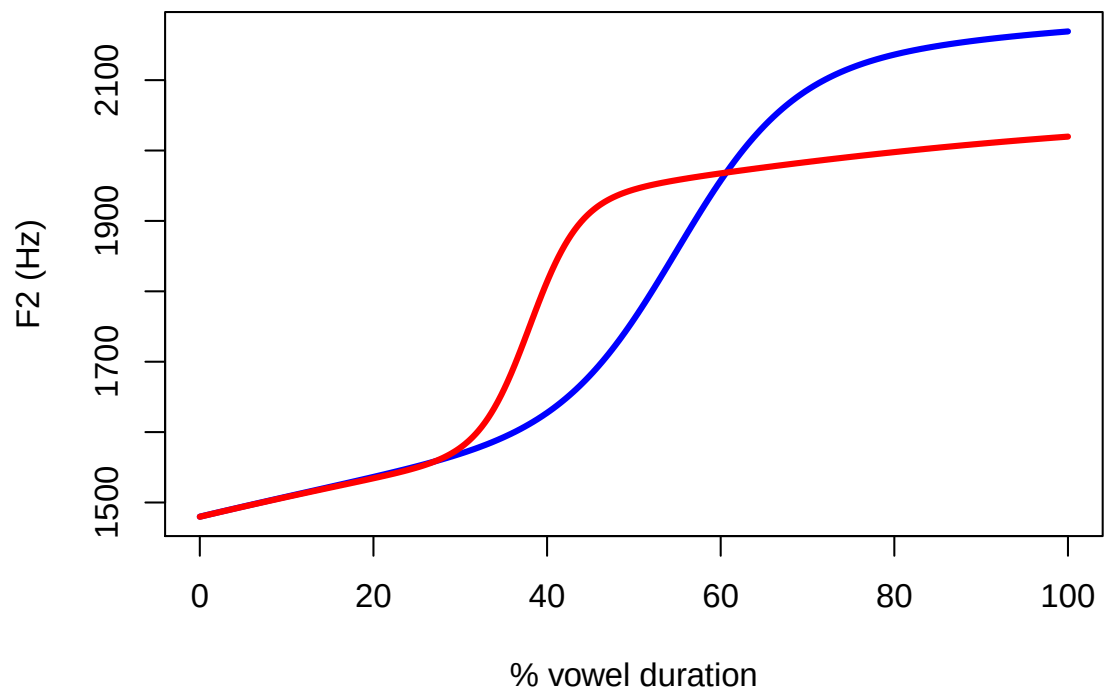
A simulation-based comparison of approaches to significance testing using generalised additive mixed models

Márton Sóskuthy
University of York

11 June 2017
PaPE, Köln

Dynamic analysis & GAMMs

- dynamic analysis: series of measurements with temporal / spatial structure
- in phonetics: formant & pitch trajectories, tongue contours, etc.



Dynamic analysis & GAMMs

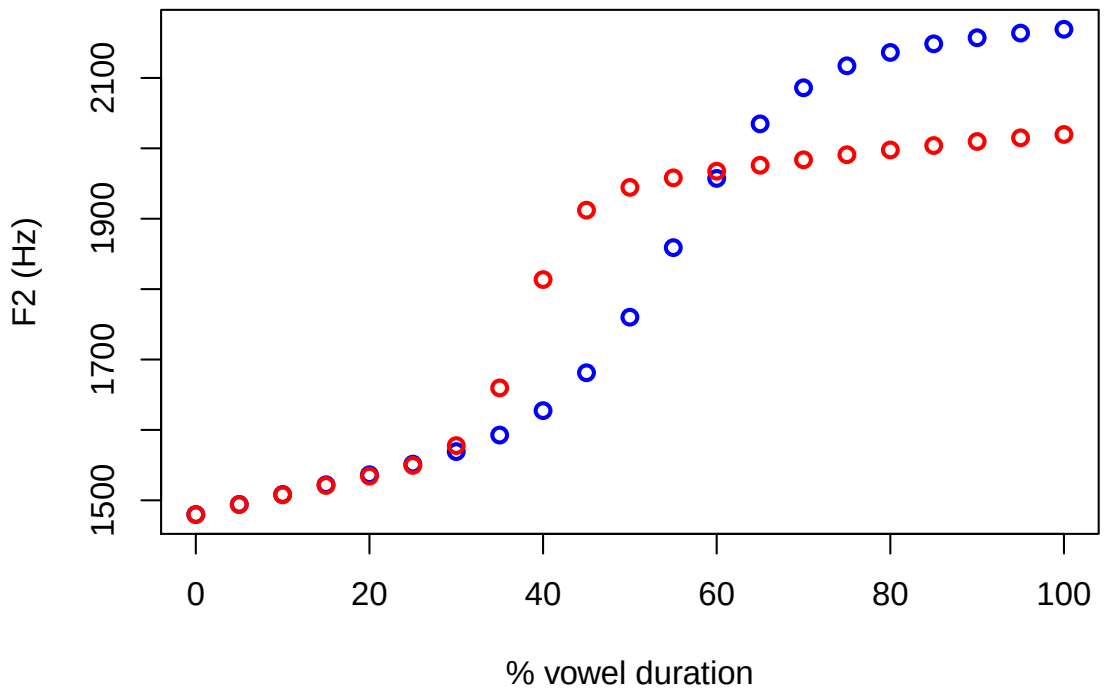
- dynamic analysis is difficult. . .
- what aspects of a curve should we focus on?
  - duration?
  - timing of transition?
  - size of movements?
  - . . .
- how do we account for dependencies within curves? (cf. later)

- Generalised Additive (Mixed) Models (GAMMs): a very general solution to this problem

Dynamic analysis & GAMMs

- linear regression models:

