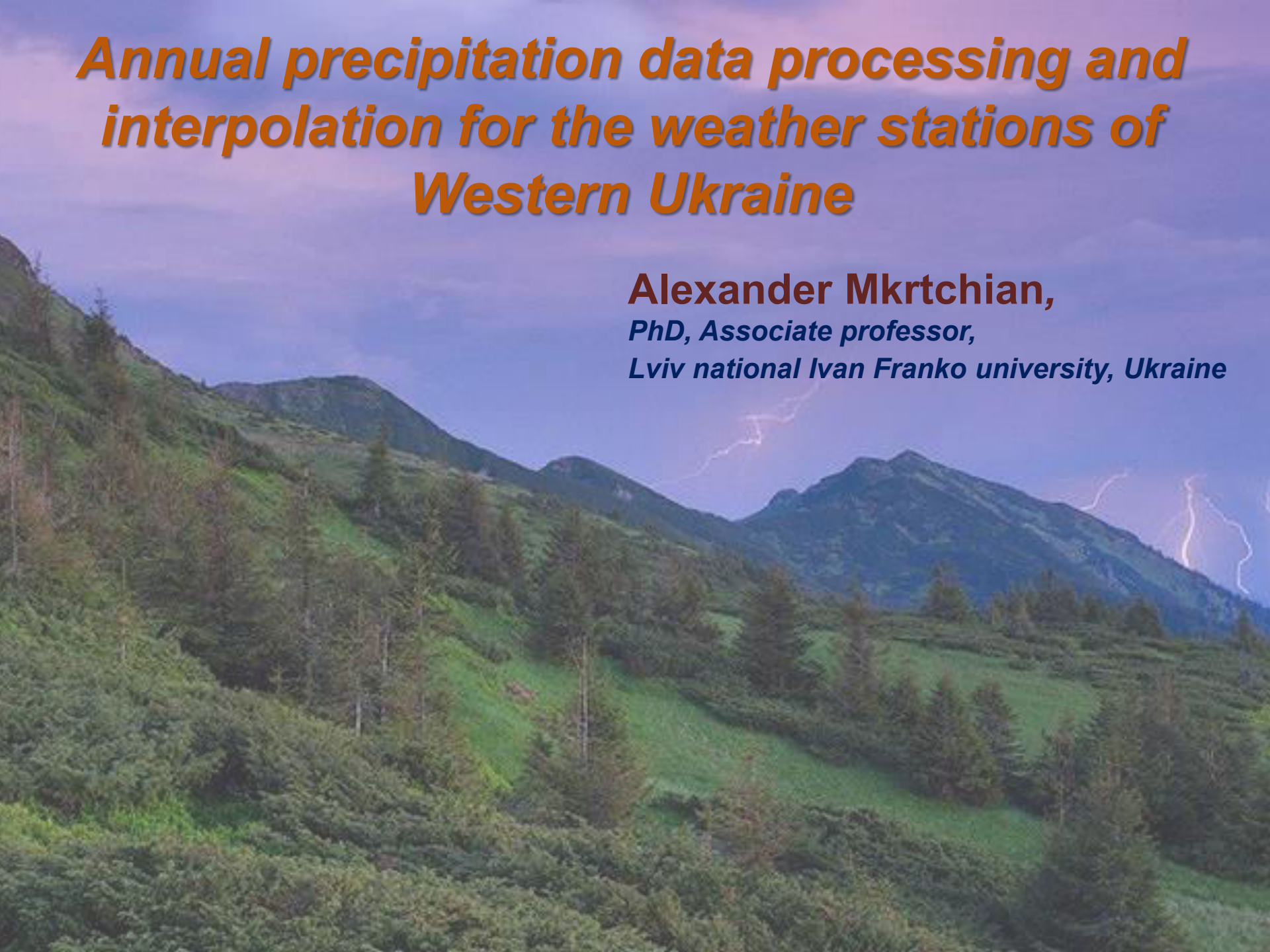
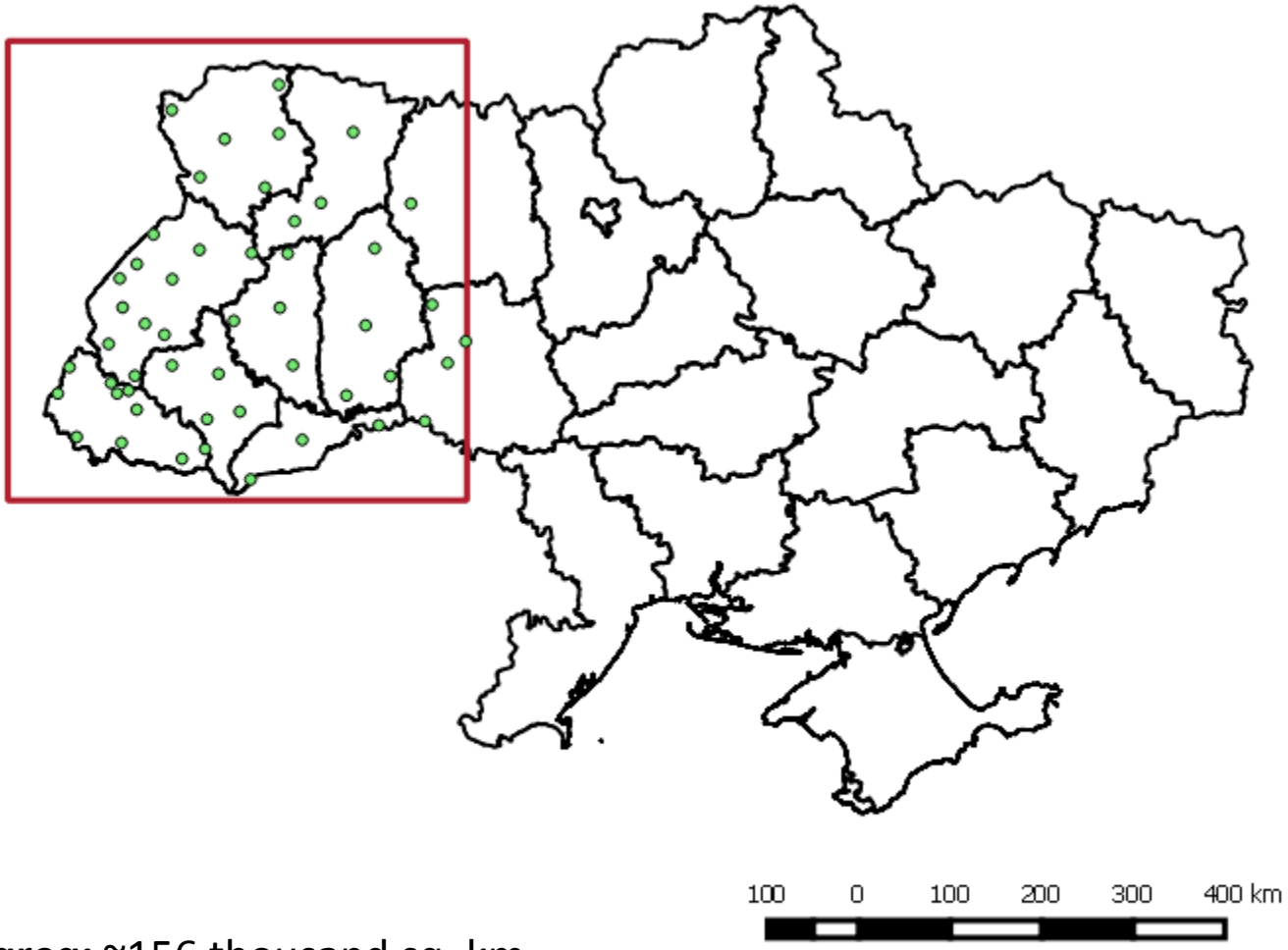
Annual precipitation data processing and interpolation for the weather stations of Western Ukraine



