StanCon 2026

International Conference on Bayesian Inference and Probabilistic Programming

17-21 August, 2026

Uppsala, Sweden

Contents

Oral

When: 2026-08-19, 13:20 - 13:40, Where: Ångström Laboratory

A scalable hierarchical nonparametric stochastic differential equation framework for reconstructing stability landscapes from sparse microbiome data

Chandler Ross[1]
Ville Laitinen[1], Guilhem Sommeria-Klein[2], Leo Lahti[1]
[1] University of Turku, Finland
[2] Inria, University of Bordeaux, France

Abstract: Reconstructing the dynamics of complex systems from short, disjoint time series is a major challenge often faced in statistics. Due to ethical or practical constraints, many datasets contain sparse longitudinal data across many individuals, making it difficult to estimate key dynamical properties such as stability, resilience, and the presence of alternative stable states. Existing approaches rely on long, dense, and continuous time series, often fail to characterize uncertainty, and assume predefined functional forms.

An alternative approach is to reconstruct the dynamical landscape using non-parametric models. We overcome the above limitations by combining information across multiple short time series using Bayesian inference. Our approach decomposes the dynamics into the drift (deterministic) and diffusion (stochastic) components of a stochastic differential equation (SDE). Both components are represented as two-dimensional Gaussian processes over the system state variable and a covariate. This structure allows for partial pooling across the covariate of choice: each point on the grid is correlated with every other point to create a smoothly evolving landscape not sensitive to outliers and that extrapolates to regions with limited to no observations. To further constrain the SDE components and facilitate inference, we utilize cross-sectional data points to inform the stationary density through hierarchical priors, enabling us to benefit from entire datasets. To accommodate larger datasets and improve scalability, we implemented the Hilbert space basis function approximation. Inference was carried out in Stan using Hamiltonian Monte Carlo.

We apply this framework to study the dynamical properties of components that summarize the compositions of communities in the human gut microbiome called enterosignatures. These signatures are associated with health and disease outcomes and are not well understood from a stability perspective. Our approach allows us to reconstruct two-dimensional stability landscapes across the natural lifespan without requiring dense, individual time series. More broadly, it demonstrates how scalable non-parametric SDE models with partial pooling can be implemented in Stan to study complex dynamical systems from sparse biological data.

When: 2026-08-20, 11:40 - 12:00, Where: Ångström Laboratory

An Introduction to Probabilistic Data Analysis with Stan on the Carpentries Platform

Ville Laitinen[1]
Leo Lahti[1]
[1] University of Turku

Abstract: We present an openly developed learning resource for applied Bayesian data analysis with Stan. The material is built on the The Carpentries platform, a community-driven initiative that creates openly available coding lessons based on evidence-based pedagogical principles.

The aim is to provide an approachable introduction to probabilistic data analysis and Stan for learners from diverse backgrounds. The emphasis is on practice-oriented model building rather than formal theory. Learners work directly with full probabilistic programs in Stan, encouraging engagement with model structure rather than reliance on higher-level interfaces.

The content is organized into modular "Episodes." After a brief introduction to core Bayesian ideas, subsequent episodes are largely self-contained and can be flexibly combined for workshops or longer courses. Interactivity and experimentation are central design principles. Participants are encouraged to modify models, test assumptions, explore prior choices, and observe how such changes affect inference. Theoretical details are included as optional side material, allowing motivated learners to dig deeper while keeping the main learning trajectory practical and accessible. In addition, references to widely used textbooks are provided for further study.

The material has been used for several years in a university course and continues to evolve based on systematic student feedback and classroom experience. Furthermore, additional episodes are continuously under development.

In this presentation, we will discuss the pedagogical design of the resource, lessons learned from teaching Stan to heterogeneous audiences, and how interactive, practice-first instruction can lower barriers to principled Bayesian modeling.

When: 2026-08-18, 11:20 - 11:40, Where: Ångström Laboratory

Bayesian Distributional Structural Equation Modeling in Stan

Luna Fazio[1]
Paul Bürkner[1]
[1] TU Dortmund University

Abstract: Accounting for the complexity of psychological theories requires methods that can predict not only changes in the means of latent variables -- such as personality factors, creativity, or intelligence -- but also changes in their variances. Similarily, the outcomes of behavioral experiments often need to be modeled by distributions with several parameters, each being informative about or even directly defining their own latent variables.

Structural equation modeling (SEM) is the framework of choice for estimating and analyzing complex relationships among latent variables. We have developed a Bayesian framework for Distributional SEM which broadens the scope of feasible models for distributional modeling in both the measurement and the latent model parts. We implement all our models in Stan and use statistical simulation to validate our framework across several distinct model structures. We demonstrate that reliable statistical inferences can be achieved and that computation can be performed with sufficient efficiency for practical everyday use. At the same time, we also find several interesting patterns and challenges that arise when estimating distributional SEMs, which we believe are relevant for a much wider class of models than those we investigated.

When: 2026-08-19, 11:40 - 12:00, Where: Ångström Laboratory

Blending ensemble dispersal modelling with dynamic stage-structured population models: A coherent framework for sea lice infestation dynamics

Tim M. Szewczyk[1]
Dmitry Aleynik[1], Andrew C. Dale[1], Philip A. Gillibrand[2], Kim S. Last[1], Helena C. Reinardy[1]
[1] The Scottish Association for Marine Science
[2] Mowi Scotland Ltd.

Abstract: Salmon lice (*Lepeotheirus salmonis*) are a chronic challenge for sustainable salmon farming in the North Atlantic. As external parasites of farmed and wild salmonids, their impact on fish health poses a welfare and economic threat to aquaculture industries and a conservation concern for wild fish populations. However, monitoring and predicting infestations is challenging. The microscopic planktonic larvae disperse for several days, driven by ocean currents interacting with complex, uncertain swimming behaviour. Larval concentrations are generally too low for traditional monitoring methods to be feasible, and detection of early attached stages on fish hosts is highly error prone.

To address these hurdles, we use Bayesian modelling to combine ensembles of particle-tracking dispersal models with dynamic stage-structured population models. In this framework, uncertainty in the larval spatiotemporal distribution is captured by a suite of particle-tracking models representing plausible physical and biological parameterizations given current knowledge. In Stan, a latent ensemble infestation pressure at each salmon farm is calculated as a weighted average, with host attachment rates estimated as a function of farm stocking density and hourly environmental conditions. A stage-structured population model then projects the daily mean number of lice per fish through time, with environmentally dependent survival and development rates, and anti-lice treatments implemented as increases in mortality. Finally, the latent abundance of each life stage serves as the mean for the weekly sample of counts of attached lice per fish performed by the industry, accounting for sampling effort and stage-specific detection probabilities. Prior distributions for environmental effects and demographic rates reflect the corresponding strength of empirical evidence.

This framework demonstrates how dispersal ensembles and hierarchical population models can be fused to coherently quantify uncertainty across physical, biological, and observational processes using Bayesian inference. By estimating infestation pressure and lice population trajectories jointly, the model enables robust inference on ecological processes and promises to improve predictive capabilities even when early-stage monitoring is sparse or unreliable. The approach is readily extensible to operational forecasting, evaluation of management strategies, and integration of additional data streams such as *in situ* larval sampling or advances in on-farm monitoring technologies.

When: 2026-08-18, 11:40 - 12:00, Where: Ångström Laboratory

Embedded Laplace Approximation in Stan

Charles Margossian[1]
Steve Bronder[2], Aki Vehtari[3], Brian Ward[2]
[1] University of British Columbia, Department of Statistics
[2] Flatiron Institute, Center for Computational Mathematics
[3] Aalto University, Department of Computer Science

Abstract: Stan's latest release provides an embedded Laplace approximation to fit latent Gaussian models, a popular class of hierarchical models with a multivariate normal prior. Examples include Gaussian processes, mixed effect models and regression with a horseshoe prior. It is well known that the posterior geometry of hierarchical models can frustrate most "off-the-shelf" inference algorithms. To address this challenge, the embedded Laplace approximation exploits the well-behaved geometry of conditional posteriors to approximately marginalize out the Gaussian latent variables. Inference can then be conducted on the marginal posterior of the hyperparameters, a distribution with an often well behaved geometry. While there already exist high-performance software with an embedded Laplace approximation, building this feature inside of Stan has a few benefits. First, Stan provides a great deal of flexibility when specifying the latent Gaussian model, including the prior on hyperparameters and the observational model. Secondly, Stan's MCMC can be used to approximate the marginal posterior, which is especially useful when the hyperparameters are high-dimensional. Furthermore, other gradient-based algorithms in the Stan ecosystem, such as pathfinder variational inference, can now be considered to do inference on the marginal posterior distribution.

When: 2026-08-19, 10:10 - 10:30, Where: Ångström Laboratory

Embedding geographic concepts of place in probabilistic modelling of temporal activity patterns using bayesian dirichlet regression

Jesse Piburn[1]
Sarah Walters[1], Marie Urban[1], Robert Stewart[1], **Dr. Pranav Sanke**[1]
[1] Oak Ridge National Laboratory

Abstract: Point of Interest (POI) datasets enriched with hourly temporal activity patterns—such as Google Popular Times—provide a granular view into the weekly rhythms of human behavior at individual locations, offering valuable insights into how places are used across spatial, cultural, and socioeconomic contexts. However, current analytical approaches are largely descriptive and lack the capacity to preserve variation, quantify uncertainty, or support statistical inference. To address this gap, we introduce a probabilistic framework based on Bayesian Dirichlet regression, modeling each POI's temporal activity pattern as a 168-dimensional probability vector—one value per hour of the week—drawn from a Dirichlet distribution. This framework includes a regression structure on both the mean vector $\boldsymbol{\mu}$, representing expected activity patterns, and the precision parameter Φ, capturing behavioral consistency across places. We also propose the use of the Continuous Ranked Probability Score (CRPS) to assess functional alignment by measuring how typical or atypical a place's behavior is relative to its assigned category. Grounded in human geography, we propose interpretive mappings between model parameters and conceptual constructs: activity ($\boldsymbol{\mu}$), affordance (Φ), and function (via CRPS). Through a case study using Google Popular Times data, we show how this framework enables generalizable, uncertainty-aware inference, supports domain-driven hypothesis testing, and offers predictive capabilities when extended with contextual covariates. This work contributes a statistically rigorous and interpretable approach to temporal modeling of place, advancing the methodological foundations of platial analysis and supporting practical applications in energy modeling, disaster response, and population dynamics.

When: 2026-08-18, 11:00 - 11:20, Where: Ångström Laboratory

From Mass Matrix Adaptation to Normalizing Flows

Adrian Seyboldt[1]
[1] pymc-labs

Abstract: Despite its robustness, Hamiltonian Monte Carlo (HMC) still struggles with strongly curved posterior distributions. These distributions lead to poor sampling performance or even the inability to sample, often indicated by poor effective sample sizes or divergences.

HMC adapts a mass matrix during the warmup phase, effectively applying a linear transformation to the posterior such that it becomes easier to sample. But since this transformation is linear, it can only correct global scaling or correlations, not locally varying curvature, such as in a funnel.

In Nutpie, mass matrix adaptation is formulated as minimizing a Fisher divergence objective, which has proven effective in practice. We generalize this idea from linear to nonlinear transformations. This yields an efficient inference method that combines Markov chain Monte Carlo (MCMC) and variational inference. A nonlinear transformation, such as a normalizing flow, is trained on recent MCMC draws and their score information, and progressively reshapes the posterior as seen by the sampler. As the learned transformation improves, the sampler becomes increasingly efficient.

I will also show how this approach works in practice in Nutpie, a sampling library that supports PyMC and Stan models. It makes posteriors tractable that cause standard HMC to break down, while keeping gradient costs practical.

When: 2026-08-19, 09:50 - 10:10, Where: Ångström Laboratory

Hierarchical Riemannian manifold Hamiltonian Monte Carlo algorithms

Jonas Wallin[1]
[1] Lund University

Abstract: Hamiltonian Monte Carlo (HMC) methods and their dynamic extensions such as the No-U-Turn Sampler (NUTS) are powerful Markov chain Monte Carlo techniques for sampling from complex probability distributions. The Riemannian manifold HMC (RMHMC) is an extension of HMC where the mass matrix depends on the position. This dependence can lead to great improvements in mixing, but its practical implementation is challenging. We present a hierarchical version of RMHMC, which is well-suited for many hierarchical sampling problems. Unlike the general RMHMC, it admits a closed form explicit leapfrog integrator, which is straightforward to implement efficiently. The method can be used within dynamic HMC algorithms such as the NUTS. We introduce an adaptive hierarchical RMHMC algorithm which adjusts the parameters of the hierarchical mass matrix automatically during simulation.

When: 2026-08-20, 10:10 - 10:30, Where: Ångström Laboratory

Incorporating Response Times into Item Response Models: A Case Study with Special Functions in Stan

John Ashley Burgoyne[1]
[1] Institute for Logic, Language, and Computation, University of Amsterdam

Abstract: Psychological measurement often involves estimating quantities based not only on participants' responses but also on how quickly they respond. An archetypical example is the old–new paradigm for short-term memory, in which participants view a set of stimuli and then judge a second set as old or new. Traditional models (e.g., signal detection theory) rely solely on accuracy, but response time also carries information: if Alex takes twice as long as Jordan to classify items, a lower memory score is warranted even if their accuracies are identical.

Item response theory (IRT), the mainstay of psychological measurement, has struggled to incorporate response times where simple sufficient statistics are desirable. Process models like Ratcliff's (1978) diffusion model do not lend themselves to sufficiency (Van der Maas et al. 2011), and models with clean sufficient statistics like the signed residual time (SRT) model do not always imply a satisfactory process (Maris and Van der Maas 2012). The standard advice is to model accuracy and response time separately (Van der Linden 2018).

