Spatial Statistical Modeling and Prediction in R Using spmodel

Spatial Statistics 2023: Climate and the Environment

Michael Dumelle (presenting)
EPA (USA)

Matt Higham
St. Lawrence University (USA)

Jay M Ver Hoef
NOAA (USA)

Disclaimer

The views expressed in this workshop are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency or the U.S. National Oceanic and Atmospheric Administration. Any mention of trade names, products, or services does not imply an endorsement by the U.S. government, the U.S. Environmental Protection Agency, or the U.S. National Oceanic and Atmospheric Administration. The U.S. Environmental Protection Agency and the U.S. National Oceanic and Atmospheric Administration do not endorse any commercial products, services, or enterprises.

What is `spmodel`?

`spmodel` is an `R` package to fit, summarize, and predict for a variety of spatial statistical models. Some of the things that `spmodel` can do include:

- Fit spatial linear and generalized linear models for point-referenced and areal (lattice) data
- Compare model fits and inspect model diagnostics
- Predict at unobserved spatial locations (i.e., Kriging)
- And much more!

Why use `spmodel`?

There are many great spatial modeling packages in `R`. A few reasons to use `spmodel` for spatial analysis are that:

A Basic Overview

Goals

1. Fit a spatial linear model using `splm()`.
2. Tidy, glance at, and augment the fitted model.
3. Predict for unobserved locations (i.e., Kriging).
4. Explore other `spmodel` features and provide resources to learn more

The Sulfate Data

The `sulfate` data in `spmodel` contains data on 197 sulfate measurements in the continental United States

```
1 head(sulfate)
```

```
Simple feature collection with 6 features and 1 field
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: 162932.8 ymin: 1080571 xmax: 914593.6 ymax: 1453636
Projected CRS: NAD83 / Conus Albers
  sulfate                  geometry
1  12.925 POINT (817738.8 1080571)
2  20.170 POINT (914593.6 1295545)
3  16.822 POINT (359574.1 1178228)
4  16.227 POINT (265331.9 1239089)
5   7.858 POINT (304528.8 1453636)
6  15.358 POINT (162932.8 1451625)
```

