were developed, 111 EHR-based data collection tools built and 16.000 patients on registries. The authors found that the combination of Agile and MDD allows rapidly gaining consensus, acceptance, clarification on requirements and removing ambiguity. They recognized useful design patterns for re-use on subsequent projects.

Blagoj et al. [57] presents the national e-health system implemented in Serbia as "MojDoktor" and in Macedonia as "MojTermin". These systems are based on the same integrated health information platform. For its fast development and adoption, its architecture was designed with Agile methodologies. However, this system suffers problems such as: its extensibility, soundness, interoperability and standardization. To solve these issues, the authors proposed a formalization of the system design and its implementation as a Model Driven Architecture. The iterative and gradual development agile practice was adopted for the development. The models proposed involve a set of formal models on several abstraction levels, a model for data transformation that provides interfaces for interoperability between systems. The system were modeled by adopting the CIM, PIM and PSM:

- Computational Independent Model: consists of the workflow diagrams used to describe the business logic, and the document description.
- Platform Independent Model: on the one hand, the authors used several tools to model the system: class diagrams, abstract state machine to capture behavior formalism. On the other hand, they proposed a framework for describing the software architecture that is based on creating models, meta-models and constraints between meta-models. The models layer contains models that describe specific elements of the system (e.g., services, entities, process). The meta-model layer defines generic models for specific elements (e.g., services, entities, process).
- Platform Specific Model: this layer involves the source code that implements the different elements modeled by the PIM. For each model of the PIM, a platform specific implementation exists. These implementations must conform to the models, and the models must conform to the meta-models.

Using this organization, the authors demonstrated how the model built can be used to model and generate RESTful APIs. The authors used Visual Paradigm for UML for code generation. The code generated was REST API and API documentation. The main benefits highlighted by

System	Domain	Problem	Agile practice	MDE tools	Code gen.
Kannan et al. [50]	EHR	Misunderstanding requirements, Frequent product change requests, Product defects, Costly and time consuming development	User Stories, User Interface Storyboards, Use Case Text	Class Diagram, Object Diagram, Use Case Diagrams, Decision Threes, Swimlane Workflow diagrams, State Diagrams	No
Blagoj et al. [57]	e-health	Extensibility, Soundness, Interoperability, Standardization	Iterative and gradual development	CIM, PIM, PSM - Workflow diagrams, Class diagram, Abstract state machine diagram, Visual Paradigm	Yes (partial code generation)
Our work	Epi. Surv.	Requirements cannot be fixed and change generally	Sprint, Scrum boards, Scrum meetings, Scrum Team, Product Backlog, Iterative and Incremental	CIM, PIM, PSM - Class Diagram, Eclipse, Imogene	Yes

Table 8: Comparison of existing systems to our system

this new system are the abstraction of the complexity of the system through the MDA model, the standardization of the system to make it easy to exchange information with other systems, and the reduction of the development costs regarding the future improvements of the system.

We summarized this related work section in table 8. It presents the key comparison points of the two related work papers to our system. Given to this table, our work is the only work that proposes a tool for the automatic generation of a whole system. On the other hand, the work we proposed is the only work exploiting the whole potential of the agile methodologies and presenting step by step how a health information system can be modeled and generated using an Agile process and a MDA framework.

VIII CONCLUSION

The main objective of this research was to answer the research question: "how to combine Scrum and MDA to develop and maintain epidemiological surveillance software?" Thus, we put together three empirical methods in software engineering to investigate the combination of Scrum and MDA for the development of epidemiological surveillance software and provide an answer to this research question (see Section 5.4). The survey was used to collect information on the system in place and how the solution should look like. Using Case Study research, we portrayed the development the software for epidemiological surveillance of tuberculosis so that it can be reused for other diseases. Action research allowed us to resolve the problem of failed software developed for epidemiological surveillance by involving health workers in the process of finding the solution.

Globally, we found that when the MDA tool is mature enough to generate the entire source code, the combination of Scrum and MDA allows effective development of epidemiological surveillance software. In effect, combining Scrum and MDA with an appropriate MDA tool has a positive impact on programmer productivity and satisfaction, development cost-effectiveness, timelines, and customer satisfaction. Scrum practices and models developed showed to be efficient practices to validate the evolution of the development and, therefore, facilitates more productive software development because customers are integrated and misunderstanding problems are solved as soon as possible. This was particularly observed during the validation of the Platform Independent Model by the National Tuberculosis Control Program staff and the validation of the software by end users. Thus, to combine Scrum and MDA to develop epidemiological surveillance software, Scrum practices (Involvement of stakeholders, Product Backlog definition, Sprint definition, Scrum iteration, Scrum meetings) must be used to build models and generate related software. The primary novelty of this work is the ability to use Scrum Process

in the modeling and the generation of software. To our knowledge, the combination of Scrum and MDA for the complete development of health software has not been attempted previously.

The direct perspectives to this work are: (1) the application of the approach presented in Section 5.4 into many other settings in order to come up with a generic approach that can be applied in any epidemiological surveillance and other industrial settings. To this end, we are currently working with nutritionists on a system for nutrition (overnutrition and undernutrition) surveillance; (2) generally, many people whose domain experts and computer scientists intervene in the development of epidemiological surveillance systems. We are currently developing a Web version that can allow the collaborative modeling of the system (the same idea as Google Docs). On the other hand, we are working on the automatic generation of the PSM (imogene model) from the PIM (class diagram) and vice versa. To this end, we are exploring solutions such are srcML[21] and papyrus[22].

REFERENCES

Publications

[1] F. D. Davis. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". In: *MIS Q.* 13.3 (Sept. 1989), pages 319–340. ISSN: 0276-7783.

[2] B. Kitchenham, L. Pickard, and S. Pfleeger. "Case studies for method and tool evaluation". In: *IEEE Software* 12.4 (1995), pages 52–62.

[3] D. Damian, A. Eberlein, M. Shaw, and B. Gaines. "Using different communication media in requirements negotiation". In: *IEEE Software* 17.3 (2000), pages 28–36.

[4] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta. "Agile Software Development Methods: Review and Analysis". In: *CoRR* abs/1709.08439 (2002).

[5] L. Bratthall and M. Jørgensen. "Can You Trust a Single Data Source Exploratory Software Engineering Case Study?" In: *Empirical Softw. Engg.* 7.1 (Mar. 2002), pages 9–26. ISSN: 1382-3256.

[6] M. Lenzerini. "Data integration: A theoretical perspective". In: *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. 2002.

[7] C. Larman and V. R. Basili. "Iterative and Incremental Development: A Brief History". In: *Computer* 36.6 (June 2003), pages 47–56. ISSN: 0018-9162.

[8] F. David and P. John. *The Mda Journal: Model Driven Architecture Straight From The Masters*. Meghan Kiffer, 2004. ISBN: 0929652258.

[9] K. Schwaber. *Agile Project Management With Scrum*. Redmond, WA, USA: Microsoft Press, 2004. ISBN: 073561993X.

[10] X. Blanc. *MDA en action, Ingénierie logicielle guidée par les mod'eles*. EYROLLES, 2005.

[11] V. Jones, A. Rensink, and E. Brinksma. "Modelling mobile health systems: an application of augmented MDA for the extended healthcare enterprise." In: *Ninth IEEE International Enterprise Distributed Object Computing Conference (EDOC 2005), 19-23 September 2005, Enschede, The Netherlands*. 2005, pages 58–69.

[12] S. Kemmis and R. Mctaggart. "Participatory Action Research: Communicative Action and the Public Sphere." In: *Handbook of qualitative research* (Jan. 2005).

[21] https://www.srcml.org/

[22] https://www.eclipse.org/papyrus/

[13] Mapatano and Piripiri. “Quelques erreurs courantes d’analyse d’un syst‘eme d’information sanitaire(RD Congo)”. In: *Santé Publique 2005*. Volume 17. 2005.

[14] K. S. Rubin, T. Beale, and B. Blobel. “Modeling for Health Care”. In: *Person-Centered Health Records: Toward HealthePeople™*. Edited by G. A. Demetriades James E.and Christopherson and R. M. Kolodner. New York, NY: Springer New York, 2005, pages 125–146.

[15] F. Truyen. “The Fast Guide to Model Driven Architecture The Basics of Model Driven Architecture”. In: (Jan. 2006).

[16] B. Blobel and P. Pharow. “A model driven approach for the German health telematics architectural framework and security infrastructure”. In: *International journal of medical informatics* 76.2-3 (2007), pages 169–175. ISSN: 1386-5056.

[17] S. Easterbrook, J. Singer, M.-A. Storey, and D. Damian. “Selecting Empirical Methods for Software Engineering Research”. In: *Guide to Advanced Empirical Software Engineering*. Edited by F. Shull, J. Singer, and D. I. K. Sjøberg. London: Springer London, 2008, pages 285–311. ISBN: 978-1-84800-044-5.

[18] J. Mathenge Kanyaru, M. Coles, S. Jeary, and K. Phalp. “Using visualisation to elicit domain information as part of the Model Driven Architecture Approach”. In: *CEUR Workshop Proceedings* 376 (Jan. 2008).

[19] M. Pikkarainen, J. Haikara, O. Salo, P. Abrahamsson, and J. Still. “The impact of agile practices on communication in software development”. In: *Empirical Software Engineering* 13 (June 2008), pages 303–337.

[20] W. Raghupathi and A. Umar. “Exploring a model-driven architecture (MDA) approach to health care information systems development”. In: *International Journal of Medical Informatics* 77.5 (2008), pages 305–314. ISSN: 1386-5056.

[21] P. A. Rottier and V. Rodrigues. “Agile development in a medical device company”. In: *Proceedings - Agile 2008 Conference*. 2008. ISBN: 9780769533216.

[22] R. K. Yin. *Case Study Research: Design and Methods (Applied Social Research Methods)*. Sage Publications, 2008. ISBN: 1412960991.

[23] D. Gasevic, D. Djuric, and V. Devedzic. *Model Driven Engineering and Ontology Development*. 2nd. Springer Publishing Company, Incorporated, 2009.

[24] W. Raghupathi and A. Umar. “Integrated Digital Health Systems Design”. In: *International Journal of Information Technologies and Systems Approach* 2 (June 2009), pages 15–33.

[25] P. Runeson and M. Höst. “Guidelines for conducting and reporting case study research in software engineering”. In: *Empirical Software Engineering* (2009). ISSN: 13823256.

[26] R. Matinnejad. “Agile model driven development: An intelligent compromise”. In: *Proceedings - 2011 9th International Conference on Software Engineering Research, Management and Applications, SERA 2011*. 2011. ISBN: 9780769544908.

[27] AAMI. *Guidance on the use of agile practices in the development of medical device software*. AAMI, 2012. ISBN: 1570204454.

[28] B. C K Choi. “The Past, Present, and Future of Public Health Surveillance”. In: *Scientifica* 2012 (Aug. 2012), page 875253.

[29] T. Dingsøyr, S. Nerur, V. Balijepally, and N. B. Moe. “A Decade of Agile Methodologies”. In: *J. Syst. Softw.* 85.6 (June 2012), pages 1213–1221. ISSN: 0164-1212.

[30] M. McHugh, F. McCaffery, and V. Casey. “Barriers to Adopting Agile Practices When Developing Medical Device Software”. In: *Software Process Improvement and Capability Determination*. Edited by A. Mas, A. Mesquida, T. Rout, R. V. O’Connor, and A. Dorling. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pages 141–147.

[31] M. Mc Hugh, O. Cawley, F. McCaffcry, I. Richardson, and X. Wang. "An agile V-model for medical device software development to overcome the challenges with plan-driven software development lifecycles". In: *2013 5th International Workshop on Software Engineering in Health Care, SEHC 2013 - Proceedings*. 2013. ISBN: 9781467362825.

[32] P. Mohagheghi, W. Gilani, A. Stefanescu, and M. A. Fernandez. "An empirical study of the state of the practice and acceptance of model-driven engineering in four industrial cases". In: *Empirical Software Engineering* (2013). ISSN: 13823256.

[33] P. Mohagheghi, W. Gilani, A. Stefanescu, and M. A. Fernandez. "An empirical study of the state of the practice and acceptance of model-driven engineering in four industrial cases". In: *Empirical Software Engineering* (2013). ISSN: 13823256.

[34] T. Özcan, S. Kocak, and P. Brune. "Agile Software Development with Open Source Software in a Hospital Environment – Case Study of an eCRF-System for Orthopaedical Studies". In: *Web Engineering*. Edited by F. Daniel, P. Dolog, and Q. Li. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pages 439–451.

[35] E. Stringer. *Action Research*. SAGE Publications, 2013. ISBN: 9781483320731.

[36] V. Curcin, T. Woodcock, A. J. Poots, A. Majeed, and D. Bell. "Model-driven approach to data collection and reporting for quality improvement". In: *Journal of Biomedical Informatics* 52 (2014). Special Section: Methods in Clinical Research Informatics, pages 151–162. ISSN: 1532-0464.

[37] U. Eliasson, R. Heldal, J. Lantz, and C. Berger. "Agile model-driven engineering in mechatronic systems - an industrial case study". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2014). ISSN: 16113349.

[38] H. C. Estler, M. Nordio, C. A. Furia, B. Meyer, and J. Schneider. "Agile vs. structured distributed software development: A case study". In: *Empirical Software Engineering* (2014). ISSN: 15737616.

[39] Y. Martínez, C. Cachero, and S. Meliá. "Empirical study on the maintainability of Web applications: Model-driven Engineering vs Code-centric". In: *Empirical Software Engineering* (2014). ISSN: 15737616.

[40] G. Mussbacher, D. Amyot, R. Breu, J.-M. Bruel, B. H. C. Cheng, P. Collet, B. Combemale, R. B. France, R. Heldal, J. Hill, J. Kienzle, M. Schöttle, F. Steimann, D. Stikkolorum, and J. Whittle. "The Relevance of Model-Driven Engineering Thirty Years from Now". In: *Model-Driven Engineering Languages and Systems*. Edited by J. Dingel, W. Schulte, I. Ramos, S. Abrahão, and E. Insfran. Cham: Springer International Publishing, 2014, pages 183–200.

[41] B. Perisic. "Model Driven Software Development – State of The Art and Perspectives". In: *INFOTEH-JAHORINA*. Volume 13. Mar. 2014, pages 1237–1248.

[42] D. Salah, R. F. Paige, and P. Cairns. "A Systematic Literature Review for Agile Development Processes and User Centred Design Integration". In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. EASE '14. London, England, United Kingdom: ACM, 2014, 5:1–5:10. ISBN: 978-1-4503-2476-2.

[43] F. P. Basso, R. M. Pillat, F. Roos-Frantz, and R. Z. Frantz. "Combining MDE and scrum on the rapid prototyping of web information systems". In: *International Journal of Web Engineering and Technology* (2015). ISSN: 17419212.

[44] A. Hajou, R. S. Batenburg, and S. Jansen. "An Insight into the Difficulties of Software Development Projects in the Pharmaceutical Industry". In: *Lecture Notes on Software Engineering* (2015). ISSN: 23013559.

[45] I. Inayat, S. S. Salim, S. Marczak, M. Daneva, and S. Shamshirband. "A Systematic Literature Review on Agile Requirements Engineering Practices and Challenges". In: *Comput. Hum. Behav.* 51.PB (Oct. 2015), pages 915–929. ISSN: 0747-5632.

[46] R. S. Pressman and B. R. Maxim. *Software Engineering : A Practitioner's Approach, Eighth Edition*. 2015. ISBN: 2008048802.

[47] H. Schlieter, M. Burwitz, O. Schonherr, and M. Benedict. "Towards Model Driven Architecture in HealthCare Information System Development". In: *Wirtschaftsinformatik*. Mar. 2015, pages 497–511.

[48] A. R. da Silva. "Model-driven engineering: A survey supported by the unified conceptual model". In: *Computer Languages, Systems Structures* 43 (2015), pages 139–155.

[49] D. A. Henderson. "The Development of Surveillance Systems". In: *American Journal of Epidemiology* 183.5 (Feb. 2016), pages 381–386.

[50] V. Kannan, J. C. Fish, and D. L. Willett. "Agile model driven development of electronic health record-based specialty population registries". In: *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2016, pages 465–468.

[51] H. Alfraihi and K. Lano. "The integration of agile development and model driven development: A systematic literature review". In: *MODELSWARD 2017 - Proceedings of the 5th International Conference on Model-Driven Engineering and Software Development*. 2017. ISBN: 9789897582103.

[52] H. A. A. Alfraihi and K. C. Lano. "The integration of agile development and model driven development: A systematic literature review". In: *The 5th International Confrence on Model-Driven Engineeing and Software Development* (2017).

[53] B. Atansovski, M. Bogdanović, G. Velinov, L. Stoimenov, D. Sahpaski, I. Skrceska, M. Kon-Popovska, D. Janković, and B. Jakimovski. "Transforming an Enterprise E-Health System from Process Oriented to Model Driven Architecture". In: *7th International Conference on Information Society and Technology ICIST*. May 2017, pages 159–162.

[54] H. Bagherian, M. Farahbakhsh, R. Rabiei, H. Moghaddasi, and F. Asadi. "National Communicable Disease Surveillance System: A review on Information and Organizational Structures in Developed Countries". In: *Acta informatica medica* (Dec. 2017).

[55] H. Bagherian, M. Farahbaksh, R. Rabiei, H. Moghaddasi, and F. Asadi. "National Communicable Disease Surveillance System: A review on Information and Organizational Structures in Developed Countries". In: *Acta Informatica Medica* 25 (Dec. 2017), pages 271–276.

[56] C. L. Richards, M. F. Iademarco, D. Atkinson, R. W. Pinner, P. Yoon, W. R. M. Kenzie, B. Lee, J. R. Qualters, and T. R. Frieden. "Advances in Public Health Surveillance and Information Dissemination at the Centers for Disease Control and Prevention". In: *Public Health Reports* 132.4 (2017), pages 403–410.

[57] B. Atanasovski, M. Bogdanovic, G. Velinov, L. Stoimenov, A. S. Dimovski, B. Koteska, D. Jankovic, I. Skrceska, M. Kon-Popovska, and B. Jakimovski. "On defining a model driven architecture for an enterprise e-health system". In: *Enterprise Information Systems* 12.8-9 (2018), pages 915–941.

[58] I. Essebaa and S. Chantit. "Model driven architecture and agile methodologies: Reflexion and discussion of their combination". In: *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, FedCSIS 2018*. 2018. ISBN: 9788394941970.

[59] F. Ricca, M. Torchiano, M. Leotta, A. Tiso, G. Guerrini, and G. Reggio. "On the impact of state-based model-driven development on maintainability: a family of experiments using UniMod". In: *Empirical Software Engineering* (2018). ISSN: 15737616.

[60] F. J. Azanzi, G. Camara, and M. Tchuente. "Extracting ontological knowledge from Java source code using Hidden Markov Models". In: *Open Computer Science* 9.1 (2019), pages 181–199.

[61] Bin, N. Ali, E. Engström, M. Taromirad, M. R. Mousavi, N. M. Minhas, D. Helgesson, S. Kunze, and M. Varshosaz. "On the search for industry-relevant regression testing research". In: *Empirical Software Engineering* (2019). ISSN: 15737616.

[62] A. Jiomekong and G. Camara. "Model-Driven Architecture Based Software Development for Epidemiological Surveillance Systems". In: *Studies in health technology and informatics* 264 (Aug. 2019), pages 531–535. ISSN: 0926-9630.

[63] A. Alami, P. Nielsen, and A. Wasowski. "A Tailored Participatory Action Research for FOSS Communities". English. In: *Empirical Software Engineering* 25.5 (Sept. 2020), pages 3639–3670. ISSN: 1382-3256.

[64] A. Bucchiarone, J. Cabot, R. F. Paige, and A. Pierantonio. "Grand challenges in model-driven engineering: an analysis of the state of the research". In: *Software and Systems Modeling* (2020). ISSN: 16191374.

