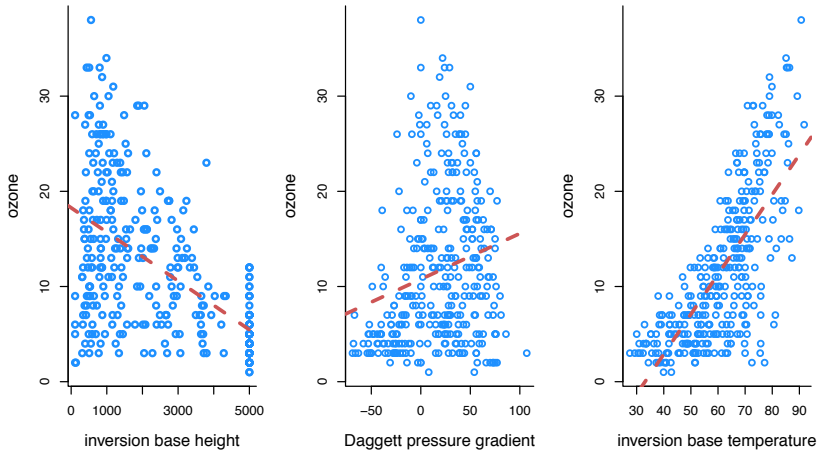
37458

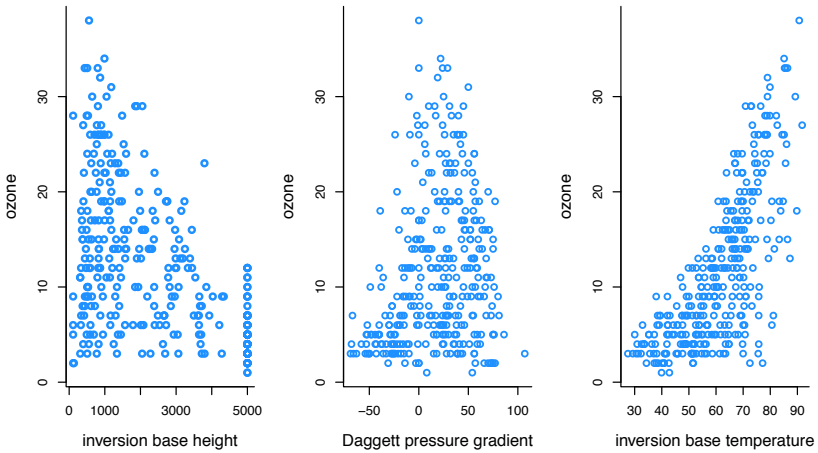
Advanced Bayesian Methods

Generalized Additive Models

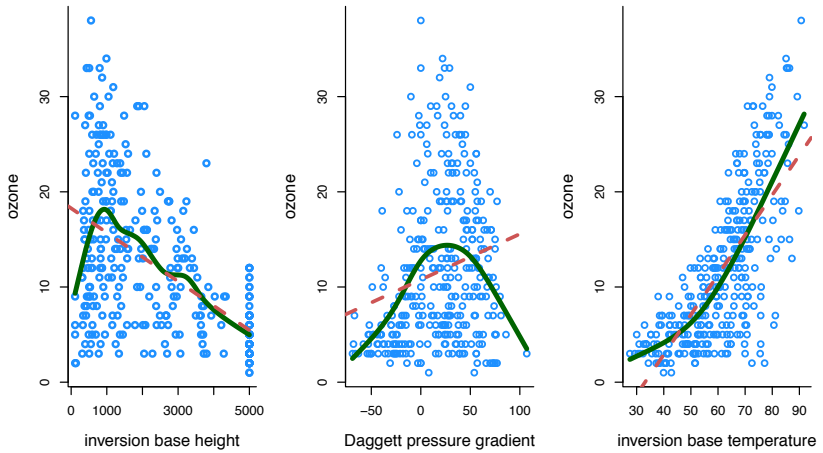
Linear Model Fits



Data from Ozone Study from California, U.S.A.