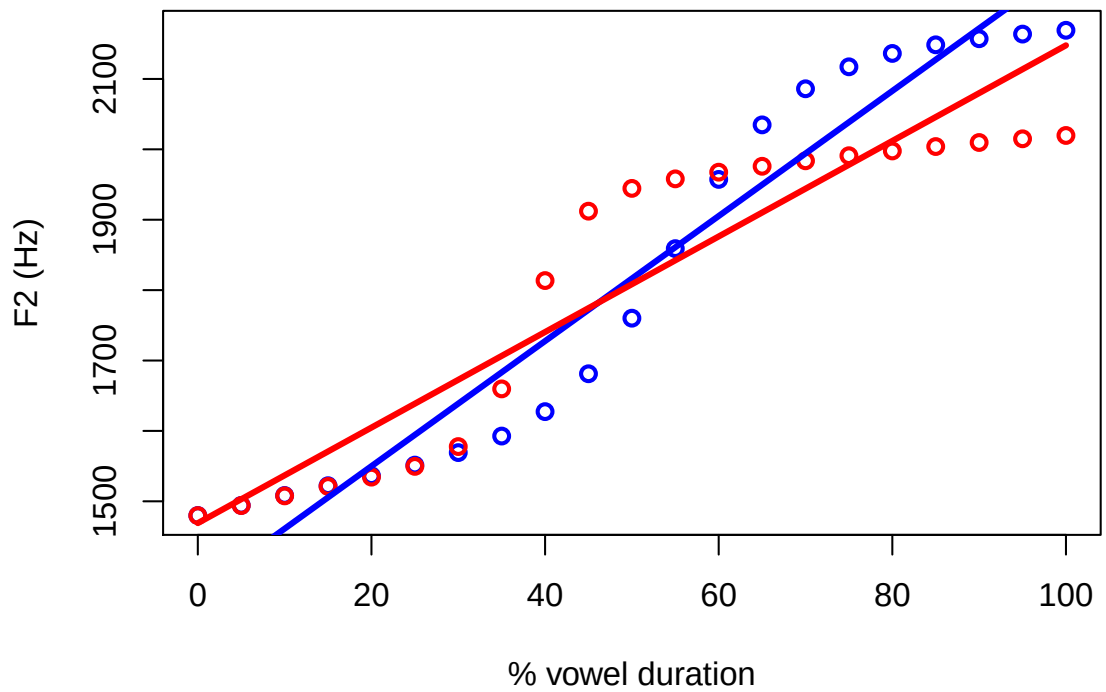
$$\mathbf{y} = \alpha + \beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + \ldots + \epsilon \quad (1)$$



Dynamic analysis & GAMMs

- traditional regression models:

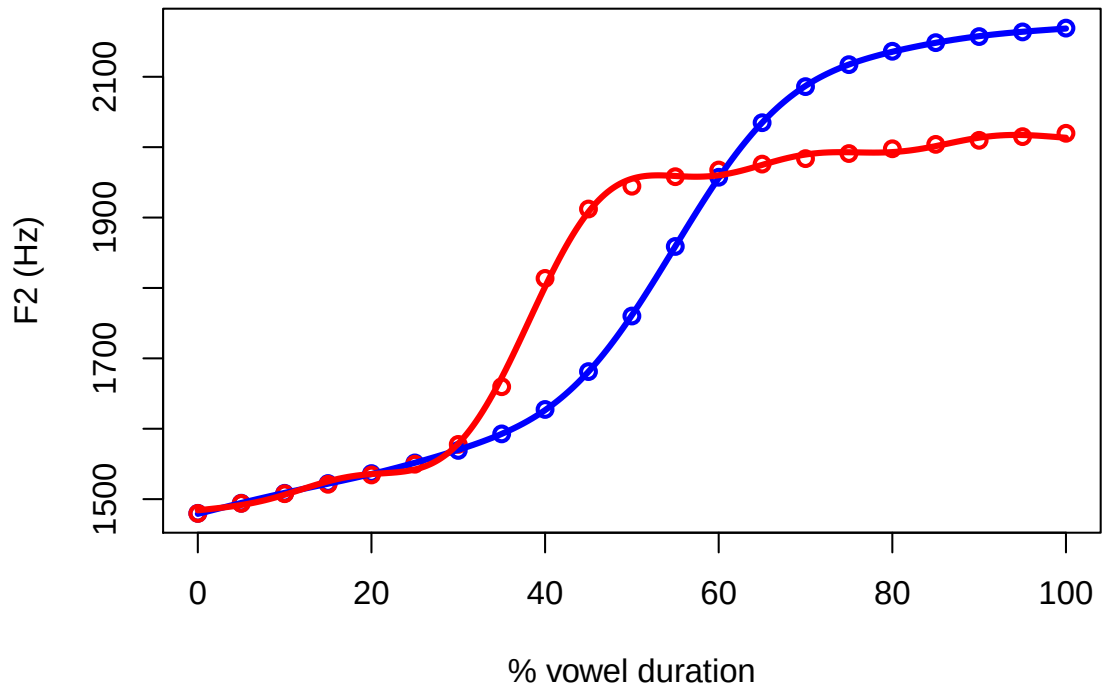
$$\mathbf{y} = \alpha + \beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + \ldots + \epsilon \quad (2)$$



Dynamic analysis & GAMMs

- (generalised) additive models:

$$\mathbf{y} = \alpha + \beta_1 \mathbf{x_1} + \ldots + f_1(\mathbf{x_1}) + f_2(\mathbf{x_2}) + \ldots + \epsilon \quad (3)$$



