量測空間相依性：半變異元分析

Measuring spatial dependency: Semi-variogram analysis

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

複習：空間相依性或空間自相關

Moran's I coefficient

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

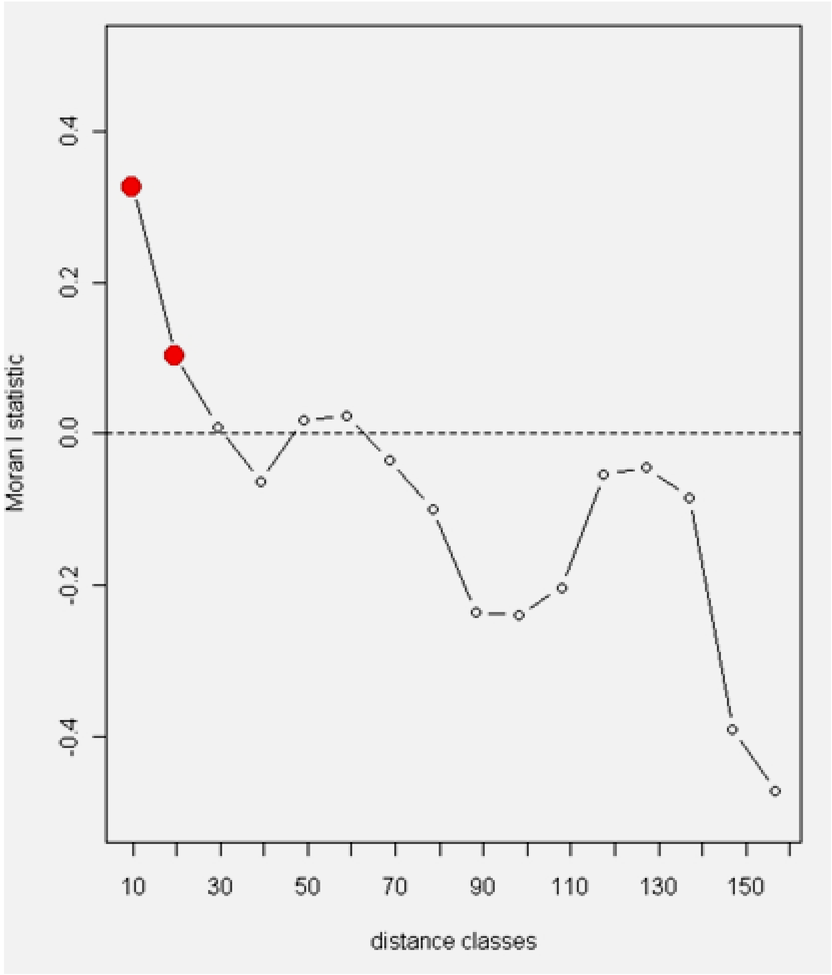
N: no. of spatial units

$w_{i,j}$: a matrix of spatial weights

$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$ (sum of all $w_{i,j}$)

複習：Moran Correlogram

Correlogram: plot distance on X-axis against correlation coefficient on Y-axis

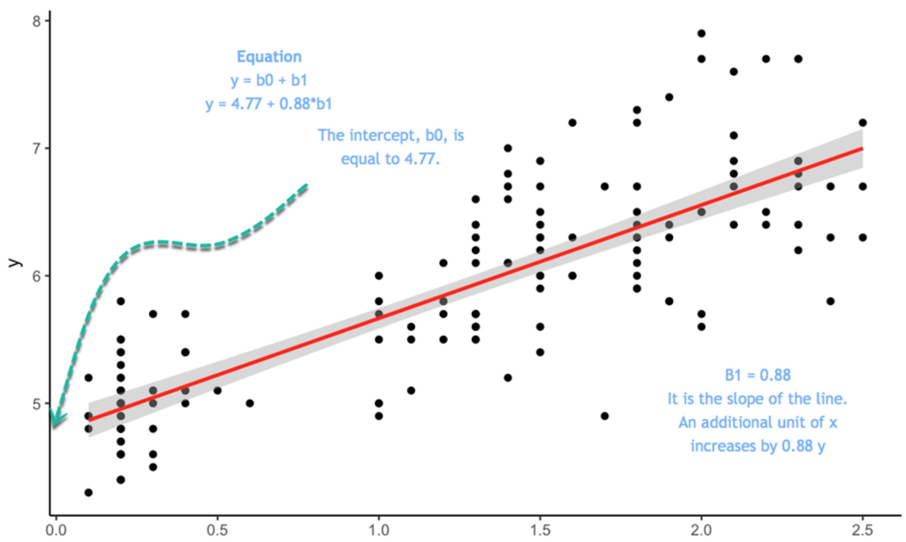


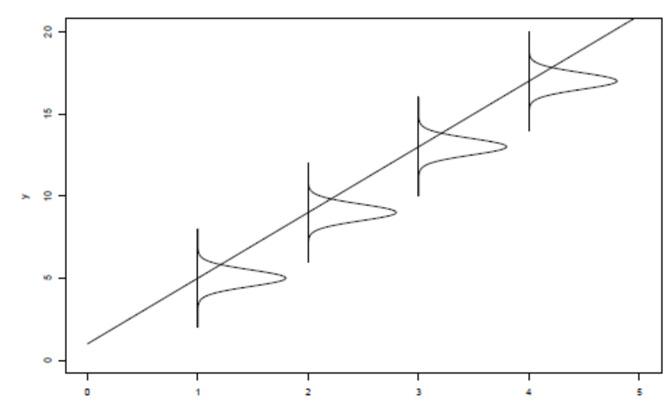
複習：Simple Linear Regression Model

$$y_i = \alpha + \beta x_i + \varepsilon_i .$$

$$Y_i \sim n(\alpha + \beta x_i, \sigma^2)$$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad y \in \mathbf{R}^n$$



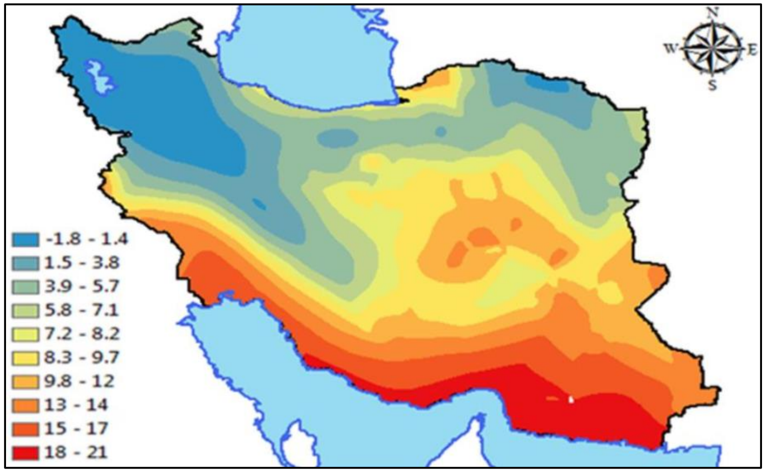


Geostatistical approach: Random field (隨機場)

Z(s): a regionalized random variable that is associated with a true measurement, z(s), that characterizes the quantity of a variable at point s.

$$Z(s) = \mu(s) + \eta(s) + \varepsilon(s)\ ;\quad s \in D\,,$$

random variable, mu = 0



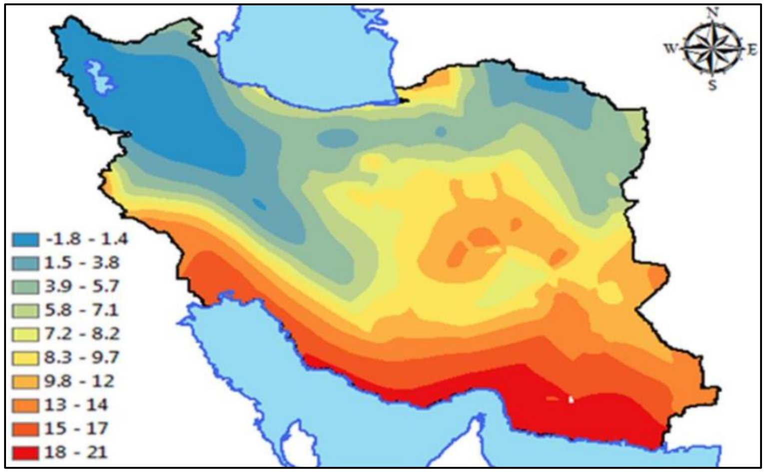
Temperature

Stationary 穩定態

$$E(\eta(\boldsymbol{s}+\boldsymbol{h})-\eta(\boldsymbol{s}))=0\,,$$

$$\operatorname{var}(\eta(\boldsymbol{s}+\boldsymbol{h})-\eta(\boldsymbol{s}))=2\gamma(\boldsymbol{h})\,.$$

Variogram (變異元)

