INFLUENCE OF SPATIAL HETEROGENEITY AND SCALING ON LEAF AREA INDEX ESTIMATES FROM REMOTE SENSING DATA

S.GARRIGUES[1], D.ALLARD[2], M.WEISS[3], F.BARET[1],S.MARNI[4],H.JEANJEAN[5]

1. INRA, Avignon, France: sgarrig@avignon.inra.fr, baret@avignon.inra.fr
2. INRA-Biométrie, Avignon, France, allard@avignon.inra.fr
3. NOVELTIS, Toulouse, France, weiss@avignon.inra.fr
4. ALCATEL SPACE, Cannes, France, Sandrine.Marni@space.alcatel.fr
5. CNES, Toulouse, France, Herve.Jeanjean@cnes.fr

ABSTRACT –To monitor terrestrial surfaces, quantitative information like Leaf Area Index (LAI) are retrieved from remote sensing data. Because of technology constraints the sensors currently used have coarse resolution. Characterizing spatial heterogeneity at coarse resolution is important to improve LAI estimates. The aim of this study is to propose a methodology to characterize spatial structure of remote sensing data. Variogram models are fitted for different variables (RED and NIR reflectance, NDVI vegetation index and LAI) for four contrasted landscapes (crop, pine forest, Mediterranean vegetation and tropical forest). The comparison of the variogram model parameters (sill, range, integral range) between landscapes allows to draw a typology of spatial heterogeneity. Then spatial heterogeneity is related to scaling issue to emphasize the influence of resolution with application to the estimation of LAI.

1 INTRODUCTION

Terrestrial surfaces are observed frequently and globally by series of large swath sensors such as NOAA/AVHRR,VEGETATION, MERIS, MODIS, MISR and POLDER. To monitor phenological changes and dynamic processes such as primary production, studies have to be carried using high time frequency data. Because of technological constraints, these sensors are associated to coarse spatial resolution, *i.e*, in the resolution from few hundreds of meters up to few kilometers. However, at these scales, spatial heterogeneity may have a great influence on land surface characteristics estimation from remotely sensed data (Leaf Area Index: LAI, vegetation cover, canopy chlorophyll content…), particularly if the relationship between the considered variable and the radiometric data are non linear, which is the case for LAI. Indeed the transfer function used to derive LAI from remote sensing data (reflectance, NDVI) is generally built to be applied to homogeneous surface whereas the remote sensing measurement is a spatial average of radiative signals over the pixel. Applying such transfer functions at coarse resolution to derive LAI does not account for the subpixel variation and leads to error on retrieving quantitative information LAI. Therefore the spatial heterogeneity issue has to be dealt with the problem of scaling and the non linearity of the transfer function.

Depending on the kind of study, spatial heterogeneity can have different meanings. In this work, spatial heterogeneity is related to a quantitative information that characterizes the spatial structure of remote sensing data (vegetation description) for a given geographic domain (ie image). Spatial heterogeneity can be quantitatively characterized by different techniques. A first way consists in computing textural parameters (variance, covariance, skewness..) to characterize spatial data variation within an image (Haralick,1996). Different authors show that geostatistical theory (Cressie, 1991) can be applied to remote sensing image to describe data spatial variation by variogram function (Woodcock *et al* 1987, Jupp *et al.* 1988a, Jupp *et al.* 1988b, Woodcock *et al* 1988, Atkinson 1997, Milne *et al.* 1999, Treizt 2000, Beaufort 2000…). Fractal dimension (Mandelbrot, 1983) was used to monitor spatial pattern change in multiscale analysis (De Cola 1989, Bian and Walsh 1993, Bian 1997, Xia *et al.* 1997, Cao 1997…). More recently wavelet transform is used in multiresolution analysis (Ranchin and Wald, 1993). However Chen (1999) underlines that contextural parameters (size and organization of the objects within the image) are more effective than textural parameter to describe the effect of spatial discontinuity on scaling. Moreover different studies characterize spatial heterogeneity with scaling issue. Bian (1997) uses fractal dimension to emphasize data aggregation influence on spatial pattern. Milne *et al.* (1999) used geostatistical theory to monitor data regularization evolution with resolution. Spatial heterogeneity is also studied as an influence on LAI error estimates at coarse resolution. Chen (1999) shows that the error estimates depends not only on the non linearity of the derivation algorithm but also on the mixed pixel spatial pattern organization. Friedl (1997) shows with a simulation study that the LAI error estimates results from the interaction between spatial resolution of the sensor and the scale of spatial variation in the ground scene.