The Sulfate Data

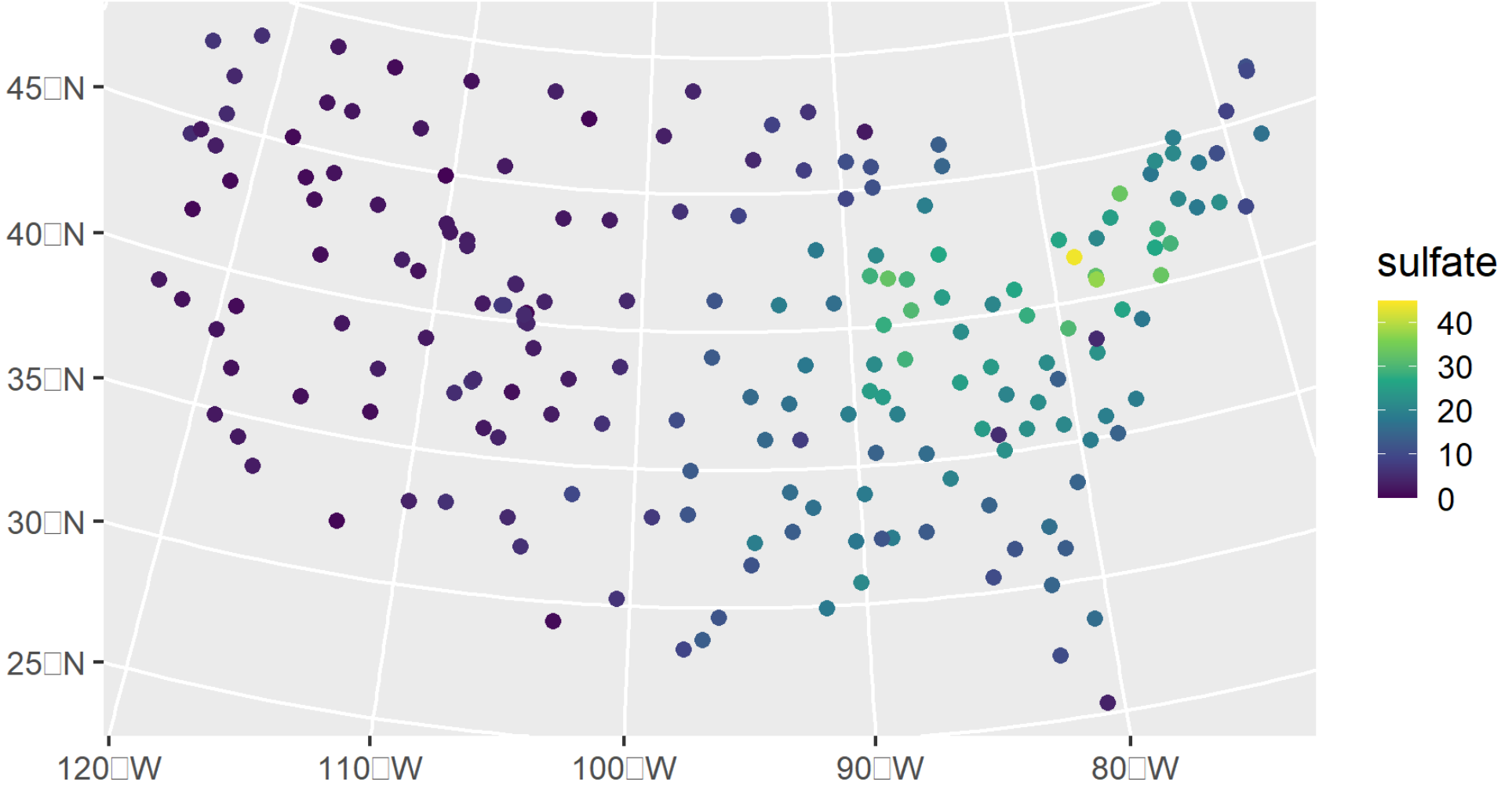

Figure 1: Distribution of sulfate data.

Fitting a Model

We fit and summarize a spatial linear model with an intercept by running

```
spmod <- splm(sulfate ~ 1, data = sulfate, spcov_type = "exponential")
summary(spmod)
```

```
Call:
splm(formula = sulfate ~ 1, data = sulfate, spcov_type = "exponential")

Residuals:
   Min     1Q Median     3Q    Max
-5.738 -2.605  4.900 13.323 38.099

Coefficients (fixed):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    5.924      6.529   0.907    0.364

Coefficients (exponential spatial covariance):
       de         ie      range
     85.8       10.4 3105165.7
```

The broom Functions

Tidy the fixed effect output

```
1 tidy(spmod)
```

```
# A tibble: 1 × 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)     5.92      6.53     0.907   0.364
```

Glance at the model fit

```
1 glance(spmod)
```

```
# A tibble: 1 × 9
      n     p  npar value   AIC  AICc logLik deviance pseudo.r.squared
  <int> <dbl> <int> <dbl> <dbl> <dbl>  <dbl>    <dbl>              <dbl>
1   197     1     3 1140. 1146. 1146.  -570.     196.                  0
```

Augment the data with model diagnostics

```
1 augment(spmod)
```

The broom Functions

```
Simple feature collection with 197 features and 6 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: -2292550 ymin: 386181.1 xmax: 2173345 ymax: 3090370
Projected CRS: NAD83 / Conus Albers
# A tibble: 197 × 7
   sulfate .fitted .resid    .hat   .cooksd .std.resid           geometry
 *   <dbl>   <dbl>  <dbl>   <dbl>     <dbl>      <dbl>        <POINT [m]>
 1  12.9      5.92   7.00 0.00334 0.00161      -0.694  (817738.8 1080571)
 2  20.2      5.92  14.2  0.00256 0.00192       0.865  (914593.6 1295545)
 3  16.8      5.92  10.9  0.00259 0.000395      0.390  (359574.1 1178228)
 4  16.2      5.92  10.3  0.00239 0.000363      0.390  (265331.9 1239089)
 5   7.86     5.92   1.93 0.00202 0.00871      -2.07   (304528.8 1453636)
 6  15.4      5.92   9.43 0.00201 0.000240      0.345  (162932.8 1451625)
 7   0.986    5.92  -4.94 0.00380 0.000966     -0.503  (-1437776 1568022)
```

Prediction (i.e., Kriging)

```
predict(spmod, newdata = sulfate_preds)
```

Augment prediction data

```
augment(spmod, newdata = sulfate_preds)
```

```
Simple feature collection with 100 features and 1 field
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: -2283774 ymin: 582930.5 xmax: 1985906 ymax: 3037173
Projected CRS: NAD83 / Conus Albers
# A tibble: 100 × 2
   .fitted             geometry
 *   <dbl>          <POINT [m]>
 1    1.62  (-1771413 1752976)
 2   24.4    (1018112 1867127)
 3    8.95 (-291256.8 1553212)
 4   16.5    (1274293 1267835)
 5    4.93 (-547437.6 1638825)
 6   26.8    (1445080 1981278)
 7    2.87  (-1629090 3037173)
 8   14.2    (1202757 1930524)
```

