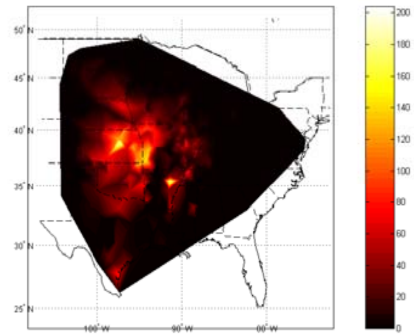
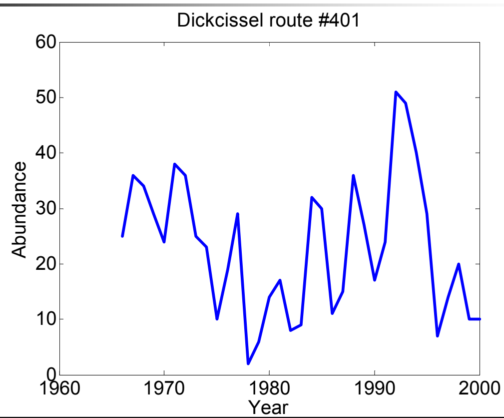
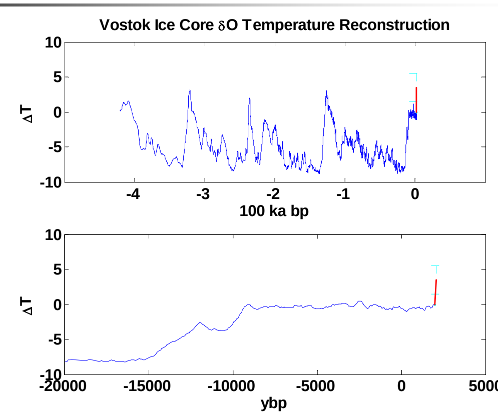
Space and time

Spatial data



Temporal data





Are time & space unique?

- Or can we just do a model:
  - y=f(lat,lon)+ε
  - y=f(temp(lat,lon),precip(lat,lon))+ε
  - y=f(t)+ε
  - y=f(temp(t))+ε

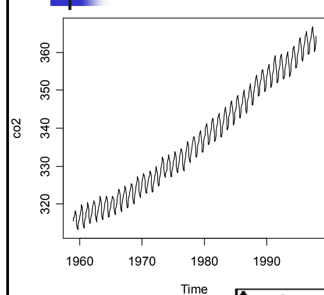
Space & time different

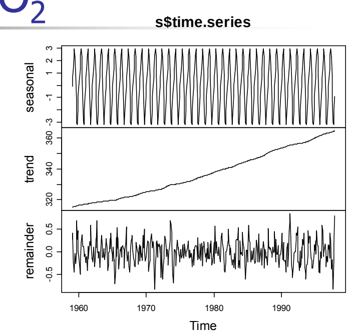
- Key assumption of GLM is independence of data points
- In space/time there is a particular form of nonindependence:
  - Nearby points are very similar
  - Called autocorrelation (self-correlated)
- World’s simplest weather forecast:
  - Tomorrow’s weather the same as today
  - Actually has a high degree of accuracy
- Also have a well-defined sense of “distance” between points

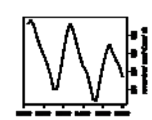
Three components of a time series

- A time series has:
  - Trend
  - Seasonality (or cyclicity)
  - Noise

Mauna Loa CO_2







Stationarity

- To analyze a time series, want to assume stationarity
- Stationarity=
  - Mean constant over time
  - Variance constant over time
  - Covariance structure constant
- Must remove trend & seasonality to make stationary
  - Removing trend:
    - Fit model (e.g. straight line)
    - Diff (t_i-t_{i-1})
  - Removing seasonality
    - Use routine

In R

- Creating timeseries

```
d=read.table('/temp/dickabund.csv',h=T)
dts=ts(d$Abund,start=1966)
```

- Plotting

```
plot(dts)
```

- Subscripting

```
window(dts,start=1980,end=1989)
window(dts,start=1980,end=1989,freq=0.5)
```

- Decomposing

```
data(co2)
dec=stl(co2,s.window='periodic')#s.window=12
Str(dec)
plot(dec$time.series)
plot(window(co2,1990,c(1991,6)) #Jan90-Jun91
```

