

TITRE DE LA THESE : Élaboration d'un langage dédié pour la connexion de données de terrain aux modèles épidémiologiques mécanistes (DatEpiLang)

Direction de thèse : Massimo TISI (IMT Atlantique), Sébastien Picault (INRAE/BIOEPAR)

Co-encadrement de thèse : Erwan BOUSSE (Nantes Université)

Laboratoires : LS2N (équipe Naomod), BIOEPAR

Contexte :

Les modèles épidémiologiques mécanistes ont montré leur intérêt majeur pour comprendre, prédire et anticiper la propagation d'agents pathogènes. Les données de terrain jouent des rôles croissants et diversifiés dans les modèles épidémiologiques : forçage de conditions environnementales, estimation de paramètres non observables, sélection de scénarios pertinents, ou identification des conditions initiales. Leur complexité vient aussi de la diversité des échelles auxquelles elles sont collectées, au regard de la diversité des échelles mobilisées dans les modèles, tant spatiales (animal ; lot ou troupeau ; ambiance, bâtiment ou environnement) que temporelles (de l'ordre de l'année pour la surveillance réglementaire à la seconde pour des capteurs en élevage). Dans ce contexte, le couplage entre données et modèles reste un problème difficile, avec des méthodes *ad hoc*, un temps de développement important et une faible réutilisabilité.

Objectifs de la thèse :

L'objectif de cette thèse est de concevoir un langage dédié qui permettra de relier efficacement les données de terrain aux modèles épidémiologiques mécanistes. L'objectif est double : 1) standardiser le processus d'intégration de données pour améliorer la fiabilité et la pertinence des prédictions des modèles épidémiologiques, et 2) rendre ce processus plus accessible aux chercheurs et professionnels, augmentant ainsi la réactivité et l'efficacité des interventions en cas d'épidémie. Cette innovation méthodologique vise à transformer la complexité et l'hétérogénéité des données de terrain en informations exploitables, capables d'alimenter avec précision des modèles prédictifs. La question pratique de l'intégration de données recueillies en élevage (échographies pulmonaires, vidéos et son) dans les modèles de maladies respiratoires fait l'objet de la thèse CIFRE de Loïc Eyango (Adventiel), aux moyens de méthodes spécifiques. La thèse visera à généraliser et systématiser ces approches et à les valider par des cas d'usage concrets, en s'appuyant sur les compétences complémentaires de modélisation (BIOEPAR Dynamo) et d'ingénierie des modelés (LS2N Naomod) des co-directeurs. Naomod notamment étendra sa théorie pour la transformation incrémentale, pour permettre la définition de langages de transformation dédiés. La thèse contribuera à ouvrir un nouveau terrain d'application pour la plateforme low-code de Naomod (Lowcomote) dans le domaine de la santé.

Compétences attendues du ou de la candidat·e :

Étudiant·e en M2 informatique ou équivalent. Formation solide en génie logiciel (notamment model-driven engineering) et notions d'intelligence artificielle (notamment systèmes multi-agents). Des compétences rédactionnelles et un intérêt pour la modélisation et les sciences du vivant seront nécessaires pour faciliter le dialogue avec les modélisateurs et épidémiologistes, intégrer leurs connaissances expertes dans le langage dédié et documenter les choix des méthodes informatiques mobilisées.

Références :

[1] Eyango Tabi TGL, Rouault M, Potdevin V, L'Hostis X, Assié S, Picault S, Parisey N (2024) Deep mechanistic model: Integrating deep learning and stochastic mechanistic approaches for Bovine Respiratory Diseases diagnosis and epidemiological forecasting. In: Proceedings of the Conference of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM). Uppsala, Sweden

- [2] Picault S, Huang Y-L, Sicard V, Arnoux S, Beaunée G, Ezanno P (2019) EMULSION: Transparent and flexible multiscale stochastic models in human, animal and plant epidemiology. *PLoS Comput Biol* 15:e1007342. <https://doi.org/10.1371/journal.pcbi.1007342>
- [3] Picault S, Niang G, Sicard V, Sorin-Dupont B, Assié S, Ezanno P (2023) Leveraging artificial intelligence and software engineering methods in epidemiology for the co-creation of decision-support tools based on mechanistic models. *bioRxiv*. <https://doi.org/10.1101/2023.09.03.555060>
- [4] Josselin Enet, Erwan Bousse, Massimo Tisi, Gerson Sunyé, On the Suitability of LSP and DAP for Domain-Specific Languages. *MoDELS (Companion) 2023*: 353-363
- [5] Davide Di Ruscio, Dimitrios S. Kolovos, Juan de Lara, Alfonso Pierantonio, Massimo Tisi, Manuel Wimmer: Low-code development and model-driven engineering: Two sides of the same coin? *Softw. Syst. Model.* 21(2): 437-446 (2022)
- [6] Zahra Rajaei, Shekoufeh Kolahtouz Rahimi, Massimo Tisi, Frédéric Jouault: A DSL for Encoding Models for Graph-Learning Processes. *STAF Workshops 2021*: 149-161