

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

پروژه درس آمار فضایی
بررسی مجموعه داده پارانا

مدرس: دکتر محسن محمدزاده

دانشجو: علیرضا کبورانی

بهار ۱۴۰۱

۱.مقدمه.....	۲
۲. معرفی مجموعه داده مورد مطالعه.....	۳
۳. بررسی نرمال بودن داده‌ها.....	۴
۴. مانایی در میانگین.....	۶
۵. زدودن روند داده‌ها.....	۷
۶. بررسی فضایی بودن داده‌ها.....	۹
۷. همسان‌گردی.....	۹
۸. برازش تغییرنگار به مجموعه داده Parana.....	۱۰
۹. برآورد استوار تغییرنگار.....	۱۱
۱۰. کریگیدن عادی و عام.....	۱۲
۱۱. تحلیل استوار داده های فضایی.....	۱۴
۱۲. کریگیدن نشانگر.....	۱۹
۱۳. کریگیدن احتمالی.....	۲۱
۱۴. پیشگویی فضایی بیزی.....	۲۲
۱۵. جدول نهایی مقایسه‌ی مقادیر CVMSE.....	۲۴

۱. مقدمه

در اغلب مسائل آماری فرض بر این است که مشاهدات تحت شرایط یکسان و به صورت مستقل از هم جمع‌آوری شده‌اند. فرض استقلال کمک شایانی به تسهیل مبانی نظری می‌کند. اما در عمل ممکن است این فرض ما را از موقعیت دور کرده و موجب از بین رفتن اطلاعات زیادی می‌شود.

موارد زیادی وجود دارد که مشاهدات مستقل نبوده و برحسب موقعیت قرار گرفتن خود در فضای مورد بررسی به یکدیگر وابسته هستند. اگر این وابستگی تابعی از فاصله‌ی بین موقعیت‌های مشاهدات باشد، به گونه‌ای که مشاهدات نزدیک به هم وابسته‌تر و مشاهدات دورتر از هم وابستگی کم‌تری داشته باشند، این‌گونه مشاهدات داده‌های فضایی نامیده می‌شوند.

کاربردهای این دانش در علوم و صنعت‌های مختلفی چون هواشناسی، نقشه‌برداری، معدن و اکتشاف، عمران و شهرسازی و ... دیده شده است.

در این پژوهش سعی شده است تا انواع رویکردهای مبتنی بر حل مسئله‌هایی که برحسب موقعیت فضایی به هم وابسته‌اند، با استفاده از تحلیل فضایی یک مجموعه داده و به کمک نرم‌افزار R بررسی شوند.

۲. معرفی مجموعه داده مورد مطالعه

این گزارش به تحلیل فضایی "داده‌های بارندگی از ایالت پارانا در برزیل" پرداخته است. این مجموعه داده توسط دیگل و ریبریو (۲۰۰۱) در فضای دو بعدی گردآوری و معرفی شده است. داده‌ها به میانگین بارندگی در سال‌های مختلف برای دوره‌ی می تا ژوئن (فصل خشک) اشاره دارد. داده‌ها از ۱۴۳ ایستگاه باران‌سنجی در سراسر ایالت پارانا در برزیل جمع‌آوری شده‌اند.

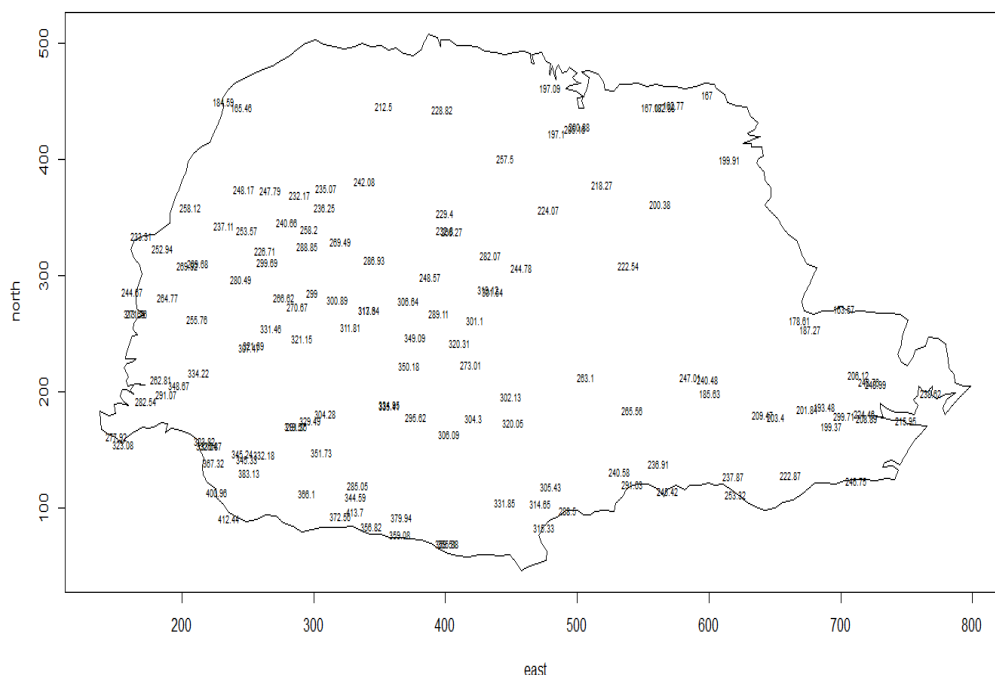
در این مجموعه داده ستون X طول جغرافیایی، ستون Y عرض جغرافیایی و ستون Z متوسط بارش سالانه‌ی باران در ایالت پارانا در برزیل است.

به عنوان نمونه ۶ مشاهده‌ی اول این مجموعه داده در ادامه آمده است :

```
> df = data.frame(x = parana$coords[,1], y = parana$coords[,2], z = parana$data)
> head(df)
```

	x	y	z
1	402.9529	164.5284	306.09
2	501.7049	428.7710	200.88
3	556.3262	445.2706	167.07
4	573.4043	447.0418	162.77
5	702.4228	272.2959	163.57
6	668.5442	261.6707	178.61

نمودار شکل زیر نقشه‌ی هوایی ایالت پارانای برزیل، به همراه مشاهدات ایستگاه‌های هواشناسی است :



۳. بررسی نرمال بودن داده‌ها

برای داشتن درک بهتری از داده‌ها ابتدا لازم است توزیع داده‌ها مشخص شود. برای بررسی توزیع داده‌ها، ابتدا از ابزار شهودی کمک گرفته می‌شود.

۱.۳. نمودار شاخه و برگ

نمودار شاخه و برگ یک روش بررسی توزیع داده‌ها به طور شهودی است که این نمودار برای مجموعه داده‌ها به صورت زیر است :

The decimal point is 1 digit(s) to the right of the |

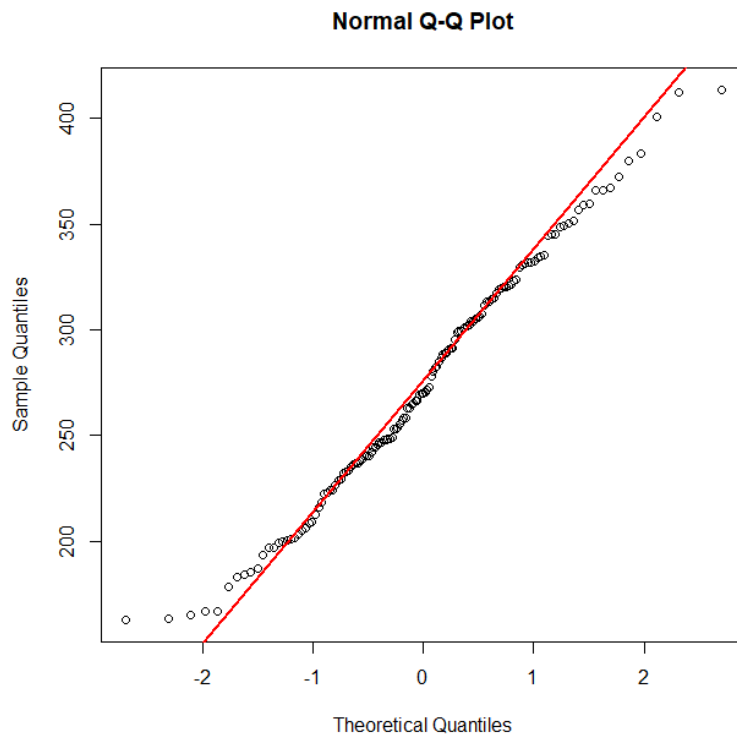
16		345779
18		35673779
20		001235699368
22		3344799233567789
24		011255577889993346888
26		3356679001238
28		0235799911269
30		0011224456772445579
32		0001234911222455
34		555990279
36		06673
38		03
40		124

از آن جایی که این نمودار تقریباً متقارن است، پس به طور شهودی می‌توان دریافت که توزیع داده‌ها متقارن است.

۲-۳. نمودار Q-Q

نمودار چندک-چندک نیز یک روش برای بررسی توزیع داده‌ها است. این نمودار نشان می‌دهد که چگونه یک توزیع نظری می‌تواند به خوبی داده‌های تجربی را الگوبندی کند. در واقع مقادیر مرتب شده‌ی یک نمودار را با چارک‌های یک توزیع خاص مقایسه می‌کند. اگر این دو توزیع با هم هماهنگ باشند، در آن صورت نقاط روی نمودار یک الگوی خطی را تشکیل می‌دهند.

نمودار Q-Q مجموعه داده‌ها رسم شده است. با توجه به این نمودار می‌توان ادعا کرد که به طور شهودی داده‌ها از یک میدان تصادفی گاوسی آمده‌اند.

