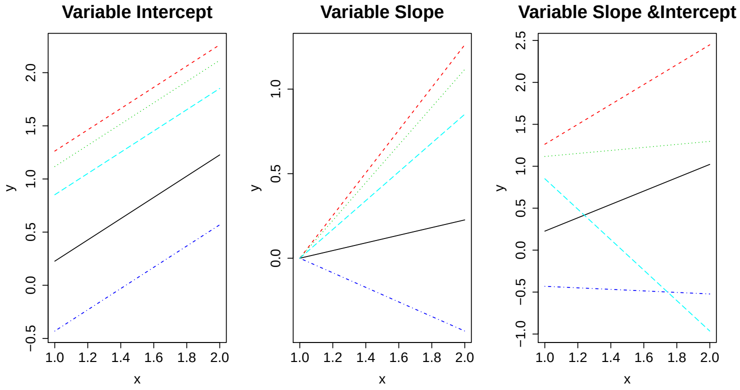
Multilevel Models & Timeseries Modeling

Types of Multilevel Models



Variable Intercept Models Useful with Group Level Predictors

$$y_i = \alpha_{j[i]} + \beta_i x_i + \epsilon_{ij}$$

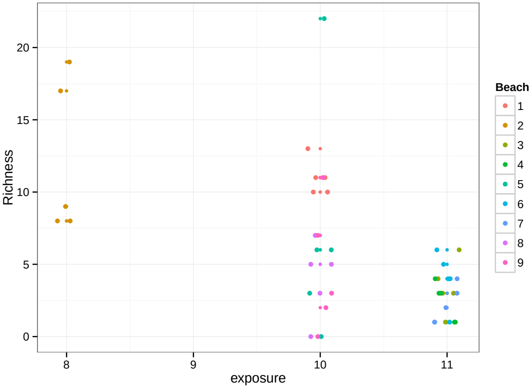
$$\alpha_{j[i]} \sim N(\mu_\alpha + x_j, \sigma_\alpha^2)$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

where i = individual sample, j = group

Each Site has a Unique Exposure - How does it Affect Species Richness?

Each Site has a Unique Exposure - How does it Affect Species Richness?



Data from Zuur et al. 2009

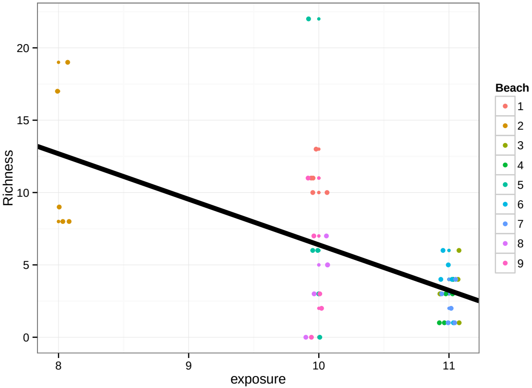
A Variable Intercept Model for Wave Exposure

```
varIntExposure<- lme(Richness ~ exposure,
                     random =~ 1 |
                       Beach, data=rikz)
anova(varIntExposure, type="m")

#             numDF denDF F-value p-value
# (Intercept)     1    36   21.28  <.0001
# exposure        1     7   15.49  0.0056
```

Plot Fit Using Extracted Components

```
adf <- as.data.frame(t(fixef(varIntExposure)))
#
names(adf)[1] <- "intercept"
#
p+  geom_abline(data=adf,
              mapping=aes(intercept=intercept, slope=exposure),
              lwd=2)
```

