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 features:

- Fit spatial linear and generalized linear models for point-referenced data; geostatistical models
- Fit spatial linear and generalized linear models for areal (lattice) data; autoregressive models
- 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

```r
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

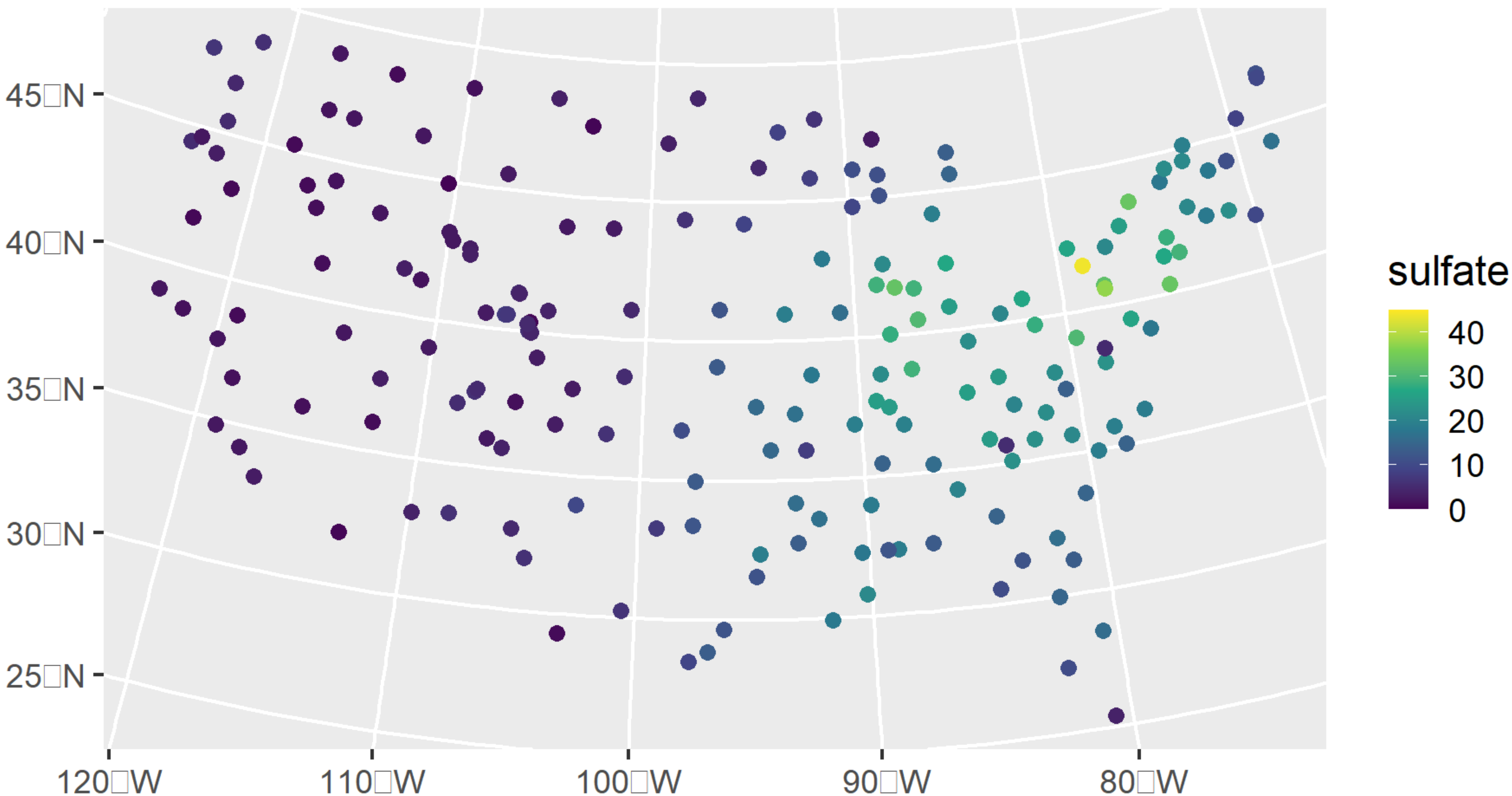

Figure 1: Distribution of sulfate data.

Fitting a Non-Spatial Model

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