This study consists in developing a methodology to characterize quantitatively spatial heterogeneity for different variables (reflectance or biophysical variable derived from remote sensing) for different landscapes. To this purpose, the study objectives are:

1. describing spatial heterogeneity at high resolution (20m resolution) for different vegetation sites and variables
2. analyzing scaling effect in relation with spatial heterogeneity
3. observing heterogeneity influence on LAI error estimates at coarse resolution.

2 STUDIED DATA

2.1 Studied sites

The data come from the VALERI database (http:\www.avignon.inra.fr\valeri, Baret *et al.* 2002). It consists in a network of sites used to validate large swath satellite biophysical products (including LAI). For each site, high resolution images (SPOT) are available. Four SPOT images of 4 contrasted landscape are used. The size of each image is the same: 3km*3km and the SPOT pixel resolution is 20m. Figure 1 shows the NDVI image at SPOT resolution for the different sites.

2.2 Studied variables

a) SPOT reflectances

NIR and RED bands will be investigated. Moreover the NDVI index is computed:

$$NDVI = \frac{NIR - RED}{PIR + RED} \qquad (1)$$

The remote sensing data are not corrected from atmospheric effects, a cloud mask was applied only on Counami image, the acquisitions for the other sites being cloud free.

b) Biophysical variable

A LAI image of each site is performed by applying a semi-empirical expression (Baret and Guyot, 1991). The relationship is fitted on a learning data set composed of LAI variable and corresponding reflectance values. The data set was generated with SAIL radiative transfer model simulations (Weiss, 2002) for a wide range of simulations.

$$LAI = \frac{\log((NDVI - NDVI_{\infty})/(NDVI_s - NDVI_{\infty}))}{-K_{lai}} \qquad (2)$$

- K_{lai} is the extinction coefficient
- $NDVI_{\infty}$ is the asymptotic value of NDVI (corresponding to infinite LAI)
- $NDVI_s$ is the bare soil NDVI value

For each variable (LAI, NDVI, RED, NIR) and each site, different resolution images are computed by aggregating (arithmetic mean) the 20m resolution image.

3 METHODOLOGY

3.1 Geostatistical tools

Geostatistics (Cressie 1991, Wackernagel 1995) allow to characterize spatial distributions of one or more variables. Variable spatial distribution is characterized by the variogram function that measures the spatial dependence of neighbouring observations. Since the variable describes a spatial phenomenon, this variable is called regionalized variable. In our case the variable are reflectance or LAI derived from reflectance. Since reflectance values of an image are a function of spatial position they can be considered as regionalized variables. For this study, the variable support is the SPOT pixel that is a 20m*20m square. Although the value associated to the pixel is a spatial average of radiative signals over the square we will consider the value associated to the pixel as punctual. This induces the assumption of homogeneous SPOT pixels (intrinsic characteristics of VALERI site). Moreover since the pixel value results from a spatial averaging process (Point Spread Function) the remote sensing variable can be considered continuous on the image. Spatial variation and spatial dependence are described by the experimental semivariogram $\tilde{a}_e(h)$ (commonly called experimental variogram) which is the mean squared deviation of a variable at locations separated by a given lag distance:

$$g_e(h) = \frac{1}{2*N(h)} * \sum_{i=1}^{N(h)} \left(z(x_i) - z(x_{i+h})\right) \qquad (3)$$

$z(x_i)$ is the variable at location i, $z(x_{i+h})$ is the variable value at lag h from x and N(h) is the number of pairs of points separated by the distance h. Figure 2 presents the experimental variograms performed over the 3km*3km SPOT image for different variables and different sites. Two main parameters characterize the variogram:

- The total sill (σ^2) corresponds to the true variance of the data.
- The range (r) is the lag at which the variogram reaches the sill. Up to this distance data are spatially autocorrelated,

beyond this distance samples are spatially independent .

