Sampling of anomalies within LPIS:

towards the combination of ecological methodologies and GIS technologies

MARS Unit, JRC Ispra

Why to develop a method to sample anomalies within LPIS?

Regarding the budget yearly dedicated to 1st pillar subsidies

- **EC and MS could decide to verify conformity of the payments made to farmers according to the eligible reference areas they declared within LPIS**
- **MS could plan to assess the spatial correctness of their LPIS to perform more accurate risk analysis in the frame of CwRS**
- **EC, MS and contractors could decide to determine relative individual liability in case of ineffective and/or inappropriate functioning of LPIS**

⇒ **LPIS quality assessment is necessary**

if LPIS quality assessment is globally addressed when developing LPIS Core Model, in situ LPIS quality needs an accurate quantification of LPIS anomalies from year to year

What does mean "*in situ* LPIS anomalies quantification"?

⇒ **Frequency of anomalies** = to **count** each one of the significant area discrepancies existing between *"parcel area from up-to-date VHR imagery"* vs. *"reference parcel area registered within LPIS"*

⇒ **Magnitude of anomalies** = to **measure** (in ha) the area discrepancy pointed out previously

⇒ **Exclusion/reduction process** = then, to decide of the way **to correct** and fine-tune farmers' aids calculation

Problems:

- To be 100% accurate, all the parcels inside national LPIS should be controlled and anomalies pointed out and measured **UNREALISTIC**

Solution:

- To sample a single part of LPIS larger enough to obtain an accurate and acceptable estimation of the true LPIS quality **FEASIBLE**

What is a LPIS anomaly (in the frame of this study)?

⇒ **LPIS irregularity** = ***significant discrepancy between the area stated in the farmer's annual declaration and the reference area in the LPIS, calling for withdrawal of undue payments. This implies control procedures, measurement and monetary valuation of the discrepancy in area triggering a reduction/exclusion/correction process (WikiCAP)***

No area discrepancy was measured, only the presence (counting) was done

Only area discrepancies equal to or higher than 0.01 ha.parcel^{-1} were considered

Substitution

→ **LPIS anomaly** = ***a more general term in our work which doesn't imply measurement, monetary valuation and reduction/exclusion process but only area discrepancies observation and counting (≥of 0.01 ha.parc^{-1})***

Examples of LPIS anomaly

Farm land contained within reference parcel

≥0.01 ha.parcel^{-1}

Part of a building contained within reference parcel

Examples of LPIS anomaly

Construction wastes deposition within reference parcel

≥0.01 ha.parcel^{-1}

New buildings contained within reference parcel

Examples of LPIS anomaly

Buildings

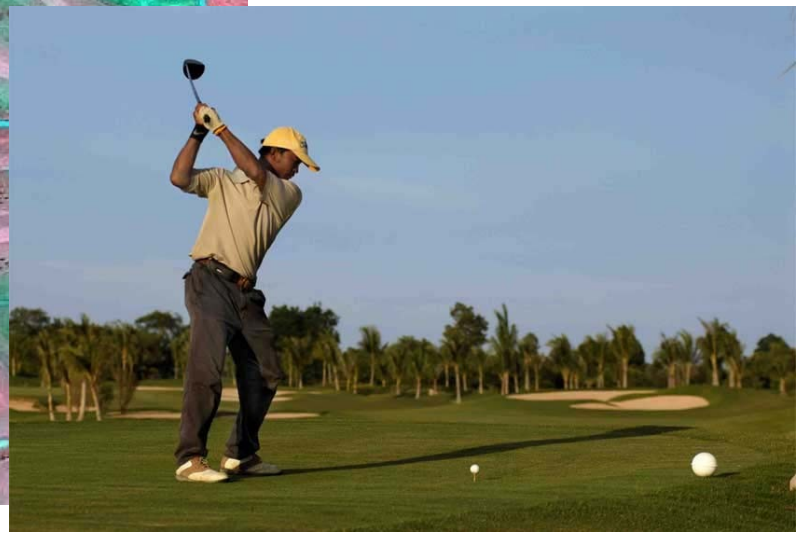
≥0.01 ha.parcel^{-1}

Asphalt roads

Examples of LPIS anomaly

Any idea??

≥0.01 ha.parcel^{-1}



Ex-ante typology of LPIS anomalies

'Trees' related	'Man-made' related	'Water' related	'Parcel boundaries' related
(TRA1) Patch of the trees	(MRA1) Building	(WRA1) Pond	(BRA1) Shift
(TRA2) Row of the trees	(MRA2) Other Man-Made objects	(WRA2) Ditch	(BRA2) Not agricultural land use
(TRA3) Forest	(MRA3) Ground road	(WRA3) River	(BRA3) Boundary not following stable land features
(TRA4) Dense single trees	(MRA4) Asphalt road	(WRA4) Marshy zone	(BRA4) Parcel selected for check overlap with another parcel
	(MRA5) Farm land		

LPIS anomaly status

From ecology

LPIS anomalies are **diverse** (different types) → species diversity or biodiversity

LPIS anomalies are **immobile** (at least during the CwRS campaign) → as plants !

LPIS anomalies have to be **counted** to obtain anomalous parcel density ($.km^{-2}$) → absolute density is expected !