[65] L. T. Heeager and P. A. Nielsen. "Meshing agile and plan-driven development in safety-critical software: a case study". In: *Empirical Software Engineering* (2020). ISSN: 15737616.

[66] P. Ralph, N. b. Ali, S. Baltes, D. Bianculli, J. Diaz, Y. Dittrich, N. Ernst, M. Felderer, R. Feldt, A. Filieri, B. B. N. de França, C. A. Furia, G. Gay, N. Gold, D. Graziotin, P. He, R. Hoda, N. Juristo, B. Kitchenham, V. Lenarduzzi, J. Martínez, J. Melegati, D. Mendez, T. Menzies, J. Molleri, D. Pfahl, R. Robbes, D. Russo, N. Saarimäki, F. Sarro, D. Taibi, J. Siegmund, D. Spinellis, M. Staron, K. Stol, M.-A. Storey, D. Taibi, D. Tamburri, M. Torchiano, C. Treude, B. Turhan, X. Wang, and S. Vegas. *Empirical Standards for Software Engineering Research*. 2020.

[67] A. S. I. G. on Software Engineering. *Empirical Standards*. 2020.

[68] OMG. *MDA*. `https://www.omg.org/mda/`. 2021.

[69] A. Jiomekong. *EPICAM supports*. July 2022.

[70] A. Jiomekong, H. Tapamo, and G. Camara. "Combining Scrum and Model Driven Architecture for the development of the EPICAM platform". In: *CARI 2022*. Yaounde, Cameroon, Oct. 2022.

A SYSTEM ARCHITECTURE

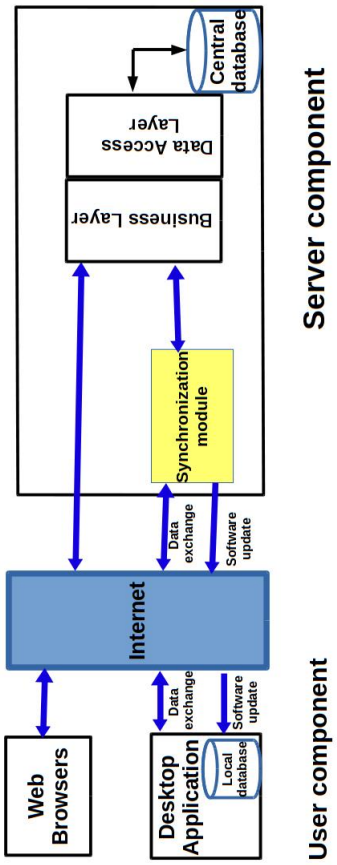

Figure 3: Epidemiological surveillance platform architecture

B PLATFORM INDEPENDENT MODEL

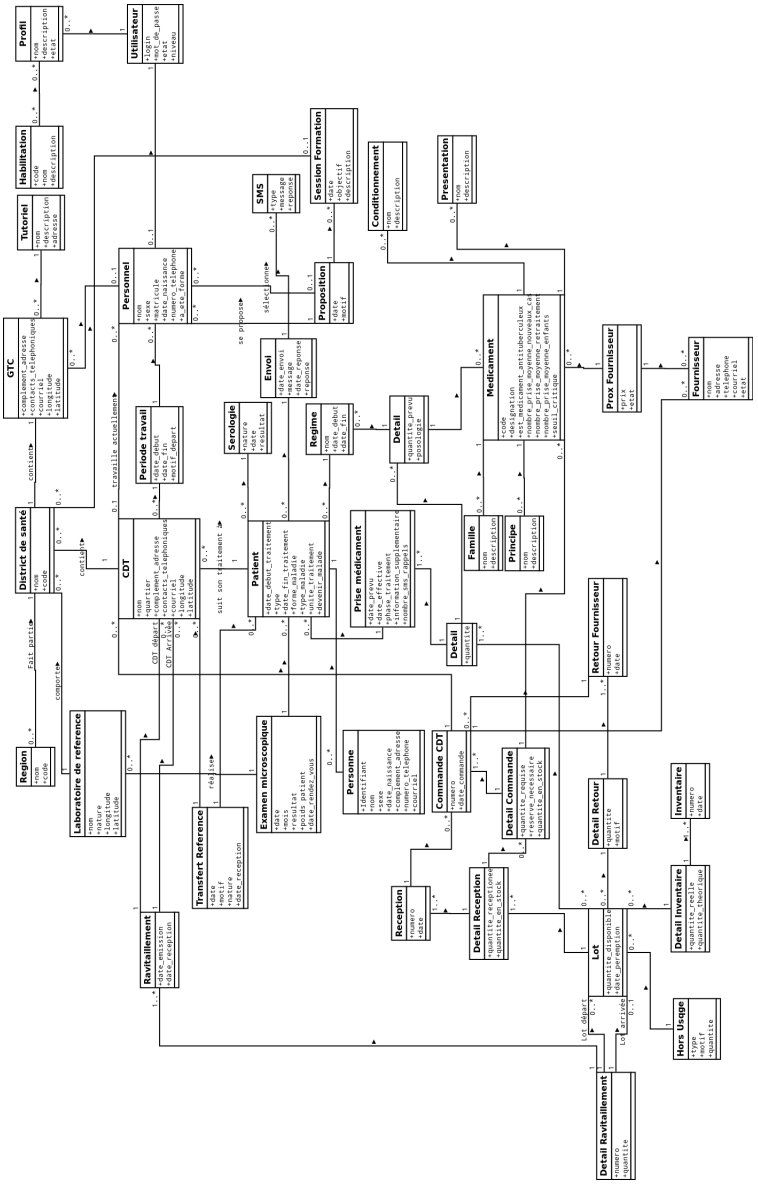

Figure 4: The Platform Independent Model of EPICAM

C PLATFORM SPECIFIC MODEL

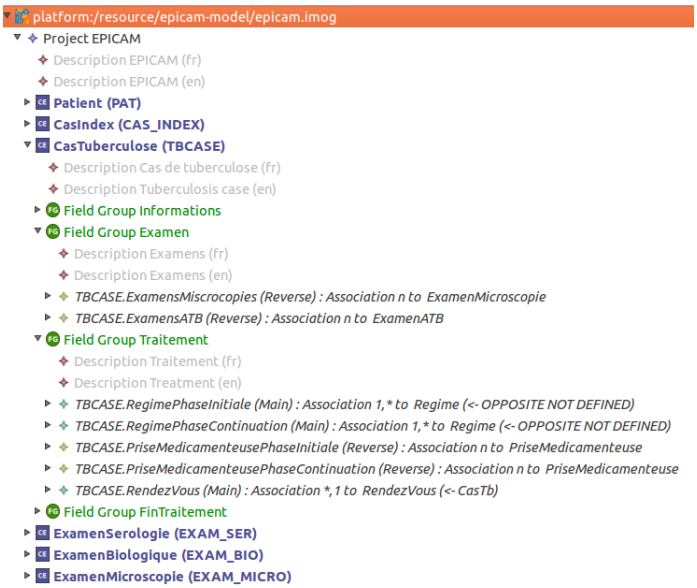

Figure 5: The Platform Specific Model of EPICAM constructed using the PIM

D SOME SCREENSHOTS OF THE APPLICATION GENERATED

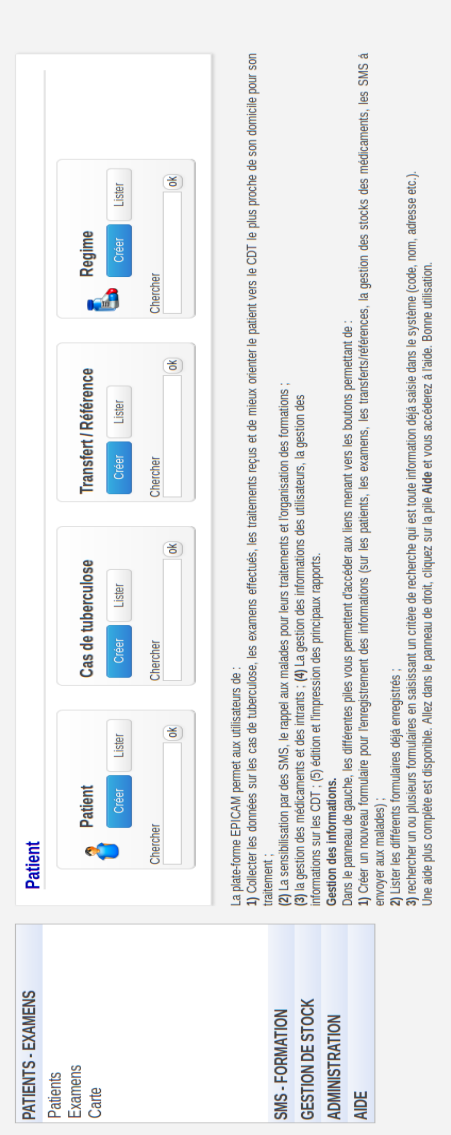

Figure 6: The entry point of the EPICAM platform

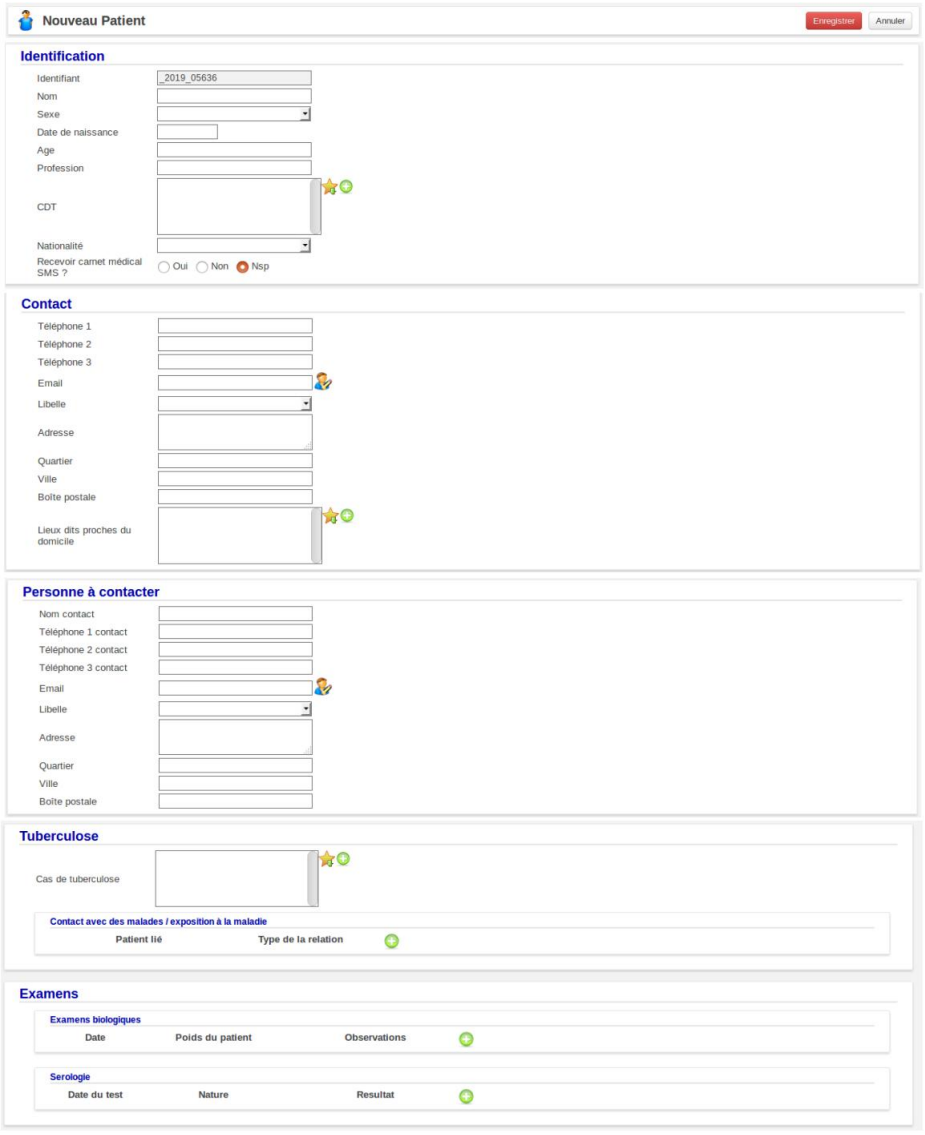

Figure 7: The patient registration form

E SCREENSHOTS FOR USER MANAGEMENT INTERFACE

The following figures presents the screenshots of the UML model of user management and the user interface generated thereafter by the Imogene platform.

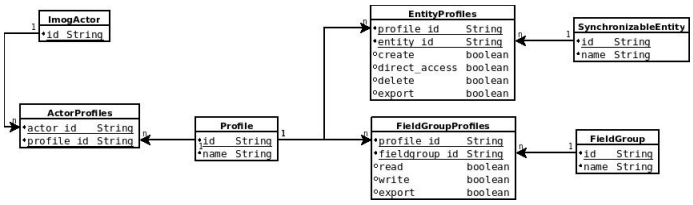

Figure 8: The class diagram of users management in the EPICAM platform

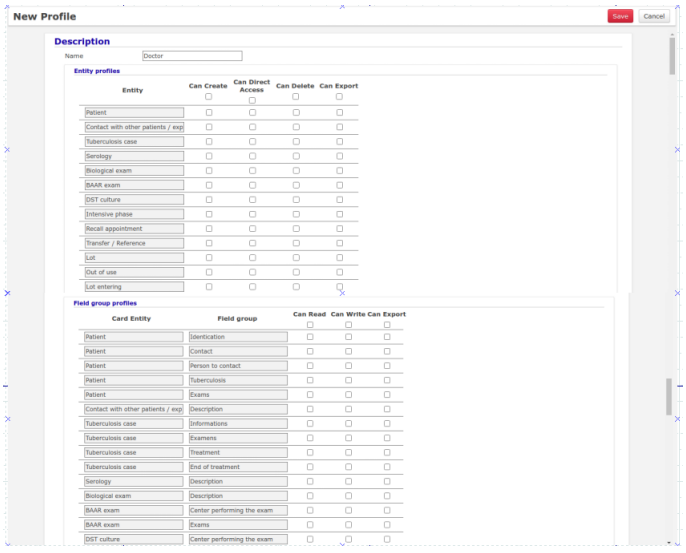

Figure 9: Interface presenting how rights are assigned to users

F RESOURCE AVAILABILITY STATEMENT

This paper is the extended version of the paper [70] accepted at CARI'2022. Additional material [69] that supports this research is available on Zenodo.

G ACKNOWLEDGEMENTS

The EPICAM project was partially supported by grants from PRODESO (Programme Franco-Camerounais pour un Développement Solidaire).

We would like to thank all the health workers for their feedback and constructive comments during the experimentation and the use of the EPICAM platform. Beside, our gratitude goes also to Dr. Nolna Désiré and Dr. Abena Jean Louis from the National Tuberculosis Control Program in Cameroon; Dr. Gaetan Texier from Centre Pasteur du Cameroon; Pr. Maurice Tchuente, from UMMISCO in Cameroon; Mr. Laurent Braak from MEDES in France and finally, all those who contributed to the EPICAM project.

This research was selected as an innovation at "Forum Afrique - 100 innovations pour un développement durable[23]."

[23] Results of selected innovations available at www.diplomatie.gouv.fr/IMG/pdf/Document-Multimedia-FR_cle0857a7.pdf

épisciences

Quantitative stability estimate for inverse coefficients problem in linear elasticity

Houcine Meftahi[1,*] **and Taher Rezgui**[2]

[1]Université de Jendouba/Isikef, Avenue de l'Union du Maghreb Arabe 8189 Jendouba, Tunisie
[1,2]ENIT- Lamsin, BP 37, 1002 Tunis Le Belvédère, Tunisie

*E-mail : houcine.meftahi@enit.utm.tn

DOI : 10.46298/arima.9346
Submitted on April 14, 2022 - Published on December 14, 2023
Volume : **39** - Year : **2023**

Editors : *Mathieu Roche, Nabil Gmati, Bruce Watson, Clémentin Tayou Djamegni*

Abstract

In this article we consider the inverse problem of reconstructing the Lamé coefficients assumed to be piecewise constants from boundary measurements. We reformulate the inverse problem into a minimization one using a Kohn-Vogelius type functional. We study the stability of the parameters when the jump of the discontinuity is perturbed. Using tools of shape calculus, we give a quantitative stability result for local optimal solution.

Keywords

Lamé parameters; Inverse problem; Shape derivative; Stability analysis

I INTRODUCTION

The inverse problem of recovering Lamé parameters consists in finding the parameters (λ, μ) from the knowledge of the so-called Neumann-to-Dirichlet operator $\Lambda_{\lambda,\mu} : H^{-1/2}(\partial\Omega)^2 \to H^{1/2}(\partial\Omega)^2$. The operator $\Lambda_{\lambda,\mu}$ maps a boundary input $g \in H^{-1/2}(\partial\Omega)^2$ to the value $u|_{\partial\Omega}$ of the solution u of an elliptic problem with Neumann boundary condition g:

$$\Lambda_{\lambda,\mu}(g) = u|_{\partial\Omega}.$$

The two major questions for the inverse problem are the uniqueness and the stability of a solution λ, μ. The question of the uniqueness in the case of perfect data, i.e. when the Neumann-to-Dirichlet operator is completely known, has led to many difficult and interesting mathematical problems and an abundant literature on the topic is available.

The question of stability consists in studying the continuous dependence of the solution on the data. Stability is necessary to ensure that a variation of the given data in a sufficiently small range leads to an arbitrarily small change in the solution, which is primordial for applications.

This concept was introduced by Hadamard in 1902 in connection with the study of boundary value problems for partial differential equations [2].

There is a extended literature on stability issues for the inverse problem of Lamé parameters, in particular for interior stability estimates, which consist in proving that, given two maps $\Lambda_{\lambda,\mu}$ and $\Lambda_{\tilde{\lambda},\tilde{\mu}}$, one has

$$\max\left\{\|\lambda-\tilde{\lambda}\|_{L^\infty(\Omega)}, \|\mu-\tilde{\mu}\|_{L^\infty(\Omega)}\right\} \leq \omega\left(\|\Lambda_{\lambda,\mu}-\Lambda_{\tilde{\lambda},\tilde{\mu}}\|_*\right),$$

where $\|.\|_*$ is the operator norm of $\mathcal{L}(H^{-1/2}(\Omega)^2, H^{1/2}(\Omega)^2)$ and $\omega : \mathbb{R}^+ \to \mathbb{R}$ is continuous with $\omega(t) \to 0$ as $t \to 0$. Recently a Lipschitz stability result has been proved in [14]. The proof relies in the monotonicity relation between the Lamé parameters, the Neumann-to-Dirichlet operator and the techniques of Localized potentials.

The inverse problem of reconstructing piecewise constants Lamé parameters λ, μ and their jump sets Γ simultaneously has been considered in [12] and the question of stability is not so well-studied.

In this contribution, we aim at giving qualitative properties of the stability in this case. We transform the inverse problem into a minimization one and study how stable is the reconstruction of the parameters λ, μ when Γ is known approximately. More precisely, we quantify the first-order stability properties of a local optimal solution of the minimization problem.

The stability estimates derived in [14] is obtained when the full Neumann-to-Dirichlet map is known. For our quantitative stability estimate we only have a finite number of measurements. To our best knowledge there are no results in this kind of stability for the problem under consideration.

In the field of PDE-constrained optimization, the differentiability and stability properties of optimal solutions with respect to parameters of the problem have been studied in details theoretically and numerically; see for instance [6–11]. Some of the ideas developed in these papers can be used for our stability analysis, unfortunately the well-known particular difficulties of differentiating with respect to the shape, in particular the fact that the set of shapes is not a vector space, prevents a direct transposition of the methods used in these papers. For our analysis on proving quantitative stability of local optimal solution, we make use of shape calculus in the spirit of [13] .

The paper is organized as follows. In section 2 we describe the setting of the direct problem, the inverse problem and the minimization problem. The stability analysis of a local optimal solution is performed in section 3.

