Geostatistical Data

Matt Kramer

kramerm@ba.ars.usda.gov

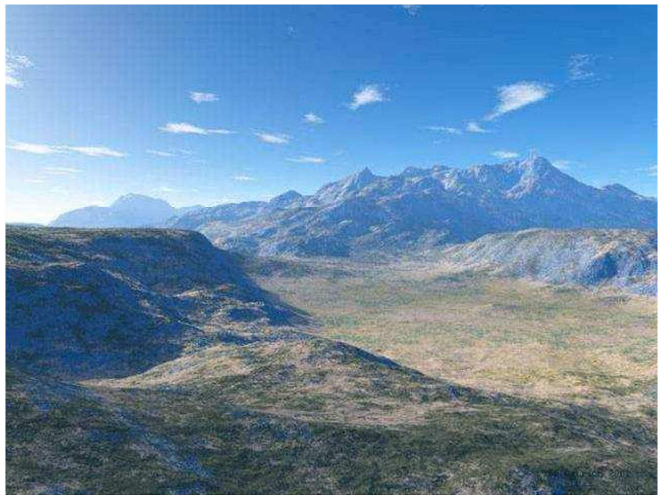
Biometrical Consulting Service, ARS/BARC/USDA

Outline

- Spatial landscapes
- Realizations
- Decomposition of the landscape
- Stationarity
- Variograms
- Ordinary kriging
- Prediction
- Universal kriging
- Important concepts not covered

Spatial landscapes

Fractal terrain by Rolf Lakaemper

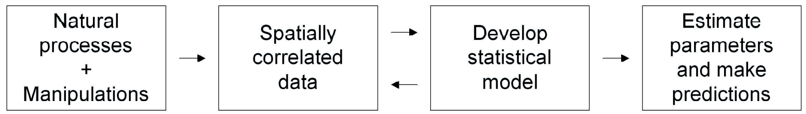


Generating fractal terrain

- The key concept behind fractals is self-similarity
- When a small region of a fractal is magnified, it looks similar to the whole region from which it was taken
- Terrain has this property (loosely defined), which is why fractal algorithms are commonly used to generate “realistic” landscapes
- The property of scale is important for field work, spatial correlation occurs at all scales and how we choose to describe it will depend on the organism (or process) being studied and the crudeness of the tools available.
- We typically classify the variation in the landscape we see into large scale variation, which we might try to explain with regression type variables (e.g. elevation), and small scale variation, which we try to explain using a model of spatial dependency (e.g. kriging).

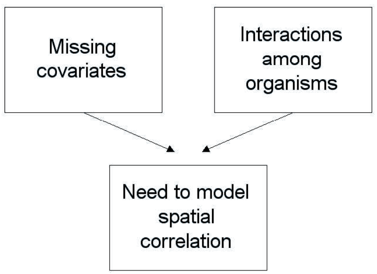
Spatial data as a process

- We observe data generated from some underlying process we are trying to understand
- These data may be observational (e.g. bird counts in a forest) or the researcher may have had a hand in the outcome (e.g. designed experiment where different treatments were applied to various locations)
- We decide on a statistical model that we believe captures the effects we are interested in
- We estimate its parameters and possibly try to interpret them



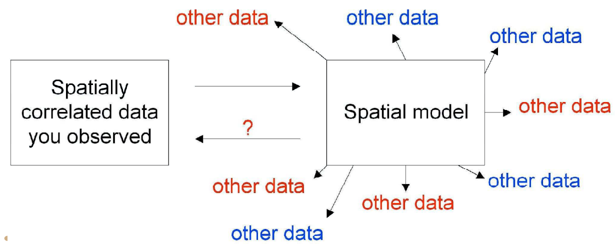
Causes of spatial correlation

- The spatial correlation in the data may be partly (or completely) due to our not having suitable variables to explain why observations closer together are more similar
- Sometimes the spatial correlation is due to interactions among the organisms themselves (e.g. root competition, aggregation), so additional covariates (predictor variables) would not help