Prediction (i.e., Kriging)

Visualize predictions

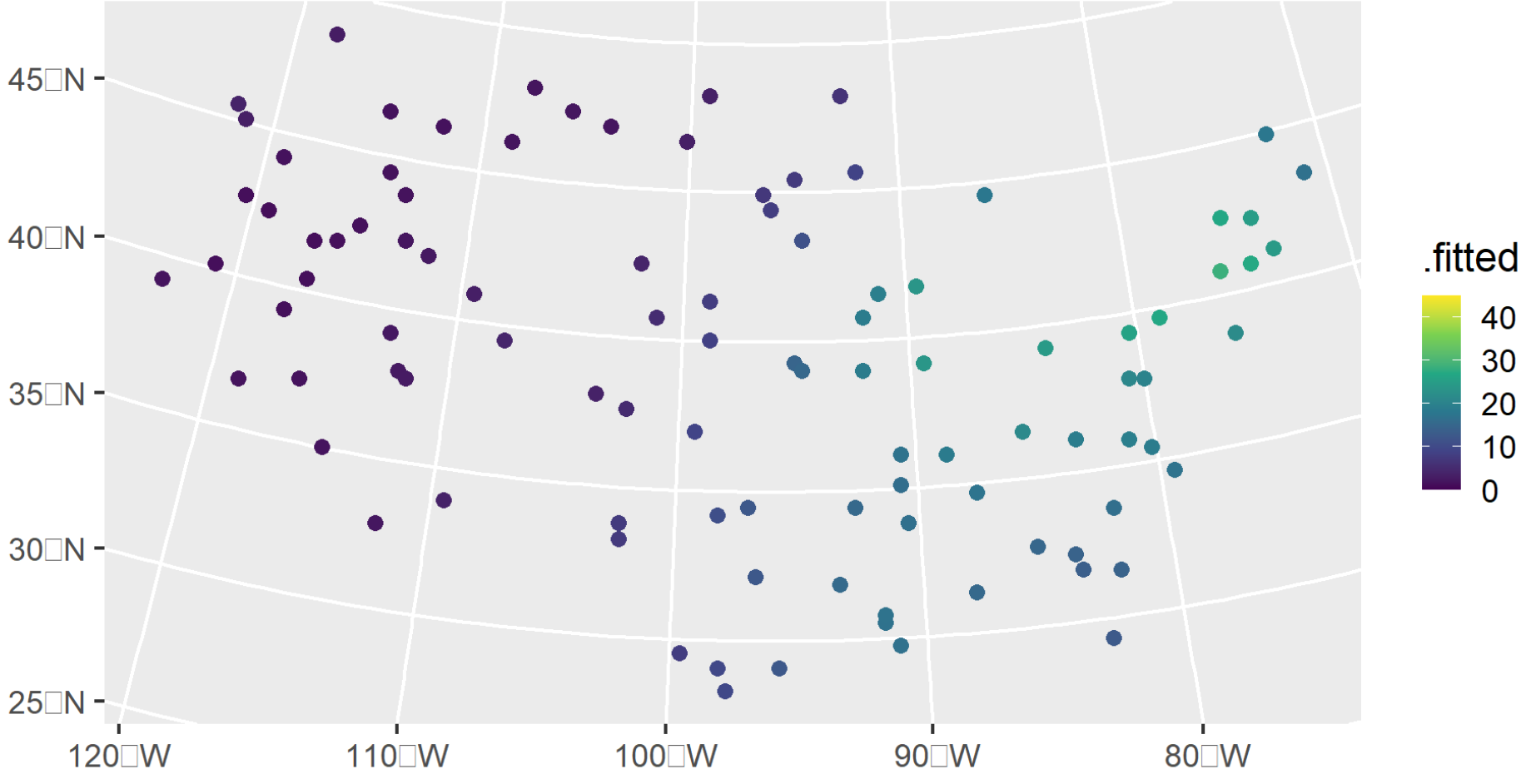

Figure 2: Distribution of sulfate data predictions.

Other Features

Other `spmodel` features include:

1. Support for non-spatial random effects, anisotropy, and large data sets
2. Support for spatial generalized linear models (`spglm()`)
3. Support for areal (i.e., lattice) data (`spautor()`; `spgautor()`)
4. Simulating spatially-dependent data from various response distributions (e.g., `sprnorm()`)
5. Much more!

Learn More

- Visit our website at https://usepa.github.io/spmodel/
- Visit our workshop workbook at https://usepa.github.io/spmodel.spatialstat2023/
- Please reach out with comments / questions / suggestions / bugs (Dumelle.Michael@epa.gov)
- Thank you for attending and enjoy the rest of the conference!