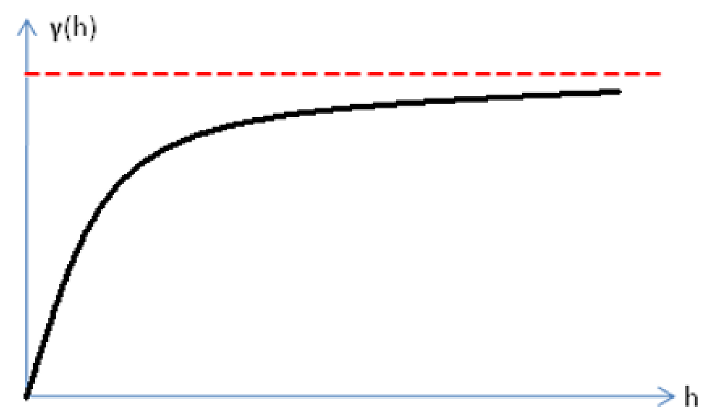


Variogram

- A variogram might be thought of as "***dissimilarity*** between point values as a function of distance", such that the dissimilarity is greater for points that are farther apart

$$2\gamma(h) = E\{[Z(x+h) - Z(x)]^2\}$$

Variogram (變異元)



Variogram γ(h)

Variogram: Mathematical definition

$$2\gamma(h) = E\{[Z(x+h) - Z(x)]^2\}$$

⬇

$$2\gamma(h) = average\left[\left(Z(i) - Z(j)\right)^2\right]$$

⬇

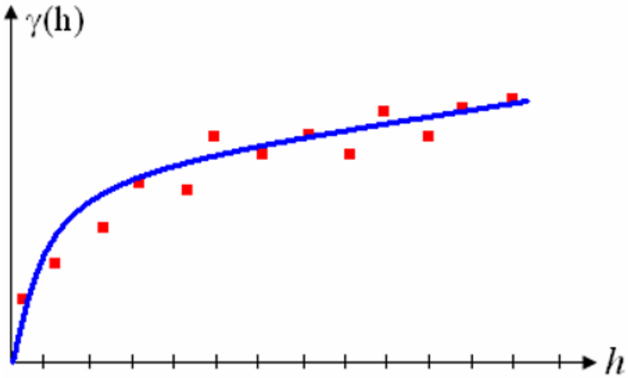
$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{N(h)} (Z(s_i) - Z(s_j))^2,$$

N(h): the number of paired comparisons at lag h.

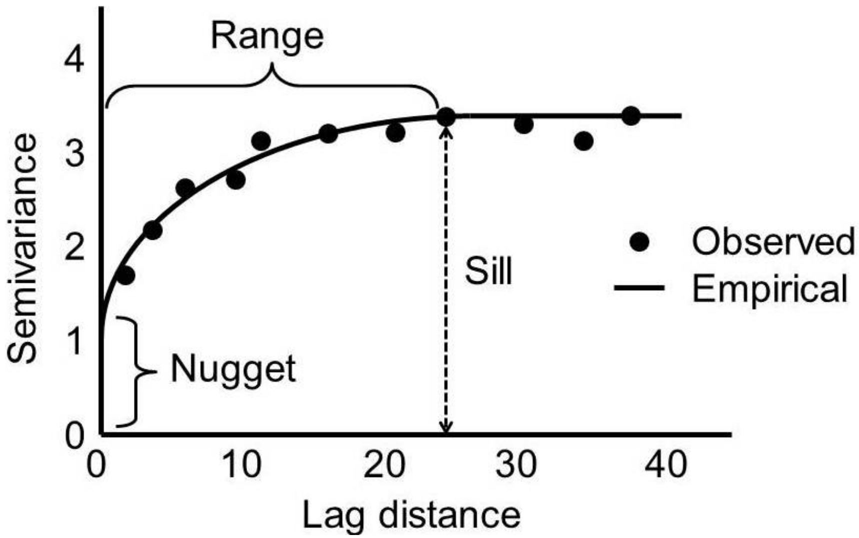
Semivariogram 半變異元圖 $\gamma(h)$

$$\gamma(h)=\frac{1}{2n(h)}\sum_{i=1}^{n(h)}\left[z(x_i+h)-Z(x_i)\ \right]^2$$

where n is the number of sample points, $Z(x_i)$ is the measured sample value at location x_i, $Z(x_{i+h})$ is the sample value at location x_{i+h}, regionalized variable $Z(x)$, and $n(h)$ is the number of pairs of observations a distance h apart.

