Additive Models Represented as Mixed Linear Models

This is another class of models with an extensive literature, which can be fit into the MLM framework

- at the price of sacrificing some techniques in that literature but
- sacrificing little if any modeling power.

This illustrates the "modularity of spline models" (RWC Sec. 12.3.1):

> *[C]oncepts like main effects, interaction effects, generalized regression, and the mixed model formulation with smoothing parameter selection by REML can be viewed as modules and put together into an almost endless variety of statistical models. . . . [O]ne can easily tailor a model to a specific application. (pp. 226–227)*

I'd say this is an understatement.

Example: Mechanical properties of pig jawbone

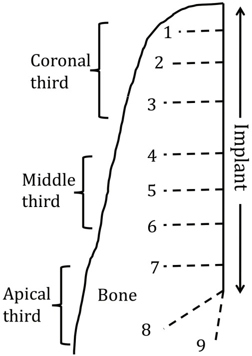
Dental implants are placed in human jaw bones and used to support prosthetic teeth or dentures.

Standard of care [in early '00s] was to place the implants and wait several months for the bone to heal and osseointegrate the implant.

Problem: If bone is not stressed, it tends to lose calcium or volume or both; older people especially can lose a lot of bone in 4 months.

Proposed solution (C-C Ko et al): Load the implants much sooner.

Here is the dataset



7 mini-pigs had a tooth extracted and replaced by an implant, which was covered by a device that simulated chewing.

After loading, each pig was sacrificed and samples of jawbone next to the implant were measured for elastic modulus.

Each sample had 9 transects.

Along a transect, bone was measured every 15 μm if bone was present, to 1500 μm.

I'll show analyses of one pig's data (Whitey).

Scientific questions:

- How does elastic modulus depend on distance from the implant?
- Does that relationship differ for the coronal, middle, and apical thirds or for the different transects within those sections?
- Considering all the pigs, how does elastic modulus depend on healing time?

I'll show a series of fits to demonstrate the modularity of additive models expressed as MLMs, and to show some things that can happen.

This sequence of fits is not intended to be an efficient data analysis.

(In Chang et al. 2003, I reduced distance to a 4-level categorical factor.)

What is an additive model?

A penalized spline fits a smooth function of a scalar x: $y_i = f(x_i) + \epsilon_i$.

An additive model fits a smooth function of a vector $\mathbf{x}$: $y_i = f(\mathbf{x}_i) + \epsilon_i$.

If the x_k are continuous, an additive model can be defined as

$$y_i = \alpha + \sum_{k=1}^{p} f_k(x_{ki}) + \epsilon_i,$$

where each $f_k(x_k)$ is a smooth function of its argument –

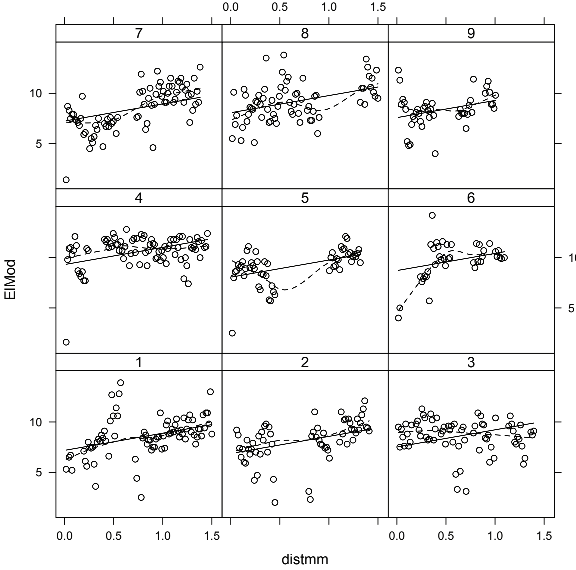
this is the simplest such model that allows a flexible shape for $f(\mathbf{x}_i)$.

If x_k is categorical, then $f_k(x_k)$ takes a different value for each x_k value.

Usually some further condition is required to identify the model.

A simple additive model, fit to Whitey's data

Dashes are loess fits; straight lines are the additive fit.



The previous slide's simple additive model

Let $x_{ij} = 15j$, the distance from the implant to location j^{th} on transect i:

This model for elastic modulus (EM) is

$$EM_{ij} = \sum_{l=1}^{9} \beta_l \mathtt{I}(l = i) + \beta_d x_{ij} + \epsilon_{ij},\ \epsilon_{ij} \sim N(0, \sigma_e^2).$$

Each transect has its own intercept, with a common slope in distance x_{ij}.

It's an <u>additive</u> model: no transect $\times$ distance interaction.

