

*Rapport sur la thèse de doctorat présentée par*  
**Ludovic MOULINE**

*Titre de la thèse :*  
**Towards a Modelling Framework with Temporal and Uncertain Data  
for Adaptive Systems**

Philippe Collet

**Contexte.** La thèse de doctorat défendue par M. Ludovic MOULINE concerne l'amélioration de la qualité des systèmes adaptatifs, actuellement omniprésents et sous des formes variées dans les grandes infrastructures logicielles. Ces systèmes prennent des décisions d'adaptation souvent très perfectibles, et ce, à cause de d'information de qualité plus ou moins bonne, que l'on peut définir comme de l'incertitude sur les données, mais aussi par l'absence de connaissance précise sur les actions de reconfiguration en cours à plus ou moins long terme, ce qui provoque des décisions répétées, inutiles et souvent erronées. La thèse se focalise sur un domaine émergent des systèmes adaptatifs par une application directe aux *smart grids*.

Dans ce contexte, les motivations des travaux présentés sont ainsi de faciliter, par des abstractions pertinentes et efficaces à l'exécution, la représentation d'informations incertaines et d'actions en cours dans ces systèmes adaptatifs. Les aspects temporels et l'incertitude sont alors traités comme des concepts de première classe dans un framework unifié de modélisation, s'appuyant sur des techniques d'ingénierie dirigée par les modèles à l'exécution.

Les propositions sont donc doubles : *i)* un langage qui gère l'incertitude dans les données, en permettant d'attacher une distribution de probabilité représentant l'incertitude, et de la propager automatiquement à l'aide des opérateurs booléens ou arithmétiques classiques, *ii)* un modèle temporel de contexte, qui permet de capturer les actions passées, en cours, et futures, ce qui permet aux développeurs de connaître les actions en cours et leur futurs impacts, mais aussi de naviguer dans les décisions passées pour comprendre pourquoi elles ont amenées à un état insatisfaisant.

**Document et analyse.** Le mémoire proposé par Ludovic MOULINE est composé de 8 chapitres, organisé en 3 parties, le tout rédigé en anglais. Le document de 220 pages est très bien organisé, et se lit avec facilité.

La première partie regroupe le contexte, les défis et naturellement, le positionnement vis-à-vis de l'état de l'art.

Le premier chapitre (12 pages) effectue naturellement l'introduction du travail, en présentant rapidement le contexte et les défis de haut niveau autour de l'ingénierie logicielle, en tenant compte de données incertaines, et en raisonnant sur des actions adaptatives à long terme. Des défis plus précis sont aussi discutés concernant la représentation et le stockage des différentes abstractions : d'une part les états et comportements inconsistants, et d'autre part, les données temporelles. Viennent ensuite les trois défis spécifiques à la thèse : faciliter la manipulation de l'incertitude des données, permettre le raisonnement sur les actions en cours et leurs effets, modéliser les décisions d'adaptation pour les diagnostiquer. Le chapitre se termine par une première présentation des contributions, de leur validation, et classiquement du plan du document. Il est à

noter que ce premier chapitre est déjà très illustré, ce qui permet au lecteur de concrétiser des problématiques de haut niveau par des illustrations pertinentes.

Le chapitre 2 (20 pages) effectue une mise en contexte de l'ensemble des éléments scientifiques nécessaires à la compréhension de la problématique et des techniques utilisées pour développer les contributions de la thèse. Le chapitre présente ainsi les systèmes adaptatifs, l'ingénierie dirigée par les modèles, l'ingénierie des langages, ainsi que des éléments de théorie des probabilités. Il est intéressant de noter que chacune des sections se termine par une explication sur l'utilisation faite de chaque élément de contexte dans la thèse. Cela positionne déjà les contributions en termes d'avancées ou de réutilisation vis-à-vis de l'état de l'art.

Le choix est fait dans cette thèse de détailler un exemple complet d'illustration dans le chapitre 3 (15 pages), cet exemple servant ensuite dans les sections de validation. L'exemple est basé sur la *smart grid* déployée au Luxembourg par une entreprise avec laquelle une collaboration a été menée pendant les travaux. Cet exemple a l'avantage d'illustrer des besoins concrets, comme l'estimation de la charge sur un câble, mais surtout les impacts d'un raisonnement sans tenir compte de l'incertitude des données, ainsi que la difficulté à gérer cette incertitude sans le support fourni par les contributions proposées. Des exemples d'actions à long terme sont aussi introduits, démontrant aussi la pertinence de la problématique considérée. Le chapitre se termine par l'émergence de trois scénarios de cas d'utilisation.