II PROBLEM FORMULATION

Let Ω be a bounded domain in $\mathbb{R}^2$ with smooth boundary $\partial\Omega = \Gamma_N \cup \Gamma_D$ and $\Gamma \in \mathcal{D}_{ad}$ with

$$\mathcal{D}_{ad} := \{\partial D \text{ where } D \text{ is open of class } C^2, D \subset \Omega, \inf_{x\in\partial D, y\in\partial\Omega} |x-y| > \varepsilon\}$$

for some $\varepsilon > 0$. Let Ω_2 be such that $\Gamma := \partial\Omega_2$ and $\Omega_1 = \Omega \setminus \overline{\Omega_2}$. Throughout the paper we work with piecewise constant parameters

$$\lambda = \lambda_1\chi_{\Omega_1} + \lambda_2\chi_{\Omega_2}, \qquad \mu = \mu_1\chi_{\Omega_1} + \mu_2\chi_{\Omega_2}$$

where $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}_+$ and χ denotes an indicator function. Let $0 < K_0 \leq \lambda_1, \mu_1 \leq K_1$(The parameters λ_1, μ_1 are given and not to be identified), then we define a set of piecewise constant parameters $\mathcal{P}_{ad}$ as :

$$\mathcal{P}_{ad} = \{(\lambda, \mu) = (\lambda_1 \chi_{\Omega_1} + \lambda_2 \chi_{\Omega_2}, \mu_1 \chi_{\Omega_1} + \mu_2 \chi_{\Omega_2}) : c_0 \leq, \lambda_2, \mu_2 \leq c_1\},$$

where $c_0, c_1 \in \mathbb{R}_+$. For a given load $g \in H^{-1/2}(\Gamma_N)^2$ the displacement u satisfies the following problem

$$\begin{cases} -\text{div}(\mathbb{C}\hat{\nabla} u) = 0 & \text{in } \Omega \setminus \Gamma, \\ [\![u]\!] = 0 & \text{on } \Gamma, \\ [\![(\mathbb{C}\hat{\nabla} u)\nu]\!] = 0 & \text{on } \Gamma, \\ (\mathbb{C}\hat{\nabla} u)\nu = g & \text{on } \Gamma_N, \\ u = f & \text{on } \Gamma_N, \\ u = 0 & \text{on } \Gamma_D, \end{cases} \tag{1}$$

where ν is the unit normal vector to the interface Γ or $\partial\Omega$ pointing outward of Ω_2 or Ω, respectively, $f \in H^{1/2}(\Gamma_N)^2$ and $[\![u]\!]$ denotes the jump of u across the interface Γ. The linearized strain tensor $\hat{\nabla} u$ and the stress tensor $\mathbb{C}\hat{\nabla} u$ are given by

$$\hat{\nabla} u = \frac{1}{2}\left(\nabla u + (\nabla u)^T\right), \qquad \mathbb{C}\hat{\nabla} u = \left(\sum_{k,l=1}^{2} \mathbb{C}_{ijkl} \frac{\partial u_k}{\partial x_l}\right)_{1 \leq i,j \leq 2},$$

with

$$\mathbb{C}_{ijkl} := \lambda \delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right).$$

It is well-known that problem (1) has a unique solution $u \in \mathcal{V}$, where

$$\mathcal{V} := \left\{u \in H^1(\Omega)^2 : \quad u|_{\Gamma_D} = 0\right\}.$$

We assume that the parameters λ_1, μ_1 are known and we consider the inverse problem of recovering the parameters λ_2, μ_2 from the measurements (f, g). We are interested in the case where the interface Γ is known, but with some uncertainty. We aim to study how this uncertainty affects the reconstruction of λ_2^*, μ_2^*. Note that the results could be straightforwardly extended to the case of more than just two phases, thus for clarity it is better to study the case of two phases.

A typical approach to solve the inverse problem in practice is to consider the so-called *Kohn-Vogelius functional*. We add a regularization term which is required due to the ill-posedness of the problem

$$J((\lambda_2, \mu_2), \Gamma, u_n, u_d) := \int_\Omega \mathbb{C}\hat{\nabla}(u_n - u_d) : \nabla(u_n - u_d)\, dx + \eta \int_{\Omega_2} (\lambda_2^2 + \mu_2^2)\, dx,$$

where $\eta > 0$. Here u_n is the solution of the Neumann problem

$$\begin{cases} -\text{div}(\mathbb{C}\hat{\nabla} u_n) = 0 & \text{in } \Omega \setminus \Gamma, \\ [\![u_n]\!] = 0 & \text{on } \Gamma, \\ [\![(\mathbb{C}\hat{\nabla} u_n)\nu]\!] = 0 & \text{on } \Gamma, \\ (\mathbb{C}\hat{\nabla} u_n)\nu = g & \text{on } \Gamma_N, \\ u_n = 0 & \text{on } \Gamma_D, \end{cases} \tag{2}$$

and u_d is the solution of the Dirichlet problem

$$\begin{cases} -\operatorname{div}(\mathbb{C}\hat{\nabla}u_d) = 0 & \text{in } \Omega \setminus \Gamma, \\ [\![u_d]\!] = 0 & \text{on } \Gamma, \\ [\![(\mathbb{C}\hat{\nabla}u_d)\nu]\!] = 0 & \text{on } \Gamma, \\ u_d = f & \text{on } \Gamma_N, \\ u_d = 0 & \text{on } \Gamma_D. \end{cases} \tag{3}$$

The variational formulations of the problems (2) and (3) read respectively

$$\text{Find } u \in \mathcal{V} \text{ such that } \int_\Omega \mathbb{C}\hat{\nabla}u : \nabla v \, dx - \int_{\Gamma_N} g \cdot v \, ds = 0 \quad \forall v \in H^1(\Omega), \tag{4}$$

$$\text{Find } u \in \mathcal{V} \text{ such that } \int_\Omega \mathbb{C}\hat{\nabla}u : \nabla v \, dx - \int_{\Gamma_N} (u-f)(\mathbb{C}\hat{\nabla}u)\nu v \, ds = 0 \quad \forall v \in H^1_0(\Omega)^2. \tag{5}$$

We associate with (4) the operator $E_n((\lambda_2, \mu_2), \Gamma, u_n) : [c_0, c_1]^2 \times \mathcal{D}_{ad} \times \mathcal{V} \to \mathcal{L}(H^1(\Omega), \mathbb{R})$ and with (5) the operator $E_d((\lambda_2, \mu_2), \Gamma, u_d) : [c_0, c_1]^2 \times \mathcal{D}_{ad} \times \mathcal{V} \to \mathcal{L}(H^1_0(\Omega), \mathbb{R}) = H^{-1}(\Omega)$. We also introduce

$$E((\lambda_2, \mu_2), \Gamma, u_n, u_d) := \left(E_n((\lambda_2, \mu_2), \Gamma, u_n), E_d((\lambda_2, \mu_2), \Gamma, u_d)\right).$$

We are interested in the situation when the interface Γ is known and one wishes to reconstruct the coefficients λ_2, μ_2, the corresponding minimization problem is:

$$\begin{cases} \text{minimize } J((\lambda_2, \mu_2), \Gamma, u_n, u_d) := \displaystyle\int_\Omega \mathbb{C}\hat{\nabla}(u_n - u_d) : \hat{\nabla}(u_n - u_d) \, dx + \eta \int_{\Omega_2} (\lambda_2^2 + \mu_2^2) \, dx \\ \text{subject to } c_0 \le \lambda_2, \mu_2 \le c_1 \text{ and } E((\lambda_2, \mu_2), \Gamma, u_n, u_d) = (0, 0). \end{cases} \tag{6}$$

Note an optimal solution $(\lambda_2^*, \mu_2^*) = (\lambda_2^*(\Gamma), \mu_2^*(\Gamma))$ of (6), if it exists, depends on Γ through the state equations.

Let us define the reduced functional corresponding to problem (6).

$$\mathcal{J}((\lambda_2, \mu_2), \Gamma) := J(\lambda_2, \mu_2, \Gamma, u_n, u_d)$$

We start with the following theorem.

Theorem 1:
The minimization problem (6) has at least one solution.

Proof. In this proof we only write $\mathcal{J}(\lambda_2, \mu_2)$ instead of $\mathcal{J}((\lambda_2, \mu_2), \Gamma)$ for simplicity since Γ is fixed. It is clear that $\inf \mathcal{J}(\lambda_2, \mu_2)$ is finite. Therefore there exists a minimizing sequence

$$\left(\lambda^k, \mu^k\right) = \left(\lambda_1 \chi_{\Omega_1} + \lambda_2^k \chi_{\Omega_2}, \mu_1 \chi_{\Omega_1} + \mu_2^k \chi_{\Omega_2}\right) \in \mathcal{P}_{ad}$$

such that

$$\lim_{k \to +\infty} \mathcal{J}(\lambda_2^k, \mu_2^k) = \inf_{c_0 \le \lambda_2, \mu_2 \le c_1} \mathcal{J}(\lambda_2, \mu_2).$$

The sequence (λ_2^k, μ_2^k) is bounded, thus there exists a subsequence still denoted (λ_2^k, μ_2^k) and some $(\lambda_2^*, \mu_2^*) \in \mathbb{R}^2$ such that

$$\lim_{k\to\infty} \left(\lambda^k, \mu^k\right) = \left(\lambda_1\chi_{\Omega_1} + \lambda_2^*\chi_{\Omega_2}, \mu_1\chi_{\Omega_1} + \mu_2^*\chi_{\Omega_2}\right).$$

By definition $u_n(\lambda_2^k, \mu_2^k)$ satisfies

$$\int_\Omega \mathbb{C}(\lambda_2^k, \mu_2^k)\hat{\nabla} u_n(\lambda_2^k, \mu_2^k) : \nabla v \, dx = \int_{\Gamma_N} g \cdot v \, ds \quad \forall v \in \mathcal{V}. \tag{7}$$

Taking $v = u_n(\lambda_2^k, \mu_2^k)$ in (7), we obtain

$$\int_\Omega \mathbb{C}(\lambda_2^k, \mu_2^k)\hat{\nabla} u_n(\lambda_2^k, \mu_2^n) : \nabla u_n(\lambda_2^k, \mu_2^n) \, dx = \int_{\Gamma_N} g \cdot u_n(\lambda_2^k, \mu_2^k) \, ds.$$

Using the Korn's inequality and a trace theorem for the right-hand side of the above equation, we deduce the existence of a constant $c > 0$ such that

$$\|u_n(\lambda_2^k, \mu_2^k)\|_{H^1(\Omega)^2} \leq c\|g\|_{L^2(\Gamma_N)^2}.$$

Therefore there exists a subsequence of $u_n(\lambda_2^k, \mu_2^k)$ still denoted $u_n(\lambda_2^k, \mu_2^k)$ such that

$$\lim_{k\to\infty} u_n(\lambda_2^k, \mu_2^k) = u_n^* \text{ weakly in } H^1(\Omega)^d,$$

and

$$\lim_{k\to\infty} u_n(\lambda_2^k, \mu_2^k) = u_n^* \text{ strongly in } L^2(\Omega)^d.$$

Letting n go to infinity in equation (7), we conclude that u_n^* satisfies

$$\int_\Omega \mathbb{C}(\lambda_2^*, \mu_2^*)\hat{\nabla} u_n^* : \nabla v \, dx = \int_{\Gamma_N} g \cdot v \, ds, \quad \forall v \in \mathcal{V}.$$

Due to the uniqueness of the weak limit we get $u_n(\lambda_2^*, \mu_2^*) = u_n^*$. This means that

$$\lim_{k\to\infty} u_n(\lambda_2^k, \mu_2^k) = u_n(\lambda_2^*, \mu_2^*) \text{ weakly in } H^1(\Omega)^2 \text{ and strongly in } L^2(\Omega)^2.$$

In the same way, we can prove that

$$\lim_{k\to\infty} u_d(\lambda_2^k, \mu_2^k) = u_d(\lambda_2^*, \mu_2^*) \text{ weakly in } H^1(\Omega)^2 \text{ and strongly in } L^2(\Omega)^2.$$

Using the lower semi-continuity of the H^1-norm yields

$$\mathcal{J}(\lambda_2^*, \mu_2^*) \leq \liminf_{k\to\infty} \mathcal{J}(\lambda_2^k, \mu_2^k) = \mathcal{J}(\lambda_2, \mu_2),$$

which concludes the proof. □

2.1 Elements of shape calculus

In this subsection we recall some basic facts about the perturbation of the identity method from shape optimization used to calculate the shape derivative. The reader is referred to [3, 4] for more details.

Let $V \in \mathcal{D}^1(\Omega, \mathbb{R}^2)$, the space of continuous differentiable functions with compact support in Ω. Introduce the perturbation of identity

$$T_t = I + tV : \Omega \to \mathbb{R}^2. \tag{8}$$

Then there exists $\varepsilon > 0$ such that T_t is a diffeomorphism and $T_t(\Omega) = \Omega$ for $t \in [0, \varepsilon)$. We denote the perturbed interfaces

$$\Gamma_t := T_t(\Gamma).$$

For $t \in [0, \varepsilon)$, T_t is invertible. Furthermore, the Jacobian $\xi(t)$ is strictly positive

$$\forall\, t \in [0, \varepsilon), \quad \xi(t) = |\text{det}DT_t| > 0, \tag{9}$$

where DT_t is the Jacobian matrix of the transformation T_t associated with the velocity field V. In the sequel, we use the notation M^{-1} for the inverse of M and M^{-*} for the transpose of its inverse. We also denote by

$$w(t) = \xi(t)|(DT_t)^{-*}\nu| \tag{10}$$

the tangential Jacobian of T_t on Γ.

Definition 1: Eulerian derivative
Suppose we are given a real valued shape function J defined on a subset D of $\mathbb{R}^2$. We say that J is *Eulerian semi-differentiable* at $\Omega \subset D$ in the direction V if the following limit exists in $\mathbb{R}$

$$DJ(\Omega)(V) := \lim_{t \searrow 0} \frac{J(T_t(\Omega)) - J(\Omega)}{t}.$$

If the map $V \longrightarrow DJ(\Omega)(V)$ is linear and continuous with respect to the topology of $\mathcal{D}^1(\Omega, \mathbb{R}^2)$, then J is said to be shape differentiable at Ω and $DJ(\Omega)(V)$ is called the shape derivative of J.

The Eulerian derivative is only a directional derivative. In this paper we also need the stronger notion of *Fréchet derivative*. To this end let $V \in \mathcal{V}_0$ and consider perturbations of identity $I + V$ where $\mathcal{V}_0$ is in a neighborhood of 0 in $\mathcal{D}^1(\Omega, \mathbb{R}^2)$ so that $T_V := I + V$ is a bi-Lipschitz homeomorphism. In what follows we will denote by

$$\Omega_V := T_V(\Omega)$$

Definition 2: Fréchet derivative
The functional $J(\Omega)$ is Fréchet-differentiable at Ω if there exists a linear and continuous functional $dJ(\Omega)$ from $\mathcal{D}^1(\Omega, \mathbb{R}^2)$ to $\mathbb{R}$ called *shape gradient* such that

$$J(\Omega_V) = J(\Omega) + dJ(\Omega)(V) + r(V),$$

where $|r(V)|/\|V\| \to 0$ as $\|V\| \to 0$.

In what follows we will mostly compute Fréchet derivatives given by Definition 2, however to obtain the expression of the shape derivative, it is convenient to simply compute the directional derivative given by Definition 1.

2.2 Inf-sup formulation

From the definition of the functional $\mathcal{J}$, and applying Green's formula once, we have

$$\mathcal{J}(\lambda_2,\mu_2) := J(\lambda_2,\mu_2,u_n,u_d) = J_0(\lambda_2,\mu_2,u_n) + J_0(\lambda_2,\mu_2,u_d) + J_1,$$

where

$$J_0(\lambda_2,\mu_2,u) = \int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}u : \hat{\nabla}u\,dx, \qquad J_1 = -2\int_{\Gamma_N} f\cdot g\,ds.$$

We introduce the Lagrangian functionals

$$G_n(\lambda_2,\mu_2,\varphi,\psi) = J_0(\lambda_2,\mu_2,\varphi) + \int_\Omega \mathbb{C}\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx - \int_{\Gamma_N} g\cdot\psi\,ds \quad \text{for all } \varphi,\psi\in\mathcal{V},$$

$$G_d(\lambda_2,\mu_2,\varphi,\psi) = J_0(\lambda_2,\mu_2,\varphi) + \int_\Omega \mathbb{C}\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx + \int_{\Gamma_N} (f-u_D)\cdot(\mathbb{C}\hat{\nabla}\psi)\nu\,ds,$$

for all $\varphi\in\mathcal{V},\psi\in H_0^1(\Omega;\mathbb{R}^2)$. Then, it is easy to check that

$$J_0(\lambda_2,\mu_2,u_n) = \inf_{\varphi\in\mathcal{V}}\sup_{\psi\in\mathcal{V}} G_n(\lambda_2,\mu_2,\varphi,\psi),$$

$$J_0(\lambda_2,\mu_2,u_d) = \inf_{\varphi\in\mathcal{V}}\sup_{\psi\in H_0^1(\Omega;\mathbb{R}^2)} G_d(\lambda_2,\mu_2,\varphi,\psi),$$

since

$$\sup_{\psi\in\mathcal{V}} G_n(\lambda_2,\mu_2,\varphi,\psi) = \begin{cases} J_0(\lambda_2,\mu_2,u_n) & \text{if } \varphi = u_n),\\ +\infty & \text{otherwise,}\end{cases}$$

$$\sup_{\psi\in H_0^1(\Omega;\mathbb{R}^2)} G_d(\mathbb{C},\varphi,\psi) = \begin{cases} J_0(\lambda_2,\mu_2,u_d) & \text{if } \varphi = u_d),\\ +\infty & \text{otherwise.}\end{cases}$$

It is easily shown that the functional G_n (respectively G_d) is convex continuous with respect to φ and concave continuous with respect to ψ. Therefore, according to Ekeland and Temam [5], the functional G_n has a saddle point (u_n,v_n) if and only if (u_n,v_n) solves the following system:

$$\partial_\psi G_n(\lambda_2,\mu_2,u_n,v_n;\hat{\psi}) = 0,$$
$$\partial_\varphi G_n(\lambda_2,\mu_2,u_n,v_n;\hat{\varphi}) = 0,$$

for all $\hat{\psi}\in\mathcal{V}$ and $\hat{\varphi}\in\mathcal{V}$. This yields that G_n has a saddle point (u_n,v_d), where the state u_n is the unique solution of (2) and the adjoint state v_n is the solution of $\partial_\varphi G_n(\lambda_2,\mu_2,u_n,v_n;\hat{\varphi}) = 0$, or equivalently:

$$\begin{cases} -\text{div}(\mathbb{C}\hat{\nabla}v_n) = 0 & \text{in } \Omega,\\ (\mathbb{C}\hat{\nabla}v_n)\nu = -2g & \text{on } \Gamma_N,\\ v_n = 0 & \text{on } \Gamma_D.\end{cases} \tag{11}$$

Similarly, the Lagrangian G_d has a unique saddle point (u_d,v_d) where the direct state u_d is the solution of the problem (3) and the adjoint state v_d is the unique solution of the following adjoint problem

$$\begin{cases} -\text{div}(\mathbb{C}\hat{\nabla}v_d) = 0 & \text{in } \Omega,\\ v_d = 0 & \text{on } \Gamma_N,\\ v_d = 0 & \text{on } \Gamma_D.\end{cases} \tag{12}$$

Summarizing the above, we have obtained

Theorem 2:
The functionals $J_0(\lambda_2, \mu_2, u_n)$ and $J_0(\lambda_2, \mu_2, u_d)$ are given as

$$J_0(\lambda_2, \mu_2, u_n) = \inf_{\varphi \in \mathcal{V}} \sup_{\psi \in \mathcal{V}} G_n(\lambda_2, \mu_2, \varphi, \psi), \tag{13}$$

$$J_0(\lambda_2, \mu_2, u_d) = \inf_{\varphi \in \mathcal{V}} \sup_{\psi \in H_0^1(\Omega;\mathbb{R}^d)} G_d(\lambda_2, \mu_2, \varphi, \psi). \tag{14}$$

The unique saddle points for G_n and G_d are respectively given by (u_n, v_n) and (u_d, v_d), where $v_n = -2u_n$ and $v_d = 0$.

