European Environment Agency

Accuracy assessment of the high resolution soil sealing map for France

MEEDDAT/SOeS
Philippe Dorelon
January 2009

Table of contents

1 CONTEXT....3

2 DELIVERABLES....3

3 METHODOLOGY....3

3.1 Preparation of data....4
3.2 Choice of areas to be checked....4

4 RESULTS....5

4.1 Calculation of errors....5
4.2 Calculation of associated confidence intervals....6
4.3 Error analysis....6
4.4 Discussion of calculated coefficients....7
4.5 Discussion of 80% threshold....7

5 FINAL APPRAISAL....8

1 Context

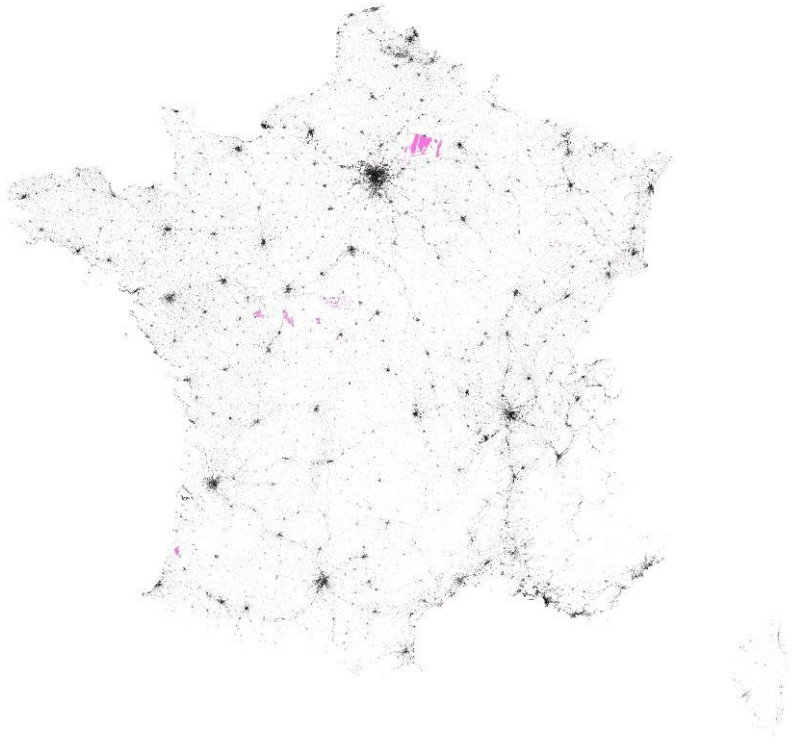
The High Resolution Soil Sealing database is an integral part of the Corine Land Cover 2006 project. The aim is to allow more accurate characterisation of areas with artificial cover not mapped in CLC.
The database was produced by a service provider selected by the EEA on the basis of images from the IMAGE2006 project, the same images having served for CLC 2006.
Each country is to check the parts relevant to its territory.

2 Deliverables

The data were provided by the service provider, Infoterra, in September 2008.
The characteristics are as follows:

- Raster database
- Resolution: 20 metres
- Projection: Lambert II étendu
- Classification :
    - From 0 to 100 : degree of imperviousness
    - 254: not classified
    - 255: no data

Overview of the database



Urban areas are visible on the map. Relatively large un-interpreted areas are also noticeable (in pink), not interpreted because of clouds in the satellite images.

3 Methodology

The methodology used is that presented at the Workshop on Quality Control and Validation of Land Cover Data (Copenhagen, 12−13 November 2007) and described in the document *Recommendations for Quantitative Assessment of High-resolution Soil Sealing Layer*, by G. Maucha and G. Büttner.

Assessment makes use of a 100 metre grid; areas with a coefficient greater than 80% are considered as having artificial cover.