LPIS anomalies as area discrepancies are **not measured** → No individual LPIS anomaly information such as area is expected

Method of sampling = QUADRAT COUNT METHOD

Main objectives of the study:

From a VHR (1m resolution) imagery dataset used for CwRS (year 2005, 11 different MS, 12 zones 50*50 km, 4 types of LPIS)…

1- Quadrat count method of sampling of LPIS anomalies has been calibrated and validated on one single zone

2- Then, it has been used to assess the overall quality of LPIS by MS and by type of LPIS onto the 12 zones selected

3- Main LPIS anomalies identified have been ranked to address main LPIS anomalies in future studies

4- Correctness of the LPIS anomalies typology proposed ex-ante has been verified

1- Quadrat Count Method (QCM) calibration and validation:

Zone used = **FB2** (Farmer Block N°2) because
= EU averaged size of the parcel
= EU traditional landscape
= one of highest LPIS quality

Assumption: *sampling method calibrated from medium agricultural and highest quality situation (FB2) should provide effective tool in most of the EU situations, especially when LPIS quality is lower*

FB2 has been completely checked, parcel by parcel, and all LPIS occurence (types, number, >0.01ha) have been pointed out in GIS

= TRUE diversity of LPIS anomalies in FB2

⟺

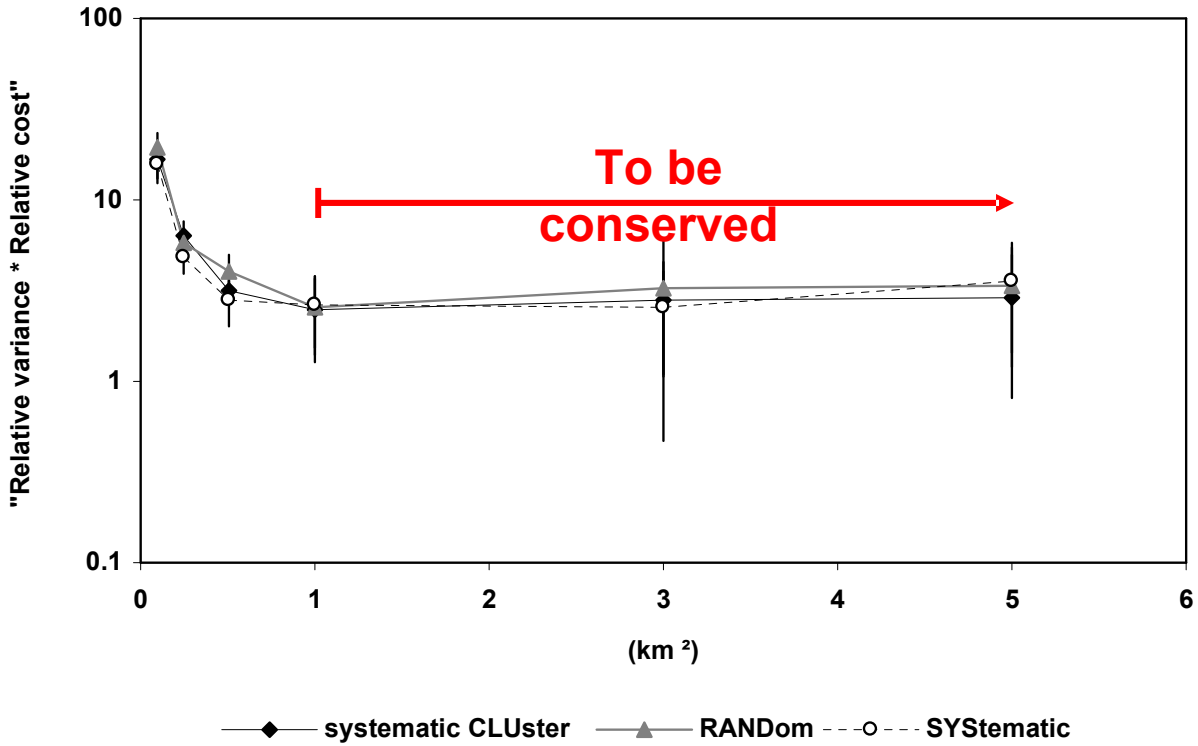
Part of FB2 is then sampled according to different QCM attributes and compared to the true values

= SAMPLED diversities of LPIS anomalies in FB2

1- Quadrat Count Method calibration and validation:

→ ***Quadrat size:***

-we tested 6 different sizes of quadrat from 0.1 up to 5 km² (20 replicates)
-they were compared by estimating the relative "variance * cost" product (low variance of the LPIS anomaly frequency is expected at lower check cost)



Quadrat sizes ≥ **1 km²** provide lowest variance of LPIS anomaly frequency at lowest cost

Remark: not visible because of the graph scale, but the product slightly increase for quadrat size higher than 1 km²

1- Quadrat Count Method calibration and validation:

→ ***Sampling size / number of quadrats:***

-indirectly calculated from the minimum number of parcels to be checked in FB2 where 5.76% (p) of the parcels in the zone were anomalous and 94.24% (q=1-p) were non anomalous

-we expected a minimum number of parcels to be checked accurate enough to provide anomaly frequency at 2% of TRUE at 90% confidence level