Modern Bayesian methods offer a way forward. Psychometricians have long considered exponential families for response times (McGill and Gibbon 1965), and later work clarified their psychometric interpretation (Maris 1993). Both fall within the generalised inverse Gaussian family (GGIG; Jørgensen 1981), which offers plain response times, log response times, and response speed as (jointly) sufficient statistics, enabling straightforward combination with Rasch models from IRT.

Such Rasch–GGIG models are of interest to the Stan community because they highlight strengths and limitations of current Stan implementations. Both the GGIG and its more specific variants can be implemented as user-defined distributions in principle. In practice, the most general versions require numerically stable derivatives of incomplete gamma or modified Bessel functions. Stan has improved here, but cannot yet reliably sample these models in full generality.

This talk presents a case study using data from a musical memory game (Burgoyne et al. 2013), showcasing which Rasch–GGIG models do and do not sample well with current implementations. Successes offer practical tools for the Stan psychometrics community; failures identify targets for future development in the math libraries or with alternative samplers.

When: 2026-08-18, 13:00 - 13:20, Where: Ångström Laboratory

Intuitive R2 based prior specification for high-dimensional models

Javier Enrique Aguilar[1]
Paul-Christian Bürkner[1]
[1] TU Dortmund

Abstract: Many Bayesian regression analyses still rely on default coefficient priors that are difficult to interpret and can encode surprisingly strong beliefs about model fit, particularly when models are high-dimensional or hierarchical. R2-based priors offer an alternative. Instead of starting from priors on the coefficients, they start from an interpretable predictive quantity, the proportion of variance explained, and translate that into joint regularization of regression effects.

This talk surveys what "R2-based prior specification" means in practice, illustrating how the framework intuitively encodes structure, from hierarchical groupings to coefficient constraints, more effectively than standard priors. We will also discuss the practical implementation and specific challenges of these priors in Stan. Ultimately, we argue that this approach is able to provide a more transparent and robust standard for modern high-dimensional regression.

When: 2026-08-20, 13:20 - 13:40, Where: Ångström Laboratory

Latent variable estimation with composite Hilbert space Gaussian processes

Soham Mukherjee[1, 2]
Javier Enrique Aguilar[1], Marcello Zago[2, 3], Manfred Claassen[2, 3], Paul-Christian Bürkner[1]
[1] TU Dortmund
[2] University of Tübingen
[3] University Hospital Tübingen

Abstract: We develop a scalable class of models for latent variable estimation using composite Gaussian processes, with a focus on derivative Gaussian processes. We jointly model multiple data sources as outputs to improve the accuracy of latent variable inference under a single probabilistic framework. Similarly specified exact Gaussian processes scale poorly with large datasets. To overcome this, we extend the recently developed Hilbert space approximation methods for Gaussian processes to obtain a reduced-rank representation of the composite covariance function through its spectral decomposition. Specifically, we derive and analyze the spectral decomposition of derivative covariance functions and further study their properties theoretically. Using these spectral decompositions, our methods easily scale up to data scenarios involving thousands of samples. We validate our methods in terms of latent variable estimation accuracy, uncertainty calibration, and inference speed across diverse simulation scenarios. Finally, using a real world case study from single-cell biology, we demonstrate the potential of our models in estimating latent cellular ordering given gene expression levels, thus enhancing our understanding of the underlying biological process.

When: 2026-08-19, 13:40 - 14:00, Where: Ångström Laboratory

Luck, skill, and the battle of the sexes: analyzing tennis

Paul Goldsmith-Pinkham[1]
Tobias Moskowitz[1], **Nils Rudi**[1]
[1] Yale University

Abstract: Identifying skill from luck is a challenge facing most fields, yet it is critical for informed decision making. We decompose outcomes in tennis (match results, tournament results, ranking points, prize money) into luck versus skill using a structural, hierarchical model estimated with point-by-point data. This is done combining Bayesian inference in Stan with probability models. We identify three sources of luck: match luck, luck of the path a player faces given the draw, and the luck of the draw itself. We demonstrate how match luck hurts the more skilled players, but path luck helps them. We show that all three types of luck have a significant impact on a player's match, tournament, and career success. Comparing the men's and women's game, because women only play best out of 3 sets (while men play best out of 5) in grand slams, there is more match and path luck in the WTA. But women also show a wider range of skill across players, which leads to more dominant players at the top, partially mitigate luck. Our model matches well both realized outcomes and betting odds at both the match and tournament levels. Finally, the model allows us to conduct counter-factual analysis to assess what happens to the distribution of outcomes, and their relation to skill and luck, if rules changed (i.e., men play best out of 3 or women play best out of 5, no-ad scoring, one serve only).

When: 2026-08-18, 13:20 - 13:40, Where: Ångström Laboratory

Population Effect Estimation in Bayesian Hierarchical Models using Sum-to-Zero Constraints

Sean Pinkney[1]
Nikolai Vetr[2]
[1] Publicis Groupe
[2] Stanford University

Abstract: Relative effects in statistical models - including multilevel and hierarchical models, models with categorical groups, multinomial responses, and simplex-valued parameters in compositional data analysis - require an identification assumption. In practice, convenience or researcher preference often dictates an anchor chosen before model estimation such as fixing one category to zero, omitting a category (treatment coding), or imposing a sum-to-zero constraint. Each of these reduces the effective degrees of freedom by one and yields an identified parameterization.

In Bayesian hierarchical models, it is common to avoid explicit constraints on the group-level parameters. Instead, identification is achieved by positing a *population distribution* from which the hierarchical effects are drawn, treating the effects as random realizations from this hypothetical superpopulation. This overparameterized formulation is often viewed as incompatible with a sum-to-zero construction for multilevel parameters, leading to the perception that sum-to-zero constraints preclude estimation of the population distribution.

In this work we show that the standard Bayesian approach is overparameterized for the purpose of recovering the population effects: the same population distribution can be recovered from a sum-to-zero parameterization.

When: 2026-08-20, 13:00 - 13:20, Where: Ångström Laboratory

Probabilistic Matrix Completion for Chemical Engineering Applications

Zeno Romero[1]
Nicolas Hayer[1], Hans Hasse[1], Fabian Jirasek[1]
[1] Laboratory of Engineering Thermodynamics (LTD), RPTU Kaiserslautern, Germany

Abstract: Chemical engineering deals with the conversion of feedstocks into valuable end products, like polymers, fertilizers, and drugs, making it a pillar of many important industries, including the chemical, pharmaceutical, and biotechnological sectors. The successful design and optimization of these processes critically depend on reliable knowledge of the physicochemical properties of the involved mixtures.

Despite their importance, experimental property data are scarce, mainly due to the high cost and effort of laboratory measurements and the combinatorial diversity of relevant mixtures, which makes it impossible to study each mixture experimentally. Consequently, reliable prediction methods are indispensable [1].

Binary mixture properties are particularly important because they capture fundamental pair interactions that can be used to model multicomponent mixtures based on physical concepts. Data for binary mixtures can be organized in large matrices, where rows and columns represent the substances and entries contain the respective properties. Since only a small fraction of all combinations has been measured experimentally, these matrices are typically more than 90% empty.

In this work, we present probabilistic matrix completion methods (MCMs), implemented in Stan, to infer missing mixture-property data. We combine these MCMs with established physical models and apply them to various properties [2–8]. The results emphasize how probabilistic programming enables a principled integration of sparse experimental data with physical prior knowledge. Through case studies on gas solubilities [4,7], activity coefficients [2,3,6], and diffusion coefficients [5,8], we show that the resulting hybrid models can surpass established approaches while covering a much larger chemical space. Overall, the results demonstrate that probabilistic matrix completion supports transparent, extensible, and uncertainty-aware modeling without sacrificing physical insight.

References:

[1] Jirasek et al., Annu. Rev. Chem. Biomol. Eng., 2023, 14, 31-51.

[2] Jirasek et al., Chem. Commun., 2020, 56, 12407-12410.

[3] Damay et al., Ind. Eng. Chem. Res., 2021, 60, 14564-14578.

[4] Hayer et al., AIChE J., 2022, 68, e17753.

[5] Großmann et al., Digit. Discov., 2022, 1, 886-897.

[6] Jirasek et al., Chem. Sci., 2022, 13(17), 4854-4862.

[7] Hayer et al., J. Phys. Chem. B, 2025, 129, 409-416.

[8] Romero et al., J. Phys. Chem. B, 2025, 129, 9219-9228.

When: 2026-08-20, 11:00 - 11:20, Where: Ångström Laboratory

Probabilistic Ordinal Regression for Hierarchical Survey Data

Aleksi Lahtinen[1]
Leo Lahti[1]
[1] Department of Computing, University of Turku, Finland

Abstract: Despite computational advances in implementing Bayesian models, probabilistic methods remain underutilized in social science and humanities research. Survey and administrative register data are central data sources in social sciences especially, and a large proportion of key variables, such as attitudes, preferences, and self-reported behaviours, are measured on ordinal scales. At the same time, these datasets frequently exhibit hierarchical structures (e.g., individuals within countries), which require multilevel modelling approaches to properly account for between-cluster variation.

In recent years, tools such as Stan have made probabilistic ordinal regression models, such as cumulative regression, with hierarchical extensions, easier to implement. Nevertheless, adoption of these methods in applied social science and humanities remains limited. Researchers often rely on classical ordinal regression, potentially overlooking uncertainty quantification and modelling flexibility.

We analyse multiple survey datasets collected in several European countries and cities to study participation in live music events and cultural consumption. The surveys include ordinal outcomes such as attendance frequency, agreement levels, and engagement intensity, measured on different scales, providing a realistic setting for demonstrating probabilistic ordinal regression on hierarchical data.

Our analyses show that probabilistic ordinal models provide more robust estimation and more informative inference than classical approaches, particularly in small samples where sparse or empty response categories can cause instability in modelling. By incorporating prior distributions and partial pooling, these models regularize sparse data, induce shrinkage, and improve uncertainty quantification when the number of hierarchical clusters is limited. We illustrate implementations for different ordinal scales and hierarchical structures, discuss prior choice and diagnostics, and compare modelling strategies. Overall, the presentation demonstrates the practical advantages of a fully probabilistic framework for analysing ordinal survey data, highlighting gains in robustness, interpretability, and multilevel inference across diverse datasets and small-sample settings.

When: 2026-08-20, 09:50 - 10:10, Where: Ångström Laboratory

simplexmcmc: an auxiliary-variable framework for mixed continuous-discrete posterior distributions

jakob torgander[1]
måns magnusson[1], jonas wallin[2]
[1] uppsala university
[2] lund university

Abstract: *This talk introduces SimplexMCMC, a Markov chain Monte Carlo (MCMC) framework for sampling from mixed continuous-discrete posterior distributions. Discrete sampling is reformulated as sampling from the interior of the probability simplex, leading to an auxiliary-variable MCMC construction that admits the target discrete distribution as its stationary distribution. The framework is extended to Bayesian models with mixed discrete and continuous parameters and admits an HMC-based instantiation that makes use of the Gumbel–Softmax distribution. This instantiation, referred to as Relaxed Discrete Hamiltonian Monte Carlo (RDHMC), provides a practical algorithm for Bayesian inference in mixed continuous-discrete models. Correctness of the SimplexMCMC framework is established through theoretical convergence results, and the practical behaviour of RDHMC is assessed empirically on simulated and real data.*

When: 2026-08-18, 13:40 - 14:00, Where: Ångström Laboratory

StanBlocks.jl: a Julia-based frontend for writing and composing Stan models

Nikolas Siccha[1]
[1] Generable Inc.

Abstract: Writing Stan models by hand is expressive but demanding: block structure must be specified manually, types and shapes declared explicitly, and generated quantities written separately from the model itself. For users who also work in Julia, this friction is compounded by the need to context-switch out of the Julia ecosystem whenever a model needs to be modified or extended. These costs become particularly significant when building complex, bespoke models — such as population pharmacokinetic/pharmacodynamic (PKPD) models — where composability and reusability of model components are essential.

StanBlocks.jl is a Julia package that addresses this by providing a lightweight frontend for Stan model authorship, building on and inspired by Maria Gorinova's SlicStan. Its core is the `@slic` macro, which accepts a Julia-flavored model definition and automatically generates valid Stan code. Block assignment (`data`, `transformed data`, `parameters`, `transformed parameters`, `model`, `generated quantities`) is inferred from static analysis of the model, as are types, shapes, and parameter constraints. Pointwise log-likelihood and posterior predictive quantities are generated automatically. The resulting Stan code can be inspected, compiled, and sampled with CmdStan in the usual way: StanBlocks.jl is a frontend, not a replacement for Stan.

A key motivation for the Julia-based approach is access to language features that substantially raise the ceiling on what can be expressed cleanly. StanBlocks.jl supports higher-order user-defined functions, variadic arguments, custom struct-like types including named tuples, and — most importantly — reusable submodels that can be composed into larger models. Post-hoc model adjustment, such as swapping priors or replacing submodels entirely without touching the underlying definition, is also supported, as is automated generation of leave-one-unit-out cross-validation variants of existing models. Together, these features have been used in production at Generable to generate complex bespoke population PKPD models.

Looking ahead, potentially planned features include closures, keyword and default arguments, and Julia-style metaprogramming via macros — enabling syntactic sugar such as ergonomic runtime assertions that would be cumbersome to express directly in Stan.

StanBlocks.jl is open source and available at https://github.com/nsiccha/StanBlocks.jl.

When: 2026-08-19, 11:00 - 11:20, Where: Ångström Laboratory

To select or not to select - predictively consistent priors instead of model selection

Anna Elisabeth Riha[1]
Leevi Lindgren[2], David Kohns[1], Aki Vehtari[1]
[1] Department of Computer Science, Aalto University, Finland
[2] work done while at Department of Computer Science, Aalto University, Finland

Abstract: Bayesian modelling workflows often require the consideration of different candidate models. Approaches for model selection in the Bayesian framework aim to support the modeller in navigating potential trade-offs between model complexity and generalisability of the results to yet unobserved data. In this work, we propose a change of perspective towards choosing predictively consistent priors, instead of relying on model selection after the fact. We revisit the issue of overfitting, and clarify why model selection is not necessarily needed and can even be harmful in case of finite data. When integrating over the posterior and using predictively consistent priors, even if those priors can be considered weakly informative, we can often safely use flexible models with a large number of parameters. Using extensive numerical experiments, we illustrate the relevance of appropriate prior choices, as well as the limitations and alternatives for model selection in different modelling tasks, including forward variable selection and increasing the complexity of nonlinear models. We also investigate different implications for the predictive consistency of the priors depending on the chosen nonlinear model and provide detailed examples for adjusting R^2-based priors for logistic regression models and models that include a known relevant treatment covariate in the analysis of randomised controlled trials.

