Chrome_&_Zinc_spatial_analysis

May 25, 2017

- 0.1 Seminario :: Analisi statistica spaziale di alcune proprietà dei suoli
- 0.1.1 Laurea magistrale in scienze forestali ed ambientali
- 0.1.2 Corso Siti Contaminati

Giuliano Langella [glangella@unina.it]

0.1.3 Step #2: Esempio applicativo, analisi spaziale nella "Terra dei Fuochi" [Cromo & Zinco]

0.1.4 Premessa

Ordinary kriging (OK) The random field (RF) $\mathbf{Z}(\mathbf{u})$ is assumed to be intrinsic secondorder stationary if the first two moments (i.e. mean m and semivariance $\gamma(\mathbf{h})$) of the twopoint RF increments exist and are invariant under translation and rotation within a bounded area \mathfrak{D} (Goovaerts1997_book, Wackernagel2003_book):

$$m = E\{Z(\mathbf{u})\}\$$

$$\gamma(\mathbf{h}) = \frac{1}{2}E\{[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})]^2\}$$
(1)

with theoretically infinite points locations $u(\mathbf{x}) \in \mathfrak{D}$, and random variables (RV) $Z(\mathbf{u})$ and $Z(\mathbf{u} + \mathbf{h})$ separated by the distance vector $h(\mathbf{x})$, where \mathbf{x} represents the coordinates $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \Re^3$.

0.1.5 Inquadramento

0.1.6 DEFs

```
In [34]: # Gives count, mean, standard deviation, standard error of the mean, and confidence int
         summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,</pre>
                               conf.interval=.95, .drop=TRUE) {
             library(plyr)
             # New version of length which can handle NA's: if na.rm == T, don't count them
             length2 <- function (x, na.rm=FALSE) {</pre>
                 if (na.rm) sum(!is.na(x))
                 else
                            length(x)
             }
             # This does the summary. For each group's data frame, return a vector with
             # N, mean, and sd
             datac <- ddply(data, groupvars, .drop=.drop,</pre>
               .fun = function(xx, col) {
                       = length2(xx[[col]], na.rm=na.rm),
                   mean = mean (xx[[col]], na.rm=na.rm),
                       = sd (xx[[col]], na.rm=na.rm)
                 )
               },
               measurevar
             # Rename the "mean" column
             datac <- rename(datac, c("mean" = measurevar))</pre>
             datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error of the mean
             # Confidence interval multiplier for standard error
             # Calculate t-statistic for confidence interval:
             \# e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1
             ciMult <- qt(conf.interval/2 + .5, datac$N-1)</pre>
             datac$ci <- datac$se * ciMult
             return(datac)
         }
   Requirements:
In []: #_____
        # P A C K A G E S
        require("ggplot2")
        require("gridExtra")
        require("dismo")
        library("gstat")
        require("rgdal")
```

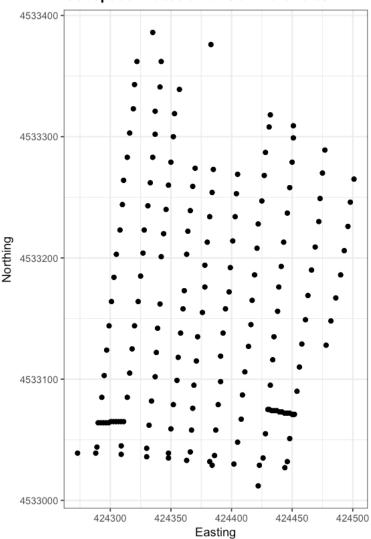
```
library("sp")
library("raster")
library("geoR")
```

Import data (San Giuseppiello Proj):

Geospatial location of soil pits (Biosoil Proj):

```
In [6]: # see
    # http://stackoverflow.com/questions/41787313/how-to-create-a-grid-of-spatial-points
    # https://pakillo.github.io/R-GIS-tutorial/
    it <- getData('GADM', country = 'IT', level = 1)
    campania <- it[it$NAME_1 == "Campania",]
    qplot(Easting, Northing, data=zinc) + coord_fixed(ratio = 1) + theme_bw() + ggtitle("Geo.")</pre>
```

Geospatial Location of Soil Boreholes



Analisi statistica preliminare:

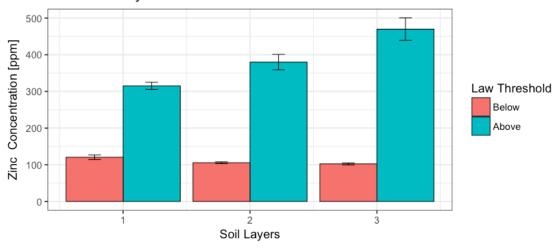
Istogramma delle concentrazioni in funzione delle profondità