Generalized Additive Model Improvement



Linear Model Versus (Generalized) Additive Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

$$y_i = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \varepsilon_i$$

f_1, f_2 and f_3 are unspecified "smooth" functions chosen by the data.

Exension to Three Predictors

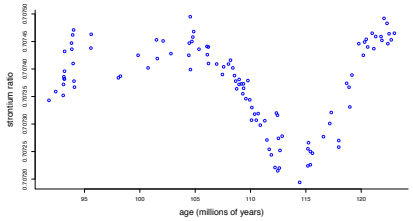
$$f_1(x_1) = \beta_1 x_1 + \sum_{k=1}^{K_1} u_{1k} z_{1k}(x_1), \quad u_{1k}|\sigma^2_{u1} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{u1}),$$

$$f_2(x_2) = \beta_2 x_2 + \sum_{k=1}^{K_2} u_{2k} z_{2k}(x_1), \quad u_{2k}|\sigma^2_{u2} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{u2}),$$

$$f_3(x_3) = \beta_3 x_3 + \sum_{k=1}^{K_3} u_{3k} z_{3k}(x_1), \quad u_{3k}|\sigma^2_{u3} \overset{\text{ind.}}{\sim} N(0, \sigma^2_{u3}),$$

Tricking Mixed Models into Penalized Spline Fitting

First consider the single predictor case:

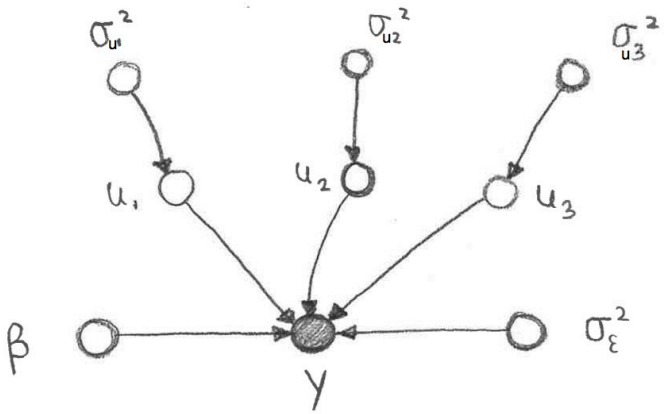


Spline model: $y = f(x_i) + \varepsilon_i$ where $f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k z_k(x)$

The $z_1(\cdot), \ldots, z_K(\cdot)$ are suitable spline basis functions.

TRICK TO AVOID OVERFITTING: $u_k \overset{\text{ind.}}{\sim} N(0, \sigma^2_u)$ (Sec. 2.7 of pen. spline notes)

Just Another Bayesian Model with this DAG



Laboratory 3 later today shows how to fit using Stan.

Look at ozoneCalifViaStan.Rs

Bayesian versus Non-Bayesian GAM Analysis

As we will see in Laboratory 3, the Bayesian inference engine approach is much slower.

HOWEVER the Bayesian engine approach:

- has more accurate inference for binary response data,
- can handle arbitrary response distributions,
- allows non-constant variances,
- can handle missing data properly.

Laboratory 4 will provide illustation in a few weeks from now.

Non-Bayesian Generalized Additive Model Software

Most generalized additive model analyses in R are non-Bayesian, with packages such as:

- gam
- mgcv
- VGAM
- gamlss

Laboratory 3 contains illustration using mgcv.

ASIDE: Bayesian Analysis with Multiple Chains

All BUGS and Stan analyses to date have used

single Markov chain Monte Carlo samples ("single chains")

It is often advocated to use

multiple chains

and then the

Brooks-Gelman-Rubin convergence diagnostic.

See pages 50-51 of the *Penalized Spline* notes...

9/3=3

$\Longrightarrow$ it's time for Laboratory 3

♦ ♦ ♦