When: 2026-08-20, 11:20 - 11:40, Where: Ångström Laboratory

Understanding Stan so you can write JAX

Brian Ward[1]
[1] Flatiron Institute

Abstract: The Stan language provides a powerful abstraction for expressing models, but the Stan ecosystem can be too limiting when the need arises to use new algorithms or accelerated hardware such as GPUs. In this talk, I will cover the basics of how a Stan model works "under the hood", and illustrate how understanding this can empower users to translate their modelling ideas into JAX code -- first, at a low level, and later in a way that almost matches Stan's level of abstraction by using some carefully crafted helper functions from a forthcoming Python package.

When: 2026-08-19, 11:20 - 11:40, Where: Ångström Laboratory

variance deltas: interactive visualizations for explaining bayesian uncertainty

collin cademartori[1]
[1] wake forest university

Abstract: When posterior uncertainty for a quantity of interest is large, it is often useful to try to explain that uncertainty in terms of other unknown quantities. Uncertainty about a mean might be explained in terms of a poorly identified measurement bias parameter, and uncertainty about a causal effect might be explained in terms of a particular type of missing data, for instance. However, in complex models with many sources of variation and many unknowns, identifying useful explanations of uncertainty becomes increasingly difficult, and there is no broadly applicable workflow for extracting such explanations from a posterior distribution. To address this, we develop an interactive system for generating and visualizing explanatory sets of unknowns. This system, which we term variance deltas, organizes sets of potentially explanatory quantities into a tree structure rooted at the quantity of interest. Paths in this tree encode conditional independence relationships between model quantities, leveraging the model structure to constrain the search space of potential explanations. Variance deltas can be generated semi-automatically from a small number of inputs, and can be modified through a series of interactive operations which extend, subdivide and merge branches in the tree. We present a simple worked example which demonstrates the potential of this system to quickly identify useful, interpretable explanations of uncertainty (while also ruling out other, nonviable explanations). We also briefly discuss a small Stan-like language for expressing the factorization of a posterior distribution which we use for specifying some of the inputs needed to construct variance deltas, and which may be of separate interest for assisting in other tasks in Bayesian workflow.

When: 2026-08-19, 13:00 - 13:20, Where: Ångström Laboratory

Within-orbit adaptive leapfrog sampler (WALNUTS) with continuous Nutpie adaptation

Bob Carpenter[1]
[1] Center for Computational Mathematics, Flatiron Institute

Abstract: I will introduce three recent innovations for fitting Stan models.

First, I will outline the WALNUTS sampling algorithm, which generalizes Stan's no-U-turn sampler (NUTS) by adapting the step size each leapfrog step to ensure the Hamiltonian is conserved to within a threshold. Acceptance is modified to a Metropolis-within-Gibbs scheme following our new Gibbs self tuning (GIST) strategy. With step size tuning, WALNUTS is able to to effectively sample from stiff/multiscale distributions such as Neal's funnel, even with really bad initializations way out in the tail of the mouth or way down in the neck of the funnel; it is even faster for high-dimensional Gaussians when initialized at the mode (a poor, but common choice of initialization for HMC). It has higher effective sample size per iteration than NUTS for most distributions, but can be more costly in terms of gradients for problems that are already very efficiently sampled by NUTS.

Second, I will describe a generalization of the warmup and initialization scheme of the Nutpie sampler used by PyMC. Nutpie minimizes Fisher divergence (like KL, but for gradient norms) between a normal distribution and the mass-matrix preconditioned target density. The generalization adapts continuously by exponentially discounting the past according to a schedule that continuously mimic the discrete block size history of Stan's Phase II warmup (this is the stage where mass matrices are estimated). Initialization is generalized from a scale-sensitive unit mass matrix (as in Stan), to a scale-free estimate based on gradient outer products (as in L-BFGS). Further, I have upgraded the step size adaptation algorithm from dual averaging, which is used by both Stan and Nutpie, to Adam, which is both faster to converge and more stable in the sense of having lower variance across iterations.

Third, I will lay out a mechanism for lock-free monitoring and convergence detection for both warmup (by comparing individual chains to cross-chain averages) and sampling (via online R-hat).

I will demonstrate with our reference C++20 implementation (flatironinstitue/walnuts on GitHub), which provides a simple command-line interface (CLI) to fit Stan models through BridgeStan.

Poster

A Graph-Based Convergence Diagnostic for Multimodal Posteriors Using R-hat

Cici Chen Gu[1]
Sara Hamis[1], Eszter Lakatos[2]
[1] Department of Information Technology, Uppsala University, Uppsala, Sweden.
[2] Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden.

Abstract: Multimodality is prevalent in many scientific and engineering applications. However, convergence diagnostics for Markov chain Monte Carlo (MCMC) commonly penalise multimodality. To address this issue, we develop a graph-based diagnostic that computes R-hat separately for each pair of chains. The resulting graph is then used to identify groups of chains that appear to explore the same posterior region. We implement this method in the probabilistic programming framework Stan and provide worked examples to demonstrate its applicability. The proposed diagnostic identifies multimodality without automatically treating it as a sampling failure, providing an informative summary of MCMC behaviour for multimodal posteriors.

ArviZ: a modular and flexible library for exploratory analysis of Bayesian models

Oriol Abril-Pla[1]
Osvaldo A Martin[2], Jordan Deklerk[3], Seth D Axen[4], Colin Carroll[1], Ari Hartikainen[1], Aki Vehtari[2, 5]
[1] arviz-devs
[2] Aalto University, Espoo, Finland
[3] DICK's Sporting Goods, Coraopolis, Pennsylvania
[4] University of Tübingen
[5] ELLIS Institute Finland

Abstract: When working with Bayesian models, a range of related tasks must be addressed beyond inference itself. These include diagnosing the quality of Markov chain Monte Carlo (MCMC) samples, model criticism, model comparison, etc. We collectively refer to these activities as exploratory analysis of Bayesian models.
In this work, we present a redesigned version of ArviZ, a Python package for exploratory analysis of Bayesian models (EABM). The redesign emphasizes greater user control and modularity. This redesign delivers a more flexible and efficient toolkit for exploratory analysis of Bayesian models. With its renewed focus on modularity and usability, ArviZ is well-positioned to remain an essential tool for Bayesian modelers in both research and applied settings

A Bayesian state-space model for tool wear estimation in drilling of Inconel 718

Reinaldo Espina[1]
[1] Basque Center for Applied Mathematics (BCAM)

Abstract: Bayesian inference provides a strong framework for modelling complex industrial processes by combining prior knowledge with observed data, even when data and direct measurements of the process are limited. In industrial machining, tool condition monitoring is particularly challenging because the tool cannot be directly observed during operation. As a result, experimental tests are required to describe tool degradation and associate it with the corresponding sensor response. However, such experiments are often costly and therefore restricted to a limited number of tools and cutting conditions.

In this work, we study tool wear under different drilling conditions, using the spindle motor current as a sensing signal. The process dynamics were modelled using a nonlinear state-space model, where tool wear is treated as a latent and monotonically increasing state over time.

To address tool-to-tool variability, which is often neglected in related work, we propose two hierarchical Bayesian formulations using fixed and random effects that account for the differences across tools while preserving individual wear dynamics. All models are implemented and fitted using the Stan probabilistic programming language, allowing full Bayesian inference and rigorous uncertainty quantification for complex hierarchical non-linear models.

We conclude that hierarchical nonlinear state-space models fitted in Stan constitute an effective approach for capturing complex degradation processes while accounting for tool variability. The proposed formulations demonstrate how principled Bayesian modeling can address uncertainty, and data constraints in applied industrial settings.

A computationally-tractable measure of global sensitivity for Bayesian inference

Arina Odnoblyudova[1]
Charita Dellaporta[1], Francois-Xavier Briol[1]
[1] UCL

Abstract: Bayesian inference should ideally not be overly sensitive to the choice of prior or other modelling hyperparameters, such as the learning rate, but even defining and measuring this sensitivity is challenging. Existing global sensitivity measures typically involve significant trade-offs between strength of the measure, interpretability, and computational tractability. Unfortunately, most methods are unable to serve the needs of modern Bayesian inference due to their high computational cost and poor performance in multiple dimensions. To address these limitations, we introduce a new approach to global sensitivity analysis based on the Fisher divergence. We cast sensitivity analysis as an optimisation problem in which the objective is a Fisher divergence subject to structured neighbourhood constraints, such as box or quadratic constraints. Our method only requires a set of samples from a reference posterior and the ability to evaluate score functions, making it broadly computationally tractable. Under mild regularity conditions, it controls changes in the whole posterior, and provides a bound on the impact of perturbations on the first two moments. We demonstrate our proposed method on challenging Bayesian inference problems which are practically out of reach of existing approaches, including models with non-parametric priors, Bayesian inference for heavy-tailed time series, simulation-based inference for problems in telecommunications engineering, and generalised Bayesian inference for doubly-intractable models.

A hidden Markov model sheds light on problem-solving strategies in the common raven (Corvus corax)

Simon Steiger[1]
Katarzyna Bobrowicz[2]
[1] Karolinska Institutet
[2] University of Luxembourg

Abstract: Background

Recent research on problem solving has shifted from focusing on outcomes to analyzing processes through fine-grained interaction data, but this approach remains overlooked in non-human animals. We demonstrate its value by applying a hidden Markov model to study physical problem solving and social learning in common ravens.

Methods

Interactions of seven ravens (six female, six captive) with twenty different puzzle boxes, aimed at retrieving a reward from within the box, were recorded across two experimental conditions: without any training (solo test), or after observing another raven solving the problem (social test).
The ravens were tested again several weeks later (memory).
Using data on interactions with goal-relevant and irrelevant components of the puzzle box, we fit a multilevel discrete-time hierarchical hidden Markov (MHMM) model with two hidden states (trial-and-error, goal-oriented problem solving) and two absorbing states (success, failure).
We calculated the relative probability of solving the puzzle box by condition, with solo test as the reference.

Results

We analyzed 76 trials of which 22 were solo test trials, 24 were social test trials, 20 were solo memory trials, and 10 were social memory trials.
The median number of interactions per trial was 31 (minimum 3, maximum 143).
Ravens in the social test condition were 3.3 times more likely to succeed compared to ravens in the solo test condition (highest density interval, HDI: 1.2 to 7.0), with similar differences in memory trials.
Transition probabilities revealed that ravens in the social condition were more likely to switch to the goal-oriented state.

Conclusion

The MHMM revealed that social observation improves problem-solving by accelerating entry into a goal-oriented state – a procedural finding inaccessible to outcome-focused analyses.
Probabilistic programming languages like Stan are essential tools in enabling these bespoke models.

A workflow for evaluating data-source importance in integrated ecological models

Noa Kallioinen[1]
Jarno Vanhatalo[1, 2]
[1] Department of Organismal and Evolutionary Biology, Faculty of Biological and Environmental Sciences, University of Helsinki
[2] Department of Mathematics and Statistics, Faculty of Science, University of Helsinki

Abstract:

Integrating several distinct data sources into a single Bayesian model is becoming more commonplace in ecology. Frameworks such as integrated population models and integrated species distribution models make use of distinct sources of observations with the intention of improving the estimation of a shared underlying process. However, the inclusion of additional data sources increases model complexity and costs associated with data collection and preparation. Therefore, it is worth understanding the relative importance of data sources for estimating the quantities of interest and answering the research question. In this work we present a workflow for evaluating the importance of each data-source in an integrated model. The workflow combines efficient likelihood sensitivity checks and exact leave-data-source-out, coupled with model-appropriate cross-validation schemes. We present a case study of the workflow on an integrated population model of Baltic grey seals.

Analysis of Congo Basin rainforest regrowth trajectories by land use history

Andrés Martínez De Velasco[1]
Félicien Meunier[1, 2], Viktor Van de Velde[1], Sacha Delecluse[3], Pierre Defourny[3], Hans Verbeeck[1], Marijn Bauters[1]
[1] Ghent University
[2] Vrije Universiteit Brussel
[3] Université catholique de Louvain

Abstract: As the world's second largest rainforest, the Congo Basin rainforest plays a crucial role in the global carbon cycle. It is a vital resource for local livelihoods and regional climate regulation [1]. Increasing human disturbance to this rainforest due to demographic growth is amplifying uncertainties in the regional carbon balance, mainly due to limited understanding of forest regrowth trajectories. We aim to better understand regional regrowth trajectories following slash-and-burn agriculture (the region's dominant cultivation system) as a function of prior land use.

Here we present results related to the Bayesian modeling of secondary forest regrowth based on satellite remote sensing data using a space-for-time approach, where forest patches of different age are coupled with their above ground biomass estimate [2,3]. The coupled age/biomass data is grouped by land use history classes and used as input to fit local (1-degree grid cell) and regional (by floristic type) sigmoidal regrowth curves using the PyMC package [4] across the Congo Basin. We use informative priors for the model parameters based on field and satellite data, and implement the dependent variable (forest age) as a latent variable. We combine the posterior distributions into aggregated regrowth parameters (e.g. time to 90% recovery) at the 1-degree level and use them as targets for a RandomForest variable importance analysis with gridded bioclimactic variables and land use history as features. Ultimately, we aim to use such local regrowth curves to calibrate the Ecosystem Demography Biosphere model (version 2) to carry out mechanistic modeling of forest regrowth in the Congo Basin under different climate change and demographic growth scenarios.