Alexander Mkrtchian,
PhD, Associate professor,
Lviv national Ivan Franko university, Ukraine

Study area



Study area: ~156 thousand sq. km
Includes 8 first rank administrative regions of Ukraine (totally)
and another 2 (partially)

Task: to create accurate and reliable map of average annual precipitation distribution based on available data

Problem points:

- ✓ sparse network of weather stations (separated by tens of km)
- ✓ discontinuous observation series, different stations having non-overlapping observation periods
- ✓ data are often classified, hardly accessible or expensive, often only exists in “paper” form
- ✓ researchers often lack funds to access chargeable data and proprietary software

Data source:

Global Historical Climatology Network (GHCN):

Online weather data archives:
https://www.ncdc.noaa.gov
http://www.ecad.eu

Peterson, Thomas C. and Russell S. Vose (1997). An overview of the Global Historical Climatology Network temperature data base. Bulletin of the American Meteorological Society 78 (12): 2837–2849

Total data consists of **33512** daily observations of precipitation on **50** weather stations,
for the period of 1924 - 2011 yrs.

Selected for the analysis: **3432** observations .
Selection criteria:

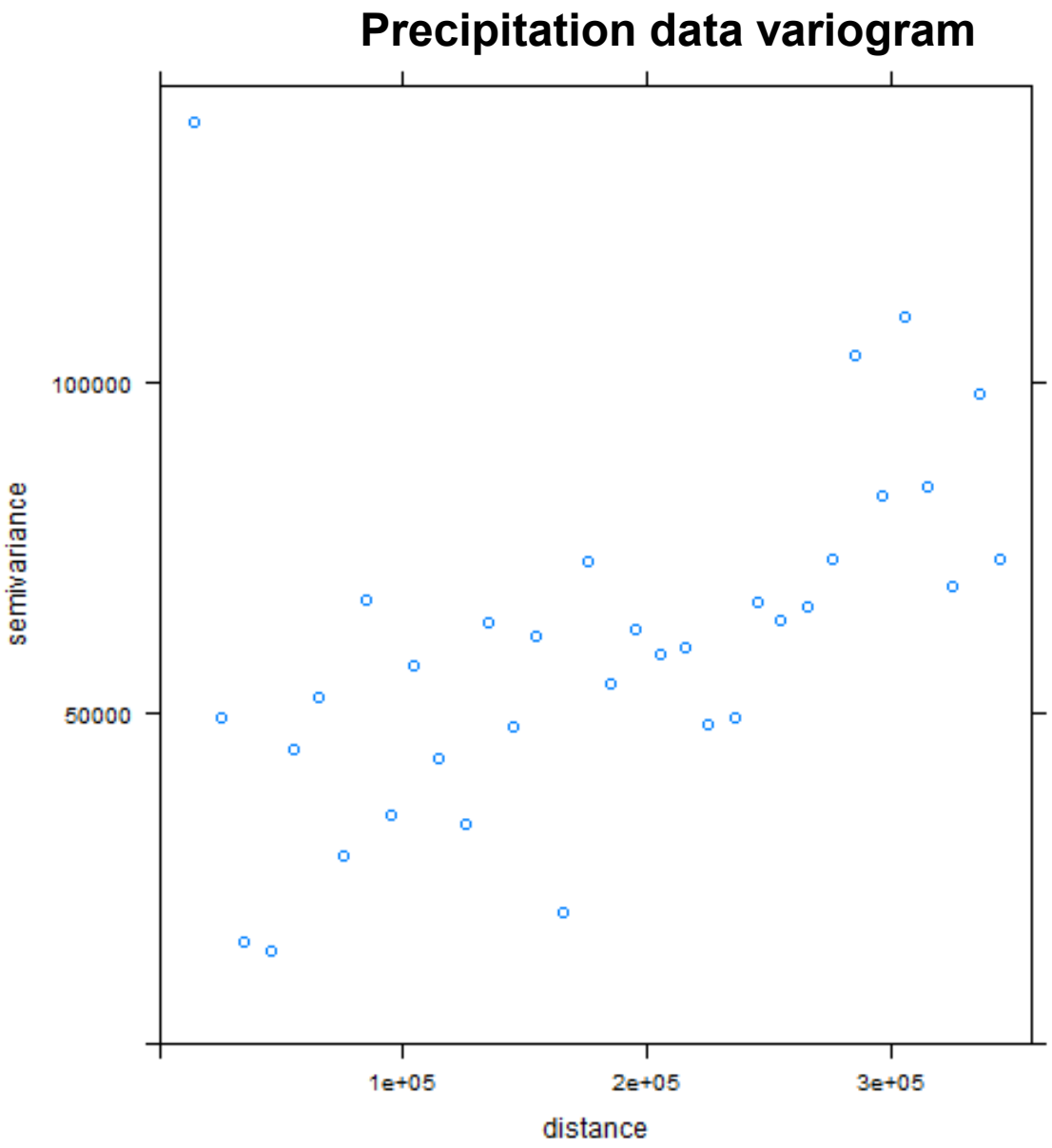
- 1961-1990 period
- daily data availability for all 50 weather stations

Data file example

```
EUROPEAN CLIMATE ASSESSMENT & DATASET (ECA&D), file created on 03-02-2016
THESE DATA CAN BE USED FREELY PROVIDED THAT THE FOLLOWING SOURCE IS ACKNOWLEDGED:

Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface
air temperature and precipitation series for the European Climate Assessment.
Int. J. of Climatol., 22, 1441-1453.
Data and metadata available at http://www.ecad.eu

FILE FORMAT (MISSING VALUE CODE IS -9999):

01-06 STAID: Station identifier
08-13 SOUID: Source identifier
15-22 DATE : Date YYYYMMDD
24-28 RR   : precipitation amount in 0.1 mm
30-34 Q_RR : Quality code for RR (0='valid'; 1='suspect'; 9='missing')

This is the series (SOUID: 103621) of UKRAINE, BEREGOVO (STAID: 1528)
See file sources.txt for more info.


 STAID, SOUID,    DATE,   RR, Q_RR
  1528,103621,19460801,    0,    0
  1528,103621,19460802,    0,    0
  1528,103621,19460803,    0,    0
  1528,103621,19460804,    0,    0
  1528,103621,19460805,    0,    0
  1528,103621,19460806,    0,    0
  1528,103621,19460807,    0,    0
  1528,103621,19460808,    0,    0
  1528,103621,19460809,    0,    0
  1528,103621,19460810,    0,    0
  1528,103621,19460811,    0,    0
  1528,103621,19460812,    0,    0
  1528,103621,19460813,    0,    0
  1528,103621,19460814,-9999,    9
  1528,103621,19460815,    0,    0
  1528,103621,19460816,   25,    0
  1528,103621,19460817,    0,    0
  1528,103621,19460818,    0,    0
  1528,103621,19460819,-9999,    9
  1528,103621,19460820,-9999,    9
  1528,103621,19460821,    5,    0
  1528,103621,19460822,    0,    0
  1528,103621,19460823,   79,    0
  1528,103621,19460824,    0,    0
  1528,103621,19460825,  138,    0
  1528,103621,19460826,-9999,    9
```