As the regionalized variable z(x) is difficult to model with a simple deterministic function, it is seen as an outcome of a random function Z(x). In this context and assuming a second-order stationarity hypothesis for Z(x), the theoretical variogram describing the spatial dependence and spatial variation of the random function is introduced:

$$g(h)=0.5*E\left[\left(Z(x)-Z(x+h)\right)^2\right] \qquad (4)$$

In practice a theoretical model is fitted on the experimental variogram.

3.2 Theoretical variogram models:

Two kinds of basic variogram functions were used:

- *spherical model*:

$$g(h)=\sigma^2\left[1.5\left(\frac{h}{r}\right)-0.5*\left(\frac{h}{r}\right)^3\right] \text{ for } h<r \qquad (5)$$

$$=s^2 \qquad \text{for } h\geq r$$

- *exponential model*:

$$g(h)=\sigma^2\left[1-\exp\left(\frac{-3h}{r}\right)\right] \qquad (6)$$

The fitted models used here are nested structure, *i.e.,* weighted sums of one or two simple models cited above. So each model is characterized by two or four parameters (σ_1^2, σ_2^2, r_1, r_2). For each site and for each studied variable (RED,NIR,NDVI,LAI) one model is fitted (figure 2).

4 HETEROGENEITY CHARACTERIZATION

4.1 Vegetation type influence on spatial structure

The NDVI variograms allow to draw spatial structure differences between sites. Indeed the total sill ($\sigma_1^2+\sigma_2^2$) value indicates the true variance and gives a first heterogeneity criterion. The Counami site (tropical forest) has the lowest sill value and is the most homogeneous site whereas the Alpilles site (crop) presents the highest value and appears as the most heterogeneous site. On the Counami site, the first structure has a very short range (70m) and accounts for most of the variance (σ_1^2 is 85% of the total sill). It can be considered that there is almost no spatial correlation after 70m. Hence it is the most spatially homogeneous site. The other sites present a first range at larger distances, between 200m and 300m. The Alpilles site and the Nezer site have patchy spatial patterns. Their first range variogram can be associated to the intra-field variability and the discontinuities between fields. Moreover it might represent the mean size of the fields. On the one hand, on Alpilles, differences between field values is particularly high with a succession of crop field with high NDVI value and bare soil field with low NDVI values leading to a very high sill. On the other hand, Nezer presents a NDVI value distribution between fields more homogeneous leading to lower sill. Moreover the size and the shape of the fields are more regular for the Nezer site than those of the Alpilles site. This confirms Chen (1999) results, *i.e.*, the importance of accounting for contextural parameter (spatial pattern organization) to characterize spatial heterogeneity. Finally it is more difficult to characterize the second range of the model. It can represent:

- a large scale continuous phenomenon : for agricultural site like Alpilles it can be soil properties, for Puechabon the north west part presents different geomorphologic characteristics than the rest of the site.
- some image singularities: Puechabon presents a quarry at the north part of the site.

The variogram model was computed on the whole 3km*3km site without any mask. So it reflects the real spatial structure of the site including singularities like quarry, river... Variogram model do not reflect only the spatial structure of the vegetation making the site comparison more difficult. In addition some noise linked with the remote sensing chain processing can influence the variogram ranges (r_1,r_2): remaining non detected clouds, atmospheric effect, topography or view angles.