References:
1: White et al., Nature, 2021
2: C. Vancutsem *et al.*, *Sci. Adv.***7**,eabe1603 (2021).
3: Wan L, *et al.*(2025) Front. Remote Sens. 6:1724950
4: *PyMC: A Modern and Comprehensive Probabilistic Programming Framework in Python*, Abril-Pla O, et al. (2023)

AutoStan: LLM agents for Bayesian modeling in Stan — a jagged intelligence that still needs supervision

Oliver Dürr[1, 2]
[1] Konstanz University of Applied Sciences, Germany
[2] Thurgau Institute for Digital Transformation (TIDIT), Switzerland

Abstract: Coding agents can now write and iteratively improve software toward an objective. The question is whether this extends to probabilistic programming in Stan. To this end, we created synthetic benchmark datasets and a simple agentic loop, AutoStan: the agent starts from a deliberately simple baseline, writes Stan code that is executed, and receives the negative log predictive density (NLPD), a strictly proper scoring rule on held-out data. The agent uses both to self-improve.

We run as many as 35 LLM models across datasets spanning robust regression, overdispersed counts, and spatial structure. We find that the agents can improve code in the majority of cases: the necessary Stan knowledge "is in the weights." They progress from the naive baseline to appropriate structure—Gaussian to negative binomial for counts, non-spatial to spatial—and tune and diagnose their own sampling. But because the loop optimizes against held-out NLPD, small datasets invite overfitting: on a demanding 1-D regression, the bias is significant with 30 test points and vanishes with 200. Further, recent and larger models are usually better. But even these can fail in surprising, subtle ways. Worse, some failures are silent: the program compiles and returns a plausible NLPD while being wrong, for example, a missing Jacobian or a hardcoded log_lik.

The proposed system works in principle, but sometimes imperfectly, requiring human oversight. Fortunately, its output is readable, inspectable Stan plus diagnostics, not a black box. The imperfection is manageable for the right user: a statistically educated practitioner who reads Stan but would rather not babysit samplers. This makes Stan an ideal language for communicating with the AI. In addition, it opens Bayesian modeling to those who know statistics but would rather not deal with the technicalities and subtleties involved in sampling. Code and datasets are public; we welcome harder problems.

baye-ites: validated per-patient treatment selection via Stan penalized hierarchical Bernoulli in a clinical trial setting studying acute stroke

Gabriele Gut[1]
Petr Semenov[1]
[1] Novartis, One Data & Digital, data42

Abstract: Individualized treatment effect simulation (ITES) requires not only accurate predictions but validated treatment selection, demonstrating that model-guided arm assignment improves real patient outcomes. We present baye-ites, a Stan-based penalized hierarchical Bernoulli-logit model with arm-specific covariate effects under partial pooling, designed to produce per-patient counterfactual survival probabilities with calibrated uncertainty. A key methodological innovation is the inter-arm penalty, which encourages similar but non-identical treatment-covariate interactions Using the International Stroke Trial (12,855 patients; aspirin, heparin, control; 6-month survival; patients dying of non-stroke related causes exclude), we sample via CmdStanPy and generate posterior predictive counterfactuals for 2,571 held-out patients on all three arms. Per-arm survival is summarized as the posterior predictive mean probability. Patients are classified as treatment-sensitive ("at-play") when arm-specific probabilities disagree across the decision boundary. Our model achieves Brier Skill Score +0.165 versus −0.083 for a doubly-robust causal forest baseline (EconML CausalForestDML), demonstrating superior calibration. It also detects 40% more mortality events (recall 23% vs 16% at matched precision ~60%, death-class F1 0.332 vs 0.256), which is the decisive comparison given the 88% base survival rate. The model stratifies patients into Global-Responders (94%, survival 91%), At-Play (3.3%, survival 54%), and Global-Non-Responders (2.9%, survival 40%), representing a 51% separation concentrating actionable uncertainty into a small stratum. Among 85 At-Play patients, those randomized to predicted-good arm survived at 67% versus 40% on predicted-bad arm (OR=3.05, p=0.017, NNT=8), consistent across all arms. Feature analysis reveals coherent profiles: aspirin benefits patients with atrial fibrillation and CT-confirmed infarct, a signal that replicates across the At-Play subgroup (p=0.048 and p=0.032) and the full Aspirin-arm population (p<0.001 visible infarct). Severe hemispheric strokes (total anterior circulation, impaired consciousness) are harmed by active treatment and heparin response has no simple feature combination achieving discriminative power. The causal forest cannot perform per-patient arm selection, as it lacks counterfactual arm-specific probabilities.

Bayesian Aggregation Methods for Federated Brain Tumor Segmentation

Khaoula El Mekkaoui[1]
Muhammad Irfan Khan[2], Suleiman A. Khan[2], Samuel Kaski[1]
[1] Aalto University
[2] Turku University of Applied Sciences

Abstract: Federated learning (FL) enables privacy-preserving collaborative training across medical institutions, yet mostly standard aggregation strategies such as SimAgg are deterministic, treating client updates as point estimates and not quantifying uncertainty in parameters associated with institutional heterogeneity. In medical imaging tasks where data distributions vary substantially across sites, explicitly modeling uncertainty through stochastic methods may help support medical practitioner confidence in decision-making.

We propose a Bayesian aggregation methodology for federated brain tumor segmentation. We evaluate simple Bayesian inference using Hamiltonian Monte Carlo to model posterior uncertainty over global parameters and introduce hierarchical Bayesian aggregation incorporating institution-level random effects to capture cross-site variability. These aggregation schemes are combined with graph-based client selection strategies (Dijkstra and Viterbi) to further improve training efficiency and stability.

We evaluate our approach on the FeTS Challenge dataset through 90 controlled experiments (10 runs per configuration). Across all clinically relevant tumor subregions, our Bayesian methods consistently outperform SimAgg in per-label DICE score. Hierarchical and simple Bayesian aggregation improve robustness to institutional heterogeneity, yielding gains particularly in necrotic core and edema segmentation. Bayesian aggregation strategies also provide more balanced improvements across labels compared to SimAgg.

Our findings suggest that Bayesian aggregation better captures institutional variability in federated medical imaging and that convergence-based evaluation metrics alone may obscure clinically meaningful improvements. Although SimAgg attains a higher projected convergence score, it yields lower per-label DICE performance

Bayesian constructive goodness-of-fit for static distributions and time series data

Søren Vad Iversen[1]
Michael Lomholt[1]
[1] University of Southern Denmark

Abstract: Model selection is a cornerstone of statistical inference, and any candidate model must be subjected to rigorous goodness-of-fit testing to assess its predictive validity. While the Bayesian framework has emerged as a compelling alternative to classical frequentist methods, a practically useful Bayesian goodness-of-fit test remains an largely open problem [1] — leaving the Bayesian framework incomplete as an alternative to frequentistic method.

We present a computationally efficient method for Bayesian goodness-of-fit testing applicable to both static distributions and time series. Inspired by [2], our approach constructs an extended hypermodel to test the candidate model against, not only allowing us to evaluate whether a given model fits the data, but also — in the case of a poor fit — provides interpretable insight into how the model may be improved. Crucially, the method requires no Monte Carlo sampling, making it significantly faster than existing approaches and practical for routine use.

References:

[1] P. Diaconis and G. Wang. Bayesian goodness of fit tests: a conversation for david mumford. Ann. Math. Sci. Appl., 3:287, 2018.

[2] I. Verdinelli and L. Wasserman. Bayesian goodness-of-fit testing using infinite-dimensional exponential families. Ann. Statist., 26(4):1215–1241, 08 1998

Bayesian models of Li ion battery aging

Martin Lindén[1]
Katarzyna Paczos[1, 2]
[1] Traton Group R&D
[2] Stockholm University, MSc Programme in Statistics

Abstract: Traton Group, with its brands Scania, MAN, International, and Volkswagen Truck & Bus, is one of the world's leading manufacturers of commercial heavy duty vehicles, with battery-electric vehicles at the core of its decarbonization strategy. The battery constitutes a substantial part of the total cost of ownership for a battery electric vehicle. Accurate forecasts of battery aging is therefore important for customer profitability and risk management, and an input to business decisions throughout the product life cycle.

Battery aging is a complex process that cannot be predicted quantitatively from first principles. Instead, we rely on a mix of domain knowledge, lab testing, and data driven modelling techniques to produce lifetime forecasts that support various business- and operational decisions. In this poster, we introduce the battery aging problem, showcase some common non-linear regression and dynamical battery aging models using publicly available data, and present results from a thesis project at Traton Group R&D that apply Bayesian methods in Stan to model and predict battery aging.

bayesian MRP for modeling attitudes toward immigration and religious outgroups in france

Kota Mori[1]
[1] University of Paris 8, CRESPPA-GTM

Abstract: This study examines negative attitudes towards Muslims and immigrants in France using data from European Values Study (EVS, 2018), drawing on Intergroup Threat Theory (ITT) as its theoretical basis. Its primary contribution lies in the application of Multilevel Regression with Poststratification (MRP) in a Bayesian framework implemented via Stan, which is still relatively underused in social sciences.

Two binary outcomes are modeled: rejection of Muslim and rejection of immigrants/people of different race as neighbors. For each, a sequence of four nested regression models is estimated, progressively adding residency factor, unemployment status, two standardized threat scales (realistic/symbolic), and political partisanship, controlling for religious and ethnic origin. Partial pooling is achieved through hierarchically structured random effects for four demographic variables (age, sex, education, profession).

The poststratification grid is constructed from census-based population counts provided by the French National Institute of Statistics and Economic Studies (INSEE). For each posterior draw, posterior predictive probabilities are computed for all poststratification cells and aggregated using population cell counts from the census. This process yields a posterior distribution of the population prevalence (φ), with proper uncertainty propagation. This makes Stan a natural fit for MRP workflows, where population-level quantities can be derived alongside individual-level parameters within a single coherent inferential framework, without any post-hoc correction.

Model comparison was executed via LOO-CV, and prior sensitivity was assessed with four prior specifications (Student-t with ν=7, ν=3, Cauchy, and Normal), following recommendations of existing literature. Results are substantively consistent across all specifications, assuring robust inference. The findings of the analyses align with ITT literature, yet suggesting different mechanism shaping attitudes towards Muslims and Immigrants in France.

This poster also presents a simple way to implement MRP within a Bayesian workflow that effectively addresses challenges in survey-based research common in social sciences: non-representative samples, small-cell sparsity, and quantification of uncertainty.

Bayesian multivariate hierarchical modelling of vegetation change in boreal and subarctic reindeer ranges

Henri Wallen[1]
Sari Stark[1], Mika Kurkilahti[2], Antti-Juhani Pekkarinen[2], Jouko Kumpula[2]
[1] University of Lapland
[2] Natural Resources Institute Finland

Abstract: Disentangling the joint effects of climate, herbivory, and plant competition on vegetation change requires models that simultaneously handle correlated responses, non-linear effects, and hierarchical spatial structure. We present a Bayesian multivariate hierarchical model, fitted with brms and Stan, applied to a decade-apart repeated inventory of dwarf shrub and lichen cover and height across 617 field sites in northernmost Finland. Four vegetation response variables were modelled jointly using Student-t likelihoods to account for heavy-tailed residual variation. Non-linear effects of reindeer grazing intensity were represented using seasonally varying penalised thin-plate splines, while climatic covariates were included in the model as linear terms. Spatially nested variation was captured through site-level random effects, specified in a multivariate structure to estimate covariance among responses. The multivariate framework captures both the modelled effects of plant–plant interactions and the residual covariation among vegetation components at the site level, allowing competitive and facilitative relationships to emerge from the correlation structure without requiring them to be fully specified a priori.

Bayesian spatial modelling of regional spread of Omicron wave using the colleague connectivity matrix across the Netherlands

PingPing Song[1]
Sake de Vlas[1], Luc Coffeng[1]
[1] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

Abstract: Infectious diseases do not unfold evenly across regions, as the type and frequency of inter-regional contacts through which infections spread vary spatially. Workplace connections are essential to such spatial heterogeneity, as colleague connections are large in volume and wide in spatial reach. However, the role of population connectivity in regional epidemic spread is not well quantified. We therefore aimed to estimate how geographical colleague connectivity derived from the Dutch working population (8 million) contributed to the spread of Omicron waves in the Netherlands.

To this end, our analysis proceeded in three steps. First, we fitted a logistic model to regional SARS-CoV-2 sequencing data to estimate the proportion and number of Omicron cases per province. Second, we built five BYM2 (a reparameterization of the Besag-York-Mollié model) spatial models with increasing complexity, using log-transformed and normalised colleague connections per residential province pair as the adjacency weight. These models combined a spatially structured intrinsic conditional auto-regressive (ICAR) component and an unstructured province-level random intercept. The most parsimonious model was selected based on approximate leave-one-out cross-validation and included province-specific growth rate deviations and a quadratic time trend to account for sub-exponential growth. Finally, we fitted this model to 100 randomly sampled posterior draws of provincial Omicron incident cases from the first step using a negative binomial likelihood.

With a scaling factor of 0.18, colleague connectivity spatial structure explains 41.7% of the variation in province-level Omicron incidence at wave onset (95% credible interval 6.9% to 85.7%). By encoding the connectivity matrix into the spatial structure of a Bayesian growth model, our findings suggest that colleague connectivity plays a moderate and nonnegligible role in shaping early regional epidemic risks. Our findings support using population connectivity for early signalling and prioritising regions with strong workplace ties to already affected regions at the onset of future epidemic waves.

