MORE EFFICIENT AND ACCURATE METHODS FOR USING REMOTE SENSING

R. Benedetti, P. Postiglione
University of Chieti-Pescara
Federica Piersimoni
ISTAT

Introduction

- Remote sensing is an important tool in the study of natural resources and environment.
- Remote sensing is defined as the technique of deriving information about the earth surface features and the estimation of their geo-bio-physical properties using electromagnetic radiation as a medium of interaction. This acquisition is without a physically contact with earth.
- Remote sensing techniques are widely used in agriculture and agronomy.
- As well-stated by Carfagna and Gallego (2005), the spectral response and the identification of crops are not in one-to-one correspondence.
- To solve this problem, the literature has explored some methods for the correct identification of the crops. We refer to this group of techniques as classification methods.

1. **Unsupervised classification divides the image into unknown classes. Unsupervised classifications seek to group pixels having similar spectral reflective characteristics into distinct clusters.**
2. Supervised classification uses a set of user-defined spectral signatures to classify an image. The spectral signatures are derived from training areas (or sets) that are created by depicting features of interest on an image.

<u>The main difference between the two approaches is that in unsupervised classification, the classes need not be defined a priori.</u>

Introduction

Remote sensing techniques may represent a suitable tool for particular problems in agricultural survey as, for example:

- reliability of data;
- incomplete sample frame and sample size;
- methods of units' selection;
- measurement of area, non sampling errors;
- gap in geographical coverage;
- non availability of statistics at disaggregated level.

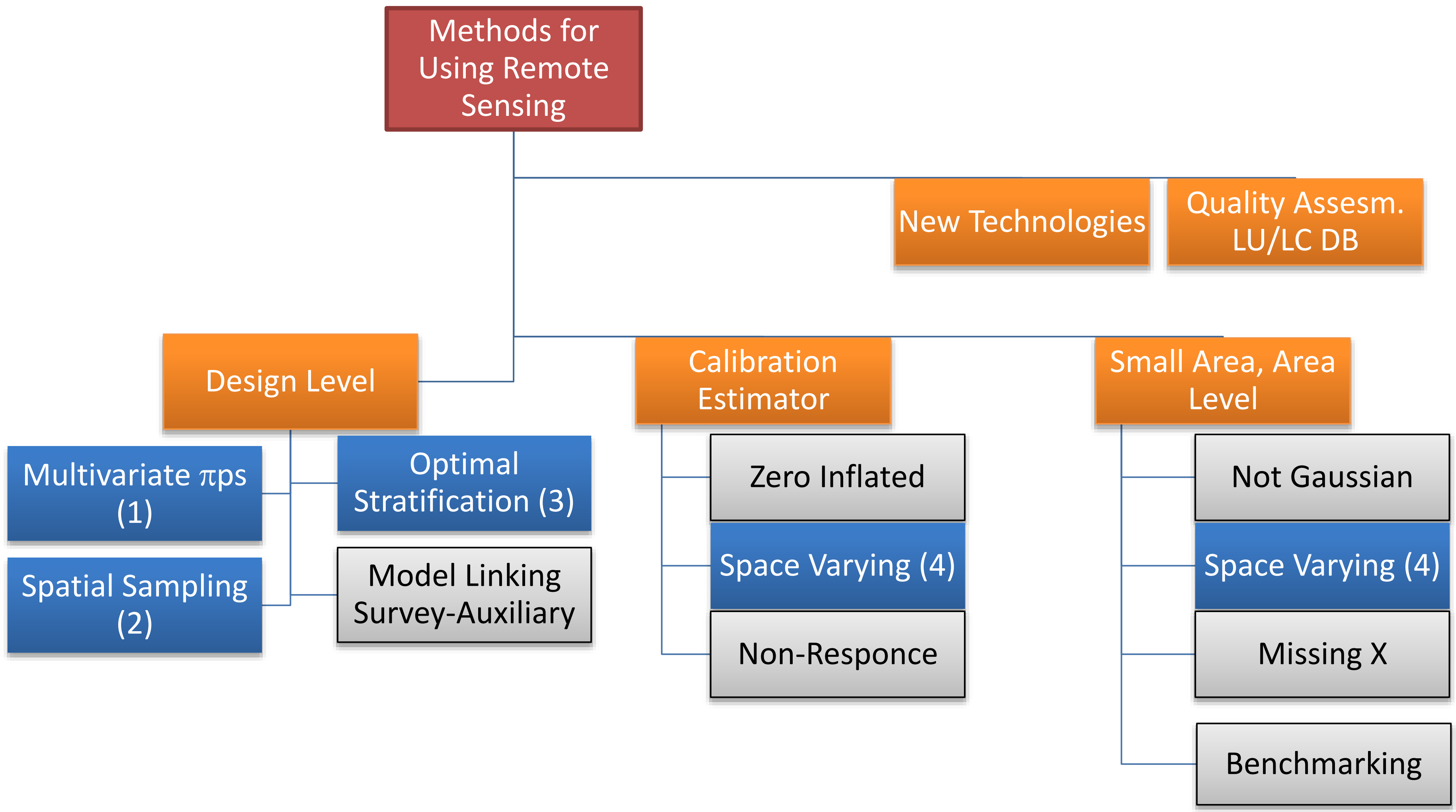
remote sensing technologies on the different phases of sampling spatial statistical units:

- *ex- ante* use (i.e., at design level)
- *ex-post use* (i.e., at estimation level)

In the case of auxiliary information, two specific issues must be discussed (Benedetti et al. 2015):

- the definition of the statistical units is not unique;
- a rich set of auxiliary variables, other than dimensional variables, is available: consider, in our case, the information provided by airplane or satellite remote sensing.