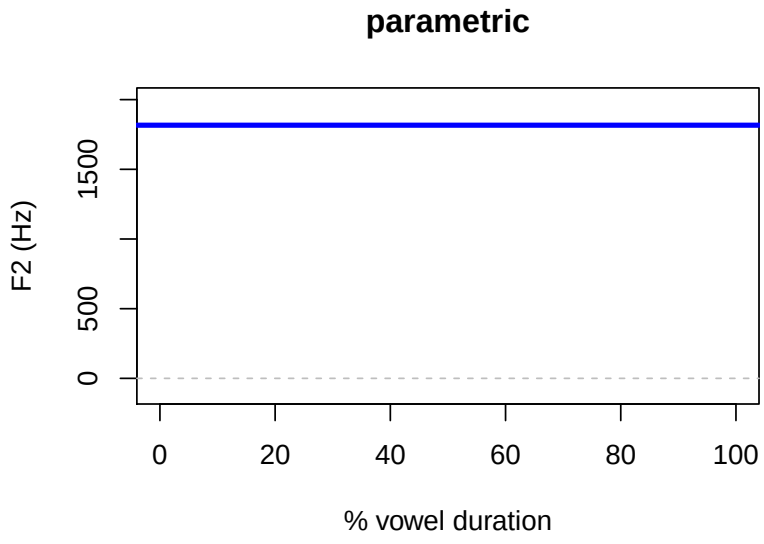
Dynamic analysis & GAMMs

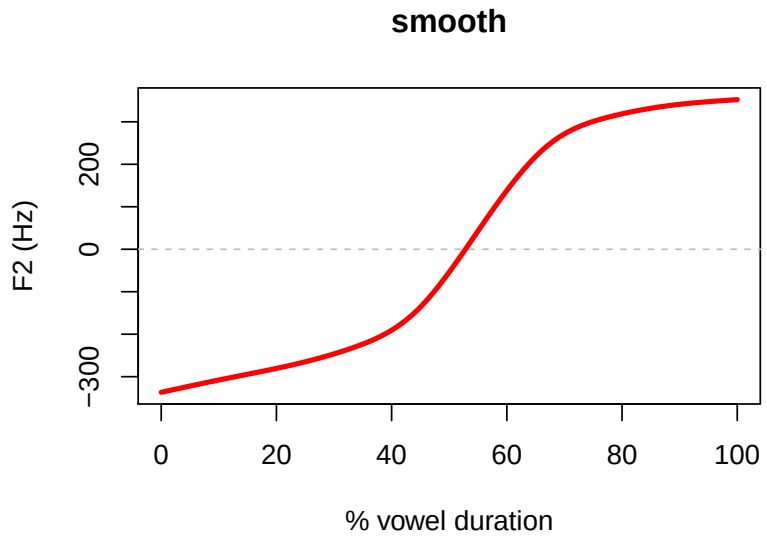
- (generalised) additive models:

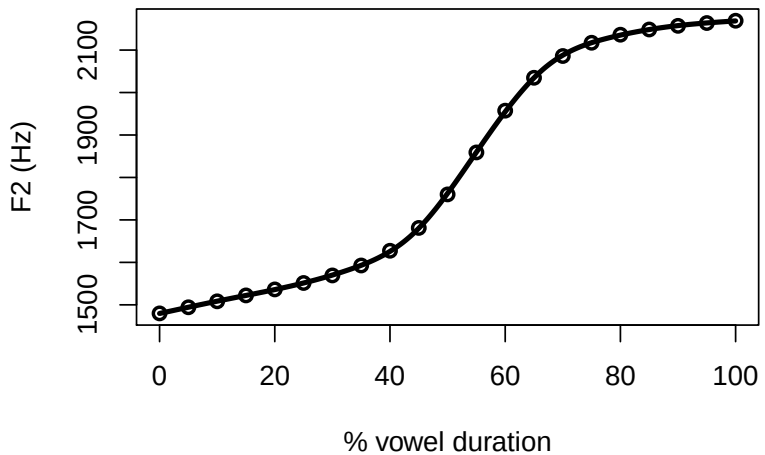
$$\mathbf{y} = \alpha + \beta_1 \mathbf{x_1} + \ldots + f_1(\mathbf{x_1}) + f_2(\mathbf{x_2}) + \ldots + \epsilon \quad (4)$$

- parametric terms: (usually) capturing the *height* of trajectories
- smooth terms: capturing the *shape* of trajectories

