Applied spatial data analysis with R: status and prospects

Roger Bivand

19 February 2019

Introduction

Outline

- Spatial and spatio-temporal data are characterised by structures that distinguish them from typical tabular data
- The geometric structures also have spatial reference system information, and can adhere to standards, which may ease geometrical operations
- Satellite data and numerical model output data typically have regular grid structures, but these are often domain-specific
- Computationally intensive tasks include interpolation, upsampling, focal operations, change of support and handling vector data with very detailed boundaries, as well as modelling using Bayesian inference
- A further challenge to modelling using training sets with spatial data is how to split the observations in the presence of spatial dependence

Spatial and spatio-temporal data

Spatial data

Spatial data typically combine position data in 2D (or 3D), attribute data and metadata related to the position data. Much spatial data could be called map data or GIS data. We collect and handle much more position data since global navigation satellite system (GNSS) like GPS came on stream 20 years ago, earth observation satellites have been providing data for longer.

```
> library(osmdata)
## Data (c) OpenStreetMap contributors, ODbL 1.0. http://www.openstreetmap.o
> library(sf)
## Linking to GEOS 3.7.1, GDAL 2.4.0, PROJ 5.2.0
> bbox <- opq(bbox = 'bergen norway')
> byb0 <- osmdata_sf(add_osm_feature(bbox, key = 'railway',
+   value = 'light_rail'))$osm_lines
> tram <- osmdata_sf(add_osm_feature(bbox, key = 'railway',
+   value = 'tram'))$osm_lines
> byb1 <- tram[!is.na(tram$name),]
> o <- intersect(names(byb0), names(byb1))
> byb <- rbind(byb0[,o], byb1[,o])
> library(mapview)
> mapview(byb)
```

Vector data

Spatial vector data is based on points, from which other geometries are constructed. The light rail tracks are 2D vector data. The points themselves are stored as double precision floating point numbers, typically without recorded measures of accuracy (GNSS provides a measure of accuracy). Here, lines are constructed from points.

```
> all(st_is(byb, "XY"))
## [1] TRUE
> str(st_coordinates(st_geometry(byb)[[1]]))
##  num [1:14, 1:3] 5.33 5.33 5.33 5.33 5.33 ...
##  - attr(*, "dimnames")=List of 2
##  ..$ : chr [1:14] "4870663682" "331531217" "331531216" "331531215" ...
##   ..$ : chr [1:3] "X" "Y" "L1"
```

Raster data

Spatial raster data is observed using rectangular (often square) cells, within which attribute data are observed. We probably do not know where within the raster cell the observed value is correct; all we know is that at the chosen resolution, this is the value representing the whole cell area.

```
> library(elevatr)
> elevation <- get_elev_raster(as(byb, "Spatial"), z = 14)
## Merging DEMs
## Reprojecting DEM to original projection
## Note: Elevation units are in meters.
## Note: The coordinate reference system is:
##  +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
> elevation
## class       : RasterLayer
## dimensions  : 5663, 3595, 20358485  (nrow, ncol, ncell)
## resolution  : 4.29e-05, 2.12e-05  (x, y)
## extent     : 5.229278, 5.383503, 60.2723, 60.39236  (xmin, xmax, ymin, yma
## coord. ref. : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0
##    data    source    :    /tmp/RtmptJNayw/raster/r_tmp_2019-02-
13_152847_7095_99381.grd
## names       : layer
## values      : -5.561122, 605.8524  (min, max)
```

Spatio-temporal data

Representing spatial vector data in R (sp)

```
> library(sp)
> str(slot(as(st_geometry(byb), "Spatial"), "lines")[[1]])
## Formal class 'Lines' [package "sp"] with 2 slots
##   ..@ Lines:List of 1
##   .. ..$ :Formal class 'Line' [package "sp"] with 1 slot
##  .. .. .. ..@ coords: num [1:14, 1:2] 5.33 5.33 5.33 5.33 5.33 ...
##   .. .. .. .. ..- attr(*, "dimnames")=List of 2
##  .. .. .. .. .. ..$ : chr [1:14] "4870663682" "331531217" "331531216" "331
##   .. .. .. .. .. ..$ : chr [1:2] "lon" "lat"
##   ..@ ID   : chr "ID1"
```

Representing spatial vector data in R (sf)

```
> strwrap(st_as_text(st_geometry(byb)[[1]]))
## [1] "LINESTRING (5.333375 60.30436, 5.333386"
## [2] "60.30439, 5.333512 60.30463, 5.333664"
## [3] "60.30487, 5.3342 60.30559, 5.334472"
## [4] "60.30589, 5.334727 60.30613, 5.334901"
## [5] "60.30628, 5.33523 60.30652, 5.335494"
## [6] "60.30667, 5.335813 60.30682, 5.336282"
## [7] "60.30701, 5.336615 60.3071, 5.336872"
## [8] "60.30716)"
```

Representing spatial raster data in R (sp and raster)

Representing spatial raster data in R (sf and stars)

Spatial reference systems

Baseline WKT and PROJ4

Escaping the WGS84 hub: PROJ4 6.0 and WKT2

Earth observation

Volume and frequency

Array storage

Cloud storage

Computationally intensive tasks

Data handling

Interpolation

Inference

Inference under heterogeneity

Inference under dependence

Spatial dependence and subsetting data

Heterogeneity and dependence

Conclusions

Conclusions