```r
lmod <- lm(sulfate ~ 1, data = sulfate)
summary(lmod)
```

```

Call:
lm(formula = sulfate ~ 1, data = sulfate)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.997  -8.864  -1.359   7.064  31.840 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   12.183      0.683   17.84   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.587 on 196 degrees of freedom
```

Fitting a Spatial Model

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

```r
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

```
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

```
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
```

The broom Functions

Augment the data with model diagnostics

```r
augment(spmod)
```

```
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)
```

The Sulfate Prediction Data

The `sulfate_preds` data contains 100 locations at which to predict sulfate

```
1  head(sulfate_preds)
```

```
Simple feature collection with 6 features and 0 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: -1771413 ymin: 1267835 xmax: 1445080 ymax: 1981278
Projected CRS: NAD83 / Conus Albers
# A tibble: 6 × 1
             geometry
          <POINT [m]>
1  (-1771413 1752976)
2   (1018112 1867127)
3 (-291256.8 1553212)
4   (1274293 1267835)
5 (-547437.6 1638825)
6   (1445080 1981278)
```

The Sulfate Prediction Data

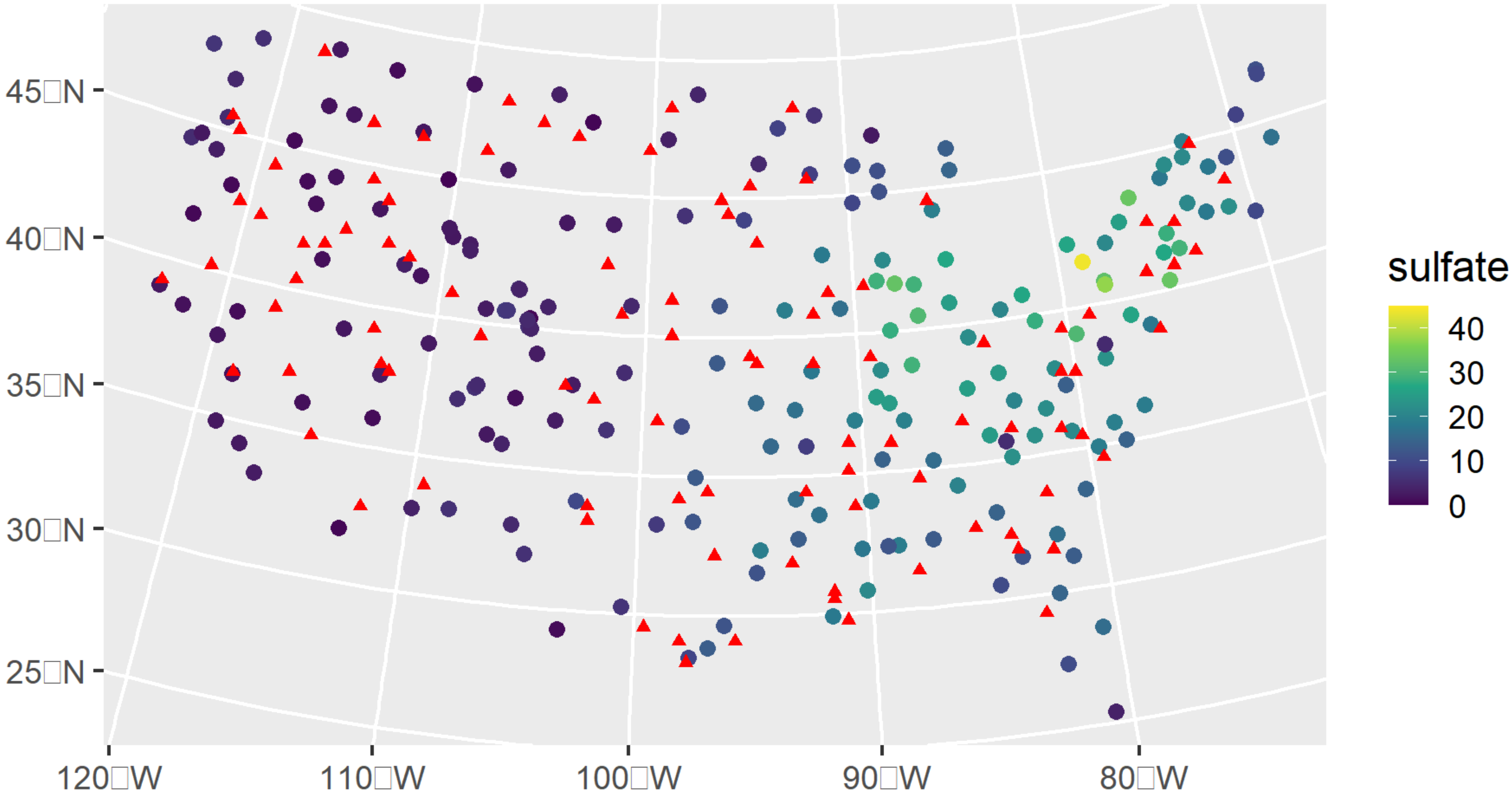

Figure 2: Distribution of sulfate data and prediction locations.

Prediction (i.e., Kriging)

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

Augment prediction data

