Spatial analysis with R

Lucy Bastin
Aston University
General plan

• Spatial questions
• Spatial data formats & types
• R packages for
  – Spatial data analysis
  – Map visualisation
• Research examples
  – Disease mapping
  – Web-based interpolation
  – Conservation planning
Spatial questions

• Metric space
  – How far? What direction? What area?...
  – Affected by map projection
  – Image processing, Vector graphics, CAD...

• Topological space
  – Connected? Within? Accessible from?
  – Graph theory, optimal routing...
Vector models of the world

- Coordinates: points, lines, polygons etc..
- Great for discrete, homogeneous objects.
Raster models of the world

• Regular sampling: ideal for continuous phenomena - aerial photo or a satellite image.

The Human Footprint Index

The Human Footprint Index (HF) expresses as a percentage the relative human influence in each terrestrial biome. HF values range from 0 to 100. A value of zero represents the least influenced - the "most wild" part of the biome with value of 100 representing the most influenced (least wild) part of the biome.
Vector and raster data in R

• Vector formats
  – Shapefiles, csv and text files...

• Raster formats
  – Geotiff, netCDF..

sp, rgdal, raster, maptools, RArcInfo, ncdf

fundamental data types and reading/writing
Spatial analysis in R

gstat, geoR, intamap, spacetime, akima
spatial / spatiotemporal kriging / interpolation
rgeos
topological operations: e.g., intersection...
PBSmapping, maps, GEOmap, mapproj
point-in-polygon, re-projection...
ecespa, spatstat, spatial, splancs
point pattern analysis
trip, adehabitat, rangeMapper
animal tracking and territory mapping
landsat
topographic / other correction of satellite imagery
glmmBUGS, Mondrian, spdep, RPyGeo...
Visualisation in R

RgoogleMaps, plotKML, OpenStreetMap
overlay data on Google Maps or OSM

**ggmap** (enhances ggplot2)
overlay on Google Maps, Stamen, OSM or CloudMade

rasterVis
interactive image mapping (esp. large rasters)

maps, GEOmap, rworldmap, GADM
country boundary datasets & map functions
library(sp)
con <- url("http://gadm.org/data/rda/CHE_adm1.RData")
print(load(con))
close(con)

spplot(gadm, "language", col.regions=col, main="Swiss Language Regions")

rworldmap

References, tutorials, etc.

http://www.bias-project.org.uk/ASDARcourse/unit7_slides.pdf
https://sites.google.com/site/rodriguezsanchezf/news/usingrasagis
http://www.itc.nl/~rossiter/teach/R/RSpatialIntro_ov.pdf

* https://dl.dropbox.com/u/24648660/ggmap%20useR%202012.pdf
http://www.nceas.ucsb.edu/scicomp/usecases/CreateMapsWithRGraphics
http://geography.uoregon.edu/geogr/topics/maps.htm
http://www.milanor.net/blog/?p=594
http://spatialanalysis.co.uk/r/
1. Disease mapping

- Working with local PCTs to map infection.
- Example: community-associated MRSA.
- Infections in people who have not recently visited a hospital as an in- or out-patient.
- Using demographic / environmental factors to identify SIGNIFICANT clusters of disease.
- R packages: splancs and spatstat
Simulating disease rates

- Underlying population demographics
  - E.g., differing susceptibility with age.
Simulated disease cases

Underlying population stratified by age and other risk factors

Many realisations created (e.g., to model the expected patterns if infection was a random Poisson process free of spatial autocorrelation)
Disease mapping and modelling

Using underlying data on sociodemographic and environmental factors to identify significant clusters of disease – e.g., community-associated MRSA.

These point realisations are kernel-filtered, to produce density grid maps that can be overlaid.
PCF for real MRSA cases with 95% confidence intervals from 1000 simulated sets of cases
Disease mapping and modelling

Using underlying data on sociodemographic and environmental factors to identify **significant** clusters of disease – e.g., community-associated MRSA

Grey = 97.5\textsuperscript{th} percentile of the simulated surfaces

**Significant spatio-temporal clusters**
Disease mapping and modelling
Using underlying data on sociodemographic and environmental factors to identify significant clusters of disease –
- Community-associated MRSA
Cluster locations investigated in detail:
- Unreported outbreaks at nursing homes.
- Infections in surrounding areas.
Bastin, Rollason, Hilton et. al (2007)
Specific genetic strains of MRSA clustered in specific areas of the West Midlands.

Potential for tracking & tracing infections which move between hospital & community (e.g., C. difficile).