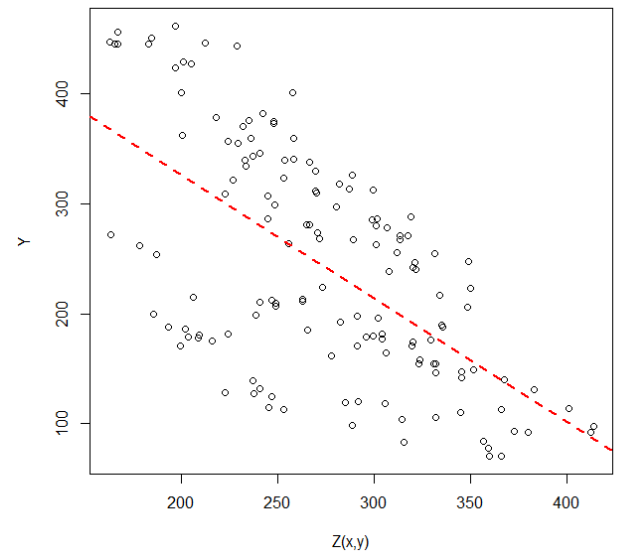
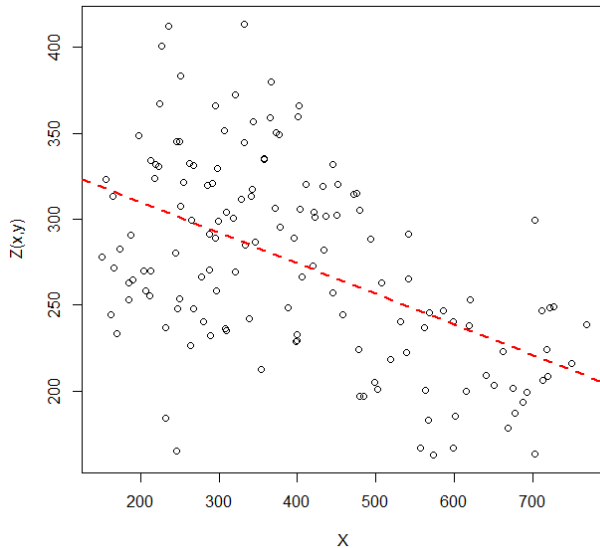


به منظور بررسی دقیق تر توزیع داده ها، آزمون شاپیرو-ویلک روی مجموعه داده انجام شده است که نتیجه ی آزمون بیانگر آن است که توزیع متغیر میزان بارش با اطمینان ۹۵ درصد نرمال است.

```
> shapiro.test(df$z) #Normal ast  
  
Shapiro-Wilk normality test  
  
data:  df$z  
W = 0.98644, p-value = 0.1738
```

۴. مانایی در میانگین

برای بررسی شهودی اینکه آیا داده‌ها مانای در میانگین هستند یا خیر، روند داده‌ها را در راستای مولفه‌های اندیس موقعیت رسم شده است :



در نمودار مشاهدات در برابر X ، داده‌ها در راستای شرق به غرب و در نمودار مشاهدات در برابر Y ، در راستای شمال به جنوب، دارای روند هستند. پس داده‌ها مانای در میانگین نیستند. بنابراین باید روند در راستای شرق به غرب و روند در راستای شمال به جنوب از داده‌ها حذف شود. خطوط نقطه‌چین در نمودارها نیز صرفاً برای مشاهده‌ی بهتر روند در داده‌ها رسم شده‌اند.

روش دیگر بررسی مانایی در میانگین استفاده از مدل رگرسیونی است. چنانچه برخی از ضرایب معنی‌دار باشند، داده‌ها مانای در میانگین نیستند. ۳ مدل متفاوت به صورت زیر به روند داده‌ها برازش شده‌اند :

$$\text{مدل ۱)} \mu(s) = \beta_0 + \beta_1 X + \beta_2 Y$$

$$\text{مدل ۲)} \mu(s) = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 X^2 + \beta_4 Y^2$$

$$\text{مدل ۳)} \mu(s) = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 X^2 + \beta_4 Y^2 + \beta_5 XY$$

برازش یک مدل، (به عنوان مثال مدل ۱) به روند داده‌ها منجر به نتایج زیر شد :

```
> summary(model1)

Call:
lm(formula = df$z ~ df$x + df$y)

Residuals:
    Min       1Q   Median       3Q      Max
-74.926 -18.078   0.293  23.435  72.749

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  441.63635     9.70838   45.49  <2e-16 ***
df$x         -0.19599     0.01620  -12.10  <2e-16 ***
df$y         -0.36618     0.02566  -14.27  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.73 on 140 degrees of freedom
Multiple R-squared:  0.6985,    Adjusted R-squared:  0.6941
F-statistic: 162.1 on 2 and 140 DF,  p-value: < 2.2e-16
```

مقدار p-value نشان می‌دهد که فرضیه صفر مبنی بر صفر بودن ضرایب رگرسیونی برای X و Y رد می‌شود. بنابراین داده‌ها در هر دو جهت دارای روند هستند و مانای در میانگین نیستند.

با بررسی مقدار ضریب تعیین تعدیل شده برای سه مدل یاد شده، می‌توان دریافت که مدل سوم برای ادامه‌ی تحلیل مدل مناسب‌تری است. زیرا مقدار ضریب تعیین تعدیل شده برای آن بیشتر شده است. بنابراین برای حذف روند داده‌ها از مدل سوم استفاده می‌شود.

```
> summary(model1)[9]
$adj.r.squared
[1] 0.6941491
```

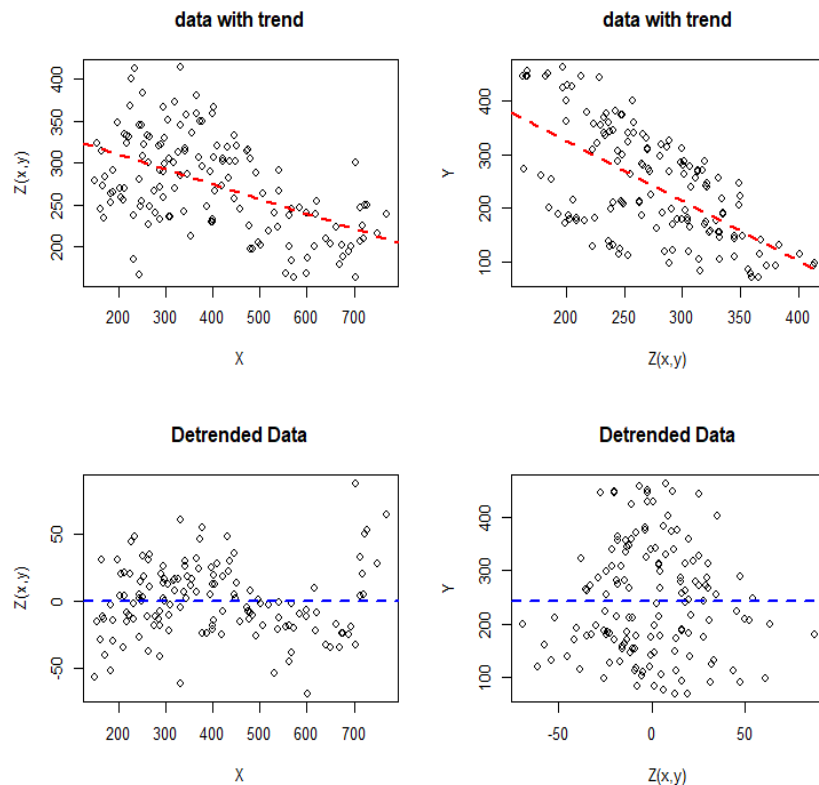
```
> summary(model2)[9]
$adj.r.squared
[1] 0.7600931
```

```
> summary(model3)[9]
$adj.r.squared
[1] 0.7794682
```

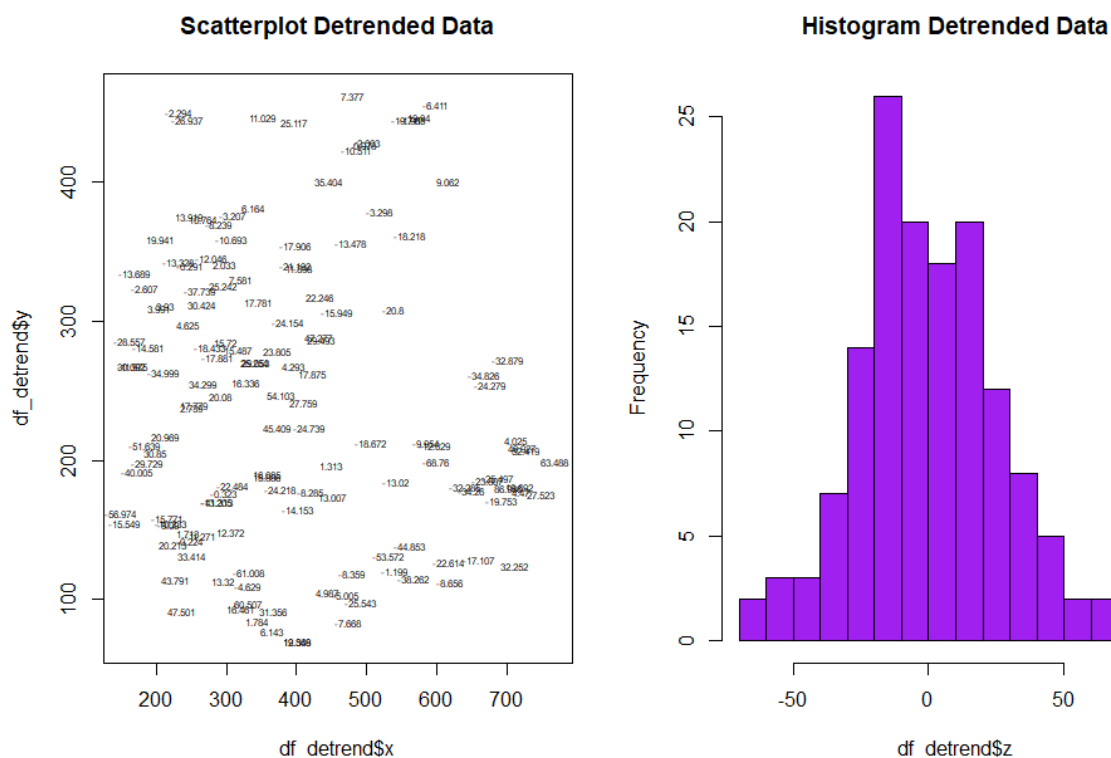
۵. زدودن روند داده‌ها

در این مرحله با کمک مدل ۳ روند داده‌ها را حذف می‌کنیم.