Primary processing of climatic data in R

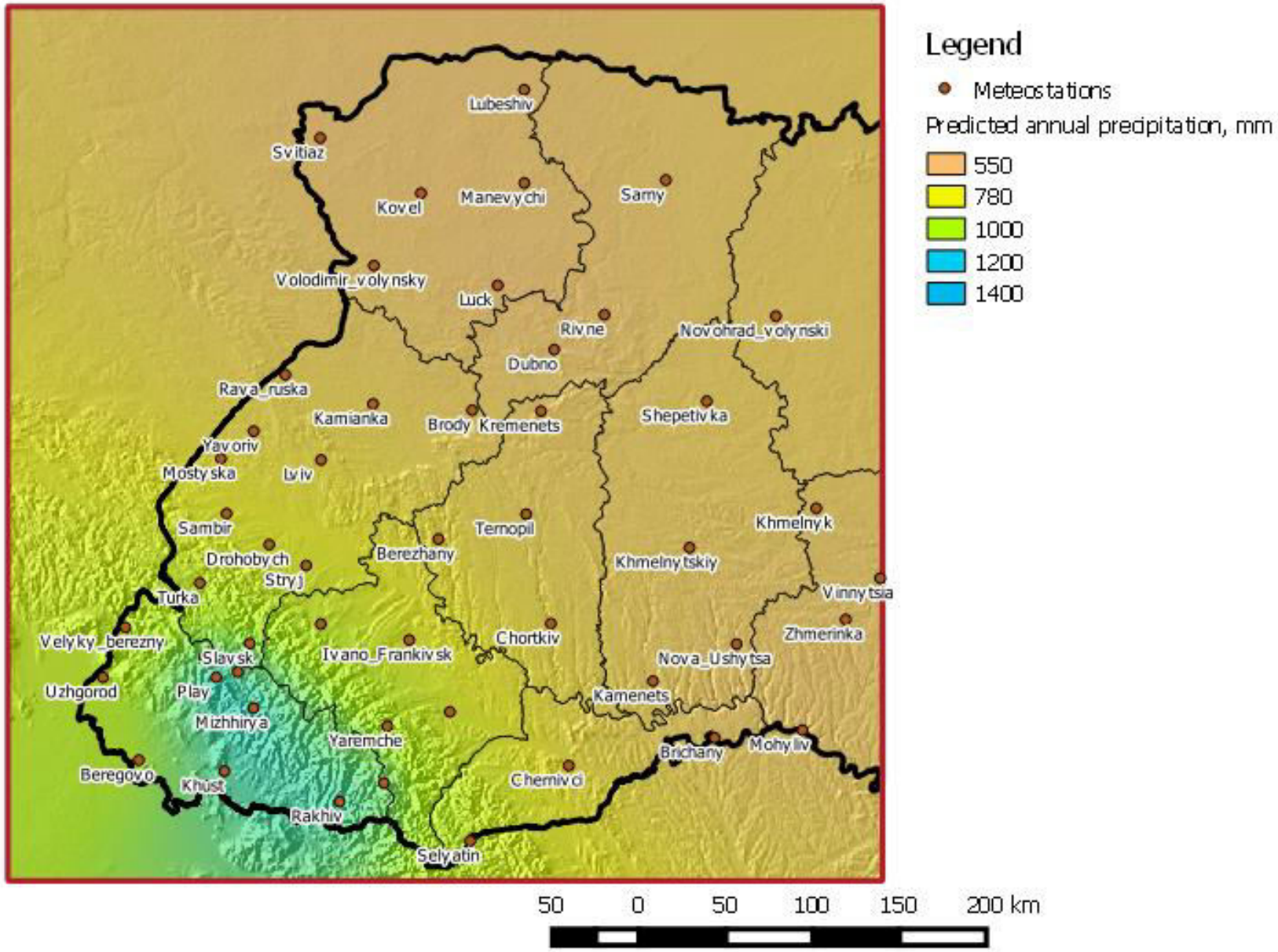
```
lf <- list.files()
b = data.frame()
for (fname in lf) {
    a <- read.csv (fname, skip=21)
    a <- a[ ,-c(1,2,5)]
    fn <- substr (fname, 1, (nchar(fname)-4))
    colnames(a)[2] <- fn
    a[which(a[[fn]] == -9999), fn] <- NA
    if (nrow(b)==0) {
        b <- a
    } else {
        b <- merge (b,a, all = TRUE)
    }
}
rm(a)
b$year <- substr (b$DATE, 1,4)
b$month <- substr (b$DATE, 5,6)
b$day <- substr (b$DATE, 7,8)
b$DATE <- NULL
b <- b[c(51:53, 1:50)]


obsm <- tapply (b$RR, b$month, function (x) (sum(!is.na(x)))
temp1 <- by(a[, 4:53], a$month, function (x) sapply (x, sum))
temp2 <- do.call(rbind, temp1)
temp3 <- temp2 * dm / obsm
temp4 <- as.data.frame(temp3)
pr_an <- sapply(temp4, sum)
```



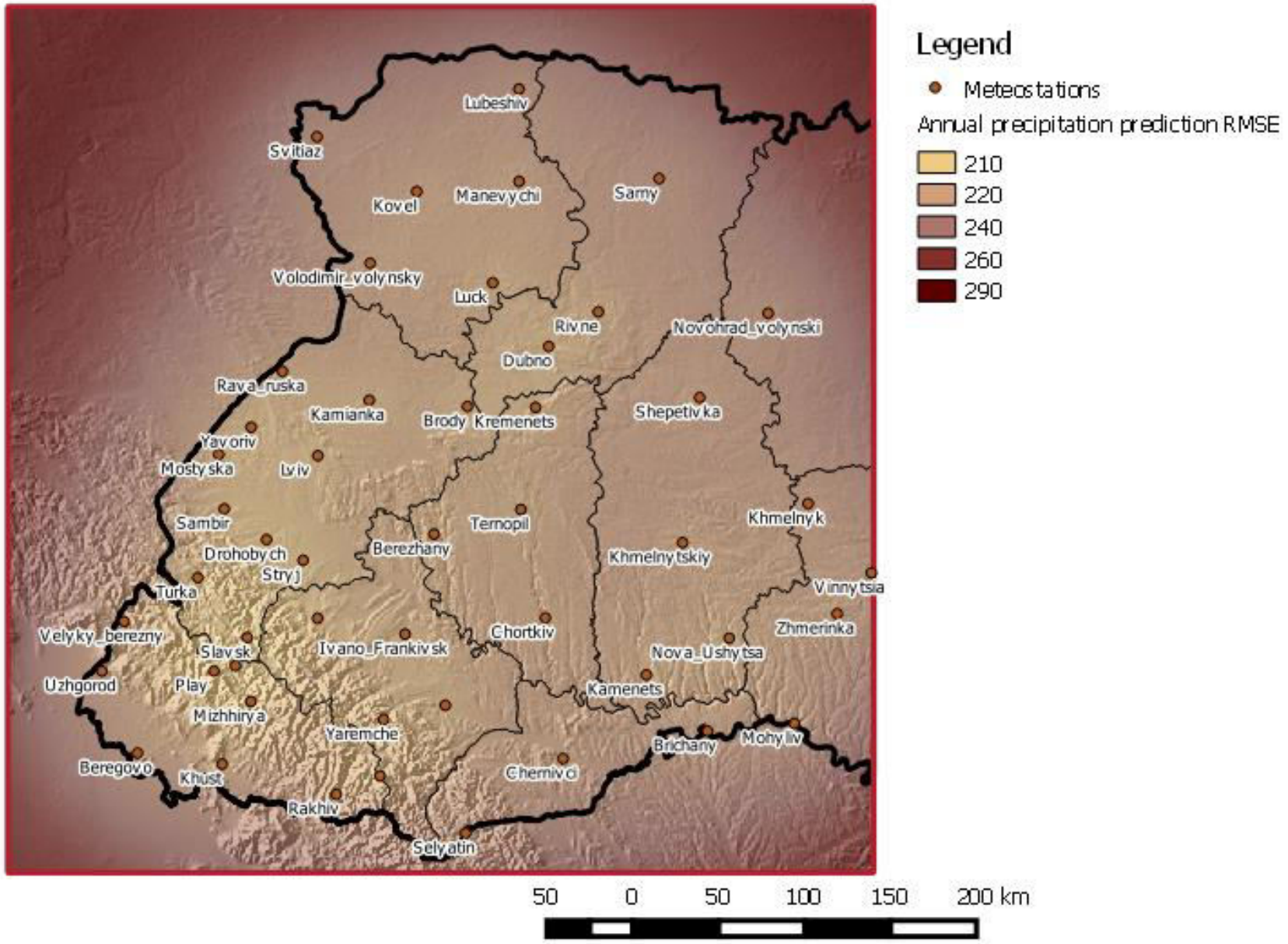
Fitted variogram parameters (exponential model):

