

Laboratorio 3

Katherine Barquero

24/9/2020

El presente laboratorio estima y evalúa diferentes métodos para realizar interpolación a datos estadísticos. Entre los métodos o modelos utilizados contemplan los siguientes:

- Proximity polygons
- Nearest neighbour
- Inverse distance weighted (IDW)
- Ordinary kriging
- Thin plate spline model
- Ponderación de métodos usando el RMSE

Question 1: Describe what each step in the code chunk above does

- Se cargan los datos y las librerías a utilizar
- Se estima la variable de interés, en este caso la precipitación anual
- Se realiza un mapa
- Se estima el modelo nulo y el valor del RMSE
- Se realiza la interpolación utilizando la técnica de proximidad de polígonos
- Se evalúa el ajuste del modelo utilizando validación cruzada, para esto los pasos a seguir fueron: se definen 5 repeticiones, se establece la muestra de datos que se usaran para el testing y la muestra para el entrenamiento, se estima el modelo de interpolación utilizando la función de Voronoi y se calcula el RMSE para evaluar el ajuste del modelo. Se compara el valor del RMSE con respecto al modelo nulo, para identificar si los datos estimados se ajustan lo mejor posible a los observados.

Question 2: How does the proximity-polygon approach compare to the NULL model?

Los polígonos de proximidad parten la zona de interés en secciones más pequeñas según la cantidad de observaciones puntuales que haya agrupadas. En cada polígono se tiene que interpolar hacia áreas más grandes. En la validación cruzada se estima el RMSE para cada una de las 5 interpolaciones y se compara con respecto al obtenido en el modelo nulo. En todos los casos se obtiene un valor del RMSE menor al modelo nulo, lo que quiere decir que al utilizar la función de Voronoi y al estimar los promedios con los datos interpolados, se obtiene un mejor ajuste.

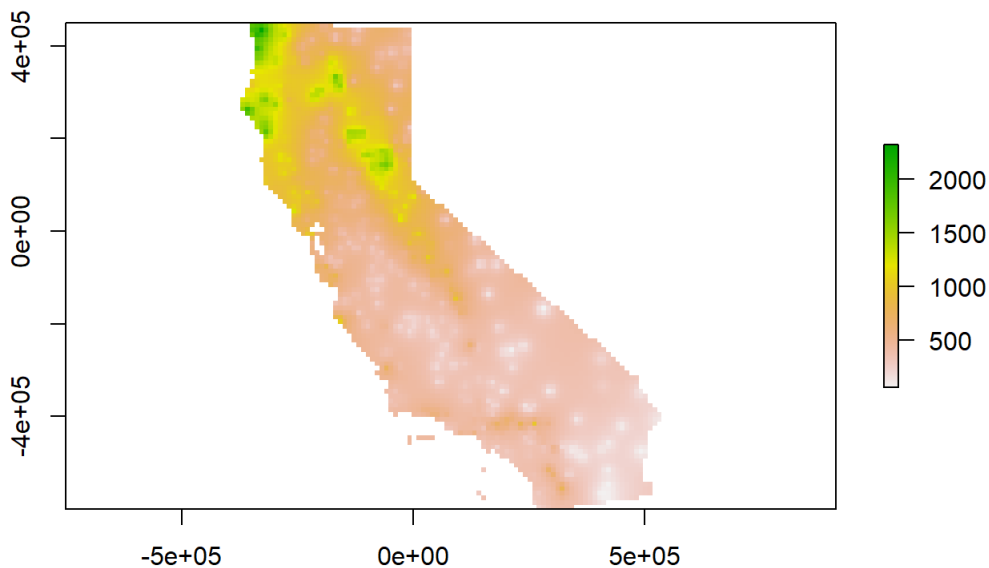
Question 3: You would not typically use proximity polygons for rainfall data. For what kind of data would you use them?

La técnica de proximidad de polígonos se utiliza para variables categóricas, además por la representación de los datos de este ejercicio podría ser mejor su uso en conjuntos de datos que presenten una cobertura más uniforme. Puede ser que los polígonos varíen si los datos cambian en el tiempo, dependerá del tipo de datos que se utilicen.

Question 4: IDW generated rasters tend to have a noticeable artefact. What is that?