```r
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)
```

Prediction (i.e., Kriging)

Visualize predictions

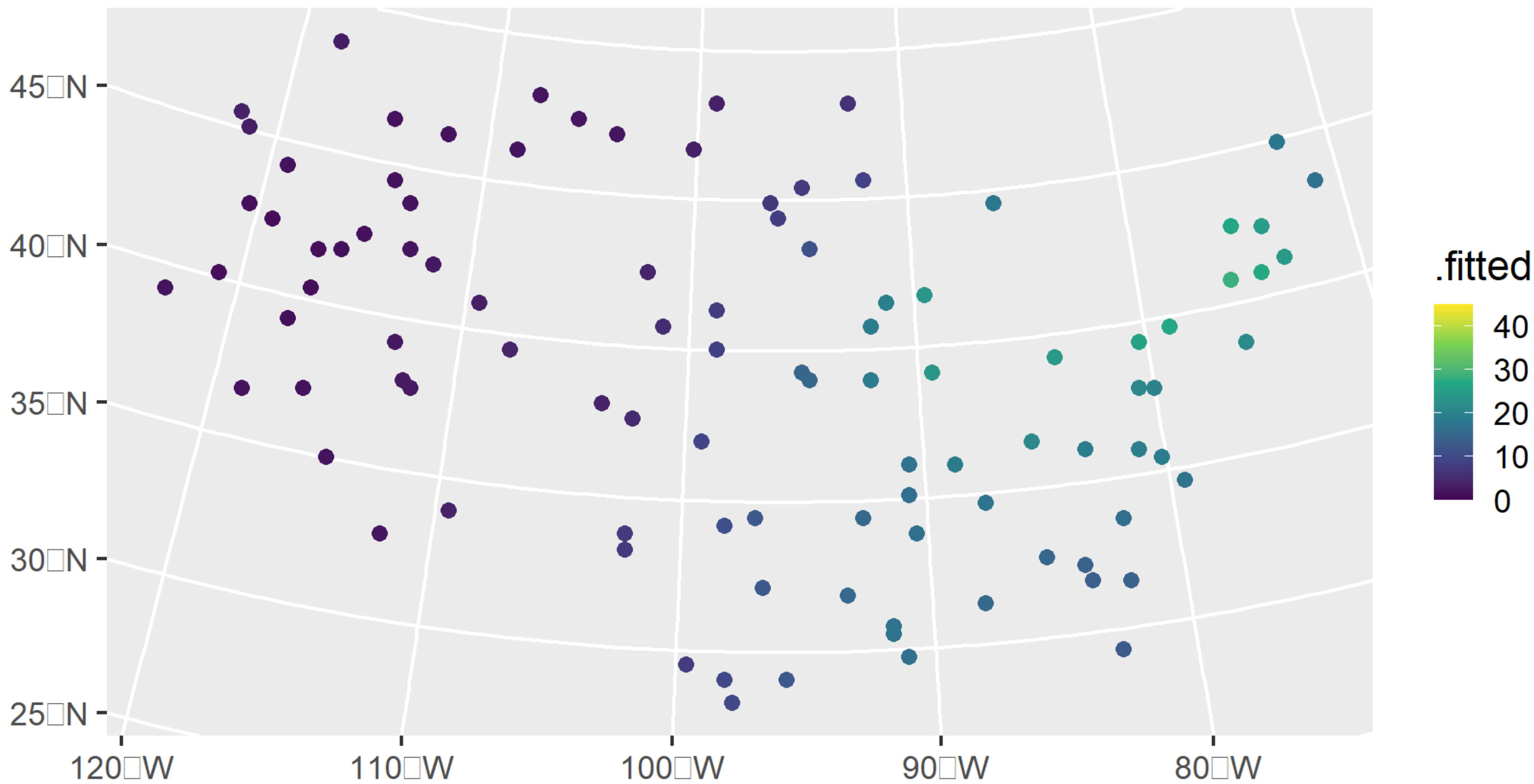

Figure 3: Distribution of sulfate data predictions.

Other Features

Other `spmodel` features include:

1. Support for fixing spatial covariance parameters, non-spatial random effects, anisotropy, and large data sets (Ver Hoef, Dumelle, et al. 2023)
2. Support for areal (i.e., lattice) data (`spautor()`)
3. Simulating spatially-dependent data from various response distributions (e.g., `sprnorm()`)
4. Spatial generalized linear models (Ver Hoef, Blagg, et al. 2023)
5. Much more!

The Moose Data

The moose data in `spmodel` contains data on 218 observations of moose presence/absence in Alaska, USA

