EUROPEAN COMMISSION
JOINT RESEARCH CENTRE - ISPRA
Institute for the Protection and Security of the Citizen
AGRICULTURE Unit

Ispra, 31/07/2007
JRC IPSC/G03/P/HKE/hke D(2007)(8218)
file://S:\FMPArchive\P\8218.doc

TECHNICAL RECOMMENDATIONS

FOR THE 2008 CAMPAIGN OF

**REMOTE-SENSING CONTROL OF
AREA-BASED SUBSIDIES**

SELECTION OF CONTROL ZONES AND RISK ANALYSIS

DRAFT GUIDANCE DOCUMENT

CONTENTS

Post: IPSC-AGRIFISH, TP 266, Joint Research Centre, I-21020 Ispra (VA), Italy
Telephone: direct line +39 0332 78 9639
Facsimile: +39 0332 78 6369
E-mail: herve.kerdiles@jrc.it URL: http://agrifish.jrc.it/

1. Introduction

1.1. Objectives of the document

1.1.1. Member States (MS) or regions (hereafter referred to as MS) are requested to provide information on their control zones (number, area, targeted number of applications to be checked with remote sensing) and the number and type of satellite images in the autumn preceding the campaign, i.e. at a time where applications have not yet been lodged.

1.1.2. The objective of this document is to provide technical guidance for the definition of the zones to be controlled with remote sensing. A second document will provide guidance on the selection of the imagery for the control of area based subsidies.

1.1.3. Although this document is intended for MS applying CwRS, some of the recommendations or techniques also apply to classical inspections for what regards the selection of the OTSC sample. An integrated strategy between the two types of On-The-Spot (OTS) checks is therefore recommended.

1.1.4. This document only provides recommendations that should be adapted to the national context. The only authoritative references are the Regulations applicable in the Member State concerned.

1.1.5. This document is available on the MARS Control with Remote Sensing (CwRS) website at http://agrifish.jrc.it/marspac/CwRS/default.htm

2. Definition of control zones

2.1. Number of control zones

2.1.1. Ideally the whole Member State would be covered with Very High Resolution (airborne or satellite) images of the year so that the selection of the sample of applications to be checked On-The-Spot could be free of any constraint, i.e. based on the sole characteristics of the applications once these are entered in the IACS. However, because of budgetary, technical and time limitations, this strategy is hardly ever feasible and controls are made on a limited number of zones or “clusters of applications”, which may reduce the quality of both the random and risk based samples. These control zones need to be defined before applications are known taking account of a number of constraints.

2.1.2. The CwRS strategy which has to be defined in the summer / autumn preceding the campaign can be characterized by the following parameters or options:

- The rate of CwRS checks with respect to the total number of OTS checks to be carried out in a given MS or region;
- The number of control zones to achieve the targeted number of CwRS checks;
- The method of selection of these control zones (at random or on the basis of risk analysis);

- The method of selection of the applications inside the control zones; although not directly related to the definition of control zones, this criterion may affect indirectly their number or extent (e.g. in case applications are selected on the basis of risk analysis inside the control zones);
- The number and type of images (airborne/satellite, resolution) requested for the checks to be carried out as well as their timing (windows).

2.1.3. There is no simple rule to define the number of control zones for a given MS. This number is usually set as the result of experience as well as of budgetary, logistical, regulatory, landscape and other constraints. MS new to CwRS are advised to start with a low number of zones to assess the pros and cons of the technique. A large number of zones may allow a better distribution of the control pressure as well as a better representativeness (in case random zones are selected) while reducing the number of classical inspections in case of failure of image acquisition over some zones. However it also increases the fixed costs of image acquisition management, image processing, training (ground truth to train interpreters)...