III STABILITY OF THE PARAMETERS WITH RESPECT TO THE INTERFACE

The optimality condition $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2, \mu_2, \Gamma)(\alpha, \beta) = 0$ for all $\alpha, \beta \in \mathbb{R}$ for problem (6) can be rewritten as

$$\text{Find } (\lambda_2^*(\Gamma), \mu_2^*(\Gamma)) \text{ such that } D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*(\Gamma), \mu_2^*(\Gamma), \Gamma) = (0, 0). \tag{15}$$

In this paper we are interested in the stability of the optimal solution $(\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V))$ of the minimization problem (6) with respect to the perturbed interface $\Gamma_V = (I + V)(\Gamma)$ for $V \in \mathcal{V}_0$, where $\mathcal{V}_0$ is a neighborhood of 0 in $\mathcal{D}^1(\Omega, \mathbb{R}^2)$.

Therefore we need to study the differentiability of the parameter-to-solution map $V \mapsto (\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V))$. This is done by applying the implicit function theorem to the optimality conditions

$$D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V), \Gamma_V) = (0, 0).$$

Essentially the idea is to linearize $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V), \Gamma_V) = (0, 0)$ to get

$$D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)\left(\partial_\Gamma\lambda_2^*(\Gamma)(V), \partial_\Gamma\mu_2^*(\Gamma)(V)\right) + \partial_\Gamma(D_{(\lambda_2,\mu_2),\Gamma}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma))(V) = (0, 0),$$

which yields, when $D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)$ is invertible,

$$(\partial_\Gamma\lambda_2^*(\Gamma)(V), \partial_\Gamma\mu_2^*(\Gamma)(V)) = -\left(D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)\right)^{-1}\partial_\Gamma\left(D_{(\lambda_2,\mu_2),\Gamma}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)\right)(V).$$

In the following theorem, we prove the existence of $(\lambda_2^*(\Gamma), \mu_2^*(\Gamma))$ and we give an explicit formula for its variation with respect to Γ.

Theorem 3:
For given $\Gamma \in \mathcal{D}_{ad}$ suppose that there exists $(\lambda_2^*(\Gamma), \mu_2^*(\Gamma))$ such that

$$D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma) = (0, 0),$$

and assume that $D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)$ is invertible. Then, there exists a neighborhood $\mathcal{W}$ of 0 in $\mathcal{D}^1(\mathbb{R}^2, \mathbb{R}^2)$, a neighborhood $\mathcal{U}$ of (λ_2^*, μ_2^*) in $\mathbb{R}^2$ and a C^1 function

$$\mathcal{W} \ni V \longmapsto (\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V)) \in \mathcal{U},$$

such that $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*(\Gamma_V), \mu_2^*(\Gamma_V), \Gamma_V) = (0, 0)$, and

$$(\partial_\Gamma\lambda_2^*(\Gamma)(V), \partial_\Gamma\mu_2^*(\Gamma)(V)) = -\left(D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)\right)^{-1}\partial_\Gamma\left(D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*, \mu_2^*, \Gamma)\right)(V). \tag{16}$$

Remark 1:
Theorem 3 says that if Γ_V is the perturbation of Γ by the transformation $I+V$, then the first order variation of the optimal solution is given by

$$(\lambda_2^*(\Gamma_V),\mu_2^*(\Gamma_V))-(\lambda_2^*(\Gamma),\mu_2^*(\Gamma)) \approx -\left(D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)\right)^{-1}\partial_\Gamma\left(D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)\right)(V).$$

Proof. According to Theorem 5 and Theorem 6, the function $D_{(\lambda_2,\mu_2)}\mathcal{J}$ is of class C^1 and in view of our assumptions satisfy $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)=(0,0)$ and $D^2_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)$ is invertible, we conclude the proof thanks to the implicit function theorem. □

3.1 First-order derivatives of the cost functional

In the following theorem, we give the derivative of the functional J with respect to the Lamé parameters (λ_2,μ_2).

Theorem 4:
The functional J is differentiable, and its derivative at (λ_2,μ_2) in the direction $(\tilde{\lambda}_2,\tilde{\mu}_2)$ is given by

$$\begin{aligned} D_{(\lambda_2,\mu_2)}J\left(\lambda_2,\mu_2,u_n,u_d\right)(\tilde{\lambda}_2,\tilde{\mu}_2) = &\int_\Omega \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}u_d : \hat{\nabla}u_d\,dx \\ &-\int_\Omega \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}u_n : \hat{\nabla}u_n\,dx. \end{aligned} \tag{17}$$

Proof. Let $\lambda_2^t=\lambda_2+t\tilde{\lambda}_2$, $\mu_2^t=\mu_2+t\tilde{\mu}_2$, $\mathbb{C}=\mathbb{C}(\lambda_2,\mu_2)$, $\tilde{\mathbb{C}}=\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)$, $\mathbb{C}_t=\mathbb{C}+t\tilde{\mathbb{C}}$ and $t\in\mathbb{R}$ is sufficiently small parameter. Under hypotheses of Theorem 5.1 of R. Correa and A. Seeger [1], we have

$$D_{(\lambda_2,\mu_2)}J\left(\lambda_2,\mu_2,u_n,u_d\right)(\tilde{\lambda}_2,\tilde{\mu}_2) = \partial_t\tilde{G}_n(t,u_n,v_n)\Big|_{t=0} + \partial_t\tilde{G}_d(t,u_d,v_d)\Big|_{t=0},$$

where

$$\tilde{G}_n(t,\varphi,\psi) := G_n(\lambda_2^t,\mu_2^t,\varphi,\psi) = J_0(\lambda_2^t,\mu_2^t,\varphi) + \int_\Omega \mathbb{C}_t\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx - \int_{\Gamma_N} g\cdot\psi\,ds,$$

$$\begin{aligned} \tilde{G}_d(t,\varphi,\psi) := G_d(\lambda_2^t,\mu_2^t,\varphi,\psi) = J_0(\lambda_2^t,\mu_2^t,\varphi) &+ \int_\Omega \mathbb{C}_t\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx \\ &+ \int_{\Gamma_N}(f-\varphi)\cdot(\mathbb{C}\hat{\nabla}\psi)\nu\,ds, \end{aligned}$$

and

$$\partial_t\tilde{G}_n(t,u_n,v_n)\Big|_{t=0} = -\int_\Omega \tilde{\mathbb{C}}\hat{\nabla}u_n : \hat{\nabla}u_n\,dx,$$

$$\partial_t\tilde{G}_d(t,u_g,v_d)\Big|_{t=0} = \int_\Omega \tilde{\mathbb{C}}\hat{\nabla}u_d : \hat{\nabla}u_d\,dx.$$

The above equations yield (17). To end the proof, we should verify the four assumptions $(H_1)-(H_4)$ of Theorem 5.1 of R. Correa and A. Seeger [1]. We introduce the sets

$$X_n(t) := \left\{ x^t \in \mathcal{V} : \sup_{y\in\mathcal{V}} \tilde{G}_n(t,x^t,y) = \inf_{x\in\mathcal{V}} \sup_{y\in\mathcal{V}} \tilde{G}_n(t,x,y) \right\},$$

$$Y_n(t) := \left\{ y^t \in \mathcal{V} : \inf_{x\in\mathcal{V}} \tilde{G}_n(t,x,y^t) = \sup_{y\in\mathcal{V}} \inf_{x\in\mathcal{V}} \tilde{G}_n(t,x,y) \right\},$$

$$X_d(t) := \left\{ x^t \in \mathcal{V} : \sup_{y\in H_0^1(\Omega;\mathbb{R}^d)} \tilde{G}_d(t,x^t,y) = \inf_{x\in\mathcal{V}} \sup_{y\in H_0^1(\Omega;\mathbb{R}^2)} \tilde{G}_d(t,x,y) \right\},$$

$$Y_d(t) := \left\{ y^t \in H_0^1(\Omega;\mathbb{R}^d) : \inf_{x\in\mathcal{V}} \tilde{G}_d(t,x,y^t) = \sup_{y\in H_0^1(\Omega;\mathbb{R}^2)} \inf_{x\in\mathcal{V}} \tilde{G}_d(t,x,y) \right\},$$

and obtain

$$\text{for all } t\in[0,\varepsilon] \quad S_n(t) = X_n(t)\times Y_n(t) = \{u_n(\mathbb{C}_t), v_n(\mathbb{C}_t)\} \neq \emptyset,$$

$$\text{for all } t\in[0,\varepsilon] \quad S_d(t) = X_d(t)\times Y_d(t) = \{u_{\mathrm{D}}(\mathbb{C}_t), v_d(\mathbb{C}_t)\} \neq \emptyset,$$

and assumption (H_1) is satisfied.

Assumption (H_2)*:* The partial derivatives $\partial_t\tilde{G}_n(t,\varphi,\psi), \partial_t\tilde{G}_d(t,\varphi,\psi)$ exist everywhere in $[0,\varepsilon)$ and the condition (H_2) is satisfied.
Assumptions (H_3) *and* (H_4)*:* We first show the boundedness of $(u_{\mathrm{N}}(\mathbb{C}_t), v_{\mathrm{N}}(\mathbb{C}_t))$. Letting $v = u_n(\mathbb{C}_t)$ in the variational equation

$$\int_\Omega \mathbb{C}_t\hat{\nabla}u_n(\mathbb{C}_t) : \hat{\nabla}v\,dx = \int_{\Gamma_{\mathrm{N}}} g\cdot v\,ds, \tag{18}$$

for all $v\in\mathcal{V}$, we obtain

$$\int_\Omega \mathbb{C}_t\hat{\nabla}u_n : \hat{\nabla}u_n\,dx \leq \|g\|_{L^2(\Gamma_{\mathrm{N}};\mathbb{R}^2)}\|u_n\|_{L^2(\Gamma_{\mathrm{N}};\mathbb{R}^2)}.$$

From Korn's inequality and the trace theorem, there exists $c>0$, depending only on Ω such that

$$\|u_{\mathrm{n}}(\mathbb{C}_t)\|_{H^1(\Omega;\mathbb{R}^d)} \leq c\|g\|_{L^2(\Gamma_{\mathrm{N}};\mathbb{R}^2)},$$

which yields

$$\sup_{t\in[0,\varepsilon)} \|u_n(\mathbb{C}_t)\|_{H^1(\Omega;\mathbb{R}^2)} \leq c\|g\|_{L^2(\Gamma_{\mathrm{N}};\mathbb{R}^2)}.$$

We apply the same technique to the variational equation

$$\int_\Omega \mathbb{C}_t\hat{\nabla}u_d(\mathbb{C}_t) : \hat{\nabla}v\,dx + \int_{\Gamma_{\mathrm{N}}} (f-u_d(\mathbb{C}_t))\cdot(\mathbb{C}\hat{\nabla}v)\nu\,ds, \tag{19}$$

for all $v \in \mathcal{V}$, and we are able to show that the function $u_d(\mathbb{C}_t)$ is bounded. The next step is to show the continuity with respect to t of $(u_n(\mathbb{C}_t), u_d(\mathbb{C}_t))$. Subtracting (18) at $t > 0$ and $t = 0$ and choosing $v = u_n(\mathbb{C}) - u_n(\mathbb{C}_t)$ yields

$$\begin{aligned}&\int_\Omega \mathbb{C}\hat{\nabla}\left(u_n(\mathbb{C}) - u_n(\mathbb{C}_t)\right) : \hat{\nabla}\left(u_n(\mathbb{C}) - u_n(\mathbb{C}_t)\right) dx \\ &= \int_\Omega \left(\mathbb{C} - \mathbb{C}_t\right)\hat{\nabla}u_n(\mathbb{C}_t) : \hat{\nabla}\left(u_n(\mathbb{C}) - u_n(\mathbb{C}_t)\right) dx.\end{aligned}$$

Furthermore due to the boundedness of $u_n(\mathbb{C}_t)$, we obtain

$$\|u_n(\mathbb{C}_t) - u_n(\mathbb{C})\|_{H^1(\Omega,\mathbb{R}^2)} \leq C d_\Omega(\mathbb{C}_t, \mathbb{C}),$$

where

$$d_\Omega(\mathbb{C}_t, \mathbb{C}) := \max\left\{\|\lambda_2^t - \lambda_2\|_\infty, \|\mu_2^t - \mu_2\|_\infty\right\}.$$

Due to the strong continuity of $\mathbb{C}_t$ as a function of t, one deduces that $u_n(\mathbb{C}_t) \to u_n(\mathbb{C})$ in $H^1(\Omega;\mathbb{R}^2)$ as $t \to 0$. Concerning the continuity of $u_d(\mathbb{C}_t)$, one may show from (19) that $u_d(\mathbb{C}_t) \to u_d(\mathbb{C})$ in $H^1(\Omega;\mathbb{R}^2)$. Finally in view of the strong continuity of

$$(t,\varphi) \to \partial_t \tilde{G}_n(t,\varphi,\psi), \quad (t,\psi) \to \partial_t \tilde{G}_n(t,\varphi,\psi),$$

$$(t,\varphi) \to \partial_t \tilde{G}_d(t,\varphi,\psi), \quad (t,\psi) \to \partial_t \tilde{G}_d(t,\varphi,\psi),$$

assumptions (H_3) and (H_4) are verified. □

3.2 Second-order derivatives of the cost functional

To compute the second order derivative of $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2,\mu_2,\Gamma)$, we use a Lagrangian method as in the previous section. We start by introducing the Lagrangian functionals

$$\begin{aligned}L_n(\lambda_2,\mu_2,\varphi,\psi) = &\int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}\varphi : \hat{\nabla}\varphi\, dx + \int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}\varphi : \hat{\nabla}\psi\, dx \\ &- \int_{\Gamma_N} g\cdot\psi\, ds \quad \text{for all } \varphi,\psi \in \mathcal{V},\end{aligned} \tag{20}$$

$$\begin{aligned}L_d(\lambda_2,\mu_2,\varphi,\psi) = &\int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}\varphi : \hat{\nabla}\varphi\, dx + \int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}\varphi : \hat{\nabla}\psi\, dx \\ &+ \int_{\Gamma_N} (f-\varphi)\cdot(\mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}\psi)\nu\, ds, \quad \text{for all } \varphi \in \mathcal{V}, \psi \in H_0^1(\Omega)^2.\end{aligned} \tag{21}$$

The saddle point equations of L_n and L_d are given by

$$\begin{aligned}&\partial_\psi L_n(\lambda_2,\mu_2,u_n,v_n;\hat{\psi}) = 0,\\ &\partial_\varphi L_n(\lambda_2,\mu_2,u_n,v_n;\hat{\varphi}) = 0,\end{aligned}$$

$$\begin{aligned}&\partial_\psi L_d(\lambda_2,\mu_2,u_d,v_d;\hat{\psi}) = 0,\\ &\partial_\varphi L_d(\lambda_2,\mu_2,u_d,v_d;\hat{\varphi}) = 0,\end{aligned}$$

or equivalently

$$\int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}u_n : \hat{\nabla}\hat{\psi}\,dx - \int_{\Gamma_N} g\cdot\hat{\psi}\,ds = 0, \tag{22}$$

$$\int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}v_n : \hat{\nabla}\hat{\varphi}\,dx = -2\int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}u_n : \hat{\nabla}\varphi\,dx, \tag{23}$$

$$\int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}u_d : \hat{\nabla}\hat{\psi}\,dx + \int_{\Gamma_N} (f-u_d)\cdot(\mathbb{C}(\lambda_2,\mu_2)\nabla\hat{\psi})\nu\,ds = 0, \tag{24}$$

$$\int_\Omega \mathbb{C}(\lambda_2,\mu_2)\hat{\nabla}v_d : \hat{\nabla}\hat{\varphi}\,dx = -2\int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}u_d : \hat{\nabla}\varphi\,dx. \tag{25}$$

We have the following two result

Theorem 5: second-order derivative
The function $D_{(\lambda_2,\mu_2)}J(\lambda_2,\mu_2,\Gamma)$ is differentiable with respect to (λ_2,μ_2) and

$$\begin{aligned} D^2_{(\lambda_2,\mu_2)}J\,(\lambda_2,\mu_2,\Gamma)\,(\tilde{\lambda}_2,\tilde{\mu}_2)(\hat{\lambda}_2,\hat{\mu}_2) = &\int_{\Omega_2} \mathbb{C}(\hat{\lambda}_2,\hat{\mu}_2)\hat{\nabla}u_n : \hat{\nabla}v_n\,dx \\ &+ \int_{\Omega_2} \mathbb{C}(\hat{\lambda}_2,\hat{\mu}_2)\hat{\nabla}u_d : \hat{\nabla}v_d\,dx, \end{aligned} \tag{26}$$

where u_n, v_n, u_d and v_d are solutions of (22), (23), (24) and (25) respectively.

Proof. As in the proof of Theorem 4, using the Lagrangian L_n and L_d for fixed Γ, and applying the results in the Appendix, we have

$$D^2_{(\lambda_2,\mu_2)}J\,(\lambda_2,\mu_2,u_n,u_d)\,(\tilde{\lambda}_2,\tilde{\mu}_2)(\hat{\lambda}_2,\hat{\mu}_2) = \partial_t\tilde{L}_n(t,u_n,v_n)\Big|_{t=0} + \partial_t\tilde{L}_d(t,u_d,v_d)\Big|_{t=0},$$

where

$$\tilde{L}_n(t,\varphi,\psi) := G_n(\lambda_2^t,\mu_2^t,\varphi,\psi) = \int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}\varphi : \hat{\nabla}\varphi\,dx + \int_\Omega \mathbb{C}_t\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx - \int_{\Gamma_N} g\cdot\psi\,ds,$$

$$\begin{aligned} \tilde{L}_d(t,\varphi,\psi) := G_d(\lambda_2^t,\mu_2^t,\varphi,\psi) = &\int_{\Omega_2} \mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\hat{\nabla}\varphi : \hat{\nabla}\varphi\,dx + \int_\Omega \mathbb{C}_t\hat{\nabla}\varphi : \hat{\nabla}\psi\,dx \\ &+ \int_{\Gamma_N} (f-\varphi)\cdot(\mathbb{C}\hat{\nabla}\psi)\nu\,ds, \end{aligned}$$

and

$$\partial_t\tilde{L}_n(t,u_n,v_n)\Big|_{t=0} = \int_{\Omega_2} \mathbb{C}(\hat{\lambda}_2,\hat{\mu}_2)\hat{\nabla}u_n : \hat{\nabla}v_n\,dx,$$

$$\partial_t\tilde{L}_d(t,u_d,v_d)\Big|_{t=0} = \int_{\Omega_2} \mathbb{C}(\hat{\lambda}_2,\hat{\mu}_2)\hat{\nabla}u_d : \hat{\nabla}v_d\,dx.$$

The hypotheses of Theorem 5.1 in [1] can be checked in a similar way as in the proof of Theorem 4. □

3.3 Shape derivative of the first-order derivatives of the cost functional

In the following theorem, we give the volume expression of the shape derivative of the first derivative of of the cost functional $\mathcal{J}$.