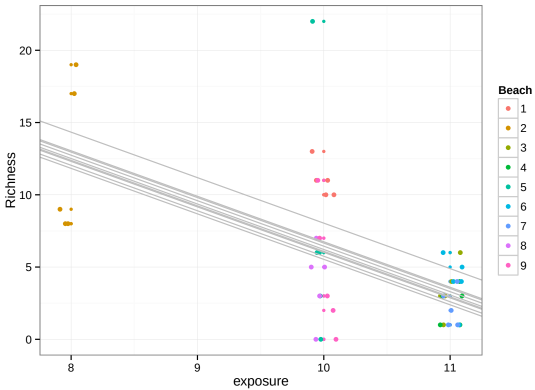
Plot Fit Using Extracted Components



Plot Fit Using Extracted Components

```
adf2 <- as.data.frame(coef(varIntExposure))
#
names(adf2)[1] <- "intercept"
#
p+ geom_abline(data=adf2,
               mapping=aes(intercept=intercept, slope=exposure),
               color="grey")
```

Plotting is a Wee Bit Tricksy...



Variable Slope-Intercept Model with No Group Level Predictors

$$y_i = \alpha_{j[i]} + \beta_{[j]i} X + \epsilon_{ij}$$

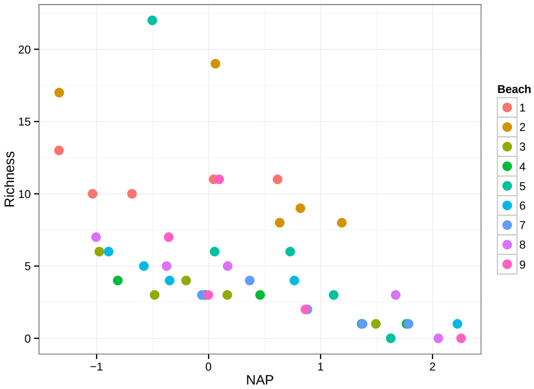
$$\begin{pmatrix} \alpha_{[i]j} \\ \beta_{[i]j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right)$$

General Protocol for Model Fitting

Variable slope? Intercept? Slope-Intercept? Why do I evaluate Fixed Effects?

1. Start with model with all fixed and random effects that may be important. Evaluate with diagnostics.
2. Evaluate random effects with full model of fixed effects (AIC, χ^2)
3. Evaluate fixed effects with reduced random effects (F Tests)
4. Model diagnostics again...
5. Draw inference from model

How Important is Tide Height?



Three Models with Different Random Effects

```
varSlope <- lme(Richness ~ NAP,
                random =~ 0 + NAP | Beach, data=rikz)
#
varInt<- lme(Richness ~ NAP,
             random =~ 1 | Beach, data=rikz)
#
varSlopeInt <- lme(Richness ~ NAP,
                   random =~ 1 + NAP | Beach, data=rikz)
```

Does Slope Vary Randomly?

```
ranef(varSlope)

#           NAP
# 1 -4.921e-10
# 2  2.659e-09
# 3  2.901e-09
# 4 -3.136e-10
# 5 -7.361e-09
# 6  5.445e-09
# 7 -4.260e-09
# 8  2.708e-09
# 9 -1.288e-09
```

SD in Variable Slope Model is Small

```
summary(varSlope)

....

# Random effects:
#  Formula: ~0 + NAP | Beach
#                NAP Residual
# StdDev: 0.0001139     4.16
#
....
```

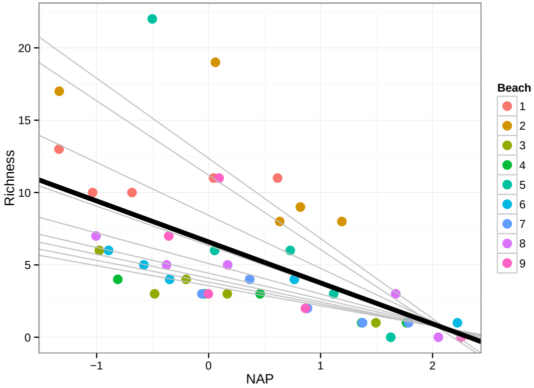
Evaluation of Different Random Effects Models

```
anova(varSlope, varSlopeInt)

#              Model df   AIC   BIC logLik   Test L.Ratio
# varSlope         1  4 260.2 267.2 -126.1
# varSlopeInt      2  6 244.4 254.9 -116.2 1 vs 2   19.82
#              p-value
# varSlope
# varSlopeInt  <.0001

anova(varSlopeInt, varInt)

#              Model df   AIC   BIC logLik   Test L.Ratio
# varSlopeInt      1  6 244.4 254.9 -116.2
# varInt           2  4 247.5 254.5 -119.7 1 vs 2   7.096
#              p-value
# varSlopeInt
# varInt        0.0288
```