- Aggregating

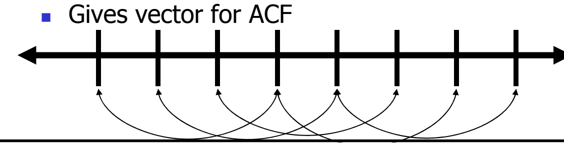
```
aggregate(co2,1,sum) #1 period/year=yearly
aggregate(co2,4,sum) #4/year=quarterly
```

Covariance structure

- Removing trend/seasonality makes
  - Mean constant
  - Variance constant (we hope)
- We hope we have a simplified covariance structure
  - Key simplification
  - $cov(e_i,e_j)=f(|i-j|)$ (i.e. depends only on time separation)
  - $cov(e_{i+t},e_i)=c_t$
- Called the ACF (Autocorrelation or autocovariance function)
  - Recall $cor(x,y)=cov(x,y)/\sigma_x\sigma_y$
    - Proportional – same shape, different heights
  - Covariance is usually useful in later calculations
  - Correlation easier to interpret (scaled –1 to 1)
  - Plot in autocorrelogram/autocovariogram

Calculating covariance

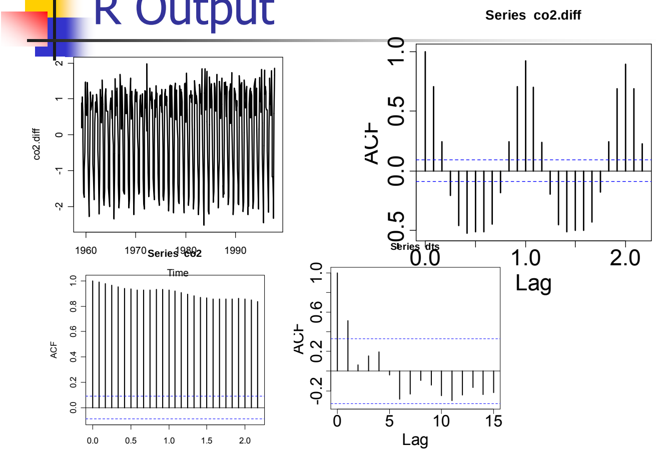
- For lag (distance in time) = t
  - Have multiple pairs
  - If N=8, t=4, then have 5 (N-t-1) pairs
  - # pairs fewer as t bigger → big CI for large t
  - Calculate covariance or correlation of these pairs:
    - $cov_4=cov(x_i,x_{i-4})=1/n\Sigma(x_{i+t}-\mu)(x_i-\mu)$
    - $cor_t=cov_t/cov_0$
- Repeat for each t
- Gives vector for ACF



In R

```
co2.diff=diff(co2)
plot(co2.diff)
acf(co2.diff)
acf(co2)
acf(dts)
```

R Output



Significance tests

- acf draws significance bands
- Problem of multiple tests
- Is 1 point outside enough? 20?
- Portmanteau test
- $Q=N\Sigma\rho(j)^2$ for j=1..K (N=length ts, K=test, e.g. 20)
- Q is Chi-square distributed with K df

Cofactors

- Endogenous
  - The time series itself
  - Self generating
    - Population dynamics, weather
  - Autocorrelation – can predict future from past
- Exogenous
  - External factors
  - E.g. weather or food for population dynamics
  - Need future of exogenous factors to predict
- Usually handled via regression with autoregressive & exogenous variables
  - $y_t \sim y_{t-1}+y_{t-2}+f_{t-1}+\varepsilon_t$

Other fancy stuff

```
NAO_SST <- read.table("Scottish_SST.txt", header = T,sep="\t")
SST=ts(NAO_SST$SST,start=1945,freq=12)
NAO=ts(NAO_SST$NAO,start=1945,freq=12)
ccf(stl(NAO,s.window="period") $time.series[,3], stl(SST,s.window="period")
    $time.series[,3])

ducks <- read.table("ducks.txt", header = T, sep = "\t", dec = ".")
library(lattice)
xyplot(Gadwall + Goldeneye + Goosander + Mallard +
       Pochard + Pintail ~ Year, data = ducks,
       type = "l", outer = T, layout = c(3, 2),
       scales = list(x = list(relation = "same"),
                     y = list(relation = "free")),
       ylab = "Abundance", xlab = "Time (years)")
library(vegan)
ducks_pca <- rda(ducks[, -c(1, 2)], scale = T)
plot(ducks_pca, scaling = 2, type = "n", xlab = "axis 1", ylab = "axis 2")
segments(x0 = 0,
         y0 = 0,
         x1 = scores(ducks_pca, display = "species", scaling = 2)[, 1],
         y1 = scores(ducks_pca, display = "species", scaling = 2)[, 2])
text(ducks_pca, display = "sp", scaling = 2, col = 2)
text(ducks_pca, display = "wa", labels = ducks[, 1], scaling = 2, pch = 19)
```