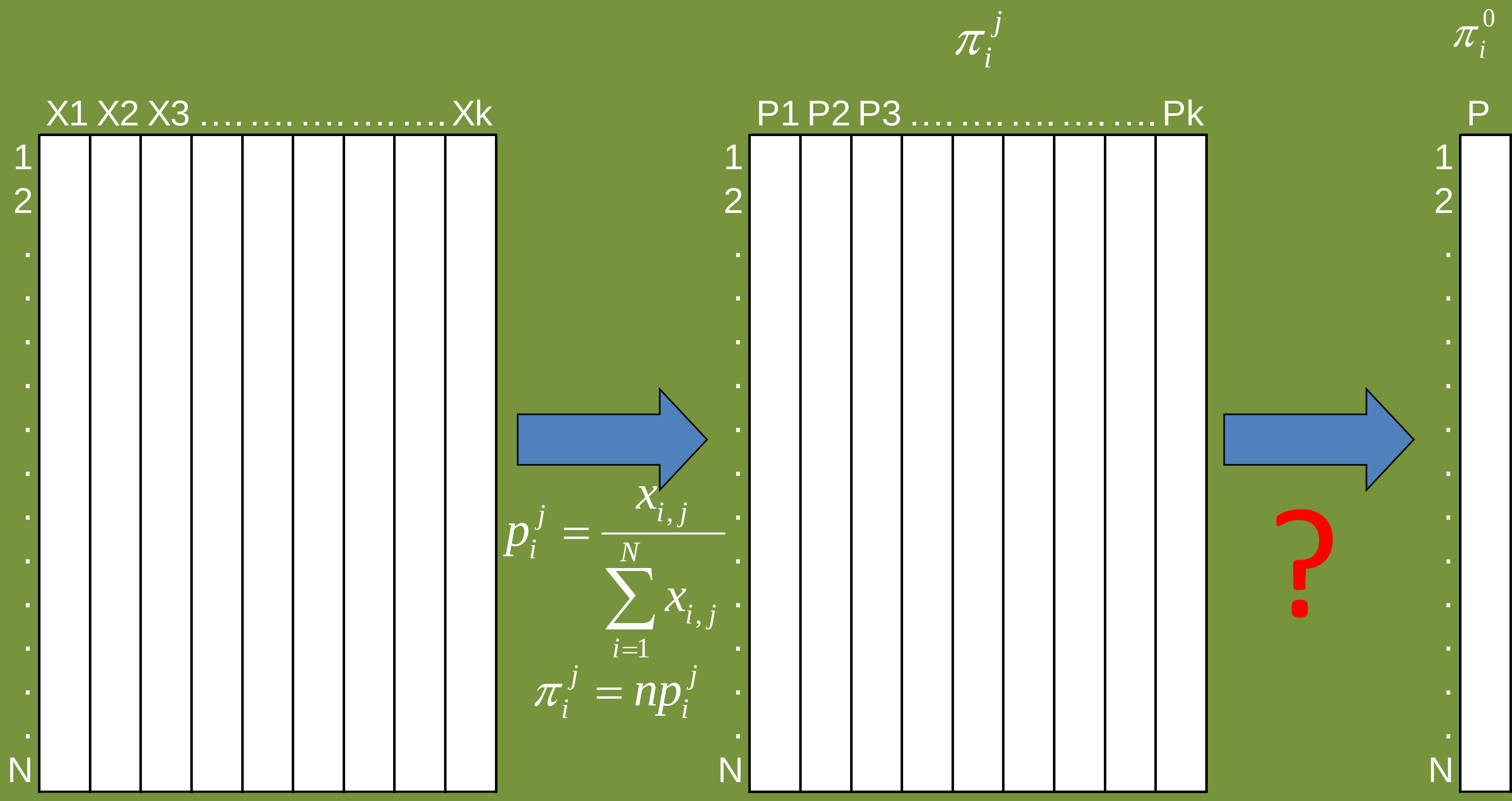
DESCRIPTION OF THE RESEARCH TOPIC



Summary of Activity & Benefits

Research Activity / Methods	What is currently used	What can be done	Benefits
Methods at the design level			
Multivariate ops	Maximal Brewer	Optimal Linear Combination	theoretical properties - flexibility more reasonable distribution of the weights
Optimal stratification	Sequence of thresholds chosen by eye or by trial and error	Optimal Algorithm given H	Considerable reduction of variance (more than 60% depending on the auxiliary)
Spatially balanced samples	Systematic when possible, or rules that rarely allow an efficient use of HT (one unit per stratum)	Spatially Balanced Sampling, in particular samples with prob. prop. To the within distance	theoretical properties - flexibility Considerable reduction of variance (0-100% on simulated data, and more than 50% on real data depending on the homogeneity of the survey variable)
Calibration estimators			
Non response	RHG on original strata	Calibration	theoretical properties - flexibility Considerable reduction of variance depending on the aux (Remote Sensing)
Models for space varying coefficients	???	Partition (post-stratification) on Calibration models	Considerable reduction of variance
Zero inflated data	???	Model Calibration	better fit of the models to the auxiliary data
Small Area			
Area Level Models	The theory exists but the operational use is very limited in spatial agricultural surveys	Simply a FH model	Promising results when used on a point frame (correlation at an area level, avoid geom. Errors)