```r
head(moose)
```

```
Simple feature collection with 6 features and 4 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: 281896.4 ymin: 1518398 xmax: 311325.3 ymax: 1541016
Projected CRS: NAD83 / Alaska Albers
      elev strat count presence                 geometry
1 468.9167     L     0        0 POINT (293542.6 1541016)
2 362.3125     L     0        0 POINT (298313.1 1533972)
3 172.7500     M     0        0 POINT (281896.4 1532516)
4 279.6250     L     0        0 POINT (298651.3 1530264)
5 619.6000     L     0        0 POINT (311325.3 1527705)
6 164.1250     M     0        0 POINT (291421.5 1518398)
```

The Moose Data

Visualize presence/absence data

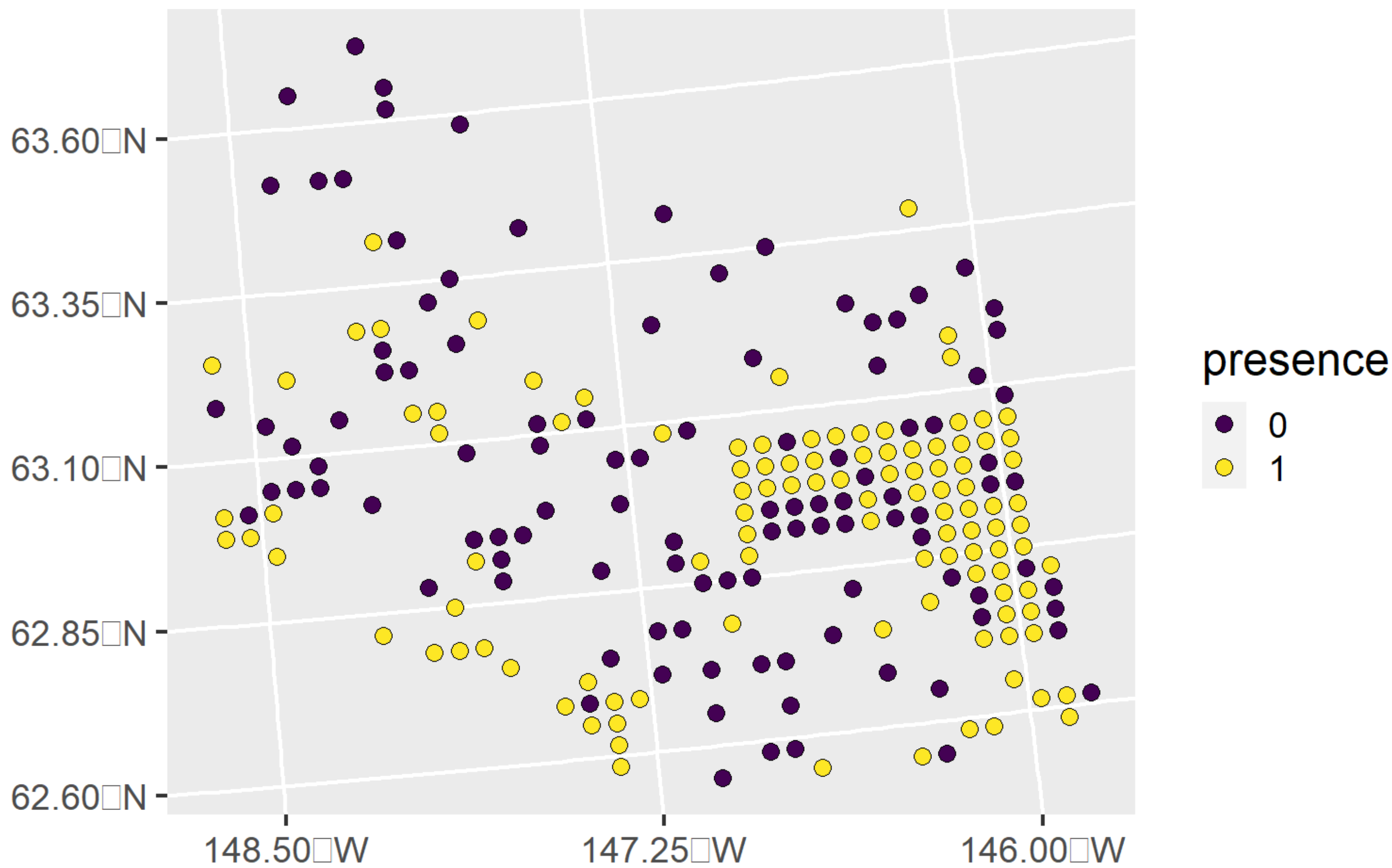

Figure 4: Distribution of moose presence/abscence.

Fitting a Non-Spatial Generalized Linear Model

Model moose presence as a function of elevation, strata, and their interaction