4.2 Integral range: a parameter derived from the theorical variogram

The integral range (A) (Chiles,1999) is a yardstick that summarizes the variogram range.

$$A=\left(\frac{1}{s^2}\right)*\int_{R^2} C(h)dh=\int_{R^2}\left(1-\frac{g(h)}{s^2}\right)dh \qquad (7)$$

where C(h) is the covariance function defined by:

$$C(h)=C(0)-g(h) \qquad (8)$$

By plotting the total sill of each model versus the integral range for the NDVI of each site, we can draw a first heterogeneity typology (figure 3): spatial heterogeneity increases from the tropical forest site and the pine forest site to the Mediterranean vegetation site and the agricultural site.

4.3 Variable effect on spatial structure

Figure 2 does not show important differences between NDVI spatial structure and LAI spatial structure. No LAI variogram was fitted for the Counami site because of saturation problem with transfer function applied at high NDVI values. The main difference is between reflectance (NIR, RED) variables and NDVI. The first structure (r_1) is detected in both reflectance and NDVI variograms, but the level of variability (total sill) and the second range (r_2) scale differ. One of the reasons could be that NDVI characterizes the amount of surface vegetation whereas the reflectance variables are affected by several other factors. Therefore, it is more relevant to use NDVI variable to describe vegetation spatial structure.

5 SCALING EFFECT

5.1 Statistical parameters evolution with scaling

A first approach to explore scaling effect is to observe the evolution of the NDVI value distribution at different ranges of resolution. Figure 3 shows the NDVI histograms for the four sites at four resolutions: 20m, 100m, 300m, 500m. If the NDVI values were independent, standard results in statistics would show that the standard deviation of block averages would be inversely proportional to their area and that the distribution, properly standardized, would tend to a Normal distribution. However Alpilles site standard deviation decreases by 32 % from 20m resolution to 300m resolution whereas Counami site standard deviation decreases by 60 % from 20m resolution to 300m resolution. Except for Counami, the distribution shows low symmetrization with scaling up. These differences between NDVI distribution evolution through scaling are explained by the difference between spatial autocorellation range. Thus, to understand the scaling effect, the spatial structure has to be accounted for.

5.2 Dispersion variance

a) Dispersion Variance definition

An other way to explore the evolution of the spatial structure with resolution consists in observing the variable variance evolution of 20m SPOT pixel (x) within coarser block size (v). Calling V the entire domain image (3000m*3000m), V is decomposed in a union of n congruent subregions v (v1…vn). We respectively call Z(V) and $Z(v_i)$ the average variable value over V and the block v_i :

Spatial variability can be compared at three different scales (Friedl, 1997):

1. the variation of the pixel value (x) with respect to the regional average value on the whole image (V)

$$S^2(x,V) = \frac{1}{V}\sum_x (Z(x) - Z(V))^2 \qquad (9)$$

$\sum_x$ represents the pixel sum over the whole image.

2. the variation of subregional average value (v block value) with respect to the regional average value on the whole image (V)

$$S^2(v,V)=\frac{1}{n}\sum_{i=1\ldots n}(Z(vi)-Z(V))^2 \qquad (10)$$

3. the average variation of the pixel values within block v

$$S^2(x,v)=\frac{1}{n}\sum_{i=1..n}\frac{1}{v}\sum_x\left(z(x)-Z(vi)\right)^2 \qquad (11)$$

$S^2(x,V)$ corresponds to the variance of the whole image, that can be seen as the variance of pixel values within blocks plus the variance of block values within the image (Krige relation, Wackernagel,1995):

$$S^2(x,V)=S^2(x,v)+S^2(v,V) \qquad (12)$$

We called $S^2(x,v)$ the experimental dispersion variance of x within v. A theorical dispersion variance can be derived from the variogram model and is obtained by taking the expectation (Wackernagel,1995):

$$s^2(x,V)=E(S^2(x,V))=s^2-C(v,v)=g(v,v) \qquad (13)$$

b) Dispersion variance evolution with resolution

Figure 5 presents the evolution of experimental and theoretical dispersion variances versus the aggregation scale. As expected, the dispersion variance increases with the resolution. The comparison of the dispersion variance evolution with resolution for the four sites confirms the first observations made in §3 . Besides it allows to makes inferences on data regularization, *ie*, with increasing resolution, data becomes more spatially continuous which is reflected by changes in geostatistical properties (lower sill, more regular variogram at the origin, dispersion variance reaching the sill):