In [42]: # S Y N T H E S I S

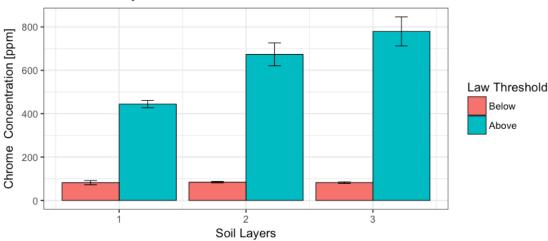
```
tgc_Z <- summarySE( zinc,measurevar="zinc",groupvars=c("Layer","zincthre") )</pre>
tgc_C <- summarySE( chrome, measurevar="chrome", groupvars=c("Layer", "chromethre") )
# bar :: aggregating previous response
plt1 <- ggplot(tgc_Z, aes(x=Layer, y=zinc, fill=factor(zincthre))) +</pre>
    geom_bar(position=position_dodge(), stat="identity",
             colour="black", # Use black outlines,
             size=.3) +
                              # Thinner lines
    geom_errorbar(aes(ymin=zinc-se, ymax=zinc+se),
                  size=.3,
                              # Thinner lines
                  width=.2,
                  position=position_dodge(.9)) +
    xlab("Soil Layers") +
    ylab( paste("Zinc ","Concentration [ppm]")) +
    scale_fill_hue(name="Law Threshold", # Legend label, use darker colors
                   breaks=c(0, 1),
                   labels=c("Below", "Above")) +
    ggtitle("The Effect of Magnitude and Soil Layer\non Uncertainty of Chemical Concent
    #scale_y_continuous(breaks=0:20*4) +
    theme_bw()
plt2 <- ggplot(tgc_C, aes(x=Layer, y=chrome, fill=factor(chromethre))) +</pre>
    geom_bar(position=position_dodge(), stat="identity",
             colour="black", # Use black outlines,
             size=.3) +
                             # Thinner lines
    geom_errorbar(aes(ymin=chrome-se, ymax=chrome+se),
                              # Thinner lines
                  size=.3,
                  width=.2.
                  position=position_dodge(.9)) +
    xlab("Soil Layers") +
    ylab( paste("Chrome ","Concentration [ppm]")) +
    scale_fill_hue(name="Law Threshold", # Legend label, use darker colors
                   breaks=c(0, 1),
                   labels=c("Below", "Above")) +
    ggtitle("The Effect of Magnitude and Soil Layer\non Uncertainty of Chemical Concent
    #scale_y_continuous(breaks=0:20*4) +
```

```
theme_bw()
grid.arrange(plt1,plt2,ncol=1,nrow=2)
```