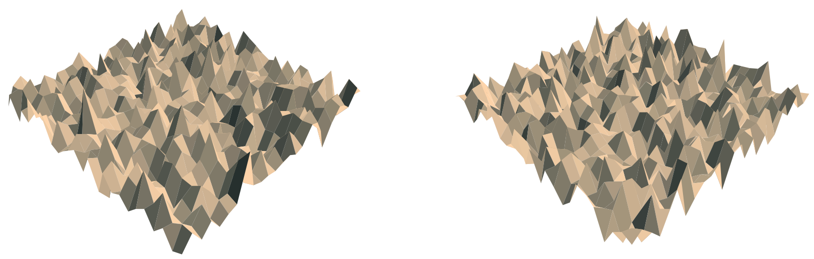


Realizations

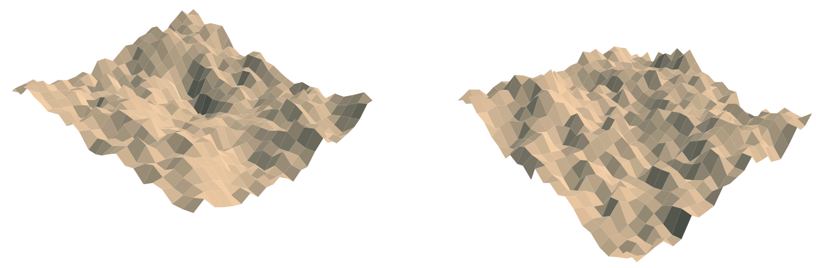
- We have formal models for describing spatial correlation
- We choose one consistant with the spatial pattern of our observations
- The data observed are not unique to that statistical model
- The data are one realization of this statistical model
- Looking at many realizations helps to better understand what kinds of sample data this model can generate



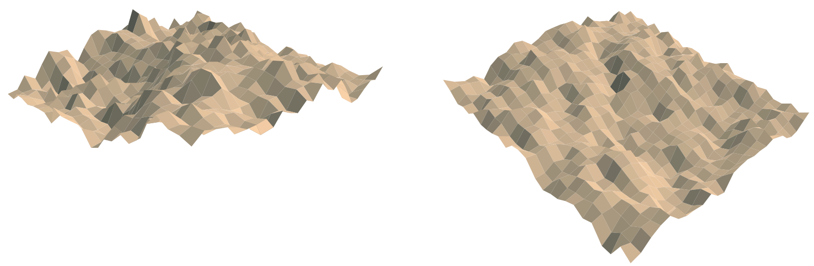
Two realizations of spacial random noise $\sim N(0,1)$



Two realizations of moderately spacially correlated data



Two realizations of strongly spacially correlated data



Realizations—what have we learned?

- spatially correlated observations look “smoother”, some of this is due to scaling
- There are regions of high and low observations with spatial correlation, this pattern may be masked by covariates or treatment effects when looking at “real” data
- You cannot determine the degree of spatial correlation by looking at these plots, we use a tool called the variogram for that
- Strongly spacially correlated data is often symptomatic of a failure to adequately model the “trend” (large scale variation)

Decomposing the landscape: Large scale variation

- Typically thought of as the trend, variation on a scale much larger than distances between observations
- Important to capture all explanatory variables making up the trend, otherwise the residuals may be "non-stationary", which will make modeling small scale variation difficult
- Especially important to capture explanatory variables that vary spatially (spatially varying covariates)
- In designed experiments, blocking is used to capture some of the large scale spatial variation and randomization within the block to reduce the impact of small scale variation
- Large scale variation is typically handled using covariates (e.g. elevation, soil characteristics, latitude and longitude) and ANOVA type variables (e.g. treatments/interventions, historical land use, type of vegetation cover)

Decomposing the landscape: Small scale variation

- Sources of variation not associated with the trend, and at a smaller scale
- Typically imagined to have two components, a smooth function which describes the covariances (correlations) between neighboring observations, and random error (or noise)
- The scale of small scale variation is larger than the smallest distance between observations (typically several times larger)
- What may be considered small scale variation in one study may be large scale variation in another.
- We ignore spatial relationships that occur at scales not captured by our data.