نمودار داده‌های روند زدوده شده و نمودار داده‌های با روند به منظور مقایسه‌ی بهتر در کنار یکدیگر به صورت زیر آورده شده‌اند :



همان طور که در نمودارها مشخص است روند داده‌ها حذف شده است (خطوط نقطه‌چین درواقع مدل رگرسیون خطی ساده بوده و نقطه چین‌های آبی در نمودارها نشان می‌دهد که داده‌ها روند زدوده شده‌اند). نمودار پراکندگی و نمودار هیستوگرام داده‌های روند زدوده به صورت زیر است :



۶. بررسی فضایی بودن داده‌ها

برای بررسی اینکه آیا داده‌ها همبسته فضایی هستند یا خیر از آزمون موران استفاده می‌شود. نتیجه آزمون موران برای مجموعه داده‌ی مورد مطالعه به صورت زیر است :

```
> Moran.I(df_detrend$z, weight = df_dist_inv)
$observed
[1] 0.06818084

$expected
[1] -0.007042254

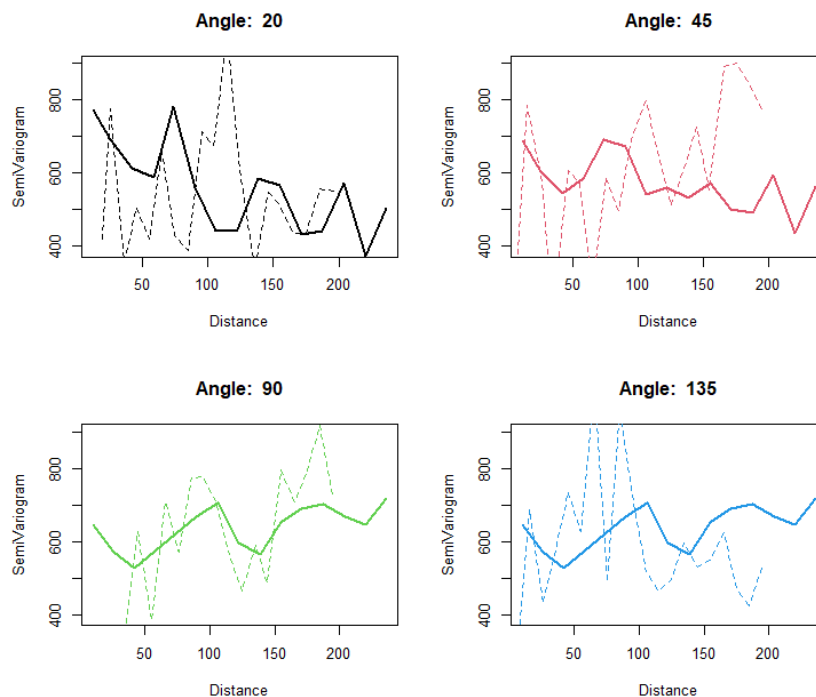
$sd
[1] 0.0186604

$p.value
[1] 5.550157e-05
```

بر اساس این نتایج و پی-مقدار آزمون می‌توان فرضیه صفر مبنی بر عدم وجود همبستگی فضایی در داده‌ها را رد کرد. بنابراین داده‌ها از نظر فضایی وابسته هستند.

۷. همسان‌گردی

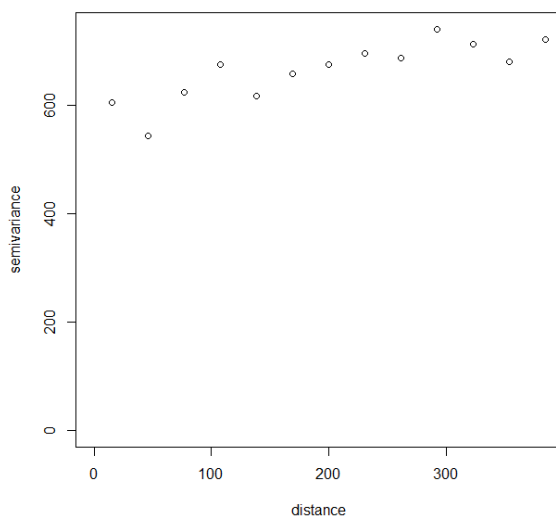
برای بررسی همسان‌گردی لازم است تغییرنگار تجربی در تمام جهات رسم شود، اگر تغییرنگارها در تمام جهات بر هم منطبق باشند، همسان‌گردی برقرار است.



خطوط نقطه‌چین نمودار تغییرنگار نیرومند و دیگری نمودار تغییرنگار تجربی است. به دلیل عدم تفاوت معنی‌دار نمودارهای تغییرنگار تجربی می‌توان گفت به طور شهودی تغییرنگار همسانگرد است.

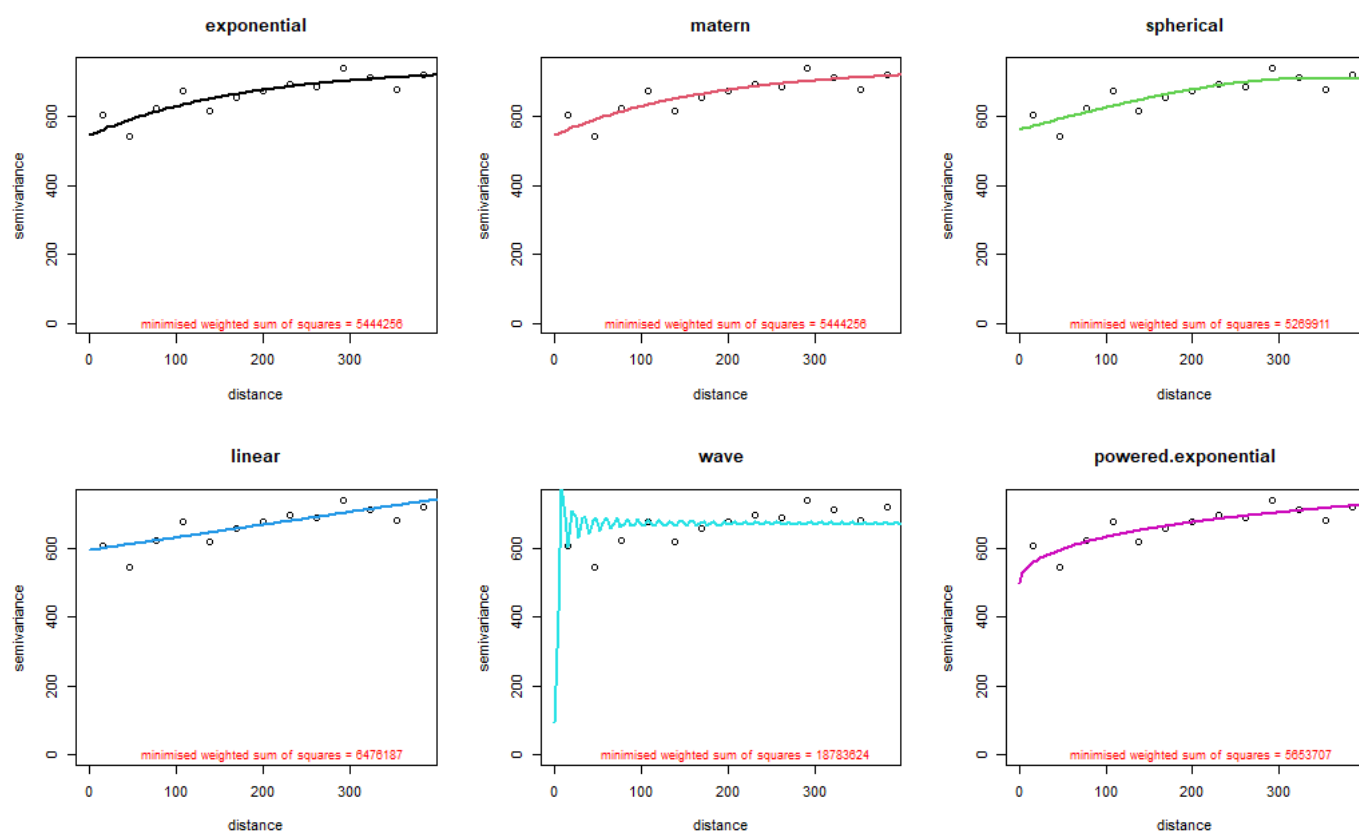
۸. برازش تغییرنگار به مجموعه داده Parana

نمودار تغییرنگار تجربی داده‌های Parana به صورت زیر است :



مدل‌های نیم‌تغییرنگار معتبر همسان‌گرد را به داده‌ها برازش می‌دهیم با روش WLS پارامترهای مدل‌های نیم‌تغییرنگار برازش داده شده به داده‌ها، برآورد شده‌اند و بر اساس معیار SSE یک مدل را به عنوان مدل برتر انتخاب می‌کنیم. مدلی که SSE کم‌تری داشته باشد، مدل برازنده تری است.

پس با توجه به این معیار، مدل نیم‌تغییرنگار کروی، نسبت به سایر مدل‌ها، مدل مناسب‌تری است.