2.1.4. The following criteria (which may be correlated) may be considered for deciding on the number of control zones:

- The effectiveness of Remote Sensing (RS) with respect to the alternative classical inspections: independently of the number of applications to be checked per zone[1], this effectiveness may depend on the landscape structure (e.g. presence of extensive agricultural areas to ensure a target controlled area of at least 25% of the control zone, large fields or large farms for which the classical field inspections are time consuming and costly) and of the control needs (e.g. type of crops or GAEC to be checked, proportion of applications for Agri-Environmental Measures for which a field visit is requested[2]);
- The number of applications to be subjected to CwRS;
- The average size of the zone (compromise between the technical capacity of the satellites, logistical constraints…) and the average number of applications per zone (to be estimated based on historical claims using the LPIS).
- Logistical constraints: care should be taken not to overload a few regional offices by concentrating most CwRS checks on their respective territory. As much as possible, CwRS applications should be provided to the contractor in digital form and after administrative cross checks to avoid time consuming requests for clarification. When this processing is made by a regional office of the Paying Agency or the Ministry of Agriculture, it is worth considering the number of applications this office can prepare within a given time period (e.g. within 1 month of the lodging deadline). Also when the follow up of the CwRS checks is ensured by a regional office, it is worth considering the number of checks this office can carry out in due time (e.g. with respect to the crop calendar or other time constraints).
- Budgetary constraints (both at EU and MS levels)

[1] It is usually more effective (faster and cheaper) to check a large number of applications in a given zone using remote sensing rather than classical inspections.

[2] Fur such applications, CwRS may appear less effective than classical inspections

2.2. Random zones versus zones selected by Risk Analysis

2.2.1. According to article 27(1) of Regulation EC N° 796/2004, the control sample for OTS checks shall be selected on the basis of a risk analysis and representativeness of the aid applications. The representativeness shall be obtained by selecting randomly 20 to 25% of the minimum number of farmers to be subjected to OTS checks.

2.2.2. For the selection of the random sample, two strategies may be applied:

- Select applications randomly (using for instance simple random sampling) from the whole list of applications. Most likely this sample will be scattered over the MS territory and will have to be checked by classical inspection for most of the claims. However applications falling in a control zone may be checked with RS (and will be counted as part of the random sample even if the zone was selected on the basis of risk analysis).
- Alternatively, select zones randomly and inside these random zones select applications systematically (e.g. all applications falling in the zone) or randomly to constitute (part or a total) of the total random sample. However, care should be taken to ensure that the random sample is **representative** of the background level of anomalies in the MS; in effect it is feared that the constraints affecting the definition of control zones (e.g. the targeted minimum % of controlled area in a zone) may bias the selection of random zones towards intensive agricultural areas which may not be representative of the whole country. Also as a general rule, it is also not advised to have the random sample concentrated in one or 2 zones. It is felt that a minimum number of **5** random zones should be defined for the representativeness of the random sample.

2.2.3. A combination of the previous two strategies is also possible, for instance in countries where two distinct strata coexist: one stratum of intensive agriculture inside which random zones could be selected for RS checks and the other of more extensive agriculture (i.e. pastures mingled with non agricultural features) in which classical inspections would be used to check the scattered (random) applications.

2.2.4. For the selection of the risk based sample, again two strategies are possible:

- Select the control zones (nearly) at random and perform RA inside the zones (provided there are enough applications in the zones to allow an efficient RA);
- Select control zones using RA and then select applications inside these zones either in a systematic way (e.g. all applications with more than 50% of their parcels falling in the zone) or using RA among the applications falling inside the zones, in case the number of applications inside the zones is larger than the targeted number. Selecting all applications inside a zone selected by RA is likely to result in a weaker RA than selecting applications individually out of the whole population of applicants. However because of budgetary and technical constraints, controlling zones selected by RA, i.e. geographical clusters of applications, is the best compromise between efficiency and cost. Moreover, controlling all applications in a given area may enable a more complete check of adjacent applications (sharing the same reference parcels).

2.2.5. Very large applications (with high risk factor) with less than 50% of their parcels inside the control zones could be also included in the sample to be checked using RS as this will increase the effectiveness of RS. Even if the check of such applications may have to be completed with classical inspection, it may be worth using RS for the part falling inside the zone.

2.3. Mapping risk factors

2.3.1. It is assumed that risk factors have been determined[3] by the MS and any application can be given a "score" per risk factor, most likely using previous year(s) data (e.g. application data from year n-1, control results from year year n-2, if the applications to be checked with RS are lodged on year n). As a result each application can be assigned a global risk score which is the weighted sum of the individual risk scores, where the weights are fixed according to the importance given to each risk factor. Alternatively, any application may be assigned a class of risk using a combination of risk factors (Cf. § 4.2). In such a case, it is advisable to convert these "qualitative" classes into quantitative scores for the next step (e.g. if class B and D are considered to be high risk, they will be assigned a high score).