We want to be 90% confident (α=.01 and df=infini)

$$n_{FB_2} = \frac{(1.645)^2(0.0576)(0.9424)}{(0.02)^2} = 367$$

To find FB2 anomaly occurrence observed (p=5.76%, 1-p=94.24%)

This, at more or less 2%

The minimum number of parcels to be checked to obtain a sampled anomaly occurrence closed to the true anomaly occurrence at ±2% is **367 parcels**

1- Quadrat Count Method calibration and validation:

→ ***Sampling size / number of quadrats:***

- these 367 parcels to sampled are then expressed in km² by multiplying by the mean parcel area observed in each of the 12 zones

LPIS types	Replicates	Mean area (ha)	km²
	AP1	4.8	17.62
AP	AP2	1.17	4.04
	AP3	3.73	12.95
	CP1	0.56	2.05
CP	CP2	3.91	14.35
	CP3	0.41	1.52
	FB1	6.66	22.87
FB	FB2	5.04	18.5
	FB3	5.12	16.92
	PB1	4.41	14.47
PB	PB2	1.82	6.68
	PB3	38.54	120.82

From FB2, 18.5 km² should be sampled

The maximum area to be sampled from PB3 would be 121 km² = unrealistic regarding time and resources available for this study (2 months/2 persons)

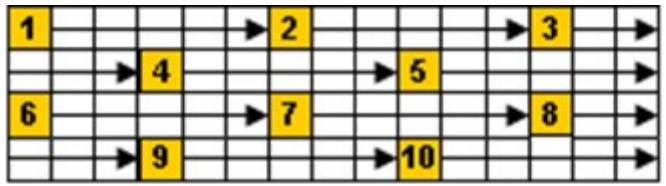
Thus, we chose the second higher area to be sampled (FB1) and we multiply by two = 45km²

It should allow for sampling accurately the first eleven zones and give an approximate of the twelfth one (PB3)

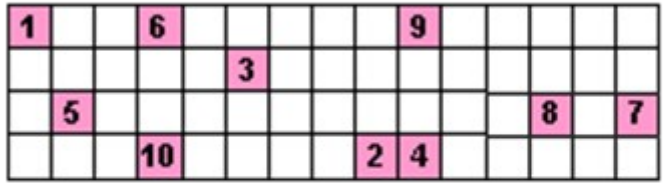
1- Quadrat Count Method calibration and validation:

→ ***Sampling design i.e. distribution of quadrats:***

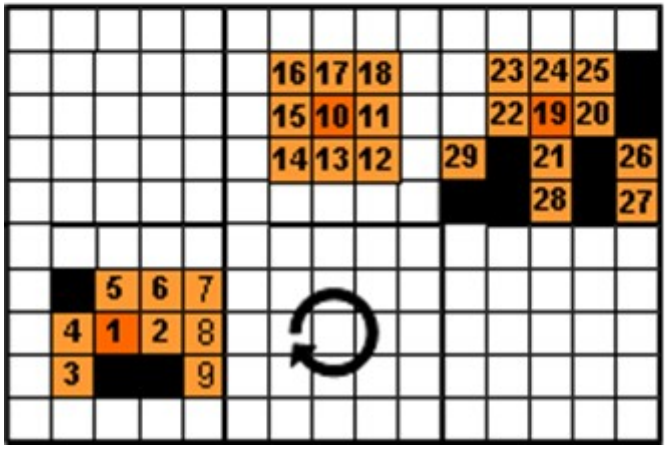
- three different QCM designs have been tested (Systematic = SYS, Random = RAND, Systematic cluster = CLU) (20 replicates)



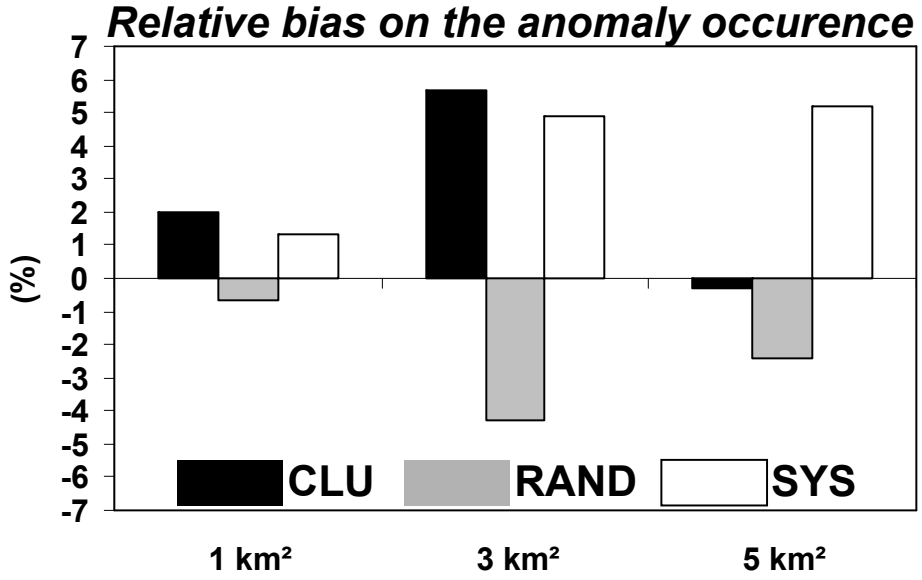
SYS



RAND



CLU



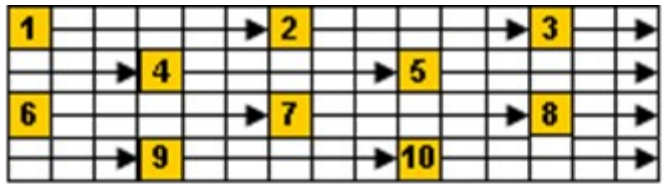
-whatever the design, 1km² quadrat size estimation is closed to true at 2%