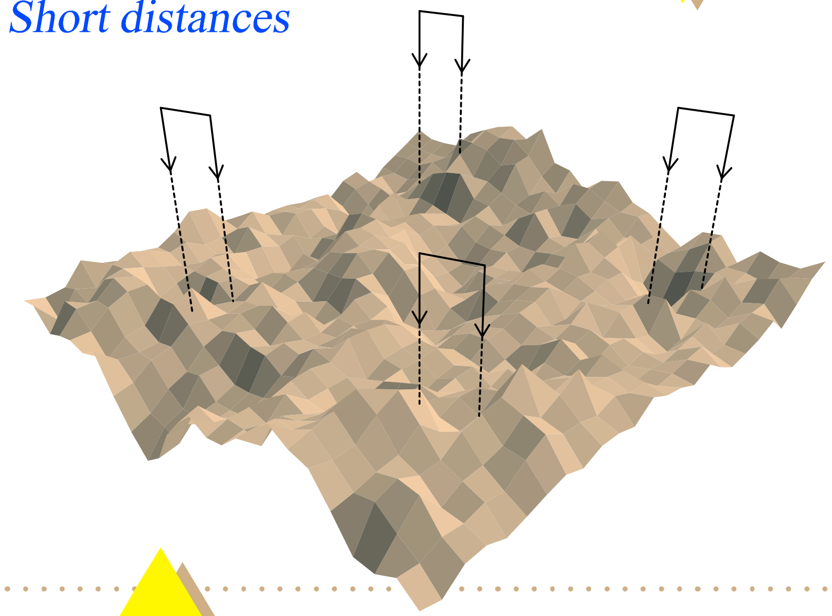
Stationarity

- We need to make simplifying assumptions to model small scale variation
- Spatial correlation necessarily involves pairs of observations
- Data sets with more than 3 observations, have more pairs of observations than observations
- We want the number of parameters in a model to be (far) less than the number of observations.
- In the simplest case, assume spatial relationships between observations are the same everywhere in the landscape, i.e. that the spatial relationships only depend on the distance between observations
- This property is stationarity

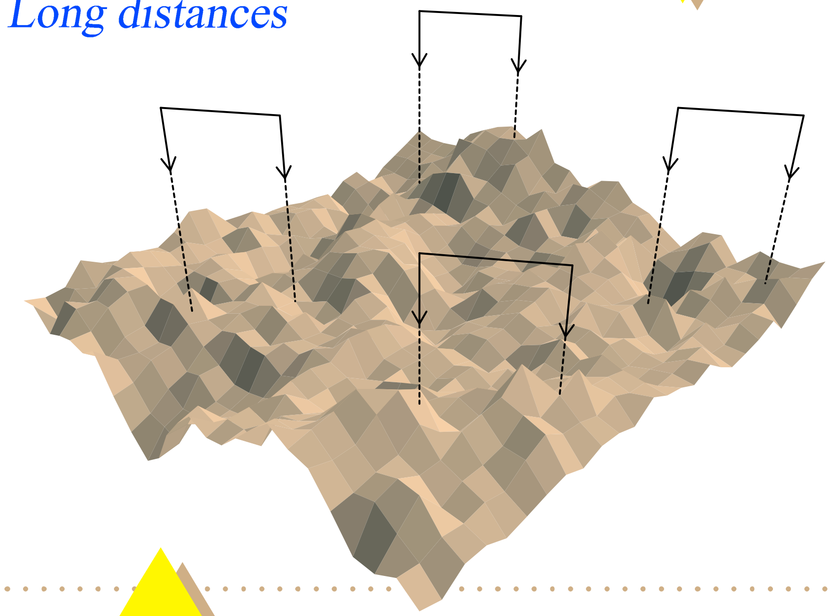
Stationarity

- Often this is not realistic, we may have to allow for spatial relationships to depend on direction (so observations may be more correlated going north to south than east to west), or for them to vary in some other way across the landscape.
- In general, raw data will not be stationary until the large scale variation is removed, so one must first deal with large scale variation before tackling small scale variation
- In the remainder of this presentation, we assume stationarity, but for real data, this would need to be verified.