2.3.2. There are several ways to map risk at MS level in order to define the control zones on the basis of RA:

- Map the risk at LPIS parcel level.
- Map the risk at administrative unit level (e.g. commune) by taking account of the applications falling inside each unit.
- Map the risk at grid level: define a regular grid for instance with a spacing fitting the swath of VHR satellite (e.g. a 10 x 10 km grid) and compute the average risk per cell by taking account of the applications falling inside each cell.

2.3.3. Mapping the risk at LPIS parcel level is straightforward when each LPIS parcel can be assigned to a single farm (case of agricultural parcel or farmer's block LPIS) as the farm risk can be directly assigned to the LPIS parcel. When the LPIS parcel is shared by several farmers, a possibility could be to compute the LPIS parcel risk score as a weighted average of the risk scores of the parcels claimed inside the LPIS parcel, with weights proportional to each parcel area (i.e. equal to the parcel area divided by the LPIS parcel area). In this proposal, an LPIS parcel with only part of its area being claimed (e.g. 10%) is likely to have a low risk score, which is consistent with maximizing the use of the imagery (i.e. the area effectively checked) over a control zone[4].

2.3.4. There may be a need to summarize the risk map derived at LPIS parcel level to a lower level of detail so as to ease the selection of zones. This could be made through a weighted average of the risk scores of each LPIS parcel falling inside an administrative unit or grid, using for instance weights proportional to the LPIS parcels area (i.e. LPIS parcel area divided by grid area).

2.3.5. Alternatively, the risk at administrative unit or grid level may be mapped as follows:

---

[3] If this is not the case, see § 4 on the assessment of the risk analysis

[4] On the other hand, such an LPIS parcel could also be considered at risk due to the inefficiency of area cross checks. It could however be selected in an extra sample based on LPIS parcels not fulfilling the 90% rule (art 6.2 of Reg 796/2004).

1. identify the farms falling inside the administrative unit or grid;

2. compute the (weighted) average risk score of the administrative unit or grid for each risk factor (a weight may be introduced to account for the different sizes of farms);

3. For each risk factor, rank the administrative units or grids; it is possible to assign them a risk class (e.g. between 1 and 5) according to their quantile;

4. For each administrative unit or grid, compute the overall risk score using a weighted average of the risk scores of the different factors. Different weights may be assigned to give more emphasis on particularly high risk factors.

2.3.6. The result of the risk map is a stratification of the MS showing for instance high, medium and low risk strata. There could also be a stratum "not suitable for CwRS" if the density of applications is too low in some parts of the MS.

2.3.7. Selecting control zones on the basis of RA does not necessarily mean selecting all zones in the high risk stratum only (which may be the same every year). Zones could also be selected in medium and low risk strata, for instance with lower sampling rates than in the high risk stratum. This strategy presents the advantage of distributing the control pressure in every stratum (therefore respecting the principle according to which "no application should have a zero probability of being checked"), which may later be useful at the time of assessing the RA.

2.3.8. Selecting the CwRS zones inside each stratum can be made "manually" or at random. The latter strategy is called stratified random sampling and may allow estimating at the same time the background level of anomalies. In other words, the RA sample may help "completing" the (simple) random sample of 1 to 1.25% of all applications. This could be particularly useful in MS where the (simple) random sample may be considered too small[5] for deriving a reliable background level of anomalies (and later carrying out a reliable RA assessment).

2.3.9. To illustrate how the RA sample could be used for estimating the background level of anomalies, here is a numerical example for the selection of a (stratified random) 3.8% RA sample[6] that could be applied for CwRS or classical inspection. Each stratum is sampled: the highest sampling rate is set in the high risk stratum; the medium risk and to a lesser extent the low risk strata are also sampled to account for the possibility that the stratification (i.e. the risk factors selected) may not be perfect.

[5] i.e. less than several hundreds applications. In small MS, the 1-1.25% sample will likely result in an unstable estimation of the background level of anomalies.