حال با استفاده از روش ماکسیسم درست‌نمایی نیز، ۶ مدل مختلف به داده‌ها برازش داده و مقدار

Bayesian Information Criterion (BIC)

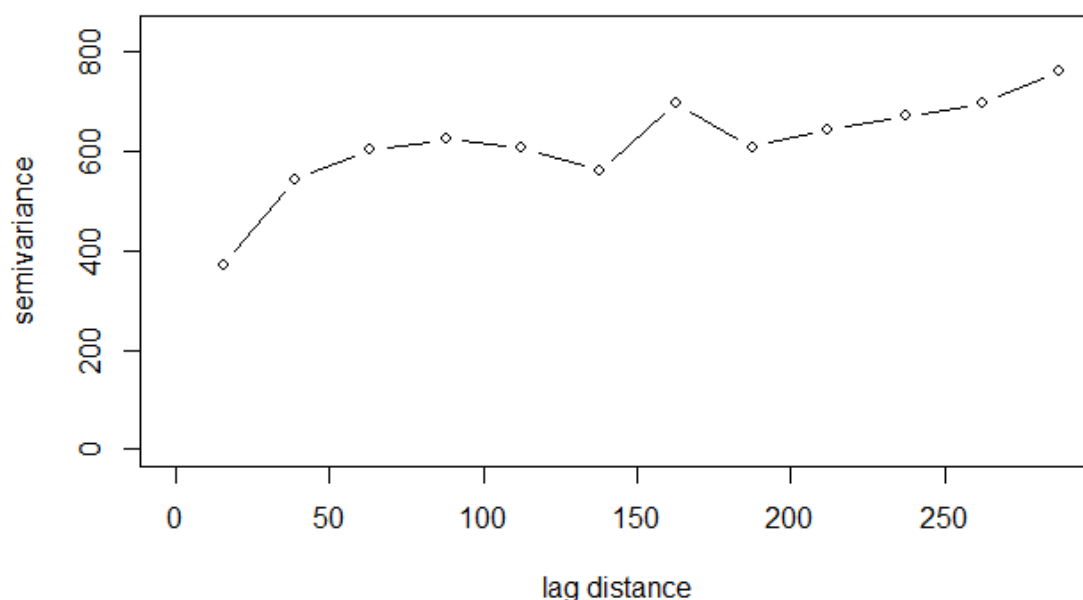
برای هر ۶ مدل به دست آمده را محاسبه می‌کنیم. مدلی که BIC کمتری داشته باشد مدل مناسب‌تری است.

Model	BIC
Exponential	۱۳۴۱.۹۲
Linear	۱۳۴۸.۰۶۶
Spherical	۱۳۴۰.۵۱۵
Wave	۱۳۴۴.۳۵۶
Gaussian	۱۳۴۰.۳۵۱
Powered Exponential	۱۳۴۲.۷۷۲

مقدار BIC برای دو مدل کروی و گاوسی تقریباً با هم برابر شد بنابراین فعلاً هر دو به عنوان مدل‌های برتر انتخاب می‌شوند تا در ادامه‌ی تحلیل بررسی شود که کدام مدل نتایج قابل قبول‌تری را ارائه می‌دهد. (شایان ذکر است که معیار ارزیابی BIC میزان اطلاع از دست رفته توسط مدل را مشخص می‌کند).

۹. برآورد استوار تغییرنگار

برای آن‌که داده‌های نامتعارف و پرت تأثیری حداقلی در برآورد تغییرنگار داشته باشند می‌توان از یک مدل تغییرنگار استوار نیز استفاده کرد، که نمودار آن به شکل زیر است :



۱۰.

کریگیدن عادی و عام

```
> vario_model_wls
  model    psill    range
1  Nug 562.1523  0.0000
2  sph 710.0041 335.9528
> vario_model_mle
  model    psill    range
1  Nug 405.2801  0.0000
2  sph 733.0784 153.3758
```

یک مسأله مهم در آمار فضایی مسأله پیشگویی است که به

معنی برآورد مقدار نامعلوم میدان تصادفی در یک نقطه با

مختصات معلوم، بر اساس نمونه ای از این میدان تصادفی است. به طور معمول با فرض مانایی میدان تصادفی،

معلوم بودن ساختار همبستگی فضایی آن و یک تابع زیان که پرکاربردترین آن درجه دوم است با مینیمم

کردن میانگین توان های دوم خطای پیشگو بدست آورد.

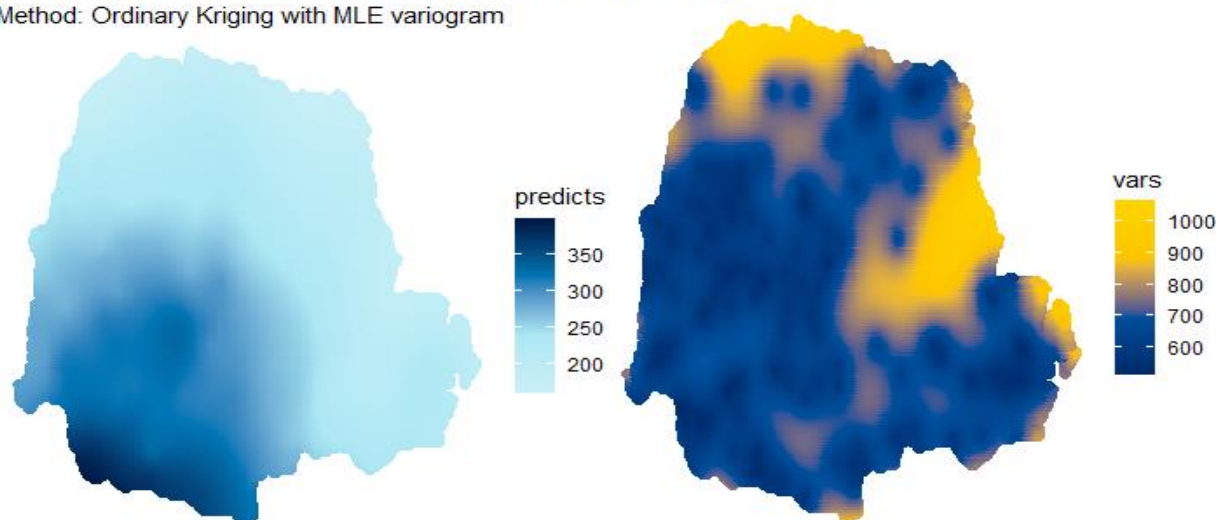
ابتدا یک تغییرنگار برای داده ها با استفاده از روشهای MLE و WLS برآورد می کنیم سپس کریگیدن عادی را با استفاده از آنها بر روی داده های خام و همچنین داده های روند زدوده پیاده سازی می کنیم.

سپس مقدار CVMSE را برای تمام روش های کریگیدن محاسبه و با کمک آن بهترین روش را برای این مجموعه داده انتخاب می کنیم:

CVMSE	روش پیشگویی روی مجموع داده پارانا
۵۷۸.۵۳	کریگیدن عادی داده های خام (WLS)
۵۳۱.۵	کریگیدن عادی داده های روند زدوده (WLS)
۵۳۰.۶۸	کریگیدن عام (WLS)
۶۴۲.۴۳	کریگیدن عادی داده های خام (MLE)
۵۲۱.۲	کریگیدن عادی داده های روند زدوده (MLE)
۵۲۸.۸۷	کریگیدن عام (MLE)

جدول مقایسه ی مقادیر CVMSE

Rainfall prediction for Parana State, Brasil Rainfall prediction var
Method: Ordinary Kriging with MLE variogram

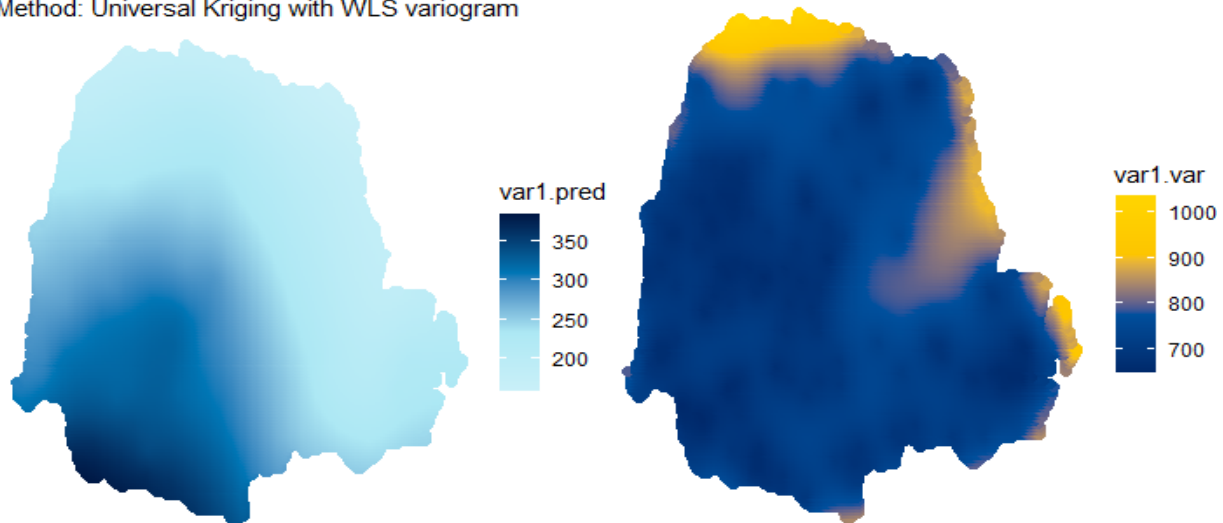


پیشگو انتخاب شده	(۶۷۳, ۱۷۰)	۲۱۴.۴	۶۰۳.۷۴۴	استفاده بهترین
کریگیدن عادی بر	(۶۵۶, ۱۸۴)	۲۰۷.۶	۵۸۱.۸۳	ازبخش قبل یعنی
زدوده شده (MLE)	(۳۵۹, ۴۰۹)	۲۲۸.۸۸	۷۴۸.۷۳	روی داده های روند

برای پنج نقطه دلخواه که مختصات آنها در جدول زیر آمده است پیشگویی و واریانس آن، محاسبه شده است. و دو نمودار مقادیر پیش بینی شده و نمودار واریانس آنها برای دو روش برتر با مقدار CVMSE کمتر است.