Theorem 6:
The functional $D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)(\tilde{\lambda}_2,\tilde{\mu}_2)$ is shape shape differentibale with respect to Γ and its shape derivative in the direction V is given by:

$$
\begin{aligned}
&\partial_\Gamma\left(D_{(\lambda_2,\mu_2)}\mathcal{J}(\lambda_2^*,\mu_2^*,\Gamma)(\tilde{\lambda}_2,\tilde{\mu}_2)\right)(V)\\
&=\frac{1}{2}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[-\nabla u_n DV-DV^T\nabla u_n^T\right]:\left[\nabla u_n+\nabla u_n^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_n+\nabla u_n^T\right]:\left[\nabla u_n+\nabla u_n^T\right]\operatorname{div}V\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[-\nabla u_n DV-DV^T\nabla u_n^T\right]:\left[\nabla v_n+\nabla v_n^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_n+\nabla u_n^T\right]:\left[-\nabla v_n DV-DV^T\nabla v_n^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_n+\nabla u_n^T\right]:\left[\nabla v_n+\nabla v_n^T\right]\operatorname{div}V\,dx\\
&+\frac{1}{2}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[-\nabla u_d DV-DV^T\nabla u_d^T\right]:\left[\nabla u_d+\nabla u_d^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_d+\nabla u_d^T\right]:\left[\nabla u_d+\nabla u_d^T\right]\operatorname{div}V\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[-\nabla u_d DV-DV^T\nabla u_d^T\right]:\left[\nabla v_d+\nabla v_d^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_d+\nabla u_d^T\right]:\left[-\nabla v_d DV-DV^T\nabla v_d^T\right]\,dx\\
&+\frac{1}{4}\int_{\Omega_2}\mathbb{C}(\tilde{\lambda}_2,\tilde{\mu}_2)\left[\nabla u_d+\nabla u_d^T\right]:\left[\nabla v_d+\nabla v_d^T\right]\operatorname{div}V\,dx.
\end{aligned}
$$

Proof. Theorem 6 can be proven using Theorem 5.1 of R. Correa and A. Seeger [1]. □

IV CONSLUSION

In this paper we have considered the inverse problem of recovering piecewise constants Lamé paremeters. To solve the inverse problem, we minimized a Khon-Vogelius type functional. When the jump of parameters is perturbed, we derived a quantitative stability estimate using shape calculus tools and the implicit function theorem.

REFERENCES

[1] R. Correa and A. Seeger. "Directional derivative of a minimax function". In: *Nonlinear Anal.* 9.1 (1985), pages 13–22. ISSN: 0362-546X.

[2] J. Baumeister. *Stable solution of inverse problems*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, 1987, pages viii+254. ISBN: 3-528-08961-X.

[3] M. C. Delfour and J.-P. Zolésio. "Shape sensitivity analysis via min max differentiability". In: *SIAM J. Control Optim.* 26.4 (1988), pages 834–862. ISSN: 0363-0129.
[4] J. Sokołowski and J.-P. Zolésio. *Introduction to shape optimization*. Volume 16. Springer Series in Computational Mathematics. Shape sensitivity analysis. Berlin: Springer-Verlag, 1992, pages ii+250. ISBN: 3-540-54177-2.
[5] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1999.
[6] K. Malanowski and F. Tröltzsch. "Lipschitz stability of solutions to parametric optimal control problems for parabolic equations". In: *Z. Anal. Anwendungen* 18.2 (1999), pages 469–489. ISSN: 0232-2064.
[7] K. Malanowski and F. Tröltzsch. "Lipschitz stability of solutions to parametric optimal control for elliptic equations". In: *Control Cybernet.* 29.1 (2000), pages 237–256. ISSN: 0324-8569.
[8] K. Malanowski. "Sensitivity analysis for parametric optimal control of semilinear parabolic equations". In: *J. Convex Anal.* 9.2 (2002). Special issue on optimization (Montpellier, 2000), pages 543–561. ISSN: 0944-6532.
[9] R. Griesse. "Parametric sensitivity analysis in optimal control of a reaction diffusion system. I. Solution differentiability". In: *Numer. Funct. Anal. Optim.* 25.1-2 (2004), pages 93–117. ISSN: 0163-0563.
[10] R. Griesse. "Parametric sensitivity analysis in optimal control of a reaction-diffusion system. II. Practical methods and examples". In: *Optim. Methods Softw.* 19.2 (2004), pages 217–242. ISSN: 1055-6788.
[11] K. Brandes and R. Griesse. "Quantitative stability analysis of optimal solutions in PDE-constrained optimization". In: *J. Comput. Appl. Math.* 206.2 (2007), pages 908–926. ISSN: 0377-0427.
[12] J. Hegemann, A. Cantarero, C. L. Richardson, and J. M. Teran. "An explicit update scheme for inverse parameter and interface estimation of piecewise constant coefficients in linear elliptic PDEs". In: *SIAM J. Sci. Comput.* 35.2 (2013), A1098–A1119. ISSN: 1064-8275.
[13] H. Kasumba, K. Kunisch, and A. Laurain. "A bilevel shape optimization problem for the exterior Bernoulli free boundary value problem". In: *Interfaces Free Bound.* 16.4 (2014), pages 459–487. ISSN: 1463-9963.
[14] S. Eberle, B. Harrach, H. Meftahi, and T. Rezgui. "Lipschitz stability estimate and reconstruction of Lamé parameters in linear elasticity". In: *Inverse Problems in Science and Engineering* (2020), pages 1–22.

ARIMA

épisciences

The one step fixed-lag particle smoother as a strategy to improve the prediction step of particle filtering

Samuel Nyobe[*1,2], **Fabien Campillo**[3], **Serge Moto**[1,2], **Vivien Rossi**[4,5]

[1]MIBA research Unity, Faculty of Science, University of Yaoundé, Cameroon
[2]UMI 209, UMMISCO-Cameroon, IRD, Sorbonne University, France
[3]Inria, MathNeuro Team, Montpellier, France
[4]RU Forets and Societies, CIRAD, Yaoundé, Cameroon
[5]National Advanced School of Engineering, University of Yaoundé, Cameroon

*E-mail : samuel.nyobe@facsciences-uy1.cm

DOI : 10.46298/arima.10784
Submitted on 6 janvier 2023 - Published on 30 novembre 2023
Volume : **39** - Year : **2023**

Editors : *Bruce Watson, Mathieu Roche, Nabil Gmati, Clémentin Tayou Djamegni*

Abstract

Sequential Monte Carlo methods have been a major advance in the field of numerical filtering for stochastic dynamical state-space systems with partial and noisy observations. However, these methods still have some weaknesses. One of its main weaknesses concerns the degeneracy of these particle filters due to the impoverishment of the particles. Indeed, during the prediction step of these filters, the particles explore the state space, and if this exploration phase is not done correctly, a large part of the particles will end up in areas that are weakly weighted by the new measurement and will be mostly eliminated. Only a few particles will remain, leading to a degeneracy of the filter. In order to improve this last step within the framework of the classic bootstrap particle filter, we propose a simple approximation of the one step fixed-lag smoother. At each time iteration, we propose to perform additional simulations during the prediction step in order to improve the likelihood of the selected particles. Note that we aim to propose an algorithm that is almost as fast and of the same order of complexity as the bootstrap particle filter, and which is robust in poorly conditioned filtering situations. We also investigate a robust version of this smoother.

Keywords

particle filter, robust particle filter, regime switching particle filter, one step fixed-lag particle smoother, extended Kalman filter, unscented Kalman filter.

I INTRODUCTION

Since the 1980s sequential Monte Carlo (SMC) methods, also called particle filter (PF) methods [2, 4, 6, 7, 11], have met with vast success and broad usage in the context of nonlinear filtering for state-space models with partial and noisy observations, also known as hidden Markov models (HMM) for state-space models. The success of these methods is due to their ability to take into account, in a numerically realistic and efficient way, the non-linearity of the dynamics and the non-Gaussianity of the underlying conditional distributions.

The exact (i.e., non approximated) dynamics of these HMMs takes the form of a sequential Bayes formula represented in Eqs. (3)-(4) also called Bayesian filter. In the case of linear models with additive Gaussian noise, the Kalman filter gives an exact solution of this problem as well as an efficient algorithm to compute it. There are a few cases where this filter can be computed explicitly in a finite-dimensional way, but in the majority of cases it is necessary to use numerical approximation methods. Historically, the first method proposed is the extended Kalman filter (EKF) [1] and its variants, which consist in linearizing the model around the current estimate and applying the Kalman filter. But in the case of strongly nonlinear models, the EKF often diverges. Since the 1980s, Monte Carlo methods have become very effective alternatives, they are now recognized as a powerful tool in the estimation with Bayesian filters in nonlinear/non-Gaussian hidden Markov models.

Among Monte Carlo methods, the PF techniques rely on an online importance sampling approximation of the sequential Bayes formula (3)-(4). At each time iteration, the PF builds a set of particles i.e. an independently and identically distributed sample from an approximation of the theoretical solution of the Bayesian filter.

A PF time iteration classically includes two steps: a prediction step consisting in exploring the state space by moving the particles according to the state equation, followed by a correction step consisting in first weighting the particles according to their correspondence with the new observation, i.e. according to their likelihood, then resampling these particles according to their weight. These two steps can be understood respectively as the mutation and selection steps of genetic algorithms.

The first really efficient PF algorithm, namely the bootstrap particle filter (BPF) or sampling-importance-resampling filter proposed in 1993 [11], follows exactly these two steps.

The resampling step is essential, without it we notice after a few time iterations an impoverishment of the particles: very few particles, even one, will concentrate all the likelihood; the other particles then become useless, the filter then loses track of the current state.

This resampling step is therefore essential, but still it does not prevent another type of particle degeneracy [2, 7, 17, 21, 27]. The latter appears when the particles explore areas of the state space that are not in agreement with the new observation. This is due to the fact that in its classical version, especially in the case of BPF, the prediction step propagates the particles in the state space without taking into account the next observation. Thus, when we associate weights with particles using their value via the likelihood function, a large proportion of the particles may have almost zero weight, this phenomenon is aggravated by the machine epsilon. In the worst case, all the particles can have a zero weight, the filter has then completely lost track of the state evolution.

To overcome this problem it is relevant to take into account the observations in the prediction step consisting in exploring the state space. The auxiliary particle filter (APF) [22, 23] is one

of the first attempts to take into account the observation in the prediction step. One can consult the review articles [9, 10] concerning the possibilities of improvement of particle filters.

Another possibility to overcome this difficulty, without using the observation following the current one, is to use robust versions of particle filters, such as the regime switching method [15] or the averaging-model approach [19, 25], which allow model uncertainties to be taken into account.

The present paper aims to propose a new PF algorithm, called predictive bootstrap particle smoother (PBPS), to further improve the prediction step but keeping the simplicity of the BPF. The principle of the PBPS is to take into account the current and the next observations for the prediction and correction steps. At each prediction step, we make additional explorations in order to better determine the likelihood according to the next observation. It can be seen as a simple approximation of the one step fixed-lag smoother. The additional explorations contribute to the correction step in order to improve the likelihood of the selected particles with the current and next observation.

We propose an algorithm that is only a little more complex than the BPF, which is why we do not consider more complex algorithms, like the smoother with a deeper time-lag. We also want to propose an efficient algorithm in case of unfavorable filtering situations such as when observability is poorly conditioned.

In Section II state-space models and the (exact) formulas for nonlinear filters and smoothers are presented. In Section III, we first present the classical bootstrap particle filter (BPF), then the predictive bootstrap particle smoother (PBPS) and its "robustfied" version called regime switching predictive bootstrap particle smoother (RS-PBPS).

In Section IV we present simulations in two test cases where observability is weakly conditioned. The first test case is a classical example in state space dimension 1 for which we compare the PBPS to the extended Kalman filter (EKF) [1], the unscented Kalman filter (UKF) [13, 26], the bootstrap particle filter (BPF) [11], and the the auxiliary particle filter (APF) [22]. The second is more realistic, it is a classical problem of bearing-only 2D tracking with a state space of dimension 4. In this case we first compare the PBPS to the BPF and the APF filters. In this example there are often model uncertainties, so it is relevant to use robust filters, and we compare the RS-PBPS to "robustified" versions of the particle filters.

II NONLINEAR FILTERING AND SMOOTHING

In order to simplify the presentation, we will make the abuse of notation, customary in this field, of representing probability distributions, e.g. $\mu(\mathrm{d}x)$, by densities, e.g. $\mu(x)\,\mathrm{d}x$; we also represent the Dirac measure $\delta_{x_0}(\mathrm{d}x)$ as $\delta_{x_0}(x)\,\mathrm{d}x$ where $\delta_{x_0}(x)$ is the Dirac function (with $\delta_{x_0}(x) = 0$ if $x \neq x_0$, $\delta_{x_0}(x) = +\infty$ if $x = x_0$, and $\int \delta_{x_0}(x)\,\mathrm{d}x = 1$).

2.1 The state space model

We consider a Markovian state-space model with state process $(X_k)_{k\geq 0}$ taking values in $\mathbb{R}^n$ and observation process $(Y_k)_{k\geq 1}$ taking values in $\mathbb{R}^d$. We suppose that conditionally on $(X_k)_{k\geq 0}$, the observations Y_k are independent.

The ingredients of the state-space model are:

$$q_k(x|x') \stackrel{\text{def}}{=} p_{X_k|X_{k-1}=x'}(x)\,, \qquad \text{(state transition kernel)}$$
$$\psi_{k,y}(x) \stackrel{\text{def}}{=} p_{Y_k|X_k=x}(y)\,, \qquad \text{(local likelihood function)}$$
$$\rho_0(x) \stackrel{\text{def}}{=} p_{X_0}(x)\,, \qquad \text{(initial distribution)}$$

for any $x, x' \in \mathbb{R}^n$, $y \in \mathbb{R}^d$.

These ingredients can be made more explicit in the case of the following state space model:

$$X_k = f_{k-1}(X_{k-1}) + g_{k-1}(X_{k-1})\, W_{k-1}\,, \tag{1}$$
$$Y_k = h_k(X_k) + V_k\,, \tag{2}$$

for $k \geq 1$, where X_k (resp. Y_k, W_k, V_k) takes values in $\mathbb{R}^n$ (resp. $\mathbb{R}^d$, $\mathbb{R}^m$, $\mathbb{R}^d$), f_k (resp. g_k, h_k) is a differentiable and at most linear growth, uniformly in k, from $\mathbb{R}^n$ to $\mathbb{R}^n$ (resp. $\mathbb{R}^d$, $\mathbb{R}^{n\times m}$, $\mathbb{R}^d$), g_k being also bounded. Random sequences W_k and V_k are independent centered white Gaussian noises with respective variances R_k^W and R_k^V; W_k, V_k, X_0 being independent. In this case:

$$q_k(x|x') = \frac{1}{\sqrt{(2\,\pi)^n\,\det R_{k-1}^W}} \times \\ \times \exp\Big(-\frac{1}{2}\big[x - f_{k-1}(x')\big]^*\,\big[g_k(x')\,R_{k-1}^W\,g_k(x')^*\big]^{-1}\,\big[x - f_{k-1}(x')\big]\Big)$$

and

$$\psi_{k,y}(x) \propto \exp\Big(-\frac{1}{2}\big[y - h_k(x)\big]^*\,[R_k^V]^{-1}\,\big[y - h_k(x)\big]\Big)$$

($\psi_{k,y}$ has to be known up to a multiplicative constant).

2.2 Nonlinear filtering

Nonlinear filtering aims at determining the conditional distribution η_k of the current state X_k given the past observations $Y_1, \ldots, Y_k$, namely:

$$\eta_k(x) \stackrel{\text{def}}{=} p_{X_k|Y_{1:k}=y_{1:k}}(x)\,,\ x \in \mathbb{R}^n$$

for any $k \geq 1$ and $y_{1:k} \in (\mathbb{R}^d)^k$, here we use the notation:

"$Y_{1:k}$" for $(Y_1, \ldots, Y_k)$

(e.g. $Y_{1:k} = y_{1:k}$ means $Y_\ell = y_\ell$ for all $\ell = 1, \ldots, k$).

The nonlinear filter allows us to determine η_k from η_{k-1} using the classical two-step recursive Bayes formula:

Prediction step. The predicted distribution $\eta_{k^-}(x) \stackrel{\text{def}}{=} p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x)$ of X_k given $Y_{1:k-1} = y_{1:k-1}$ is given by:

$$\eta_{k^-}(x) = \int_{\mathbb{R}^n} q_k(x|x')\,\eta_{k-1}(x')\,\mathrm{d}x'\,, \tag{3}$$

for $x \in \mathbb{R}^n$.

Correction step. The new observation $Y_k = y_k$ allows the predicted distribution to be updated in order to obtain η_k according to the Bayes formula:

$$\eta_k(x) = \frac{\psi_k(x)\,\eta_{k^-}(x)}{\int_{\mathbb{R}^n} \psi_k(x')\,\eta_{k^-}(x')\,\mathrm{d}x'}\,, \tag{4}$$

for $x \in \mathbb{R}^n$; here and in order to simplify the notations, $\psi_k(x)$ represents $\psi_{k,y_k}(x)$.

Note that in the correction step (4) , η_k is proportional to the product of η_{k-1} and the local likelihood function, that is:

$$\eta_k(x) \propto \psi_k(x)\,\eta_{k^-}(x)\,.$$

2.3 Nonlinear smoothing

Define $\bar{\eta}_k(x)$ the conditional distribution of X_k given $Y_{1:k+1} = y_{1:k+1}$, that is:

$$\bar{\eta}_k(x) \stackrel{\text{def}}{=} p_{X_k|Y_{1:k+1}=y_{1:k+1}}(x)\,,\ x \in \mathbb{R}^n$$

for any $k \geq 1$ and $y_{1:k+1} \in (\mathbb{R}^d)^{k+1}$.

To get the recurrence equation for $\bar{\eta}_k(x)$, we consider the extended state vector $\mathbf{X}_k = (X_k, X_{k-1})$, it's a Markov with transition:

$$\begin{aligned}
&Q_k(x', x''|x_{k-1}, x_{k-2}) = \\
&\quad = p_{X_k,X_{k-1}|X_{k-1}=x_{k-1},X_{k-2}=x_{k-2}}(x', x'') \\
&\quad = p_{X_k|X_{k-1}=x'',X_{k-1}=x_{k-1},X_{k-2}=x_{k-2}}(x')\; p_{X_{k-1}|X_{k-1}=x_{k-1},X_{k-2}=x_{k-2}}(x'') \\
&\quad = p_{X_k|X_{k-1}=x_{k-1},X_{k-2}=x_{k-2}}(x')\; \delta_{x_{k-1}}(x'') \\
&\quad = p_{X_k|X_{k-1}=x_{k-1}}(x')\; \delta_{x_{k-1}}(x'') \\
&\quad = q_k(x'|x_{k-1})\; \delta_{x_{k-1}}(x'')\,,
\end{aligned}$$

where $x'' \to \delta_{x_{k-1}}(x'')$ is the Dirac delta function in x_{k-1} (here, we adopt the usual abuse of notation, which consists in representing the Dirac measure by the Dirac delta function: $\delta_x(x') = 0$ if $x' \neq x$, $\delta_x(x') = +\infty$ if $x' = x$, and $\int \delta_x(x')\,\mathrm{d}x' = 1$.)

The conditional distribution $\boldsymbol{\eta}_k(x', x'')$ of $\mathbf{X}_k = (X_k, X_{k-1})$ given $Y_{1:k} = y_{1:k}$, we apply the previous filter formula:

$$\begin{aligned}
\boldsymbol{\eta}_{k^-}(x', x'') &= \iint Q_k(x', x''|x_{k-1}, x_{k-2})\,\boldsymbol{\eta}_{k-1}(x_{k-1}, x_{k-2})\,\mathrm{d}x_{k-1}\,\mathrm{d}x_{k-2} \\
&= \iint q_k(x'|x_{k-1})\,\delta_{x_{k-1}}(x'')\,\boldsymbol{\eta}_{k-1}(x_{k-1}, x_{k-2})\,\mathrm{d}x_{k-1}\,\mathrm{d}x_{k-2} \\
&= \iint q_k(x'|x_{k-1})\,\delta_{x_{k-1}}(x'')\,\boldsymbol{\eta}_{k-1}(x'', x_{k-2})\,\mathrm{d}x_{k-1}\,\mathrm{d}x_{k-2}\,,
\end{aligned}$$

and the distribution of Y_k given $(X_k = x_k, X_{k-1} = x_{k-1})$ is the distribution of Y_k given $X_k = x_k$, hence:

$$\boldsymbol{\eta}_k(x', x'') \propto \psi_k(x')\,\boldsymbol{\eta}_{k^-}(x', x'')\,, \tag{5}$$

and the x''-marginal distribution of this last expression gives $\bar{\eta}_{k-1}(x'')$, the conditional distribution of X_{k-1} given $Y_{1:k} = y_{1:k}$.