[6] Such a rate would leave space for a manually selected RA sample

Stratum	% of applications in each stratum	Rate of OTS checks per stratum (% of applications checked in each stratum)	% of applications checked per stratum with respect to the total number of applications	Average rate of anomalies (missing ha / declared ha)
1 high risk	10%	20%	2%	8%
2 medium risk	30%	5%	1.5%	4%
3 low risk	60%	0.5%	0,3%	2%
total	100%		3.8%	

The overall rate of checks is 3.8% (20%x10% + 5%x30% + 0.5%x60% = 2% + 1.5% + 0.3% = 3.8%)

The average rate of anomalies of this RA sample is 5.95% ((8% x 2% + 4% x 1.5% + 2% x 0.3%) / (2% + 1.5% + 0.3%) = 5.95%)

whereas the background level of anomalies as derived from this random sample is 3.20% (8% x 10% + 4% x 30% + 2% x 60% = 3.20%)

2.3.10. In practice, independently of the strategy of control (CwRS or field inspection) a possible strategy for selecting the OTSC samples could be:

- Select a 1% (20% of 5%) simple random sample from the whole population of claims; these will be likely checked by classical inspection (unless they fall in control zones);
- Divide the remaining 4% RA sample into 2 groups: one group selected manually and the other selected on the basis of stratified random sampling using varying sampling rates[7].

Such a strategy could be translated into random and risk zones for part or the total of the OTSC sample.

2.3.11. In order to meet the targeted number of CwRS applications, it is advisable to select a slightly larger number of applications as the real number of applications is not known at the time of zones definition.

2.4. Satellite technical constraints

2.4.1. Whether the control zones are selected on the basis of a risk analysis or at random, the technical constraints of satellite remote sensing sensors should be taken into account to optimize the probability of acquisition. These constraints do not apply if the zone is to be covered with aerial orthophoto.

2.4.2. The main constraint is the size and shape of the zone with respect to the coverage of the Very High Resolution (VHR) satellites: since these have narrow swaths (of the order of 10 – 15 km – see VHR Imagery Specifications for the CwRS Programme, Issue 1.2[8]), it is advisable to define a zone so it that can be acquired in one pass (or one day for satellites

[7] NB: in high risk strata or classes with relatively few claims, all the claims in the stratum may be selected.

[8] Available on line at http://agrifish.jrc.it/marspac/DCM/default.htm

able to make several adjacent passes in a short time like Ikonos) so as to avoid, weather permitting, zones covered with scenes fragments acquired with several weeks difference.

2.4.3. Accepting low elevation angles[9] (higher off-nadir view angles) for VHR satellites also increases the number of acquisition attempts, therefore likely reducing the period needed to cover the zone. However MS should ensure that the ancillary data needed to orthorectify the VHR imagery (e.g. DEM, GCPs) is of adequate accuracy over the selected zone (see Guidelines for Best Practice and Quality Checking of Ortho Imagery Issue 2.6[10]).

2.4.4. High Resolution satellites are usually not a constraint since scene extensions are significantly larger than VHR scenes. The scenes of the main HR satellites are as follows: 60 x 60 km for SPOT (SPOT 5 may also be ordered in 40x40km, 30x30km, or 20x20km scenes); 170x183 km for Landsat TM (or 55x55 km for a mini-scene); and 140x140 km for IRS-P6 LISS (or 70x70 km for a quadrant).

2.4.5. The geographic coordinates of the selected zones (e.g. in xml[11] or alternatively shape format) will be checked by the VHR image providers to assess the feasibility of acquiring the zones within the time windows set. The image providers may suggest a small adjustment to the zones in order to maximise the likely time needed for covering the zone (e.g. if a zone needs a second pass just because few km2 are not covered by the 1st pass, it may be worth considering reducing the zone so that it fits in 1 pass or increasing it significantly to take advantage of the satellite swath).

2.5. Synergy with LPIS orthoimagery

2.5.1. Control zones may fall or be chosen in regions where there is a plan acquire VHR (satellite or aerial) imagery for the updating of the Land Parcel Identification System (LPIS). In such a case, the VHR imagery should be acquired and processed in priority over these control zones. This imagery could be used either as the main VHR image or as back up (depending on the timing of the flight and processing of the images).

2.5.2. Use of by-products: The Digital Elevation Model (DEM) and the Ground Control Points used for the creation of the LPIS ortho-imagery may be used for the preprocessing of the new ortho imagery (satellite or aerial) so as to save time and resources.

2.6. Acquisition of aerial photographs

2.6.1. Acquisition of aerial photography for the controls is the responsibility of the MS Administration, i.e. is not coordinated by the Commission. The main advantage of aerial photography with respect to VHR satellite imagery is that it allows covering much larger areas (e.g. large administrative units such as full provinces) in a limited period of time. Alternatively a large number of small zones may also be covered in a given region.