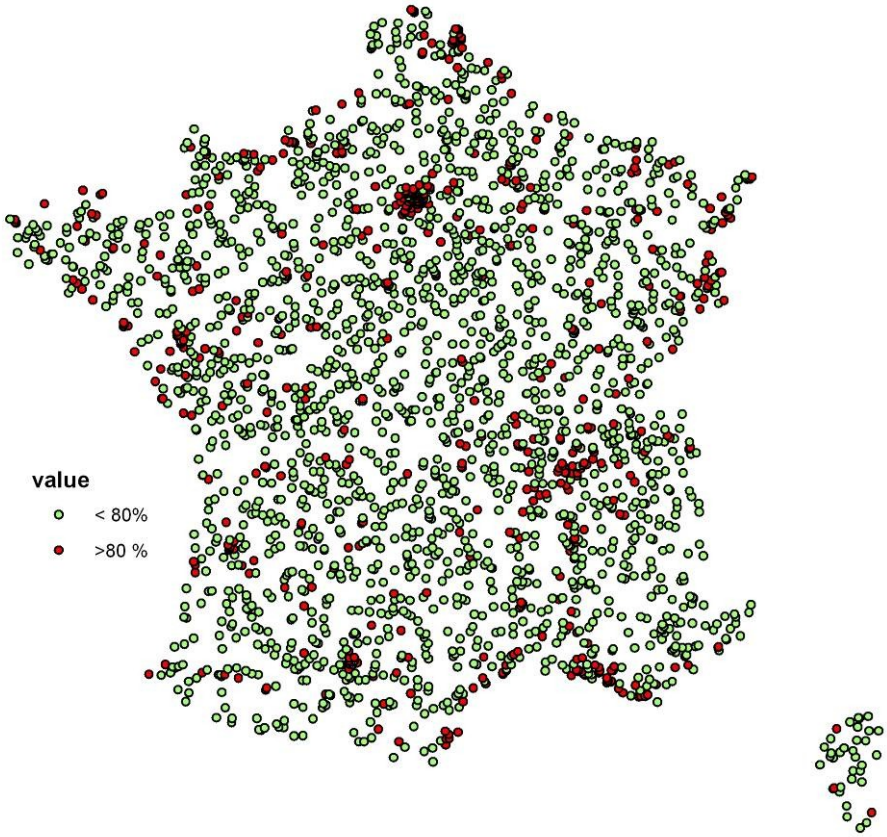
3.1 Preparation of data

The data are aggregated to form a 1 hectare (100 m x 100 m) grid by calculation of an average imperviousness coefficient. Areas with a not-interpreted or no-data pixel assume that value.

3.2 Choice of areas to be checked

The method stipulates selection of 2 500 areas to be checked, of which 500 have an imperviousness coefficient greater than 80%.
The selection draw was of the random, no-repeats type.

Overview of selected areas



Each area contains one hundred points to be classified as artificial cover or not; the control was blind, i.e. the operator did not know the value in the HR soil sealing base.
An area's coefficient is obtained by summing the number of artificial cover points.
The data used were:

- BD Ortho© IGN: priority data
  - Resolution: 0.5 m
  - Date of taking of photos: from 2001 to 2006
- Google Earth

The ArcGIS 9.2 software was used. Developments were made by the selected service provider to facilitate the work.

Example of an area to be checked

4 Results

The results matrix is as follows:

		Reference data			
		≥80%	<80%	Sum	User's accuracy
HR Soil Sealing	≥80%	269	231	500	53.8%
	<80%	6	1 994	2000	99.7%
	Sum	275	2225	2500	
		Overall accuracy		90.5%	

4.1 Calculation of errors

Proportion of sealed surfaces, calculated from complete data: $P_{class} = 7.9\%$

Proportion of commission errors: $P_{commission} = \frac{231}{500} = 46.2\%$

Proportion of omission errors $P_{omission} = \frac{6}{2000} \cdot \frac{1 - P_{class}}{P_{class}} = 3.5\%$