```
gmod <- glm(presence ~ elev * strat, family = binomial,
               data = moose)
summary(gmod)
```

Fitting a Non-Spatial Generalized Linear Model

```
Call:
glm(formula = presence ~ elev * strat, family = binomial, data = moose)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-1.347  -1.249    1.013    1.050    1.397

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.5219235  0.5081825  -1.027    0.304
elev         0.0001895  0.0023700   0.080    0.936
stratM       1.0274043  0.7149642   1.437    0.151
elev:stratM -0.0012737  0.0037689  -0.338    0.735

(Dispersion parameter for binomial family taken to be 1)
```

Fitting a Spatial Generalized Linear Model

```r
spgmod <- spglm(presence ~ elev * strat, family = binomial,
                data = moose, spcov_type = "matern")
summary(spgmod)
```

- Also for `spgautor()`

Fitting a Spatial Generalized Linear Model

```
Call:
spglm(formula = presence ~ elev * strat, family = binomial, data = moose,
    spcov_type = "matern")

Deviance Residuals:
    Min       1Q  Median       3Q      Max  
-1.9731 -0.8232  0.4238  0.7713  1.8166  

Coefficients (fixed):
             Estimate Std. Error z value Pr(>|z|)   
(Intercept) -3.181148   1.290195  -2.466  0.01368 * 
elev         0.009939   0.004021   2.472  0.01344 * 
stratM       3.336690   1.098589   3.037  0.00239 **
elev:stratM -0.011573   0.006412  -1.805  0.07110 . 
```

Learn More

- Visit our website at https://usepa.github.io/spmodel/
- Visit our workbook at https://usepa.github.io/spmodel.spatialstat2023/
- Please reach out with feedback (Dumelle.Michael@epa.gov)
- Thank you for attending!

References

Ver Hoef, Jay M, Eryn Blagg, Michael Dumelle, Philip M Dixon, Dale L Zimmerman, and Paul Conn. 2023. "Marginal Inference for Hierarchical Generalized Linear Mixed Models with Patterned Covariance Matrices Using the Laplace Approximation." *arXiv Preprint arXiv:2305.02978*.

Ver Hoef, Jay M, Michael Dumelle, Matt Higham, Erin E Peterson, and Daniel J Isaak. 2023. "Indexing and Partitioning the Spatial Linear Model for Large Data Sets." *arXiv Preprint arXiv:2305.07811*.