The Effect of Magnitude and Soil Layer on Uncertainty of Chemical Concentration

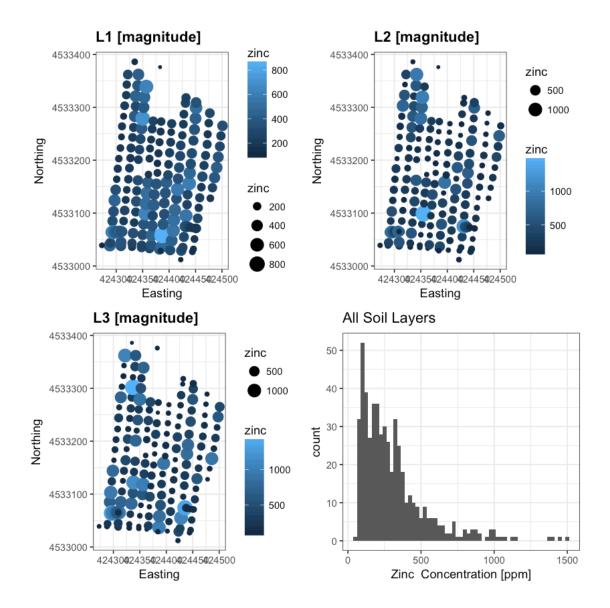


The Effect of Magnitude and Soil Layer on Uncertainty of Chemical Concentration



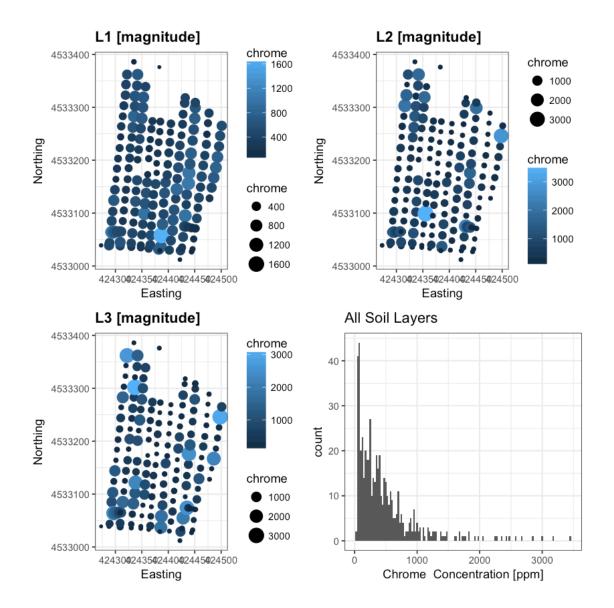
Zinc levels in soils at different depths:

```
In [43]: plt1 <- qplot(Easting, Northing, data=zinc[which(zinc$Layer==1),], colour=zinc, size=zinc plt2 <- qplot(Easting, Northing, data=zinc[which(zinc$Layer==2),], colour=zinc, size=zinc plt3 <- qplot(Easting, Northing, data=zinc[which(zinc$Layer==3),], colour=zinc, size=zinc plt4 <- qplot(zinc, data=zinc, geom="histogram", binwidth=25, main="All Soil Layers") + to grid.arrange(plt1,plt2,plt3,plt4,ncol=2,nrow=2)</pre>
```



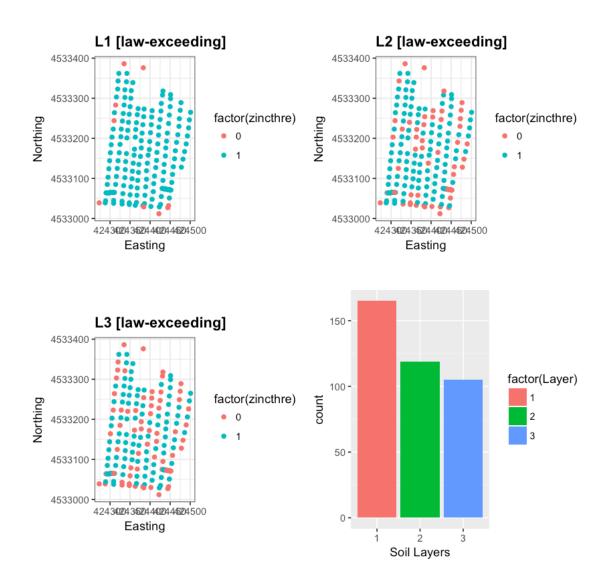
Chrome levels in soils at different depths:

```
In [44]: plt1 <- qplot(Easting, Northing, data=chrome[which(chrome$Layer==1),], colour=chrome, s
    plt2 <- qplot(Easting, Northing, data=chrome[which(chrome$Layer==2),], colour=chrome, s
    plt3 <- qplot(Easting, Northing, data=chrome[which(chrome$Layer==3),], colour=chrome, s
    plt4 <- qplot(chrome, data=chrome, geom="histogram",binwidth=25,main="All Soil Layers")
    grid.arrange(plt1,plt2,plt3,plt4,ncol=2,nrow=2)</pre>
```