2.4 Approximations

The main difficulty encountered by the nonlinear filter (3)-(4) lies in the two integrations. These integrations can be solved explicitly only in the linear/Gaussian case, leading to the Kalman filter, and in a very few other specific nonlinear/non-Gaussian cases. In the latter cases, the optimal filter can be solved *explicitly* in the form of a finite dimensional filter; hence in the vast majority of cases, it is necessary to use approximation techniques [4]. Among the approximation techniques, we will consider the extended Kalman filter (EKF) [1], the unscented Kalman filter (UKF) [13, 26] and particle filter techniques, also called sequential Monte Carlo techniques [6, 11], see [8] for a recent overview.

Concerning the particle filters, as we will see in the next section, it is important to notice that on the one hand we do not need to know the analytical expressions of the state transition kernel and of the initial distribution, we just need to be able to sample (efficiently) from them; on the other hand, we do need the analytical expression of the local likelihood function (up to a multiplicative constant), indeed, for a given y, we need to compute $\psi_{k,y}(x)$ for a very large number of x values.

III PARTICLE APPROXIMATIONS

Particle filtering, also known as sequential Monte Carlo (SMC) methods, is a set of computational techniques used for the approximation of nonlinear filters [6]. According to this approach, the nonlinear filter η_k is approximated by η_k^N of the form:

$$\eta_k^N(x) = \sum_{i=1}^{N} \omega_k^i \, \delta_{\xi_k^i}(x) \, , \; x \in \mathbb{R}^n \, .$$

composed of N particles located in ξ_k^i in $\mathbb{R}^n$ and associated weights ω_k^i, the weights are positive and sum to one [6]. Ideally the particles are sampled from η_k and the importance weight are all equal to $1/N$ (meaning that all the particles have the same "importance").

Particle filters offer two essential advantages in numerical approximation:

- they are finite-dimensional filters in the sense that they are represented using a finite number of parameters, in this case the position of the particles (after selection), which can therefore be integrated into computer memory;
- the representation of particle filters, as weighted sums of Dirac measurements, greatly facilitates the calculation of integrals according to these filters; moreover, many operations can be parallelized as a function of the particle index (or vectorized with interpreted languages like `Python`).

In the 3 particle filters presented, at each time increment between two observations, in a first so-called "mutation" step, the particles explore the state space. A weight is then associated with each of these particles according to a fitness function (here a local likelihood function), these weights will be noted $\omega_k^{1:N}$ in the 3 cases, then in a second "selection" step, particules are selected based on the fact that particles with a high weight will have a greater probability of reproduction than particles with low weights (at constant number of particles N), leading to a family of particles noted $\xi_k^{1:N}$ in the 3 cases.

3.1 The bootstrap particle filter

The bootstrap particle filter (BPF) filter is a classical sequential importance resampling method: suppose we have a good approximation $\eta_{k-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{k-1}^i}$ of η_{k-1}. We can apply the prediction step (3) to η_{k-1}^N and get the following approximation of η_{k^-}:

$$\tilde{\eta}_{k^-}(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} q_k(x|x')\, \eta_{k-1}^N(x')\, \mathrm{d}x' = \sum_{i=1}^N \omega_{k-1}^i\, q_k(x|\xi_{k-1}^i)\,,$$

and then apply the correction step (4) and get the following approximation of η_k:

$$\tilde{\eta}_k(x) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^N \omega_{k-1}^i\, \psi_k(x)\, q_k(x|\xi_{k-1}^i)}{\int_{\mathbb{R}^n} \sum_{i=1}^N \omega_{k-1}^i\, \psi_k(x')\, q_k(x'|\xi_{k-1}^i)\, \mathrm{d}x'}\,,$$

but the two last approximations $\tilde{\eta}_{k^-}$ and $\tilde{\eta}_k$ are mixtures of the continuous densities $q_k(\,\cdot\,|\xi_{k-1}^i)$ and therefore not of the particle type. The bootstrap particle filter (BPF) proposed by [11] is the simplest method to propose a particle approximation. For the prediction step we use the sampling technique:

$$\xi_{k^-}^i \sim q_k(\,\cdot\,|\xi_{k-1}^i)\,,\ i = 1:N \quad \text{(independently)}\,,$$

which is:

$$\xi_{k^-}^i = f_{k-1}(\xi_{k-1}^i) + g_{k-1}(\xi_{k-1}^i)\,\mathrm{w}^i\,,$$

where w^i are i.i.d. $N(0, R_{k-1}^W)$ samples. Then we let

$$\eta_{k^-}^N(x) \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_{k-1}^i\, \delta_{\xi_{k^-}^i}(x)\,.$$

Through the correction step (4), this approximation $\eta_{k^-}^N$ gives $\sum_{i=1}^N \omega_k^i\, \delta_{\xi_{k^-}^i}$ where:

$$\omega_k^i \stackrel{\text{def}}{=} \frac{\psi_k(\xi_{k^-}^i)\, \omega_{k^-}^i}{\sum_{j=1}^N \psi_k(\xi_{k^-}^j)\, \omega_{k^-}^j} \tag{6}$$

are the updated weights taking account of the new observation y_k through the likelihood function ψ_{k,y_k}. This correction step *must* be completed by a resampling of the particles $\xi_{k^-}^i$ according to the importance weights ω_k^i:

$$\xi_k^i \sim \sum_{j=1}^N \omega_k^j\, \delta_{\xi_{k^-}^j}\,,\ i = 1:N \quad \text{(independently)}\,, \tag{7}$$

leading the to the particle approximation:

$$\eta_k^N(x) \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_k^i\, \delta_{\xi_k^i}(x)\,,\ \text{with } \omega_k^i = \frac{1}{N}\,.$$

Algorithm III.1 gives a summary of the BPF, in this version a resampling is performed at each time interval. The multinomial resampling step (7), whose basic idea is proposed in [11], could be greatly improved, there are several resampling techniques, see [5, 16] for more details.

Algorithm III.1 Bootstrap particle filter (BPF).

1: $\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$ `# initialization`
2: **return** $\xi_0^{1:N}$
3: **for** $k = 1, 2, \ldots$ **do**
4: $\quad \xi_{k^-}^i \sim q_k(\,\cdot\,|\xi_{k-1}^i)\,,\ i = 1:N$ `# particles evolution`
5: $\quad \omega_k^i \leftarrow \psi_k(\xi_{k^-}^i)\,,\ i = 1:N$ `# likelihood`
6: $\quad \omega_k^i \leftarrow \omega_k^i / \sum_{j=1}^N \omega_k^j\,,\ i = 1:N$ `# renormalization`
7: $\quad \xi_k^{1:N} \overset{\text{iid}}{\sim} \sum_{j=1}^N \omega_k^j\, \delta_{\xi_{k^-}^j}$ `# resampling`
8: $\quad$ **return** $\xi_k^{1:N}$
9: **end for**

It is not necessary to resample the particles systematically at each time iteration as in Algorithm III.1. However, this resampling step should be done regularly in terms of time iterations to avoid degeneracy of the weights [6].

We will consider another degeneracy problem. Suppose that in step (6), the local likelihood values $\psi_k(\xi_{k^-}^i)$ associated with the predicted particles $\xi_{k^-}^i$ are all very small, or even equal to zero due to rounding in floating point arithmetic. This normalization step (6) is then impossible. This problem occurs when the filter loses track of the true state X_k, i.e. the likelihood of the predicted particles $\xi_{k^-}^i$ w.r.t. the observation y_k are all negligible, in this case the observation y_k appears as an outlier. This occurs especially when the period of time between two successive instants of observation is very large compared to the dynamics of the state process. Strategies to overcome this weakness encompass the iterated extended Kalman particle filter [18] or the iterated unscented Kalman particle filter [12]; see [10] and [17] for reviews of the subject.

3.2 The predictive bootstrap particle smoother

The purpose of the predictive bootstrap filter (PBPS) is to improve the correction step (6) of the BPF by using both observations $Y_k = y_k$ and $Y_{k+1} = y_{k+1}$ at each iteration k. In a way, the PBPS can be seen as an approximation of the distribution of X_k given $Y_{1:k+1} = y_{1:k+1}$, the one step fixed-lag smoother presented in Section 2.3.

In the proposed algorithm, the prediction step consists in a first step to propagate at time k the η_{k-1}^N particles by simulating the transition kernel sampler, then in a second step to extend these particles according to a one-step-ahead sampler. Then the correction step consists in updating the weights of the N particles of $\eta_{k^-}^N$ according to their likelihood with y_k and the likelihood of their one-step ahead offspring particles with y_{k+1}.

The iteration $k-1 \to k$ of the filter is more precisely:

Prediction step. As for the BPF we sample N particles and compute N normalized likelihood weights:

$$\tilde{\xi}_k^i \sim q_k(\,\cdot\,|\xi_{k-1}^i)\,,\ \tilde{\omega}_k^i \propto \psi_k(\xi_{k^-}^i)\,\omega_{k^-}^i\,,\ i = 1:N\,.$$

Again we will resample the particles $\tilde{\xi}_k^{1:N}$, but instead of doing so according to the weights $\tilde{\omega}_k^{1:N}$, we will first modify the latter ones. For each index i, we generate a one-step-ahead

Algorithm III.2 predictive bootstrap particle smoother (PBPS).

$\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$ # initialization
for $k = 1, 2, \ldots$ **do**
  for $i = 1 : N$ **do**
    $\tilde{\xi}_k^i \sim q_k(\cdot|\xi_{k-1}^i)$ # particles propagation
    $\tilde{\omega}_k^i \leftarrow \psi_k(\tilde{\xi}_k^i)$
    $\tilde{\xi}_{k+1}^i = f_k(\tilde{\xi}_k^i)$ # offspring
    $\tilde{\omega}_{k+1}^i \leftarrow \psi_{k+1}(\tilde{\xi}_{k+1}^i)$ # offspring weighting
    $\omega_k^i \leftarrow \tilde{\omega}_k^i \, \tilde{\omega}_{k+1}^i$ # particles weighting
  end for
  $\omega_k^i \leftarrow \omega_k^i / \sum_{i'=1}^N \omega_k^{i'}$, $i = 1 : N$ # weights normalization
  $\xi_k^{1:N} \overset{\text{iid}}{\sim} \sum_{i'=1}^N \omega_k^{i'} \, \delta_{\tilde{\xi}_k^{i'}}$ # particles resampling
  return $\xi_k^{1:N}$
end for

offspring particles and compute its weight:

$$\tilde{\xi}_{k+1}^i \sim \tilde{q}_{k+1}(\cdot|\tilde{\xi}_k^i)\,, \qquad\qquad \tilde{\omega}_{k+1}^i \overset{\text{def}}{=} \psi_{k+1}(\tilde{\xi}_{k+1}^i).$$

Note that the likelihood weights $\tilde{\omega}_{k+1}^i$ depend on the next observation y_{k+1}. The choice of the one-step-ahead sampler $\tilde{q}_{k+1}$ will be detailed at the end of this sub-section.

Correction step. We compute the weights at time k according to Eq. (5):

$$\omega_k^i \overset{\text{def}}{=} \tilde{\omega}_k^i \, \tilde{\omega}_{k+1}^i \qquad \text{for } i = 1 : N.$$

Then, N particles $\tilde{\xi}_k^{1:N}$ are resampled according to the weights $\omega_k^{1:N}$.

The PBPS is depicted in Algorithm III.2.

Choice of one-step-ahead sampler $\tilde{q}_{k+1}$. For the special case of the system (1)-(2) we can choose a simpler "deterministic sampler":

$$\tilde{\xi}_{k+1}^i = f_k(\tilde{\xi}_k^i)$$

which corresponds to the one-step-ahead mean:

$$\begin{aligned}\tilde{\xi}_{k+1}^i &= \mathbb{E}\big(X_{k+1}\big|X_k = \tilde{\xi}_k^i\big) \\ &= \mathbb{E}\big(f_k(X_k) + g_k(X_k)\, W_k\big|X_k = \tilde{\xi}_k^i\big) \\ &= f_k(\tilde{\xi}_k^i) + g_k(\tilde{\xi}_k^i) \underbrace{\mathbb{E}(W_k|X_k = \tilde{\xi}_k^i)}_{=0} = f_k(\tilde{\xi}_k^i)\,.\end{aligned}$$

This choice greatly reduces the computation burden and it is sufficient, as we will see, for a linear state equation. For a highly nonlinear state equation, we can choose $\tilde{q}_{k+1} = q_{k+1}$, which corresponds to a simulation of the state dynamic.

Algorithm III.3 Regime Switching predictive bootstrap particle smoother (RS-PBPS)

$\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$ # initialization
return $\xi_0^{1:N}$
for $k = 1, 2, \ldots$ **do**
  $\mu_{k-1}^{1:N} \overset{\text{iid}}{\sim} U(\mathcal{M})$
  $\tilde{\xi}_k^i \sim q_k(\,\cdot\,|\xi_{k-1}^i, \mu_{k-1}^i)\,,\ i = 1:N$ # particles propagation
  $\tilde{\omega}_k^i \leftarrow \psi_k(\tilde{\xi}_k^i)\,,\ i = 1:N$
  $\tilde{\xi}_{k+1}^i = f_k(\tilde{\xi}_k^i)\,,\ i = 1:N$ # offspring
  $\tilde{\omega}_{k+1}^i \leftarrow \psi_{k+1}(\tilde{\xi}_{k+1}^i)\,,\ i = 1:N$ # offspring weighting
  $\omega_k^i \leftarrow \tilde{\omega}_k^i\, \tilde{\omega}_{k+1}^i\,,\ i = 1:N$ # particles weighting
  $\omega_k^i \leftarrow \omega_k^i / \sum_{j=1}^N \omega_k^j\,,\ i = 1:N$ # renormalization
  $\xi_k^{1:N} \overset{\text{iid}}{\sim} \sum_{j=1}^N \omega_k^j\, \delta_{\tilde{\xi}_k^j}$ # resampling
  return $\xi_k^{1:N}$
end for

3.3 The regime switching predictive bootstrap particle smoother

In most applications, both the state model and the observation model are only partially known. As we will see in Section 4.2, applying a filter without care in such situations can be delicate. There are several techniques to make these filters more robust to model mismatch such as the dynamic model averaging (DMA) [19, 25] or the regime switching (RS) [15]. We will use both DMA and RS techniques in the simulations of Section 4.2, but we introduce only the RS technique in the present section by proposing a "robustified" version of PBPS, called regime switching PBPS (RS-PBPS).

The partial knowledge on the state-space model introduced in Section 2.1 can be related to the state model, i.e. to the the transition kernel q_k, or the observation model, i.e. the likelihood function ψ_k.

In the great majority of applications, the uncertainty relates to the state model, so we will restrict ourselves to this case. In order to model this uncertainty, we assume that the state dynamics corresponds to a transition kernel of the form:

$$q_k(x|x', m)\,,$$

where m is a parameter that evolves dynamically in the finite set:

$$\mathcal{M} = \{m_1, \ldots, m_L\}\,.$$

Thus $\mathcal{M}$ represents the different possible regimes of the state model over time.

The principle of these robust filters consists in making the particles and the weights evolve according to a mixture of the parameters $\mathcal{M}$. The dynamic model averaging particle filter (DMA-BPF) proposed in [25] is detailed in Algorithm A.2, we will not describe it here.

We now present one of the algorithms proposed by [15]. We use the simplest of them which gives good results according to [15]. In the BPF, we replace the prediction step (Algorithm III.1 line 4) by:

$$\mu_{k-1}^{1:N} \overset{\text{iid}}{\sim} U(\mathcal{M})\,, \quad \xi_k^i \sim q_k(\,\cdot\,|\xi_{k-1}^i, \mu_{k-1}^i)\,, \quad i = 1:N\,,$$

where $U(\mathcal{M})$ is the uniform distribution on $\mathcal{M}$; the only difference lies in the fact that the particles are sampled at random according to the various models represented by $\mathcal{M}$. The resulting method called RS-BPF is described in Algorithm A.3. The adaptation to PBPS and APF is also immediate, leading to the RS-PBPS and the RS-APF; the RS-PBPS is detailed in Algorithm III.3.

IV SIMULATION STUDIES

We compare numerically the PBPS to other filters on two space-state models where observability is weakly conditioned. The performance of the filters is compared using an empirical evaluation of the root-mean-square error (RMSE). We simulate S independent trajectories $(X_{0:K}^{(s)}, Y_{1:K}^{(s)})_{s=1:S}$ of the state-space models, where K is the length of the chronological sequence of observations. For each simulation s, we ran R times each filter $\mathcal{F} \in \{$BPF, PBPS, APF, DMA-BPF, RS-BPF, RS-APF, RS-PBPS$\}$ and we compute the root mean squared error at time k:

$$\text{RMSE}_k(\mathcal{F}) \overset{\text{def}}{=} \sqrt{\frac{1}{S}\sum_{s=1}^{S}\frac{1}{R}\sum_{r=1}^{R}\left|\hat{X}_k^{\mathcal{F}(s,r)} - X_k^{(s)}\right|^2}\,, \tag{8}$$

where:

$$\hat{X}_k^{\mathcal{F}(s,r)} \overset{\text{def}}{=} \int_{\mathbb{R}^n} x\,\eta_k^{N,\mathcal{F}(s,r)}(x)\,\mathrm{d}x = \frac{1}{N}\sum_{i=1}^{N}\xi_k^{\mathcal{F}(s,r),i}$$

is the numerical approximation of $\hat{X}_k^{(s)} = \mathbb{E}(X_k^{(s)}|Y_{1:k}^{(s)})$ by the filter $\mathcal{F}$. We also compute the global root mean squared error:

$$\text{RMSE}(\mathcal{F}) \overset{\text{def}}{=} \frac{1}{K}\sum_{k=1}^{K}\text{RMSE}_k(\mathcal{F})\,.$$

Note that for the EKF and UKF filters the summation over r in (8) is useless.

All implementations are done in R language [24] using a 2.11 GHz core i7-8650U intel running Windows 10 64bits with a 16 Go RAM.