Beyond Point Predictions: Bayesian Deep Learning for Global Scale Building Attribution

Pranav Sankhe[1]
Jesse Piburn[1], Michael Lindgren[1], Robert Stewart[1]
[1] Oak Ridge National Laboratory

Abstract: Predictions used in high stakes settings must be accompanied by well calibrated measures of uncertainty. In domains such as disaster risk modeling, climate adaptation, insurance underwriting, and urban planning, point estimates alone are insufficient for strategic decision making; confidence intervals aids practitioners assess risk, allocate resources, and quantify tail exposure. Bayesian approaches are a natural fit for such problems, offering principled probabilistic model specification, hierarchical structure, and coherent uncertainty propagation. However, when the inferential target involves tens of millions or even billions of buildings that are locally spatially correlated, fully specified hierarchical Bayesian models become computationally prohibitive.

We present a scalable deep learning framework for global building attribute imputation that preserves key Bayesian ideas while relaxing the requirement of full posterior sampling. Global building datasets are incomplete, heterogeneous, and spatially imbalanced, with complex dependence structures across geography. Our approach leverages a graph neural network to encode spatial relationships among neighboring buildings and employs Monte Carlo (MC) Dropout as a variational Bayesian approximation to capture predictive uncertainty. Interpreting dropout as approximate variational inference in neural network architecture provides a conceptual bridge between Bayesian non-parametrics and modern deep learning. By maintaining dropout at inference time and drawing multiple stochastic forward passes, we approximate the posterior predictive distribution with relatively negligible additional computational cost.

This strategy can be viewed as a pragmatic alternative to fully Bayesian neural networks or large scale spatial hierarchical models fit via HMC. While we sacrifice asymptotically exact posterior inference, we gain orders of magnitude improvements in scalability and run time efficiency enabling planetary scale deployment.

Beyond the voltmeter: recovering electrochemical parameters of Na-Ion batteries through Bayesian inversion

Gabriela Bergiel[1]
Mateusz Daniołł[1], Ryszard Sroka[1]
[1] AGH University of Krakow

Abstract: The battery technology rapidly evolves and sodium-ion cells emerge as a promising alternative for energy storage in low and extreme-temperature environments. The need for well-parameterised physics-based models becomes critical for technology selection and operational safety. However, identifying physically meaningful parameters such as diffusion coefficients, exchange current densities, or ohmic resistances without disassembly requires inferring them solely from external voltage-current measurements. This non-invasive identification is inherently ill-posed: multiple parameter combinations may reproduce observed voltage trajectories equally well. Also available Na-Ion datasets remain scarce. Bayesian inference provides a natural framework to address these challenges. It allowes for encoding physical knowledge as priors, quantifying uncertainty in parameter estimates, and enables principled inference even under data scarcity.We present a Bayesian hierarchical calibration framework for the Single Particle Model (SPM), a minimal ODE-based representation of a sodium-ion (Na-Ion) battery. Cell parameters are treated as draws from a population-level distribution, enabling partial pooling across cells, principled quantification of cell-to-cell variability, and improved inference under data scarcity. Crucially, transport parameters such as the solid-phase diffusion coefficient D_s are modelled as temperature-dependent through Arrhenius kinetics. At the current stage, D_s at reference temperature is estimated from data while activation energy Ea is informed by literature priors. Full temperature-dependent inference is awaiting a dedicated low-temperature dataset.The framework is implemented in Stan and validated on publicly available Na-Ion cycling data. In parallel, we are constructing a dedicated experimental dataset spanning temperatures down to −40°C. The test bench includes thermal chamber with IoT monitoring and metrologically characterised battery tester with dedicated cell holders. This dataset will enable full Arrhenius-informed inference across the temperature dimension.We discuss posterior geometry, identifiability diagnostics, shrinkage behaviour, and the role of physics-informed priors in this hierarchical inverse problem.

Bringing negative binomial models into exact projection predictive inference

Zhi Ling[1, 2]
Swapnil Mishra[1, 2]
[1] Saw Swee Hock School of Public Health, National University of Singapore
[2] Department of Statistics and Data Science, National University of Singapore

Abstract: Projection predictive inference is a Bayesian framework for simplifying complex models while preserving predictive performance. It has become an important tool for variable selection and model compression. For Gaussian, binomial, and Poisson responses, projection can be carried out directly and integrated into efficient model-search procedures. However, no native projector has been available for the negative binomial, one of the most widely used likelihoods for count data. The difficulty is that its projection objective involves an infinite series with no tractable closed form.

We show that, for a broad class of discrete distributions with analytically available probability generating functions (pgfs), the Kullback–Leibler (KL) divergence admits an integral representation in terms of the pgf and a kernel. This representation leads to fast and accurate quadrature approximations of the KL divergence and of its gradients and Hessians. The numerical error is controlled by the chosen quadrature rule. The class covers many commonly used families, notably discrete natural exponential families and generalized hypergeometric probability families. This reformulation replaces the original infinite series subject to uncertain truncation errors, by a one-dimensional numerical integration problem requiring only a fixed number of pgf evaluations to attain high accuracy in practice.

For negative binomial models, this gives an exact representation of the observation-space projection objective and enables efficient derivative-based optimization by quadrature. We implement the resulting algorithms in the projpred package in the Stan ecosystem. Because negative binomial likelihoods are central in applications ranging from epidemiology to genomics, this development opens the door to Bayesian variable selection and model compression for overdispersed responses. Overall, we provide a general route for extending projection predictive workflows to a wider range of discrete responses with infinite support.

bscm: an R package for Bayesian synthetic control models for panel data causal inference

Jouni Helske[1]
[1] University of Turku

Abstract: Synthetic control methods (SCM) are widely used for evaluating the causal effects of policy reforms and other interventions in panel data settings. An appealing feature of SCM is the simplicity of its core idea: counterfactual outcomes for a treated unit are constructed as a convex combination of untreated units. However, standard SCM provides only point estimates of treatment effects and offers limited tools for principled uncertainty quantification. I recast SCM within a fully Bayesian framework, introducing a Bayesian synthetic control model (BSCM) that enables coherent uncertainty quantification for treatment effect estimates, donor weights, and other quantities of interest. The model naturally accommodates time-varying covariates and multiple treated units with staggered adoption. The approach is implemented in the Stan-based bscm R package, which provides methods for estimation, diagnostics, visualization and model comparisons using variants of leave-one-out cross-validation. Through simulation studies, I compare the Bayesian approach with existing SCM variants and show improved predictive performance for treatment effect estimation. Finally, I demonstrate the method in an application estimating the mental health effects of Russia's 2022 invasion of Ukraine on Ukrainian immigrants in Finland.

building a bayesian model for investigating role-dependence of internal references in cross-modal magnitude productions

katharina naumann[1]
[1] university of tübingen

Abstract: This psychophysical experiment investigates the cross-modal correspondence of brightness and loudness in healthy human subjects utilizing a magnitude production paradigm: Subjects are presented with two stimuli, a light and a sound, and adjust the intensity of one of them such that its perceived intensity stands in a certain given ratio to the other stimulus. For example, the subject adjusts the luminance of the light in a way that its brightness is double the loudness of the sound. This is done for different stimulus levels of luminance and sound pressure, and different production ratios.
We examine two behavioral phenomena, the near-miss to commutativity and the regression effect in cross-modal matching, that can be explained by the Global Psychophysical Theory with role-dependent internal references. The theory is based on behavioral axiomatic invariances that imply a ratio scale of perceived intensity. Extending past efforts in the field which focused on the statistical test of behavioral axioms, we are taking the deterministic theory into a probabilistic framework. For that, we are implementing a Bayesian estimation of the global psychophysical model in Stan. Prior distributions of model parameters are informed by previous results in the field, and checked for the psychophysical plausibility of their predictions. The Bayesian model is developed and validated with simulation based calibration and sensitivity analyses. The model's development process and its posterior predictive ability are discussed. Role-dependence of the internal references is tested by model comparison via Bayes Factors.

Composable MCMC kernels for automated Metropolis-within-Gibbs inference algorithms

Alin Morariu[1]
Chris Jewell[1], Jess Bridgen[1]
[1] Lancaster University

Abstract: State-transition models are central to applications in epidemiology and ecology, yet statistical inference remains challenging due to high-dimensional latent state spaces, strong temporal dependence, and intractable likelihoods. Bayesian inference using Markov Chain Monte Carlo (MCMC) enables joint estimation of model parameters and unobserved event times via data augmentation. However, Metropolis-within-Gibbs (MWG) schemes that combine multiple specialized kernels are notoriously difficult to implement. Existing probabilistic programming frameworks face a fundamental trade-off: automation limits extensibility, while flexibility requires substantial implementation effort. As a result, the software ecosystem is dominated by tightly coupled, model-specific implementations that hinder reuse and extension.

We propose a composable MCMC constructor for probabilistic programming that formalizes kernel composition using a functional approach, with monadic and functorial constructs from category theory. From a programming perspective, this enables modular construction of inference algorithms. It facilitates the combination of continuous-domain parameter samplers with discrete, state-space-oriented data augmentation kernels, without manual management of shared state - a limiting factor to using MCMC in epidemic modelling. The framework is programming-language agnostic, providing a principled blueprint for implementation across probabilistic programming platforms.

We demonstrate the approach through parameter inference for partially observed epidemic models, showing how complex MWG algorithms can be expressed concisely within a probabilistic programming workflow and reused across applications. By reducing implementation effort and improving composability, our framework lowers barriers to advanced MCMC methods and strengthens the extensibility of probabilistic programming systems.

Contributing to Stan: A Developer's Guide to the Core, Interfaces, and First Contributions

Aki Vehtari[1]
Brian Ward[2], Florence Bockting, Steve Bronder
[1] ELLIS Institute Finland, Aalto University
[2] Flatiron Institute

Abstract: Would you like to contribute a new feature to Stan, or make your first contribution to the Stan ecosystem? This tutorial provides an overview of Stan from a developer's perspective, covering the Stan C++ and OCaml core, the R packages, and how the different parts of the ecosystem fit together. We will also discuss the Stan development process, including how contributions are proposed, reviewed, and merged.

The tutorial is designed to help participants take their first concrete steps toward contributing. Using a selection of good first issues as examples, we will show how to get started, navigate the codebase and development workflow, and work effectively with the Stan developer community. By the end of the tutorial, participants will have a clearer understanding of where and how they can contribute, and they will be well-positioned to continue working on a contribution after the session with support from Stan developers.

Correcting measurement error in factor score regression: a joint Bayesian implementation in Stan

Carolina Alvarez-Garavito[1, 2]
An Hotterbeekx[3], Lorenzo Maria Canziani[4], Lenny Coppers[3], Roy Gusinow[1, 2], Manuel Huth[1, 2], Gaia Maccarrone[4], Anna Górska[5], Elisa Gentilotti[4], Elisa Rossi[6], Salvatore Cataudella[6], Evelina Tacconelli[4], Samir Kumar-Singh[3], Jan Hasenauer[1, 2]
[1] Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany
[2] Bonn Center for Mathematical Life Sciences, University of Bonn, Bonn, Germany
[3] Molecular Pathology Group, Laboratory of Cell Biology Histology (CBH) and Vaccine Infectious Disease Institute (CBH), Faculty of Medicine, University of Antwerp, Antwerpen, Belgium
[4] Division of Infectious Diseases, Department of Diagnostics and Public Health, University of Verona, Verona, Italy
[5] Department of Biomedical Engineering, Wroclaw University of Science and Technology, Wroclaw, Poland
[6] CINECA Interuniversity Consortium, Bologna, Italy

Abstract: In applied biomarker research, where data are typically high-dimensional and strongly correlated, dimensionality reduction is often used to derive latent co-signaling clusters, whose estimated per-sample scores are then entered as covariates in regression models. Although standard practice, this two-stage procedure ignores measurement error in estimated scores, attenuating regression coefficients and underestimating posterior uncertainty. Here, we apply factor analysis to derive latent factor modules and estimate their association with clinical phenotypes using a Bayesian framework for covariate measurement-error correction. Specifically, we consider: (1) a structural model linking latent factor scores and clinical covariates to a binary outcome via logistic regression; (2) a measurement model in which observed measurements are linear in the latent factors, with loadings fixed at a maximum-likelihood factor solution; and (3) a conditioning model placing a covariate-informed multivariate normal prior on the latent scores. All components are estimated jointly via HMC. Latent factors are reparameterized non-centrally, with the inter-factor covariance represented through its Cholesky factor. The non-centering decouples the latent variables from their scale parameters, avoiding the funnel geometry that degrades HMC mixing in hierarchical models. The correlation factor receives an LKJ(2) prior, and regression coefficients weakly-informative Normal priors. We applied this framework to the ORCHESTRA prospective multinational study to examine the association of Long COVID phenotypes and 38 serum cytokine measurements across four follow-up periods. Comparing against a naive plug-in regression on estimated factor scores, our approach recovered phenotype associations that the naive method missed due to coefficient attenuation, with no cases in the reverse direction. The pattern is consistent with measurement-error theory and demonstrates the practical value of joint estimation in high-dimensional biomarker studies.

Detecting spatial and temporal changes in Dungeness crab abundance across the Salish Sea

Deirdre Loughnan[1, 2]
Heather Earle[2], Lauren Krzus[2], Alyssa-Lois M. Gehman[1, 2], Mary I. O'Connor[1]
[1] University of British Columbia
[2] Hakai Institute

Abstract: Climate change is increasing ocean temperatures and leading to novel stressors in marine ecosystems, which can alter species abundances, spatial distributions and shifting the timing of life events. The effects of these changes are particularly concerning for species of ecological and economic importance, such as Dungeness crab. Young Dungeness crabs are influenced by ocean conditions, but it remains unclear which environmental factors determine the timing of their arrival or spatial distributions. But the outcome of these processes is critical to sustaining crab populations and to predicting the abundance of adult crabs for local fisheries. Crab larvae, however, can travel long distances on ocean currents, making it challenging to model how spatially variable environmental conditions across regions contribute to local populations. To better understand what environmental factors are influencing the timing of young crab arrival and abundances, we combined environmental data with count data of larval crab from a network of 29 sites across the coast of British Columbia, Canada. At each site, data was collected every two days during the summer season for four years. Using Bayesian mixture models, we tested whether crab larvae are likely to originate from multiple cohorts and identified the key environmental factors that determine when larvae first arrive and peak in their abundances at each site. Our findings provide novel insights into the factors that shape dispersal and recruitment of Dungeness crab in the Canadian Pacific Northwest, and contribute to the sustainable management of this important marine species.