Short distances



Long distances



Comparison

Comparison of short (0.09 units apart) and long (0.2) distance pairs.

distance	obs. 1	obs. 2	$2\hat{\gamma}(h)_i =$ (obs. 1 − obs. 2)^2
short	-0.67	0.78	2.10
short	1.47	1.52	0.00
short	-0.82	-0.00	0.67
short	-1.12	-0.38	0.54
long	-0.67	1.40	4.27
long	1.47	2.20	0.54
long	-0.82	0.28	1.22
long	-1.12	-0.40	0.52

$2\hat{\gamma}(h)$ is the classical estimator of the variogram; h is the distance separating the observations

Variogram

If there is small scale spatial autocorrelation, we expect observations near each other to be more similar than ones further away

- This was seen in our example, $2\hat{\gamma}(0.09) = 0.83 < 2\hat{\gamma}(0.2) = 1.64$
- The pattern that emerges, if we plot distance (h) on the x-axis and $2\hat{\gamma}(h)$ (or $\hat{\gamma}(h)$, the semivariogram) on the y-axis, should tell us something about small scale variation
- $\hat{\gamma}(h)$ should be small when the distance h is small, $\hat{\gamma}(h)$ should be larger as the distance h increases
- What is the best way to do this?

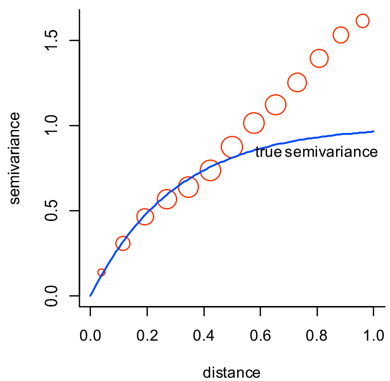
Variogram

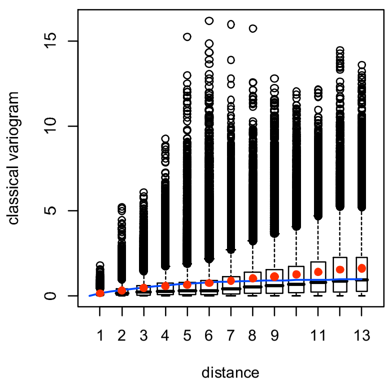
- If we look at the distribution of pairs of observations by distance apart, we find that there are far fewer pairs of observations separated by large distances
- Thus, our estimates $\hat{\gamma}(h)$ for h large will not be as good as $\hat{\gamma}(h)$ for h smaller
- If our data are not evenly spaced, we may find the same problem for h very small, there may only be a few pairs that represent the smallest distances
- This means that some regions of the semivariogram have better support than others

Variogram

- To create the semivariogram, we break h up into many distance groups (e.g. 0–0.2, 0.2–0.4, 0.4–0.6, etc.) and calculate $\hat{\gamma}(h)$ for each distance group.
- Then we can plot the average value of h for that distance group against $\hat{\gamma}(h)$
- We can also plot $\hat{\gamma}(h)_i$ for each pair of observations, this may help us decide if the average value for each h is a reasonable estimate of what the "mean" should be
- In practice, we have software that does this, though we may make decisions about how large an interval each distance group should be, and what our largest h should be (since beyond a certain h results will be rather flaky as there aren't many pairs of observations for very large h)

Variogram ($\hat{\gamma}(h)$ vs. h)





Variogram: What have we learned?

- The variogram nicely displays the similarity of neighboring observations, and how differences between observations increase with increasing distance
- Even with $n = 676$ observations, the empirical semivariances do not follow the true semivariances beyond $h = 0.5$ units (distance between the two furthest observations is 1.4 units)
- These data were generated from a known model (where we know the true parameters), yet there are still problems with the variogram
- We could regenerate data sets from this model until we created one that produced a nice variogram, but one cannot do that for "real" data

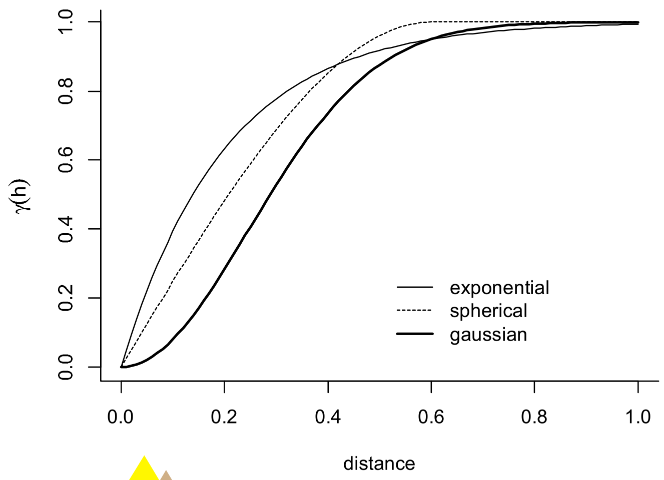
Variogram: What have we learned?