(1) MULTIVARIATE ΠPS



$$\pi_i^j$$

$$\pi_i^0$$

$$p_i^j = \frac{x_{i,j}}{\sum_{i=1}^{N} x_{i,j}}$$

$$\pi_i^j = np_i^j$$

We Assume

- $0 \le \pi_i^j \le 1$
- for at least one *j*, $\pi_i^j > 0$ (otherwise the unit is outside the frame)
- for at least one *j*, $\pi_i^j < 1$ (otherwise the problem is trivial)

We Are Looking For

$$p_i^0 = \sum_{j=1}^{k} p_i^j \alpha_j \qquad \pi_i^0 = \sum_{j=1}^{k} \pi_i^j \alpha_j$$ (α_j unknown)

Where $\sum_{j=1}^{k} \alpha_j = 1$ and $\alpha_j \in [0;1]$

in such a way that $\sum_{i=1}^{N} \pi_i^0 = n$

(2) Spatial Sampling, Why ?

- **A - *The decomposition lemma*, states that (Knottnerus, 2003, p. 87):**

$$\sigma_z^2 = Var\left(\bar{z}_s\right) + \frac{n-1}{n} E\left(\sigma v^2_{z,s}\right)$$

- **B - Yates-Grundy-Sen HT variance:** $V\left(\hat{Y}_{HT}\right) = -\frac{1}{2}\sum_{i\in U}\sum_{j\in U}\left(\pi_{ij} - \pi_i\pi_j\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$

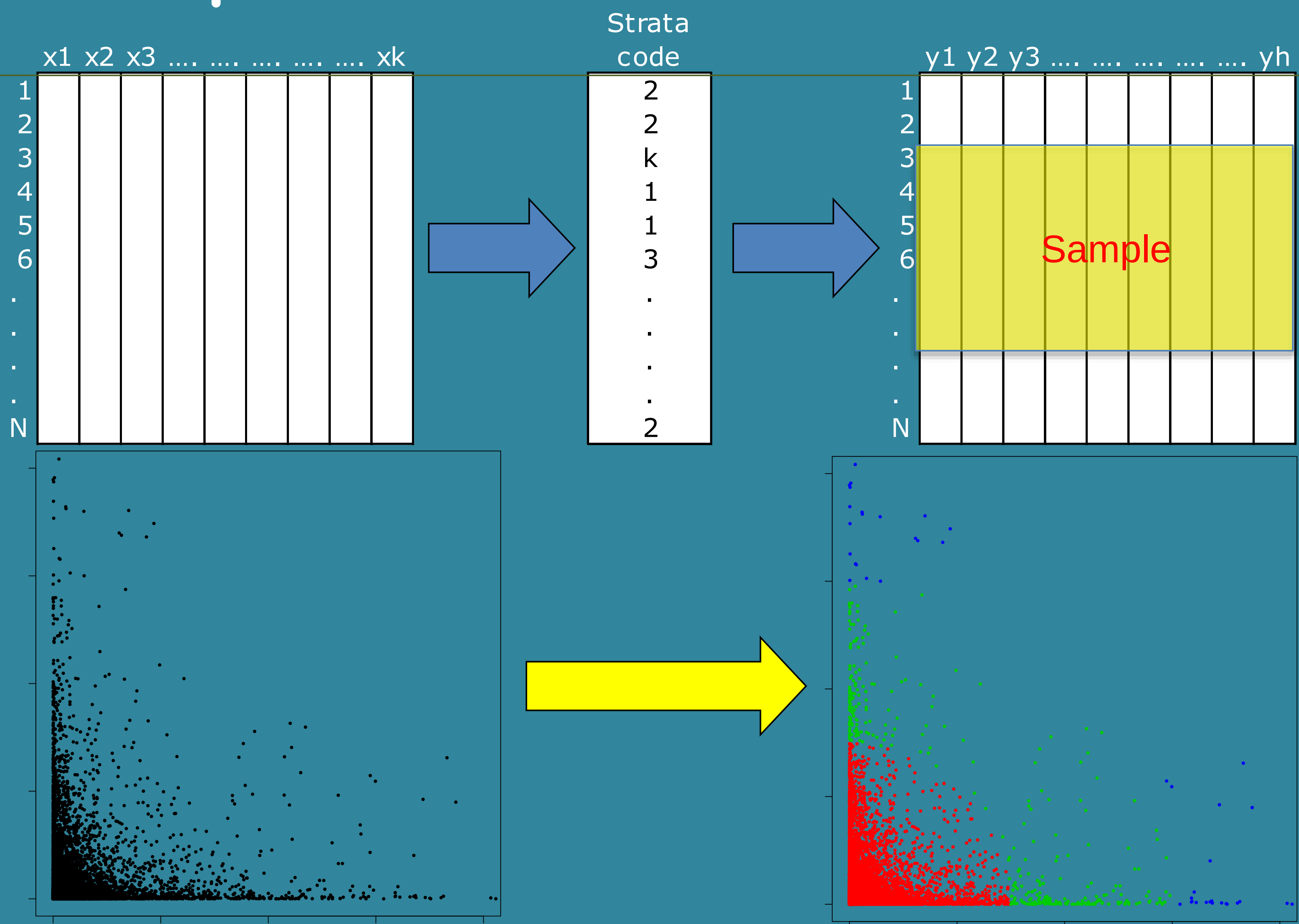
- **C – Anticipated Variance (Grafström and Tillé, 2013):**