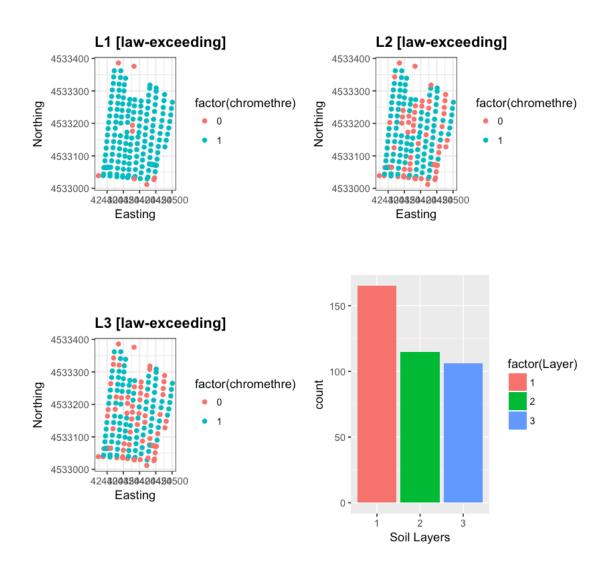


Superamento limiti di legge

Zinc



Chrome



Incertezza di misura al variare del livello di concentrazione del contaminante

Zinc

Chrome

```
In [ ]: ggplot(chrome, aes(x=1:length(chrome$chrome), y=chrome, colour=factor(chromethre))) +
           geom_errorbar(aes(ymin=chrome-std, ymax=chrome+std), width=.01) +
           xlab("Samples") +
           ylab( paste("Chrome ","Concentration [ppm]") ) +
           scale_colour_hue(name="Law Threshold",
                                                             # Legend label, use darker colors
                            breaks=c(0, 1),
                            labels=c("Below", "Above"),
                            1=40) +
                                                       # Use darker colors, lightness=40
           ggtitle("The Effect of Magnitude \n on Chemical Concentration Uncertainty") +
           expand_limits(y=0) +
                                                       # Expand y range
            #scale_y_continuous(breaks=0:20*4) +
                                                      # Set tick every 4
           theme_bw() +
           theme(legend.justification=c(1,0)
                  #, legend.position=c(.5,.8)
                                                       # Position legend in bottom right
```

0.1.7 Grids

Configurazione del contesto geospaziale:

```
In [57]: #_______

#  # G R I D S
#______

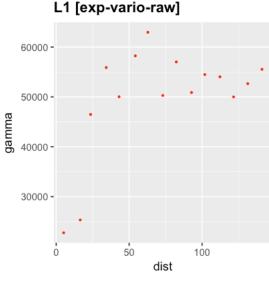
# data
  coordinates(zinc) = ~Easting+Northing
  coordinates(chrome) = ~Easting+Northing
  # grid
  coordinates(grd) = ~Easting+Northing
  gridded(grd) = TRUE
```

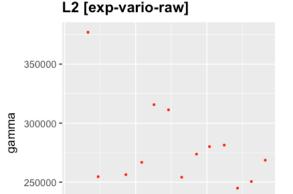
Sommario del contesto geospaziale:

0.1.8 CHROME

Variography | omnidirectional | raw

200000 -



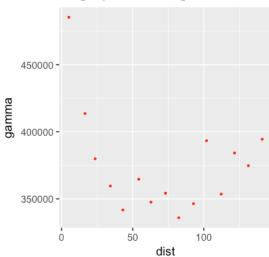


50

dist

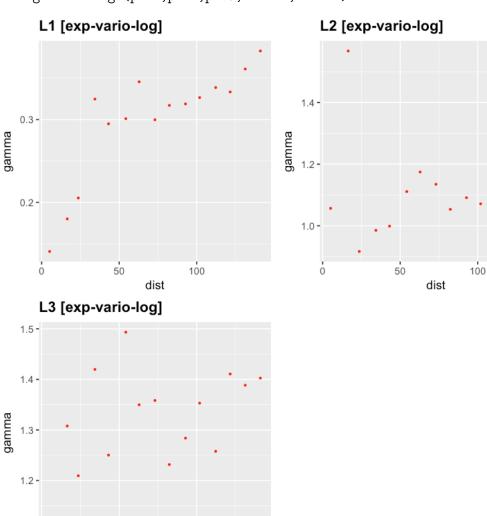
100

L3 [exp-vario-raw]



Variography | omnidirectional | log

```
plt1 <- ggplot(data=vgm.exp.log.L1,aes(x=dist,y=gamma)) + geom_point(color='red',size=0)
plt2 <- ggplot(data=vgm.exp.log.L2,aes(x=dist,y=gamma)) + geom_point(color='red',size=0)
plt3 <- ggplot(data=vgm.exp.log.L3,aes(x=dist,y=gamma)) + geom_point(color='red',size=0)
grid.arrange(plt1,plt2,plt3,ncol=2,nrow=2)</pre>
```



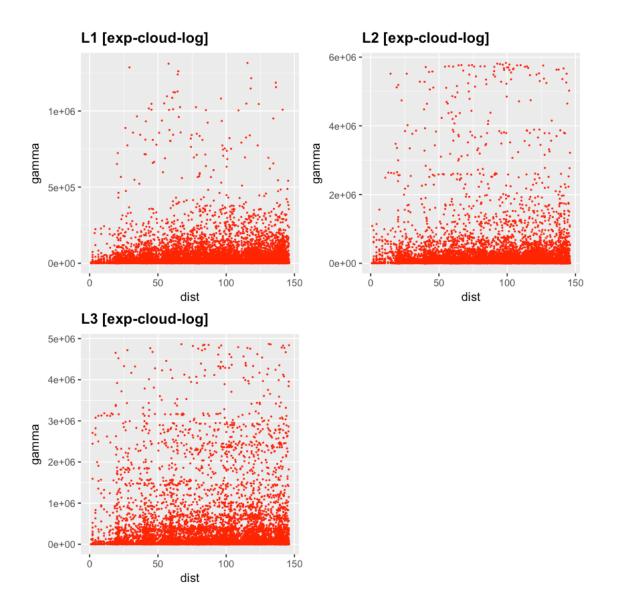
Variography | omnidirectional | log & cloud