- The box plots show how variable the individual semivariance estimates are for each distance class
- The variogram is an imperfect tool, but in practice it works well
- There are robust procedures for estimating the variogram

Variogram—what model to use?

- Software for modeling spatial data will have many different models that one can use to capture the spatial autocorrelation
- These models differ in how the strength of the correlation between observations diminishes as distance between them increases
- The data for this example were generated using an exponential model
- Many of the models produce very similar results (and you might need a lot of data to be able to discriminate between models)
- It is more important to try to capture the spatial dependencies with some model, even if you aren't sure it is the "right" model, then to ignore the spatial dependencies completely.

Three common variogram models



Variogram—estimation of model parameters

- Once we have decided on a model for the data, we need to estimate its parameters
- Many variogram models have parameters (or combinations of parameters) that can be interpreted as the range, sill, and nugget (these terms show geostatistics' mining origin)
  - The range is the minimumn distance separating observations that are (nearly) spatially independent
  - The sill is the value of $\gamma(h)$ when $h =$ range
  - A nugget effect occurs if, as h (the distance between observations) goes to zero, $\gamma(h)$ does not approach zero
  - The partial sill $=$ sill $-$ nugget

Variogram models

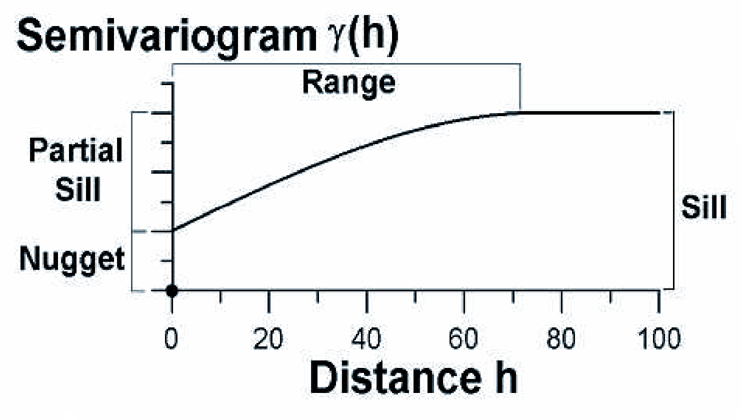
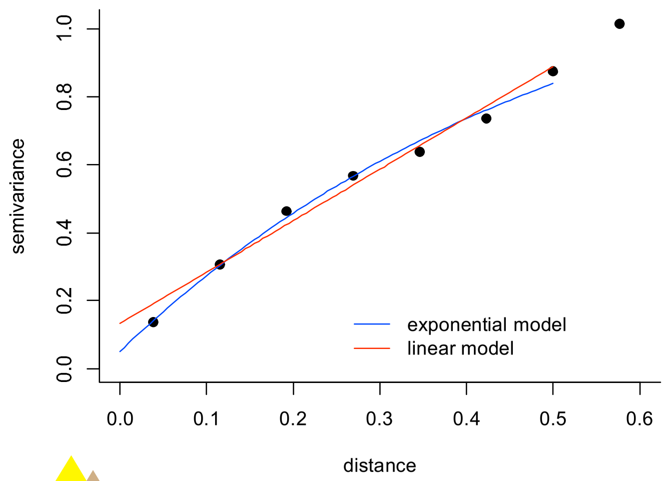


Image by Jay Ver Hoef

Variogram—estimation of model parameters

- A least squares approach (i.e. regresson equation) is common
- The least squares approach is usually modified so that it gives more weight to small h (where it is most important to have a good fit) and to areas of the variogram that have the most pairs of observations
- Robust methods have also been developed
- The software typically does this fitting, you only select the model you want to use and options for how to do the fit
- You then plot the graph against the variogram estimates (the averaged or “binned” estimates, one for each distance category) to check the fit visually