نمودار نتایج کریگیدن عام روی دیتا روند زدوده با برآورد تغییرنگار به روش WLS

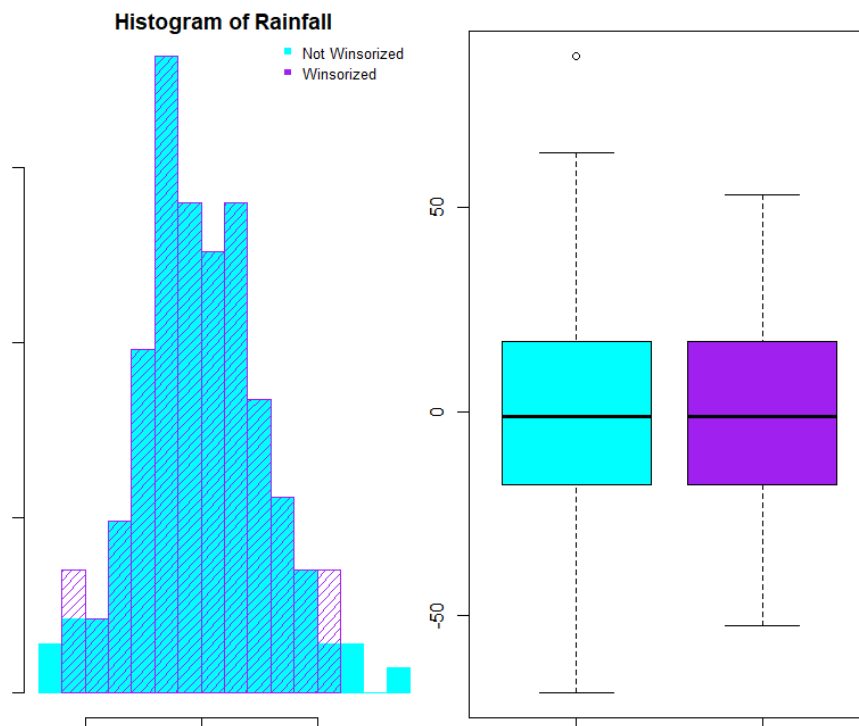
Rainfall prediction for Parana State, Brasil Rainfall prediction var
Method: Universal Kriging with WLS variogram



نمودار نتایج کریگیدن عادی روی دیتا روند زدوده با برآورد تغییرنگار به روش MLE

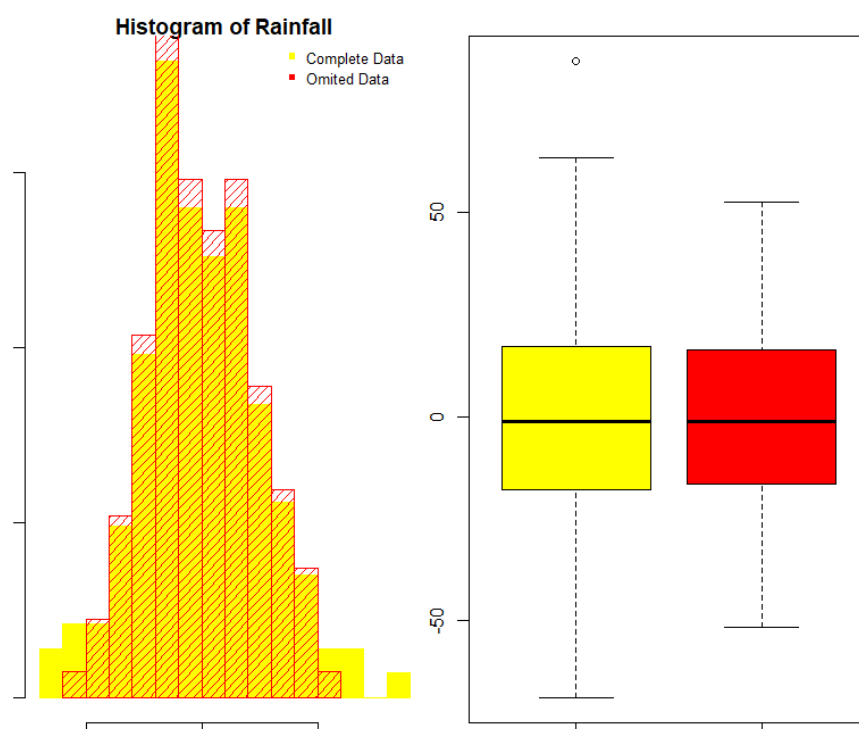
۱۱. تحلیل استوار داده‌های فضایی

یکی از اهداف روش های آمار کاهش میزان حساسیت یک مدل آماری برای برقراری پیش فرض های آن است. یکی از مهم ترین مباحث در این زمینه وجود یا عدم وجود داده های دور افتاده است زیرا وجود داده دور افتاده باعث کاهش دقت یک استنباط آماری می شود. بنابراین برای یک مدل آماری خیلی مهم است که حساسیت زیادی نسبت به این تغییرات نداشته باشد. بنابراین ما به دنبال روش های استوار تحلیل داده ها هستیم. یکی از روش های کاهش حساسیت مدل ها نسبت به داده های دور افتاده، اصلاح آنها است. یکی از روش های اصلاح داده های دور افتاده روش وینزوری کردن است. در این روش ابتدا تغییر نگار برآورد می شود. و سپس ضرایب رگرسیونی، با استفاده از آن برآورد می شد. و با یک مقدار مرجع که یک مرجع خوب میتواند میانه باشد. در شکل زیر نمودار هیستوگرام و جعبه ای داده ها را قبل و بعد از وینزوری کردن نشان می دهد. قبل وینزوری نشان می دهد که یک داده دور افتاده وجود داشته است. که با اصلاح آن هیچ مشاهده دور افتاده ای دیده نمی شود..



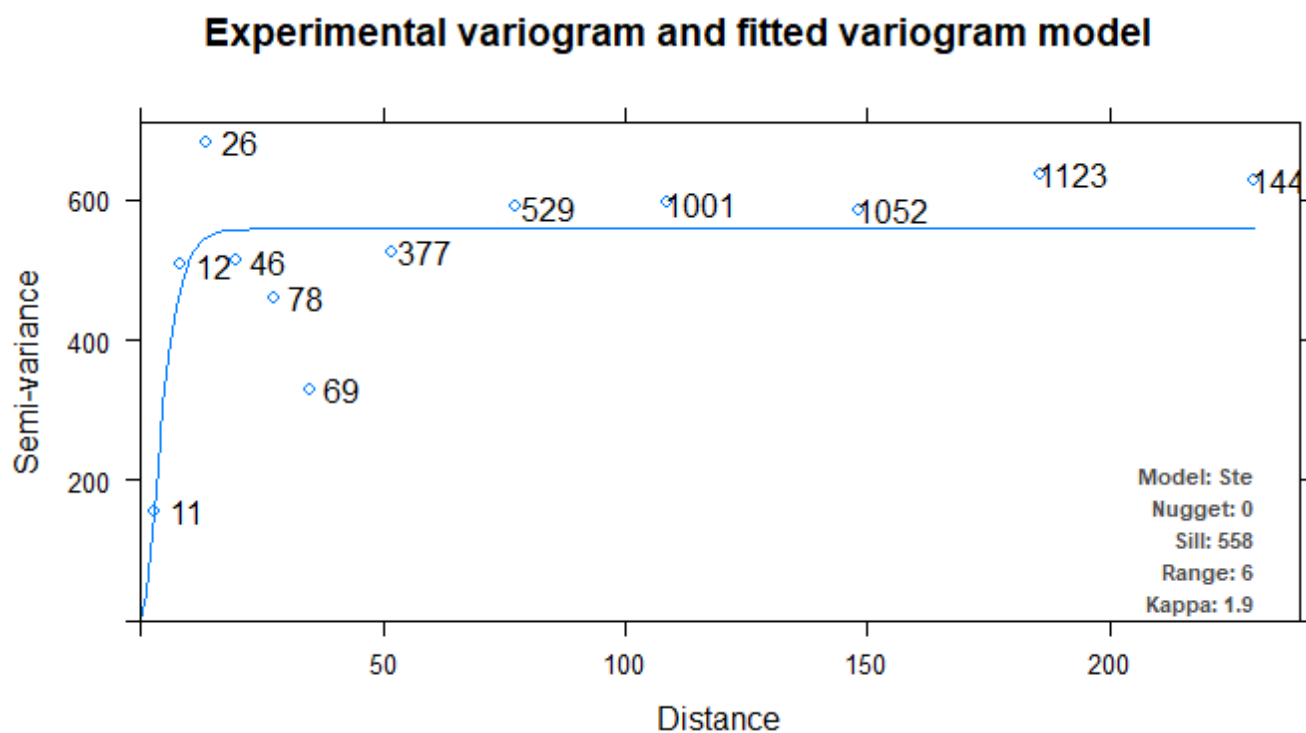
نمودار جعبه ای و هیستوگرام داده ها قبل و بعد از وینزوری کردن

یکی دیگر از روش های اصلاح داده ها پیراستن آنها است. چون با روش وینزوری کردن ممکن است توزیع داده ها عوض شود می توان برای رفع این نقص می توان داده ها را پیراسته کرد. به این صورت که بعد از مرتب کردن آنها درصدی از داده ها را از بالا و پایین آن حذف کنیم در نمودار زیر نمودار های هیستوگرام و جعبه ای آن قبل و بعد از پیراستن آنها آورده شده است



نمودار جعبه ای و هیستوگرام قبل و بعد از پیراستن داده

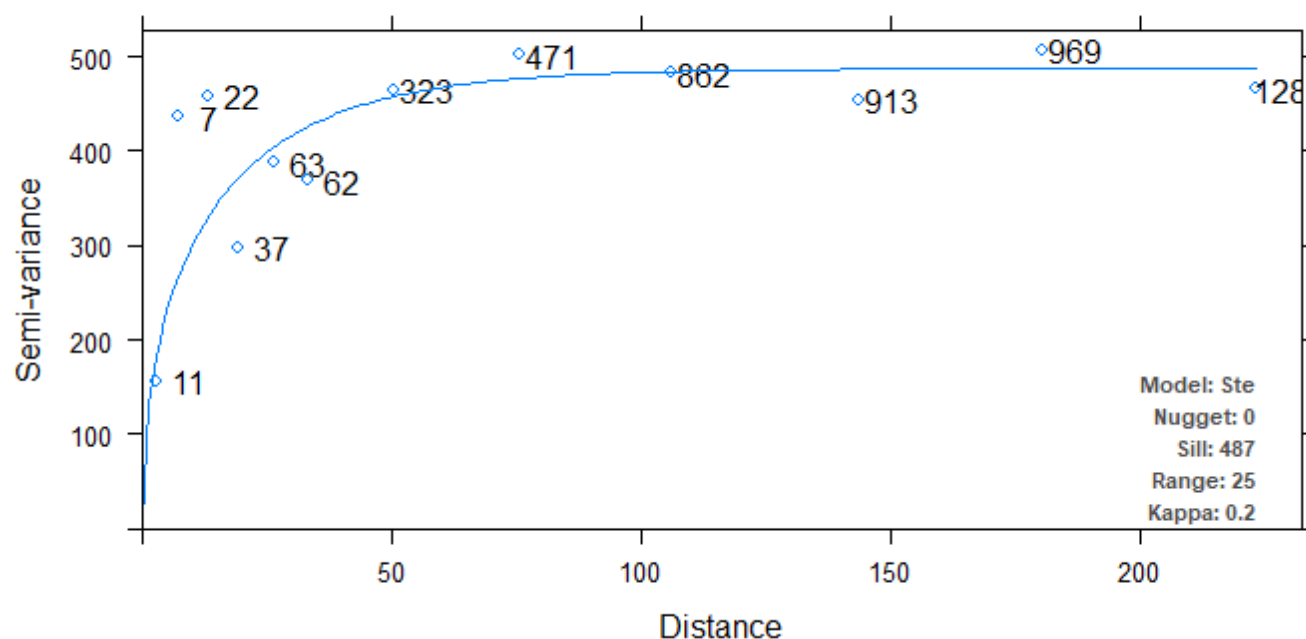
برای هر کدام از داده های وینزوری شده و پیراسته شده یک تغییرنگار و با استفاده از آنها برای پنج نقطه مورد نظر پیشگویی انجام شده است.