Dynamic analysis & GAMMs

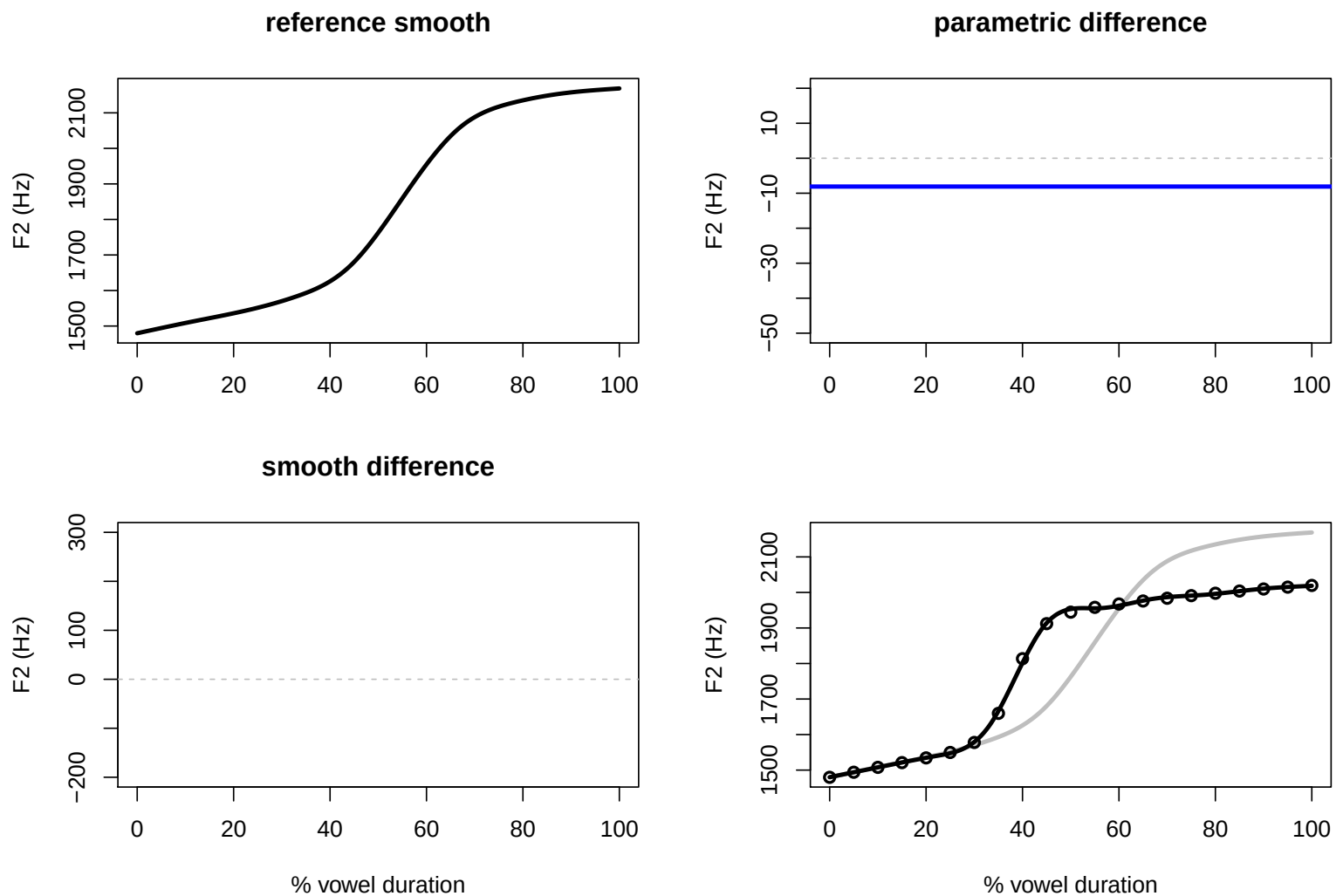






Dynamic analysis & GAMMs

- an important concept: difference smooths

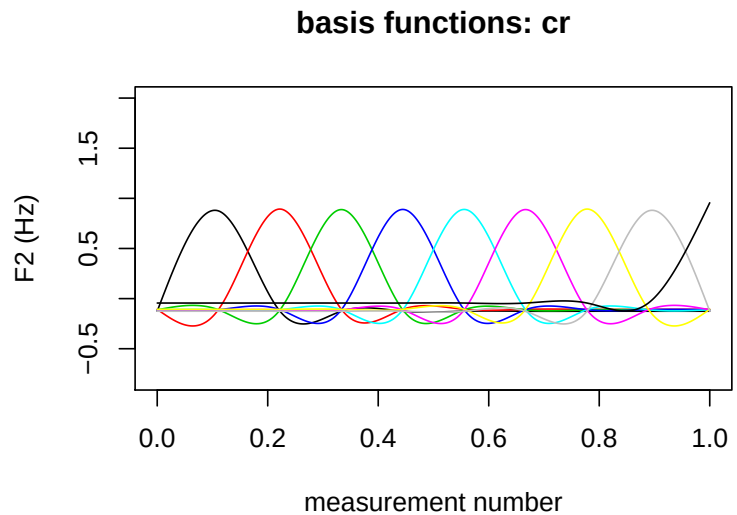


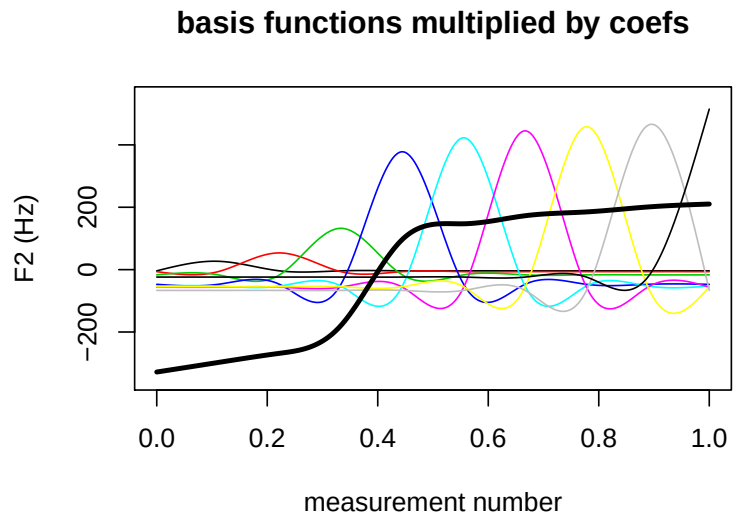
Dynamic analysis & GAMMs

- parametric terms are the same as in linear regression

→ how are smooth terms defined/estimated?