-Random design tends to underestimate when systematic designs are overestimating

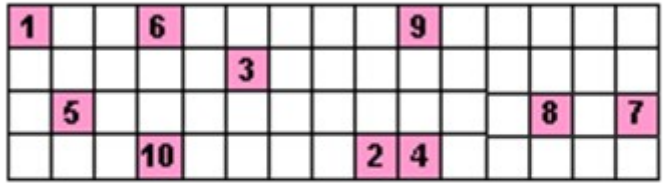
1- Quadrat Count Method calibration and validation:

→ ***Sampling design i.e. distribution of quadrats :***

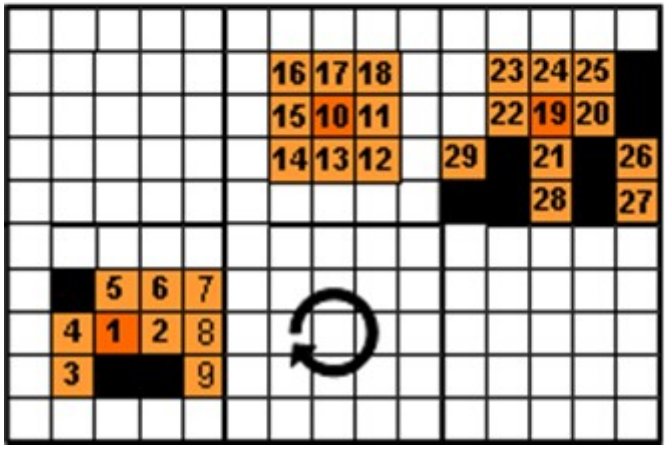
- three different QCM designs have been tested (Systematic = SYS, Random = RAND, Systematic cluster = CLU)