2.6.2. However, acquiring aerial photography has also some proper constraints such as restrictions over military zones and air traffic lanes. Cloud cover is not as restricting for

[9] the term elevation angle is used because it is not related to the sensor altitude and it better describes the sensor object geometry. Elevation angle 60° = off nadir view angle 26.9° (IK), 27.8° (QB), 27.7° (EROS), 26.2° (SPOT).

[10] Available on line at http://agrifish.jrc.it/marspac/DCM/default.htm

[11] Schema to be defined

aerial photography as for satellite imagery, but meteorological conditions are in any case affecting the radiometric quality of the photos.

2.6.3. Moreover, the lead-time in the processing of analogue aerial photography (requiring development, printout, scanning) may be longer[12] than that for satellite images. Aerial photography acquisition must therefore be organised sufficiently in advance, and the acquisition periods should be relatively early in the year. The use of (natural or infrared) colour imagery permits an easier identification of land covers, thus significantly reducing follow-up rapid field visits for crop identification.

2.6.4. It is also compulsory that aerial flights be carried out within the present state of the art: the use of GPS and inertial navigation systems (i.e., "pin-point flights") linked to the camera makes it possible to optimise the flight coverage and considerably reduce the costs of further processing.

2.6.5. If aerial photography is intended to be used, the flight plan proposed by the contractor must guarantee that the whole control area is covered. In practice, the MS Administration must approve this flight plan prior to the flight.

3. <u>Possible factors for the risk analysis</u>

3.1.1. As from 2008, according to article 1 (10) of Regulation 972/2007 amending article 27 of Regulation 796/2004, MS are responsible for the definition of the risk criteria to be used for the RA.

3.1.2. The following risk factors which were included in the previous versions of the regulation are listed here only as suggestion:

- amount of aid involved;
- number of agricultural parcels and the area or number of animals for which aid is requested;
- changes from the previous year;
- findings of checks made in past years;
- farmers who are either just above or below ceilings or limits relevant for the granting of aids;
- other factors to be defined by the Member States.

The following paragraphs report on the experience of some MS with the use of these criteria. These factors may be applied directly (e.g. select all farms or x farms with an amount of aid above a threshold) or combined through weights (giving different weights to the different factors so as to get an overall score per farm) or combined in the selection criterion (e.g. select all farms with an amount of aid above a given threshold and claiming a particular crop and having been sanctioned in any scheme in the past 4 years).

3.2. Criterion "amount of aid"

3.2.1. At first sight, the higher the amount of aid, the higher the risk. However, it was noted that this was not in general a positively correlated factor; (small) errors are more likely to be found on small farms.

---

[12] of the order of several weeks

3.2.2. Conversely, it was considered correlated to set-aside (indeed, even a "leveraged" factor due to the effect of set-aside on payments; however this observation was valid before the introduction of SPS and in any case is not valid for SAPS).

3.3. Criterion "number of parcels"

3.3.1. This criterion was generally positive when tested against previous years (more parcels to declare, more risk of error), and effective in identifying errors at the parcel level. In some MS, this factor was combined with the area (area / parcels), and only applied if >20 (for example) parcels were in the dossier. It was also noted that it is a costly criterion to control – more parcels, more to control.

3.4. Criterion "changes from the previous year"

3.4.1. This was often crop specific – an increase in pasture land was considered a useful change factor. In other MS, it was noted that the change focus is on particular (coupled) crops, where risk is positively correlated with the crop.

3.5. Criterion "findings of checks made in past years"

3.5.1. It is generally agreed that this could imply not only a "black list" of applicants (ones that had anomalies) but also a "white list" of applicants (applicants that had been checked and had presented no problems[13]).

3.5.2. It is considered that "past years" meant 3 to 4 years.

3.5.3. For instance the sum of the reductions and sanctions (in euros) or the sum of ha not determined over the last 3 years divided by the number of OTS checks could be computed for each farm or each potential control zone (administrative unit or grid).

3.5.4. Since OTS checks are limited in area, it is likely that this factor will appear as "not controlled" for many farms and potential zones or that the score of a zone may be computed over very few checks. In order to compensate for this, an additional "historical" factor based on the number of checks made for a given farm (or administrative unit) in the past n years may be used (with a higher score i.e. risk for farms or zones with less checks).