Diffusion resampling: differentiable resampling in sequential Monte Carlo

Jennifer Andersson[1]
Zheng Zhao[2]
[1] Uppsala university
[2] Linköping university

Abstract: Resampling is an essential component in sequential Monte Carlo (SMC) methods since it mitigates the weight degeneracy problem. In recent years, differentiable particle filtering, which involves learning components of particle filters and parameters in state-space models (SSMs), has become an active research area. In the context of gradient-based learning, resampling poses a challenge due to the discrete randomness inherent in classical resampling methods. In this talk, we introduce *diffusion resampling*; a new differentiable resampling method based on an ensemble score diffusion model. The new diffusion-based resampling method is not only differentiable-by-construction, but also computationally efficient and informative, and thus immediately compatible with end-to-end learning in differentiable particle filters. In our work, we prove that diffusion resampling provides a consistent resampling distribution. We showcase our method on several filtering and parameter estimation problems in SSMs, and verify empirically that diffusion resampling outperforms state-of-the-art differentiable resampling methods on these problems. In addition, we show that it achieves competitive end-to-end performance when used in a more complex learning setting with high-dimensional image observations.

DiPPER: A Bayesian approach to differential prevalence analysis with applications in microbiome studies

Juho Pelto[1, 2]
Kari Auranen[1, 3], Janne Kujala[1], Leo Lahti[2]
[1] Department of Mathematics and Statistics, University of Turku, Turku, Finland
[2] Department of Computing, University of Turku, Turku, Finland
[3] Department of Clinical Medicine, University of Turku, Turku, Finland

Abstract: Recent evidence suggests that analyzing the presence/absence of microbial features (e.g., species) can offer a compelling alternative to analyzing complete abundance profiles in microbiome studies. Specifically, differential prevalence analysis, which can be used to detect disease-associated microbial features, has been proposed as a robust and easily interpretable alternative to traditional differential abundance analysis.

Standard approaches to differential prevalence analysis, however, face challenges with boundary cases and multiple testing. To address these limitations, we developed DiPPER (Differential Prevalence via Probabilistic Estimation in R), a novel method based on Bayesian hierarchical modeling. The core idea of the method is that differential prevalence estimates are assumed to arise from a common asymmetric Laplace distribution, whose variance and skewness are informed by all features collectively.

We benchmarked DiPPER against existing differential prevalence and abundance methods using data from 57 human gut microbiome studies. Our results demonstrate that DiPPER outperforms alternative methods by combining high sensitivity with robust error control. Furthermore, by leveraging hierarchical shrinkage, DiPPER provides finite differential prevalence estimates and uncertainty intervals that are inherently adjusted for multiple testing.

DiPPER is currently implemented in Stan using a standard approach relying on conventional parameterization and default NUTS sampling. However, we are actively exploring ways to accelerate the method by optimizing the model parameterization and posterior sampling.

A pre-print of the paper introducing DiPPER is available at https://arxiv.org/abs/2602.05938.

Don't disregard the data for lack of a likelihood: Bayesian synthetic likelihood for enhanced multilevel network meta-regression

Harlan Campbell[1, 2]
Charles Margossian[1], Jeroen Jansen[2, 3], Paul Gustafson[1]
[1] Department of statistics, University of British Columbia
[2] Precision AQ
[3] Department of clinical pharmacy, University of California, San Francisco

Abstract: Multilevel network meta-regression (ML-NMR) enables population-adjusted indirect treatment comparisons by combining individual participant-level data (IPD) with aggregate-level data. When individual-level covariates are unavailable, ML-NMR marginalizes over the covariate distribution, but this strategy cannot exploit subgroup-level summary results that are often available and potentially highly informative. We propose using Bayesian Synthetic Likelihood (BSL) to leverage this ancillary summary information with implementation in Stan. At each MCMC iteration, the method imputes missing covariates by sampling from the model-implied conditional distribution, computes synthetic subgroup summaries from the imputed data and matches them to observed summaries via a multivariate normal synthetic likelihood. A key implementation challenge is that generating synthetic datasets within Stan must be done in a way that does not distort the likelihood, and sampling from discrete distributions causes non-differentiability that breaks gradient-based sampling. We address these issues by imputing missing values in the transformed parameters block using pre-drawn random numbers and continuous relaxation to approximate discrete elements. We further implement an importance sampling correction using Pareto-smoothed importance sampling (PSIS) to diagnose and adjust for the approximation error introduced. This work (1) demonstrates how BSL strategies can be efficiently implemented in Stan, (2) introduces a novel application of BSL to missing data problems where summary statistics from the complete dataset are available despite substantial missingness in the individual-level data, and (3) shows that using BSL can substantially improve upon standard ML-NMR by leveraging informative ancillary information.

Estimating petrol prices for Germany

Ariane Kehlbacher[1]
Francesca Giacco[1]
[1] German Aerospace Center DLR

Abstract: The aim of this study is to investigate changes in petrol prices over the time period from 2014 to 2025. An inverse demand function is estimated. The likelihood accounts for different types of variation including short term fluctuations, deviations from a long-term equilbrium relationship as well as structural breaks like the Covid19 pandemic and the Russian attack on Ukraine. An error correction model is estimated using rstan. It allows for cointegration between sales and prices. Currently the model runs on simulated data. But for the conference we will have the results using real data. We would use this poster to get feedback on modelling in particular in terms of issues with identification.

Estimating the prevalence of inequality constrained models of stochastic transitivity using Gibbs sampling and finite mixture models

Mattias Forsgren[1]
Gustav Karreskog Rehbinder[2, 3], Peter Juslin[1]
[1] Department of Psychology, Uppsala University
[2] Department of Economics, Uppsala University
[3] Institute for Futures Studies

Abstract: One of the most central axioms in psychology, economics, and across social science, is transitivity of preferences. It has become a received view that preferences are "often" intransitive, to the point where this is stated in the APA Dictionary of Psychology.

However, there is limited empirical evidence for this received view due to the complexities of evaluating a deterministic axiom based on noisy behavioural data. To bridge this gap, we need models of *stochastic* transitivity (ST).

We build on previous work modelling ST in two ways. (i) We develop a Gibbs sampling algorithm that directly simulates marginal likelihoods of ST models. This increases efficiency, allowing larger participant samples, and provides greater flexibility by not having to rely on existing GUI-based software. (ii) Instead of comparing models for each individual in isolation, we use STAN to fit finite mixture models that estimate prevalences of ST models in the population. This allows building a predictive (Bayesian) model and inference beyond the observed sample.

When doing so, we estimate near-universal prevalence of a model called "Strong stochastic transitivity" (SST) for everyday options (political parties, confectioneries, restaurants, charitable donations, magazines, movies, cars, dinners, holiday destinations, and fruits). SST supposes that a unitary utility is attached to each option, and that preferences are based on comparing these utilities under symmetric and homogenous noise. For monetary lotteries – the stimuli used in most previous work – we estimate a few percent to have intransitive preferences, and the rest to have SST preferences. However, intransitivity becomes very rare when we increase the difference in expected value between lotteries.

In sum, far from being "often" found, intransitivity of preferences appears to be a rare phenomenon, possibly driven by using highly similar lotteries as stimuli. The high prevalence of SST challenges the generality of theories of "constructed preferences".

Evaluating a Bayesian hierarchical meta-analysis framework for PBPK model qualification in the context of regulatory decision-making

Xiaomei Chen[1, 2]
Niels Hendrickx[1], Gustaf Wellhagen[1], Robin Svensson[3], Mats O. Karlsson[1], Andrew C. Hooker[1]
[1] Department of Pharmacy, Uppsala University, Uppsala, Sweden
[2] Unit for Pharmacokinetics and Drug Metabolism, Department of Pharmacology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[3] Swedish Medical Products Agency, Uppsala, Sweden

Abstract: Background: Physiologically based pharmacokinetic (PBPK) models are mechanistic computer models that simulate how drugs move through the body by incorporating biological parameters such as tissue volumes, blood flow rates, and enzyme activities. They are increasingly used in regulatory decisions — for instance, to replace expensive clinical trials when a model can reliably predict how one drug affects another. For such high-stakes use, regulators need evidence that a model's predictions are trustworthy. A Bayesian hierarchical meta-analysis framework was recently approved by the European Medicines Agency to quantify a PBPK model's prediction performance using a collection of historical study results. However, practical requirements for this framework — how much historical data is needed, and how robust it is — have not been systematically studied.

Methods: We conducted simulation studies using a Bayesian hierarchical meta-analysis model calibrated on 219 historical studies comparing PBPK model predictions to observed clinical outcomes. We evaluated parameter estimation accuracy, type I error, and statistical power under regulatory decision-making scenarios. The simulation studies were performed in R using Stan-based inference (*CmdStanR*) and the *SimDesign* package. We systematically varied: (1) the size of the historical dataset (4–200 studies, 5–30 subjects per study); (2) the PBPK model's performance in terms of bias and imprecision; (3) informative vs. non-informative priors; and (4) structural misspecification of the hierarchical model.

Results: Stan-based estimation converged across all scenarios and type I error was controlled throughout. With fewer than 20 historical studies, parameter estimation bias was substantial (>30%) and statistical power was low (<80%). Power exceeded 80% with 20 or more studies and plateaued beyond 50. Power was insensitive to prediction bias but declined with increasing prediction imprecision. The hierarchical model was robust to prior choice and the tested misspecifications.

Conclusions: At least 20 historical studies are required to reliably characterize PBPK model prediction uncertainty in this framework. The Bayesian hierarchical meta-analysis, implemented in Stan, proved robust across most prior specifications and to the tested forms of model misspecification, supporting its use in regulatory decision-making. These findings provide actionable guidance for designing historical validation datasets in regulatory PBPK model qualification submissions.

Exploration of the Bayesian Decision Tree Posterior via Hamiltonian Monte Carlo-based Methods

Jodie Cochrane[1, 2]
Adrian Wills[1], Sarah Johnson[1]
[1] University of Newcastle, Australia
[2] Uppsala University, Sweden

Abstract: Decision trees have found widespread application due to their flexibility and interpretability. However, a Bayesian approach to learning decision trees from data is challenging due to the need to explore both the tree structure space and the space of decision parameters associated with each tree structure. Existing attempts often use sample-based approximate inference, such as Markov chain Monte Carlo, where the proposal mechanism is based on local, random-walk characteristics, which affects the sampling efficiency and often becomes stuck in local modes.

We investigate using a Hamiltonian Monte Carlo (HMC) approach to more efficiently explore the posterior by employing a joint update scheme which exploits the geometry of the likelihood within each tree topology. We introduce two novel methods for Bayesian inference of decision trees entitled RJHMC-Tree and DCC-Tree. For both methods, HMC is used to conduct local inference within each tree structure, with two frameworks considered for global exploration of the overall space. The RJHMC-Tree approach uses a reversible-jump scheme to propose global moves. The second approach, DCC-Tree, uses a utility function similar to the Divide, Conquer, Combine (DCC) sampling strategy, whereby each distinct tree structure is instead associated with a unique set of decision parameters.

In general, HMC-based methods show better testing performance compared to existing methods. The RJHMC-Tree algorithm demonstrates significantly higher acceptance rates than existing methods. The DCC-Tree algorithm improves further on the RJHMC-Tree method by removing the observed inefficiencies associated with moving between different tree structures.

gemlib: compositional probabilistic programming for stochastic epidemics

Alin Morariu[1]
Amos Keshet[1], Lloyd Chapman[1], Jessica Bridgen[1], **Chris Jewell**[1]
[1] Lancaster University

Abstract: Infectious disease models (IDMs) have become a standard tool in epidemiological analysis. For such models to be used in practice, both parameter estimation and forward simulation algorithms are required for fully quantitative decision support. Stochastic implementations of these models represent the cutting edge for admitting detailed population structure. More importantly, they deal with large disparities in the number of infected individuals over time and space by avoiding continuous state-space approximations of their differential equation counterparts.

IDMs model a sequence of epidemiological state transitions, such as individuals' infection and removal times, conditional on the state of the population as it evolves in time. A particular challenge for parameter inference is that many such transitions (e.g. infections) are unobserved, with complex Monte Carlo methods (MCMC, SMC, ABC) being used to marginalise them out of a larger Bayesian hierarchical model. Currently probabilistic programming languages cater poorly for this situation, meaning that researchers turn to *ad hoc* computer implementations that may be slow, opaque, and difficult to debug and modify. As a solution, we present *gemlib*, a functional Python probabilistic programming library for defining a wide class of IDMs, using either discrete- or continuous-time stochastic processes.

gemlib abstracts the IDM into **State,** and one or more composable **Kernels** evolving **State** in time. Each Kernel may be considered a random variable, allowing both simulation of an IDM and automatic computation of statistical likelihood functions for principled statistical fitting. The separation of concerns between model *description* and computational logic is therefore maintained as required by a probabilistic programming framework.

gemlib is implemented in Python, providing configurable classes for arbitrarily complex IDMs which can be embedded within a larger Bayesian hierarchical model. A composable suite of MCMC samplers is provided for implementing the Metropolis-within-Gibbs samplers required for the continuous (or semi-continuous) parameter space, as well as unbiased Bayesian inference whilst accounting for censored epidemiological event data.

We demonstrate gemlib using three examples from livestock diseases: a simple individual-level epidemic model (e.g. foot and mouth disease), a network-connected metapopulation model (e.g. bovine TB), and an endemic model for anti-microbial resistance.