ARIMA

- A pre-fractal approach
  - Autoregressive
    - Random walk analog
    - $y_t=\beta_1 y_{t-1}+\beta_2 y_{t-2}+\varepsilon_t$
  - Moving average
    - White noise analog
    - $y_t=f_1\varepsilon_{t-1}+f_2\varepsilon_{t-2}+\varepsilon_t$
- AR(2)MA(2)
  - $y_t=\beta_1 y_{t-1}+\beta_2 y_{t-2}+ f_1\varepsilon_{t-1}+f_2\varepsilon_{t-2}+ \varepsilon_t$
  - Has a well-defined covariance structure (mostly 0, band along diagonal) – use GLS
- ARIMA = ARMA+differencing

GLM vs GLS

- GLM assumes:
  - $Y=Xb+\varepsilon$ w/ $\varepsilon \sim N(0,\sigma I)$
  - $cov(\varepsilon_i,\varepsilon_j)=0$
  - $cov(\varepsilon_i,\varepsilon_i)=\sigma$
- GLS relaxes
  - $Y=Xb+\varepsilon$ w/ $\varepsilon \sim N(0,\Sigma)$
  - $cov(\varepsilon_i,\varepsilon_j) \propto cor(\varepsilon_i,\varepsilon_j) \neq 0$
- In practice:
  - Too many parameters in Σ
  - Usually put some structure on Σ

σ	0	0
0	σ	0
0	0	σ

σ_{11}	σ_{12}	σ_{13}
σ_{12}	σ_{22}	σ_{23}
σ_{13}	σ_{12}	σ_{33}

GLS Timeseries

- Observe one variable over time (say population size): $y_1,y_2,y_3,\ldots y_t$
  - Want $y \sim temp_t+precip_t$ but ε autocorrelated
  - Assume $cov(\varepsilon_t, \varepsilon_s)=\rho^{|t-s|}$
  - AR1
  - Use correlation
  - 2 parameters: σ, ρ

$$
\begin{array}{c|ccccc|}
 & \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \varepsilon_5 \\
\hline
\varepsilon_1 & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\
\varepsilon_2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\
\varepsilon_3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\
\varepsilon_4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\
\varepsilon_5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \\
\end{array}
$$

Regression with timeseries

```
d<-read.table("Hawaii.txt",h=T)
str(d)
d$N<-sqrt(d$Moorhen.Kauai)
plot(d$N~d$Year)
m<-lm(N~Rainfall+Year,data=d,na.action=na.omit)
plot(m)
plot(residuals(m))
acf(residuals(m))
#uh oh!
m.gls<-
   gls(N~Rainfall+Year,na.action=na.omit,data=d,correlation=corAR1(
   form=~Year))
summary(m.gls)
AIC(m)
AIC(m.gls)
```

Example

```
d=read.table('dickabund.csv',h=T, sep=",")
dts=ts(d,start=1966)
acf(dts,lag.max=20)
m=gls(abund~tmin+tmax,data=d,
  corr=corAR1(0.4),method="ML")
m
summary(m)
```

Fractal model (vs ARIMA/GLS)

- White noise:
  - $y_t=\mu+\varepsilon_t$
- Random walk/Brownian motion/diffusion:
  - $y_t=y_{t-1}+\varepsilon_t$
- Fractal brownian motion (fBM)
  - Movie mountain scapes
  - Hurst exponent (H):
    - 0.5=white noise
    - 1= random walk
    - 0=antipersistent – flipflops regularly
    - Intermediate values

fBM

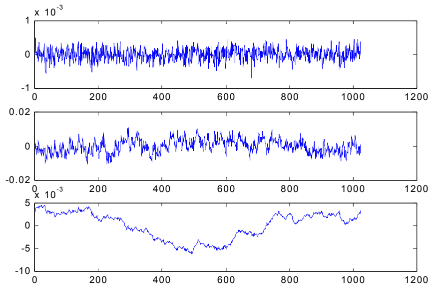
- H>0.5 → memory of system
  - Propensity for being like before
  - Weather shows memory
  - Random walk (H=1) is brown or red
  - Pink noise is in between white & red (e.g. 1/f)
  - Pinkish noise has variance constantly increasing with time!
  - H<0.5 (antipersistent) is blue noise
- Related values
  - β=2*H-1 (0=white noise, 2=random walk/brown)????
    - Power spectrum is $f(s)\propto s^{-\beta}$
  - D = Fractal dimension
- Extinction risk depends heavily on autocorrelation of noise