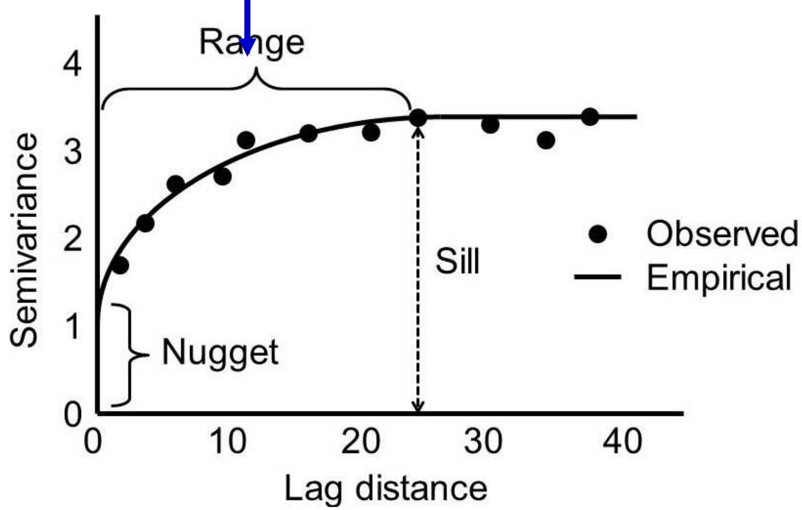


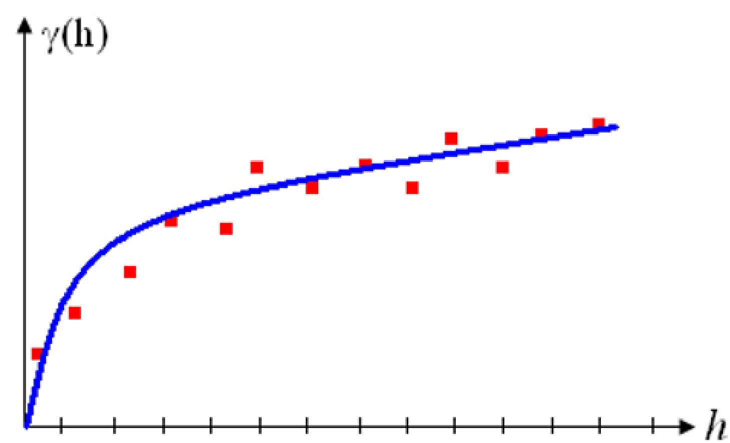
Concept of Semivariogram



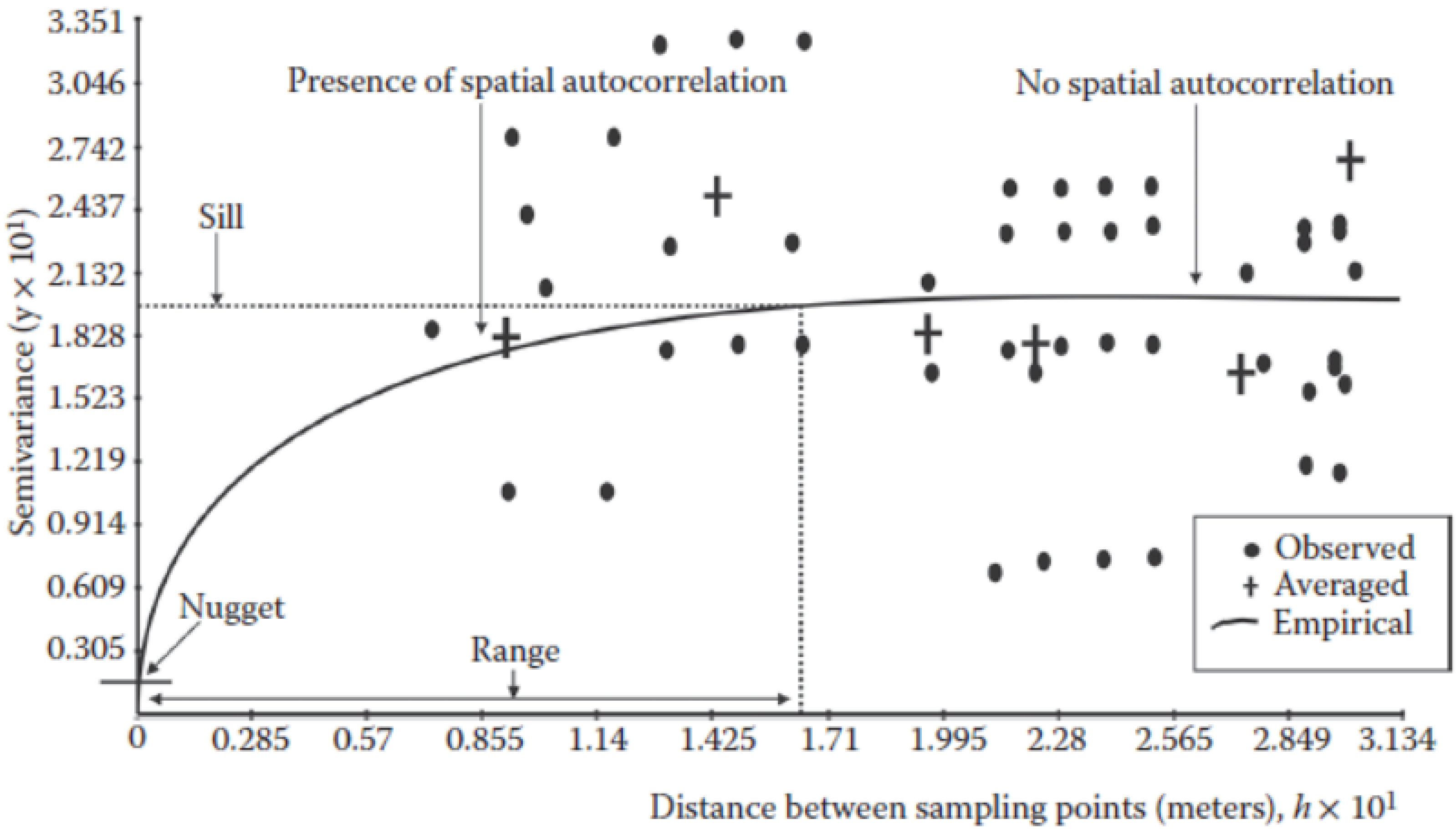
Fitting a Variogram Model

- Now, we're going to fit a variogram model (i.e., curve) to the empirical variogram
- That is, based on the shape of the empirical variogram, different variogram curves might be fit
- The curve fitting generally employs the method of least squares – the same method that's used in regression analysis





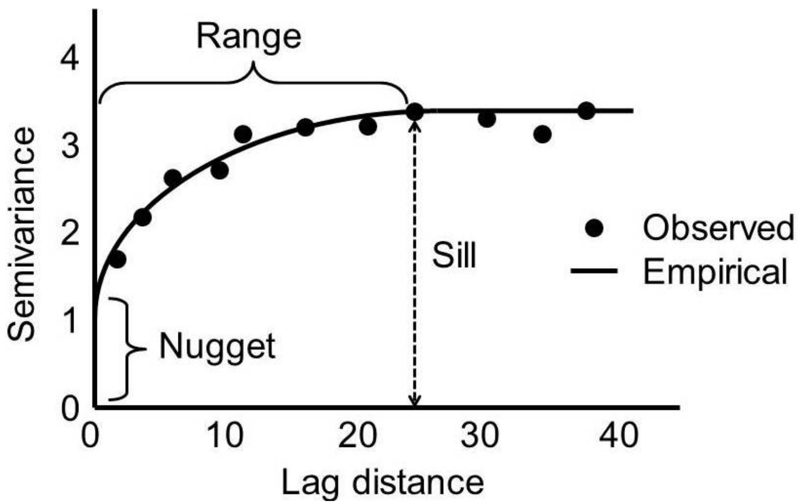
Variogram Model



The Variogram Parameters

- The variogram models are a function of three parameters, known as the range, the sill, and the nugget.
- Semivariance value where it flattens out is called a "sill."
- The distance range for which there is a slope is called the "*neighborhood*"; this is where there is positive spatial structure
- The intercept is called the "nugget" and represents random noise that is spatially independent

$$\tilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{u=1}^{N} (z(u) - z(u+h))^2$$



Variogram Models

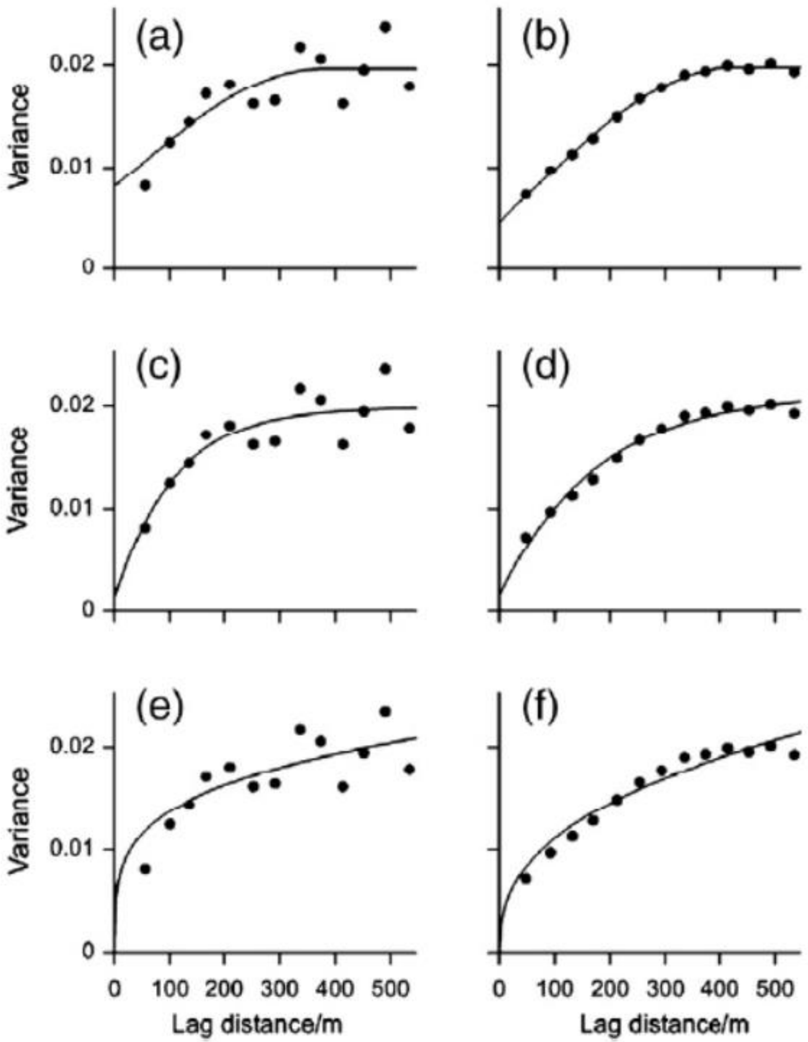
spherical model

$$\gamma(h) = c_0 + c_1\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], \quad \text{for } 0 < h < a,$$

$$\gamma(h) = c_0 + c_1, \quad \text{for } h \geq a,$$

exponential model

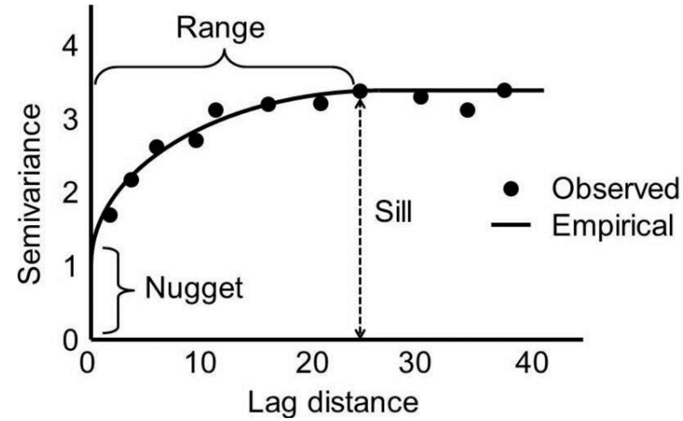
power function



Variogram Models

spherical model

- the most widely used.
- Monotonically non-decreasing:
  as *h* increases, the value of *γ(h)* does not decrease - i.e., it goes up (until *h≤r*) or stays the same (*h>r*)



nugget (sill-nugget)

$$\gamma(h) = c_0 + c_1\left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], \quad \text{for } 0 < h < a,$$

range

$$\gamma(h) = c_0 + c_1, \quad \text{for } h \geq a,$$

Variogram Models

$$\gamma(h) = c\left[1 - \exp\frac{-3h}{a}\right]$$

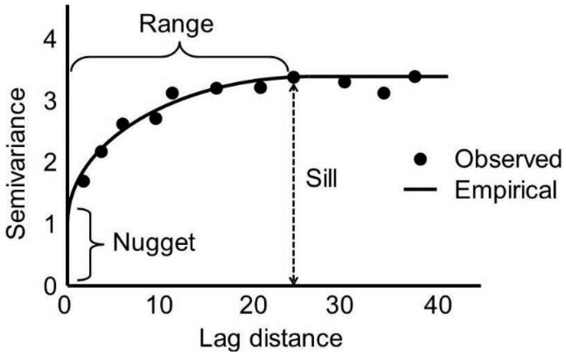
exponential model

- similar to the spherical model, but assumes that the correlation never reaches exactly zero, regardless of how great the distances between points are
- In other words, the variogram approaches the value of the sill asymptotically
- Because the sill is never actually reached, the range is generally considered to be the smallest distance after which the covariance is 5% or less of the maximum covariance
- The model is monotonically increasing
  - I.e., as *h* goes up, so does $\gamma(h)$