Variogram model parameters



Variogram model parameters

- Two models were fit, exponential and linear, to the data up to $h = 0.5$.
- Note: These fits look good only because the distance was cut off at $h = 0.5$!
- Estimates for the variogram model parameters, nugget, partial sill, range:
  - Exponential: τ^2 = 0.12, σ^2 = 1.28, ϕ = 0.52
  - Linear: τ^2 = 0.13, σ^2 = 1.51, ϕ = 1.00

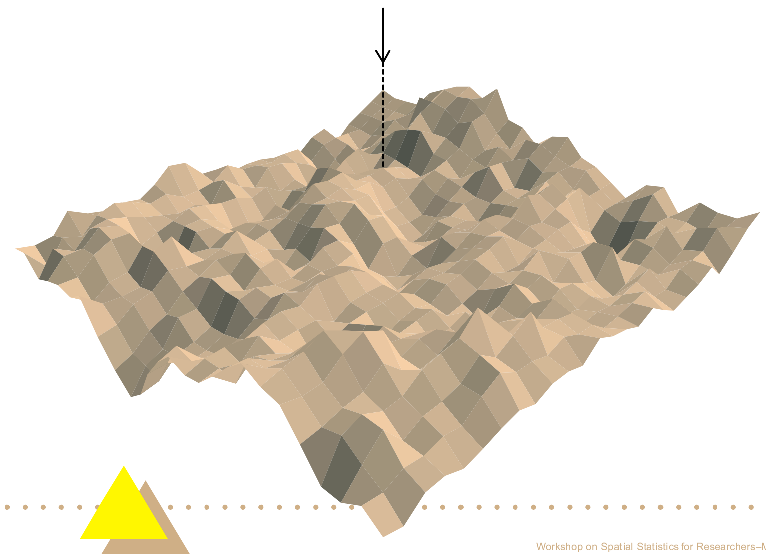
Ordinary Kriging

We now have a model for the spatial dependencies in our data.

- We can estimate a value at a particular location (which should be within the general area in which the data were collected!)
- In this case, the uncertainty associated with the estimate will depend on how far the location is from real observations and how much spatial correlation exists
- If the location is further from any real observations than the range, we get no “special” information from nearby observations and the best estimate will be the mean
- Unlike, e.g. regression, a prediction at a location where we have an observation just gives us back the value of the observation
- This is a technique that can be used for observations that are unequally spaced as well regularly spaced (the example used here is for regularly spaced data)

Prediction at $(x = 0.27, y = 0.27)$

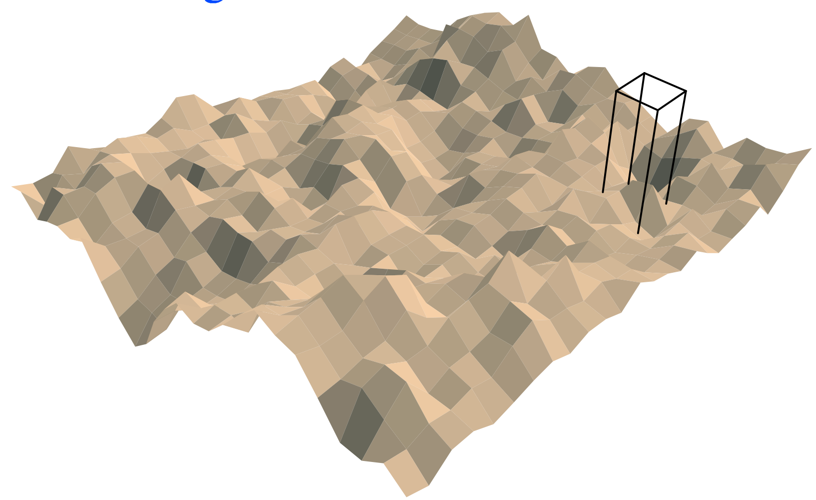
point estimate = −0.376, kriging variance = 0.044



