Geographically Weighted Regression

Roger Bivand

June 22, 2013

Geographically weighted regression (GWR) is an exploratory technique mainly intended to indicate where non-stationarity is taking place on the map, that is where locally weighted regression coefficients move away from their global values. Its basis is the concern that the fitted coefficient values of a global model, fitted to all the data, may not represent detailed local variations in the data adequately – in this it follows other local regression implementations. It differs, however, in not looking for local variation in 'data' space, but by moving a weighted window over the data, estimating one set of coefficient values at every chosen 'fit' point. The fit points are very often the points at which observations were made, but do not have to be. If the local coefficients vary in space, it can be taken as an indication of non-stationarity.

The technique is fully described by Fotheringham et al. (2002) and involves first selecting a bandwidth for an isotropic spatial weights kernel, typically a Gaussian kernel with a fixed bandwidth chosen by leave-one-out cross-validation. Choice of the bandwidth can be very demanding, as n regressions must be fitted at each step. Alternative techniques are available, for example for adaptive bandwidths, but they may often be even more compute-intensive. GWR is discussed by Schabenberger and Gotway (2005, pp. 316–317) and Waller and Gotway (2004, p. 434), and presented with examples by Lloyd (2007, pp. 79–86).

```
> library(maptools)
> library(spdep)
> owd <- getwd()
> setwd(system.file("etc/shapes", package = "spdep"))
> NY8 <- readShapeSpatial("NY8_utm18")
> setwd(owd)
```

```
> library(spgwr)
```

```
> bwG <- gwr.sel(Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8, gweight = gwr.Gauss,
+     verbose = FALSE)
> gwrG <- gwr(Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8, bandwidth = bwG,
+     gweight = gwr.Gauss, hatmatrix = TRUE)
```

```
> gwrG
```

```
Call:
gwr(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8,
    bandwidth = bwG, gweight = gwr.Gauss, hatmatrix = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 179942.6
Summary of GWR coefficient estimates:
                 Min.  1st Qu.   Median  3rd Qu.     Max.  Global
X.Intercept. -0.52220 -0.52070 -0.52020 -0.51440 -0.51110 -0.5173
PEXPOSURE     0.04718  0.04803  0.04953  0.04972  0.05048  0.0488
PCTAGE65P     3.91200  3.93400  3.95900  3.96200  3.98000  3.9509
PCTOWNHOME   -0.55940 -0.55800 -0.55770 -0.55550 -0.55460 -0.5600
Number of data points: 281
```

```
Effective number of parameters (residual: 2traceS - traceS'S): 4.39979
Effective degrees of freedom (residual: 2traceS - traceS'S): 276.6002
Sigma (residual: 2traceS - traceS'S): 0.6575073
Effective number of parameters (model: traceS): 4.206294
Effective degrees of freedom (model: traceS): 276.7937
Sigma (model: traceS): 0.6572774
Sigma (ML): 0.6523395
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 568.0103
AIC (GWR p. 96, eq. 4.22): 561.5689
Residual sum of squares: 119.5786
Quasi-global R2: 0.1934333
```

Once the bandwidth has been found, or chosen by hand, the `gwr` function may be used to fit the model with the chosen local kernel and bandwidth. If the `data` argument is passed a `SpatialPolygonsDataFrame` or a `SpatialPointsDataFrame` object, the output object will contain a component, which is an object of the same geometry populated with the local coefficient estimates. If the input objects have polygon support, the centroids of the spatial entities are taken as the basis for analysis. The function also takes a `fit.points` argument, which permits local coefficients to be created by geographically weighted regression for other support than the data points.

The basic GWR results are uninteresting for this data set, with very little local variation in coefficient values; the bandwidth is almost 180 km. Neither `gwr` nor `gwr.sel` yet take a `weights` argument, as it is unclear how non-spatial and geographical weights should be combined. A further issue that has arisen is that it seems that local collinearity can be induced, or at least observed, in GWR applications. A discussion of the issues raised is given by Wheeler and Tiefelsdorf (2005).