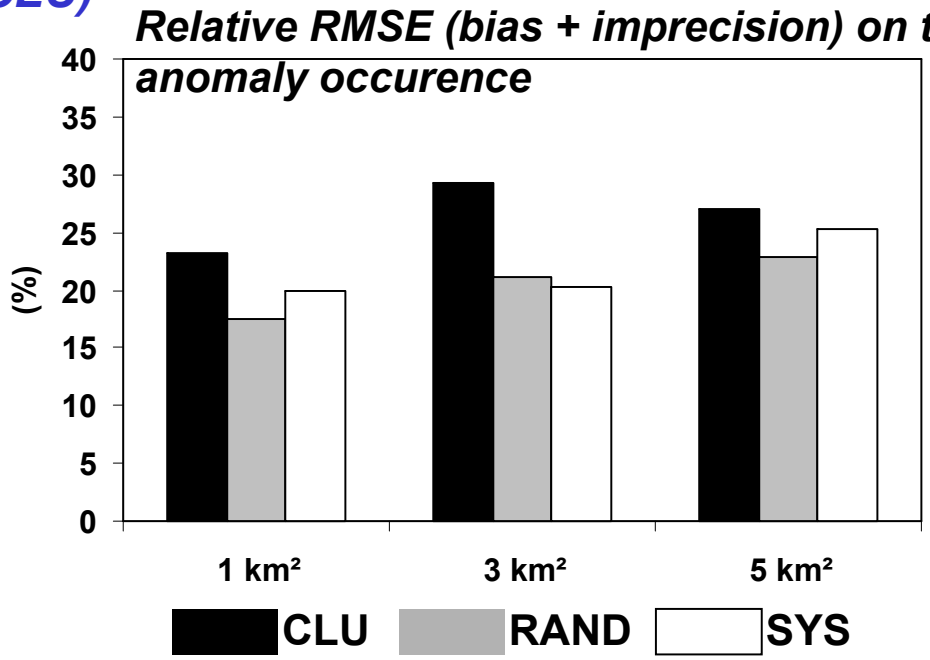
SYS



RAND



CLU



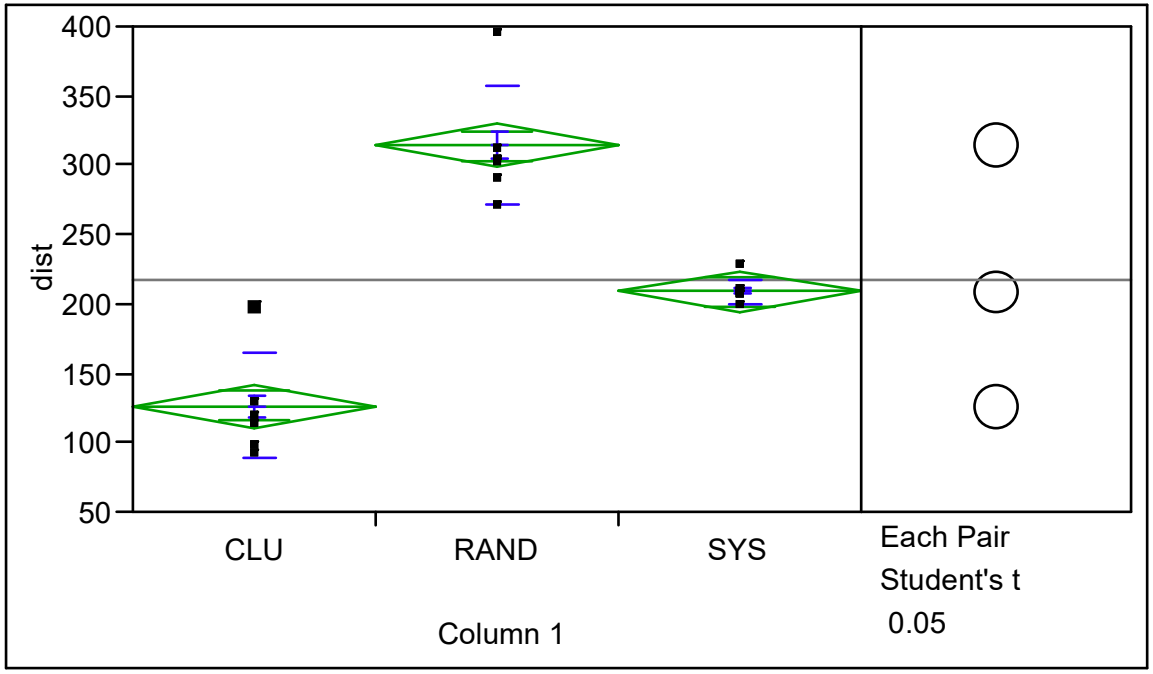
-whatever the quadrat size, CLU is slightly less accurate than other designs

- quadrat size 1km² is definitely the more accurate one

1- Quadrat Count Method calibration and validation:

→ ***Sampling design i.e. distribution of quadrats :***

- we also considered the travel costs (distance as a proxi) for a field inspector to checked all the sampled parcels in FB2 (20 replicates) according to each sampling design



CLU = 126 ± 38 km

RAND = 314 ± 43 km

SYS = 209 ± 8 km

1- Quadrat Count Method calibration and validation:

→ ***Conclusions regarding the calibration of the sampling method:***

To efficiently sample FB2 zone and reach a sampled anomaly population closed to the TRUE anomaly population

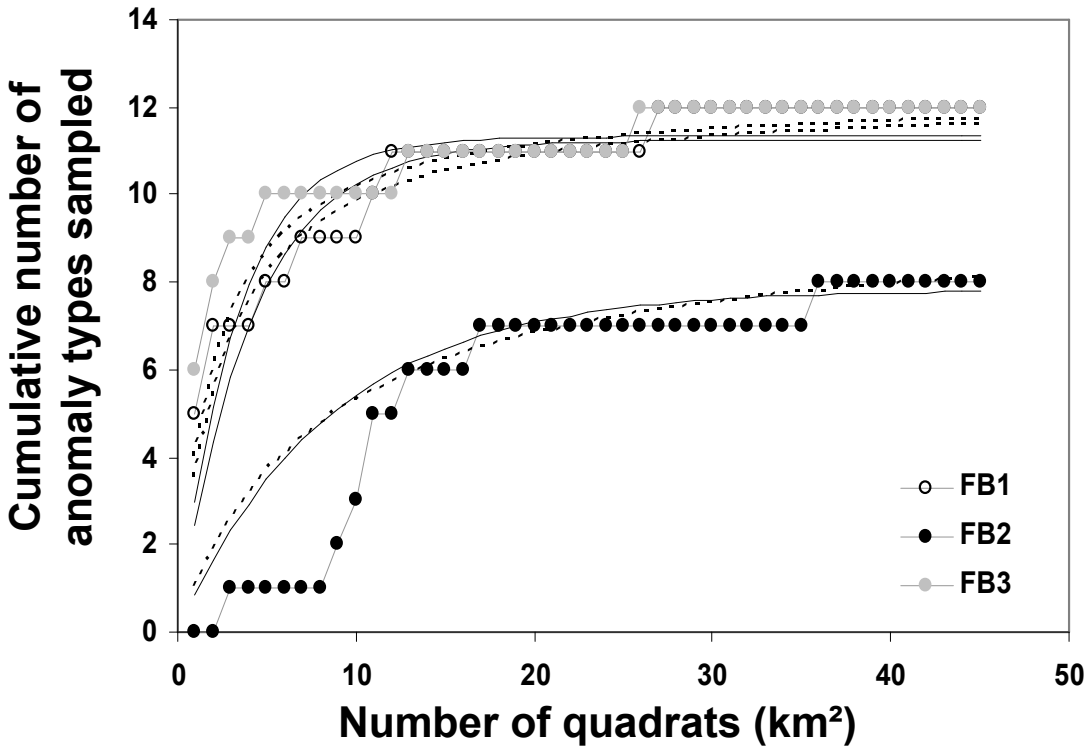
we must…

- **used square quadrat of 1 km² size**
- **used systematic cluster sampling design (a bit less accurate but very cost-effective)**