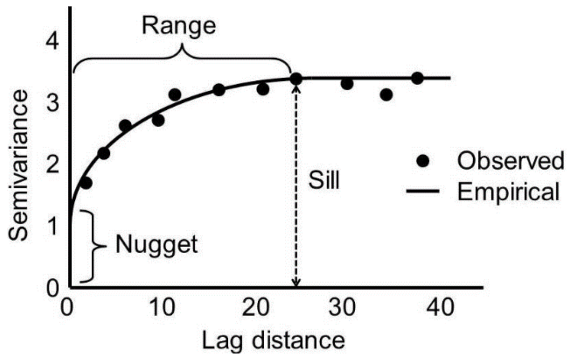
Variogram Models

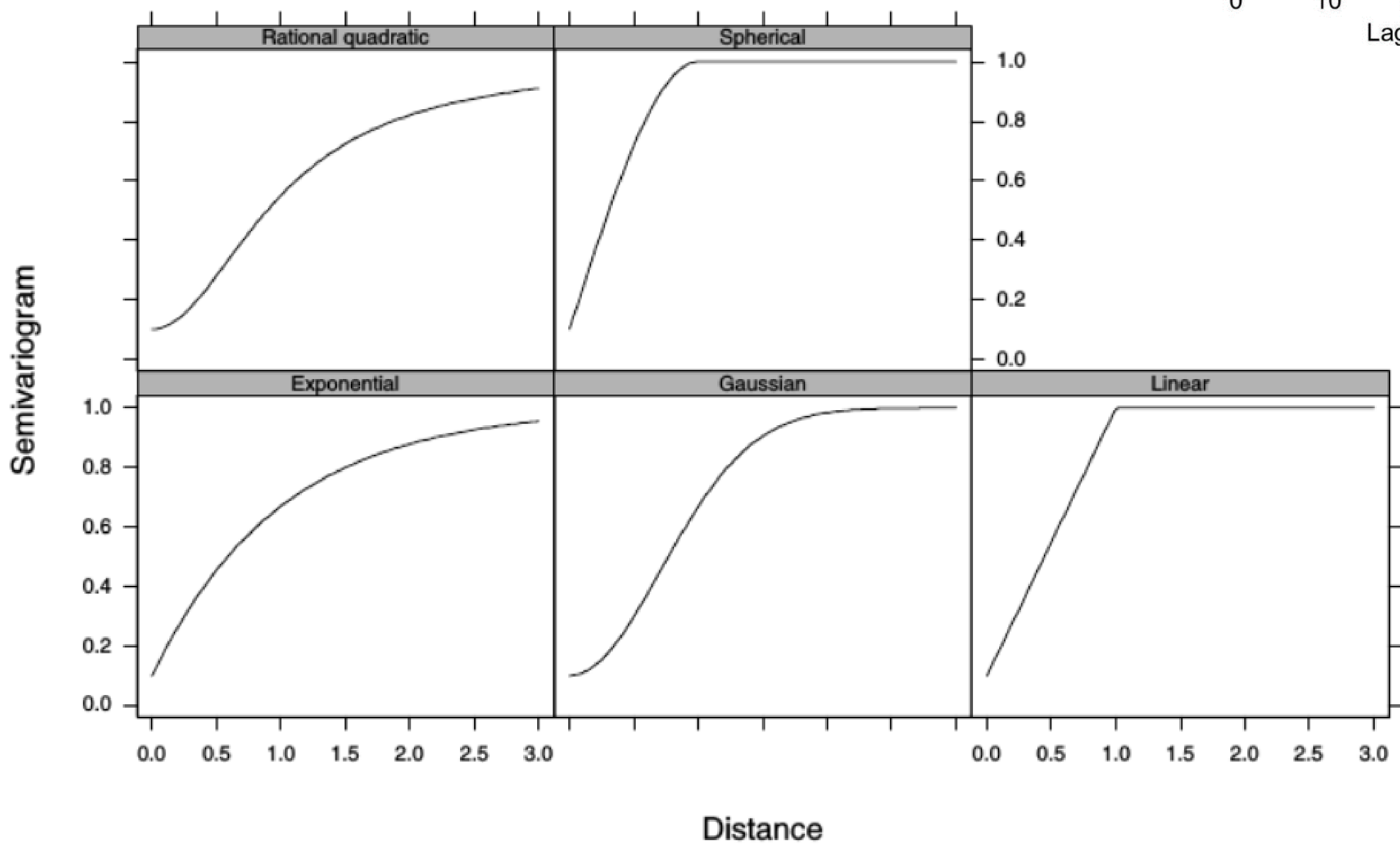
Model Variogram Types

Value	Model variogram type (from VarModel)	Equation
1	Spherical	$\gamma(h) = c\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right]$
2	Exponential	$\gamma(h) = c\left[1 - \exp\frac{-3h}{a}\right]$
3	Gaussian	$\gamma(h) = c\left[1 - \exp\left(\frac{-3h}{a}\right)^2\right]$



