

SPDB - dokumentacja

Paweł Kamiński, Robert Sawicki

Temat projektu

Zbadanie funkcjonalności dostępnej w środowisku R dotyczącej predykcji przestrzennej metodą kriging (uniwersalny, zwykły, prosty), predykcji IDW (IDW - inverse distance weighted interpolation) oraz regresji liniowej. Celem zadania jest ocena dostępnych metod pod kątem ich przydatności do analizy danych przestrzennych: poprawność działania, czas wykonania w zależności od wielkości danych, łatwość użycia, dostępność materiałów pomocniczych.

Zbiór danych - pakiet Riem

- Ten pakiet pozwala na uzyskanie danych pogodowych ze stacji ASOS (lotniska)
- Dane to pomiary ze stacji obserwacyjnych we Francji, z dnia 6 grudnia 2019 roku
- Dane zawierają obserwacje z 97 stacji
- Zostały podjęte próby aproksymacji dla innych krajów
- github.com/ropensci/riem
- cran.r-project.org/package=riem

```
# A tibble: 209 x 31
  station valid      lon    lat    tmpf
  <chr>    <dtm>    <dbl> <dbl> <dbl>
1 RJNG    2019-11-02 02:23:00 137.  35.4  66.2
2 RJNG    2019-11-02 02:48:00 137.  35.4  66.2
3 RJNG    2019-11-02 16:38:00 137.  35.4  55.4
4 RJNG    2019-11-02 23:35:00 137.  35.4  57.2
5 RJNG    2019-11-03 02:35:00 137.  35.4  68
6 RJNG    2019-11-03 13:46:00 137.  35.4  57.2
7 RJNG    2019-11-03 14:21:00 137.  35.4  55.4
8 RJNG    2019-11-03 18:23:00 137.  35.4  53.6
9 RJNG    2019-11-07 05:29:00 137.  35.4  66.2
10 RJNG   2019-11-07 09:23:00 137.  35.4  62.6
```

Kriging

Metoda aproksymacji poprzez interpolację wartości o rozkładzie normalnym. Często używana w geostatystyce do przewidywania prawdopodobieństwa wystąpienia pewnego zjawiska, na przykład złoża surowców lub określonej temperatury.

W języku R pakiet: KRIG

Kriging - odmiany

- Ordinary (zwyczajny) kriging - zakłada stałą średnią w sąsiedztwie punktu pomiarowego
- Simple (prosty) kriging - zakłada stałą średnią zmiennej w całej dziedzinie
- Universal (uniwersalny) kriging - zakłada wielomianowy trend zmian zmiennej

Inverse distance weighting

Metoda interpolacji polegająca na obliczaniu wartości w badanym punkcie poprzez obliczenie średniej ważonej punktów pomiarowych. Wagi poszczególnych składników są odwrotnie proporcjonalne do odległości między punktem nieznanym i punktem próbki.

W języku R pakiet: biblioteki gstat i sp

Regresja liniowa

Metoda obliczania wartości dla nieznanego punktu poprzez przybliżenie wszystkich punktów pomiarowych funkcją liniową.

Opis środowiska, w którym zostały wykonane testy

```
q # ~ > lscpu * pkaminski/prod @ 19:43:53
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                8
On-line CPU(s) list:   0-7
Thread(s) per core:    2
Core(s) per socket:    4
Socket(s):             1
NUMA node(s):         1
Vendor ID:             GenuineIntel
CPU family:            6
Model:                142
Model name:            Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz
Stepping:              10
CPU MHz:               800.008
CPU max MHz:           4000,0000
CPU min MHz:           400,0000
BogoMIPS:              3984.00
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              256K
L3 cache:              8192K
NUMA node0 CPU(s):     0-7
```


Miary dokładności - wartość bezwzględna różnicy temperatur

	Prosty kriging	Zwykły kriging	Uniwersalny kriging	Metoda Sheparda (IDW)	Regresja liniowa
Średnia arytmetyczna	1.495832	1.362208	1.520102	1.507976	2.413334
Odchylenie standardowe	1.0975726	0.7688928	0.9606059	1.1766775	1.6541248
Liczba pomiarów	500	500	500	500	500