1- Quadrat Count Method calibration and validation :

→ ***Is the 45 km² sampling effort (from FB2) valid for all the zones?***

- we considered here the effectiveness of sampling to capture the diversity of anomalies existing in one zone (3 Farmer Block zones as examples)



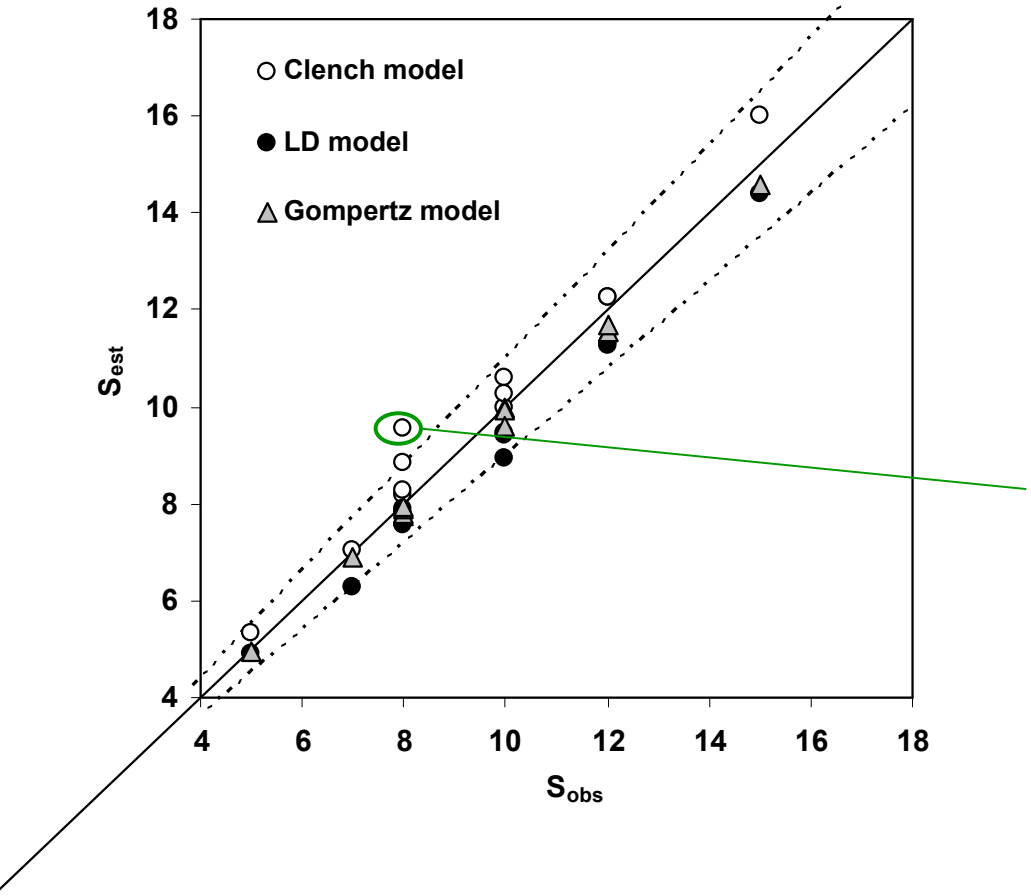
The LPIS anomaly diversity is fitted against the increasing sampling effort (from 1 to 45 km² sampled):

-If a maximum is reached = 45 km² sampling effort is sufficient (at 90% confidence) to sample the true diversity of LPIS anomalies within a zone

- If a maximum is not reached = 45 km² sampling effort is insufficient (at 90% confidence) and supplementary quadrats should be sampled

1- Quadrat Count Method calibration and validation :

→ ***Is the 45 km² sampling effort (from FB2) valid for all the zones?***



Whatever the model used to fit the observations and whatever the zone considered:

- 45 km² sampling effort was sufficient to obtain LPIS anomaly diversity closed to the TRUE (at least 90% confidence)

-Only for FB2, the sampled diversity was estimated as 85% of the true diversity

-The mean estimation, all zones and models considered, of the true diversity was 98% !!

2- Overall LPIS quality assessment:

→ ***What was the overall quality in each one of the zones selected?***

LPIS types	LPIS zones	Total No of reference parcels sampled	No of anomalous reference parcels	Total No of anomalies	Anomalous reference parcels percentage
AP	AP1	866	237	460	27.37
	AP2	1 321	167	223	12.64
	AP3	879	232	614	16.39
CP	CP1	6 233	1 762	3 061	28.27
	CP2	2 547	121	170	4.75
	CP3	6 243	504	623	8.07
FB	FB1	619	237	525	38.29
	FB2	573	48	60	8.38
	FB3	788	434	1 143	55.08
PB	PB1	547	271	578	49.54
	PB2	922	101	186	10.95
	PB3	154	103	1 662	66.88

One of the highest LPIS quality

2- Overall LPIS quality assessment:

→ ***What was the overall quality in each one of the zones selected?***

LPIS types	LPIS zones	Anomalous reference parcels percentage	Average (STDV)	significant difference
AP	AP1	27.37	18.8 (7.7SD)	(a)
	AP2	12.64		
	AP3	16.39		
CP	CP1	28.27	13.7 (12.7SD)	(a)
	CP2	4.75		
	CP3	8.07		
FB	FB1	38.29	33.9 (23.7SD)	(a)
	FB2	8.38		
	FB3	55.08		
PB	PB1	49.54	42.5 (28.6SD)	(a)
	PB2	10.95		
	PB3	66.88		

- Proportions of anomalous parcels were high (from 5% up to 67%!)