نیم تغییرنگار تجربی برای داده های وینزوری شده

نیم تغییرنگار تجربی داده های پیراسته شده

Experimental variogram and fitted variogram model



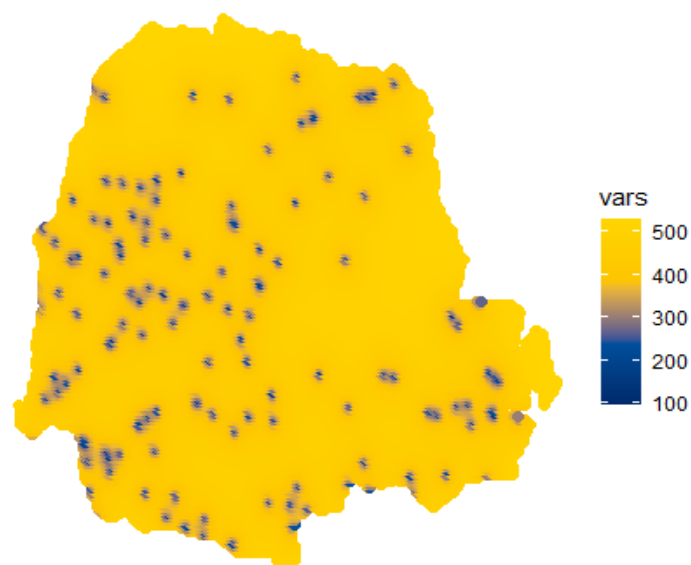
شکل زیر نمودار مقادیر پیش بینی شده و واریانس آنها بر روی داده های پیراسته شده با استفاده از روش کریگیدن عام را نشان می دهد.

Rainfall prediction for Parana State, Brasil

Method: Universal Kriging on Omitted Data

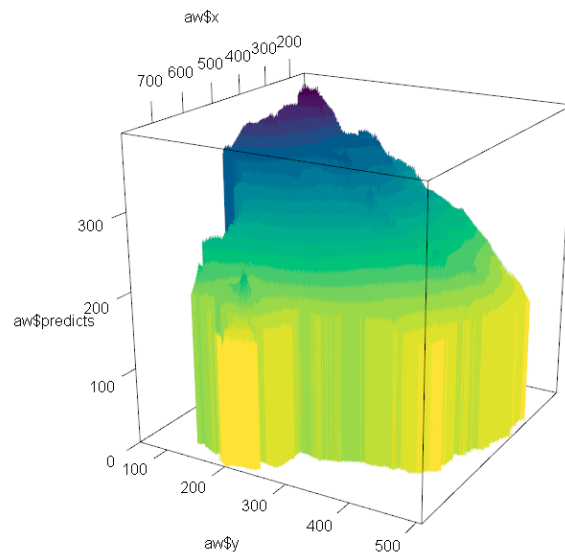


Rainfall prediction var



نمودار مقادیر پیش بینی شده و واریانس آنها بر روی داده های پیراسته

رویه پیشگویی نیز برای روش بالا به صورت زیر است.



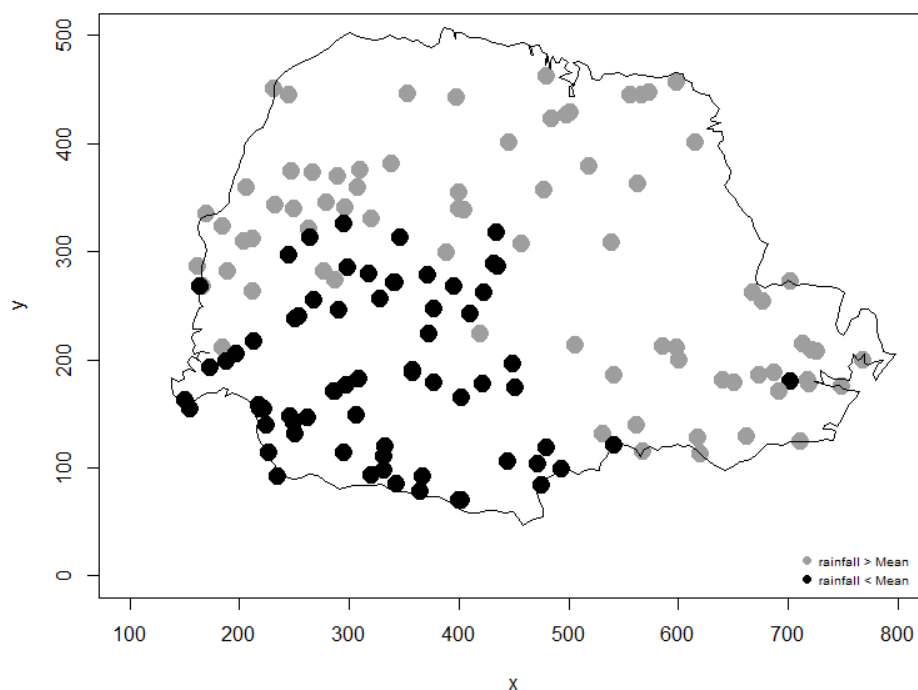
رویه پیشگو

با توجه به جدول زیر کریگیدن بعد از پیراستن داده ها بهترین عملکرد را نسبت به سایر پیشگو ها داشته است و به نظر می رسد داده های وینزوری شده باعث انحراف پیشگویی شده اند.

CVMSE	روش پیشگویی روی مجموع داده پارانا
۵۷۸.۵۳	کریگیدن عادی داده های خام (WLS)
۵۳۱.۵	کریگیدن عادی داده های روند زدوده (WLS)
۵۳۰.۶۸	کریگیدن عام (WLS)
۶۴۲.۴۳	کریگیدن عادی داده های خام (MLE)
۵۲۱.۲	کریگیدن عادی داده های روند زدوده (MLE)
۵۲۸.۸۷	کریگیدن عام (MLE)
۳۸۷.۱۱	کریگیدن عام داده های پیراسته
۵۹۶.۵۳	کریگیدن عام داده های وینزوری شده

۱۲. کریگیدن نشانگر

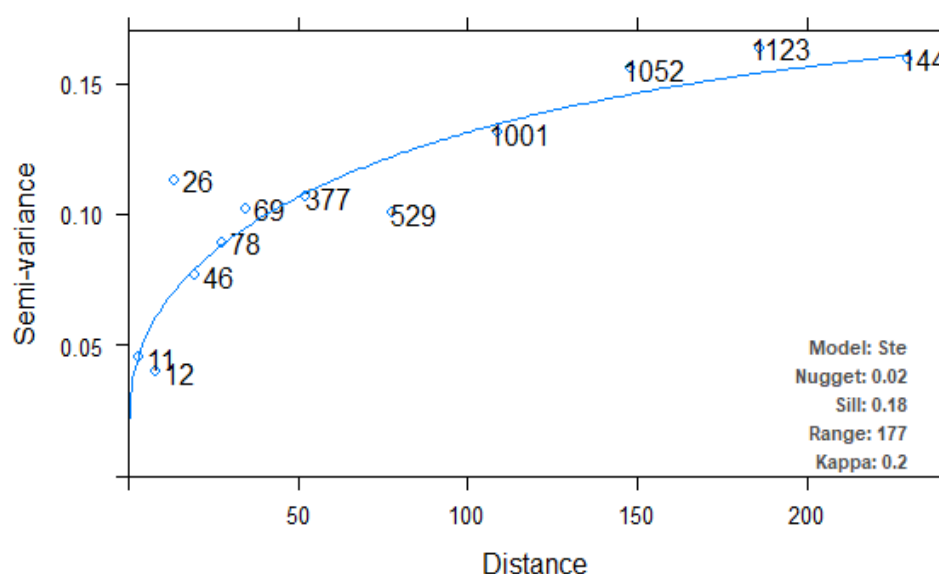
در کریگیدن نشانگر داده های فضایی به یک داده دو دویی تبدیل می شوند و سپس با استفاده از این داده های دو دویی پیشگویی انجام می شود. که از یک مقدار آستانه برای دودویی کردن داده ها استفاده می شود. در اینجا با انتخاب میانگین داده ها به عنوان آستانه دودویی کردن داده ها انجام شده و نتایج آن در نمودار زیر قابل مشاهده است.