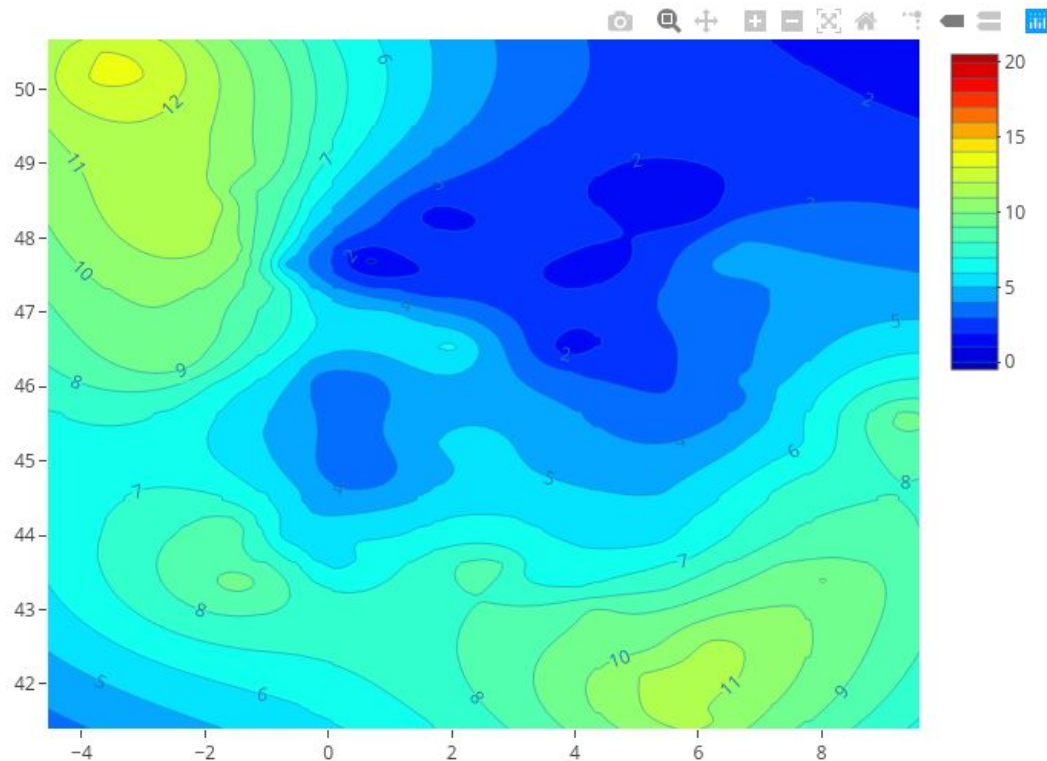
Miary dokładności - opis

- Każda metoda interpolacji została uruchomiona 100 razy za każdym razem z usunięciem 5 punktów ze zbioru pomiarowego
- Wartości temperatury dla usuniętych punktów zostały porównane z wynikami wynikającymi z metod interpolacyjnych
- Z tak otrzymanych wyników obliczono średnie arytmetyczne oraz odchylenia standardowe wartości bezwzględnej różnicy temperatur
- Na podstawie tabelki widać, że zwykły kriging średnio daje estymaty najbliższe prawdziwym wartościom
- Prosty kriging oraz uniwersalny kriging dają minimalnie gorsze estymaty
- Metoda Sheparda prowadzi do podobnej średniej co prosty i uniwersalny kriging, jednakże z większym odchyleniem
- Regresja liniowa dała najgorsze rezultaty

Wyniki pomiarów

Metoda simple kriging dla danych pomiarowych z Francji.