50

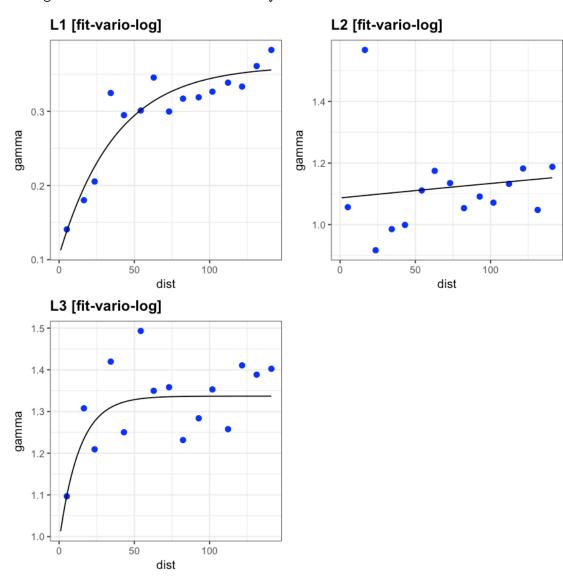
dist

100

0



Warning message in fit.variogram(vgm.exp.log.L2, model = vgm(1, "Exp", 100, 0.1), : No convergence after 200 iterations: try different initial values?



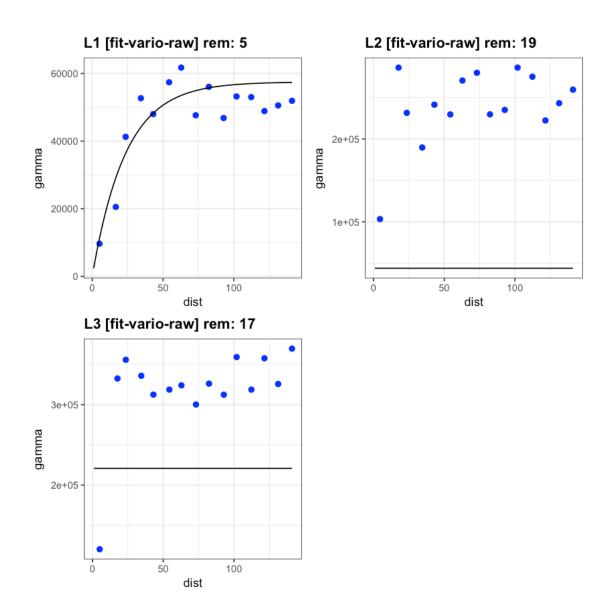
Rimozione delle coppie di punti caratterizzati da alta covarianza a corta distanza:

```
In [66]: # remove couples with high covariance but short distance apart:
              <L1>
                                                                     chrome[which(chrome$Layer==1),]
         chrome L1
                                             <- data.frame(
         coordinates(chrome_L1) = ~Easting+Northing
         vgm.exp.log.L1c_rd = variogram( chrome~1, chrome_L1, cloud=TRUE)
         short_dist
                             <- vgm.exp.log.L1c[which(vgm.exp.log.L1c_rd$dist<5.0),]</pre>
         high_gamma
                             <- short_dist[which(short_dist$gamma>100000),]
         high_gamma
                             <- data.frame(high_gamma)</pre>
         list_remove <- sort(unique( c(high_gamma$left,high_gamma$right) ))</pre>
         rem.L1 <- length(list_remove)</pre>
         if(length(list_remove)==0){
                  chrome_L1_rd <- chrome_L1</pre>
         }else{
                  chrome_L1_rd <- chrome_L1[-list_remove,]</pre>
         }
              <L2>
         chrome_L2
                                                                     chrome[which(chrome$Layer==2),]
                                             <- data.frame(
         coordinates(chrome_L2) = ~Easting+Northing
         vgm.exp.log.L2c_rd = variogram( chrome~1, chrome_L2, cloud=TRUE)
         short_dist
                             <- vgm.exp.log.L2c[which(vgm.exp.log.L2c_rd$dist<5.0),]</pre>
         high_gamma
                             <- short_dist[which(short_dist$gamma>100000),]
                             <- data.frame(high_gamma)
         high_gamma
         list_remove <- sort(unique( c(high_gamma$left,high_gamma$right) ))</pre>
         rem.L2 <- length(list_remove)</pre>
         if(length(list_remove)==0){
                  chrome_L2_rd <- chrome_L2</pre>
         }else{
                  chrome_L2_rd <- chrome_L2[-list_remove,]</pre>
         }
             <L3>
         chrome_L3
                                             <- data.frame(
                                                                     chrome[which(chrome$Layer==3),]
         coordinates(chrome_L3) = ~Easting+Northing
         vgm.exp.log.L3c_rd = variogram( chrome~1, chrome_L3, cloud=TRUE)
         short_dist
                           <- vgm.exp.log.L3c[which(vgm.exp.log.L3c_rd$dist<5.0),]</pre>
         high_gamma
                             <- short_dist[which(short_dist$gamma>100000),]
         high_gamma
                             <- data.frame(high_gamma)</pre>
         list_remove <- sort(unique( c(high_gamma$left,high_gamma$right) ))</pre>
         chrome_L3_rd <- chrome_L3[-list_remove,]</pre>
         rem.L3 <- length(list_remove)</pre>
         if(length(list_remove)==0){
                  chrome_L3_rd <- chrome_L3</pre>
         }else{
                  chrome_L3_rd <- chrome_L3[-list_remove,]</pre>
         }
```

