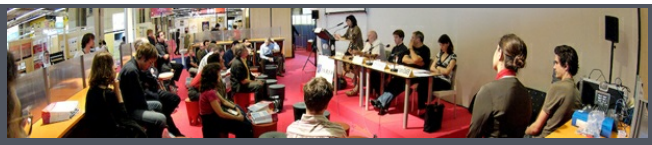
INVENTEURS DU MONDE NUMÉRIQUE

INTERNATIONAL RELATIONS



INTERNSHIP SUBJECT

2875 - Penalized Mixed-Effects Meta-Regression to Explore Sources of Het

Introduction

Meta-analysis is a statistical technique that synthesizes results from independent studies addressing a common research question by computing a combined effect size. By aggregating information, it enhances statistical power. However, a critical prerequisite is the assessment of heterogeneity—variation in effect sizes across studies—which may threaten the validity of pooled conclusions if not appropriately modeled.

A fundamental approach to modeling heterogeneity in meta-analysis is the random-effects model. Let yi denote the observed effect size from study i (for i = 1, ..., N). It is assumed that:

$y_i = \theta_i + e_i, \quad e_i \sim N(0, v_i)$

Here, θ_i is the (unknown) true effect in study i, and e_i is the sampling error, with known within-study variance v_i. Under standard assumptions, the observed effect sizes y_i are unbiased and normally distributed estimates of θ_i. To model between-study heterogeneity, the true effects θ_i are assumed to vary around a common mean μ:

$\theta_i = \mu + u, \quad u_i \sim N(0, \tau^2)$

where τ^2 represents the between-study variance. To account for potential sources of heterogeneity, study-level characteristics (moderators) can be incorporated into a mixed-effects meta-regression model:

$\theta_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_{ip} + u_i, \quad u_i \sim N(0, \tau^2)$

where x_{ij} denotes the value of the jth moderator for study i, β_j the corresponding regression coefficient, and τ^2 captures the residual heterogeneity not explained by the moderators.

These models are special cases of linear mixed-effects models with known heteroscedastic sampling variances. Estimation typically follows a two-step procedure: A. Estimate the between-study variance τ^2; B. Estimate μ or $\beta = (\beta_0, \ldots, \beta_p)$ using weighted least squares with weights: $w_i = 1 / (v_i + \tau^2)$. Standard errors and confidence intervals are computed assuming normality. Hypothesis tests such as $H_0: \tau^2 = 0$ are commonly performed using Cochran's Q-test.

Application in Microbiome Research

Advances in next-generation sequencing have revolutionized microbiome research. Meta-analyses in this field allow synthesis of microbiota–disease associations, but they are particularly prone to heterogeneity due to differences in sequencing technologies, bioinformatics pipelines, sample types, population characteristics. In areas such as gut, respiratory or oral microbiomes, where fewer studies are available and study characteristics vary widely, the resulting high dimensionality and limited sample size pose analytical challenges. Standard meta-regression models may be inadequate in these cases. To address this, researchers have applied multivariate data analysis (e.g., PCA for quantitative variables, MCA for categorical variables, and FAMD for mixed data), or Bayesian penalized regression methods to select informative moderators in high-dimensional settings.

Main activities:

- Investigate traditional mixed-effects models with a limited number of moderators using the metafor R package;
- Explore Bayesian penalized meta-regression approaches using the pema R package;
- Assess the feasibility of implementing non-Bayesian penalized meta-regression using, for example, the CVXR R package;
- Apply the studied methods to reanalyze recent meta-analyses on human microbiome alpha-diversity and disease associations, incorporating

Required Skills

Statistical and Machine Learning Knowledge
A solid foundation in statistical methods and machine learning is required. Prior knowledge of mixed-effects models and penalized regression techniques (e.g., Lasso) would be a strong advantage.

Programming Skills
The trainee should be proficient with computational tools, particularly R or Python. While deep expertise in both languages is not required, solid general programming skills are necessary to quickly adapt to the R environment at the beginning of the internship.

Interest in Biomedical Applications
The methods studied will be applied in a clinical research context, specifically in meta-analyses of human microbiome studies related to health and disease. Although prior expertise in microbiome or biomedical research is not required, a strong curiosity for biomedical topics, motivation to invest in learning, and good data interpretation skills are essential.

General Information

- **Research Theme :** Modeling and Control for Life Sciences
- **Locality :** Talence
- **Level :** Master
- **Period :** 15th January 2026 -> 14th April 2026 (3 months)

These are approximative dates. Please contact the training supervisor to know the precise period.

- **Deadline to apply :** 1st July 2025 (midnight)

Contacts

- **Training Supervisor :** Marta Avalos Fernandez / marta.avalos-fernandez@inria.fr
- **Team Manager :** Melanie Prague / melanie.prague@inria.fr

study-level characteristics as moderators to better understand heterogeneity.

References

- Avalos, M., Métayer, C., Alin, T., Thiébaut, R., Enaud, R., et al. (2022). *The respiratory microbiota alpha-diversity in chronic lung diseases: First systematic review and meta-analysis*. Respiratory Research, 23(214). https://doi.org/10.1186/s12931-022-02132-4
- Blázquez-Rincón, D., Sánchez-Meca, J., Botella, J., & Suero, M. (2023). *Heterogeneity estimation in meta-analysis of standardized mean differences when the distribution of random effects departs from normal: A Monte Carlo simulation study*. BMC Medical Research Methodology, 23(1), 19. https://doi.org/10.1186/s12874-022-01809-0
- Broderick, D., Marsh, R., Waite, D., Pillarisetti, N., Chang, A. B., & Taylor, M. W. (2023). *Realising respiratory microbiomic meta-analyses: Time for a standardised framework*. Microbiome, 11(1), 57. https://doi.org/10.1186/s40168-023-01499-w
- Cao, M., Wang, J., Liu, P., et al. (2023). *Gut microbiota composition in depressive disorder: A systematic review, meta-analysis, and meta-regression*. Translational Psychiatry, 13, 379. https://doi.org/10.1038/s41398-023-02670-5
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman & Hall/CRC. https://cran.r-project.org/web/packages/pema/vignettes/meta-analysis_tutorial.html
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.5). Cochrane. https://training.cochrane.org/handbook/current
- Kou, Z., Liu, K., Qiao, Z., Wang, Y., Li, Y., Li, Y., Yu, X., & Han, W. (2024). *The alterations of oral, airway, and intestine microbiota in chronic obstructive pulmonary disease: A systematic review and meta-analysis*. Frontiers in Immunology, 15, 1407439. https://doi.org/10.3389/fimmu.2024.1407439
- Saracco, A., Chavent, M., & Avalos, M. (2023). *Utility of multivariate data analysis and penalized meta-regression to explore sources of heterogeneity in microbiome meta-analyses*. In World of Microbiome Conference 2023, Sofia, Bulgaria. https://hal.science/hal-04260888
- Van Lissa, C. J., Clapper, E.-B., & Kuiper, R. (2024). *A tutorial on aggregating evidence from conceptual replication studies using the product Bayes factor*. Research Synthesis Methods. https://doi.org/10.1002/jrsm.1765
- Van Lissa, C. J., van Erp, S., & Clapper, E.-B. (2023). *Selecting relevant moderators with Bayesian regularized meta-regression*. Research Synthesis Methods, 14(2), 301–322. https://doi.org/10.1002/jrsm.1628
- Viechtbauer, W. (2010). *Conducting meta-analyses in R with the metafor package*. Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

More information

- **Inria Team :** SISTM
- **Inria Center :** Centre Inria de l'université de Bordeaux