nugget 30000 mm^2,
partial sill 66000 mm^2,
range 468 km

Annual precipitation interpolated by ordinary kriging



Predicted RMSE of interpolation by ordinary kriging (mm^2)



Study area terrain (absolute elevation)

Legend

- Meteostations

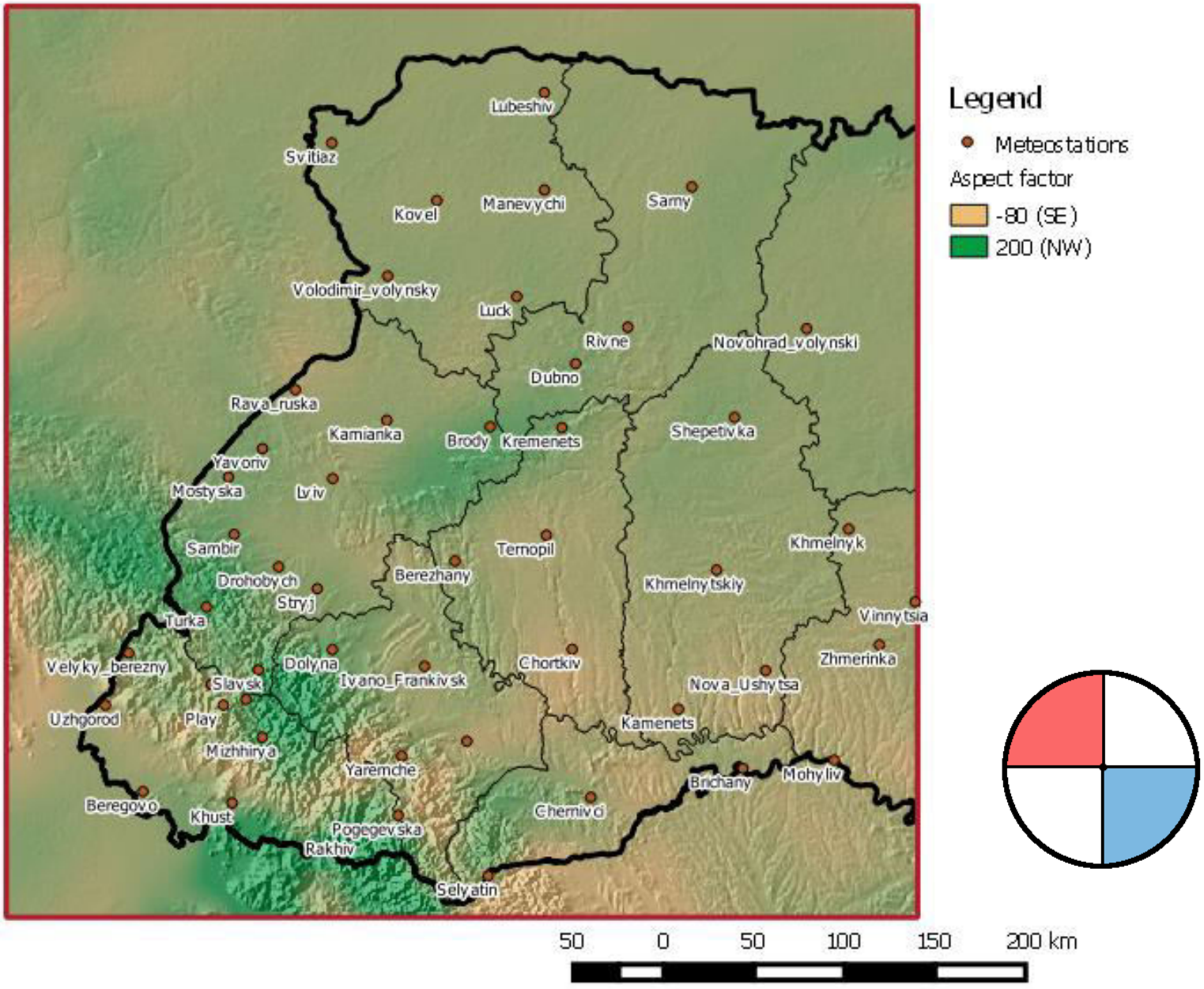
Elevation

- 100
- 350
- 600
- 900
- 1400

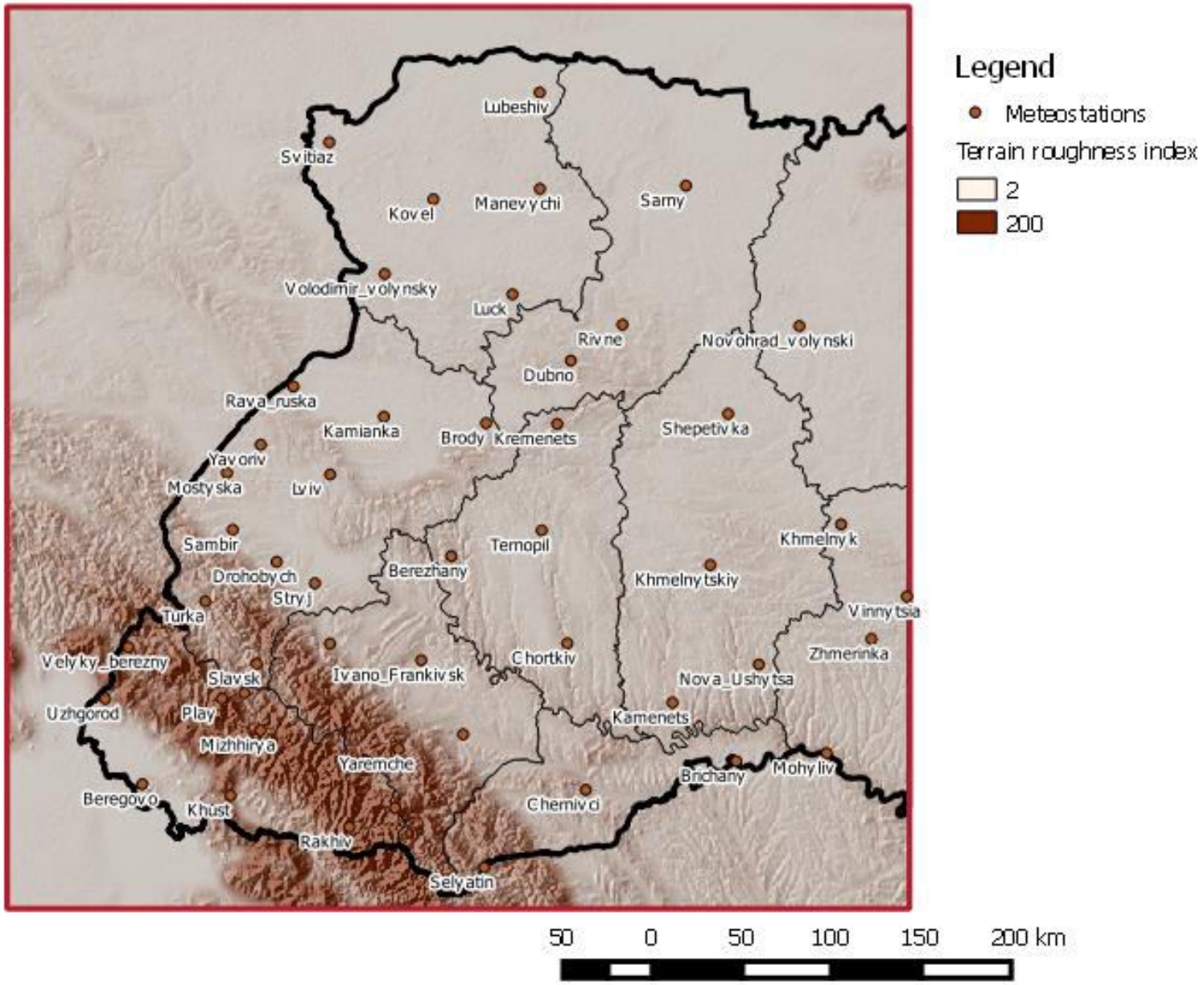
50 0 50 100 150 200 km

Source of terrain data:
SRTM data V4,
available from
http://srtm.csi.cgiar.org

Terrain Aspect factor (NW-SE, r = 50.4 km)



Terrain Roughness factor



Multiple regression model

```
Residuals:
     Min       1Q   Median       3Q      Max
-144.095  -36.397    2.025   42.174  123.160

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 532.48879   16.07466  33.126  < 2e-16 ***
dem           0.30150    0.06243   4.830 1.69e-05 ***
Dem_std       2.10120    0.25321   8.298 1.51e-10 ***
asp_nwse_7    1.68166    0.39018   4.310 9.05e-05 ***