3.6. Criterion "applicants at/near ceiling levels"

3.6.1. It is agreed that this is a potentially effective criterion, e.g. in GAECs checks or in respect of the compulsory set-aside area (if farmer is at or just over the threshold).

3.7. Criterion "other factors defined by MS"

3.7.1. A number of other ideas or practices could be listed:

- Dispersion of farms: when outlying parcels are (for example) >50kms from the main farm.
- Set-aside: farms with set-aside entitlements not inspected in the last three years.
- VHR area: All farms selected (to ensure economic efficiency)
- Claimed parcel area close to GIS (agricultural parcel) area or:

---

13 Such an approach would fit well with ISO approaches of batch checking

- Area claimed is less than 90% of gross LPIS reference
- Areas where the LPIS needs to be updated
- Inclusion of dairy farms (new to SPS)
- New applicants, systematic check (but usually not found to be problematic); late applicants; applicants not using the main tool for lodging their application (e.g. paper claims in MS where the application process is digital which allows performing systematic cross checks and informing the farmer of possible mistakes);
- New farmers requesting access to the reserve (SPS)
- No control for previous n years
- Focus on sensitive crops that pose a problem for dual use: sugar/fodder beet, beans, peas[14]
- False declarations in other schemes
- applicants having received forestry grant aid in the last 3 years

4. Assessment of the risk analysis

4.1. Assessment criteria

4.1.1. According to article 1(10) of Regulation 972/2007 amending article 27 of Regulation 796/04, the effectiveness of risk analysis shall be assessed and updated on an annual basis:

- By establishing the relevance for each risk factor;
- By comparing the results of the risk based and randomly selected samples;

This analysis should be made taking into account the specific situation in the Member State.

4.1.2. The purpose of the random sample is to permit an estimate of the “background level” of anomalies. This background level, when compared with the rate of anomalies of the RA sample, allows assessing the quality of the RA. A higher rate of anomalies found in the random sample with respect to the RA sample might indicate inappropriate risk factors.

4.1.3. There are several ways to express the rate of anomalies of a given sample. This rate can be:

- The % of applications of this sample with anomalies (i.e. for which the area determined is less than the declared area);
- The % of missing area in this sample (i.e. the total area not determined over the total claimed area). In MS applying SPS, the entitlements may be taken into account to reflect the fact that missing area does not necessarily lead to reductions or sanctions.
- The amount of aid not paid over the total claimed aid; this rate takes account of the possible sanctions and of the entitlements for MS applying SPS.

At the time risk factors have to be determined for the selection of control zones (i.e. in the autumn preceding the campaign), the percentage of missing area for each claim may be of easier access than the amount of aid (if sanctions have not yet been calculated).

---

14 This criterion is generally used to identify applications for classical inspections as the signature in RS is difficult to discriminate between these crops

4.1.4. The "background level" of anomalies should also support decisions enacting the mechanism for increasing the control rate (Cf. article 26 of regulation 796/2004 and AGRI document DS/2006/24 – REV 1).

4.1.5. Article 27(1) of regulation 796/2004 fixes the random sample between 20% and 25% of the minimum sample rate for OTS checks, i.e. between 1 and 1.25%. This may create a problem in large sized countries, where for instance more than 10,000 random dossiers would be sampled. This is likely to be much higher than necessary to estimate the "background level "(probably <5,000 could be sufficient, depending on the frequency of dossiers with problems). While this may not present directly a problem with respect to the estimate, it may have two unintended effects:

- Fewer checks are focused on dossiers that do present a specific risk, thus missing opportunities to address risk
- Since it may be assumed that randomly sampled dossiers may be somewhat more expensive to organize, overall costs may increase.

4.1.6. The Commission has made no statement of what reliability this estimate should be. It was noted that, where the proportion of dossiers with problems approach low values (5% and below), the number of samples required to make this estimate at a given level of reliability/confidence increase geometrically. Thus, a member state which manages a) to keep levels of irregularities low, and b) keeps control levels constant, could find that the "background level" estimate is not stable or reliable.

4.1.7. In brief, the above trend is that a MS that should be able to benefit from low levels of irregularity actually needs to undertake more controls to determine that low level reliably.

4.2. Identification of risk factors (CART method)

4.2.1. The identification of potential risk factors among all possible factors characterizing a claim should be made on a sample (1) representative of the whole population of claims and (2) for which the probability of selection is known.