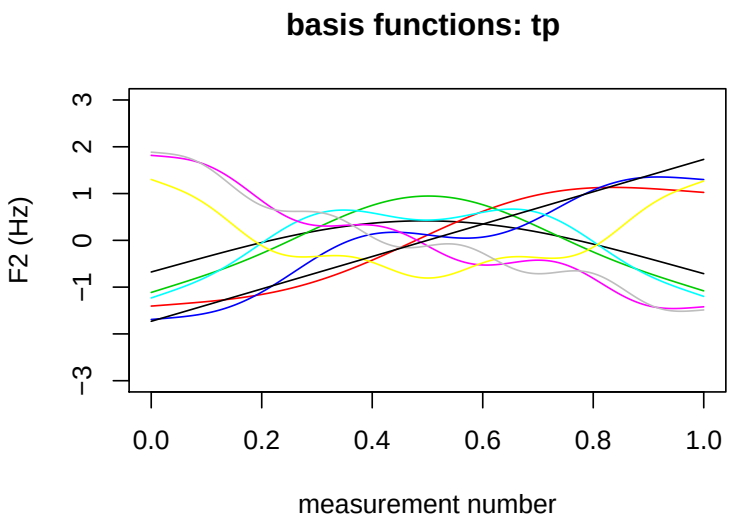
- key concept: basis functions (+ smoothing penalty. . .)
- example 1: cubic regression splines

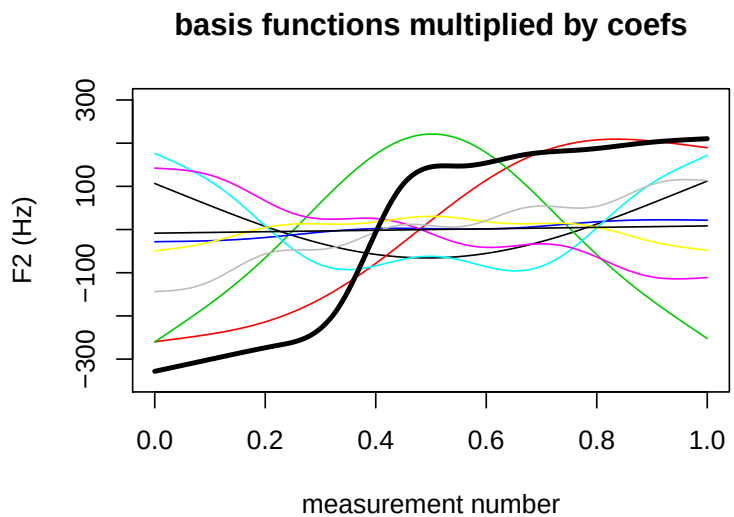




Dynamic analysis & GAMMs

- key concept: basis functions
- example 2: thin plate regression splines





Dynamic analysis & GAMMs

- can vary...
  - type of basis functions
  - number of basis functions
    (more $\rightarrow$ more potential wiggliness)

Dynamic analysis & GAMMs

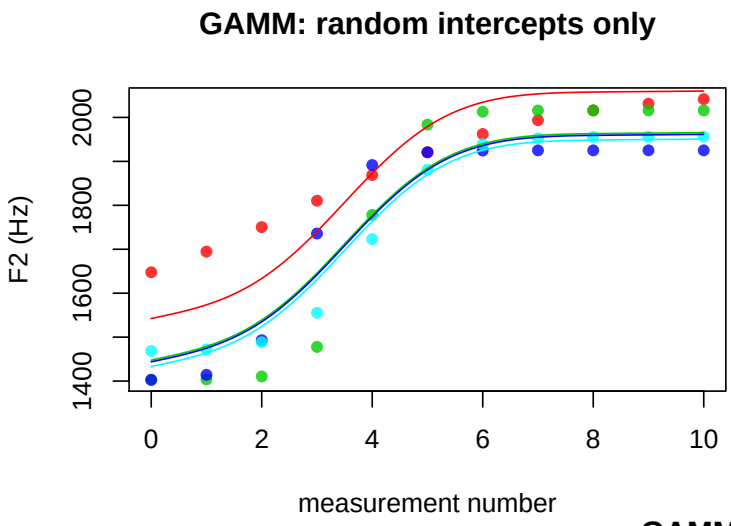
- final point about GAMMs: dependencies within trajectories

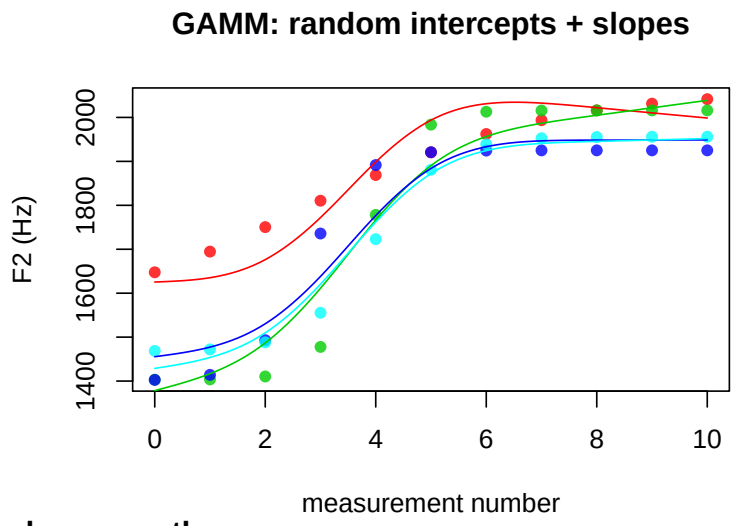
$\rightarrow$ autocorrelation in the residuals

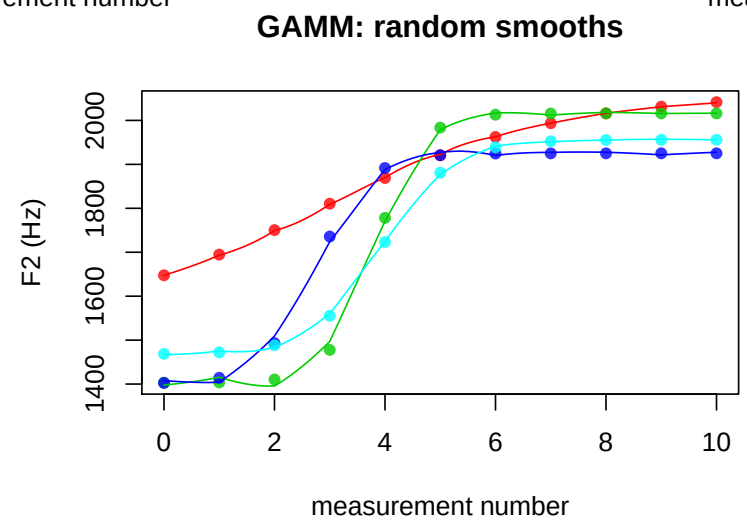
- two potential solutions ...

