Internship Proposal
Hybrid Federated Causal Discovery

Context: The Pierre Louis Institute of Epidemiology and Public Health (IPLESP), co-accredited by Inserm and Sorbonne University, brings together research strengths in epidemiology and public health within Sorbonne University. IPLESP's main objective is to produce original knowledge on pressing public health issues and related intervention effectiveness, focusing on emerging infectious diseases, chronic diseases, environmental health, and mental health. To tackle these challenges, causal inference Pearl et al. [2000], Hernan and Robins [2025] emerges as an indispensable tool. Therefore, at IPLESP, we are establishing a new team dedicated to developing advanced methodologies rooted in causal inference.
In this context, access to causal graphs is essential for estimating causal effects [Greenland et al., 1999, Savitz and Wellenius, 2016]. These graphs represent qualitative cause-and-effect relationships between exposures, health outcomes, and other variables. However, in many applications, it is challenging for a practitioner to provide such a graph. To this end, Causal Discovery Spirtes et al. [2000], an active research field of causal inference, attempts to discover a causal graph from observational data. New methods are regularly proposed, but no single method stands out. Indeed, they all rely on assumptions that may or may not be appropriate for a particular dataset Assaad et al. [2022]. In many cases, the results of causal discovery methods are still unsatisfactory in real-world applications Aït-Bachir et al. [2023]. Nevertheless, it has been shown that multiple datasets from different environments can improve causal discovery Mooij et al. [2020], Huang et al. [2020]. However, the French regulatory context, characterized by the extremely strict application of the General Data Protection Regulation (GDPR) to preserve data confidentiality, makes creating a causal graph from multiple datasets challenging. Therefore, it is important to start developing a federated causal discovery method that preserves privacy. Federated in the sense that we need to learn a causal graph (or an abstraction) from many datasets representing all these datasets and preserving confidentiality in the sense that we cannot compromise patient privacy Mian et al. [2023]. Federated causal discovery is crucial in the current context, especially in the healthcare field. This is particularly relevant when considering the existence of a causal graph representing complex relationships, not recoverable solely from data from a single environment but recoverable by combining data from multiple environments.

Proposal: The objective of the internship is to develop a hybrid method for federated causal discovery. Indeed, as shown in Bystrova et al. [2024], mixing different causal discovery methods allows to relax some of the canonical assumptions in causal discovery. To this end, the candidate will:

- Familiarize with causal discovery Spirtes et al. [2000] and federated causal discovery literature Mian et al. [2023], Li et al. [2024], Wang et al. [2023], Yang et al. [2023], Meurisse et al. [2023], Vo et al. [2021], Xiong et al. [2023].
- Implement a hybrid method for federated causal discovery.
- Validate the method on synthetic and real-world data.

Candidate profile: Highly motivated candidate with an M2 degree and a strong background in computer science, machine learning, probability, and causal inference, along with a keen interest in epidemiology and public health. Proficiency in Python programming is also required. Knowledge of the English language is required.

Location: The intern will work at IPLESP (https://iplesp.fr/), located in Paris. They will be supervised by Federico Baldo (PostDoc) and Charles Assaad.

Dates: Starting date: To be discussed, early 2026, for 5-6 months.

Contact: To apply, please send a CV and a cover letter to Federico Baldo federico.baldo@inserm.fr, only candidature in English will be examined.

References

A. Aït-Bachir, C. K. Assaad, C. de Bignicourt, E. Devijver, S. Ferreira, E. Gaussier, H. Mohanna, and L. Zan. Case studies of causal discovery from it monitoring time series. *arXiv preprint arXiv:2307.15678*, 2023.

C. K. Assaad, E. Devijver, and E. Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73:767–819, 2022.

D. Bystrova, C. K. Assaad, J. Arbel, E. Devijver, E. Gaussier, and W. Thuiller. Causal discovery from time series with hybrids of constraint-based and noise-based algorithms. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL `https://openreview.net/forum?id=PGLbZpVk2n`.

S. Greenland, J. Pearl, and J. Robins. Causal diagrams for epidemiologic research. *Epidemiology (Cambridge, Mass.)*, 10(1):37—48, January 1999. ISSN 1044-3983. doi: 10.1097/00001648-199901000-00005.

M. Hernan and J. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2025. ISBN 9781420076165. URL `https://books.google.fr/books?id=_KnHIAAACAAJ`.

B. Huang, K. Zhang, J. Zhang, J. Ramsey, R. Sanchez-Romero, C. Glymour, and B. Schölkopf. Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research*, 21(89):1–53, 2020.

L. Li, I. Ng, G. Luo, B. Huang, G. Chen, T. Liu, B. Gu, and K. Zhang. Federated causal discovery from heterogeneous data, 2024. URL `https://arxiv.org/abs/2402.13241`.

M. Meurisse, F. Estupiñán-Romero, J. González-Galindo, N. Martínez-Lizaga, S. Royo-Sierra, S. Saldner, L. Dolanski-Aghamanoukjan, A. Degelsegger-Marquez, S. Soiland-Reyes, N. Van Goethem, et al. Federated causal inference based on real-world observational data sources: application to a sars-cov-2 vaccine effectiveness assessment. *BMC medical research methodology*, 23(1):248, 2023.

O. Mian, D. Kaltenpoth, M. Kamp, and J. Vreeken. Nothing but regrets — privacy-preserving federated causal discovery. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8263–8278. PMLR, 25–27 Apr 2023. URL `https://proceedings.mlr.press/v206/mian23a.html`.

J. M. Mooij, S. Magliacane, and T. Claassen. Joint causal inference from multiple contexts. *Journal of machine learning research*, 21(99):1–108, 2020.

J. Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2):3, 2000.

D. A. Savitz and G. A. Wellenius. Causal Diagrams for Epidemiologic Inference. In *Interpreting Epidemiologic Evidence: Connecting Research to Applications*. Oxford University Press, 08 2016. ISBN 9780190243777. doi: 10.1093/acprof:oso/9780190243777.003.0003.

P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. MIT press, 2000.

T. V. Vo, T. N. Hoang, Y. Lee, and T.-Y. Leong. Federated estimation of causal effects from observational data. *arXiv preprint arXiv:2106.00456*, 2021.

Z. Wang, P. Ma, and S. Wang. *Towards Practical Federated Causal Structure Learning*, page 351–367. Springer Nature Switzerland, 2023. ISBN 9783031434150. doi: 10.1007/978-3-031-43415-0_21. URL `http://dx.doi.org/10.1007/978-3-031-43415-0_21`.

R. Xiong, A. Koenecke, M. Powell, Z. Shen, J. T. Vogelstein, and S. Athey. Federated causal inference in heterogeneous observational data. *Statistics in Medicine*, 42(24):4418–4439, 2023.

D. Yang, X. He, J. Wang, G. Yu, C. Domeniconi, and J. Zhang. Federated causality learning with explainable adaptive optimization, 2023. URL `https://arxiv.org/abs/2312.05540`.