asp_nwse_5   -1.20073    0.43651  -2.751  0.00860 **
asp_we        0.15685    0.05586   2.808  0.00741 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 65.83 on 44 degrees of freedom
Multiple R-squared:  0.9383,    Adjusted R-squared:  0.9313
F-statistic: 133.9 on 5 and 44 DF,  p-value: < 2.2e-16
```

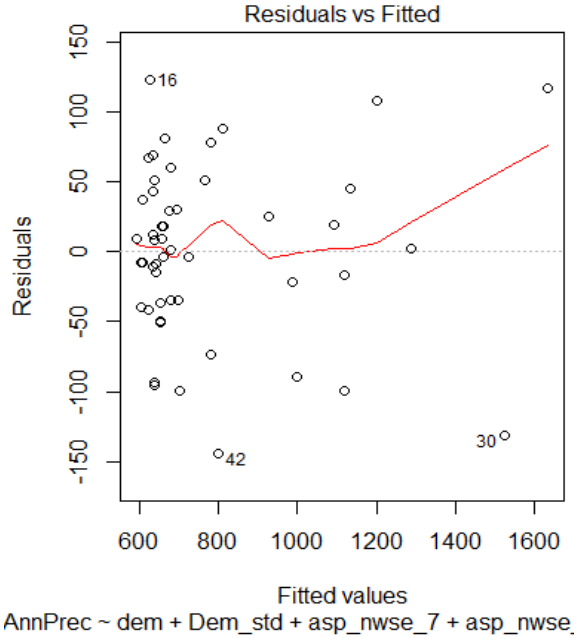
Regression model parameters

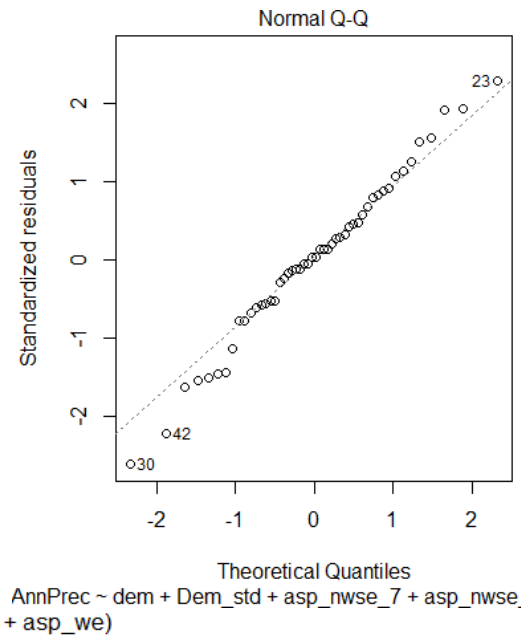
Terrain characteristic	Moving circle r, km	Regression model parameters		
		Coeff.	t value	Pr(>|t|)
Terrain elevation	-	0.3	4.83	1.7e-05
Terrain roughness	7.2	2.1	8.3	1.5e-10
Aspect factor NW/SE	36	-1.2	-2.75	0.0086
Aspect factor NW/SE	50.4	1.68	4.31	9e-05
Aspect factor W/E	36	0.16	2.81	0.0074

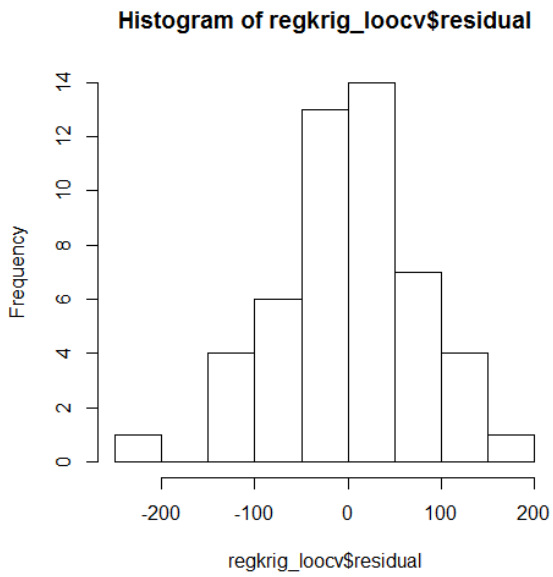
Model output:
Residual standard error: 65.83 on 44 degrees of freedom
Multiple R-squared: 0.9383, Adjusted R-squared: 0.9313
F-statistic: 133.9 on 5 and 44 DF, p-value: < 2.2e-16