Cette première partie se termine naturellement par une étude de l'état de l'art effectuée dans le chapitre 4 (29 pages). Ce chapitre est fortement structuré par deux questions de recherche, pour chaque problématique abordée par la thèse, et par des sous-questions plus précises. L'étude suit partiellement une forme de *SLR* améliorée pour déterminer les travaux à considérer.

Concernant le raisonnement sur l'état d'un système adaptatif, l'étude s'intéresse ainsi aux travaux connexes sur la modélisation de l'évolution du contexte ou du système, la modélisation des actions et de leurs effets, et bien évidemment les capacités de raisonnement sur ces modèles. Concernant l'incertitude sur les données, l'étude recherche les catégories d'incertitude considérées, leur modélisation plus ou moins explicite, et les capacités de raisonnement.

L'ensemble de l'étude couvre donc une grande diversité de travaux et conclut à l'absence de support pour le raisonnement sur l'adaptation, et à une analyse plus fine et nuancée sur l'incertitude de données. Dans sa partie descriptive, l'étude montre bien les inspirations et compromis envisagés pour la contribution qui va suivre sur la partie incertitude des données, mais moins sur la partie concernant le raisonnement de l'adaptation. Heureusement, une section de conclusion effectue une synthèse pertinente. L'ensemble reste cependant cohérent, très structuré, et démontre clairement la maîtrise de l'état de l'art par le candidat.

La partie II du document détaillent les contributions, avec trois chapitres, les deux derniers intégrant une section de validation spécifique.

Le chapitre 5 (4 pages) est très court. A partir d'une synthèse des défis déjà définis précédemment, il décrit une vision architecturale fonctionnelle du framework de modélisation proposée dans cette thèse. La figure 5.1, qui décrit cette architecture, montre les différents éléments et interactions, et on perçoit bien la portée des contributions et leur cohérence. En revanche cette figure est aussi relativement complexe, et manque un peu d'explications dans ce chapitre. Il faudra au lecteur un passage à travers les chapitres suivants pour bien assembler les éléments de contribution fournis.

Le chapitre 6 (33 pages) décrit une première contribution forte de la thèse, le langage Ain'tea, qui intègre explicitement les notions d'incertitude précédemment déterminées dans un langage de manipulation des données. Une donnée est ainsi forcément attachée à une distribution de probabilité. L'impact de cette introduction est alors étudié selon les axes syntaxiques et sémantiques. Les types de base représentés sont les booléens, les valeurs numériques et les références. Pour chacun, les opérateurs de base, qui interprètent la probabilité sur les valeurs manipulées, sont définis et illustrés. On y retrouve des opérateurs d'existence, de confiance, de comparaison d'identité et d'égalité, parfois de conversion et de navigation, et enfin des opérations calculatoires propageant automatiquement l'incertitude des opérandes. La sémantique statique est définie avec des règles de typage qui reprennent naturellement les combinaisons possibles entre distributions de probabilité. L'ensemble des définitions est bien posé et illustré, fournissant un fondement solide à l'utilisation ou à l'implémentation du langage.

Le chapitre effectue ensuite une évaluation de cette extension, en cherchant à évaluer la concision et la capacité du système de typage défini à détecter des erreurs de gestion de l'incertitude. L'implémentation d'un prototype utilisant Xtext est alors présenté et un extrait d'implémentation en Ain'tea est comparé avec deux implémentations similaires en C# et Python. De façon similaire, la détection d'une incohérence est aussi démontrée sur un exemple. Dans les deux cas, on est face à une illustration bien claire et argumentée du résultat, la concision et la propagation automatique sont montrées, comme la détection d'incohérence de type, mais je trouve cette illustration insuffisante pour conclure directement aux deux questions de recherche posées. Évaluer les qualités d'un langage est un problème extrêmement complexe. La concision, comme la facilité d'utilisation, d'un langage ne sont pas faciles à démontrer, et nous n'avons ici qu'une illustration. Concernant le typage, un ensemble de tests unitaires couvrant des cas systématiquement déterminés aurait fourni une réponse de meilleure qualité, d'autant qu'ils sont certainement présents dans l'implémentation. Si la validation est sans doute présentée dans des termes un peu ambitieux, cela n'enlève rien à la contribution tout à fait solide, qui répond bien aux exigences définies plus avant, et qui fait distinctement avancer l'état de l'art.