4.2 Calculation of associated confidence intervals

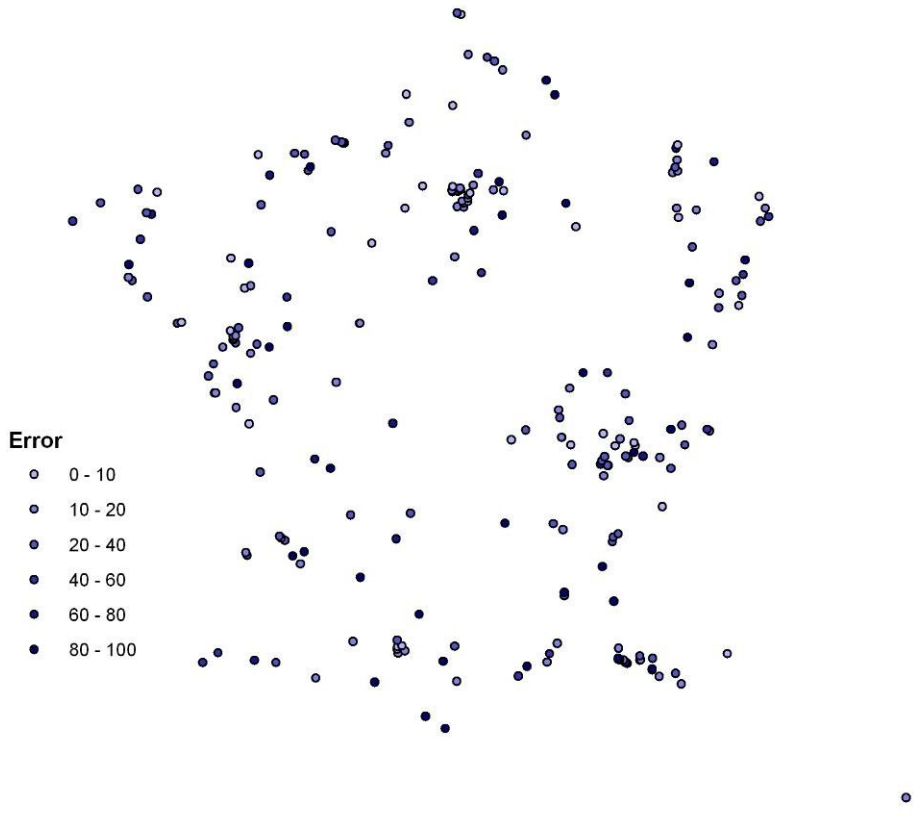
Error type	Number of all samples	Number of wrong samples	Statistical results of the validation				
			p_{min}	p_{max}	Mean error	Confidence interval	Probability of having more than 15% error real in the database
Commission	500	231	42.6%	49.8%	46.2%	±3.6%	100%
Omission	2000	6	1.2%	5.8%	3.5%	±2.3%	0%

The calculations for omission errors are incorrect as they include the proportion of artificial surfaces, widely over-estimated according to the checks.
Taking a conservative hypothesis of an error factor of two for the proportion of artificial surfaces, the level of omission errors would reach 7% and would remain significantly below 15%.

