# Internship: Bayesian neural networks in formulation development

## Description

Nutrients play an important role in human health and must be supplied in the form of food and/or nutritional supplements. When a new galenic form with a new combination of ingredients is developed, experimental studies are carried out to determine storage stability. This can result in increased development costs or delay product launches. Therefore, there is the need for predictive models for shelf-life. Especially challenging is the heterogeneity of the underlying measurements, hampering the identification of degradation drivers.

In this internship, we will use adaptive Hamiltonian Monte-Carlo (HMC) sampling to estimate not only degradation trajectories of nutritional formulations but also the associated Bayesian credible intervals. We will work with Bayesian neural networks, and actual industrial data reflecting the degradation of active ingredients over time under given storage conditions.

To capture technical replication errors while extracting as much physical insight as possible, we will explore hybrid models that combine an algebraic mechanistical sub-model with a data-driven neural network sub-model. The goal is a systematic comparison to identify the most expressive model.

To achieve this goal, we propose the following steps as part of this internship

* Implementation of dedicated predictor variables
* Implementation of a pool of models in pyro and pytorch
* Bayesian hyperparameter search with respect to model architecture and sampler parameters to control the trade-off between sampler convergence and model expressiveness

## Literature

[1] [Modeling the degradation kinetics of ascorbic acid - PubMed (nih.gov)](https://pubmed.ncbi.nlm.nih.gov/27892705/)

[2] [[2107.03342] A Survey of Uncertainty in Deep Neural Networks (arxiv.org)](https://arxiv.org/abs/2107.03342)

## Eligibility

* Enrollment at TU Dortmund University required
* Internship for Module “Project Work”; this part of the module should not have been completed yet

## Must-haves

* Strong programming skills in python, incl. pandas, numpy, pytorch, and any of matplotlib, plotly dash, or seaborn
* Background in systematic model selection, maximum likelihood estimation and feed-forward MLPs
* Interest in Markov-Chain Monte-Carlo methods
* Team mindset and creative modeling mindset

## Nice-to-haves

* Knowledge of pyro, numpyro, plotly dash, ipython notebooks, optuna
* Knowledge of HMC / NUTS, Gelman’s convergence criteria
* Background in hybrid modeling and Bayesian optimization
* Experience with Git, Linux

## Contacts

TU Dortmund University: Prof. Dr. Katja Ickstadt, Department of Statistics, [ickstadt@statistik.tu-dortmund.de](mailto:ickstadt@statistik.tu-dortmund.de)

Bayer AG: Dr. Moll Glass, Applied Mathematics, [moll.glass@bayer.com](mailto:moll.glass@bayer.com)

Applicants are asked to include their CV, preferred start and end date and preferred workload.

## Miscellaneous

We recommend a duration of at least 3 months at 80% (4 days internship, 1 day off). This is a paid internship. Bayer will also provide the necessary hardware and a shared office. Working from home is possible.