داده های دودویی شده روی نقشه

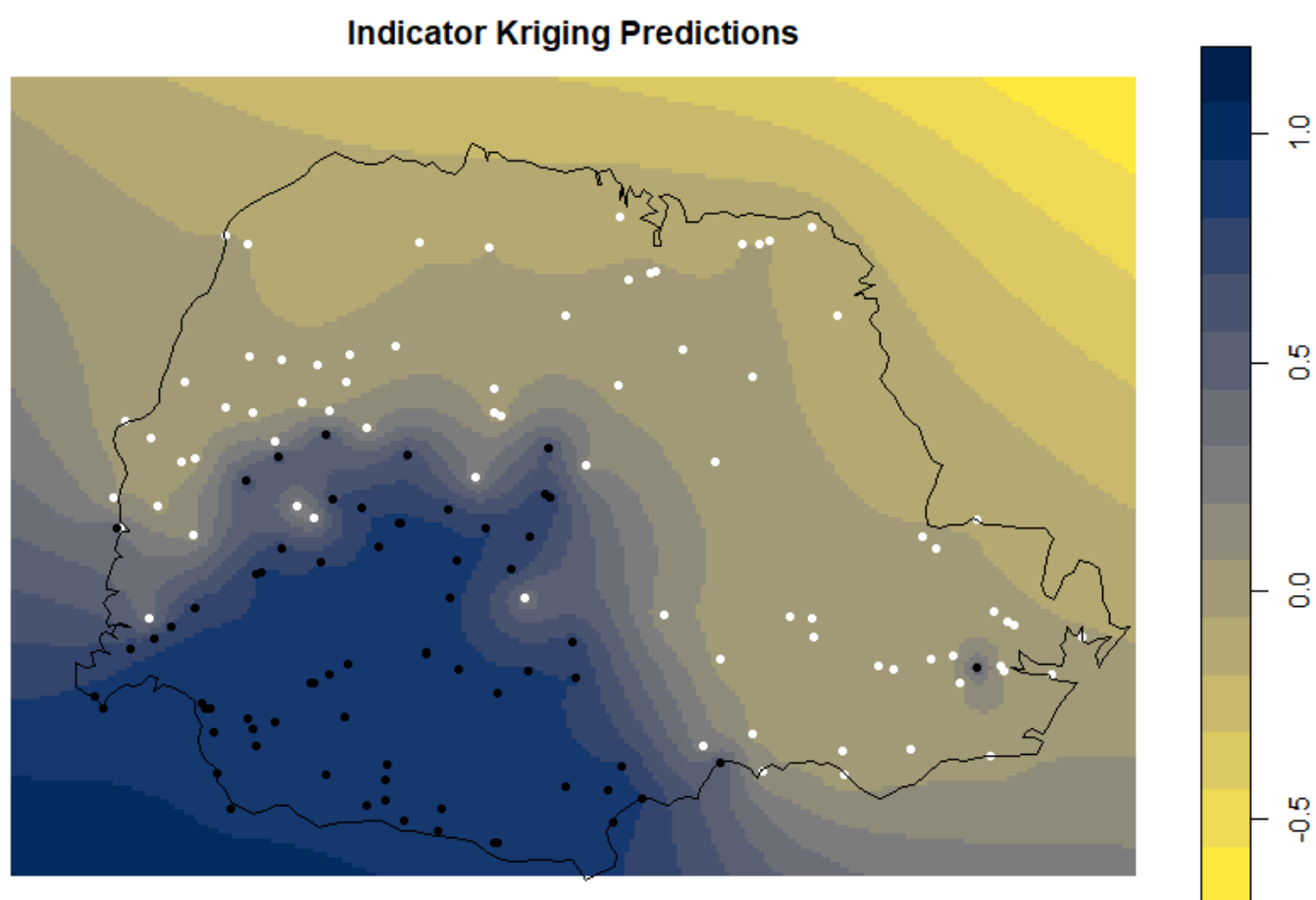
یک تغییرنگار تجربی ماترن به عنوان بهترین تغییرنگار با توجه به معیار RSS بر روی داده های دودویی برآورد شده است.

Experimental variogram and fitted variogram model



نیم تغییرنگار تجربی داده های دودویی

نقاط پیشگویی شده و واریانس آنها توسط کریگیدن نشانگر در شکل و جدول زیر آمده است. CVMSE این روش ۰.۰۸۴ است.



پیش بینی برای پنج نقطه جدید و واریانس آن در جدول زیر آمده است.

```
> ind_method
coordinates  var1.pred  var1.var
1 (448, 304) 0.472954846 0.08453952
2 (679, 88) 0.196688679 0.13312258
3 (673, 170) 0.050397676 0.08967402
4 (656, 184) 0.013251415 0.08060851
5 (359, 409) 0.006361294 0.11163799
```

۱۳. کریگیدن احتمالی

در این روش ابتدا یک توزیع حاشیه ای تعیین می شود این توزیع میتواند به صورت تجربی برآورد شود. و سپس با استفاده از هم کریگیدن یک پیشگو تعریف می شود. که در جدول سمت چپ نتیجه برآورد توزیع که یک توزیع نرمال با میانگین ۲۷۴.۱ و انحراف معیار ۵۷.۱۷ برآورد شده است و در جدول سمت راست نیز آزمون کولموگروف اسمیرنوف انجام شده که توزیع برآورد شده را تایید می کند

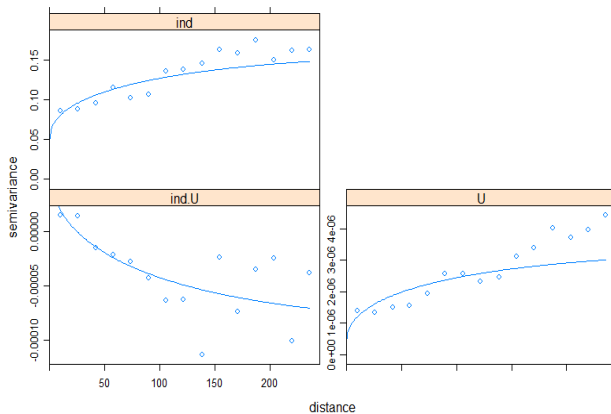
```
> summary(fitn)
Fitting of the distribution ' norm ' by maximum likelihood
Parameters :
      estimate Std. Error
mean 274.41056  4.781427
sd   57.17757  3.380980
Loglikelihood: -781.5093   AIC: 1567.019   BIC: 1572.944
Correlation matrix:
      mean sd
mean   1  0
sd     0  1
```

Asymptotic one-sample Kolmogorov-Smirnov test

data: df\$z

D = ۰,۰۶۳۳, p-value = ۰,۶۱۵۵

alternative hypothesis: two-sided

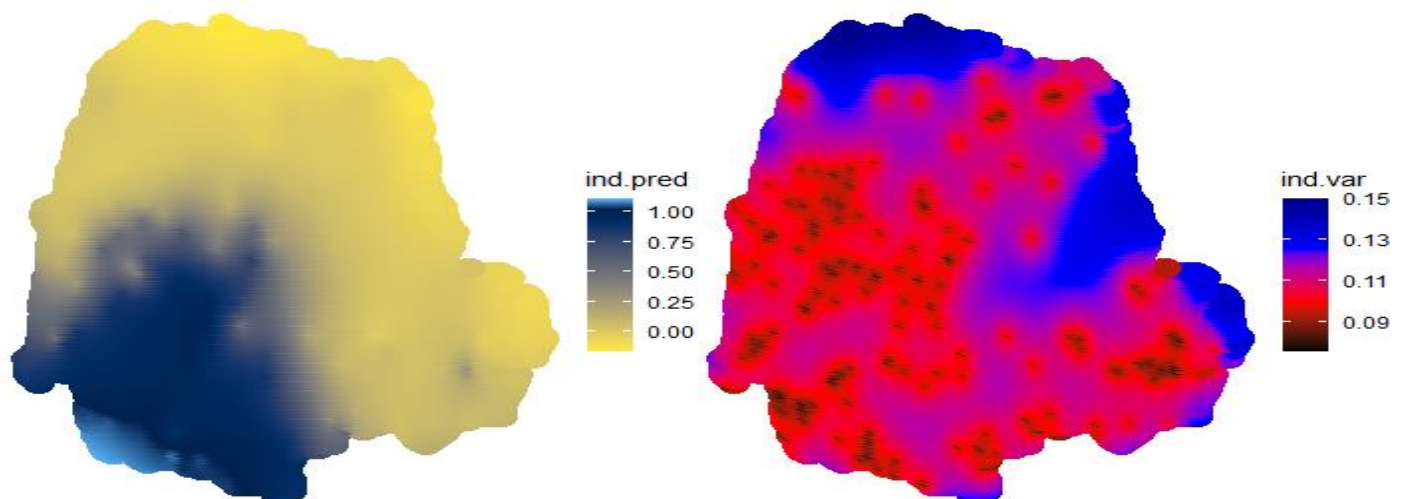


نیم تغییر نگار برزاش شده به داده ها به صورت زیر است.

```
> head(CK_df)
  x  y ind.pred ind.var  U.pred  U.var cov.ind.U
1 456 51.5 0.9683058 0.1279928 0.004463876 2.355784e-06 -3.057124e-05
2 459 51.5 0.9617964 0.1280519 0.004534902 2.355983e-06 -3.046939e-05
3 462 51.5 0.9549554 0.1281031 0.004604154 2.355932e-06 -3.035208e-05
4 465 51.5 0.9477590 0.1281633 0.004671508 2.356187e-06 -3.025523e-05
5 453 54.5 0.9738716 0.1261950 0.004399875 2.307219e-06 -2.812303e-05
6 456 54.5 0.9677327 0.1262176 0.004474983 2.306226e-06 -2.794448e-05
```

جدول بالا پیش بینی نقاط مورد نظر را با کریگیدن احتمالی نشان میدهد. در نهایت CVMSE این روش ۰.۰۹۸ بدست آمد که نسبت به کرگیدن نشانگر بیشتر است و گویای آن است که U اطلاعات مفید بیشتری برای ارتقای پیشگویی نشانگر ارائه نداده است.

نمودار پیشگو و واریانس پیشگو برای کریگیدن احتمالی به صورت زیر است



۱۴. پیشگویی فضایی بیزی

در تمامی روش های قبلی پارامترهای رگرسیونی و سایر پارامتر های موجود ثابت در نظر گرفته شده است. اما گاهی اوقات چنین فرضی ممکن است برقرار نباشد بنابراین نمیتوان از روش های فراوانی گرا برای پیشگویی استفاده کرد زیرا پارامترها نیز ممکن است متغیر تصادفی باشند. بنابراین باید برای آنها نیز یک توزیع در نظر گرفت که به عنوان توزیع پیشین شناخته می شود. سپس با بدست آوردن توزیع پسین به صورت تحلیلی یا تولید نمونه به صورت عددی میتوان پیشگویی را انجام داد. که به روش های بیزی مشهورند. انتخاب توزیع پیشین مناسب برای پارامتر های مدل یکی از مهمترین بخشهای استفاده از تحلیل بیزی است.