Examples of various noise

White

Pink

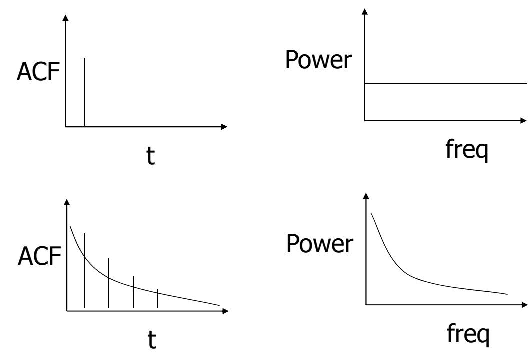
Red



ACF & power of noise

White

Brown



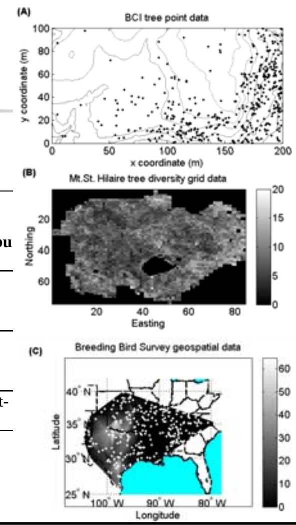
To analyze

- Spectrum analysis (from radio engineering)→Periodogram
  - Run a time series through
  - Use fourier analysis
  - Doesn't give + or – correlation – gives frequency of sine waves
  - Get "power" (energy) vs. frequency (1/lag)
- In R

```
spectrum(ts,span=oddint)
cpgram(ts)
# e.g. w/ co2
m<-stl(co2,s.window="period")
spectrum(m$time.series[,3],13)
cpgram(m$time.series[,3])
```

Space

Space – types of data



Spatial sampling design		Binary Variable (Present/Absent)	Continuous Variable (Amount/Abundance)
Sampler controlled	Regular across space	"Presence/absence quadrat"	Quadrat
	Random across space	"Presence/absence geospatial"	Geospatial
Biology controlled		Point-process	Marked point-process

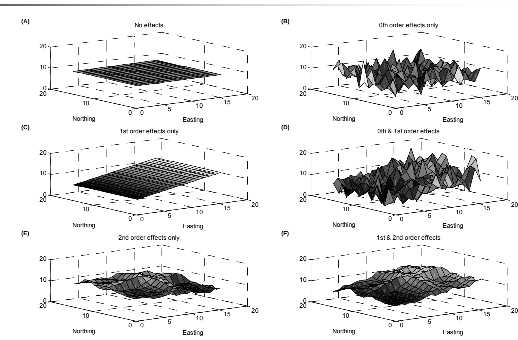
Space

- Similarities differences with time
  - Does space have trends?
  - Does space have periodicity?
  - Does space have covariance dependent only on distance?

Space very similar to time

- Main difference is that measuring points are:
  - Continuous
  - Irregular
  - 2-Dimensional
- Assume
  - $cov(e_x,e_y)=f(d(x,y))$
    Vs
  - $cov(e_{i+t},e_i)=c_t$
- Assume isotropy
  - Same in all directions

3 components

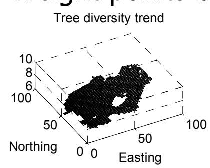


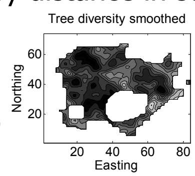
1. Noise
2. Trend
3. Spatial

- No seasonality/periodic

Trends in space (1st order)

- Several forms
- Global
  - Fit plane, quadratic
- Local=smoothing
  - Weight points by distance in some fashion





Fitting trend surface

```
library(spatial)
library(MASS)
data(topo)
```

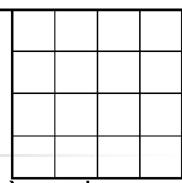
- Polynomial

```
mpo=surf.ls(degree,data) #deg=2 for
  quadratic
```