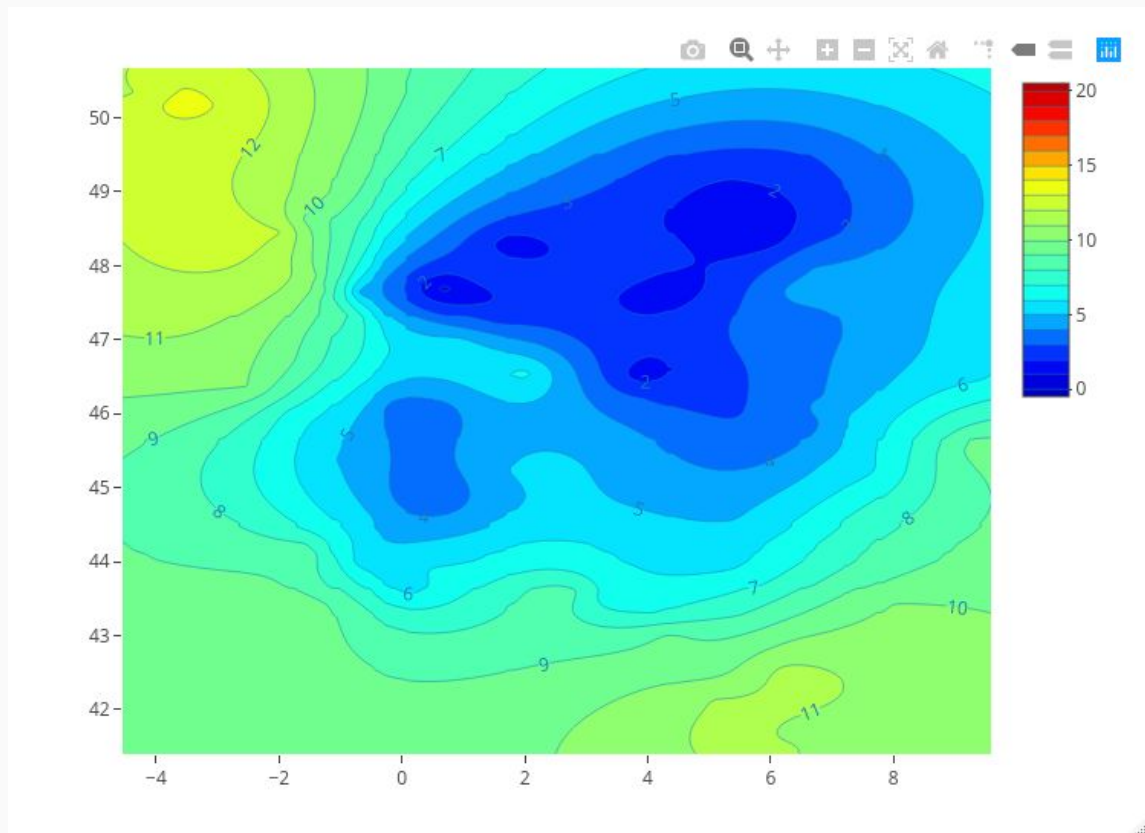
Przyglądając się wykresowi można zauważyć wystąpienie minimum lokalnego w lewym dolnym rogu.



Wyniki pomiarów c.d.