Variogram Models

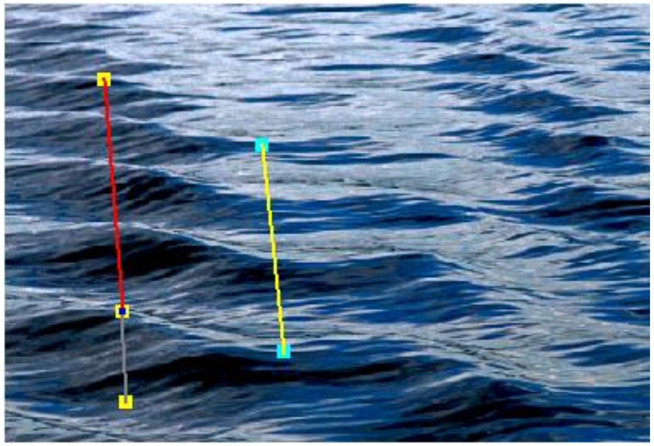


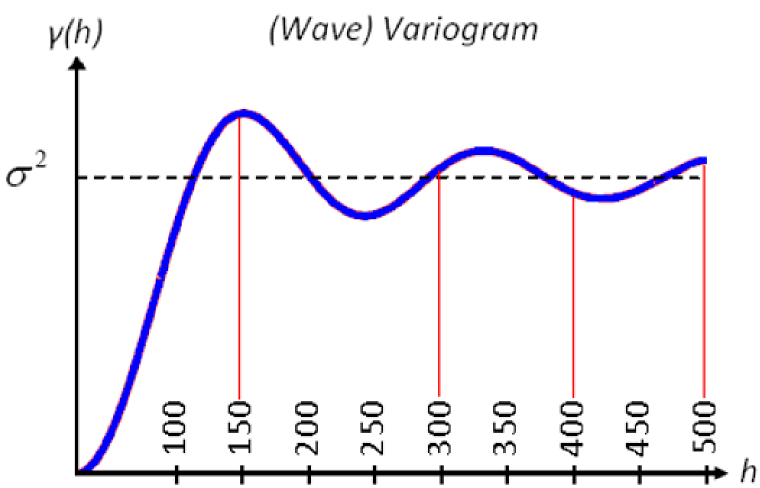


Variogram Models

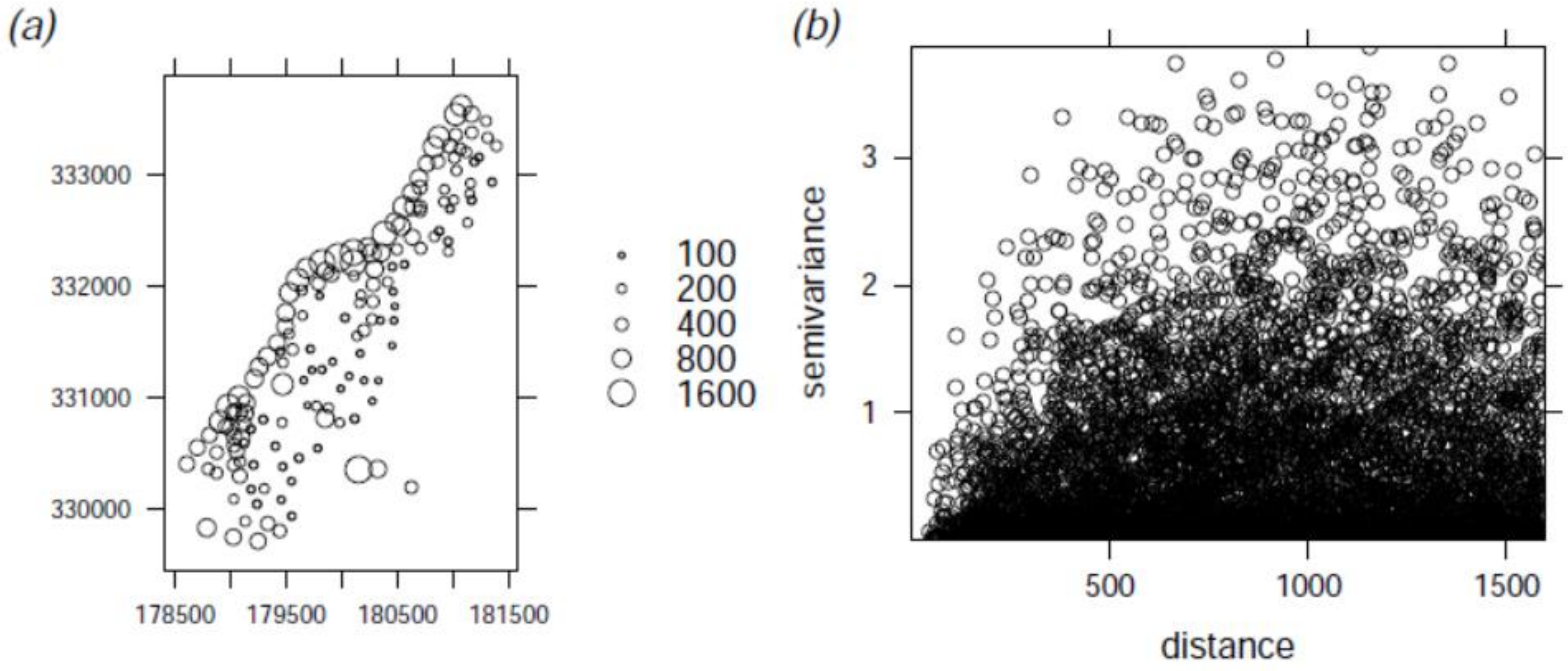
The Wave (Hole-Effect) Model

the waves exhibit a periodic pattern. A non-standard form of spatial autocorrelation applies. Peaks are similar in values to other peaks, and troughs are similar in values to other troughs. However, note the *dampening* in the covariogram and variogram below: That is, *peaks* that are closer together have values that are more correlated than peaks that are father apart (and same holds for troughs).





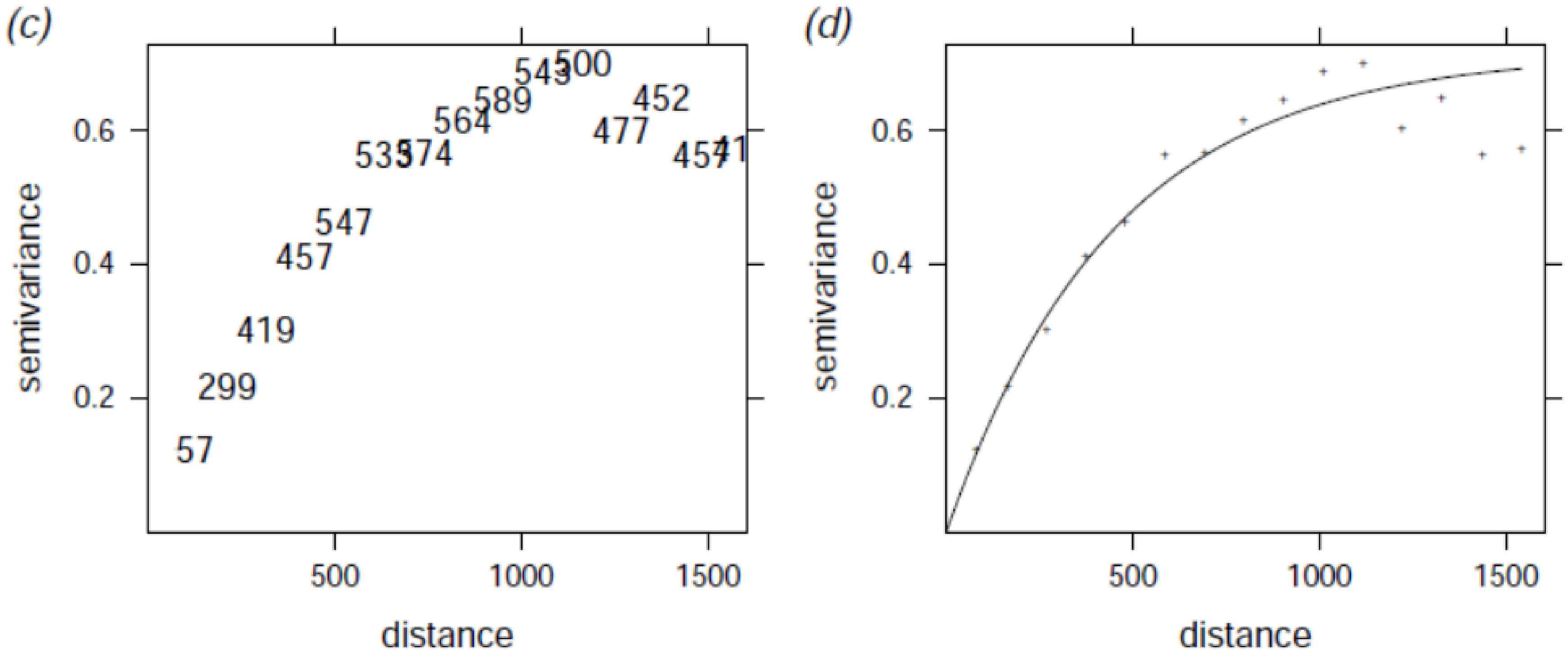
Steps of Variogram Modeling



(a) sampling locations (n=155) and measured variable

(b) variogram cloud showing semivariances for all pairs

Steps of Variogram Modeling (cont'd)



(c) semivariances aggregated to lags of about 100 m

(d) the final variogram model fitting

應用實例介紹

第38卷 第9期
2019年9月

地理科学进展
Progress in Geography

Vol.38, No.9
Sep. 2019

中国县域乡村地域多功能格局及影响因素识别

杨 忍,罗秀丽,陈燕纯

(中山大学地理科学与规划学院/广东省城市化与地理环境空间模拟重点实验室,广州 510275)

摘 要:论文以中国大陆县域为研究单元,构建乡村地域多功能评价指标体系,利用熵权法、半变异函数和地理探测器等模型,对2000—2015年中国乡村多功能指数进行综合测评,揭示中国县域乡村地域多功能的空间分异特征,定量识别乡村地域多功能空间分异的影响因素。研究结果表明:① 中国乡村地域多功能在空间上呈现出平原、东部沿海等地区高值集聚,高原、山地等地区低值集聚,各县域乡村地域多功能整体呈现出逐渐提升态势;② 经济发展功能和社会保障功能对乡村地域多功能的贡献率逐渐增大,农业生产功能和生态保育功能对乡村地域多功能的贡献率逐渐减小;③ 2000—2015年,中国大陆范围内县域乡村地域多功能空间自相关范围和强度总体呈现减小的趋势,随机性因子成为乡村地域多功能空间分异的主要驱动力;④ 县域经济整体发展水平和财政收入是影响乡村地域多功能空间分异的主导因素;各影响因素之间的两两交互作用会增强乡村地域多功能的空间分异;社会环境因素对乡村地域多功能空间分异的影响程度逐渐上升,自然环境因素的影响程度逐渐下降。

关 键 词:地域多功能;乡村发展;半变异函数;地理探测器;中国

農村地域多功能指數

目标层	准则层	指标层
乡村地域多功能	经济发展功能(0.5565)	地区生产总值
		地均生产总值
		财政贡献量
		地均财政贡献量
		产业结构
	农业生产功能(0.2707)	第一产业增加值
		人均第一产业增加值
		人均粮食占有量
		人均肉类占有量
		人均非粮农作物占有量
	生态保育功能(0.0172)	生态脆弱性
		植被覆盖度
	社会保障功能(0.1556)	医疗卫生条件
		社会福利水平
		人均储蓄存款