As Fotheringham et al. (2002) describe, GWR can also be applied in a GLM framework, and a provisional implementation permitting this has been added to the `spgwr` package providing both cross-validation bandwidth selection and geographically weighted fitting of GLM models.

```
> gbwG <- ggwr.sel(Cases ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME + offset(log(POP8)),
+     data = NY8, family = "poisson", gweight = gwr.Gauss, verbose = FALSE)
> ggwrG <- ggwr(Cases ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME + offset(log(POP8)),
+     data = NY8, family = "poisson", bandwidth = gbwG, gweight = gwr.Gauss)

> ggwrG

Call:

ggwr(formula = Cases ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME +
    offset(log(POP8)), data = NY8, bandwidth = gbwG, gweight =
    gwr.Gauss, family = "poisson")

Kernel function: gwr.Gauss
Fixed bandwidth: 179942.6
Summary of GWR coefficient estimates:
                Min. 1st Qu.  Median 3rd Qu.    Max.  Global
X.Intercept. -8.1380 -8.1360 -8.1350 -8.1350 -8.1320 -8.1344
PEXPOSURE     0.1470  0.1478  0.1491  0.1493  0.1500  0.1489
PCTAGE65P     3.9750  3.9820  3.9840  4.0060  4.0180  3.9982
PCTOWNHOME   -0.3573 -0.3552 -0.3545 -0.3489 -0.3460 -0.3571
```

The local coefficient variation seen in this fit is not large either, although from Fig. 1 it appears that slightly larger local coefficients for the closeness to TCE site covariate are found farther away from TCE sites than close to them. If, on the other hand, we consider this indication in the light of Fig. 2, it is clear that the forcing artefacts found by Wheeler and Tiefelsdorf (2005) in a different data set are replicated here.

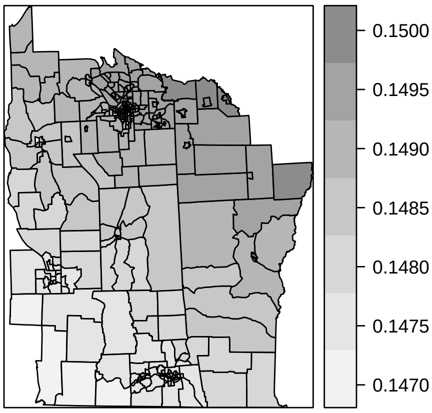

Figure 1: GWR local coefficient estimates for the exposure to TCE site covariate

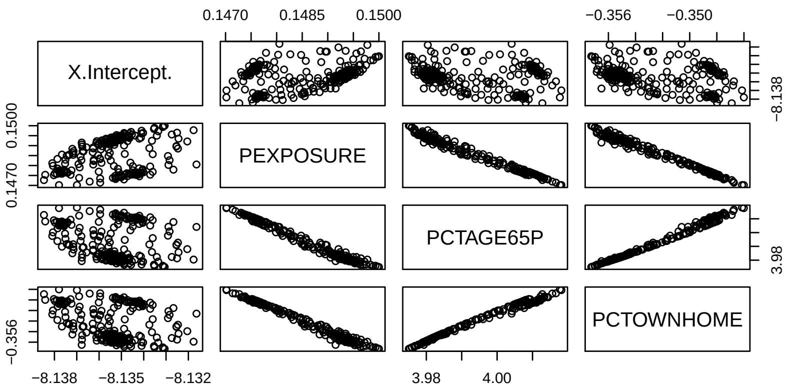

Figure 2: Pairs plots of GWR local coefficient estimates showing the effects of GWR collinearity forcing

References

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.

Lloyd, C. D. (2007). *Local Models for Spatial Analysis*. CRC, Boca Raton.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, London.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley, Hoboken, NJ.

Wheeler, D. and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7:161–187.