Dynamic analysis & GAMMs

1. random intercepts / slopes / smooths







Dynamic analysis & GAMMs

2. ~~black magic~~ autoregressive (AR) error model

Dynamic analysis & GAMMs

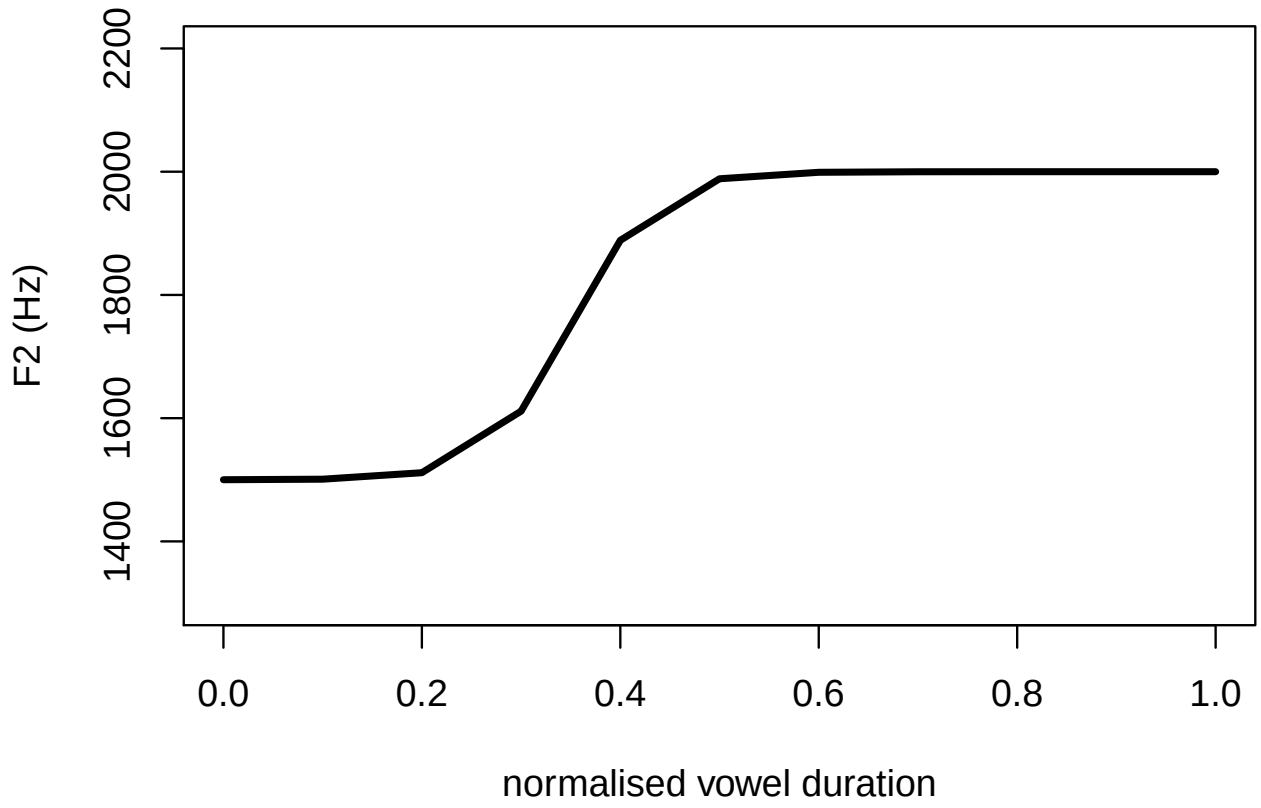
- today's questions:
  - what random structures to use?
    - intercepts? slopes? smooth?
    - what type of random smooth? how many basis functions?
    - AR model?
  - how to test for significance?
    - focus on parametric? smooth? both?
    - visual methods?
- answers provided in the form of type I and type II error simulations

Simulations

- fake data simulations along the lines of Barr et al. (2013)
- scenario:
  - two words realised with same vowel
  - a speaker reads each of them 25 times
  - 11 evenly spaced points along F2 trajectories
  - 5000 repetitions $\times$ 54 different GAMM setups ($\sim$3 months of computing time)
  - appropriate statistical test: 5% false positive rate