Metoda ordinary kriging dla pomiarów z Francji

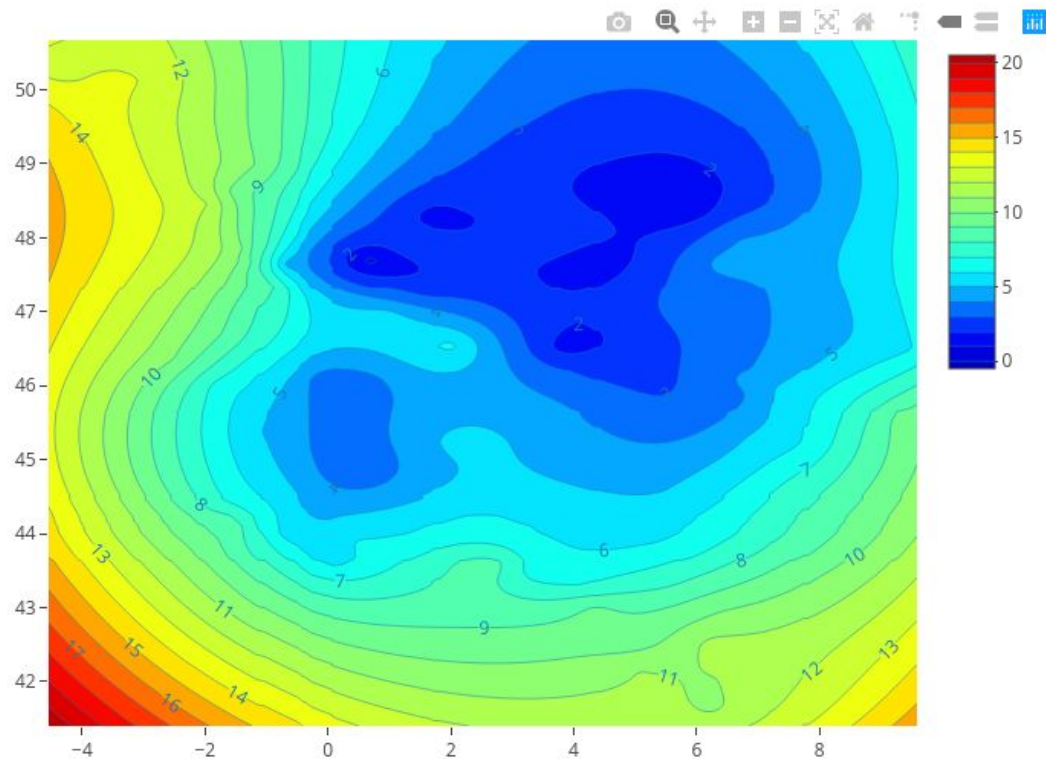
Rezultat metody ordinary kriging wizualnie przypomina rezultat metody simple kriging. Różnicą jest brak zaobserwowanego wcześniej minimum lokalnego. W tym wypadku punkty ekstrapolowane mają wartości bliższe "średniej" wartości.



Wyniki pomiarów c.d.

Metoda universal kriging dla pomiarów
z Francji

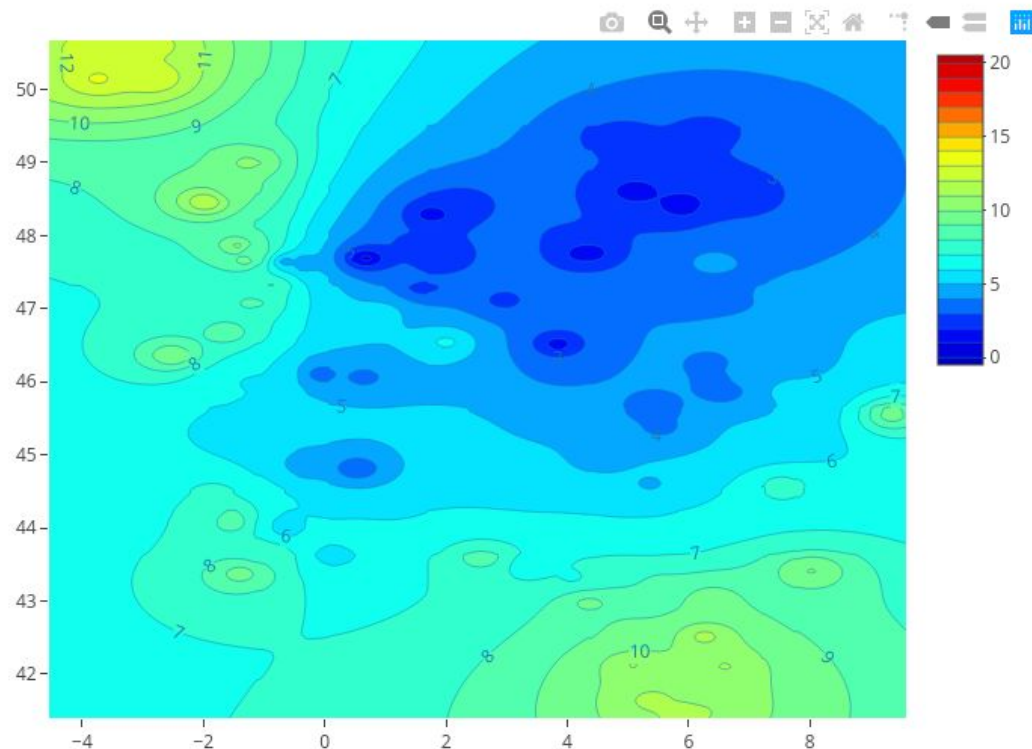
Ta metoda daje podobny rezultat do
poprzednich dwóch odmian krigingu,
jednakże widać, że punkty znajdujące
się poza obszarem wyznaczonym
przez punkty pomiarowe zostają
wyznaczone jako maksima.