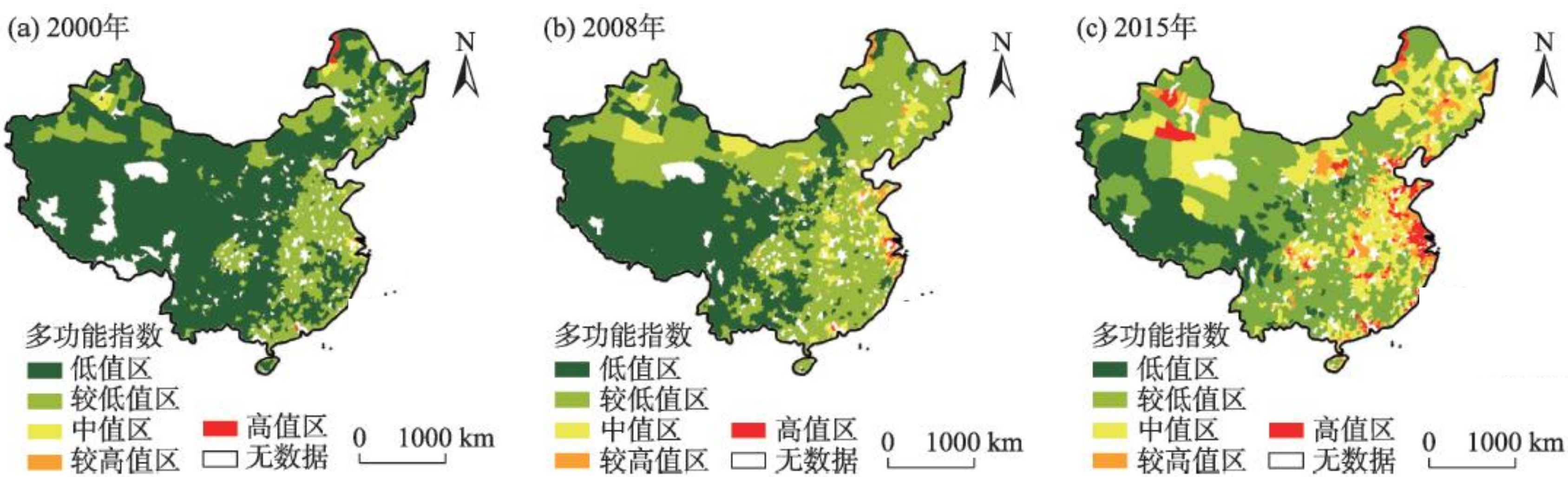


图3 2000—2015年中国乡村地域多功能指数分布

半變異元分析

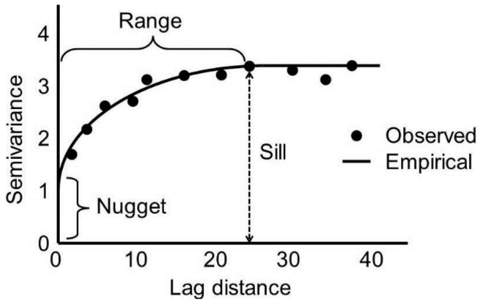


表 3 2000—2015 年乡村地域多功能半变异分析结果

range nugget sill

年份	变程/km	块金值(C_0)	基台值(C_0+C)	块金系数[$C_0/(C_0+C)$]	R^2	残差(*RSS*)
2000	2379	0.0558	0.2136	26.12%	0.9810	0.0007
2008	2395	0.0726	0.2822	25.73%	0.9850	0.0010
2015	2093	0.0981	0.3192	30.73%	0.9740	0.0019

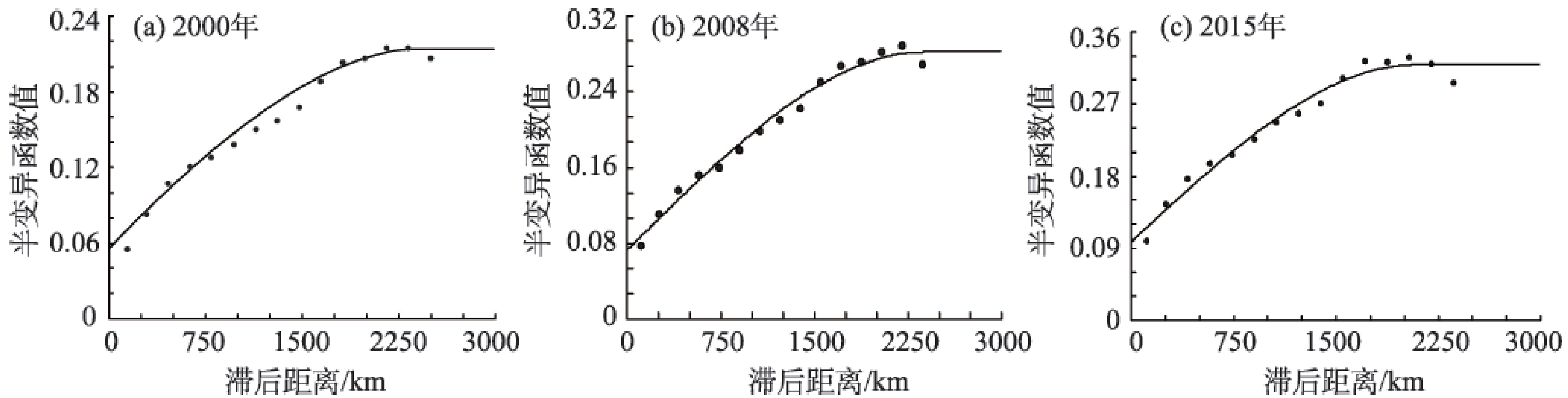


图 6 2000—2015 年乡村地域多功能半变异函数拟合结果

Fig.6 Fitting results of semi-variance function of rural multifunctionality, 2000–2015

Lab: Variogram (Exploring data)

- 安裝R套件 gstat

gstat v1.1-6 Other versions ▾

NaN Monthly downloads › 76th Percentile

by Edzer Pebesma View Source https://www.rdocumentation.org/packages/gstat Copy

Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation

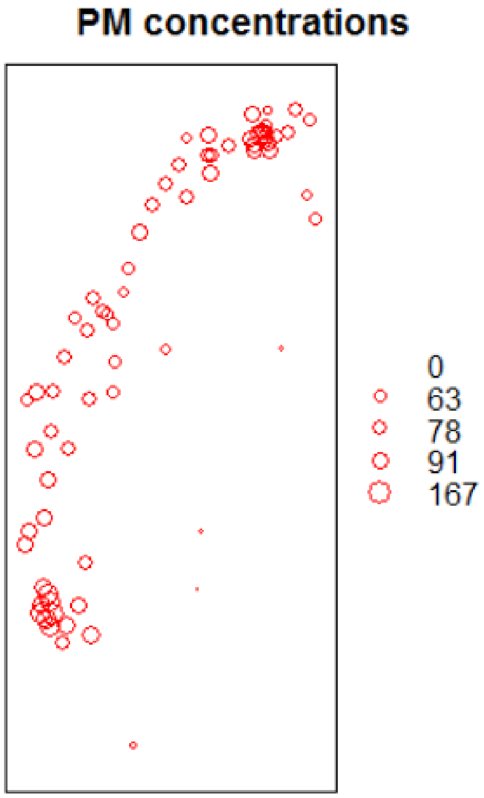
Variogram modelling; simple, ordinary and universal point or block (co)kriging; spatio-temporal kriging; sequential Gaussian or indicator (co)simulation; variogram and variogram map plotting utility functions.

Data: 台灣環保署空氣品質測站資料 (shape file)

73 obs. x 16 variables

	SiteName	SiteEngNam	AreaName	County	Township	SiteAddres	TWD97Lon	TWD97Lat	SiteType	Name	PSI	PM	O3	SO2	CO	NO2
1	二林	Erlin	中部空品區	彰化縣	二林鎮	彰化縣二林鎮萬合里江山巷1號	120.4097	23.92517	一般測站	二林	62	75	40	5.8	0.47	12.0
2	三重	Sanchong	北部空品區	新北市	三重區	新北市三重區三和路重陽路交口	121.4938	25.07261	交通測站	三重	68	102	0	3.0	1.37	36.0
3	三義	Sanyi	竹苗空品區	苗栗縣	三義鄉	苗栗縣三義鄉西湖村上湖61-1號	120.7588	24.38294	一般測站	三義	45	56	32	1.9	0.36	6.3
4	土城	Tucheng	北部空品區	新北市	土城區	新北市土城區學府路一段241號	121.4519	24.98253	一般測站	土城	62	84	30	1.9	0.51	16.0
5	士林	Shilin	北部空品區	臺北市	北投區	臺北市北投區文林北路155號	121.5154	25.10542	一般測站	士林	50	61	32	1.8	0.41	11.0
6	大同	Datong	北部空品區	臺北市	大同區	臺北市大同區重慶北路三段2號	121.5133	25.06320	交通測站	大同	61	78	0	2.3	0.84	20.0
7	大里	Dali	中部空品區	臺中市	大里區	臺中市大里區大新街36號	120.6777	24.09961	一般測站	大里	42	50	37	2.4	0.62	19.0
8	大園	Dayuan	北部空品區	桃園市	大園區	桃園市大園區中正東路160號	121.2018	25.06034	一般測站	大園	62	85	35	3.3	0.37	12.0
9	大寮	Daliao	高屏空品區	高雄市	大寮區	高雄市大寮區潮寮路61號	120.4251	22.56575	一般測站	大寮	72	109	51	8.2	0.91	38.0
10	小港	Xiaogang	高屏空品區	高雄市	小港區	高雄市小港區平和南路185號	120.3377	22.56583	一般測站	小港	86	132	34	7.1	0.72	31.0
11	中山	Zhongshan	北部空品區	臺北市	中山區	臺北市中山區林森北路511號	121.5265	25.06236	一般測站	中山	52	81	14	2.4	1.06	33.0
12	中壢	Zhongli	北部空品區	桃園市	中壢區	桃園市中壢區延平路622號	121.2217	24.95328	交通測站	中壢	67	83	26	2.4	0.91	23.0
13	仁武	Renwu	高屏空品區	高雄市	仁武區	高雄市仁武區八卦里永仁街555號	120.3326	22.68906	一般測站	仁武	91	144	32	2.7	0.70	26.0
14	斗六	Douliu	雲嘉南空品區	雲林縣	斗六市	雲林縣斗六市民生路224號	120.5450	23.71185	一般測站	斗六	59	71	41	2.5	0.49	14.0
15	冬山	Dongshan	宜蘭空品區	宜蘭縣	冬山鄉	宜蘭縣冬山鄉南興村照安路26號	121.7929	24.63220	一般測站	冬山	50	49	29	2.1	0.33	12.0
16	古亭	Guting	北部空品區	臺北市	大安區	臺北市大安區羅斯福路三段153號	121.5296	25.02061	一般測站	古亭	45	87	21	0.0	0.67	24.0
17	左營	Zuoying	高屏空品區	高雄市	左營區	高雄市左營區翠華路687號	120.2929	22.67486	一般測站	左營	81	117	40	2.5	0.59	17.0
18	平鎮	Pingzhen	北部空品區	桃園市	平鎮區	桃園市平鎮區文化街189號	121.2040	24.95279	一般測站	平鎮	60	71	31	1.6	0.44	12.0