- Counami being the most homogeneous site, its dispersion variance reaches a sill very rapidly: increasing the support leads to a rapid data regularization .
- The other sites do not reach a sill at 1000m resolution, meaning that no total regularization is reached at this scale, *ie*, data are still spatially autocorrelated. Alpilles is the most heterogeneous site, presenting high dispersion variance value.

Puechabon site presents an important regularization at short distance.

Finally, the quite good fitting of the theoretical dispersion variance on experimental dispersion variance confirms the model fitting performances.

6 INFLUENCE OF HETEROGENEITY ON LAI ERROR ESTIMATES

The LAI variable is derived from remote sensing data by a semi empirical expression (LAI=f(NDVI). However this transfer function derived from SAIL simulation is strictly valid for homogeneous surfaces. To obtain the "integrated" LAI_i value at coarser resolution, the right way consists in aggregating the LAI computed at the higher spatial resolution (20m):

$$LAI_i=\int f(ndvi) \qquad (14)$$

However the LAI at coarse resolution can be approximated by applying the transfer function f over coarse resolution NDVI data, *ie*, for values aggregated at the coarse resolution: the "mean" LAI_l or "lumped" LAI is obtained as:

$$LAI_l=f(NDVI_i) \qquad (15)$$

Where $NDVI_i$ is the integrated NDVI at coarse resolution.

Because of the non linearity of the transfer function an error is made in computing the lumped LAI_l. This error can be evaluated by the root mean square difference between the lumped variable and the distributed variable at a given resolution. Figure 6 presents a plot between the dispersion variance and the RMSE computed between lumped and integrated LAI for different resolutions. By increasing the resolution, the RMSE is increasing linearly with the dispersion variance in log-log scale. The four sites have RMSE ranges that increase according to their spatial heterogeneity degree confirming the spatial heterogeneity typology made on §3.2. However for the Puechabon site the RMSE value decreases at 1000 m resolution. Two hypothesis arise: On the one hand, a RMSE computed with 9 samples (at 1000m resolution) may not be statistically consistent. On the other hand, at 1000m resolution the Puechabon NDVI image is not stationary with some very homogeneous pixels and very heterogeneous pixels. Therefore since the NDVI range value changes with increasing resolution, the degree of non linearity of the transfer function is also changing and can influence the LAI error estimates. In conclusion, dispersion variance is a relevant parameter that reflects spatial heterogeneity importance at a given resolution and could be used to assess LAI error.

7 CONCLUSIONS AND PROSPECTS

Biophysical variable estimates error results from the non linearity of the transfer function and the spatial heterogeneity of the data. It will be important to account for these two features to improve biophysical variable estimate at coarse resolution. This study shows that geostatistical methodology allow to describe the spatial structure characteristics of different landscapes. By computing variograms, it is possible to make some inference on spatial heterogeneity between vegetation sites and to link the data autocorrelation with scaling issues. Dispersion variance provides an indication on the degree of heterogeneity for a given resolution. This parameter will be useful to account for spatial heterogeneity in biophysical variable estimates.