$$AV\left(\hat{t}_{HT} - t\right) = E_s\left[\left(\sum_{k\in s}\frac{x_k}{\pi_k} - \sum_{k\in U} x_k\right)^T \beta\right]^2 + \sum_{k\in U}\sum_{l\in U}\sigma_k\sigma_l\rho_{kl}\frac{\pi_{kl} - \pi_k\pi_l}{\pi_k\pi_l}$$

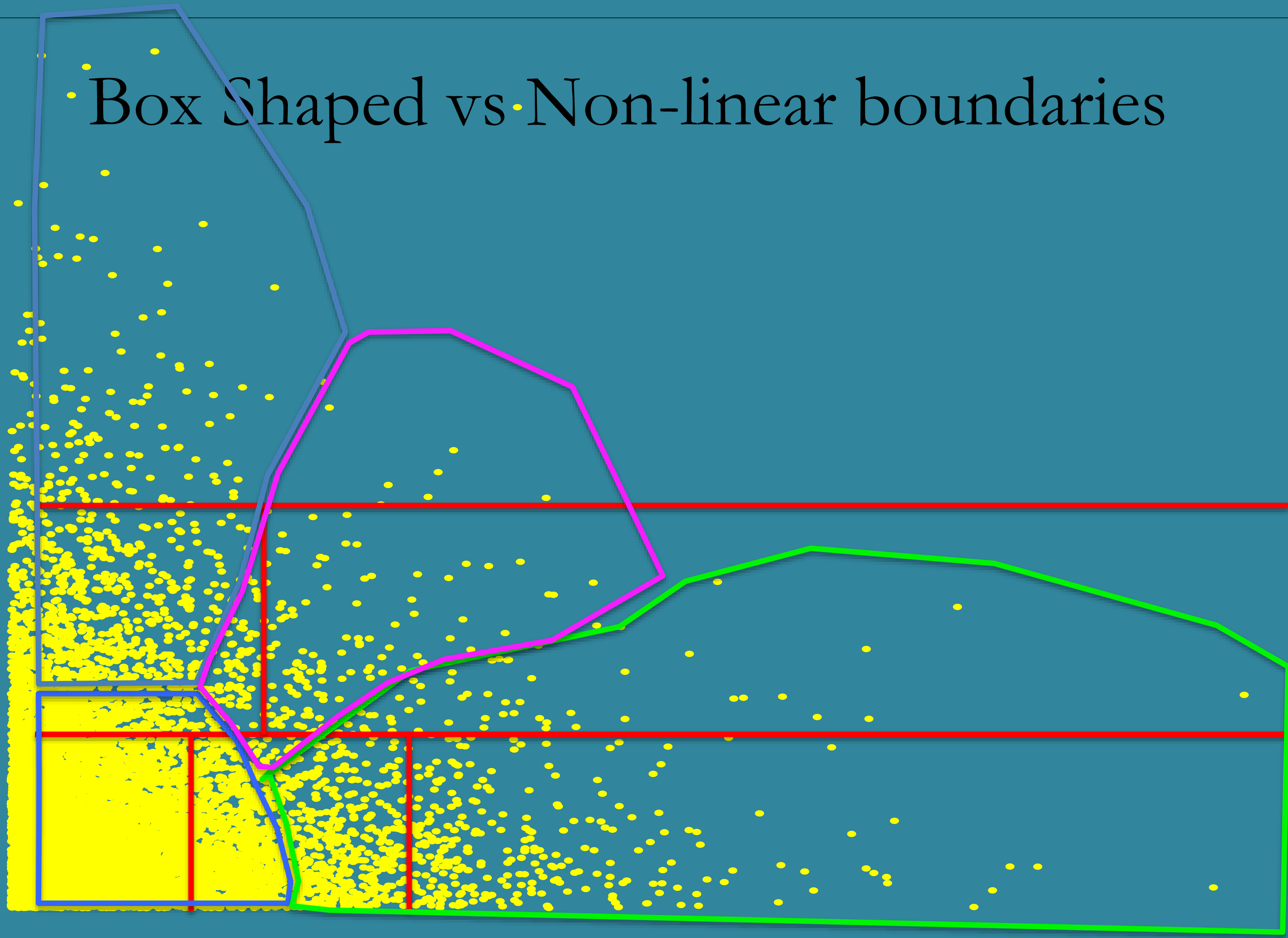
There could be a lot of different reasons why it is appropriate to select samples which are spatially well distributed:

1. **y has a linear or monotone spatial trend;**
2. **there is spatial autocorrelation, i.e. close units have data more similar than distant units;**
3. **the y shows to follow zones of local stationarity of the mean and/or of the variance, i.e. a spatial stratification exists in observed phenomenon;**
4. **the units of the population have a spatial pattern which can be clustered, i.e. the intensity of the units varies across the study region.**

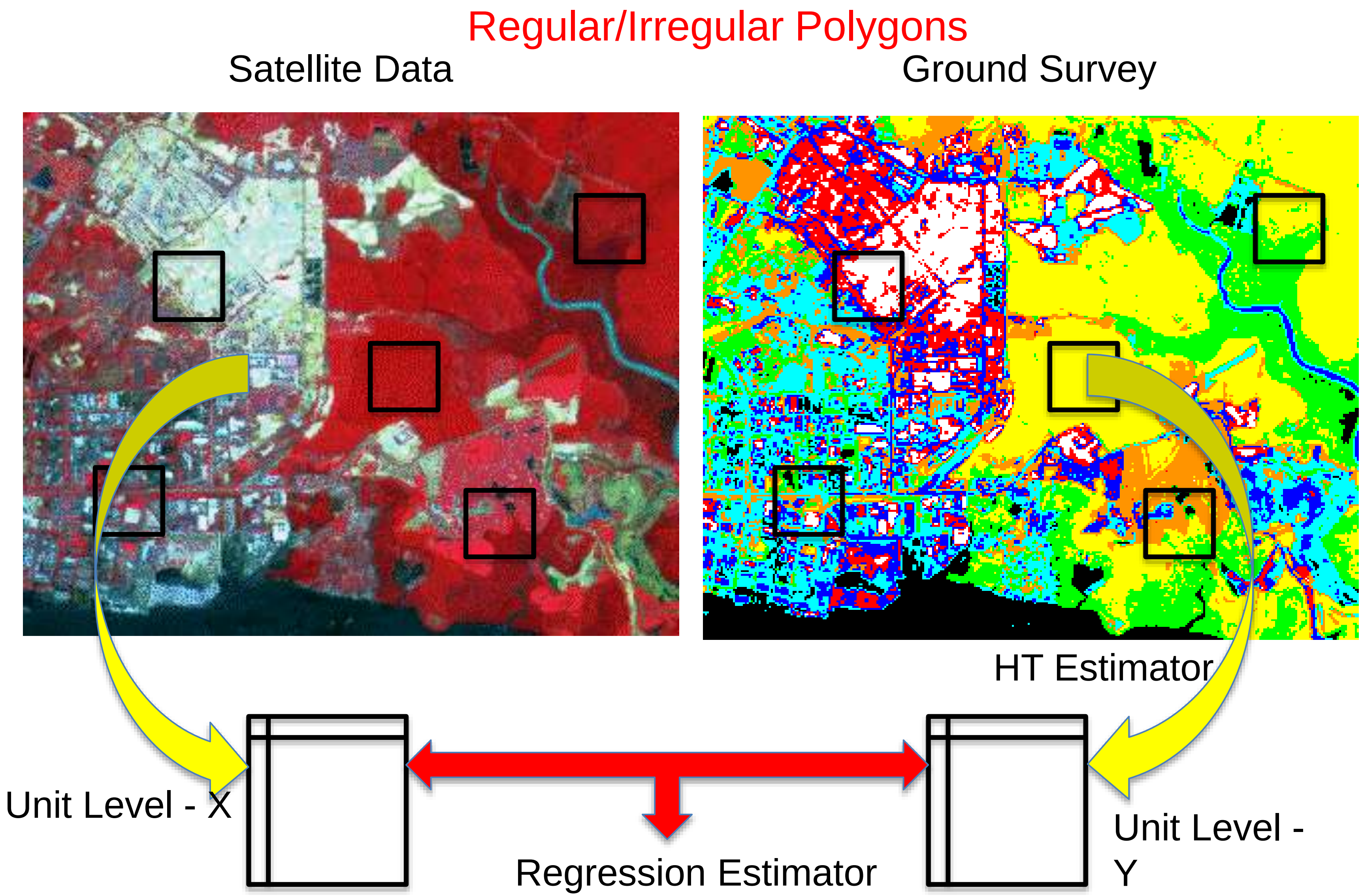
(3) Optimal Stratification



Box Shaped vs Non-linear boundaries



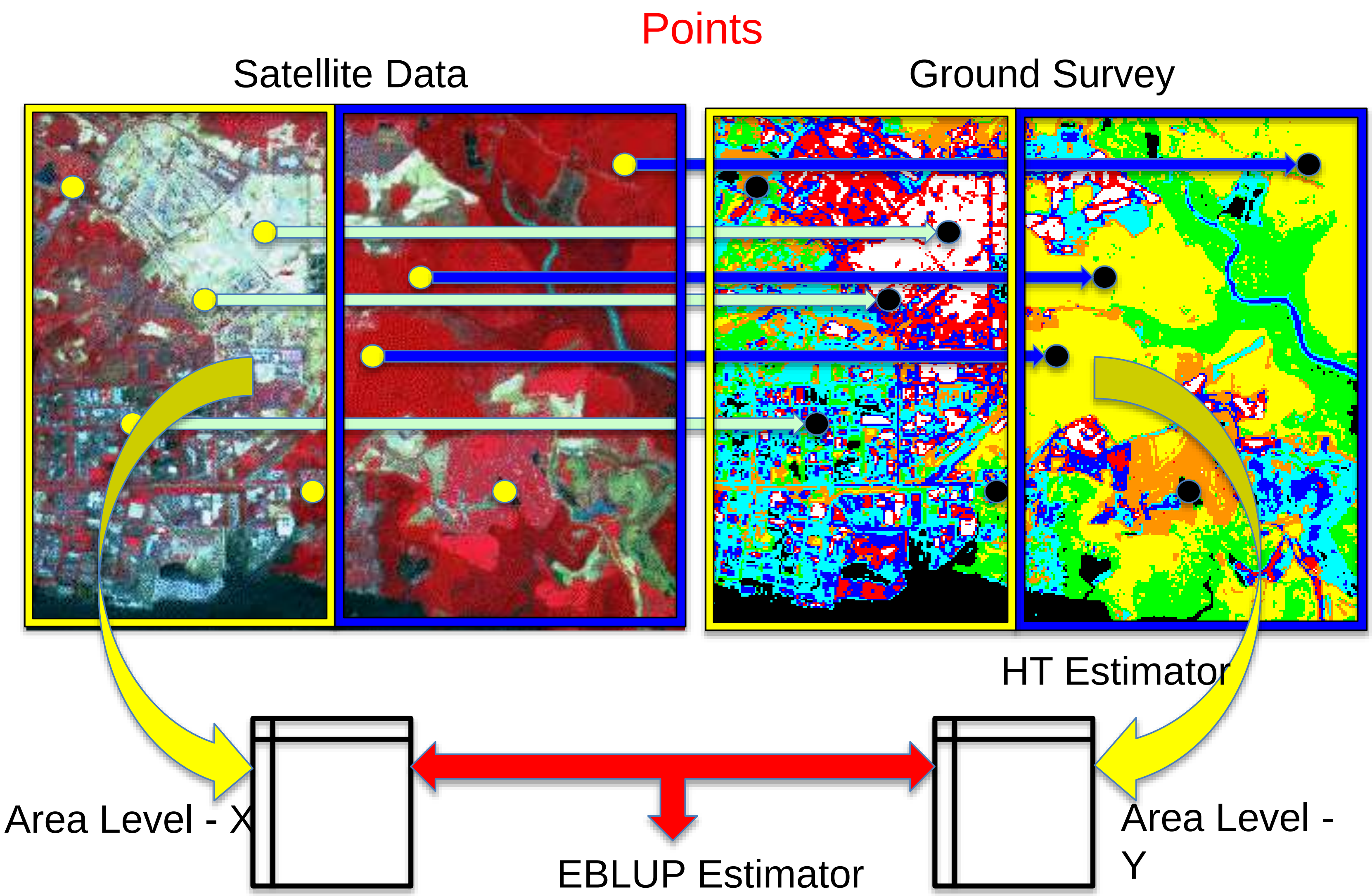
Extension of the regression or calibration estimators



Extension of the regression or calibration estimators

- Survey statisticians place considerable efforts in the design of their surveys, to be able to use the auxiliary information for producing precise and reliable estimates.
- The class of calibration estimators is an example of a very general and practical approach to incorporating auxiliary information into the estimation. These estimators are used in most surveys performed by the major National Statistical Institutes (NSIs).
- The class of calibration estimators is an instance of very general and practical approach to incorporating auxiliary information, represented by remote sensed data, into the estimation.