4.1 First case study: a one-dimensional model

We consider the following one-dimensional nonlinear model [6, 11, 14]:

$$\begin{aligned} X_k &= \frac{X_{k-1}}{2} + \frac{25\,X_{k-1}}{1+X_{k-1}^2} + 8\cos\bigl(1.2\,(k-1)\bigr) + W_{k-1}\,,\\ Y_k &= \frac{X_k^2}{20} + V_k\,, \end{aligned} \tag{9}$$

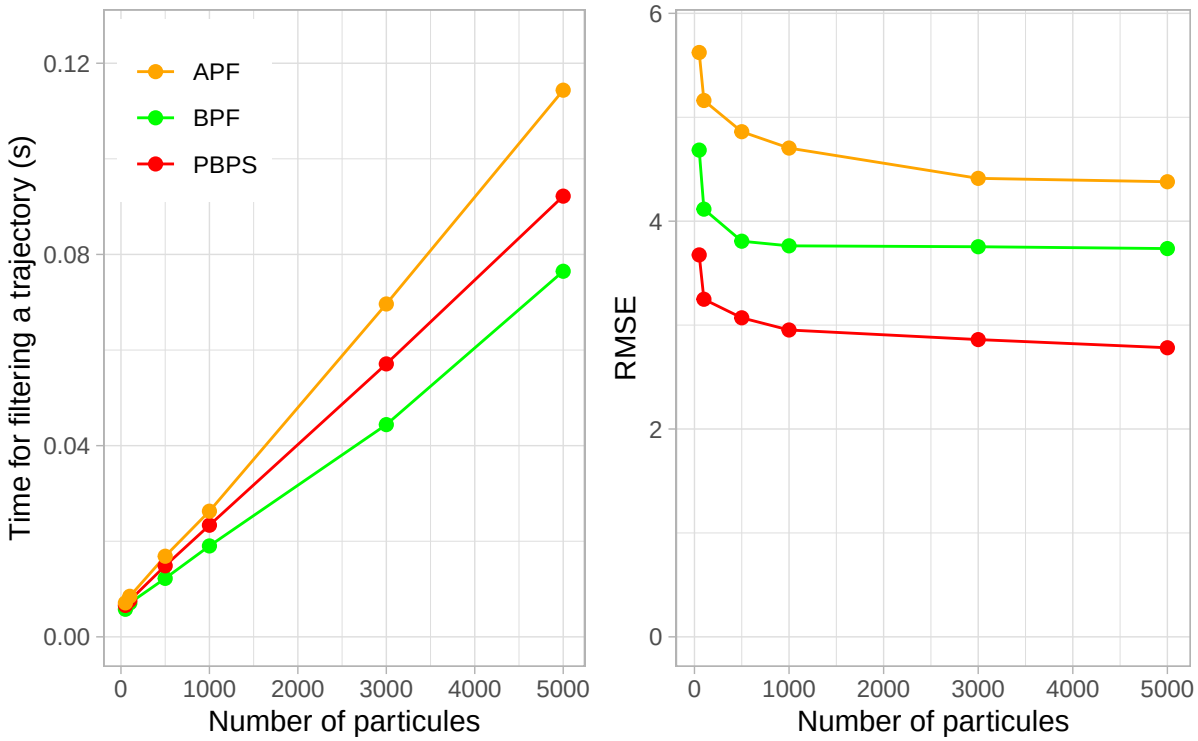

Figure 1: One-dimensional case study (9) — We plot the average computation time for filtering a trajectory (left) and the RMSE (right) of the filters PBPS, BPF and APF as a function of different values of N (50, 100, 500, 1000, 3000, 5000) (with $S = 100$ and $R = 40$). In comparison, the EKF and UKF have a respective RMSE of 16.83 and 6.88 (and a negligible computation time compared to the particle approximations).

with $1 \leq k \leq K$ ($K = 50$), $X_0 \sim \mathcal{N}(0, 1)$; $W_k \overset{\text{iid}}{\sim} \mathcal{N}(0, 3^2)$, $V_k \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$; X_0, $(W_k)_{k\geq 1}$ and $(V_k)_{k\geq 1}$ mutually independent. Note that the state process X_k is observed only through X_k^2 so the filters have difficulties in determining whether X_k is positive or negative, especially since the state process X_k regularly changes it's sign. Thus this model is regularly used as benchmark for testing filters, as filters may easily lose track of X_k.

For this example we compare the filters PBPS, BPF, APF, EKF and UKF.

In Figure 1 we plot, as a function of different values of N (50, 100, 500, 1000, 3000, 5000), on the left the average computation time for the filtering a trajectory, and on the right the RMSE (with $S = 100$, $R = 40$).

The computation times of EKF and UKF are negligible compared to those of the particle approximations. On the other hand, the RMSE of EKF (16.83) are much higher than those of the particle approximations; UKF with an RMSE of 6.88 behaves much better than EKF but is still less accurate than the particle approximations.

The computation time of PBPS is naturally higher than that of BPF, but the latter is much less accurate. For example, PBPS with $N = 50$ particles is 15 times faster than BPF with $N = 5000$ particles while being 30% more accurate.

In Figure 2, for each filter $\mathcal{F} \in \{\text{PBPS, BPF, APF, EKF, UKF}\}$ (with $N = 5000$), we plot the true state trajectory $k \to X_k$, the approximation $k \to \hat{X}_k^{\mathcal{F}}$, and the associated 95% confidence

regions. For the particle approximations the confidence interval is given empirically by the particles, for the EKF and UKF it is given as an approximated Gaussian confidence interval. The PBPS approximation has a smaller error $k \to |\hat{X}_k^{\mathcal{F}} - X_k|$ and a smaller confidence region than the other filters.

In Figure 3, we compare $k \to \text{RMSE}_k(\mathcal{F})$ for $\mathcal{F} \in \{\text{PBPS, BPF, APF, EKF, UKF}\}$ (with $N = 5000$). Again, we see that the PBPS behaves better than the other filters and that the EKF has a very erratic behavior.

The poor behavior of the EKF can be explained by the very nonlinear nature of this example and by the observability problem around 0. Note the relatively good behavior of the UKF. However, because of the observability problem at 0, both the UKF and the EKF cannot find the track of X_k once in the wrong half-plane (Figure 2 around time $k = 30$).

In Figure 4, in the case of a single simulation of the state-space system and of the BPF and PBPS filters (with $N = 5000$), we plot the true trajectory $k \to X_k$ as well as the set of particles $k \to \xi_k^{\mathcal{F},1:N}$ for $\mathcal{F} \in \{\text{BPF, PBPS}\}$. Not surprisingly, we find that the PBPS particles are more concentrated around the true value X_k and that BPF more often "loads" the symmetric part of the state space. Indeed, the smoothing allows us to reduce the variance but also to compensate the observability ambiguities; it is for this purpose that it was designed, and we reiterate that this gain is obtained with a great improvement in the computation time (provided that less particles are used).

4.2 Second case study: a four-dimensional bearings-only tracking model

We consider the following four-dimensional model [3, 11, 23]:

$$\begin{aligned} X_k &= \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} X_{k-1} + \sigma_W \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} W_{k-1}\,, \\ Y_k &\sim \text{wrapped Cauchy}\Big(\arctan\Big(\frac{X_k[1]}{X_k[2]}\Big), \rho\Big), \end{aligned} \tag{10}$$

with $1 \leq k \leq K = 20$, $X_0 \sim \mathcal{N}(\bar{X}_0, P_0)$ with:

$$\begin{aligned} \bar{X}_0 &= \big(-0.05, 0.2, 0.001, -0.055\big)^*\,, \\ P_0 &= 0.01\,\text{diag}\big(0.5^2, 0.3^2, 0.005^2, 0.01^2\big)\,, \end{aligned}$$

$\sigma_W = 0.001$, $W_k \overset{\text{iid}}{\sim} \mathcal{N}(0, I_{2\times 2})$, $\rho = 1 - 0.005^2$, X_0 and $(W_k)_{k\geq 1}$ mutually independent; conditionally on $(X_k)_{k\geq 0}$, $(Y_k)_{k\geq 1}$ and $(W_k)_{k\geq 1}$ are independent.

The state equation corresponds to a target moving on a plane. The state vector is:

$$X_k = (X_k[1], X_k[2], X_k[3], X_k[4])^* = (x_1, x_2, \dot{x}_1, \dot{x}_2)^*\,,$$

where (x_1, x_2) are the Cartesian coordinates of the target in the plane and $(\dot{x}_1, \dot{x}_2)$ are the corresponding velocities. The observer is located at the origin of the plane and accessed only to the azimuth angle $\beta = \arctan(x_1/x_2) \in [-\pi, \pi]$ corrupted by noise.

It is known that in the situation where the observer follows a rectilinear trajectory at constant speed, or zero speed as here, the filtering problem is poorly conditioned [20]. This is a situation where it is known for example that the EKF diverges very strongly.

The conditional probability density function of the measured angle Y_k given the state X_k is assumed to be a wrapped Cauchy distribution with concentration parameter ρ [23]:

$$p_{Y_k|X_k=x}(y) = \frac{(2\,\pi)^{-1}\,(1-\rho^2)}{1+\rho^2-2\,\rho\,\cos\Big(y-\arctan(\frac{x[1]}{x[2]}))\Big)} \tag{11}$$

for $-\pi \leq y < \pi$, where $\rho \in [0,1]$ is the mean resultant length. Note that the state dynamics is linear and Gaussian, but the observation dynamics is nonlinear and non-Gaussian.

Classically, in tracking studies, the trajectory model (the state equation) does not correspond to reality. In reality, the target follows a uniform straight trajectory sometimes interrupted by changes in heading and/or speed. The variance σ_W of the state equation (10) appears then as a parameter of the filter: if it is too small, the filter will have trouble tracking the target, if it is too large, the particles will spread out too much, the filter will then lose accuracy, and even lose track of the target.

For bearings-only tracking applications, the EKF and UKF filters have poor performances and will therefore not be considered in this example. On the other hand, the robust filters introduced in Section 3.3 are relevant here because there is a mismatch model.

We compare the filters in two scenarios (with $K = 40$ time steps):

- the target follows a uniform rectilinear motion;
- the target follows a uniform rectilinear motion with a change of heading at the middle of the simulation.

As we will see, the first scenario is simpler than the second. For the filter we consider two different values for σ_W: a small value $\sigma_W = 0.001$ and a large $\sigma_W = 0.003$. $\sigma_W = 0.001$ is in fact too small to be consistent with the simulated trajectory, the second value $\sigma_W = 0.003$ is more consistent with the simulated trajectory. For the robust filter we consider:

$$\mathcal{M} = \{0.0005,\, 0.001,\, 0.003,\, 0.005\}$$

as the set of possible models.

In Figure 5 we present the average computation times for filtering a trajectory as a function of different values of N (50, 100, 500, 1000, 3000, 5000) (with $S = 20$ and $R = 50$).

In Figure 6, in agreement with Figure 5, we present the evolution of the RMSE as a function of the number N of particles: the left column corresponds to the case without a turn, the right column to the case with a turn. For PBPS, BPF and APF, the first row (top) presents the RMSE for with $\sigma_W = 0.001$, the second (middle) with $\sigma_W = 0.003$. The third row (bottom) shows the RMSE for the RS-PBPS, RS-BPF, RS-APF, and DMA-BPF filters.

In Figure 7 we plot $k \to \text{RMSE}_k$ (for $N = 5000$, $R = 50$, $S = 20$) for the different filters: the left column corresponds to the case without a turn, the right column to the case with a turn; PBPS, BPF and APF for with $\sigma_W = 0.001$, on the first row; PBPS, BPF and APF for with $\sigma_W = 0.003$, on the middle row; RS-PBPS, RS-BPF, RS-APF, and DMA-BPF, on the bottom row.

In Figure 8, on a single simulation and for each filter, we plot the true trajectory of the target with the estimates of each filter: without turn in the left column, with 1 turn in the right column;

PBPS, BPS and APF filters with small $\sigma_W = 0.001$ in the top row, PBPS, BPS and APF filters with large $\sigma_W = 0.003$ in the middle row, RS-PBPS, RS-BPS, RS-APF, and DMA-BPF filters in the bottom row.

According to Figure 5, at the same number of particles, BPF is slightly faster than PBPS, both being faster than APF. The overheads for the robust version (RS) are also reasonable. In terms of complexity, the computation time is linear with the number of particles.

In order to better understand the situation, we will first comment on Figures 7 and 8: with a too small σ_W, the filters are able to track the target when it stays in a straight line (a) but they are not able to adapt when a turn occurs and continue the tracking more or less in a straight line (b). With a larger σ_W, the filters are still able to track the target when it remains in a straight line (c) with a small and increasing inaccuracy; in contrast only PBPS behaves reasonably well when a turn occurs (d). Except DMA-BPF which has a bad behavior, the RS robust filters behave better and RS-PBPS remains better than all the others (e) and (f).

From Figures 5 and 6, in the case of a turn, which is the most complex situation, we see that in order to reach PBPS/RS-PBPS accuracy with $N = 1000$ the other filters must use $N = 5000$ particles, and even in this case PBPS/RS-PBPS is at least twice as fast as the others.

V DISCUSSION AND CONCLUSION

In this paper, our aim was to propose a particle approximation method, not much more complex than the boostrap particle filter, which reduces the error variance, while being more robust in weakly conditioned situations, especially when observability conditions are poor.

We have therefore proposed an new particle filter algorithm, called predictive bootstrap particle smoother (PBPS), that deviates slightly from the strict filtering method in the sense that it takes into account the next observation in addition to the current one. We can thus consider PBPS as an approximation of the smoother with a fixed-lag interval of one step. However, the proposed algorithm is considerably simpler than the smoothing algorithms.

Furthermore, a way of making a filter more robust is to extend it to a "regime switching" strategy framework as proposed in [15]; which is what we have done by proposing a robustified version of PBPS.

In the considered examples, EKF shows very poor performance, which is perfectly understandable due to the very nonlinear nature and the poor observability conditions of the examples. It should be noted that UKF performs much better than EKF while remaining far from the performance of particle filters. Among the particle filters, APF performs slightly less well than than PBF and PBPS. Finally, even if for a given number of particles, BPF is faster than PBPS, the latter presents both a lower error variance and a much better behavior regarding observability problems. In terms of accuracy, PBPS with 1000 particles is still better than BPF with 5000 particles while being much faster. This is also the case when comparing RS-PBPS with RS-BPF

These performances are due to the fact that PBPS is a very simplified approximation of the one time step fixed-lag smoother. A strategy using a deeper fixed-lag interval of 2 or more time steps would be too costly and would not allow us to recover such performances.

In view of the performance of PBPS it would be interesting to develop strategies where the number N of particles adapts dynamically to the problem conditions. It would also be interesting to test the PBPS strategy in other cases, such as when the time between two consecutive observations is important and requires several successive prediction steps.

APPENDIX: ALGORITHMS

Algorithm A.1 Auxiliary particle filter (APF). Note that in the case of the (1)-(2), the first step, line 4, is reduced to $\bar{\xi}_k^i = \mathbb{E}(X_k|X_{k-1} = \xi_{k-1}^i) = f_{k-1}(\xi_{k-1}^i)$

1: $\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$ # initialization
2: **return** $\xi_0^{1:N}$
3: **for** $k = 1 : K$ **do**
4: $\quad \bar{\xi}_k^i = \mathbb{E}(X_k|X_{k-1} = \xi_{k-1}^i)\,,\ i = 1 : N$ # mean particle evolution
5: $\quad \bar{\omega}_k^i \leftarrow \psi_k(\bar{\xi}_k^i)\,,\ i = 1 : N$ # ... and weighting
6: $\quad \tilde{\xi}_{k-1}^{1:N} \overset{\text{iid}}{\sim} \sum_{j=1}^N \bar{\omega}_k^i\, \delta_{\xi_{k-1}^j}$ # initial particles resampling
7: $\quad \xi_{k^-}^i \sim q_k(\,\cdot\,|\tilde{\xi}_{k-1}^i)\,,\ i = 1 : N$ # particles propagation
8: $\quad \omega_k^i \leftarrow \psi_k(\xi_{k^-}^i)/\bar{\omega}_k^i\,,\ i = 1 : N$ # likelihood
9: $\quad \omega_k^i \leftarrow \omega_k^i / \sum_{j=1}^N \omega_k^j\,,\ i = 1 : N$ # renormalization
10: $\quad \xi_k^{1:N} \overset{\text{iid}}{\sim} \sum_{j=1}^N \omega_k^j\, \delta_{\xi_{k^-}^j}$ # resampling
11: $\quad$ **return** $\xi_k^{1:N}$
12: **end for**

Algorithm A.2 Dynamic model averaging bootstrap particle filter (DMA-BPF) [25].

1: $\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$, $\mu_0^{1:N} \overset{\text{iid}}{\sim} U(\mathcal{M})$ # initialization
2: **return** $\xi_0^{1:N}$
3: **for** $k = 1 : K$ **do**
4: $\quad \xi_{k^*}^i \sim q_k(\,\cdot\,|\xi_{k-1}^i, \mu_{k-1}^i), i = 1 : N$
5: $\quad \omega_k^l = \sum_{i=1}^N \psi_k(\xi_{k^*}^i) 1_{\mu_{k-1}^i = m_l}, l = 1 : L$ # likelihood of candidate models
6: $\quad N_k^{1:L} \overset{\text{iid}}{\sim} \text{Multinomial}(N; m_1, \ldots, m_L; \omega_k^1, \cdots, \omega_k^L)$
$\quad$ with $N_k^1 + \cdots + N_k^L = N$ # number of particles per candidate model
7: $\quad \hat{\xi}_{k-1}^{1,1:L} \overset{\text{iid}}{\sim} \sum_{j=1}^N \delta_{\xi_{k-1}^j} \delta_{\mu_{k-1}^j = m_l}$ # initial particles resampling per candidate model
8: $\quad (\tilde{\xi}_{k-1}^{1:N}, \mu_k^{1:N}) \leftarrow (\hat{\xi}_{k-1}^{1:N_k^1,1}, m_1), (\hat{\xi}_{k-1}^{1:N_k^2,2}, m_2), \ldots, (\hat{\xi}_{k-1}^{1:N_k^L,L}, m_L)$
9: $\quad \xi_k^i \sim q_k(\,\cdot\,|\tilde{\xi}_{k-1}^i, \mu_k^i)\,,\ i = 1 : N$ # particles propagation
10: $\quad$ **return** $\xi_k^{1:N}$
11: **end for**

Algorithm A.3 Regime Switching Bootstrap Particle Filter (RS-BPF) [15].

1: $\xi_0^{1:N} \overset{\text{iid}}{\sim} \rho_0$ # initialization
2: **return** $\xi_0^{1:N}$
3: **for** $k = 1 : K$ **do**
4: $\quad \mu_{k-1}^{1:N} \overset{\text{iid}}{\sim} U(\mathcal{M})$
5: $\quad \xi_{k^-}^i \sim q_k(\,\cdot\,|\xi_{k-1}^i, \mu_{k-1}^i), i = 1 : N$ # particles propagation
6: $\quad \omega_k^i \leftarrow \psi_k(\xi_{k^-}^i)\,,\ i = 1 : N$ # likelihood
7: $\quad \omega_k^i \leftarrow \omega_k^i / \sum_{j=1}^N \omega_k^j, i = 1 : N$ # renormalization
8: $\quad \xi_k^{1:N} \overset{\text{iid}}{\sim} \sum_{j=1}^N \omega_k^j \, \delta_{\xi_{k^-}^j}$ # resampling
9: $\quad$ **return** $\xi_k^{1:N}$
10: **end for**

REFERENCES

[1] B. D. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on Signal Processing* 50.2 (2002), pages 174–188.

[3] F. Campillo and V. Rossi. "Convolution particle filter for parameter estimation in general state-space models". In: *IEEE Transactions on Aerospace and Electronic Systems* 45.3 (2009), pages 1063–1072.

[4] D. Crisan and B. Rozovskii, editors. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.

[5] R. Douc and O. Cappé. "Comparison of resampling schemes for particle filtering". In: *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, ISPA*. 2005.

[6] A. Doucet, N. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer New York, 2001.

[7] A. Doucet and A. M. Johansen. "A Tutorial on Particle Filtering and Smoothing: Fifteen years later". In: *[4]*. 2011.

[8] A. Doucet and A. Lee. "Sequential Monte Carlo Methods". In: *Handbook of Graphical Models*. Edited by M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright. Chapman & Hall/CRC, 2018.

[9] S. Godsill. "Particle Filtering: the First 25 Years and beyond". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pages 7760–7764.

[10] S. J. Godsill and T. Clapp. "Improvement strategies for Monte Carlo particle filters". In: *[6]*. 2001, pages 139–158.

[11] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. "Novel approach to nonlinear/non–Gaussian Bayesian state estimation". In: *IEEE Proceedings F Radar and Signal Processing* 140.2 (1993), pages 107–113.

[12] W. Guo, C. Han, and M. Lei. "Improved unscented particle filter for nonlinear Bayesian estimation". In: *10th International Conference on Information Fusion*. 2007.

[13] S. J. Julier and J. K. Uhlmann. "Unscented Filtering and Nonlinear Estimation". In: *IEEE Review* 92.3 (2004), pages 401–422.