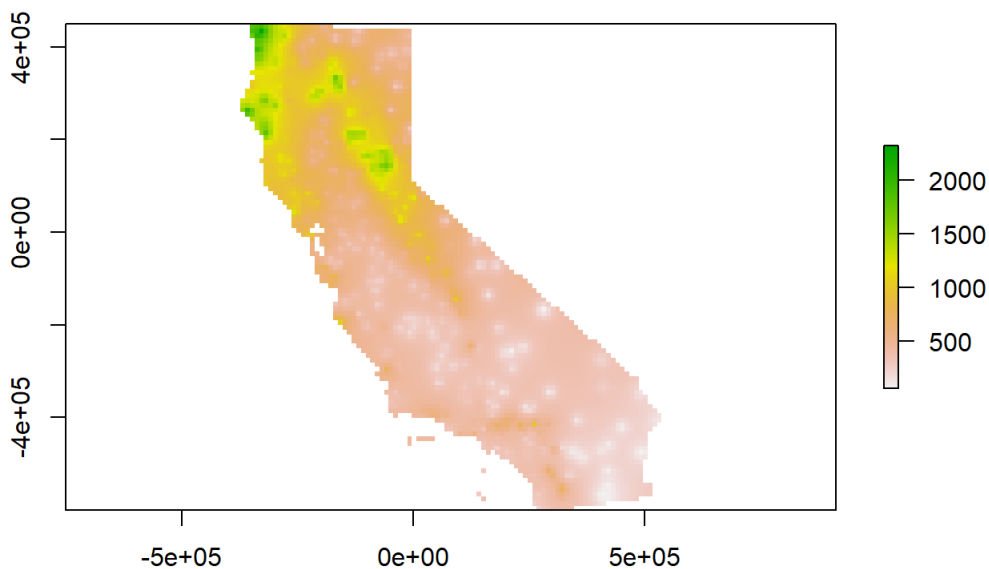
Los artefactos pueden ser entendidos como las zonas más blancas del mapa en las que se cuenta con mediciones y por lo tanto no es necesario interpolar. En este caso el IDW tiende a sobreestimar la incidencia de las precipitaciones en las zonas donde hay menos información puntual disponible.

```
## [inverse distance weighted interpolation]
```



```
## [inverse distance weighted interpolation]
```

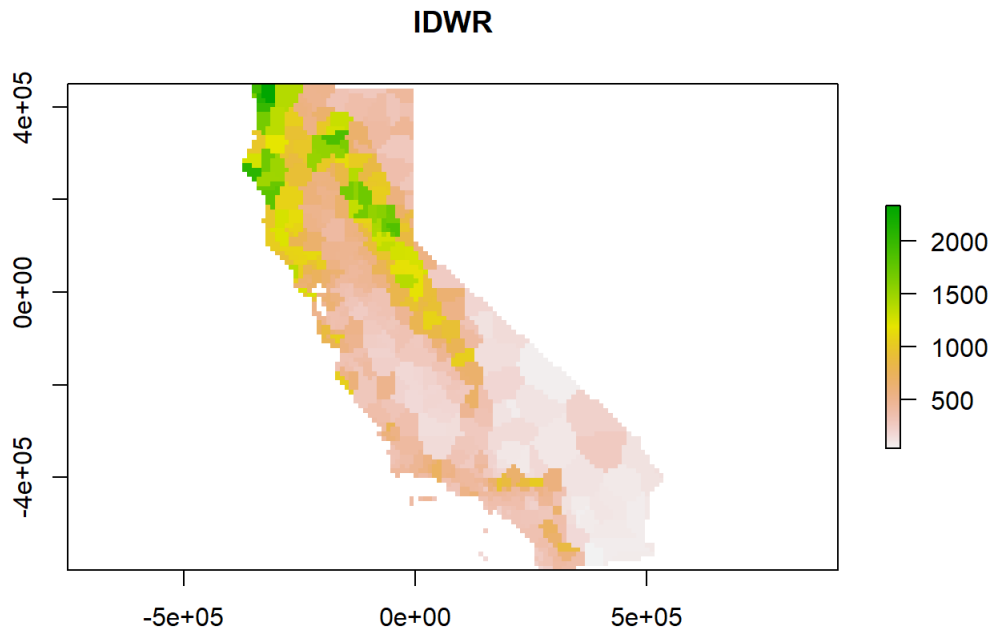
Inverse distance weighted



Question 5: Inspect the arguments used for and make a map of the IDW model below. What other name could you give to this method (IDW with these parameters)? Why?

