

A workshop on Generalised Additive Mixed Modelling for linguists

Marton Soskuthy (University of York)

Dec. 7th from 09.30-17.00 in the seminar room (room 403) at ZAS (<http://www.zas.gwz-berlin.de/148.html?&L=0>).

Please register yourself by sending an email to Marzena Zygis zygis@zas.gwz-berlin.de

The workshop will be structured in the following way:

09.30–11.00: Introducing GAMMs: what are they and when should we use them?

11.00–11.15: Coffee break

11.15–12.15: Significance testing using GAMMs

12.15–13.45: Lunch break

13.45–15.00: Fitting simple GAMMs using R

15.00–15.15: Coffee break

15.15–17.00: Fitting GAMMs to more complex data sets

You are asked to upload the following packages in R (<https://www.r-project.org>) or RStudio (<https://www.rstudio.com/products/rstudio/download3/>)

R packages:

– lme4

– mgcv

– itsadug

– ggplot2

As for readings, Marton Soskuthy has written a tutorial on GAMMs, which I've attached. You may also want to check out Bodo Winter's introduction to mixed effects modelling <http://www.bodowinter.com/tutorials.html>

This workshop provides a hands-on introduction to Generalised Additive Mixed Models (GAMMs) in the context of linguistics. GAMMs are an extension of linear regression models, which can capture (i) non-linearities in the data by using so-called smooth terms and also (ii) dependencies across data points representing the same participant, word, etc. through the use of random intercepts, slopes and smooths. GAMMs are incredibly flexible and are suitable to a large range of different linguistic data sets. However, they are also more complex than linear mixed effects regression models, which means that they can be harder to work with. This workshop will help you get started with GAMMs by tackling some fundamental concepts and also giving you some practice in fitting GAMMs to linguistic data.

We will start with a short and non-technical introduction to the general principles of GAMMs, discussing a few fundamental notions such as basis functions, smoothing penalties and random smooths. Though these concepts are typically discussed using mathematical terminology and formalisms, we will take a different approach. The focus will be on the intuitions behind these concepts, which will be amply illustrated using simple and straightforward examples from phonetics. We will then turn to the topic of significance testing

using GAMMs, which is a somewhat thorny issue and therefore deserves to be discussed in detail.

The rest of the workshop will be run as a practical, and we will work through a few examples together. Since the state-of-the-art software for fitting GAMMs is implemented in the R statistical software environment, we will use R for the practical. Although participants are not required to be R wizards, some minimal knowledge of R will be helpful. Experience with mixed effects regression models is not a must, but it would certainly be an advantage.

Since my own research is mostly on phonetics, my examples come from dynamic acoustic analysis. However, if you have a data set that is amenable to a GAMM analysis, I would be more than happy to use it for illustration. Please do let me know in advance if you're interested, as I will need some time to build your data into the practical.