در اینجا دو مرتبه پیشگویی بیزی را پیاده سازی می کنیم. یک مرتبه بر روی داده های روند زدوده شده و یک مرتبه بر روی داده های پیراسته شده پیشگویی انجام می شود. ابتدا روی داده های روند زدوده شده پیشگویی بیزی را انجام می دهیم.

برای شروع میبایست ابتدا یک مدل کوواریانس تعریف کنیم.

```
bayes_cov_model = model.control(trend.d = "cte", trend.l = "cte", model = "spherical", lambda = ۱)
```

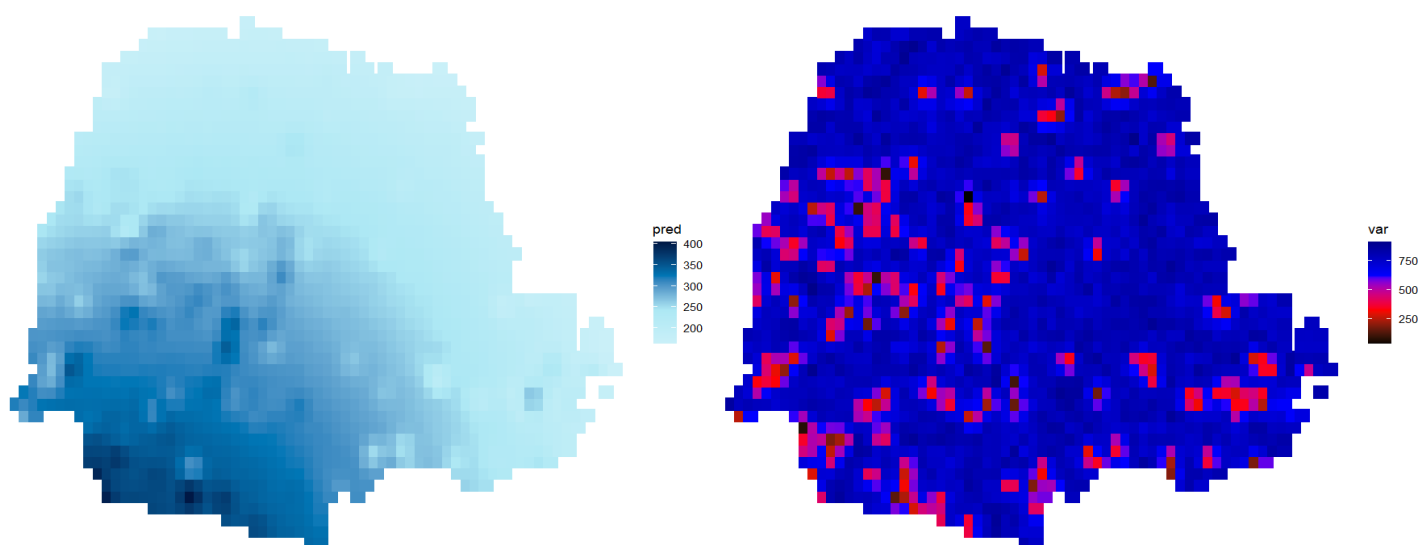
دو پارامتر اول این تابع برای حذف کردن روند به کار می رود که در اینجا چون از داده های روند زدوده استفاده شده است پس مقدار آن را "cte" یا درواقع بدون حذف روند قراردادیم. همچنین پارامتر `lambda` درواقع پارامتر تبدیل باکس کاکس بوده که در اینجا از قبل میدانیم مجموعه داده ما نرمال است پس با قرار دادن `lambda` برابر با ۱ تبدیلی انجام نمی شود.

روند کار که در اینجا پیش گرفتیم به این صورت است که ابتدا نرم افزار به صورت پیش فرض برای هر پارامتر یک توزیع پیشین در نظر میگیرد بعد ازانتخاب پیشین و انجام کریگیدن برای هر پارامتر یک توزیع پسین بدست می آید سپس با کمک توزیع های پسینی بدست آمده، توزیع های پیشین مناسب را برای داده ها پیدا کرده و دوباره پیشگویی بیز را اینبار با پیشینی هایی که پیدا کریم انجام می دهیم.

بنابراین کریگیدن بیز اولیه به هدف تشخیص توزیع پیشینی اجرا می شود. در دو خروجی پایین در سمت چپ پیشینی های پیش فرض نرم افزار و در سمت راست پیشینی هایی که به کمک پسینی های پیشگویی اولیه شناخته شده اند آورده شده است.

```
> bayes_krige$prior$beta$dist
[1] "flat"
> bayes_krige$prior$beta$dist
[1] "flat"
> bayes_krige$prior$sigma$dist
[1] "reciprocal"
> bayes_krige$prior$phi$dist
[1] "uniform"
> bayes_krige$prior$tausq.rel$dist
NULL
```

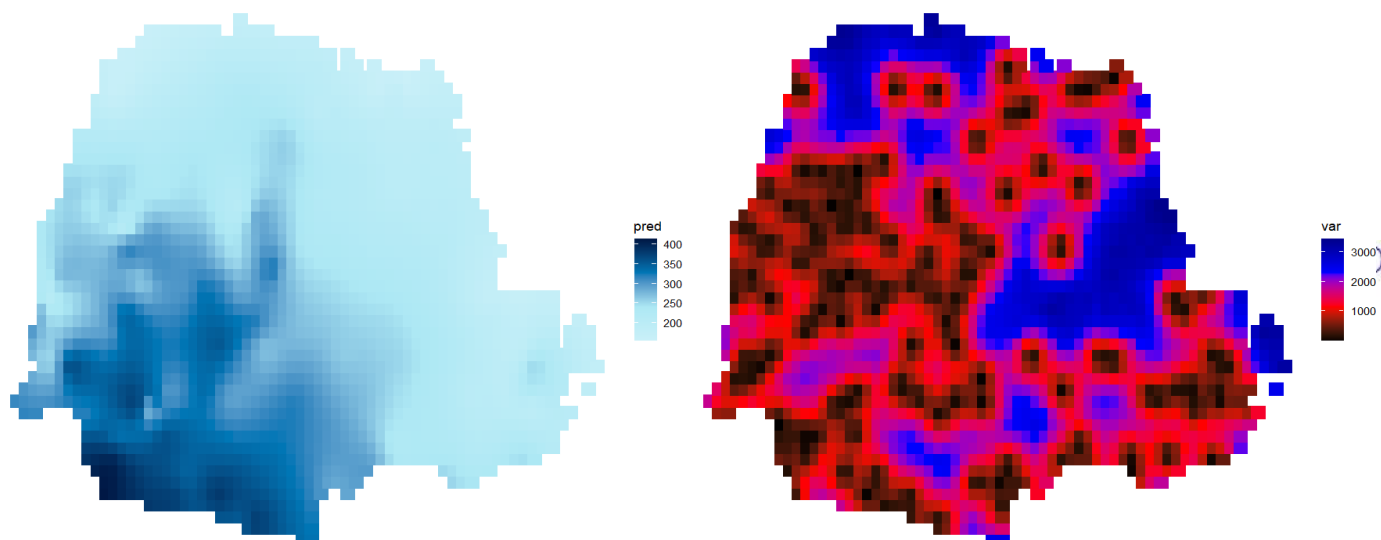
پیش بینی و واریانس کریگیدن بیزی روی داده های روند زدوده شده به صورت زیر است. که CVMSE آن ۶۳۳.۸۴ بدست آمده است.



```
> bayes_krige$prior$beta$dist
[1] "flat"
> bayes_krige$prior$sigma$dist
[1] "reciprocal"
> bayes_krige$prior$phi$dist
[1] "uniform"
> bayes_krige$prior$tausq.rel$dist
NULL
> bayes_prior_model = post2prior(bayes_krige)
> bayes_prior_model$beta.prior
[1] "normal"
> bayes_prior_model$sigma$dist
[1] "sc.inv.chisq"
> bayes_prior_model$phi.prior
[1] "user"
> bayes_prior_model$tausq.rel.prior
[1] "fixed"
```

روند کار بالا را در ادامه برای داده های پیراسته شده طی می کنیم. پیشینی های اولیه برای داده های پیراسته و پیشینی های بدست آمده از پسینی های اولین مرحله برای این داده ها به شکل رو به رو است.

پیشگویی ها و واریانس های پیشگوی بیزی روی داده های پیراسته در شکل زیر آمده است. معیار CVMSE این روش برابر با ۵۰۲.۵۵ می باشد، که در مقایسه با پیشگوهای قبلی (اگر کریگیدن عام داده های پیراسته را در نظر نگیریم) عملکرد مناسب تری دارد. بنابراین اطلاعات پیشینی باعث ارتقای پیشگویی ها شده اند.



۱۵. جدول نهایی مقایسه‌ی مقادیر CVMSE برای تمام روش‌های اجرا شده:

CVMSE	روش پیشگویی روی مجموع داده پارانا
۳۸۷.۱۱*	کریگیدن عام داده‌های پیراسته
۵۰۲.۵۵	کریگیدن بیز روی داده‌های پیراسته
۵۲۱.۲	کریگیدن عادی داده‌های روند زدوده (MLE)
۵۲۸.۸۷	کریگیدن عام (MLE)
۵۳۰.۶۸	کریگیدن عام (WLS)
۵۳۱.۵	کریگیدن عادی داده‌های روند زدوده (WLS)
۵۷۸.۵۳	کریگیدن عادی داده‌های خام (WLS)
۵۹۶.۵۳	کریگیدن عام داده‌های وینزوری شده
۶۳۳.۸۴	کریگیدن بیز روی داده‌های روند زدوده
۶۴۲.۴۳	کریگیدن عادی داده‌های خام (MLE)

* (لازم به ذکر است درباره‌ی مقدار CVMSE روش پیراسته به دلیل ارتقای خیلی زیاد شبّهاتی وجود دارد که متأسفانه دلیل این ارتقای زیاد برما مشخص نشد)

با توجه به نتایج جدول بالا به نظر می‌رسد داده‌های دورافتاده تاثیر زیادی بر پیشگویی‌های ما داشته اند و روش‌های نیرومند عملکرد بهتری بر روی این مجموعه داشته اند.

کدهای تکمیلی R تمامی خروجی‌ها ضمیمه شده است.