R code: loading data



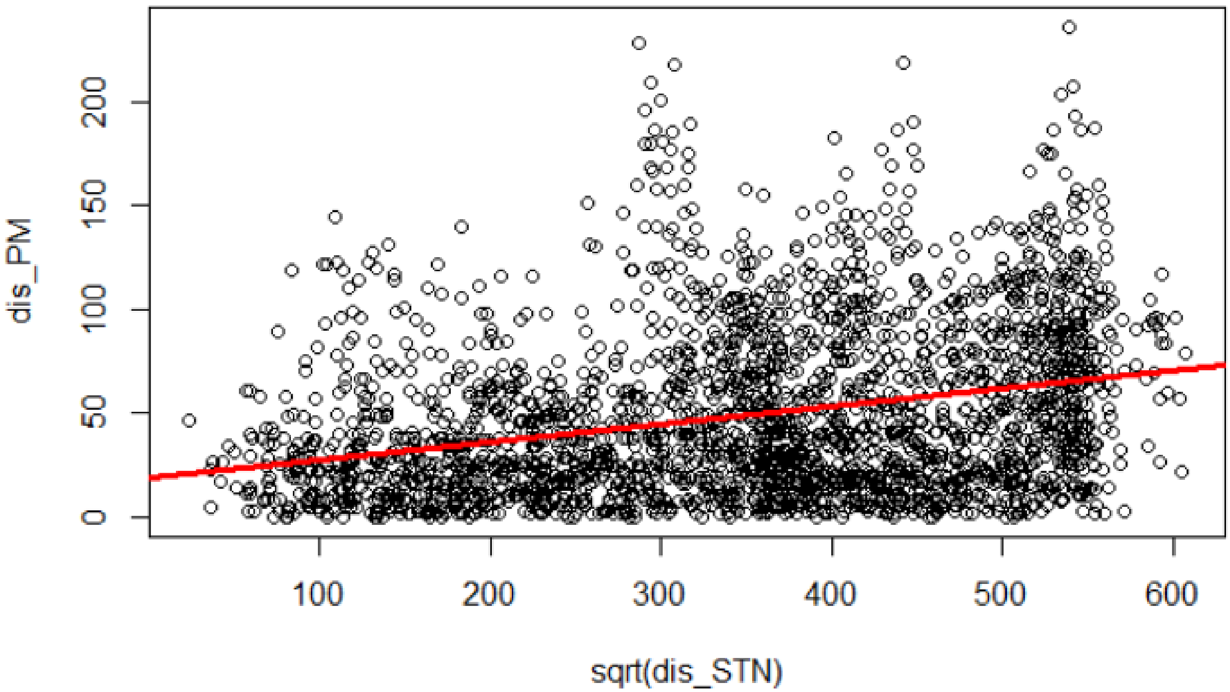

```r
library(rgdal)

rm(list=ls())
setwd("C:/Wen_Files/SA_Labs")
EPA_STN <- readOGR(dsn = "./data", layer = "EPA_STN1", encoding="utf-8")
plot(EPA_STN); head(EPA_STN)

data= EPA_STN@data
PMdata=EPA_STN@data["PM"]

bubble(EPA_STN, "PM", col="red", fill=FALSE, maxsize = 1.5, main = "PM concentrations")
```

Exploring distance vs. variance

$$2\gamma(h) = average\left[\left(Z(i) - Z(j)\right)^2\right]$$

Dist.(h) vs. $[z(x+h)-z(x)]^2$

variogram cloud



Exploring distance vs. variance (R code)

$$2\gamma(h) = average\left[\left(Z(i) - Z(j)\right)^2\right]$$

Dist.(h) vs. $[z(x+h)-z(x)]^2$

```
x= coordinates(EPA_STN)[,1]
y= coordinates(EPA_STN)[,2]

STNDF = cbind(x,y)
dis_STN= dist(STNDF)

pm= EPA_STN@data[,12]

PMDF= cbind(pm,pm)
dis_PM = dist(PMDF)

plot(dis_PM~sqrt(dis_STN))
abline(lm(dis_PM~sqrt(dis_STN)), lwd=3, col='red')
```

Using variogram() function in R

variogram {gstat} R Documentation

Calculate Sample or Residual Variogram or Variogram Cloud

Description

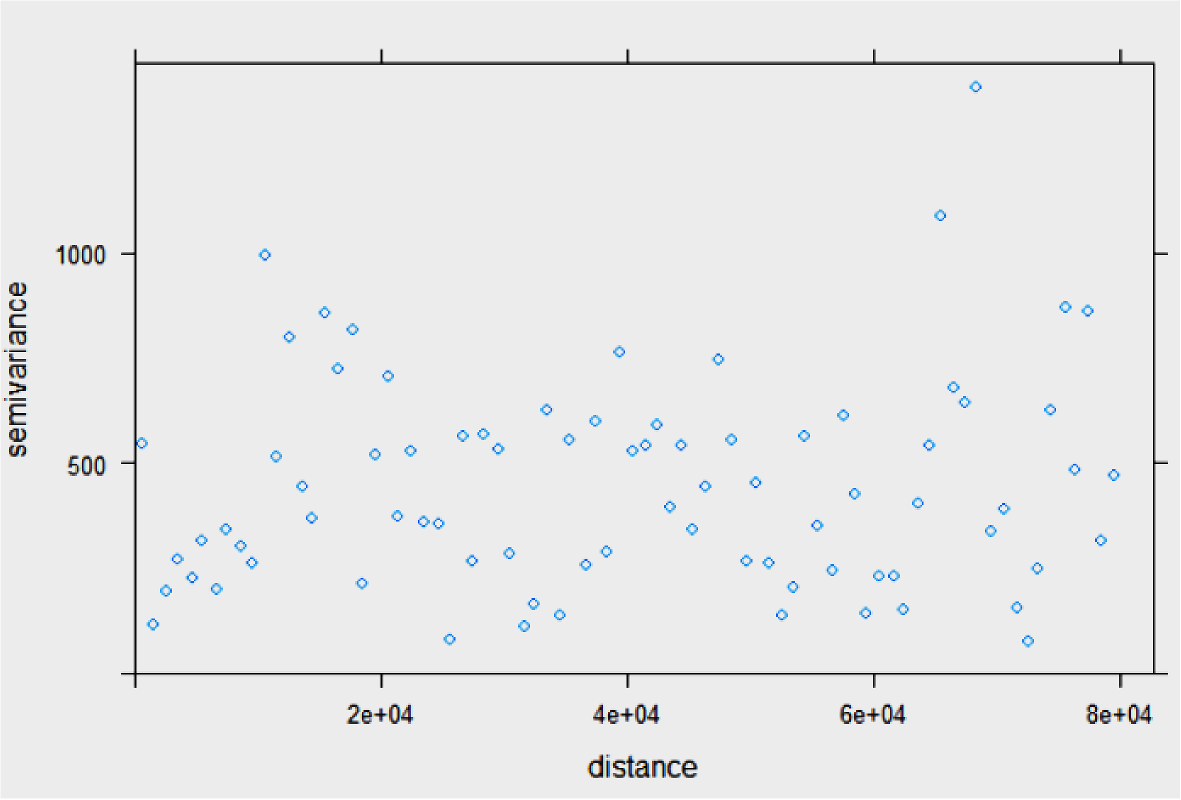
Calculates the sample variogram from data, or in case of a linear model is given, for the residuals, with options for directional, robust, and pooled variogram, and for irregular distance intervals.

In case spatio-temporal data is provided, the function `variogramST` is called with a different set of parameters.