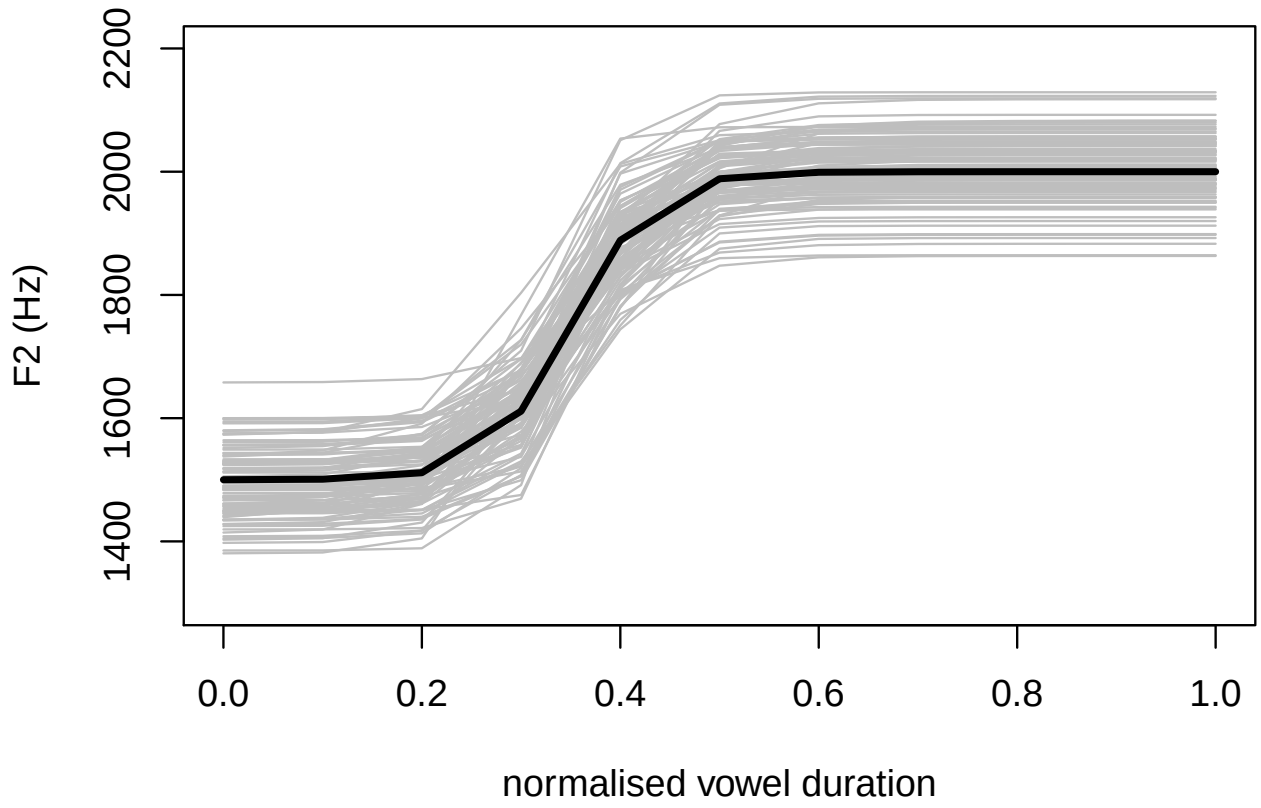
Simulations

- the underlying curve:



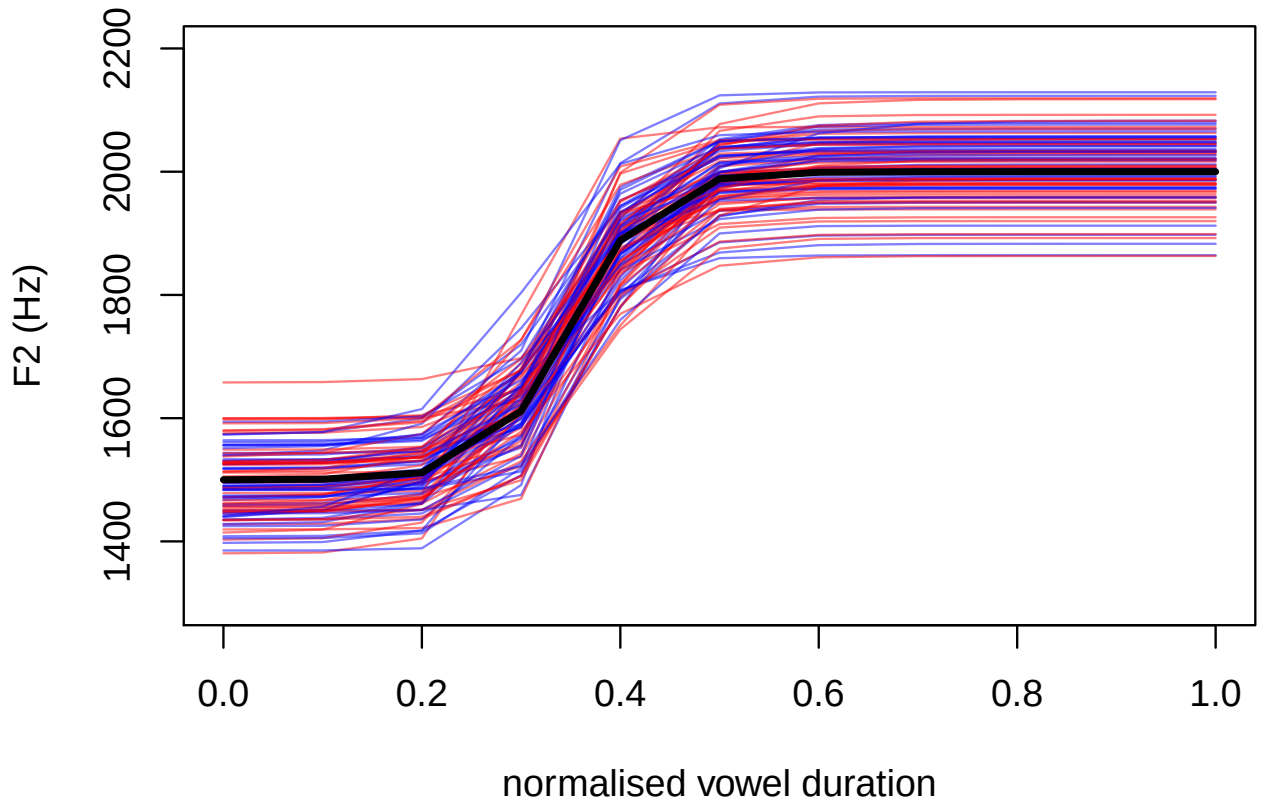
Simulations

- sample 50 random curves with some variation:



Simulations

- assigned to two words randomly:



Simulations

- let's dig in... how about random effects?
- significance based on parametric difference term:

	NONE	INTERCEPT	+ SLOPE
NO AR	0.39	0.05	0.00
AR	0.10	0.05	0.01

- significance based on smooth difference term:

	NONE	INTERCEPT	+ SLOPE
NO AR	0.24	0.38	0.16
AR	0.09	0.19	0.10

Simulations

- how about random smooths?
- significance based on smooth difference term:

	CR 3	CR 5	CR 10	TP 3	TP 5	TP 10
NO AR	0.12	0.10	0.09	0.10	0.06	0.06
AR	0.09	0.10	0.09	0.08	0.07	0.05

Simulations

- interim summary:
  - TP random smooths with sufficient wiggliness are ideal
  - autoregressive error model does fairly well ('only' double error rate)
  - not necessarily generalisable to other types of curves...

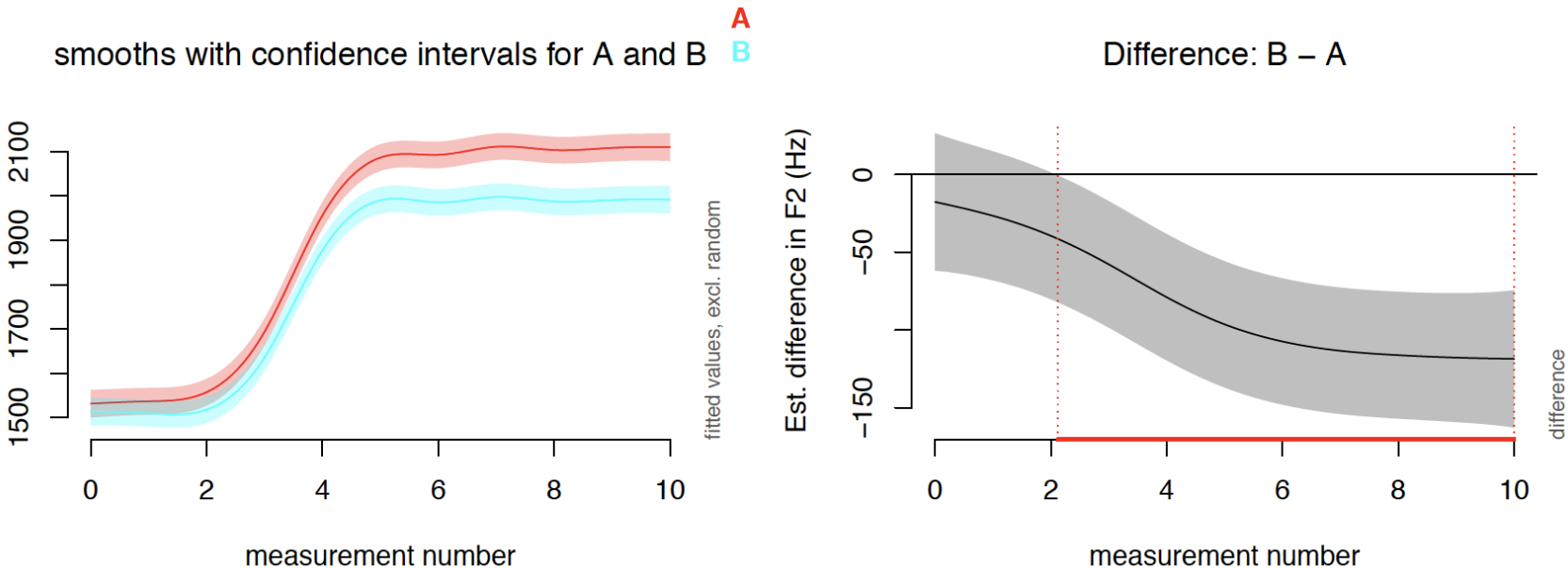
Simulations

- how about different methods of significance testing?

	TYPE I		TYPE II	
	AR	TP 10	AR	TP 10
PARAMETRIC ONLY	0.10	0.05	0.84	0.72
SMOOTH ONLY	0.09	0.06	0.66	0.52
EITHER	0.17	0.10	–	–
MODEL COMPARISON	0.06	0.03	0.87	0.71

Simulations

- ‘visual’ methods of significance testing:
  1. no overlap between confidence intervals of individual trajectories
  2. confidence interval around estimated difference excludes zero



Simulations

- question: is a significant difference along e.g. 1% of the trajectory 'overall' significant?
- using different % cut-off values for overall significance

	TYPE I		TYPE II	
	AR	TP 10	AR	TP 10
NO OVERLAP: 10%	0.03	0.01	0.81	0.64
NO OVERLAP: 20%	0.01	0.01	0.77	0.61
NO OVERLAP: 50%	0.00	0.00	0.54	0.36
EXCLUDES 0: 10%	0.22	0.13	0.95	0.90
EXCLUDES 0: 20%	0.19	0.10	0.95	0.89
EXCLUDES 0: 50%	0.08	0.04	0.87	0.77

Recommendations based on simulations

- if possible, use random smooths by trajectory with a sufficient number of basis functions (small data sets, $<$ 500 trajectories)
- otherwise, use an AR model, but be careful about *p*-values close to 0.05
- use model comparison (dropping parametric and smooth simultaneously) unless your hypothesis is specifically about parametric / smooth terms
- visual comparison should only be used as a supportive tool *after* significance testing via model comparison

Additional stuff

- key concept: smoothing penalty