[14] G. Kitagawa. "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models". In: *Journal of Computational and Graphical Statistics* 5.1 (1996), pages 1–25.
[15] Y. El-Laham, L. Yang, P. M. Djurić, and M. F. Bugallo. "Particle Filtering Under General Regime Switching". In: *2020 28th European Signal Processing Conference (EUSIPCO)*. 2021, pages 2378–2382.
[16] T. Li, M. Bolic, and P. M. Djuric. "Resampling Methods for Particle Filtering: Classification, implementation, and strategies". In: *IEEE Signal Processing Magazine* 32.3 (2015), pages 70–86.
[17] T. Li, S. Sun, T. P. Sattar, and J. M. Corchado. "Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches". In: *Expert Systems with Applications* 41.8 (2014), pages 3944–3954.
[18] L. Liang-Qun, J. Hong-Bing, and L. Jun-Hui. "The iterated extended Kalman particle filter". In: *IEEE International Symposium on Communications and Information Technology*. Volume 2. 2005, pages 1213–1216.
[19] B. Liu. "Robust particle filter by dynamic averaging of multiple noise models". In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pages 4034–4038.
[20] S. C. Nardone and V. J. Aidala. "Observability Criteria for Bearings-Only Target Motion Analysis". In: *IEEE Transactions on Aerospace and Electronic Systems* AES-17.2 (1981), pages 162–166.
[21] S. Park, J. P. Hwang, E. Kim, and H.-J. Kang. "A New Evolutionary Particle Filter for the Prevention of Sample Impoverishment". In: *IEEE Transactions on Evolutionary Computation* 13.4 (2009), pages 801–809.
[22] M. K. Pitt and N. Shephard. "Auxiliary Variable Based Particle Filters". In: *[6]*. 2001, pages 273–293.
[23] M. K. Pitt and N. Shephard. "Filtering via simulation: Auxiliary particle filters". In: *Journal of the American Statistical Association* 94.446 (1999), pages 590–599.
[24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021.
[25] I. Urteaga, M. F. Bugallo, and P. M. Djurić. "Sequential Monte Carlo methods under model uncertainty". In: *2016 IEEE Statistical Signal Processing Workshop (SSP)*. 2016, pages 1–5.
[26] E. A. Wan and R. van der Merwe. "The unscented Kalman filter for nonlinear estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*. 2000, pages 153–158.
[27] J. Zuo. "Dynamic resampling for alleviating sample impoverishment of particle filter". In: *IET Radar, Sonar Navigation* 7.9 (2013), pages 968–977.

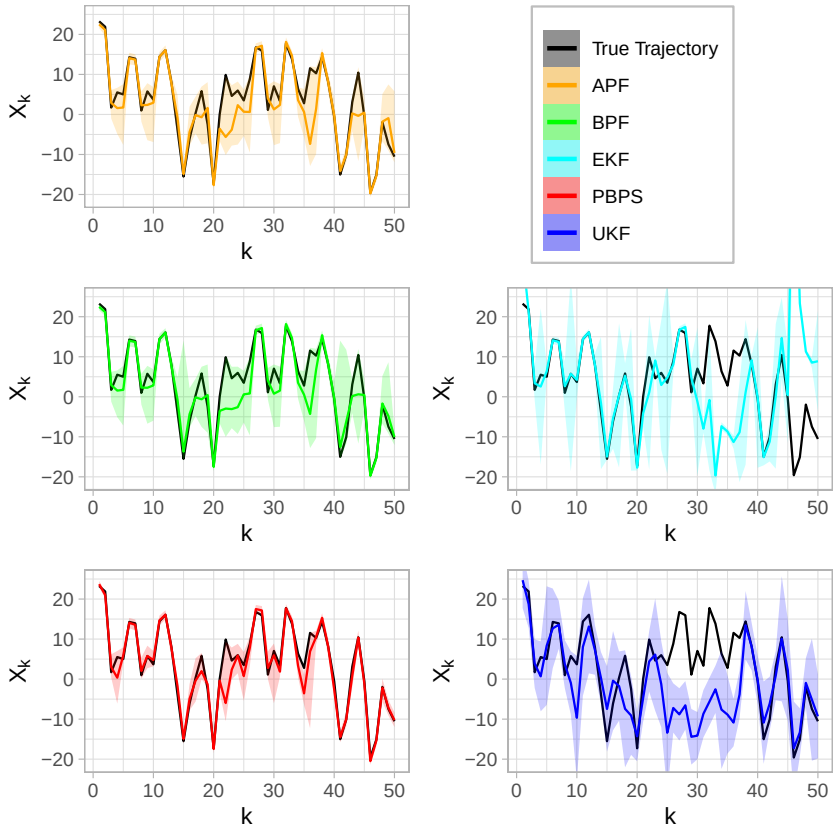

Figure 2: One-dimensional case study (9) — On a single simulation, for each filter, we plot the true state trajectory $k \to X_k$ and the approximation $k \to \hat{X}_k$ with the associated 95% confidence region ($N = 5000$).

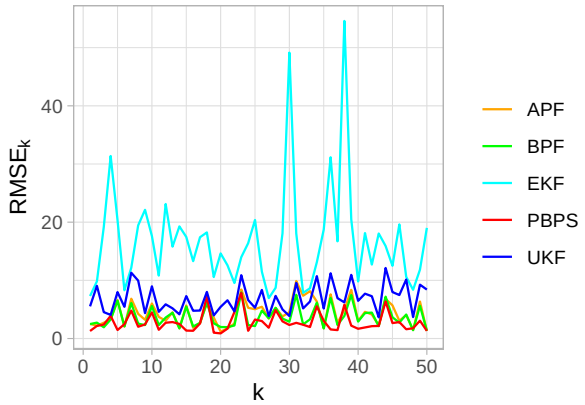

Figure 3: One-dimensional case study (9) — We compare $k \to \text{RMSE}_k$ for each filter ($S = 100$, $R = 40$, $N = 5000$).

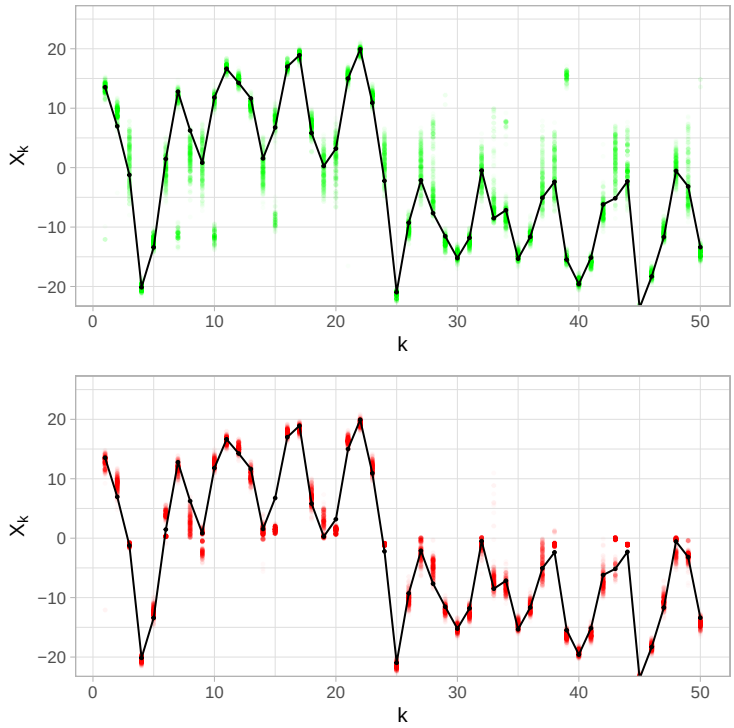

Figure 4: One-dimensional case study (9) — On a single simulation, we plot the true value $k \to X_k$ with the evolution of the set of $N = 500$ particles $k \to \xi_k^{1:N}$ for the BPF (top) and the PBPS (bottom); the particles are represented in transparency to better reflect their density.

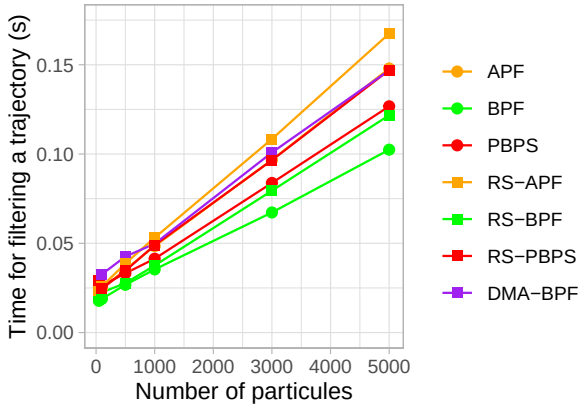

Figure 5: Four-dimensional case study (11) — We plot the average computation time for filtering a trajectory of the filters PBPS, BPF, APF, RS-PBPS, RS-BPF, RS-APF and DMA-BPF as a function of different values of N (50, 100, 500, 1000, 3000, 5000) (with $S = 20$ and $R = 50$).

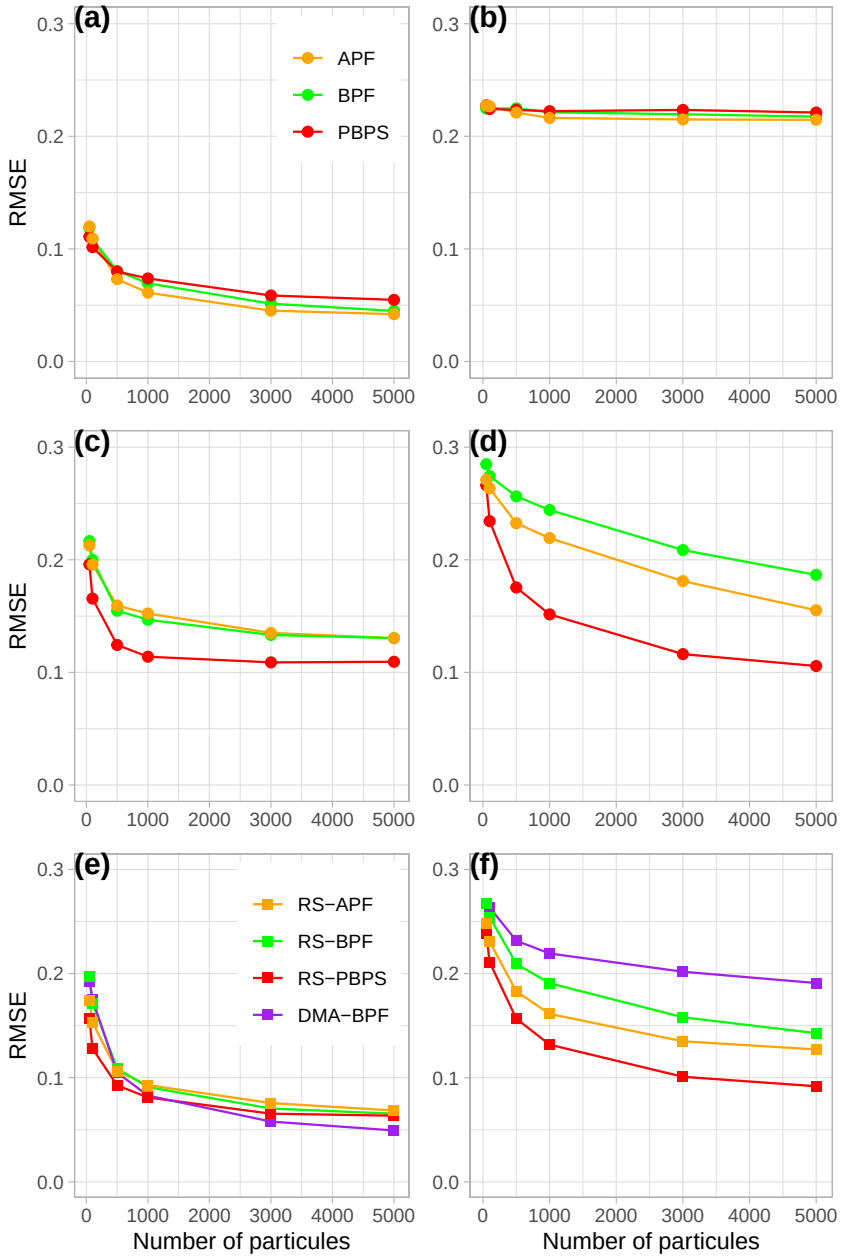

Figure 6: Four-dimensional case study (11) — Evolution of the RMSE as a function of the number N of particles: the left column corresponds to the case without a turn, the right column to the case with a turn. For PBPS, BPF and APF, the first row (top) presents the RMSE for with $\sigma_W = 0.001$, the second (middle) with $\sigma_W = 0.003$. The third row (bottom) shows the RMSE for the RS-PBPS, RS-BPF, RS-APF, and DMA-BPF filters.

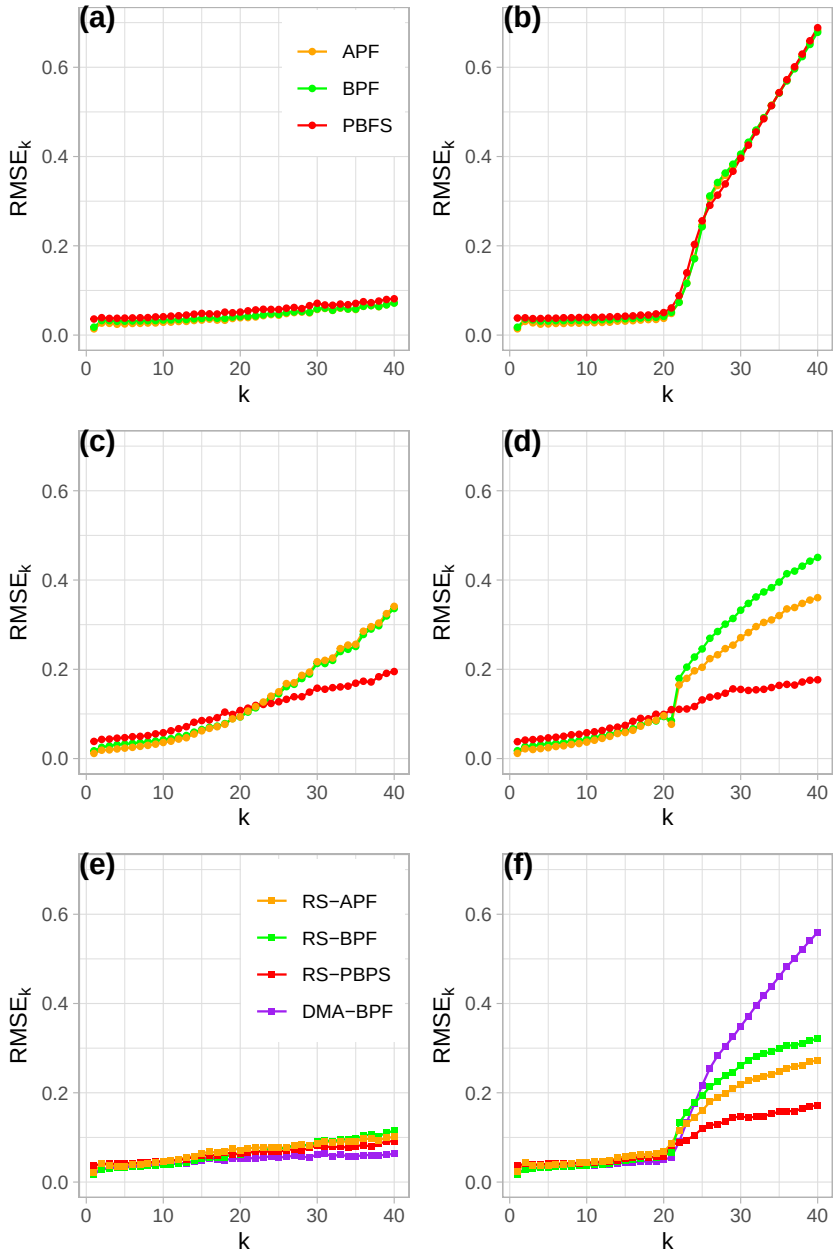

Figure 7: Four-dimensional case study (11) — $k \rightarrow \text{RMSE}_k$ (for $N = 5000$, $R = 50$, $S = 20$) for the different filters: the left column corresponds to the case without a turn, the right column to the case with a turn; PBPS, BPF and APF for with $\sigma_W = 0.001$, on the first row; PBPS, BPF and APF for with $\sigma_W = 0.003$, on the middle row; RS-PBPS, RS-BPF, RS-APF, and DMA-BPF, on the bottom row.

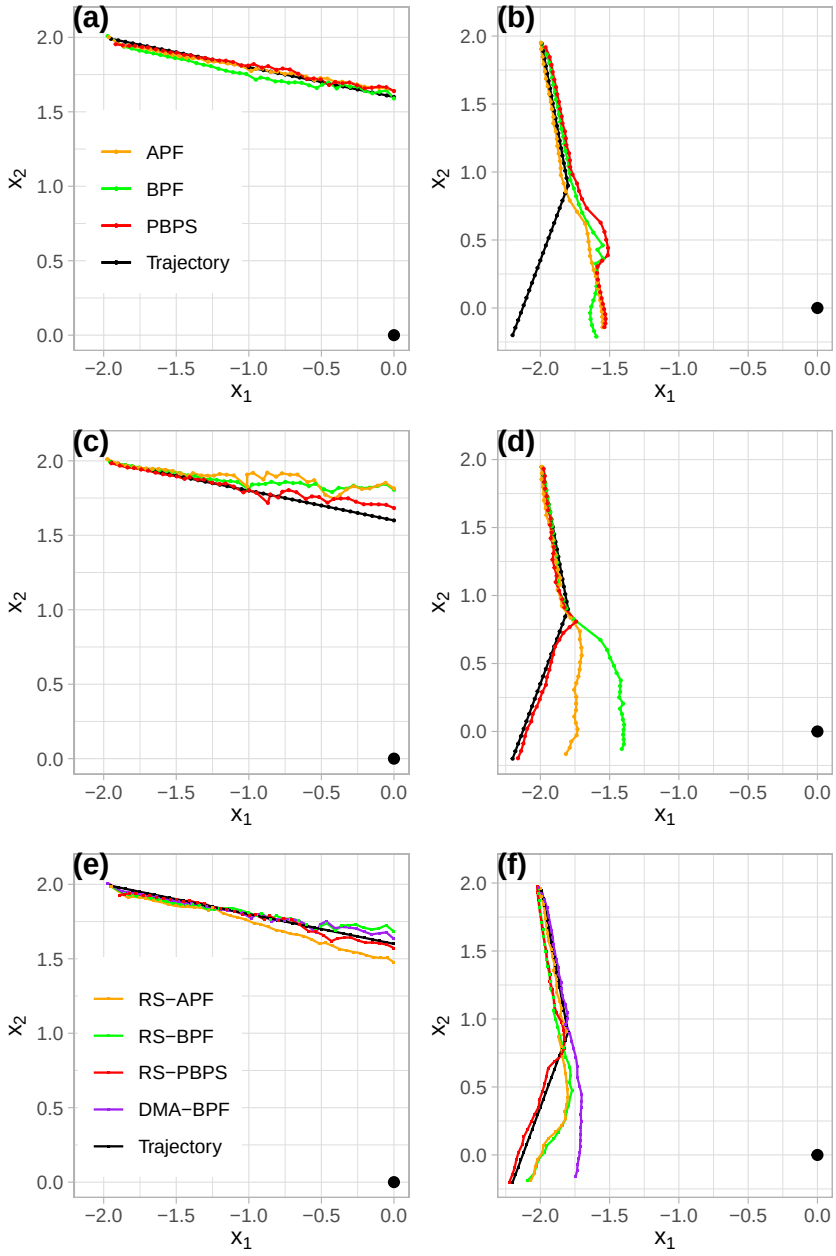

Figure 8: Four-dimensional case study (11) — On a single simulation and for each filter, we plot the true trajectory of the target with the estimates of each filter: without turn in the left column, with 1 turn in the right column; PBPS, BPS and APF filters with small $\sigma_W = 0.001$ in the top row, PBPS, BPS and APF filters with large $\sigma_W = 0.003$ in the middle row, RS-PBPS, RS-BPS, RS-APF, and DMA-BPF filters in the bottom row. The observer is in position $(0, 0)$.