Wyniki pomiarów c.d.

Metoda inverse distance weighting dla
pomiarów z Francji

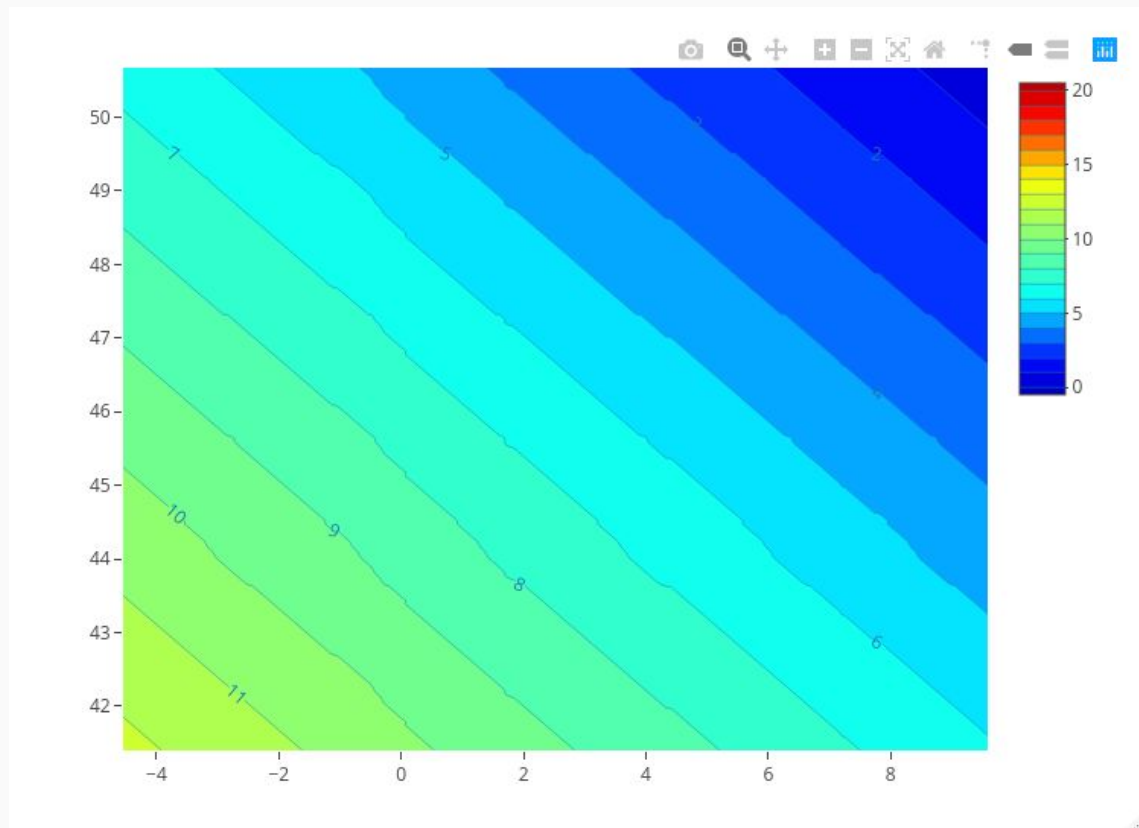
Metoda idw daje podobny efekt jak
metoda kriging, jeżeli skupimy się na
ogólnym trendzie wykresu. Z drugiej
strony w jej efekcie powstaje dużo
więcej ekstremów lokalnych.



Wyniki pomiarów c.d.

Metoda regresji liniowej dla pomiarów z Francji.