Ethical issues - patient confidentiality!!!
2. Web-based interpolation

- Real-time automatic mapping of environmental variables.
- **Web Services** expose R algorithms to interpolate from point observations, with clearly-quantified reliability.
- [http://www.intamap.org](http://www.intamap.org)
- Used Java to interface with R, but....
Case study – radiation

EURDEP radiation monitoring network

Web Service for interpolating and mapping risk.
Sensor networks

Sensor brand A:
Precision ± 0.3
No bias

Sensor brand B:
Gaussian error ± 0.8
Potential bias due to elevation.

Sensor brand C:
Positive exponential noise, varies with level.
Bias unknown.

Network is HETEROGENEOUS in spatial pattern and accuracy of measurements.

Projected sequential Gaussian process kriging (Cornford et al., StatGIS, 2009)
Using INTAMAP

Try INTAMAP

For a demonstration of the INTAMAP interpolation service, paste the contents of a CSV file containing observations you wish to interpolate in the box below and click 'Interpolate'. Each row in the file should represent an observation and have three or four columns: the x coordinate of the feature of interest, the y coordinate of the feature of interest, the result, and optionally the standard deviation. The CSV file should not contain a header.

Request Details

Load sample data (smaller data sets take less time - indicated time is for the psgp method):

```
-5.38, 50.22, 12.55, 0.98
-5.05, 50.42, 11.78, 2.26
-4.22, 50.40, 12.65, 1.01
-4.07, 50.33, 11.39, 1.93
-4.09, 50.36, 11.62, 1.62
-4.03, 50.33, 11.28, 2.07
-3.80, 50.24, 11.44, 2.67
-4.35, 50.81, 11.88, 2.10
-3.52, 50.38, 11.99, 0.96
-2.56, 49.47, 11.35, 4.09
-2.54, 49.45, 11.36, 3.49
```

Click to visualise
Request Details

Load sample data (smaller data sets take less time - indicated time is for the psgp method):

<table>
<thead>
<tr>
<th>Weather Underground temperatures (438 obs. with error column, time 30s)</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.38, 50.22, 12.55, 0.98</td>
<td></td>
</tr>
<tr>
<td>-5.05, 50.42, 11.78, 2.26</td>
<td></td>
</tr>
<tr>
<td>-4.22, 50.40, 12.65, 1.01</td>
<td></td>
</tr>
<tr>
<td>-4.07, 50.33, 11.39, 1.93</td>
<td></td>
</tr>
<tr>
<td>-4.09, 50.36, 11.62, 1.62</td>
<td></td>
</tr>
<tr>
<td>-4.03, 50.33, 11.28, 2.07</td>
<td></td>
</tr>
<tr>
<td>-3.80, 50.24, 11.44, 2.67</td>
<td></td>
</tr>
<tr>
<td>-4.35, 50.81, 11.88, 2.10</td>
<td></td>
</tr>
<tr>
<td>-3.52, 50.38, 11.99, 0.96</td>
<td></td>
</tr>
<tr>
<td>-2.56, 49.47, 11.35, 4.09</td>
<td></td>
</tr>
<tr>
<td>-2.54, 49.45, 11.36, 3.49</td>
<td></td>
</tr>
</tbody>
</table>

Click to visualise

- Perform basic outlier detection
- Enable Google Earth visualisation - requires observations EPSG code: 4326

WARNING: INTAMAP does not currently support interpolation in lat/lon, however when Google Earth visualisation is enabled all observation coordinates are transformed to lat/lon. Lat/lon is then treated by INTAMAP as if they are Euclidean coordinates. This implies a rather unusual covariance function is being used, which we do not expect to be optimal, particularly for data sets which span large ranges in latitude (>30 degrees) or are near the poles. The visualisation results should be used indicatively only. Full results can be determined using the API or WPS directly.

Interpolation method: psgp
Service location: Aston

Interpolate
Interpolation Results

Spatial prediction using the method psgp

Show overlay:  ● Mean  ○ Variance  ○ None /  □ Show observations

Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image © 2010 TerraMetrics
© 2010 Cnes/Spot Image

This work is funded by the European Commission, under the Sixth Framework Programme, by Contract 033811 with DG INFSO, action Line IST-2005-2.5.12 ICT for Environmental Risk Management.
3. Conservation planning

- Select sites for protection based on their potential, now *and in the future*.
- Data on species and available habitat is prone to error.
- The future context may be hard to predict.
- Can simulate various outcomes: climate change, development pressure, policy...
Identify species of interest

Example PA:
Ngorongoro
Species extent:
50% with most restricted ranges
• Identify biome ‘envelope’
  - Climatic variables from Holdridge’s lifezones

(Biotemperature, precipitation, PET/precipitation)
eHabitat - 1. Climatic similarity and forecasting
Institute for Environment and Sustainability