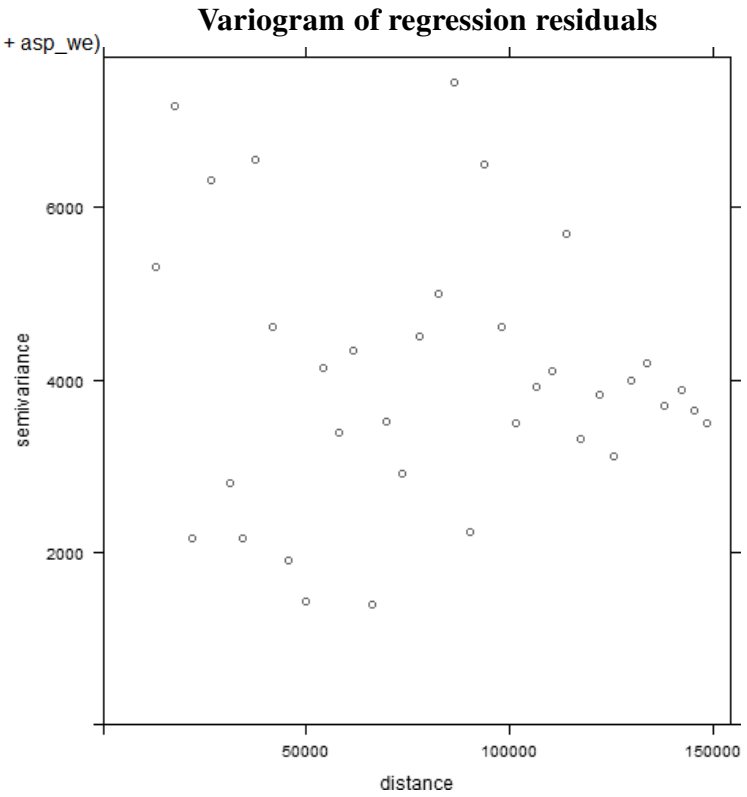
Shapiro-Wilk normality test of residuals:
W = 0.98597, p-value = 0.8125