- The technique of estimation by calibration was introduced by Deville and Särndal (1992). The idea behind is to use auxiliary information to obtain new sampling weights, called calibration weights that make the estimates agree with known totals. The estimates are generally design consistent and with smaller variance than the HT estimator.
- The estimators are obtained through a minimization of an appropriate distance function.
- A limitation of the calibration estimator is that it relies on an implicit linear relationship between the variable **y** under study and the auxiliary variables **X**. Model Calibration

Small area estimators



Small area estimators

A basic area level model that uses area level covariates has two components:

- A direct estimate of the target parameter (e.g. the mean)
- A linking model that connects the direct estimates to the auxiliary variable

The Fay-Herriot (1979) small area model is then

$$t_i = X_i^t \beta + v_i + \epsilon_i$$

Finally, the EBLUP of T_i is

$$t_i^{EBLUP} = \hat{g}_i \tilde{t}_i + (1 - \hat{g}_i) X_i^t \hat{b}$$

where

$$\hat{g}_i = \frac{s_v^2}{s_v^2 + y_i}, \quad \hat{b} = \left(\sum_i \hat{g}_i X_i X_i^t \right)^{-1} \left(\sum_i \hat{g}_i X_i t_i \right)$$

Small area estimators

Zero Inflated models

In several agricultural surveys, it is very common for the response variable to contain a substantial proportion of zeros.