Everything is a fixed effect: the fit has 10 DF: 9 intercepts + 1 slope.

Modularity of splines: unplug "$\beta_d x_{ij}$", plug in a spline.

Plug in: A spline in distance

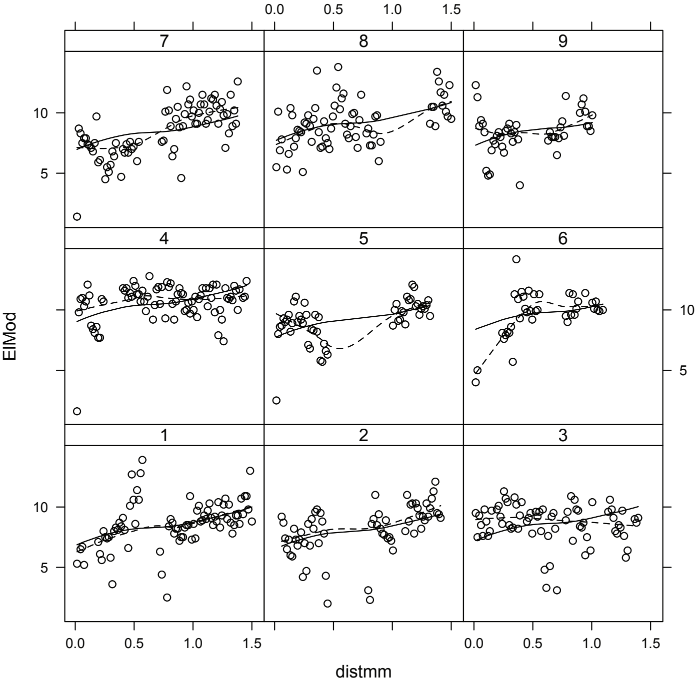
Use the truncated quadratic basis, 25 equally-spaced knots:

$$
\begin{aligned}
EM_{ij} &= \sum_{l=1}^{9} \beta_l \mathtt{I}(l = i) \\
&\quad + \beta_{21} x_{ij} + \beta_{22} x_{ij}^2 + \sum_{k=1}^{K} u_k (x_{ij} - \kappa_k)_+^2 + \epsilon_{ij}, \\
&\quad u_k \sim N(0, \sigma_s^2), \epsilon_{ij} \sim N(0, \sigma_e^2)
\end{aligned}
$$

It's still an additive model (no interaction).

The model fit has 11.9 DF (1.9 more than the simple model):

- 9 DF intercepts (fixed effect)
- 2.9 DF for the function of distance:
  - 2 DF for the linear & quadratic fixed effects,
  - 0.9 DF for the spline random effects.



7
8
9
4
5
6
1
2
3
ElMod
distmm
0.0
0.5
1.0
1.5
5
10

Plug in: Intercepts as a random effect

The transect-specific intercepts look similar; shrink them and save DF?

$$\begin{aligned} EM_{ij} &= \beta_0 + u_{0i} \\ & \quad + \beta_{11} x_{ij} + \beta_{12} x_{ij}^2 + \sum_{k=1}^{K} u_{1k} (x_{ij} - \kappa_k)_+^2 + \epsilon_{ij}, \\ & \quad u_{0i} \sim N(0, \sigma_{s0}^2), u_k \sim N(0, \sigma_{s1}^2), \epsilon_{ij} \sim N(0, \sigma_e^2) \end{aligned}$$

It's still an additive model (no interaction).

The fit hardly changes! It has 11.3 DF (0.6 less than the last model):

- 3 DF for fixed effects (1 intercept, 2 for spline)
- 7.4 DF for intercept random effect ⇒ 8.4 total DF for intercepts.
- 0.9 DF for the spline-in-distance random effect.

Plug in: Intercepts as a random walk down the transects

Shrink neighboring transect intercepts toward each other:

$$
\begin{aligned}
EM_{ij} &= \beta_0 + \xi_i \\
&\quad + \beta_{11} x_{ij} + \beta_{12} x_{ij}^2 + \sum_{k=1}^{K} u_{1k} (x_{ij} - \kappa_k)_+^2 + \epsilon_{ij}, \\
&\quad u_k \sim N(0, \sigma_{s1}^2), \epsilon_{ij} \sim N(0, \sigma_e^2)
\end{aligned}
$$

where the ξ_i are smoothed using a random walk:

$\xi_i = \xi_{i-1} + \delta_i, i = 2, \ldots, 9$, with δ_i iid $N(0, \sigma_{s0}^2)$.