REMOVED [fit-vario-raw]:

```
In [67]: #coordinates(xrf_L1_rd) = ~Easting+Northing
         vgm.exp.raw.L1_rd = variogram( chrome~1, chrome_L1_rd) # vgm( psill, model, range, nugge
                        = max(vgm.exp.raw.L1_rd$gamma) - min(vgm.exp.raw.L1_rd$gamma)
         psill
                        = max(vgm.exp.raw.L1_rd$dist)/6
         range
                        = vgm.exp.raw.L1_rd$gamma[1]/2
         nugget
         vgm.fit.raw.L1_rd = fit.variogram( vgm.exp.raw.L1_rd, model=vgm(psill, 'Exp', range, nugge
         #coordinates(xrf_L2_rd) = ~Easting+Northing
         vgm.exp.raw.L2_rd = variogram( chrome~1, chrome_L2_rd)
                        = max(vgm.exp.raw.L2_rd$gamma) - min(vgm.exp.raw.L2_rd$gamma)
         psill
                        = max(vgm.exp.raw.L2_rd$dist)/6
         range
                         = vgm.exp.raw.L2_rd$gamma[1]/2
         nugget
         vgm.fit.raw.L2_rd = fit.variogram( vgm.exp.raw.L2_rd, model=vgm(psill, 'Exp', range, nugge
         #coordinates(xrf_L3_rd) = ~Easting+Northing
         vgm.exp.raw.L3_rd = variogram( chrome~1, chrome_L3_rd)
                        = max(vgm.exp.raw.L3_rd$gamma) - min(vgm.exp.raw.L3_rd$gamma)
         psill
                       = max(vgm.exp.raw.L3_rd$dist)/6
         range
                         = vgm.exp.raw.L3_rd$gamma[1]
         nugget
         vgm.fit.raw.L3_rd = fit.variogram( vgm.exp.raw.L3_rd, model=vgm(psill, 'Exp', range, nugge
         #plot(vgm.exp_log,vgm.fit_log, main = paste("[OMNI-DIR] Fitted Variogram",ChElstr,"[rau
         # L1
         maxDist <- round(max(vgm.exp.raw.L1_rd$dist))</pre>
         vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.raw.L1_rd$psill[1] + vgm.fit.raw.</pre>
         names(vgm.fit.plot) <- c("dist", "gamma")</pre>
         plt1 <- ggplot(data=vgm.exp.raw.L1_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si</pre>
                                          ggtitle(paste("L1 [fit-vario-raw] rem:",rem.L1)) + them
         maxDist <- round(max(vgm.exp.raw.L2_rd$dist))</pre>
         vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.raw.L2_rd$psill[1] + vgm.fit.raw.</pre>
         names(vgm.fit.plot) <- c("dist", "gamma")</pre>
         plt2 <- ggplot(data=vgm.exp.raw.L2_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si
                                          ggtitle(paste("L2 [fit-vario-raw] rem:",rem.L2)) + them
         maxDist <- round(max(vgm.exp.raw.L3_rd$dist))</pre>
         vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.raw.L3_rd$psill[1] + vgm.fit.raw.</pre>
         names(vgm.fit.plot) <- c("dist", "gamma")</pre>
         plt3 <- ggplot(data=vgm.exp.raw.L3_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si
                                          ggtitle(paste("L3 [fit-vario-raw] rem:",rem.L3)) + them
         grid.arrange(plt1,plt2,plt3,ncol=2,nrow=2)
```