Predict a region

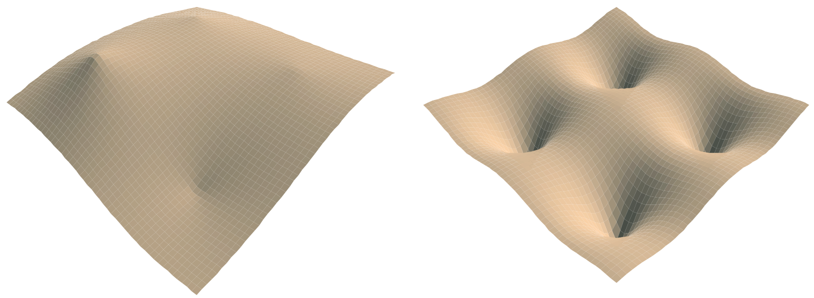
- We can also create an estimate for the region (or some subset of the region) in which the data were collected, e.g. the average value
- The uncertainty associated with this estimate will depend on the density of real observations in the region and how much spatial correlation exists
- These kinds of estimates are performed by software, we need to specify the model and what output we want

Predict a region



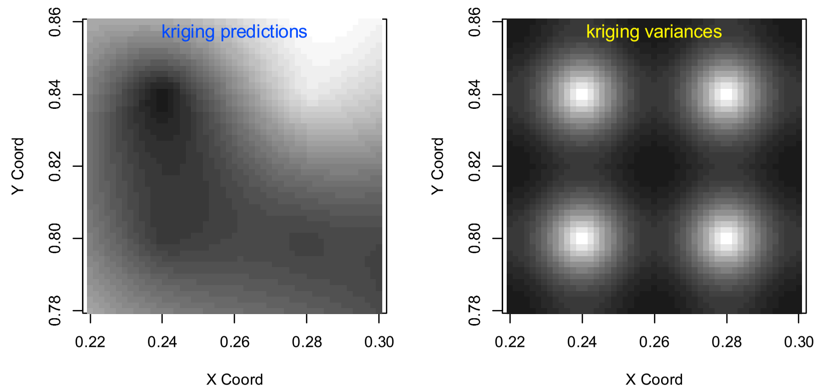
Predictions & variances—perspective view

- Left plot: krigged surface (note how smooth it is!)
- Right plot: kriging variances (variance is zero where data were taken unless there is a nugget effect)



Predictions & variances—typical output

Relative prediction and variance values coded by intensity (black = large values, white = low values)

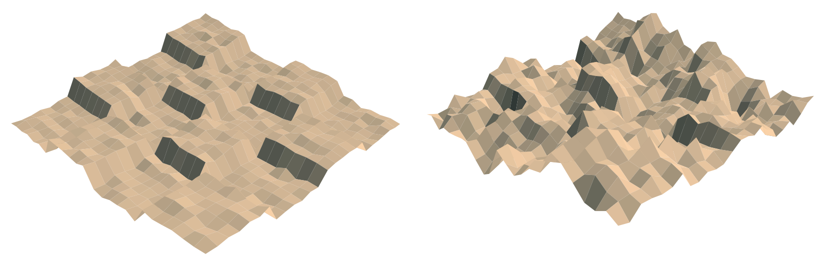


Universal Kriging—estimation strategy

- We often have other information about the landscape we are modeling, such as covariates or factors (e.g. treatment effects), in which case we have a mixed model
- If we can subtract out these effects, then we can use the strategy just discussed to model the spatially correlated residuals
- For the most common geostatistical models, mixed models software can estimate all the parameters of the model (covariates, factors, spatial covariance parameters)
- Unfortunately, there are deficiencies in the software (limited spatial models, lacking good diagnostics)

Universal Kriging—trend and noise

- Left plot: trend (covariate + two-level factor) (note: covariate effect not easy to see because it, in part, tilts the plane surface)
- Right plot: trend + noise (noise = spatially correlated residuals)



Universal Kriging—estimation strategy

- Added a covariate and factor effect to the spatially correlated observations
- We assume the spatial correlation is unrelated to these effects
- If we had no idea of the pattern of spatial correlation (of the residuals), we might start out by
  - assuming that residuals are uncorrelated and estimate the covariate and factor effect using a linear model
  - subtract out their effects from the data
  - determine if the residuals are stationary, and if so
  - use a variogram to determine their pattern of spatial covariance
  - re-estimate the model using mixed models software

Universal Kriging—estimate trend

- Although we already know the function to use for spatial correlation of the residuals, we'll pretend we don't
- First estimate the trend assuming uncorrelated residuals.