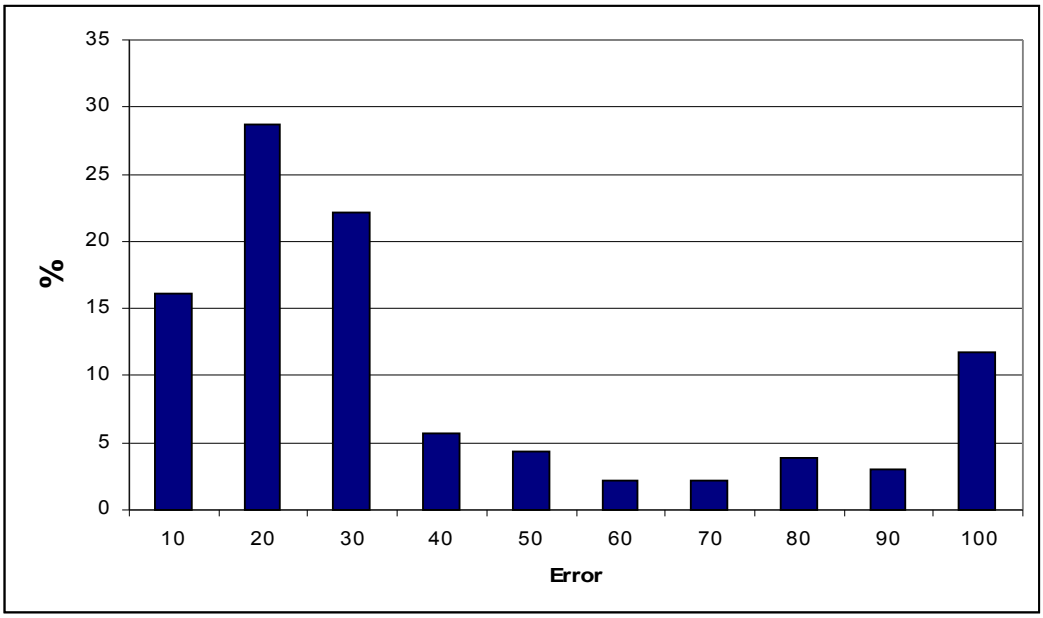
4.3 Error analysis

For France, errors are mostly commission errors, i.e. the base over-estimates artificial cover. There is no apparent spatial correlation of errors, they are distributed throughout the territory.

Overview of commission errors



Distribution of commission errors



The distribution is skewed towards small errors: 45% of errors are lower than 20 and 67% lower than 30, proving correlation between calculated and real values in spite of classification errors.
Conversely, 12% are greater than 90. In many cases, these large errors concern quarries or work sites, where the design of the product itself causes the problem: artificial cover or not?

4.4 Discussion of calculated coefficients

The results raise the question as to whether the survey method measures the same thing as the coefficient provided by the base.
Linear regression indicates a significant link between the two, and an equation of the following type:

$$Coeff_{reference} = 2.66 + 0.79 \cdot Coeff_{HRSoilSealing}$$

The 0.79 coefficient confirms the existence of a bias and unclearness as to what the base actually represents.

4.5 Discussion of 80% threshold

The artificial/non-artificial cover threshold, set at 80%, seems too high for France; the quality assessment has already indicated this.
With this threshold, the proportion of artificial cover is slightly less than 8%. Other surveys (Teruti, Lucas) estimate this proportion at 9%.
A lowering of this threshold would therefore give a more accurate picture of reality and would improve results.

A simulation in the light of the discussions above—lowering of threshold to 76%, taking account of bias (reference threshold at 62%)—gave the following results:

		Reference data			
		≥62%	<62%	Sum	User's accuracy
HR Soil Sealing	≥76%	409	94	503	81.3%
	<76%	19	1 978	1997	99.0%
	Sum	275	2225	2500	
		Overall accuracy		95.5%	

The level of commission errors is improved, even though it remains above 15%, and the level of omission errors remains contained.

5 Final appraisal

- The database characterises areas without artificial cover accurately, with an omission error level less than 15%.
- Conversely, commission errors are much too great: the database over-estimates areas with artificial cover by 46%, a level much greater than the 15% threshold
- However, the errors detected are often minor. Better definition of the 80% threshold and addition of a correcting coefficient could significantly improve the database.
- Similarly, the problem observed for quarries and work sites confirms the base's ambiguity regarding the concepts of artificial cover, built-up areas and soil sealing.
- Paradoxically, the results appear to contradict the qualitative validation which indicated many areas with artificial cover overlooked. Here again, the problem arises from the 80% threshold which is not appropriate for France. With a lower threshold, the artificial cover would be improved. The difficulty lies in establishing a satisfactory compromise between commission and omission errors.