Hybrid deterministic–Bayesian deep learning model for pediatric bone age estimation

Paul-Philipp Jacobs[1]
Daniel Gräfe[2], Johanna Pape[2], Timm Denecke[1]
[1] Department of Diagnostic and Interventional Radiology, University of Leipzig, Leipzig, Germany
[2] Department of Pediatric Radiology, University Hospital, Leipzig, Germany

Abstract: With the growing integration of artificial intelligence into clinical workflows, particularly in radiology, model outputs increasingly inform clinical decision-making. Conventional machine-learning (ML) methods provide deterministic point estimates and do not inherently quantify uncertainty. Model performance is typically evaluated post hoc using aggregate error metrics on fixed datasets, without accounting for uncertainty in model parameters or the underlying data-generating process.

Probabilistic modeling addresses these limitations by treating model parameters as random variables with probability distributions rather than fixed values. Using Bayesian inference, posterior parameter distributions are estimated from observed data, enabling posterior predictive distributions for unseen samples that incorporate parameter uncertainty. In contrast to potentially overconfident deterministic point estimates, probabilistic approaches provide calibrated predictive uncertainty, which may be critical for clinical decision-making.

This study investigates automated bone age (BA) assessment in pediatric populations. The prevailing clinical reference standard is the Greulich and Pyle (G&P) Atlas, developed in the 1950s from left hand and wrist radiographs of upper-middle-class children and adolescents in Cleveland, Ohio.

Contemporary ML based approaches typically frame BA assessment as a supervised classification task using hand radiographs as input and the G&P BA, annotated by expert radiologists, as the discretized target variable. These models generally produce deterministic point estimates derived from convolutional neural networks trained to approximate the atlas-based reference standard.

In contrast, our approach differs in two principal aspects. First, we replaced the G&P BA proxy with the patient's chronological age as the target variable, thereby avoiding reliance on atlas-based annotations. Second, we implemented a probabilistic modeling framework to enable predictive uncertainty quantification. The proposed architecture comprises a ResNet-like backbone serving as a deterministic feature extractor, followed by fully connected layers that incorporate patient sex as an additional covariate. While the convolutional layers are treated deterministically, the subsequent fully connected layers are formulated within a Bayesian framework, modeling their weights as probability distributions and trained via stochastic variational inference. The extracted image features and sex information are propagated through these Bayesian layers to generate posterior predictive distributions. BA estimation is thereby formulated as a continuous regression problem, yielding probabilistic predictions rather than single point estimates.

Investigating the robustness of a Bayesian meta-analysis of mesdopetam's anti-dyskinetic efficacy to different modelling choices

Erik Werner[1]
Fredrik Hansson[1], Susanna Waters[1]
[1] IRLAB

Abstract: Mesdopetam is a dopamine D3 receptor antagonist in clinical development for the treatment of levodopa-induced dyskinesia in Parkinson's disease. The clinical efficacy of mesdopetam has been evaluated in a phase IIa and a phase IIb study, both randomized controlled trials comparing mesdopetam against placebo. A Bayesian meta-analysis of the results has previously been published as part of a poster at the 2024 International Congress of the International Parkinson and Movement Disorder Society. However, while that work consists of a single meta-analysis, in this work we investigate how results vary under a large number of different plausible model specifications.

Pooling the two studies is complicated by the fact that they differ in many important respects, including the patient population, the duration of the study (4 vs. 12 weeks) and the dosing regimen (initially 7.5 mg BID with the possibility of increasing or decreasing the dose vs. 2.5, 5 and 7.5 mg with the possibility of decreasing the dose). Further, even for a single study there are a large number of researcher degrees of freedom in terms of e.g. the statistical model used for each endpoint, the selection of a prior distribution, and the handling of missing data and dropouts. In this analysis, we construct and sample from a large number of Bayesian models to investigate how sensitive our efficacy estimates are to different modelling choices. While the exact numerical estimates depend on the model specification, the general finding that mesdopetam has a clear anti-dyskinetic effect is robust.

LOO-PIT predictive model checking

Herman Tesso[1]
Aki Vehtari[1, 2]
[1] Aalto university
[2] ELLIS institute Finland

Abstract: We consider predictive checking for Bayesian model assessment using leave-one-out probability integral transform (LOO-PIT). LOO-PIT values are conditional cumulative predictive probabilities given LOO predictive distributions and corresponding left out observations. For a well-calibrated model, LOO-PIT values should be near uniformly distributed, but in the finite sample case they are not independent, due to LOO predictive distributions being determined by nearly the same data (all but one observation). We prove that this dependency is non-negligible in the finite case and depends on model complexity. We propose three testing procedures that can be used for continuous and discrete dependent uniform values. We also propose an automated graphical method for visualizing local departures from the null. Extensive numerical experiments on simulated and real datasets demonstrate that the proposed tests perform (at least) competitively against existing methods aimed at mitigating the effect of data-induced dependence in PIT values and have much better power than standard independence-based uniformity tests that often lead to lower than expected rejection rate.

Measuring neutrino oscillation in the NOvA experiment with Stan

Cullen Sullivan[1]
[1] Tufts University

Abstract: The neutrino is a fundamental particle that exhibits a peculiar quantum-mechanical behavior called oscillation, wherein a neutrino can change its state as it time-evolves. An experiment like NOvA, discussed in this poster, uses an artificial neutrino beam to detect transitions between the states to measure associated parameters. These parameters are among the 26 fundamental parameters of the Standard Model of Particle Physics. However, the nonlinear and degenerate character of the oscillation parameter space and the quantity and complexity of systematic uncertainties make extraction of the posterior computationally challenging.

To overcome these challenges, the NOvA collaboration uses Stan's NUTS HMC algorithm to explore 4 oscillation parameters and about 120 nuisance parameters associated with systematic uncertainties. Because of ambiguities in the parameter space, the posterior is composed of two fully disjoint subspaces designated by a binary "mass ordering" choice which is difficult to sample with HMC. δ_{CP}, a parameter which governs matter-antimatter asymmetry in neutrino oscillation, is cyclic, so naive sampling disrupts the NUTS algorithm. Finally, many systematic parameters relate to uncertainties in how neutrinos interact with matter with downstream effects on experimental observables that are too complicated to express in analytic functional form. Instead, a Monte-Carlo method called "event generation" is used to simulate the entire dataset under systematic variations. This poster will discuss how NOvA solves all of these problems to obtain efficient sampling of both oscillation and nuisance parameters with Stan.

Modeling the complexity of tree growth with hierarchical Gaussian processes and structured shocks

Victor Van der Meersch[1]
Michael Betancourt[2], Benjamin Cook[3, 4], E. M. Wolkovich[1]
[1] Faculty of Forestry, Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC, Canada
[2] Symplectomorphic LLC, New York, NY, USA
[3] Goddard Institute for Space Studies, NASA, New York, NY, USA
[4] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Abstract: Accurate models of tree growth are critical to many fields, from managing forests to modeling carbon storage. Tree growth, however, is influenced by many factors including long-term trends due to allometry, shorter-term effects due to disturbance, and broad-ranging effects from climate. Traditionally, most approaches have avoided this complexity by detrending time series and aggregating well-behaved trees from proximate locations. While this has worked in earlier analyses, shifts in forest dynamics with climate change have made it problematic.

We address this with a Bayesian model that jointly captures the critical influences and apply it to ring width data from Western US. At the tree level, two GPs capture variation not directly related to climate: a component for long-term trajectories shaped by species-specific allometry, and a short-term component for canopy dynamics. At the forest level, three covariates (growing degree-days, soil moisture, vapor pressure deficit) represent the main climatic drivers of growth. A third GP captures remaining forest-level dynamics attributable to missing climatic and biotic disturbances.

To account for episodic extreme growth reductions, we also introduced nested shocks to the model. A latent state governs whether a forest experienced an extreme event in a given year. Conditional on this state, individual trees may respond concordantly, exhibiting a major growth suppression modeled as additional variance. This nested structure, where individual shocks depend on a forest-level state, is designed to capture extreme events whose amplitude is not explained by the rest of the model.

This shock model explains up to 25% of growth decrease in extreme years, with most events linked to major droughts with severe impacts that are not well captured by standard climatic covariates. This suggests fundamental gaps in our understanding of tree responses to climate. As droughts intensify under climate change, failing to account for these non-linear shocks could underestimate forest vulnerability.

Modelling individual-level mobility, colocation networks, and higher-order structure

Jessica Bridgen[1]
James Chirombo[2], Derek Cummings[3], Chris Jewell[1], Chao Qiang Jiang[4], Kin On Kwok[5], Justin Lessler[3, 6], Steven Riley[7], Jonathan Read[1]
[1] Lancaster University
[2] Liverpool School of Tropical Medicine
[3] Johns Hopkins Bloomberg School of Public Health
[4] Guangzhou No. 12 Hospital
[5] The Chinese University of Hong Kong
[6] University of North Carolina, Chapel Hill
[7] Imperial College London

Abstract: The relationship between human mobility, social contact, and infectious disease transmission has long been recognised. We present a framework which links individual-level mobility to colocation networks, utilising four annual contact surveys from the FluScape cohort study conducted in Guangzhou, China. Colocation networks connect individuals who reported visiting the same geographic location, serving as a proxy for respiratory virus transmission opportunities. Representing social contact events as a higher-order network moves beyond pairwise descriptions to capture group-level structure of social mixing, offering insight into the mechanisms that drive transmission.

Using a Bayesian hierarchical logistic regression model, we found that mobility was primarily governed by distance decay and destination-specific attractiveness, with consistent effects across study years. The study region was discretised into 100m grid cells, each representing a potential destination. Beyond distance and population density, destinations exhibited unique attractiveness and we found substantial heterogeneity in individual mobility. From model posterior samples, simulated colocation networks broadly reproduced observed network structure across study years. These results provide a step towards understanding how individual mobility shapes the contact networks underpinning respiratory virus transmission.

Modelling Short-Term Voting Propensity Shifts as Trajectories on the Simplex: A Bayesian State-Space Imputation Model Using High-Frequency Panels

Johann-Friedrich Salzmann[1]
[1] MZES, University of Mannheim

Abstract: In modern multi-party democracies, it has become increasingly common for multiple voting intention polls to be conducted per week, providing near-continuous snapshots of public opinion, shaping public discourses. Yet it remains unclear which underlying individual changes in citizens' opinions drive the visible shifts in aggregate party support. While existing electoral research has extensively studied vote switching between elections, little is known about how flows of potential voters unfold in the short term.

This study introduces a novel approach to modelling voting intention change based on individual-level preference trajectories observed in high-frequency panel data. I propose a Bayesian state-space imputation model combining individual-level preference stability and time-specific opinion dynamics to estimate wave-to-wave preference transitions between parties, or into and out of non-voting or undecidedness, correcting for panel attrition and unbalanced participation. Employing a multi-level Dirichlet structure, the proposed model - implemented in STAN - operates in the high-dimensional simplex space without requiring a log-ratio transformation, directly representing trajectories of individual latent voting intention propensities. I introduce a scaling function and convex combination operators that account for compositional constraints, agnostic to whether the underlying space is interpreted in real or Aitchison geometric terms.

The proposed method is applied to data from the ongoing bi-weekly SOSEC panel study in Germany (n=1500, online access panel) with a focus on the German 2025 federal election campaign. It is demonstrated how the sudden increase in the Left party's support in the weeks prior to the election can be decomposed into interpretable underlying voter flows. The study finds that the Left party both successfully attracted votes from centre-left parties and was able to mobilise numerous previous undecided and non-voters.

This approach both highlights the value of high-frequency panel data for studying the micro-dynamics of political behaviour and offers an alternative, transformation-free modelling approach for data-generating processes on the simplex.

Multiscale resource availability and spatiotemporal synchrony shape bumblebee foraging patterns

Jenna Melanson[1]
Claire Kremen[1]
[1] University of British Columbia

Abstract: We present a spatially explicit genetic capture-recapture model implemented in Stan to investigate differences in foraging behaviour between two bumblebee species: a widespread native species (*Bombus mixtus*) and a rapidly expanding introduced species (*B. impatiens*). The model is motivated by a central question in pollination ecology—how does the spatial distribution of floral resources in the vicinity of a nest site shape foraging decisions— and by the practical challenge of inferring foraging ranges from genetic data of realistic quality. We first show that genetic capture-recapture data of realistic quality carry more information regarding relative than absolute foraging distance, and describe strategies for preventing overestimation of foraging range by using informative priors on colony locations. Using the best-performing model, we show that introduced *B. impatiens* forages at a greater scale than native *B. mixtus*. We then extend the model to incorporate local- and landscape-scale resource availability as predictors of relative foraging range. Results suggest that both species travel greater distances in landscapes with higher overall resource availability and to visit locations with higher abundance of a single floral species. In contrast, only *B. mixtus* shortens its foraging range in landscapes where habitat types are more closely interspersed—a pattern consistent with greater reliance on spatiotemporal complementation of resources across habitat types. This pattern is further supported by evidence of temporal turnover in colony-specific foraging locations throughout the season. Our results indicate that pollinator foraging behaviour is shaped not only by local patch quality, but by the quantity and configuration of habitat across the landscape, and that foraging kernel locations change dynamically throughout the season.

Posterior Sampling of Word Embeddings

Väinö Yrjänäinen[1]
Isac Boström[2], Johan Jonasson[2], Måns Magnusson[1]
[1] Department of Statistics, Uppsala University
[2] Department of Mathematical Sciences, Chalmers University of Technology

Abstract: We study Bayesian inference for probabilistic word embedding models, including skip-gram with negative sampling (SGNS) and continuous bag-of-words (CBoW). These models are widely used for inference, but their posteriors are non-identifiable under general linear transformations, which prevents meaningful Bayesian uncertainty quantification. We provide a simple and principled identifiability constraint that yields a well-defined posterior without altering the data model. Building on this, we develop scalable, fully Bayesian posterior sampling algorithms for SGNS and CBoW using Pólya–Gamma augmentation. Our block Gibbs samplers yield exact samples from the posterior and scale well to large corpora due to their straightforward parallelisation. We discuss the efficient implementation of the Gibbs sampler, and provide a fast TensorFlow reference implementation. In a combination of simulations and empirical experiments, we demonstrate that the sampler is both accurate and computationally efficient. It matches the accuracy of NUTS-HMC at a fraction of the computational cost and avoids the variance underestimation of mean-field variational inference. On large real datasets, we show that posterior means yield improved predictive performance, while posterior credible intervals provide meaningful uncertainty quantification for semantic similarity.