- Se le asigna todo el ponderador al único vecino.
- Cambia al algoritmo de idp (inverse distance power) a 1 y solo un vecino
- El nombre que podría dársele sería el vecino más cercado
- Toma el vecino o unidad más cercana para realizar la interpolación. Al fijar el idp en 1, todo el peso se le otorga a ese único vecino. Fija la estimación al vecino más cercano.

```
## [inverse distance weighted interpolation]
```



Question 6: Which method performed best?

Se estiman tres métodos de interpolación: inverse distance weighted interpolation (idw), el kriging ordinario y el ultimo thin plate spline.

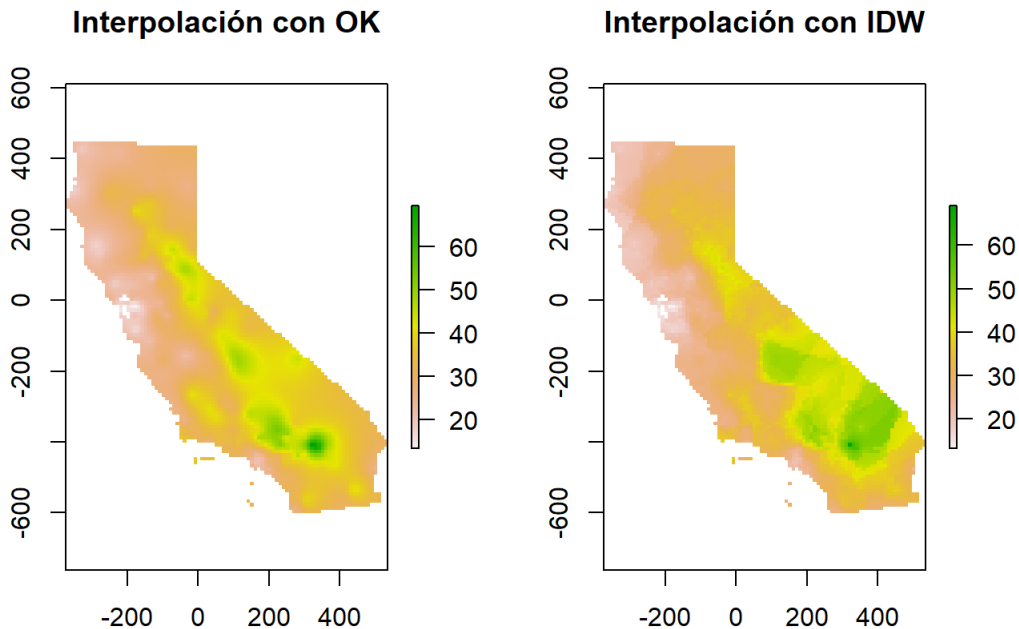
Se observa a partir de la validación cruzada que el modelo con mejor ajuste es el último mencionado, mientras que el que presenta peor desempeño es el Kriging ordinario. Además se estima un modelo ponderado de los tres métodos (llamado en este caso Ensrms) y es el que presenta un mejor ajuste.

Resultados del RMSE para cada modelo estimado

| Modelo | RMSE |
|--------|------------------|
| IDWR | 8.04130523969651 |
| Tps | 8.30723543024476 |
| Krig | 7.93079916633226 |
| Ensrms | 7.8580510818181 |

Question 7: Show where the largest difference exist between IDW and OK.