-No significant difference between LPIS types

-But a trend to observe higher anomalous parcels proportion in "blocks" (FB, PB) than in single parcels (AP, CP)

-***Remark:*** ***it doesn't mean that the LPIS are of very bad quality – do not forget that it was anomalies and not irregularities !!!!!!!!!!***

3- Ranking of the identified LPIS anomaly types

→ ***which anomaly types have predominated?***

Main sources of anomaly within LPIS:

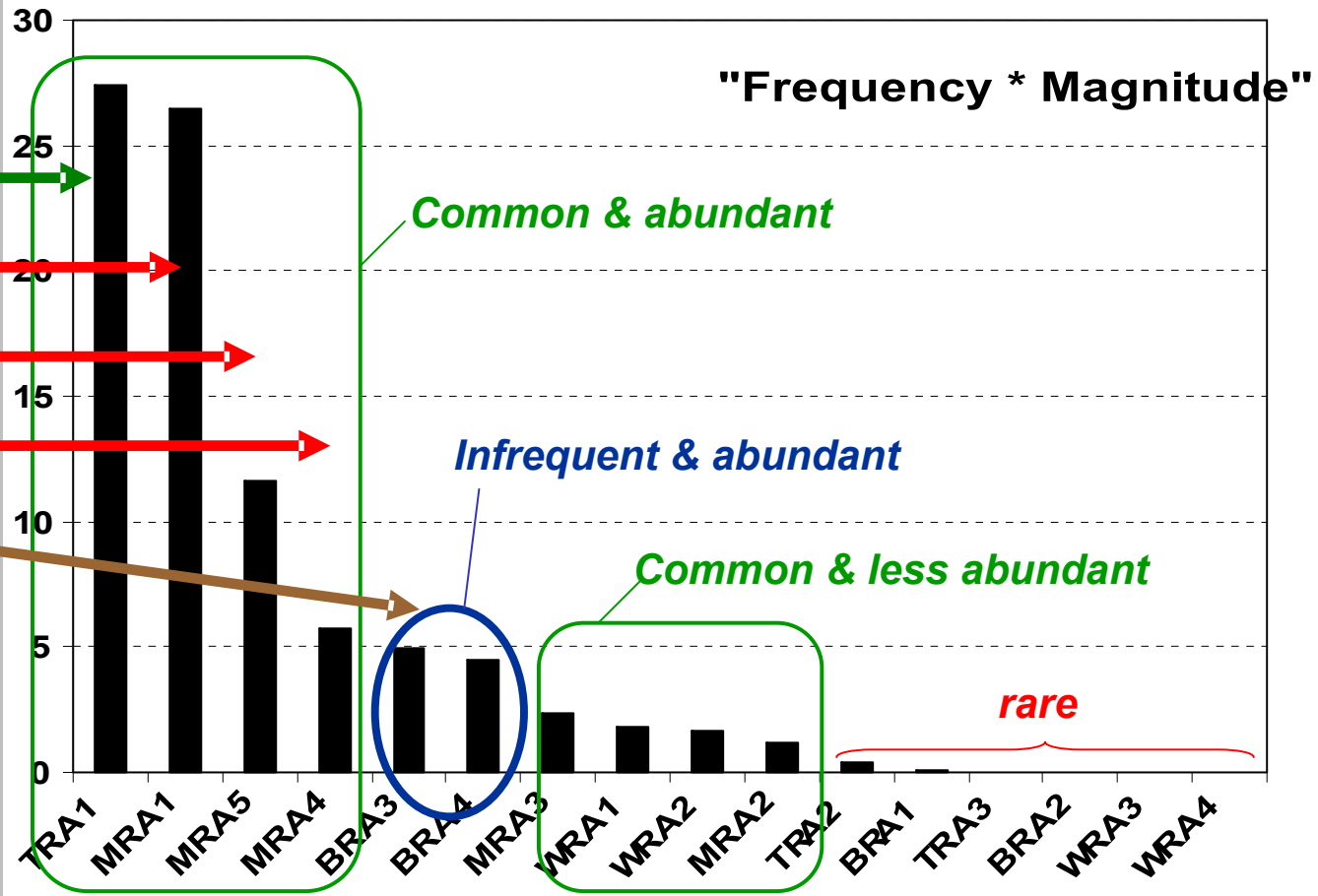
-Patch of trees

-Buildings

-Farm land

-Asphalt road

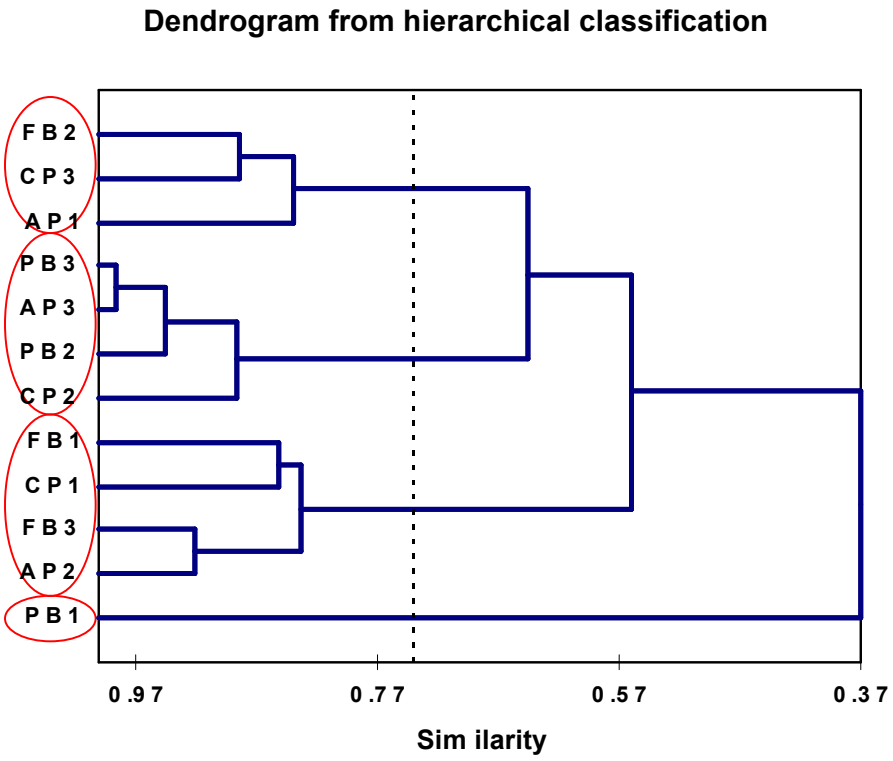
-Significant overlap between contiguous parcels



3- Ranking of the zones

→ ***Were the 12 zones similar regarding their relative anomaly composition?***

4 clusters identified



Single anomalies	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	AP2, FB1, FB3, CP1	AP3, PB2, PB3, CP2	AP1, CP3, FB2	PB1
TRA1	41.6	29.55	16.85	18.69
TRA2	1.61	1	0.41	0
TRA3	0.29	0.02	0.18	0
MRA1	17.5	15.28	49.85	4.33
MRA2	4.35	0.63	1.35	0
MRA3	5.52	0.82	3.26	0.87
MRA4	2.64	5.39	6.91	11.42
MRA5	12.83	7.69	16.53	3.98
WRA1	3.99	1.13	1.21	1.9
WRA2	0.84	0.57	1.82	17.3
WRA3	0.58	0	0	0
WRA4	0.07	0	0	0
BRA1	1.45	0.09	0	0
BRA2	0.58	0.05	0	0
BRA3	6.16	5.55	1.6	41.52
BRA4	0	32.23	0.04	0

As % by cluster

4- Correctness of the LPIS anomaly typology:

→ ***Are all the possible anomaly types (17 proposed) identified?***

Except TRA4, all the anomaly types proposed ex-ante have been pointed out during the study

(TRA4)
Dense single trees

Assumption 1: TRA4 situation was not present in the zones selected but exist somewhere else in EU

Assumption 2: TRA4 doesn't exist in all EU

The ex-ante typology of LPIS anomaly is robust and can be conserved – it asks for supplementary checks to validate or exclude TRA4

Conclusions

- To sample LPIS anomalies diversity correctly:

  Quadrat Count Method is the one

  45 km² sampling effort sufficient to sample at least 90% of the LPIS anomaly diversity