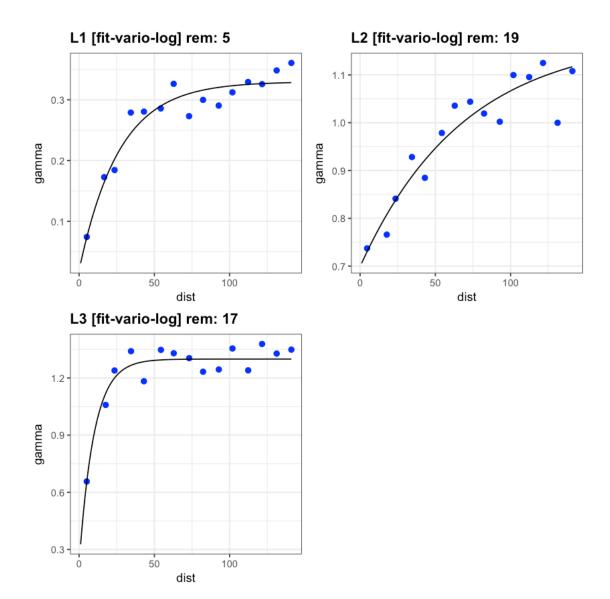
Warning message in fit.variogram(vgm.exp.raw.L1_rd, model = vgm(psill, "Exp", range, : partial sill or nugget fixed at zero valueWarning message in fit.variogram(vgm.exp.raw.L2_rd, mosingular model in variogram fitWarning message in fit.variogram(vgm.exp.raw.L3_rd, model = vgm(psill = vgm) = vgm). No convergence after 200 iterations: try different initial values?Warning message in fit.variogram partial sill or nugget fixed at zero valueWarning message in fit.variogram(object, model, fit.si singular model in variogram fit



REMOVED [fit-vario-log]:

```
In [68]: #coordinates(xrf_L1_rd) = ~Easting+Northing
    vgm.exp.log.L1_rd = variogram( log(chrome)~1, chrome_L1_rd)
    vgm.fit.log.L1_rd = fit.variogram( vgm.exp.log.L1_rd, model=vgm(0.90,'Exp',100,0.1), fit
    #coordinates(xrf_L2_rd) = ~Easting+Northing
    vgm.exp.log.L2_rd = variogram( log(chrome)~1, chrome_L2_rd)
    vgm.fit.log.L2_rd = fit.variogram( vgm.exp.log.L2_rd, model=vgm(0.90,'Exp',100,0.1), fit
    #coordinates(xrf_L3_rd) = ~Easting+Northing
    vgm.exp.log.L3_rd = variogram( log(chrome)~1, chrome_L3_rd)
    vgm.fit.log.L3_rd = fit.variogram( vgm.exp.log.L3_rd, model=vgm(0.90,'Exp',100,0.1), fit
    #plot(vgm.exp_log,vgm.fit_log, main = paste("[OMNI-DIR] Fitted Variogram",ChElstr,"[log#L1]
```

```
maxDist <- round(max(vgm.exp.log.L1_rd$dist))</pre>
vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.log.L1_rd$psill[1] + vgm.fit.log.</pre>
names(vgm.fit.plot) <- c("dist", "gamma")</pre>
plt1 <- ggplot(data=vgm.exp.log.L1_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si</pre>
                                  ggtitle(paste("L1 [fit-vario-log] rem:",rem.L1)) + them
maxDist <- round(max(vgm.exp.log.L2_rd$dist))</pre>
vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.log.L2_rd$psill[1] + vgm.fit.log.</pre>
names(vgm.fit.plot) <- c("dist", "gamma")</pre>
plt2 <- ggplot(data=vgm.exp.log.L2_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si</pre>
                                  ggtitle(paste("L2 [fit-vario-log] rem:",rem.L2)) + them
# L3
maxDist <- round(max(vgm.exp.log.L3_rd$dist))</pre>
vgm.fit.plot <- data.frame(1:maxDist,g_exp <- vgm.fit.log.L3_rd$psill[1] + vgm.fit.log.</pre>
names(vgm.fit.plot) <- c("dist", "gamma")</pre>
plt3 <- ggplot(data=vgm.exp.log.L3_rd,aes(x=dist,y=gamma)) + geom_point(color='blue',si
                                  ggtitle(paste("L3 [fit-vario-log] rem:",rem.L3)) + them
grid.arrange(plt1,plt2,plt3,ncol=2,nrow=2)
```



0.1.9 Predictions

Assign projection to point data

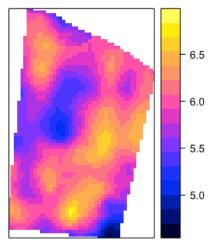
Ordinary Kriging RAW

In []: # ...