In Zero-Inflated (ZI) models, the response variable is modeled as a mixture of distributions:

$$\begin{cases} 0 & \text{with probability } 1-p_k \\ f(y_k / \mathbf{x}_k) & \text{with probability } p_k \end{cases}$$

Benchmarking the estimators

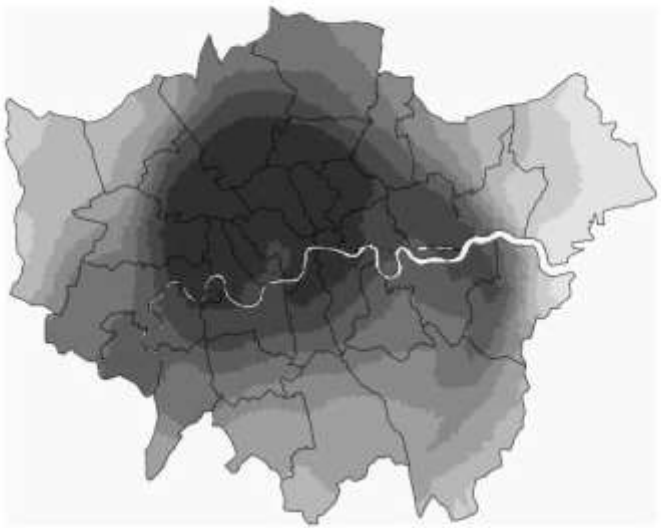
- When a reliable direct estimate is available for an aggregate of small areas, the small area estimates must be consistent with the direct estimate for the larger area.

$$\sum_{j=1}^{D} g_j \hat{\theta}_{j,model} = \sum_{j=1}^{D} g_j \hat{\theta}_{j,design}$$

- One benchmarking method that is often used in practical applications is (Pfeffermann 2013)

$$\hat{\theta}_{d,1}^{\text{bench}} = \left(\frac{\sum_{j=1}^{D} g_j \hat{\theta}_{j,design}}{\sum_{j=1}^{D} g_j \hat{\theta}_{j,model}} \right) \times \hat{\theta}_{d,model}$$

(4) Space Varying, How ?



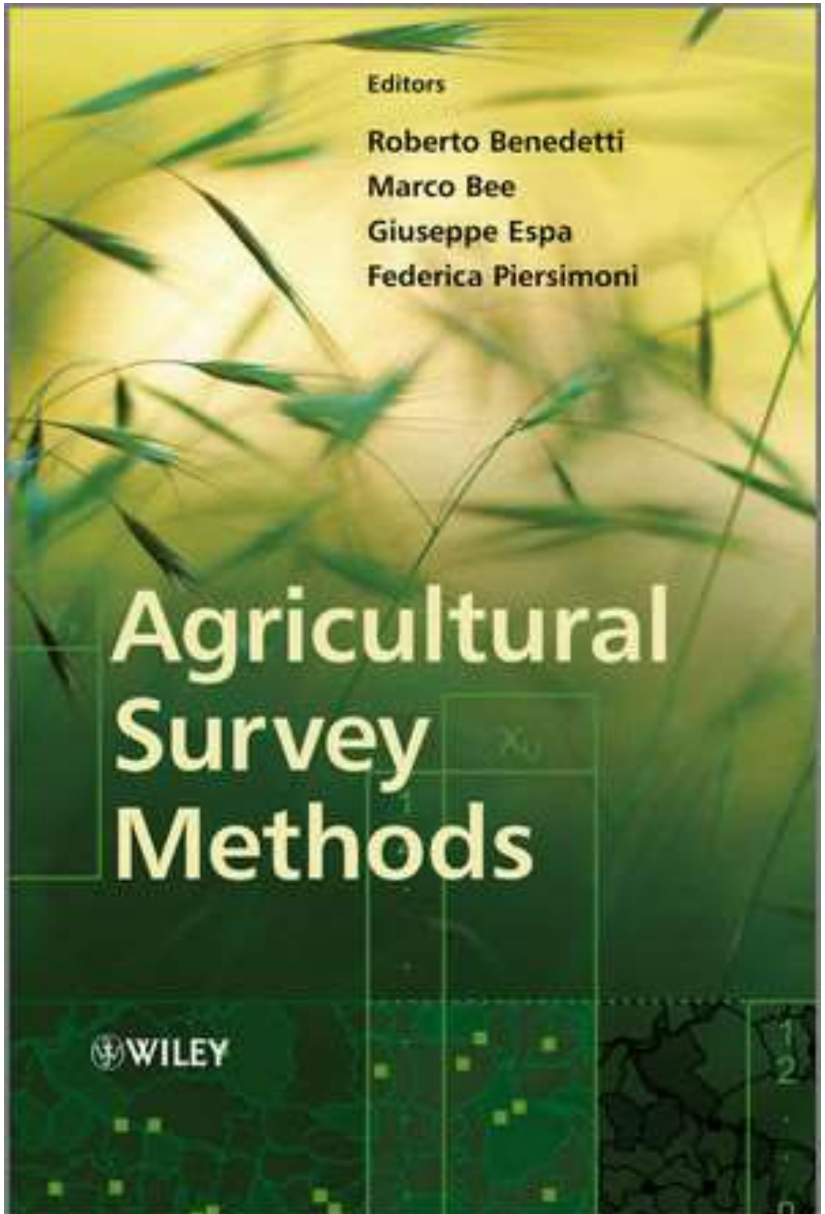
Parameters of any model in a model-based (Small Area) or assisted (calibration) framework