  Square quadrat of 1 km² provide acceptable imprecision and bias of the estimation

  Systematic cluster sampling design as less accurate but more cost-effective for further in situ verification by field inspectors

- LPIS quality depicted:

  high proportion of anomalous parcels whatever the LPIS type considered

  independently of LPIS type, zones can be grouped in 4 clusters mainly explained by the nature of the 1st or two first anomaly types pointed out

  man-made- and tree-related anomalies are predominant in all LPIS

Limits and perspectives

- Limits … LPIS Anomaly concept:

  Anomaly IS NOT irregularity

  Proportion of anomalous parcels in a MS should be corrected according to national regulations *(some tree-related features are accepted in some MS)*

  0.01 km².parcel-1 threshold is debatable and should be adapted to EU regulations *(0.1 km² for exclusion)* or JRC LPIS guideline *(0.01km² for declaration)*

  Despite this possible adaptations, the sampling method is valid !

- Perspectives:

  To apply same method onto different zones to validate LPIS typology (TRA4)

  To follow over years (time series) same zones to assess the efficiency of LPIS up-dating process

  To distinguish the relative influence of “landscape” and “LPIS creation/updating process” onto the presence of LPIS anomalies

  To promulgate the method as one possible tool to precise places where to yearly conduct CwRS risk analysis in MS

Final remark

All this study has been performed outside any EC framework – and cannot be perceived as a possible new constraint for MS in the frame of CAP

at the opposite

It has been developed keeping in mind that some MS and DG AGRI are actually searching or are expecting help from JRC to define effective and relevant LPIS quality control methods

So it is !

Thank you for attention

If you want to obtain the document (pdf version):

http://mars.jrc.it/marspac/GEO-INF/meetings/LPISworkshop2008/index.htm