We have two intercepts, β_0 and ξ_1; set $\xi_1 = 0 \Rightarrow \xi_i = \sum_{l=2}^{i} \delta_i$.

The fit hardly changes! It has 10.8 DF (0.5 less than the last model):

- 3 DF for fixed effects (1 intercept, 2 for spline)
- 6.9 DF for intercept random effect $\Rightarrow$ 7.9 total DF for intercepts.
- 0.9 DF for the spline-in-distance random effect.

Plug in: Fixed effect for coronal v. middle v. apical

$$
\begin{aligned}
EM_{ij} &= \beta_c \mathtt{I}(\text{coronal}) + \beta_m \mathtt{I}(\text{mid}) + \beta_a \mathtt{I}(\text{apical}) \\
&\quad + \xi_i + \beta_{11} x_{ij} + \beta_{12} x_{ij}^2 + \sum_{k=1}^{K} u_{1k}(x_{ij} - \kappa_k)_+^2 + \epsilon_{ij}, \\
&\quad u_k \sim N(0, \sigma_{s1}^2), \epsilon_{ij} \sim N(0, \sigma_e^2)
\end{aligned}
$$

where the ξ_i are still smoothed using a random walk:

$\xi_i = \sum_{l=2}^{i} \delta_i, i = 2, \ldots, 9$, with δ_i iid $N(0, \sigma_{s0}^2)$.

The fit is nearly identical to the preceding:

- 5 DF for fixed effects (3 for transects, 2 for spline)
- 4.9 DF for intercept random effect $\Rightarrow$ 7.9 total DF for intercepts.
- 0.9 DF for the spline-in-distance random effect.

This fit has just moved 2 DF from the intercept RE to the section FE.

Plug in: Serially correlated errors within transect

Use the same model as the last slide except the error covariance:

$$\mathbf{R} = \begin{bmatrix} \Sigma_1 & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \Sigma_2 & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Sigma_9 \end{bmatrix}, \text{ where } \begin{cases} \Sigma_{i,jl} = \sigma_e^2 \exp(-\theta d_{jl}) \\ d_{jl} = 1000|x_{ij} - x_{il}| \ \mu\text{m} \end{cases}$$

High autocorrelation: $\hat{\theta} = 42.6$, so adjacent measures have $\hat{\rho} = 0.53$.

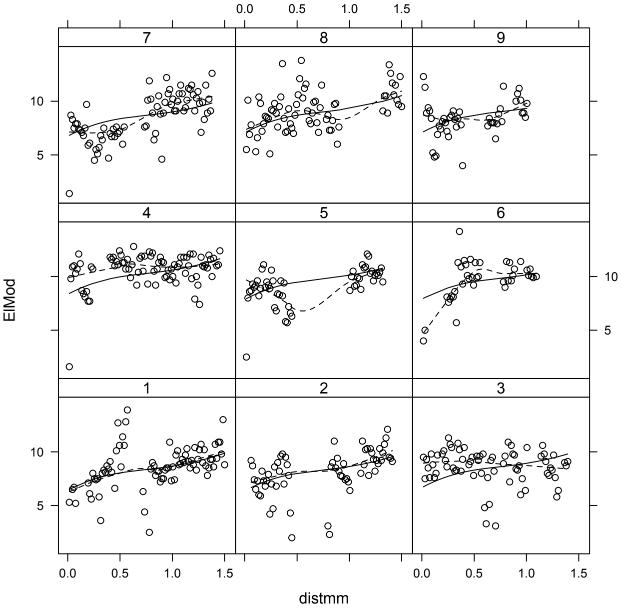
This changes the fit, to 8.2 total DF:

- 5 DF for fixed effects (3 for transects, 2 for spline)
- 2.4 DF for intercept random effect $\Rightarrow$ 5.4 total DF for intercepts.
- 0.8 DF for the spline-in-distance random effect.

We've shaved 3.6 DF off the intercepts and 0.1 off the spline.

Nobody knows why this happens.

