Overview of Statistical Analysis of Spatial Data
Geog 210C
Spatial Interpolation

Chris Funk
Outline

- Truth-generating processes
- Why does spatial information matter?
- What qualifies as a ‘good’ interpolation
- Interpolation example: African MAMJ Temperatures
- Interpolation techniques
  - Inverse distance weighting
  - Splines and radial basis functions
  - Simple Kriging
Truth generating processes

Maps are representations of the physical world
- models ‘of’ reality as it ‘really is’ and
- Models ‘for’ reality as we interact politically, socially, and environmentally with the world
  - Clifford Geertz

‘Data’ is not a passive object residing ‘out there’ preformed, rather ‘data’ arises through our dynamic inter-locution with the world.
  - Thomas Kuhn, “The Structure of Scientific Revolutions”

Applied statistics can provide us with NEW empirical knowledge, when applied coherently, guided by the an appropriate knowledge of the system being studied
  - Typically, there is an iterative process of discovery and refinement
From one data, many maps ...

How do we choose?
Characteristic of good interpolations

 진행한 대화로 추정된 내용으로는 다음과 같습니다:

- Qualities of a ‘good’ mapping approach
  - Results are unbiased
  - Results have low standard errors
  - Recreates the distribution of the underlying data

- The methodology …
  - Is transparent
  - Is physically defensible and informed
  - Is reproducible
  - Commensurate with the amount of available data
  - Produces credible estimates of errors and uncertainties
Example: IPCC AR4 Report

Can we do better than this?
MAMJ Temperature Anomalies

MAM Temp Anoms

ET
KE
SU
MA
NI
JJAS Temperature Anomalies

JJAS Temp Anoms

-1.0
-0.5
0.0
0.5
1.0
1.5


ET
KE
SU
MA
NI
Adaptation Strategies

From Re-Greening the Sahel, Reij, Tappan, and Smale

Zai techniques improve soil fertility

Improving soil fertility through rehabilitation of degraded lands
Adaptation Strategies-II

Adaptive Capacity

Exposure
Climate = What you expect
Weather= What you get
Non-stationarity = the new ‘normal’ ?

Model
\
\[ T_{2000-2009} = \mu_{1960-1989} + \Delta T_{2000-2009} \]

\[ \mu_{1960-1989} = F(q,p,elev,IR,LST,RFE) + \text{krig}(u') \]
\[ \Delta T_{2000-2009} = F(q,p,elev,IR,LST,RFE) + \text{krig}(\Delta T') \]
\[ R = \text{cor}(x'\cdot\text{wts}, y'\cdot\text{wts}) \]

- \(x'\) and \(y'\) indicate centered variables
Correlations with lat & lon

Circle magnitude: 0.3, 0.6, 0.9
Local R with Lat
fallintemp. 1960-79 MAMJ

negative positive
$\mu_{1960-1989}$ correlations with 10th percentile IR & DEM
μ_{1960-1989} correlations with LST and RFE
$\mu_{1960-1989} = \text{regress}(X'*wts, y'*wts)$

- $X'$ and $y'$ indicate centered variables
- $X$ is a set of spatially exhaustive predictor variables

Tricube Function Weights
Moving Window Regression results
\[ \mu_{1960-1989} = F(q,p,elev,IR,LST,RFE) + \text{krig}(u') \]

\[ \text{Variance}(\mu_{1960-1989}) + \text{Var}(F(\theta,\lambda,\text{elev},\text{IR},\text{LST},\text{RFE}) + \text{Var}(\text{krig}(u'))) \]
Kriged Standard Errors
Correlations with Lat and Lon

northern areas warming more

southern areas warming more
Correlations with Elevation and RFE2

Very important: negative correlations show that drier and lower locations are warming faster than high-wet locations.
> summary(bg)

<table>
<thead>
<tr>
<th></th>
<th>lon</th>
<th>lat</th>
<th>dem</th>
<th>rfe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>-19.500</td>
<td>-5.0</td>
<td>0.0</td>
<td>0.000</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-1.625</td>
<td>0.5</td>
<td>0.0</td>
<td>7.711</td>
</tr>
<tr>
<td>Median</td>
<td>16.250</td>
<td>6.0</td>
<td>318.0</td>
<td>48.174</td>
</tr>
<tr>
<td>Mean</td>
<td>16.250</td>
<td>6.0</td>
<td>402.1</td>
<td>154.865</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>34.125</td>
<td>11.5</td>
<td>561.0</td>
<td>230.974</td>
</tr>
<tr>
<td>Max.</td>
<td>52.000</td>
<td>17.0</td>
<td>4277.0</td>
<td>2061.375</td>
</tr>
</tbody>
</table>

> mn.mod = run.region(as.matrix(bg), stn.chg, xy.loc, centers, dmax=2500, min=2)

There were 50 or more warnings (use warnings() to see the first 50)

> mn.est = mf.region.predict(mn.mod$coef, mn.mod$cent, ll, gha.bg, dmax=width)

There were 50 or more warnings (use warnings() to see the first 50)

> gha.bg = as.matrix(cbind(ll[,1], ll[,2], fc.dem^0.5, fc.rfe))
> bg = cbind(xy.loc[,2], xy.loc[,1], stn.dem^0.5, stn.rfe^0.5)
> mn.cv = mf.region.cv(bg, as.vector(stn.chg), xy.loc, dmax=width)

[1] 0.5626124

There were 44 warnings (use warnings() to see them)

> plot(mn.cv$fitted.values, stn.chg)

There were 50 or more warnings (use warnings() to see the first 50)

> mn.mod = run.region(as.matrix(bg), stn.chg, xy.loc, centers, dmax=width, min=2)

There were 50 or more warnings (use warnings() to see the first 50)

> mn.est = mf.region.predict(mn.mod$coef, mn.mod$cent, ll, gha.bg, dmax=width)

There were 50 or more warnings (use warnings() to see the first 50)

> bg.est[bg.est < -2] = -2
> bg.est[bg.est > 4] = 4
> output(bg.est*gha.msk, outdf0.reg)
Cross-validated change estimates

RMSE=1, Mean Bias Error=0, Mean Absolute Error = 0.78 deg C
2000-2009 Temperature Anomalies
Standard Error Estimates
Characteristic of good interpolations

- Qualities of a ‘good’ mapping approach
  - Results are unbiased
  - Results have low standard errors
  - Recreates the distribution of the underlying data

- The methodology …
  - Is transparent
  - Is physically defensible and informed
  - Is reproducible
  - Commensurate with the amount of available data
  - Produces credible estimates of errors and uncertainties