REFERENCES

[1] Aman, A., Randriamamantena, H P., Podaire, A., Frouin, R. (1992), Upscale Integration of Normalized Difference Vegetation Index: The Problem of Spatial Heterogeneity. *IEEE Transactions on Geoscience and Remote Sens., 30 (2):326-338.*

[2] Atkinson, P.M. (1997), On estimating measurement error in remotely sensed images with the variogram. *Int. J. Remote Sens., 18:3075-3084.*

[3] Baret, F. and Guyot, G., (1991), Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens. Environ., 35:161-173.*

[4] Baret, F, Weiss, M., Garrigues, S ., Allard, D., **et** al., (2002), VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite product. *Remote Sens. Environ., Submitted.*

[5] Bian, L. and Walsh,, S.J. (1993), Scale dependencies of vegetation and topography in mountainous environment of Montana. *Prof. Geogr., 45,1.*

[6] Bian, L. (1997), Multiscale Nature of Spatial Data in Scaling Up Environmental Models. *In Scale in Remote Sensing and GIS. Quattrochi and Goodchild Eds, Lewis Publisher, pp 13-26 .*

[7] Cao,C. and Lam, N S. (1997), Understanding the Scale and Resolution Effects in Remote Sensing and GIS. *In Scale in Remote Sensing and GIS. Quattrochi and Goodchild Eds. Lewis Publisher, pp 57-72.*

[8] Chen, J.M. (1999), Spatial Scaling of a Remotely Sensed Surface Parameter by Contexture. *Remote Sens. Environ., 69 :30-42.*

[9] Chilès, J.P. and Delfiner P. (1999), Geostatistics : Modeling Spatial Uncertainty. *Wiley, New-York*.

[10] Cressie, N. A. C. (1991), Statistics for Spatial Data, *Wiley, New York.*

[11] De Beaufort, L. (2000), Definition d'une méthode de cartographie d'indice foliaire destinée à la validation de produits de capteurs satellites. *Master report from ENSAR engineering school.*

[12] De Cola, L. (1989), Fractal analysis of a classified Lansat scene. *Photogr. Eng. Remote Sens., 55(5):601-610..*

[13] Friedl, M A. (1997), Examining the effects of Sensor Resolution and Sub-Pixel Heterogeneity on Spetral Vegetation Indices: Implications for Biophysical Modeling. *In Scale in Remote Sensing and GIS., Quattrochi and Goodchild Eds., Lewis Publisher, pp 113-140 .*

[14] Haralick, R.M. (1986), Statistical image texture analysis. *In Handbook of Pattern Recognition and Image Processing, T.Y. Yong and K.S. Fu, Eds., Academic, New York, pp 247-280.*

[15] Jupp, D. L. B., Strahler, A.H and Woodcock, C.E (1988a), Autocorrelation and regularization in digital images I: Basic Theory. *IEEE Trans. Geosci. Remote Sens., 26: 463-473.*

[16] Jupp, D. L. B., Strahler, A.H and Woodcock, C.E (1988b). Autocorrelation and regularization in digital images II: Simple image models. *IEEE Trans. Geosci. Remote Sens., 27: 247-258.*

[17] Mandelbrot, B. B. (1983), The fractal geometry of nature. *W.H Freeman and Company, New York.*

[18] Milne, B. T. and C., W. B. (1999), Multiscale Assessment of Binary and Continuous Lancover Variables for MODIS Validation, Mapping and Modelling Applications. *Remote Sens. Environ., 70: 82-90.*

[19] Myers, D E. (1997), Statistical Models for Multiple-Scaled Analysis. *In Scale in Remote Sensing and GIS., Quattrochi and Goodchild Eds,Lewis Publisher, pp 273-294 .*

[20] Ranchin, T. and Wald, L. (1993), The wavelet transform for the analysis of remotely sensed data. *Int. J. Remote Sens., 14(3) :615-619.*

[21] Treizt, P. (2001), Variogram Analysis of High Spatial Resolution Remote Sensing Data: An Examination of Boreal Forest. *Int. J. Remote Sens., 22(18): 3895-3900.*

[22] Wackernagel, H. (1995), Multivariate Geostatistics., *Springer-Verlag, Berlin, Heidelberg, 256p.*

[23] Weiss, M. and Baret, F. (1999), Evaluation of Canopy Biophysical Variable Retrieval Performances from the Accumulation of Large Swath Satellite Data. *Remote Sens. Environ., 70:293-306.*

[24] Weiss, M., Baret, F., Leroy, M., Hautecoeur, O., Bacour, C., Prevot, L., and Bruguier, N. (2002), Validation of Neural Net techniques to estimate canopy biophysical from remote sensing data. *Agronomie : Agriculture and Environnement, In press.*

[25] Woodcock, Strahler, Jupp (1988). The use of variograms in remote sensing: a/ scene models and simulated images b/ Real digital images. *Remote Sens. Environ., 25: 323-379*.