Evaluation of Fixed Effects

```
anova(varSlopeInt, type="m")

#               numDF denDF F-value p-value
# (Intercept)       1    35   27.14  <.0001
# NAP               1    35   15.32   4e-04
```

Final Model



Exercise: RIKZ Tide Height and Shoreline Angle

- Evaluate the effect of angle1 (sample angle) & NAP on Richness
- Note: You already know the slope of the NAP relationship doesn't vary randomly
- Check for a NAP*angle1 interaction

nlme versus lme4

- nlme - can work like nls for flexible nonlinear specification
- nlme - can accomodate specified correlation structures
- lmer - can fit more complex models
- lmer - can fit Generalized Linear Mixed Models (GLMM)

lmer for a GLMM

```
library(lme4)
angI_lmer <- lmer(Richness ~ angle1*NAP + (1  | Beach),
                  data=rikz, family=poisson(link="log"))
Anova(angI_lmer)

# Analysis of Deviance Table (Type II Wald chisquare tests)
#
# Response: Richness
#             Chisq Df Pr(>Chisq)
# angle1       2.60  1      0.107
# NAP         41.36  1    1.3e-10
# angle1:NAP  4.95  1      0.026



#nestNest
data(Oats)
Oats$nitro <- ordered(Oats$nitro)
```

Modeling Error Structures with Generalized Least Squares

$$Y = \beta X + \epsilon$$

Mixed models can handle clustered data, but what about other violations assumptions about ϵ?

1) Error variance is not constant

2) Error is temporally or spatially autocorrelated

Modeling Error Structures with Generalized Least Squares

$$Y = \beta X + \epsilon$$

Mixed models can handle clustered data, but what about other violations assumptions about ϵ?
1) Error variance is not constant
2) Error is temporally or spatially autocorrelated

Generalized Least Squares - (of which OLS is a special case)

What's in that Epsilon?

$$\epsilon \sim N(0, \sigma^2)$$

What's in that Epsilon?

$$\epsilon \sim N(0, \sigma^2)$$

if n=3...

$$\epsilon \sim N\left(0, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right)$$

What if $sigma^2$ is not Constant?

$$\epsilon \sim N\left(0, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right)$$

for n=3

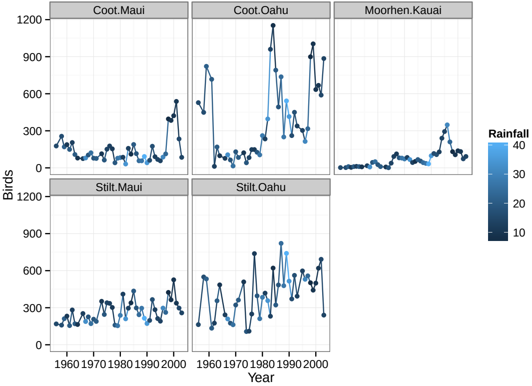
Commonly, we weight by 1/SD of a response variable when we know something about measurement precision. E.g. in R `lm(y~x, weight=1/sd(y))`. Other options include modeling σ^2 explicitly as a response. In R we use `varFixed` or other functions in conjunction with the `weights` argument with gls or lme.

What if the off diagonals are not 0?

$$\epsilon \sim N\left(0, \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}\right)$$

- Temporal or Physical distance between sampling points can induce correlation between data points.
- If we have measured EVERY relevant variable, we may account for this, but not always.