Jak widać regresja liniowa jest dość naiwnym algorytmem, który skutkuje narysowaniem nachylonej płaszczyzny. Wbrew pozorom takie przybliżenie mimo swojej naiwności daje całkiem dobre wyniki.

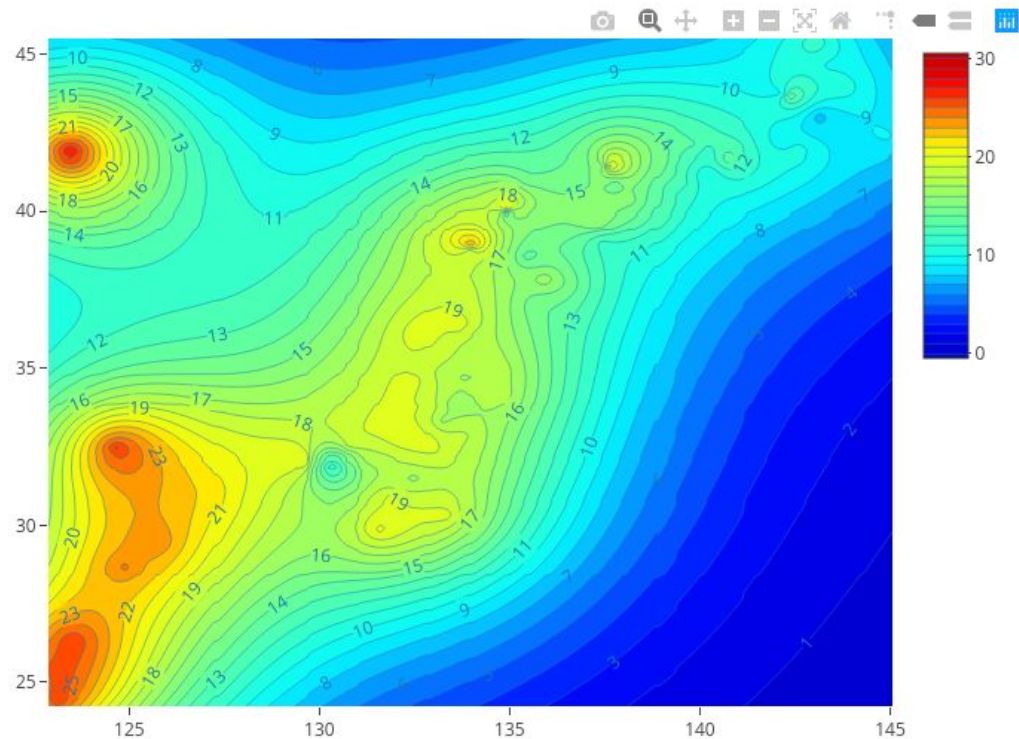


Studium przypadku

W ramach eksperymentu przeprowadzono badanie dla pomiarów z Japonii.

Łatwo zauważyć maksimum występujące w lewym górnym rogu wykresu. Po porównaniu współrzędnych geograficznych z mapą świata można dojść do wniosku, że występuje ono na terenie Korei. Podobnie było w przypadku wszystkich algorytmów.

Wniosek: Interpolacja daje niezłe wyniki, natomiast ekstrapolacja dużo gorsze.



Wnioski

- Odmiany krigingu dają lepsze wyniki niż pozostałe dwie metody, jednakże czas obliczeń wzrasta w dużym tempie wraz z liczbą generowanych punktów (konieczność policzenia kowariancji), natomiast IDW oraz regresja liniowa prowadzą do odpowiedzi dużo szybciej.
- Wszystkie metody skutkują dobrymi przybliżeniami faktycznych wartości dla danych przestrzennych, dotyczących temperatury.

Wnioski c.d.

- Każda z metod jest łatwa w użyciu przy użyciu funkcji dostępnych w pakietach języka R.
- Bez problemu można znaleźć różne implementacje oraz przykłady, wraz z opisami, zastosowania algorytmów do analizy danych.

Bibliografia

- <https://cran.r-project.org/web/packages/KRIG/>
- <https://rpubs.com/hungle510/202761>
- <http://r-statistics.co/Linear-Regression.html>
- <https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm>
- <https://www.rdocumentation.org/>