- LOESS

```
mlo=loess(z~x*y,data=?,degree=2,span=0.25,
  normalize=F)
```

Plotting Trend



- Most plots (including contour) need a regularly spaced grid
- Most data is not regularly spaced
- Getting a grid
  - Polynomial

```
surf=trmat(mpo,xlo,xhi,ylo,yhi,n)
```

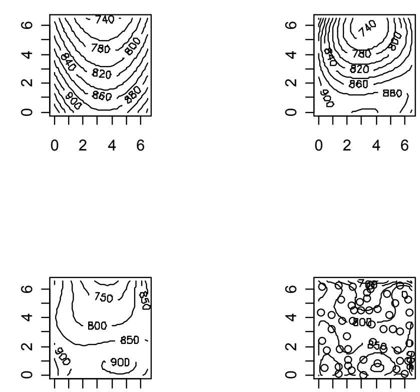
  - Loess

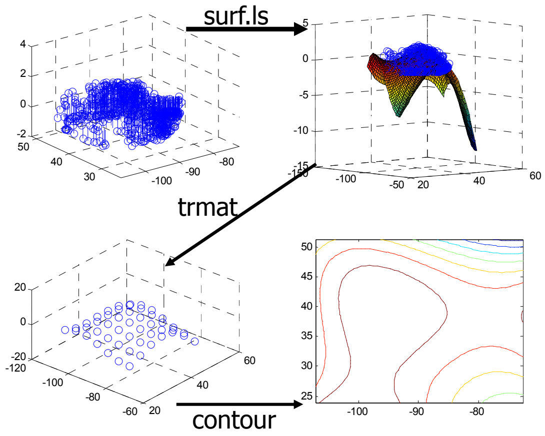
```
topogrid=list(x=seq(0,6.5,0.1),y=seq(0,6.5,0.1))
surf=predict(mlo,expand.grid(topogrid),se=T)
```

- Plotting

```
contour(surf)
contour(topogrid$x,topogrid$y,surf$fit)
```

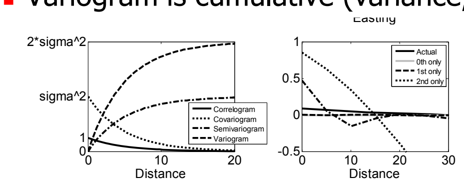
Contour of deg 2,3,4,6





Autocorrelation in space "grams"

- Space continous, not discrete
  - Cannot do exactly ACF
- Use "grams"
  - Correlogram uses correlation ~ACF
  - Variogram is cumulative (variance, not correl)



Covariograms

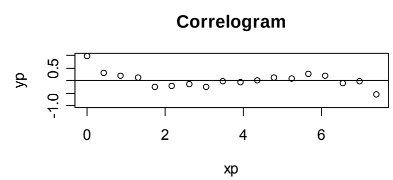
- Correlogram
  - Correlation between points at various distances
  - Distances lumped into bins
- Should you run a correlogram on the detrended surface?
  - You may want to run a model (topo.ls or topo.gls) of degree 0 (constant) and do correlogram of errors for this

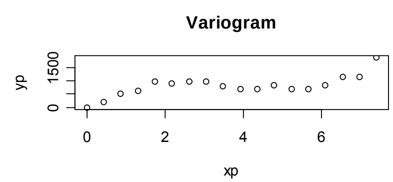
```
mpo0=surf.ls(0,topo)
mpo2=surf.ls(2,topo)
correlogram(mpo0,25)
correlogram(mpo2,25)
```

- Variogram
  - Covariance – not scaled

```
variogram(m2,nbins)
```

Output





GLS - Continuous Spatial

- Assume cov(ε_i,ε_j)=d_{ij} – only a function of distance
- Typically ρ_{ij}=exp(-kd_{ij})
- So estimate σ, k