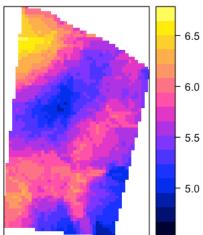
LOG

[using ordinary kriging]
[using ordinary kriging]
[using ordinary kriging]

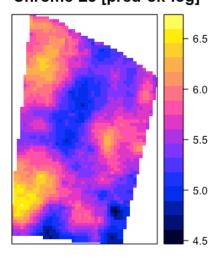
Chrome L1 [pred-ok-log]



Chrome L2 [pred-ok-log]



Chrome L3 [pred-ok-log]

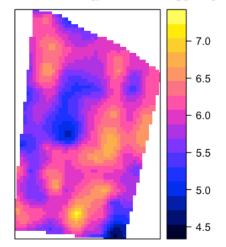


R A W | After Removal

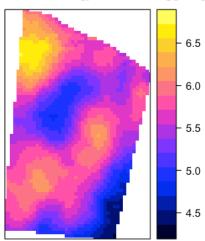
In []: # ...

LOG | After Removal

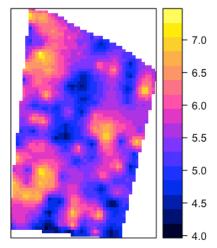
Chrome L1 [pred-ok-log] adj



Chrome L2 [pred-ok-log] adj



Chrome L3 [pred-ok-log] adj



Sintesi:

rimozione delle coppie di punti "outliers", fitting del variogramma sperimentale con un modello di variogramma ammissibile,

sia sulla variabile (=Cromo) tal quale,
 sia sulla variabile (=Cromo) in scale logaritmica,

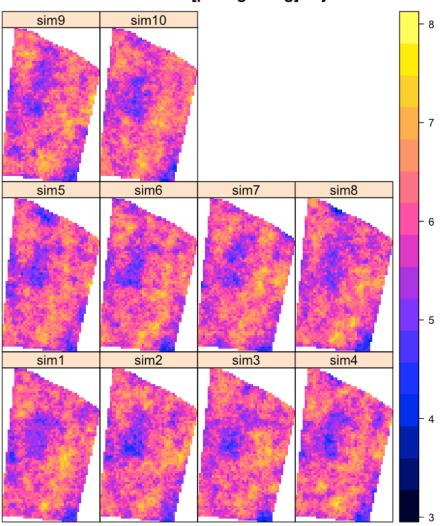
interpolazione mediante kriging ordinario in entrambi i casi, le mappe ottenute "dopo rimozione delle coppie di punti",

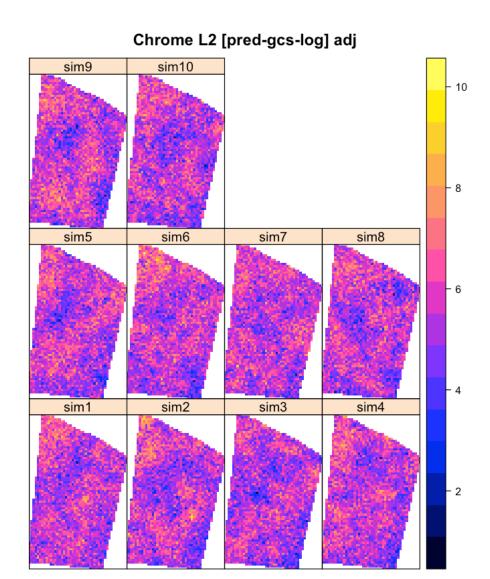
presentano un maggiore dettaglio spaziale,

sono caratterizzare da minore rumore.

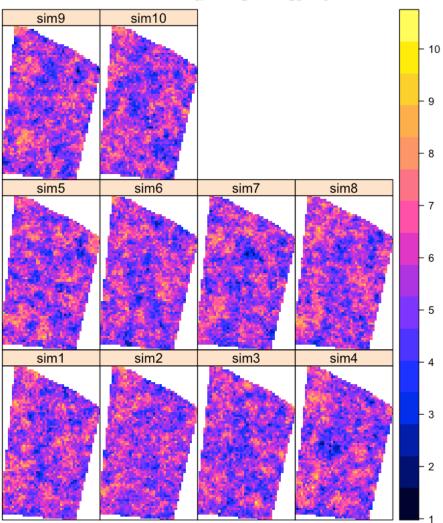
0.1.10 Conditional Gaussian Simulations

Chrome L1 [pred-gcs-log] adj









S A V E | Analysis of Uncertainty

[using conditional Gaussian simulation]