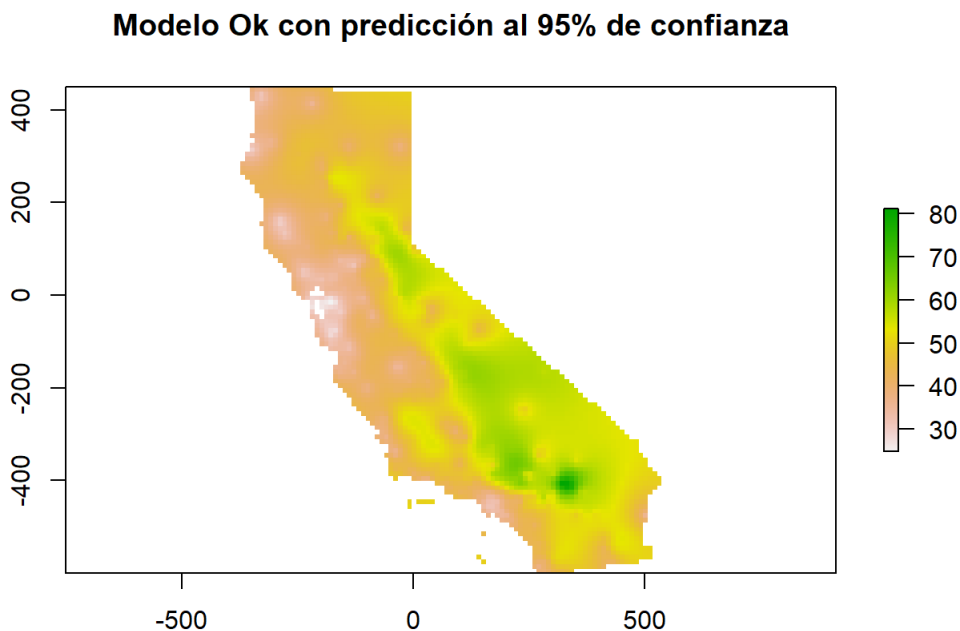
IDW tiende a interpolar puntos con valores más altos en comparación con el OK que si presenta un patrón similar a los modelos que obtuvieron un menor valor del RMSEA. Como se observa en el gráfico adjunto, si se realiza la interpolación el modelo IDW los datos tienden a concentrarse en la parte inferior derecha del mapa, captando pocos datos en la parte norte en comparación con el modelo OK.



Question 8: Show where the difference between IDW and OK is within the 95% confidence limit of the OK prediction.

Al utilizar el porcentaje de confianza, tal y como se muestra en el gráfico adjunto, la interpolación es más similar a la obtenida a partir del modelo IDW, como se mostró en el gráfico anterior. Sin embargo, aun así el modelo sigue concentrando mas datos en la parte norte que no es obtenido a través del método IDW.

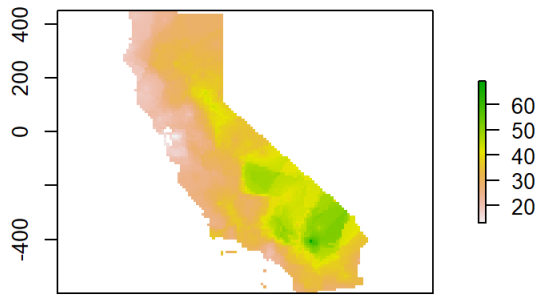
```
## [using ordinary kriging]
```



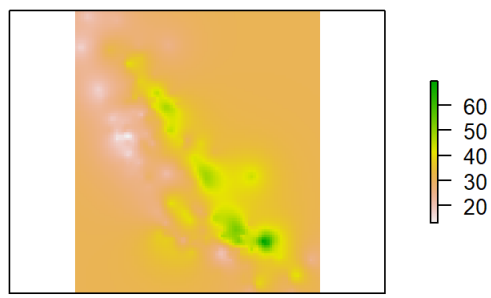
Question 9: Can you describe the pattern we are seeing, and speculate about what is causing it?

Tres de los cuatro modelos que se están comparando muestran un mejor ajuste de los datos, tienen patrones similares en el centro, sin embargo en la parte norte que es donde hay menos datos se estiman menos puntos. La especulación es que la falta de datos puede afectar la interpolación en la zona norte o bien es necesario que al estudiar el fenómeno desde la teoría y consultar a los expertos para verificar si la zona teóricamente tiene sentido que tenga menos contaminación.

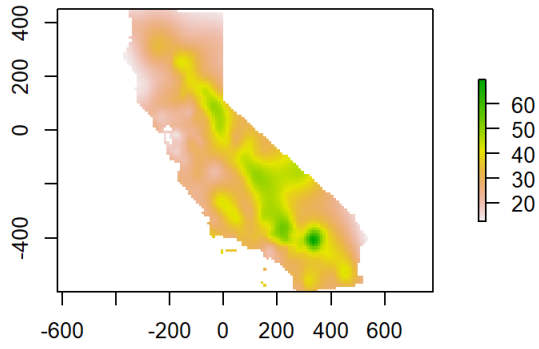
IDW



OK



TPS



Ensemble