Well . . . it doesn't look that different



Non-additive model: Categorical-by-continuous interaction

This next model fits a <u>different</u> smooth for each transect.

$$
\begin{aligned}
EM_{ij} &= \sum_{l=1}^{9} I(l=i)\left\{\beta_{0i}+\beta_{1i}x_{ij}+\beta_{2i}x_{ij}^2\right\} \\
&+ \sum_{l=1}^{9} I(l=i)\left\{\sum_{k=1}^{K} v_k^l (x_{ij}-\kappa_k)_+^2\right\} + \epsilon_{ij},
\end{aligned}
$$

$$\mathbf{v}^l = (v_1^l, \ldots, v_K^l)', l = 1, \ldots, 9 \sim \text{iid } N(\mathbf{u}, \Sigma_{vl}) \text{ for } \mathbf{u} = (u_1, \ldots, u_K)'$$

Alternatively,

$$
\begin{aligned}
EM_{ij} &= \sum_{l=1}^{9} I(l=i)\left\{\beta_{0i}+\beta_{1i}x_{ij}+\beta_{2i}x_{ij}^2\right\} \\
&+ \sum_{k=1}^{K} u_k (x_{ij}-\kappa_k)_+^2 \qquad \text{distance main effect} \\
&+ \sum_{l=1}^{9} I(l=i)\left\{\sum_{k=1}^{K} w_k^l (x_{ij}-\kappa_k)_+^2\right\} + \epsilon_{ij}, \qquad \text{interaction}
\end{aligned}
$$

In the main-effect & interaction formulation:

$$\begin{aligned} EM_{ij} &= \sum_{l=1}^{9} I(l=i)\left\{\beta_{0i} + \beta_{1i}x_{ij} + \beta_{2i}x_{ij}^2\right\} \\ &+ \sum_{k=1}^{K} u_k(x_{ij} - \kappa_k)_+^2 \\ &+ \sum_{l=1}^{9} I(l=i)\left\{\sum_{k=1}^{K} w_k^l(x_{ij} - \kappa_k)_+^2\right\} + \epsilon_{ij}, \end{aligned}$$

General model: $\mathbf{w}^l \sim \text{iid } N(\mathbf{0}, \Sigma_{vl})$, $\mathbf{u} \sim N(\mathbf{0}, \Sigma_u)$

This is hopeless without some more structure.

RWC (Sec 12.3): $\Sigma_u = \sigma_u^2 \mathbf{I}_K$ and $\Sigma_{vl} = \sigma_{vl}^2 \mathbf{I}_K$.
Different σ_{vl}^2 in each transect $\Rightarrow$ different smoothness.

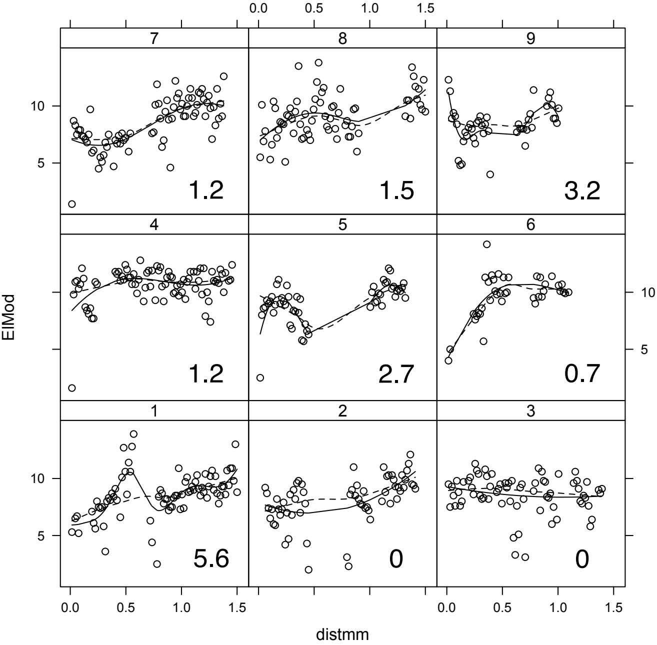
Here's a naive version of this model for the pig data

I used 25 knots per transect; RWC's default is 10 to 21.

The RL maximizer reached a true convergence and a false convergence (whatever that means).

This fit has total DF 43.2 (compared to 8.2 for non-diagonal $\mathbf{R}$):

- 27 DF for FEs (intercept, linear, quadratic per transect)
- 16.2 DF for interaction REs, split among the 9 transects
- $\approx$0 DF for the distance main effect.



7
8
9
4
5
6
1
2
3
1.2
1.5
3.2
1.2
2.7
0.7
5.6
0
0
ElMod
distmm

Many less naive models are possible

- Use fewer knots.
  - It's OK (if less convenient) to use different numbers of knots
    - for distance main effect and interaction;
    - for interaction smooth in each transect.
  - Any of these choices is still a MLM.

- Set $\sigma^2_{vl} \equiv \sigma^2_v$; this is simple but over- and under-smooths
  some transects.

- Let $\sigma^2_{vl} \sim$ inverse gamma distribution with unknown parameters.

  This shrinks the σ^2_{vl} toward a common value.

  This is no longer a MLM but a Bayesian analysis is straightforward.