Enter Repeated Measures & Time Series



Data from Zuur et al. 2009

If this was Just Repeated Measures...

```
allbirds_repeated <- lme(Birds ~ Rainfall, random= ~1|Site,
                         data=allbirds)
```

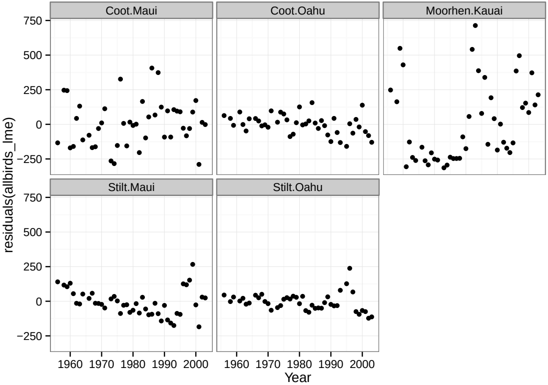
No temporal autocorrelation. No effect of time.
Assumes variation in time is purely random.

But, We Want to Look for a Temporal Trend

```
allbirds_lme <- lme(Birds ~ Rainfall + Year, random= ~1|Site,
                    data=allbirds)
```

Note: Time could have had a nonlinear effect, could have interacted with Rainfall, and could have been a factor if we didn't want to assume a functional form to the time effect.

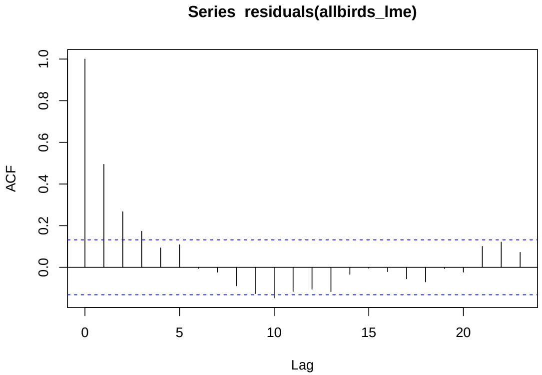
Temporal Autocorrelation in the Residuals



Need to examine cor$(X, X_t - \tau)$ to be certain.

Autocorrelation of Residuals

```
acf(residuals(allbirds_lme))
```



We Must Incorporate Autcorrelation into ϵ

$$cor(\epsilon) \sim N\left(0, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

Alternatives?

We Must Incorporate Autcorrelation into ϵ

$$cor(\epsilon) \sim N\left(0, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

Alternatives?

$$\epsilon \sim N\left(0, \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}\right)$$

Compound Symmetric Structure - often too simple

Autoregressive Error Structure - AR1

$$\epsilon_t = \rho\epsilon_{t-1} + \zeta_t$$

which produces

$$\epsilon \sim N\left(0, \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}\right)$$

for n=3 time steps

Other structures as well (AutoRegressive Moving Average, etc.)

Implementing an AR1 Structure with the Birds Time Series

```
birds_corAR <- corAR1 (form =~ Year)
#
allbirds_lme_ar <- lme(Birds ~ Rainfall + Year, random= ~1|Site,
                       data=allbirds, correlation=birds_corAR)
```

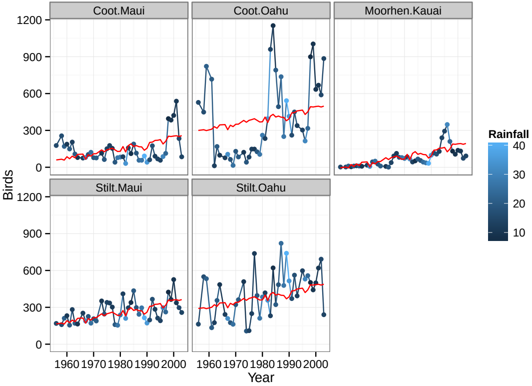
Does AR1 Improve Fit?

```
anova(allbirds_lme_ar, allbirds_lme)

#                 Model df  AIC  BIC logLik   Test L.Ratio
# allbirds_lme_ar     1  6 2815 2836  -1402
# allbirds_lme        2  5 2884 2901  -1437 1 vs 2    70.8
#                 p-value
# allbirds_lme_ar
# allbirds_lme     <.0001
```