4.2.2. The random sample is suitable for this analysis provided it is large enough to cover the main types of applications (with a minimum of 100 applications per class for major classes after the classification process; it is estimated that a minimum sample of several hundreds of applications should be used for this analysis[15]). The RA sample (or part of this sample) could be suitable to complete the random sample provided it respects the two criteria defined previously (in particular, it should not be restricted to specific categories of claims). The use of this sample or part of it for the identification of risk factors still needs to be investigated.

4.2.3. Among the factors characterizing a claim, some are qualitative or categorical variables (e.g. existence of anomalies in the past controls, claim for crop X, closeness to minima or maxima, lodging of application after deadline, paper application without cross checks, claim for other schemes such as Less favoured Areas or Agri Environment Measures…) whereas others are continuous variables (e.g. amount of aid, number of parcels,…). The proposed method will identify criteria and thresholds for classifying the claims, therefore there is no need to convert continuous factors into categorical variables.

---

[15] In other words for MS with less than roughly 50 000 applicants, the assessment of the RA based on the sole random sample may be unreliable.

4.2.4. The analysis should assess the potential risk factors with respect to a given criterion. Since the “risk for the fund” is understood in monetary terms, it seems logical to take the rate of anomalies in terms of amount of aid not paid, if available in time, or of area not determined as assessment criterion for this analysis. In other words, the analysis should identify the categories of claims for which the area not determined or the aid not paid is the highest rather than the categories of claims for which the number of anomalies is the highest.

4.2.5. CART (Classification And Regression Trees) algorithms are suggested for assessing the factors (or combinations or factors) associated with the highest rates of anomalies. These algorithms work as tree classifiers: at each node of the tree, they split the population present at this node into two (or more) groups though the identification of the factor that best divide the claims in this population with respect to the assessment criterion (e.g. % of missing area): starting from the whole sample, the 1st node will divide this sample into at least 2 classes of claims, which in turn will be split into at least 2 classes and so on. The split process stops on the basis of a statistical test. In the end, the initial population is divided into n classes, each one corresponding to a combination of factors (e.g. “claims above x euros with request to access the national reserve and with claim for protein crops”; such a class would be identified after 3 nodes).

4.2.6. In practice it may be useful to rank the combinations of factors identified according to the selected criterion in descending order. The proportion of claims in the whole population which would correspond to the 1st combination may then be estimated (hence the need for a representative sample) and so on with each successive combination of factors.

4.2.7. Once the risk factors have been identified, two strategies may be used to select the RA sample from the whole population of claims (case possible for MS using classical inspections only or if a full VHR ortho coverage of the MS exists):

- Take all claims meeting combination 1, then combination 2 and so on till the targeted number of RA application is reached. The advantage of this strategy is that it should allow picking up the highest anomalies (assuming the sample used for the identification of the risk factors is representative and there is no change with respect to the previous year). However on the other hand this sample is not representative of the whole population of claims (since a number of combinations are left out of the sample) and will hence not be usable for next year’s RA assessment.

- Sample in each stratum or class with varying sample rates, for instance from 100% for combination 1 (highest risk, provided that a reasonable number of applications fall in this class) to 1% for combination n (lowest risk class containing the majority of the claims).

4.2.8. For MS using control zones, the previous strategy must be adapted e.g. following § 2.3.

5. References

1. Commission Regulation (EC) No. 796/2004 of 21 April, 2004 laying down detailed rules for the implementation of cross-compliance, modulation and the IACS provided for in the Council Regulation (EC) No. 1782/2003 establishing rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers.

2. European Commission (2004). Discussion document concerning Risk analysis for dossier selection, OTS and CwRS checks, Issue 2 In English, MARS Ref. JRC IPSC/G03/P/SKA/ska D(2004)(3698)
3. European Commission (2003). CwRS Site definition: demonstration of the use of LPIS, Olivier Leo, Csaba Wirnhardt, 9th conference on Controls with Remote sensing, Köln, 28-29 November 2003[16].
4. Leo Breiman, J. Friedman, R. Olshen, C. Stone, 1984, Classification and Regression Trees, Chapman & Hall, 368 pages.

(End of document)

[16] Presentation available on the following web page http://agrifish.jrc.it/marspac/DCM/default.htm