Validity testing of regression model

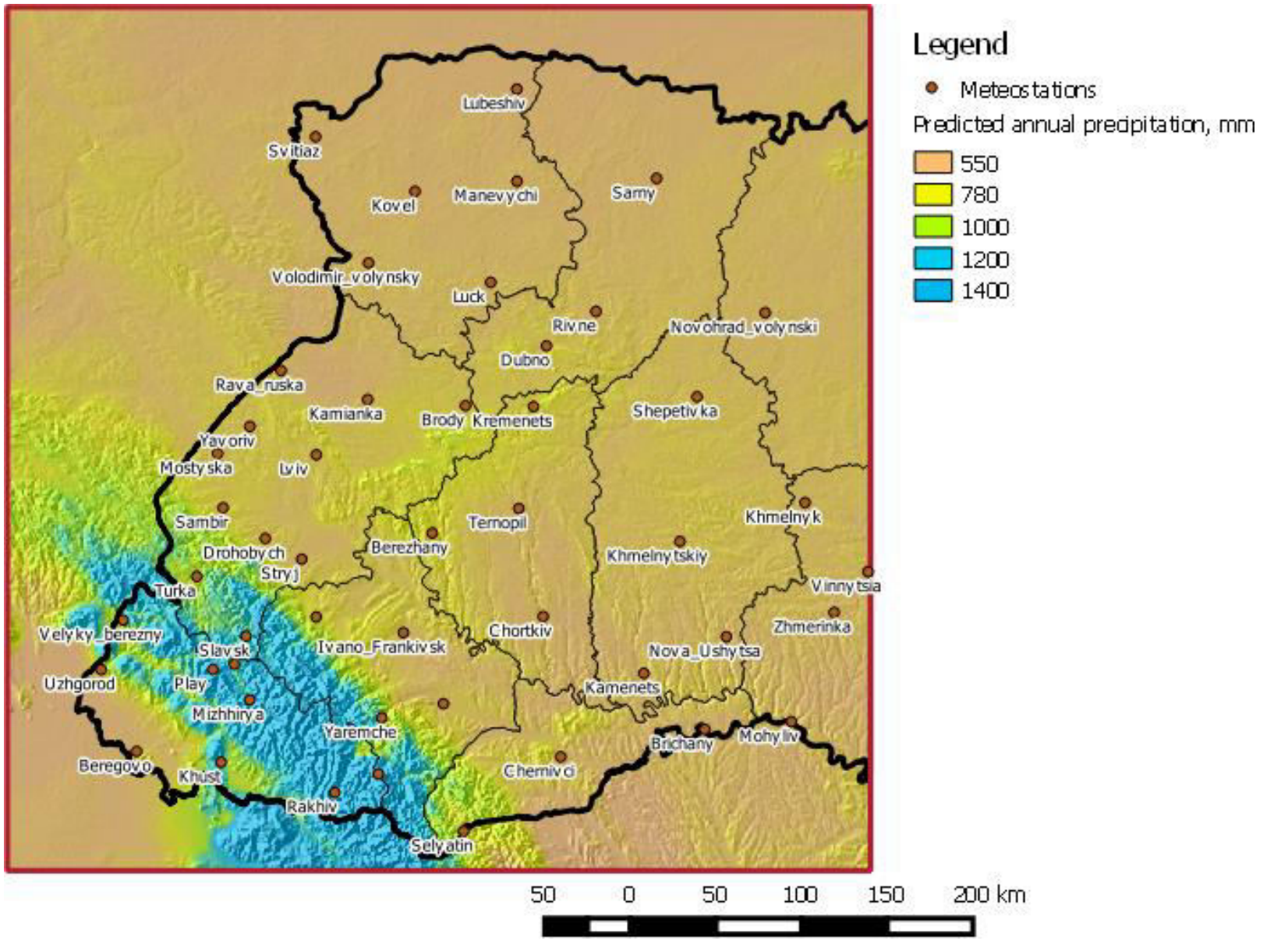




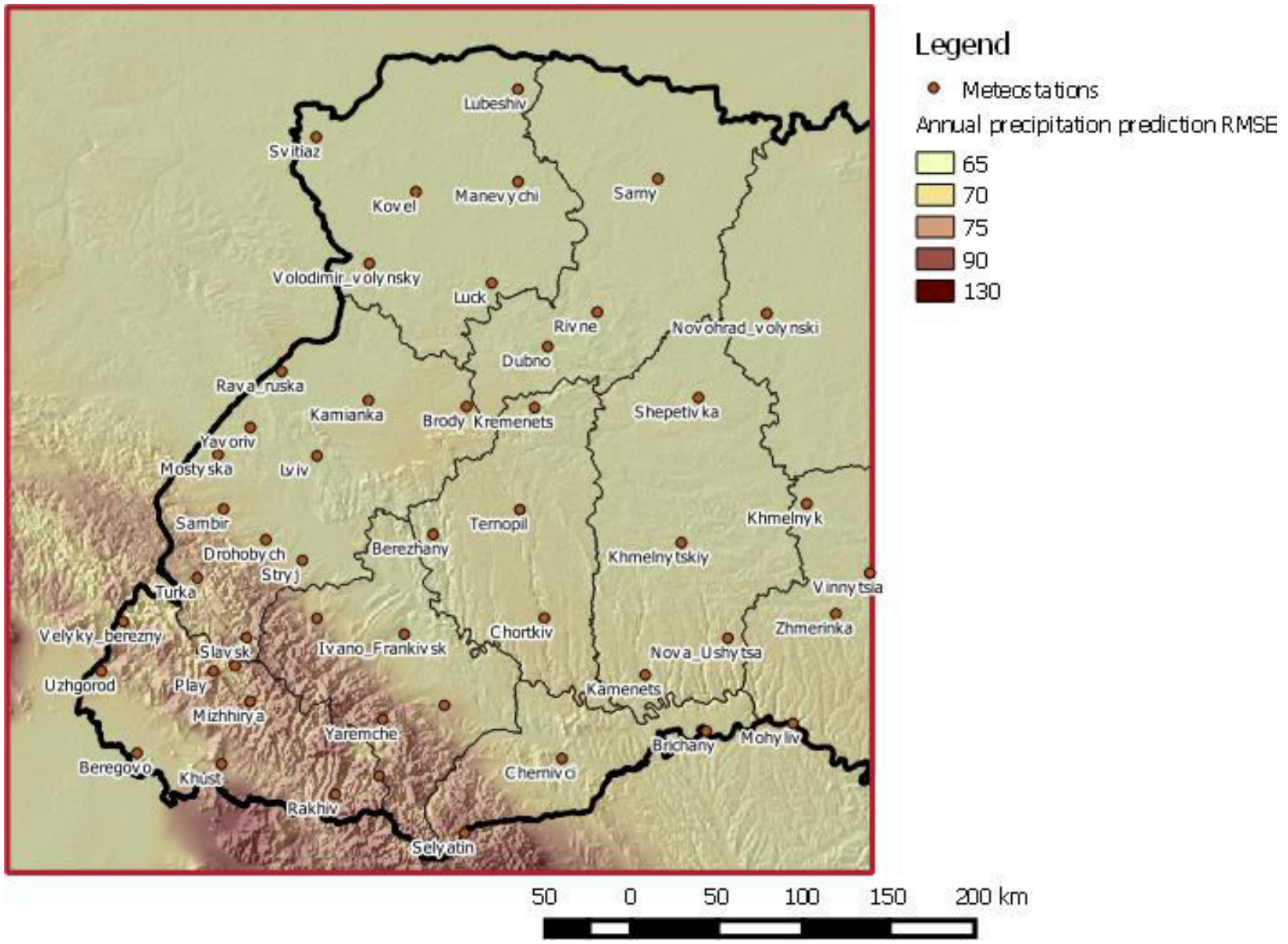




Annual precipitation interpolated by multiple regression model

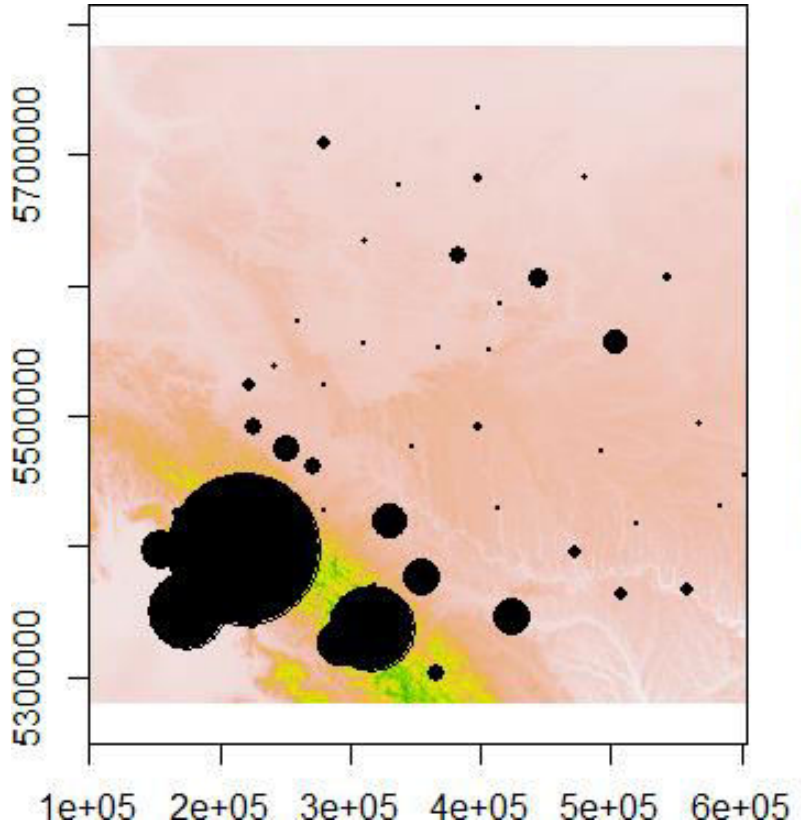


Predicted RMSE of interpolation by multiple regression model (mm^2)