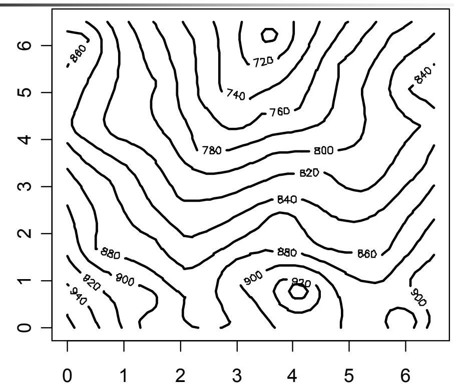
1	exp(-kd_{12})	exp(-kd_{13})	exp(-kd_{14})	exp(-kd_{15})
exp(-kd_{21})	1	exp(-kd_{23})	exp(-kd_{24})	exp(-kd_{25})
exp(-kd_{31})	exp(-kd_{32})	1	exp(-kd_{34})	exp(-kd_{35})
exp(-kd_{41})	exp(-kd_{42})	exp(-kd_{43})	1	exp(-kd_{45})
exp(-kd_{51})	exp(-kd_{52})	exp(-kd_{53})	exp(-kd_{54})	1

Kriging

- A form of estimating values at points (interpolation)
  - Either on grid or point of special interest
  - Weighted smoothing w/ weights given by variogram!
- Uses not just trend but guess at error term
  - Error term is a weighted average of nearby error terms with weighting based on distance and covariogram
  - Generally more accurate
- In R:

```
m2gls=surf.gls(2,expcov,topo,d=0.7)#GLS, not GLM,
   expcov is model for covariogram
surf=prmat(m2gls,0,6.5,0,6.5,50) #prmat vs trmat
contour(surf)
```

Kriged surface



Spatial regression

- Models:
  - S_i= β_0+β_1*t_i+γWS+αVt+ [x,y]ζ+η+ϵ_i
  - S (species richness) depends on t (temperature)
  - But also on:
    - Space per se [x,y] (trend effects) – traditional (trend surface)
    - Moving average=lagged dependent: S elsewhere (e.g dispersal) = Autocovariate, SAR
    - Autoregressive: depends on residuals (S-Shat) elsewhere = CAR, SAR
    - Lagged independent:t elsewhere (e.g. effects on juveniles or parents), SAR
    - Covariance of residuals (η)=GLS, LMM, GEE
  - See Dormann et al 2007!
- In R
  - Autoregressive, CAR, SAR in spatdep
  - Covariance in nlme, gee
- New PIC?
  - Estimates (b) unbiased, efficient
  - Does overestimate df (points not independent)
    - Duteillel's adjustment to p-value

Regression with spatial & exogenous factors

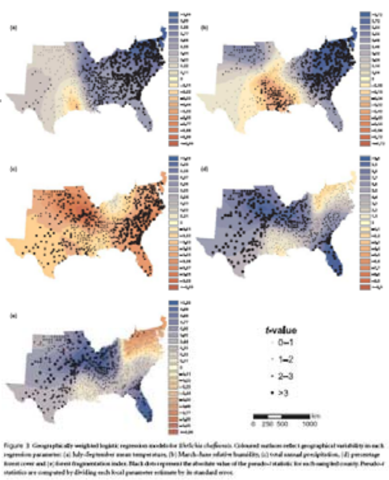
```
library(AED)
data(Boreality)
#boreal species over all species
Boreality$Bor<-sqrt(1000*(Boreality$nBor+1)/Boreality$nTot)
str(Boreality)
library(akima)
zz<-interp(Boreality$x,Boreality$y,Boreality$Bor)
image(zz)
contour(zz,add=T)
filled.contour(zz)
#simple model
m<-lm(Bor~Wet,data=Boreality)
summary(m)library(gstat)
E<-rstandard(m) #residuals
mydata<-data.frame(E,Boreality$x,Boreality$y)
coordinates(mydata)<-c("Boreality.x","Boreality.y")
bubble(mydata,"E",col=c("black","grey"))
#errors non-independent!
variogram(Bor~Wet,~x+y,data=Boreality)
```

In R II

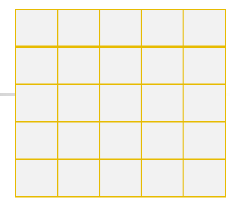
```
library(spatial) #different form of variogram +correlogram
m0<-surf.ls(0,Boreality$x,Boreality$y,Boreality$Bor)
variogram(m0,300)
correlogram(m0,300)
library(nlme)
m<-
   gls(Bor~Wet,correlation=corSpher(form=~x+y,nugget=TRUE),data=Bo
   reality)
#a 3rd form of variogram that works with nlme/gls
v<-Variogram(m,form=~x+y,robust=TRUE,maxDist=2000)
plot(v)
```

Non-stationarity

- Is the assumption of constancy over space realistic?
- Can we turn it into a tool?
- GWR (geographically weighted regression)
  - Wimberly et al 2008