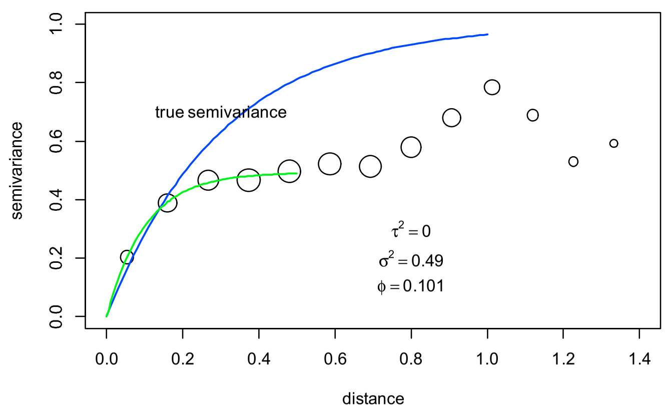
```
> fit1 <- lm (dat1 ~ as.factor(f1) + covar1 - 1)
> summary(fit1)

                Estimate Std. Error t value Pr(>|t|)
as.factor(f1)0 -0.59867    0.05293 -11.310  < 2e-16 ***
as.factor(f1)1  0.22169    0.05497   4.033 6.13e-05 ***
covar1          1.75290    0.03131  55.988  < 2e-16 ***

Residual standard error: 0.7184 on 673 degrees of freedom
```


Universal Kriging—model noise

Semivariogram of the residuals



Universal Kriging—estimate full model

R software, *geoR* package

```
gdat2 <- as.geodata(cbind(x,y,dat1))

ts1 <- trend.spatial(trend= ~ as.factor(f1) + covar1 - 1)

fit2REML <- likfit (gdat2, trend=ts1, ini.cov.pars=expfit2$cov.pars,
          fix.nugget = FALSE, cov.model="exp", method.lik = "REML")
```

Universal Kriging—estimation results

```
 beta0  beta1  beta2
0.2051 1.2540 1.0864

Parameters of the spatial component:
   correlation function: exponential
      (estimated) variance parameter sigmasq (partial sill) =  1.118
      (estimated) cor. fct. parameter phi (range parameter)  =  0.3689
Parameter of the error component:      (estimated) nugget =  0

> sqrt(diag(fit2REML$beta.var))
0.5106032 0.5107846 0.1098927
```

Estimates ignoring spatial correlation:

```
               Estimate Std. Error t value Pr(>|t|)
as.factor(f1)0 -0.59867    0.05293 -11.310  < 2e-16 ***
as.factor(f1)1  0.22169    0.05497   4.033 6.13e-05 ***
covar1          1.75290    0.03131  55.988  < 2e-16 ***
```


Universal Kriging—estimation results

These results closely match those using the *nlme* R package:

```
> fit3 <- gls (dat1 ~ as.factor(f1) + covar1 - 1, corr =
               corExp(c(1,0.1), form = ~ x + y, nugget = TRUE))
> summary(fit3)
Generalized least squares fit by REML
Correlation Structure: Exponential spatial correlation
 Formula: ~x + y
 Parameter estimate(s):
       range       nugget
3.688905e-01 3.302637e-09

                   Value Std.Error  t-value p-value
as.factor(f1)0 0.2050859 0.5105995 0.401657  0.6881
as.factor(f1)1 1.2539898 0.5107810 2.455044  0.0143
covar1         1.0863683 0.1098926 9.885727  0.0000
Residual standard error: 1.057395
```


Universal Kriging—model comparison

Comparison of results from ignoring spatial correlations versus incorporating them into the model

- for the fixed part of the model (covariate + factor), parameter estimates and standard errors differ
- differences in parameter estimates are not that large once centering has been taken into account
- standard errors are much larger for model with correlated residuals, this shows that ignoring spatial autocorrelation produces incorrect tests on factors (e.g. treatment effects)
- estimation time for the linear model was < 1 sec., for the model with autocorrelated residuals, > 10 min. ($n = 676$)

Important concepts not covered

- Isotropy—anisotropy
- non-Euclidean distance measures
- Diagnostics
- Transforming data that are not normal
- Robust methods
- Variances/standard errors for kriged estimates

THE END