Predictive assessment and comparison of Bayesian survival models for cancer recurrence

Saku Suorsa[1, 2]
Aki Vehtari[1, 2]
[1] Department of Computer Science, Aalto University, Finland
[2] ELLIS Institute Finland

Abstract: Complex data features, such as unmodelled censored event times and variables with time-dependent effects, are common in cancer recurrence studies and pose challenges for Bayesian survival modelling. Current methodologies for predictive model checking and comparison often fail to adequately address these features. This paper bridges that gap by introducing new, targeted recommendations for predictive assessment and comparison of Bayesian survival models. Our recommendations cover a variety of different scenarios and models. Accompanying code together with our implementations to open source software help in replicating the results and applying our recommendations in practice.

Probabilistic Estimation of Agent Based Models using Hamiltonian Monte Carlo

Benjamin Olsson[1]
Måns Magnusson[1], Sara Hamis[1]
[1] Uppsala University

Abstract: Agent-based models (ABMs) are increasingly used to study complex systems across many fields, but their intractable likelihoods make standard inference methods such as maximum likelihood estimation or Markov chain Monte Carlo (MCMC) difficult to apply. Likelihood-free methods such as approximate Bayesian computation and Bayesian synthetic likelihood (BSL) address this through simulation, but at high computational cost. Recent work has sought to reduce this cost by replacing simulations with learned Gaussian process (GP) emulators. We build upon these works and exploit the differentiability of a GP emulator within the BSL framework to enable efficient parameter inference via Hamiltonian Monte Carlo (HMC). Using recent results from GP convergence and BSL asymptotics, we theoretically establish the asymptotic accuracy of the proposed emulator. We then apply the method to a variety of simulated birth-death-movement ABMs. For these models, given an identical GP emulator, gradient-based HMC sampling improves exploration of the posterior compared to gradient-free MCMC sampling.

Quantifying uncertainty: a methodological comparison of Bayesian and frequentist meta-analyses of studies on lead exposure and children's IQ

Paulina Sell[1]
Margaux Sanchez[2], Philippe Palmont[2], Simon Steiger[3], Dietrich Plass[1]
[1] German Environment Agency, Environmental Hygiene, Berlin, Germany
[2] French National Agency for Food, Environmental and Occupational Health & Safety, Maisons-Alfort, France
[3] Karolinska Institutet, Department of Medicine Solna, Division of Clinical Epidemiology, Stockholm, Sweden

Abstract: Objective: Uncertainty quantification is central to evidence synthesis for risk assessment and policy-making, but frequentist meta-analytic approaches - currently predominant in epidemiology - often fall short in this regard. We compared frequentist and Bayesian frameworks to evaluate differences in uncertainty quantification and applicability for risk assessment, using lead exposure and children's IQ as a case study.

Methods: We searched PubMed and Scopus for epidemiological studies on blood lead level (BLL) and IQ score in children. After screening, risk of bias (RoB) was assessed and regression coefficients were harmonized. In both meta-analyses, we included the same effect estimates, assumed a multilevel structure and used random-effects models. The Bayesian meta-analysis employed a hierarchical model implemented in brms with weakly-informative priors. Sensitivity analyses examined robustness to RoB and to prior specification. τ^2 and I^2 were estimated as measures of between-study heterogeneity.

Results: Twelve studies were included in the meta-analyses. The frequentist approach yielded a pooled estimate of -2.45 (95% confidence interval: -3.7; -1.2) IQ points per unit increase in $\log_e$-transformed BLL (µg/dL). The Bayesian model yielded a mean pooled estimate of -2.28 (95% credible interval: -3.56; -1.06). Between-study heterogeneity was substantial (frequentist: $\tau^2 = 3.3$, $I^2 = 85\%$; Bayesian: $\tau^2 = 3.8$, $I^2 = 83.3\%$). The Bayesian approach provided a full posterior distribution for heterogeneity, directly quantifying uncertainty. Sensitivity analyses showed robustness to RoB and prior specification. The Bayesian approach was less sensitive to outlier studies reporting substantial uncertainty, resulting in a lower pooled effect compared to the frequentist approach. This regularisation is particularly valuable in small meta-analyses, where frequentist estimates are unstable.

Conclusion: We provide updated pooled estimates for the effect of lead exposure on children's IQ. The Bayesian approach provides posterior distributions, enabling direct probability statements, e.g. about exceeding risk thresholds, and comparative assessments across pollutants. These features are of particular value regarding risk assessment and policy-making.

Sequential Bayesian Causal Evaluation of Shared Investments in Nature-Based Assets: Thompson Sampling for Mechanism Learning under Uncertainty

Hanna Fiegenbaum[1]
Juan Camilo Orduz[2]
[1] Leipzig University
[2] PyMC Labs

Abstract: Shared investment in nature-based real assets is critical for strengthening nature and biodiversity as well as climate resilience, yet real-world performance depends not only on mobilized capital but on the institutional mechanisms that coordinate diverse actors over time. Existing policy design lacks tools to compare and adapt investment mechanisms under uncertainty and does not account for behavioral and action stability and ecological risk developing over time. This project introduces a synthetic, simulation-based Bayesian mechanism learning framework employing Thompson sampling bandits for counterfactual causal analysis and sequential adaptation of shared investment mechanisms to explore ecological, behavioral, action-based and economic performance trade-offs.

Visualising ROC curves for Bayesian classification models: a decision-space view of summary curves, parameterisation, and coverage

Massyle Oumessaoud[1]
Ludvig Hult[1], Nicolas Pielawski[1]
[1] Uppsala University

Abstract: Bayesian classifiers, from logistic regression and Gaussian process classifiers to Bayesian neural networks, induce a posterior over ROC curves. Yet published figures almost always show a single curve, collapsing the posterior through a summarisation that is rarely made explicit. We show this step is consequential, and decompose it into three orthogonal axes needed to understand the discriminative behaviour of a Bayesian classifier.

The first axis is the summary curve. The default is the pointwise expected ROC, but alternatives, pointwise medians, AUROC-matched draws, AUROC-weighted means, AUROC-constrained isotonic or concave projections, trade off monotonicity, concavity, agreement with the posterior AUROC, and proximity to draws.

The second axis is parameterisation. The standard view fixes a false positive rate and treats the true positive rate as random. The alternative fixes the threshold τ and treats both FPR(τ) and TPR(τ) as random, giving a cloud rather than a band. Bayesian inference makes both practical: two ways of summarising one posterior.

The third axis is the coverage guarantee, where posterior uncertainty becomes visible. Uncertainty can be reported pointwise (a credible interval at each FPR or threshold) or simultaneously, as a tube with joint coverage; the choice determines what probabilistic statement the figure supports. Pointwise intervals, sup-t / tGKF / Fair simultaneous bands, modified-band-depth and extremal-depth envelopes, and ellipsoid tubes (with adaptive and cross-correlated variants we introduce) trade width for joint strength.

We implement existing and two new methods across four posterior-sampling families (PyMC/NUTS logistic regression, Erkanli's Bayesian bootstrap, Parker's Beta-density framework, binormal conjugate) and evaluate coverage on synthetic and Pima Indians data. The axes are largely orthogonal: changing the parameterisation can flip which method is widest at fixed coverage, and AUROC-consistency or concavity constraints can move the summary curve outside any pointwise band. The poster presents the decision space as a navigable map with reporting recommendations.

‘A fine balance’: Exploring the interplay of reproductive strategies and growth dynamics in trees

Xiaomao Wang[1]
Victor Van der Meersch[1], Elizabeth Wolkovich[1]
[1] Faculty of Forestry, Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC, Canada

Abstract: Masting is the phenomenon of synchronized and highly variable seed production within plant populations. It plays a critical role in ecosystem-level functions, influencing both plant regeneration and the population dynamics of seed predators. Despite its importance, models of masting are poorly predictive, likely because of the complex temporal dependencies and the difficulty of integrating environmental variation with internal plant dynamics. One of the most important internal dynamics hypothesized to drive masting is a trade-off between vegetative growth (leaves, stems, wood) and reproduction (seeds). However, this trade-off is often obscured by environmental variation simultaneously. Here, we leverage a rarely available coupled dataset over time and space, along with a new hierarchical Bayesian generative model, to test for evidence of this trade-off while accounting for variation in the environment.The model explicitly treats annual resource allocation as a latent process occurring at the individual tree level, informed by year-to-year environmental variation and two distinct data sources: annual tree-ring widths as a proxy for growth and annual seed production records as a proxy for reproduction, collected from Mount Rainier National Park since 2008. Our approach reconstructs hidden resource pools from observed ecological data and tests how they shift in response to environmental variation during masting events, specifically modeling shifts in snowpack that shape moisture over the growing season and temperatures that drive growing season warmth. Uniquely, our dataset includes multiple stands across an elevational gradient, allowing us to model how environmental variation both between years and across space within years may shape resource pools and allocation. We utilize hierarchical structures to pool information across individuals, stands, and species to help estimate latent state transitions and how they vary across scales. This model demonstrates the use of Bayesian modeling and latent variable frameworks helping decipher unobservable physiological processes from observable ecological data.

Poster, early acceptance

Tutorial

Bayesian model diagnostics: Workflows and software tools

Osvaldo Martin[1]
Teemu Säilynoja[1, 2], Noa Kallioinen[1]
[1] Aalto University
[2] PyMC Labs

Abstract: Course description
This tutorial provides an overview and hands-on experience of diagnostics for Bayesian models. We will present software tools for using concepts from recent Bayesian workflow research in practice. Along with gaining an understanding of the underlying concepts, participants will learn to use software such as the R packages posterior, bayesplot, and priorsense, and the Python package ArviZ. Exercises will be provided in both R and Python to accommodate each participant's preferred language.

Topics
The material is structured around a modern Bayesian workflow, emphasizing the iterative nature of model building, fitting, evaluation, and critique. Rather than treating diagnostics as isolated checks, we frame them as integral tools for guiding model revision and improving the reliability of our models.

We will work primarily with posterior draws that have already been generated. This allows us to focus on diagnostics and workflow rather than on the details of fitting individual models. Using these draws, participants will explore key components of this workflow, including basic manipulation of posterior draws, checking convergence with R-hat (including rank-normalized and nested variants), effective sample size (ESS), and Pareto-k diagnostics, and model evaluation and critique with prior sensitivity checks and visual predictive checks, including PIT ECDFs, calibration plots, and interval-based summaries.

The focus will be on providing the latest recommendations based on recent research, and showing how to translate these into practice.

Instructors
Osvaldo Martin – Research fellow at Aalto University. Core developer of ArviZ and other packages for Bayesian statistics
and probabilistic programming.

Teemu Säilynoja – Data Scientist at PyMC Labs, and a doctoral candidate focusing on model calibration assessment.
Contributor to bayesplot, posterior, SBC, as well as PyMC-Marketing.

Noa Kallioinen – Doctoral researcher in Bayesian workflow at Aalto University with focus on prior specification and sensitivity checking. Lead developer of priorsense, and contributor to posterior.

Live "Learning Bayesian Statistics" Episode

Alexandre Andorra[1]
[1] Learning Bayesian Statistics

Abstract: Hello dear StanCon team, and thank you so much for your hard work!

I'm reaching out because I'd be honored to take part, if possible. More precisely, I'm thinking about doing one (or even two!) live LBS episode (https://learnbayesstats.com/) there -- that'd be a blast!

I know the names of attendees are not out yet, but I'm flexible and will customize a topic to what StanCon wants to focus on this year. As an example, here are the live shows I did at StanCon 2024:

1. #118 Exploring the Future of Stan, with Charles Margossian & Brian Ward: https://www.youtube.com/watch?v=nTECQ_uvOBc&t=3s
2. #120 Innovations in Infectious Disease Modeling, with Liza Semenova & Chris Wymant: https://www.youtube.com/watch?v=m2ID2GvHib4&t=2s

If that's of interest to you, I'd be happy to coordinate contacting the panelists and preparing the topics and questions.

Looking forward to hearing from you, and best Bayesian wishes,
Alex Andorra

Workshop

Bayesian inference for sparsity-promoting and edge-preserving priors

Jan Glaubitz[1]
Yiqiu Dong[2], Lassi Roininen[3]
[1] Linköping University
[2] Technical University of Denmark
[3] LUT University

Abstract: Inverse problems arise throughout science and engineering when experimental or observational data are used to infer unknown parameters, states, or structures in mathematical models. The Bayesian approach provides a principled framework for solving such problems by modeling the unknown quantities probabilistically and combining prior information with data through Bayes' rule. This perspective not only yields point estimators but also quantifies uncertainty in reconstructions and predictions—essential for reliable decision-making across applications such as medical imaging, geophysics, climate modeling, weather forecasting, and materials science.

A central theme of this workshop is the development of efficient methods for Bayesian inference for inverse problems with sparsity-promoting and edge-preserving priors. Such priors often lead to heavy-tailed and multimodal posterior distributions, which are particularly challenging to explore. In this context, topics of interest include scalable algorithms for high-dimensional posterior inference, hierarchical prior models, mixture models, diffusion- and transport-based sampling methods, and techniques for uncertainty quantification. Closely related areas such as data assimilation, experimental design, Bayesian neural networks, discovering biochemical reactions, and the quantification of rare or extreme events are also within the workshop's scope.

The goal of the workshop is to bring together researchers in applied mathematics, statistics, computational science and engineering, and machine learning who work on complementary aspects of Bayesian inverse problems and uncertainty quantification. By fostering interactions across traditionally separate communities—such as inverse problems, numerical analysis, Bayesian computation, and data assimilation—the workshop aims to catalyze new collaborations and stimulate the development of unified mathematical and computational frameworks for Bayesian inference in complex scientific systems.