Areal data



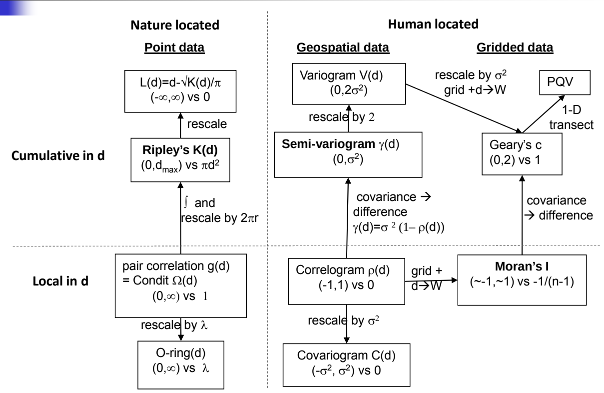
- Grids
  - Distance=# cells
    - Rook or Queen
  - Variogram->Geary c
  - Correlogram-> Moran I
  - GLS 2-D AR1
- Maps (counties, states)
  - Distance = midpoint-midpoint
  - Adjacency matrix W

Point processes

- Record location (x,y coordinates) of "events"
  - E.g. plant growing
- Idea of overdispersed/random/clumped
- Poisson process (rate=λ)
  - K(t)=# points within distance t/λ
  - If Poisson K(t)=πt^2
  - Study L(t)=sqrt(K(t)/π) should be straight line
- Great deal of machinery
- Also works on quadrats – count in each quadrat is Poisson distributed
- Alternatives to Poisson give clumped/overdispersed

```
library(spatstat)
d<-read.table("ragwortmap2006.txt",h=T)
ragwort<-
   ppp(d$xcoord,d$ycoord,c(0,3000),c(0,1500),marks=d$type)
plot(ragwort)
summary(ragwort)
plot(quadratcount(ragwort))
Z <- density.ppp(ragwort)
plot(Z,main="")
K <- Kest(ragwort)
plot(K) #note observed (black)>red null→clumped
pc <- pcf(ragwort)
plot(pc)
```

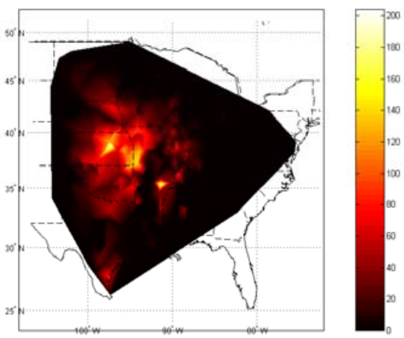
Spatial summary



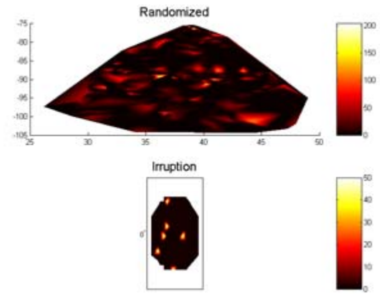
Time & space summary

- Independent error terms is clearly violated
- Study of covariance of error terms with distance interesting

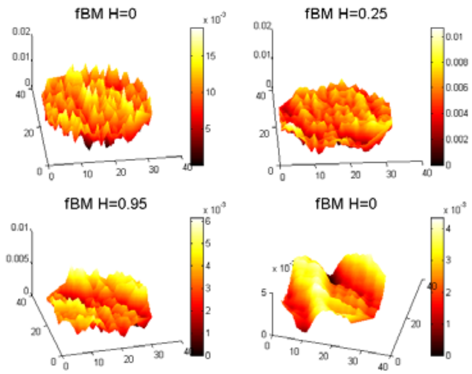
Spatial data



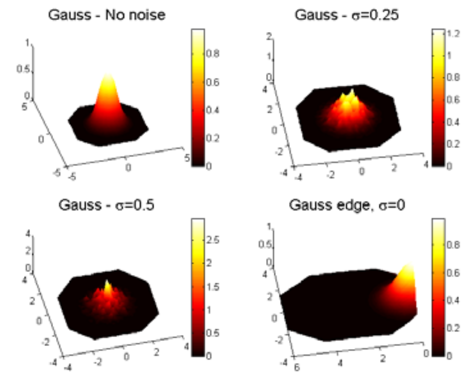
Random null models



fBM null models



Gauss null models



Results V - Empirical autocorrelation