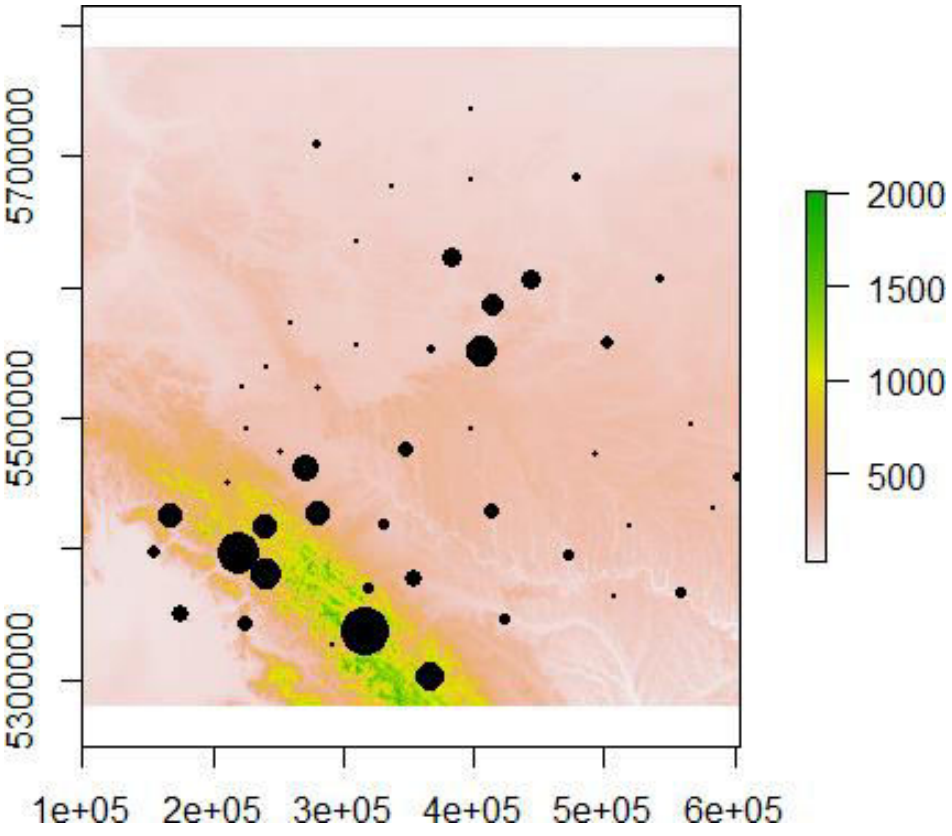


Cross-validation of interpolation results

(by krige.cv {gstat}, **leave-one-out cross validation)**



Simple kriging

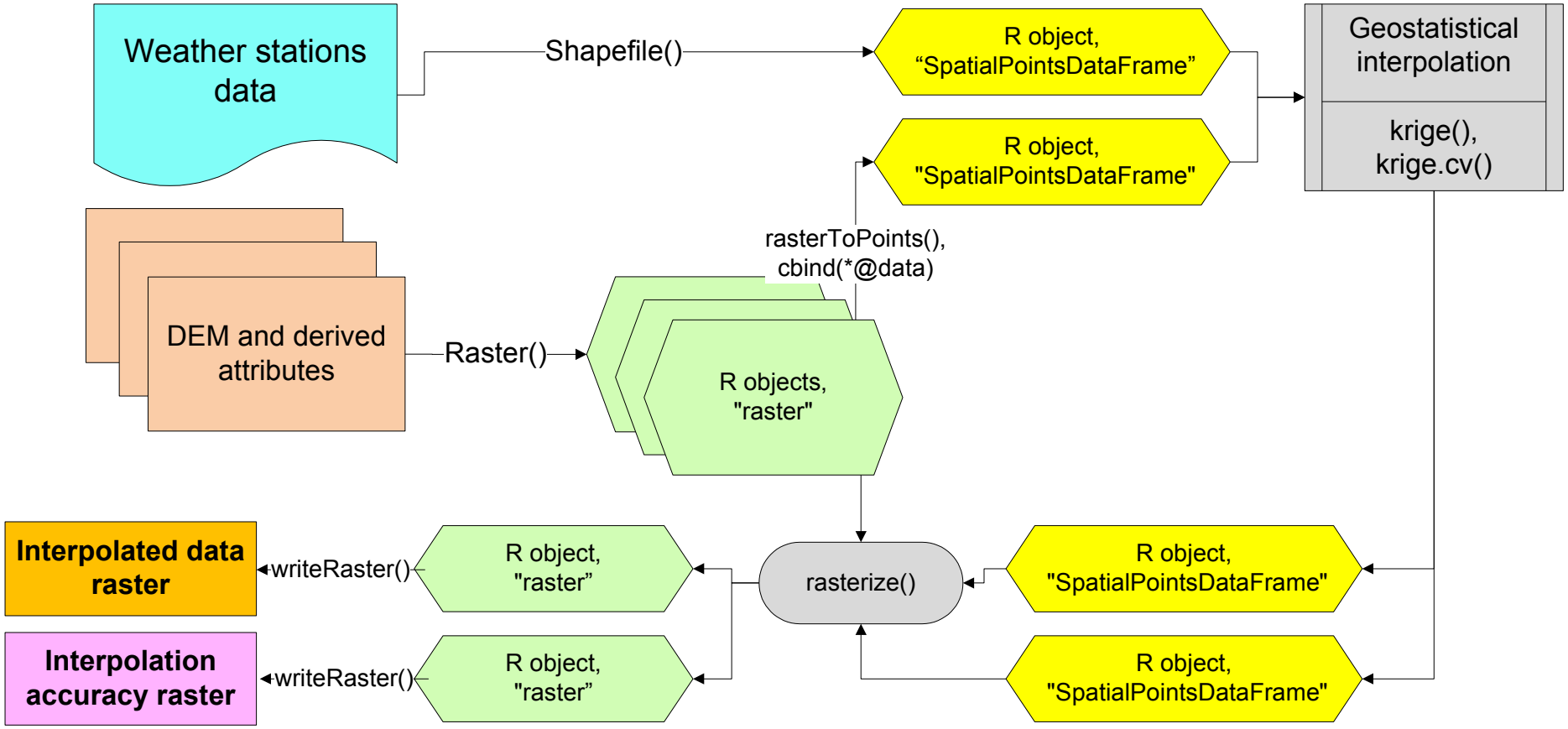


Multiple regression

Annual precipitation variance, mm²:

Initial	After ordinary kriging	After regression modeling
63096	22856 (36,2%)	6087(9,6%)

General workflow



Prospective ways of refining the interpolation (directions of further research)

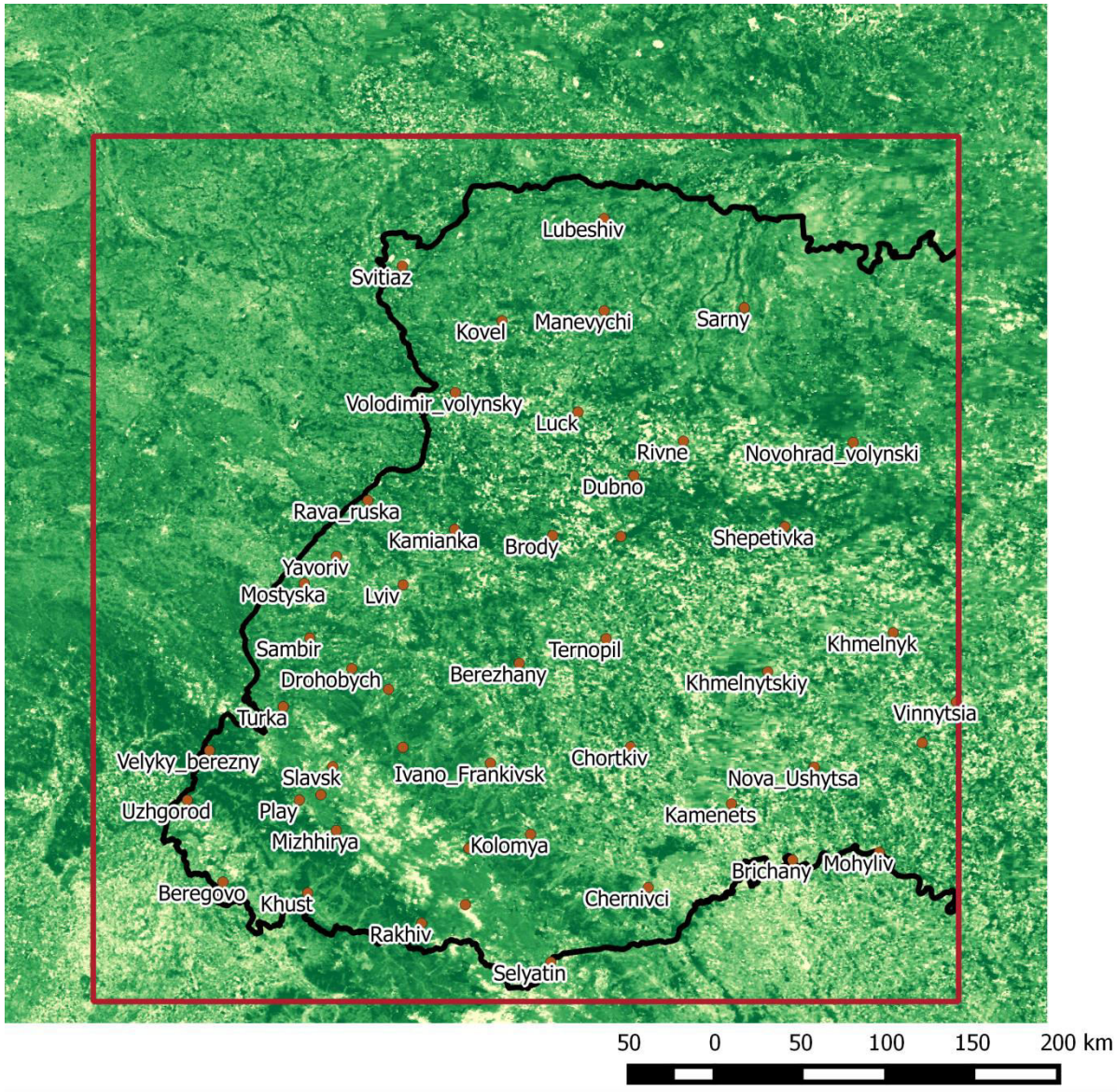
Data refining

- utilizing rain gauges data (higher spatial density, however usually short and interrupted series, more prone to errors, inadequate georeferencing)
- utilizing data from neighboring countries (requires cross-border harmonization)
- advanced methods for dealing with discontinuous observation series (automatic completion of missing data, correction for gaps and inhomogeneities in data series)

Model refining

- incorporating additional explanatory variables, connected with terrain attributes as well as land cover character
- engaging more sophisticated statistical and geostatistical models
- advanced methods for dealing with scaling and finding the optimal spatial scale for dependent variables
- engaging theoretical models that explicate processes of precipitation formation and redistribution

MODIS 500m 16 days NDVI



Processing and geostatistical interpolation of annual precipitation data for western part of Ukraine

Alexander Mkrtchian, PhD

Associate professor,
Lviv national Ivan Franko university

alemkrt@gmail.com