Le chapitre 7 (26 pages) présente un métamodèle pour décrire des connaissances de systèmes adaptatifs (contexte, besoins, actions) en tenant compte de notions temporelles. S'appuyant sur les défis identifiés dans la première partie du document, le chapitre décrit d'abord une formalisation de représentation basée sur des graphes augmentés par la dimension temporelle. La formalisation est solide, bien argumentée (requêtes, actions, etc.), mais aussi bien illustrée. La section se termine par une instanciation du modèle proposé sur un cas d'utilisation du chapitre 3.

Le métamodèle est ensuite introduit. Il est découpé en 4 packages, ce qui permet de bien comprendre son organisation autour de la connaissance pour ensuite représenter le contexte, les besoins, et les actions. Je trouve particulièrement appréciable que chacune de ces parties soit aussi discutée en termes de positionnement par rapport à l'état de l'art, motivant les choix de représentation.

Une section de validation suit, elle s'appuie sur le développement d'un prototype, disponible sur github, lui-même basé sur le framework GreyCat. La section revient sur les deux questions de recherche attenantes à cette partie. Il est ainsi illustré comment le diagnostic d'une *smart grid* peut être capturé par un modèle obtenu à l'aide du métamodèle qui structure ainsi les données à l'exécution, et comment il peut être effectué en requêtant le modèle. Un algorithme de raisonnement est aussi fourni comme illustration de l'exploitation de ce modèle sur un des cas d'utilisation, répondant ainsi à la deuxième question. Une évaluation de performance complète cette section en étudiant comment les performances de navigation dans le graphe, l'opération la plus utilisée pour un diagnostic, sont impactées par sa taille, la taille de l'entrée et le nombre d'éléments traversés. L'ensemble est convaincant, et complété par une discussion sur la potentielle

application sur la *smart grid* luxembourgeoise. Ceci montre bien la faisabilité, les capacités de passage à l'échelle et l'applicabilité de l'ensemble du framework proposé.

Globalement ce chapitre constitue une seconde contribution solide, bien fondée et bien argumentée, qui fait clairement avancer l'état de l'art. La partie de validation est à la fois pertinente et convaincante. Sur la forme, je me pose la question du positionnement du fondement théorique des graphes temporels par rapport à celui utilisé par GreyCat. Qu'il soit identique ou pas, il aurait été judicieux de bien positionner la chose.

La troisième et dernière partie s'intéresse naturellement aux conclusions et aux perspectives de recherche à travers un seul chapitre.

Le chapitre 8 (11 pages) effectue d'abord un résumé des contributions relativement bref, mais bien construit. Suit une partie tout à fait intéressante sur les perspectives ouvertes par ces travaux, tant au niveau d'une intégration plus poussée de l'incertitude dans les langages, qu'une meilleure explication, la plus automatisée possible, des systèmes adaptatifs. Je suis particulièrement sensible à ces dernières perspectives, que la question de la complexité résiduelle à gérer par les développeurs m'est clairement apparue dans le chapitre 7 : l'ensemble des concepts fournis ne fournissent-ils pas actuellement une API seulement maîtrisable par des développeurs très experts ?

Sur le fond, le travail effectué par Ludovic MOULINE est considérable, rigoureux, et apporte des contributions tout à fait significatives dans les domaines des systèmes adaptatifs et de l'ingénierie des modèles à l'exécution. Ce travail fournit des propositions consistantes qui font clairement avancer l'état de l'art. Les solutions proposées sont à la fois très pertinentes et bien fondées, comme les éléments de validation du manuscrit et les publications scientifiques dans des conférences de référence le démontrent. Sur la forme, le manuscrit, très bien construit et rédigé avec un souci constant de l'illustration, contient tous les éléments nécessaires pour évaluer un travail de thèse de grande qualité.

**Conclusion.** Au vu de l'ensemble des arguments présentés dans ce rapport, **j'émet un avis très favorable à la soutenance de thèse de M. Ludovic MOULINE en vue de l'obtention du diplôme de Docteur de l'Université de Rennes 1 spécialité Informatique.**

Sophia-Antipolis, le 31 octobre 2019



Philippe Collet, Professeur des Universités  
Université Côte d'Azur, CNRS / I3S