But - SS of Predictors Decreased

```
anova(allbirds_lme_ar, allbirds_lme)

#                  Model df  AIC  BIC logLik   Test L.Ratio
# allbirds_lme_ar      1  6 2815 2836  -1402
# allbirds_lme         2  5 2884 2901  -1437 1 vs 2    70.8
#                  p-value
# allbirds_lme_ar
# allbirds_lme      <.0001
```

Predictions Capture Major Trend...

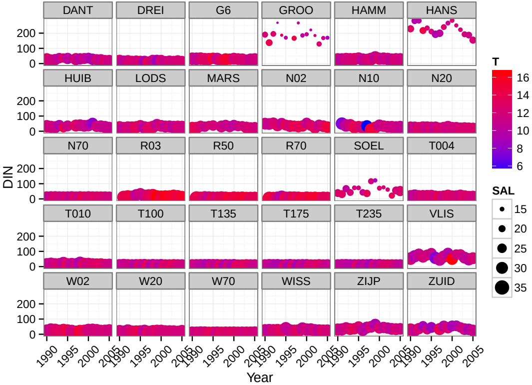



```
# [1] "R^2 = 1 - RSS/TSS = 0.47"
```

Predictions Capture Major Trend...

```
allbirds$fit <- predict(allbirds_lme_ar)
birdPlot + geom_line(color="red", data=allbirds, mapping=aes(y=fit))
r2 <- 1-sum(residuals(allbirds_lme_ar)^2)/sum((allbirds$Birds - mean(al
paste("R^2 = 1 - RSS/TSS = ",round(r2,2), sep="")
```

Exercise: Model DIN, DIP, or CHLFa in the Plankton Data Set



How well can you model the time series with the measurements at hand? Data extrapolated from Zuur et al. 2009

Example: DIN T + SAL + Year

```
plankLME_nocorr<-lme(DIN ~ Year+SAL + T,
                     random =~ 1|Station, data=plankton)
#
plankLME<-lme(DIN ~ Year+SAL + T, random =~ 1|Station,
              correlation=corAR1(form=~Year), data=plankton)
#
plankLME_IS<-lme(DIN ~ Year+SAL + T, random =~ 1+T|Station,
                 correlation=corAR1(form=~Year), data=plankton)
```

Variable Slope Intercept Model

```
anova(plankLME_nocorr, plankLME)

#                  Model df  AIC  BIC logLik   Test L.Ratio
# plankLME_nocorr      1  6 3654 3679  -1821
# plankLME             2  7 3546 3575  -1766 1 vs 2   109.2
#                  p-value
# plankLME_nocorr
# plankLME          <.0001

anova(plankLME, plankLME_IS)

#              Model df  AIC  BIC logLik   Test L.Ratio
# plankLME         1  7 3546 3575  -1766
# plankLME_IS      2  9 3532 3569  -1757 1 vs 2    18.6
#              p-value
# plankLME
# plankLME_IS   1e-04
```

All Predictors Important

```
summary(plankLME_IS)

....

#
# Correlation Structure: ARMA(1,0)
#  Formula: ~Year | Station
#  Parameter estimate(s):
#  Phi1
# 0.615
# Fixed effects: DIN ~ Year + SAL + T
#              Value Std.Error  DF t-value p-value
# (Intercept) 1504.3     353.6 438   4.254   0e+00
# Year          -0.6       0.2 438  -3.554   4e-04
# SAL           -6.3       0.4 438 -16.747   0e+00
# T             -2.4       0.5 438  -5.089   0e+00
....
```

All Predictors Important