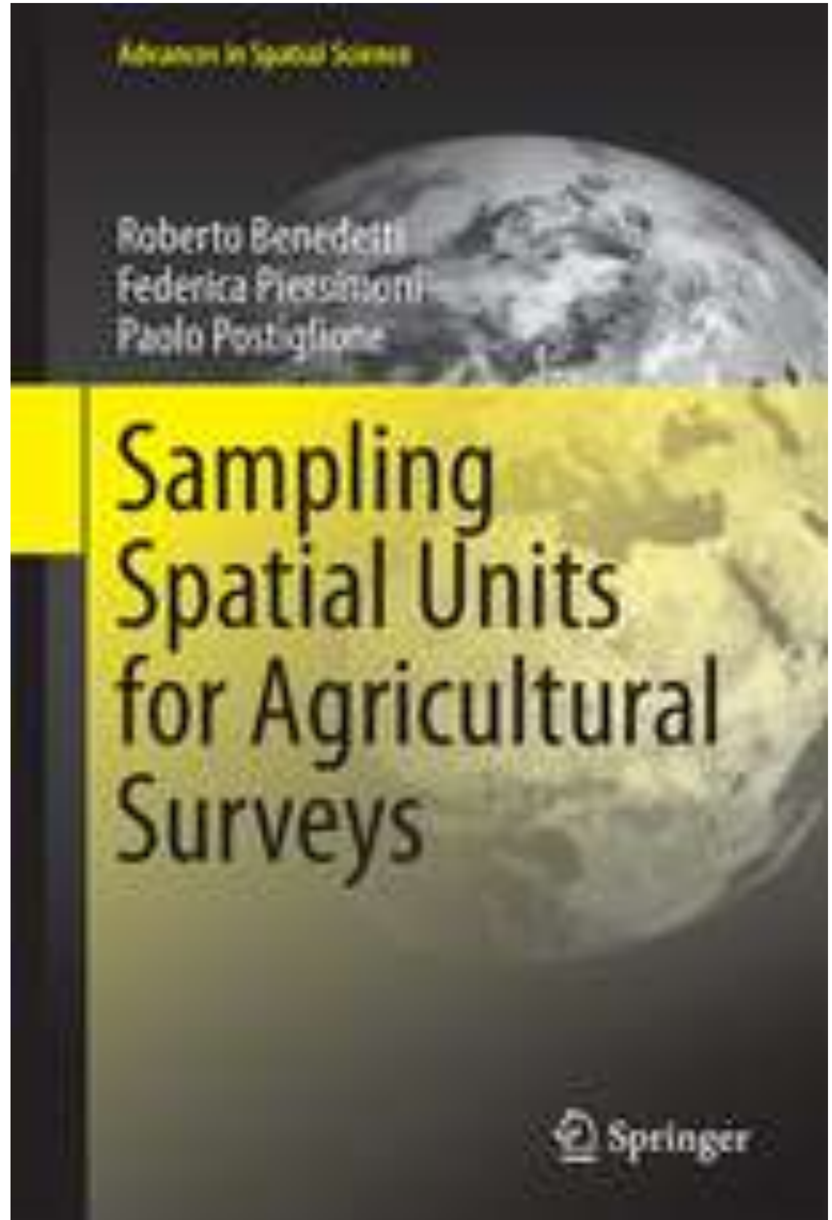
Geographically Weighted Regressio (GWR)

Smooth surface of parameters

- Local stationarity, say *k* strata
- Post-Stratification (endogenous)
- Algorithms: Iterated (GWR) or
- Simulated Annealing

References


Editors
Roberto Benedetti
Marco Bee
Giuseppe Espa
Federica Piersimoni
Agricultural Survey Methods
WILEY


Roberto Benedetti
Federica Piersimoni
Paolo Postiglione
Sampling Spatial Units for Agricultural Surveys
Springer

Concluding remarks

- We represent a further step in assessing the possibility of using remotely sensed images in sampling design and estimation.
- Remote sensing provides information that is available for several countries at a certain spatial and temporal resolution.
- Applicability depends essentially on the availability of images, which may not always be satisfied in developing countries.
- The new satellite technology will provide images that will solve this problem in the very near future.

Thanks for your attention !!!!!

Roberto Benedetti benedett@unich.it

Federica Piersimoni piersimo@istat.it

Paolo Postiglione postigli@unich.it