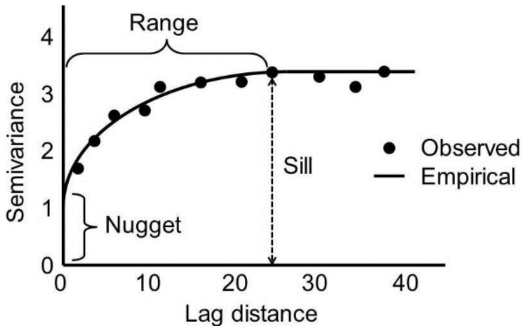
```
library(gstat)
pm.vgm = variogram(PM~1, EPA_STN,cutoff=80000, width=1000)
```

Using variogram() function in R

```
library(gstat)
pm.vgm = variogram(PM~1, EPA_STN,cutoff=80000, width=1000)
plot(pm.vgm)
```



Fitting a Variogram Model



fit.variogram {gstat}

Fit a Variogram Model to a Sample Variogram

Description

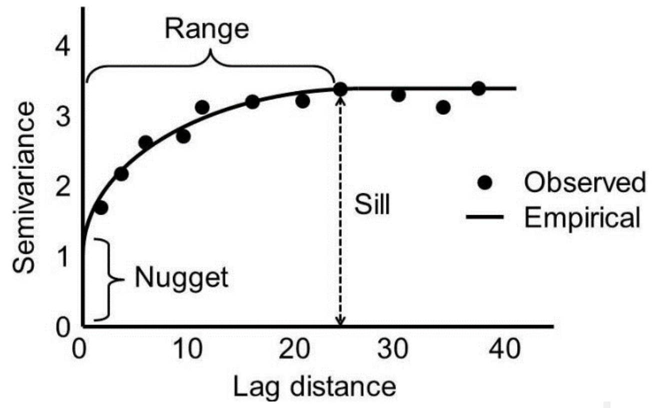
Fit ranges and/or sills from a simple or nested variogram model to a sample variogram

Usage

```
fit.variogram(object, model, fit.sills = TRUE, fit.ranges = TRUE,
        fit.method = 7, debug.level = 1, warn.if.neg = FALSE, fit.kappa = FALSE)
```

Sill, func. range, nugget

```
pm.fit = fit.variogram(pm.vgm, model = vgm(1000, "Exp",2000,1) )
```



vgm {gstat}

Generate, or Add to Variogram Model

Description

Generates a variogram model, or adds to an existing model. `print.variogramModel` prints the essence of a variogram model.

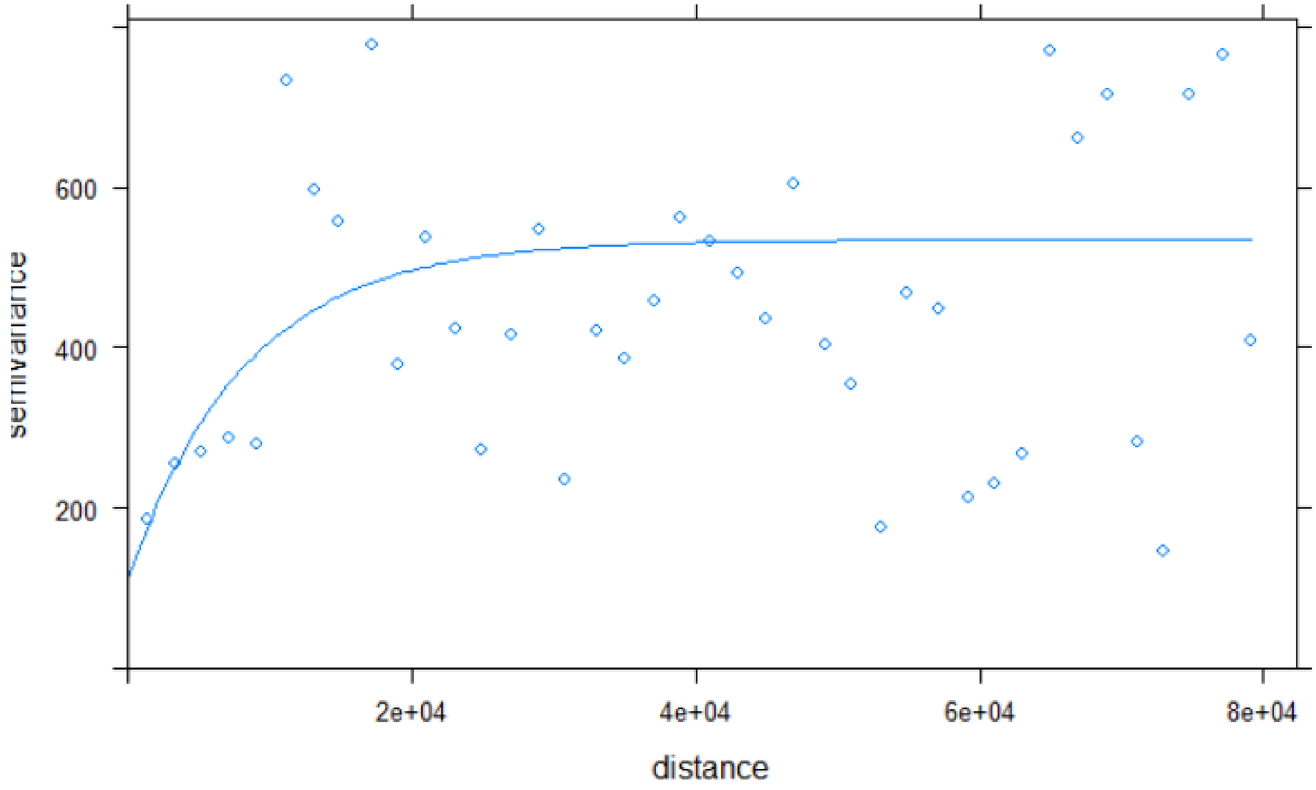
Usage

```
vgm(psill = NA, model, range = NA, nugget, add.to, anis, kappa = 0.5, ..., covtabl
        Err = 0)
## S3 method for class 'variogramModel'
print(x, ...)
as.vgm.variomodel(m)
```

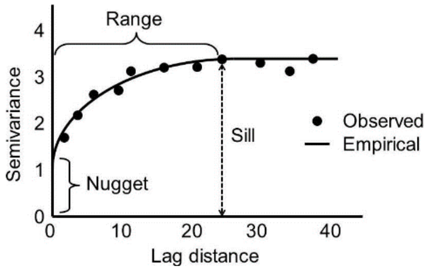
Sill, func. range, nugget

```
pm.fit = fit.variogram(pm.vgm, model = vgm(1000, "Exp",2000,1) )
```

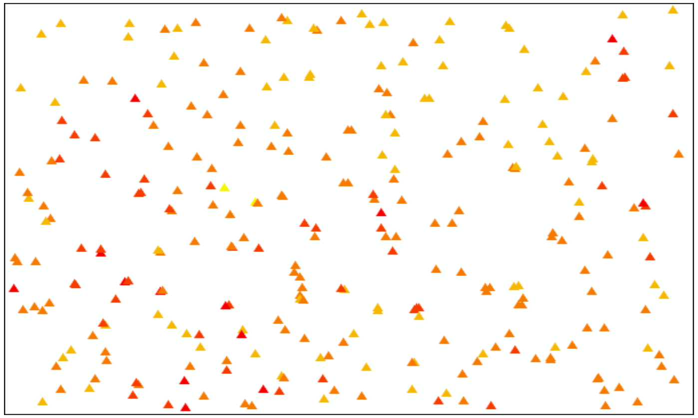
```
library(gstat)
pm.vgm = variogram(PM~1, EPA_STN,cutoff=80000, width=2000)
pm.fit = fit.variogram(pm.vgm, model = vgm(1000, "Exp",20000,1) )
plot(pm.vgm,pm.fit)
```

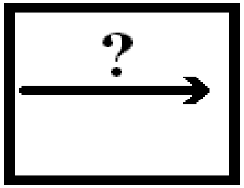


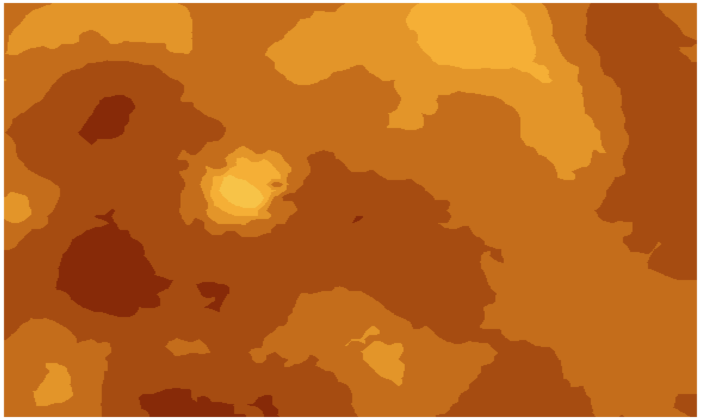
Geostatistical Approach to Spatial Interpolation: using semivariogram



Input *Process* *Output*







期末考重點：空間相依的特性

- Spatial autocorrelation: Moran’s I index
  - Spatial weighting matrix
  - Monte-Carlo significance test
  - Moran scatter plot and correlograms
- Local Moran (LISA) and Gi*(d) statistics
- Semivariogram analysis