[26] Woodcock, Strahler. (1987). The factor of Scale in remote sensing. *Remote Sens. Environ., 21: 311-332.*

[27] Xia, Z-G. and Clarke,C. K. (1997) Approaches to scaling og Geo-Spatial Data. *In Scale in Remote Sensing and GIS., Quattrochi and Goodchild Eds,Lewis Publisher, pp 309-360.*

APPENDIX

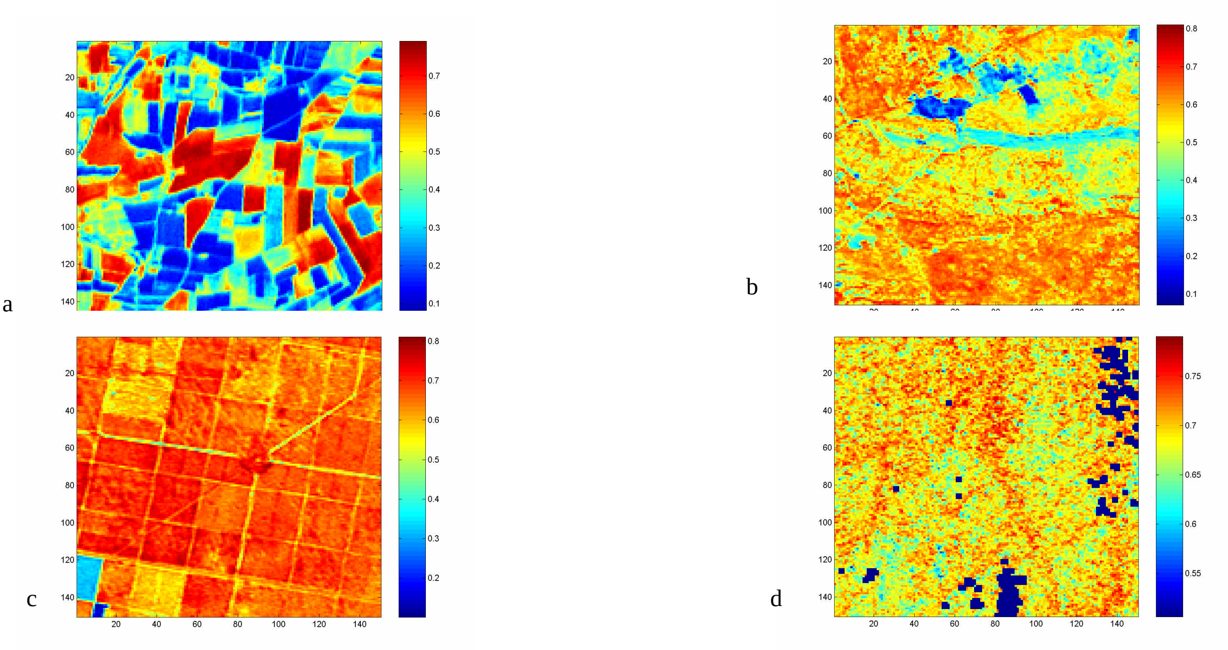

Figure1: NDVI images of the studied sites (3km*3km): a Alpilles, Crop Site (March), mean NDVI=0.41, std NDVI=0.19; b: Puechabon (France), Mediterranean Forest (June), mean=0.54, std=0.1; c: Nezer (France), Pine Forest (June), mean=0.65, std=0.06; d: Counami (French Guyana), Tropical Forest (October), mean=0.69, std=0.029

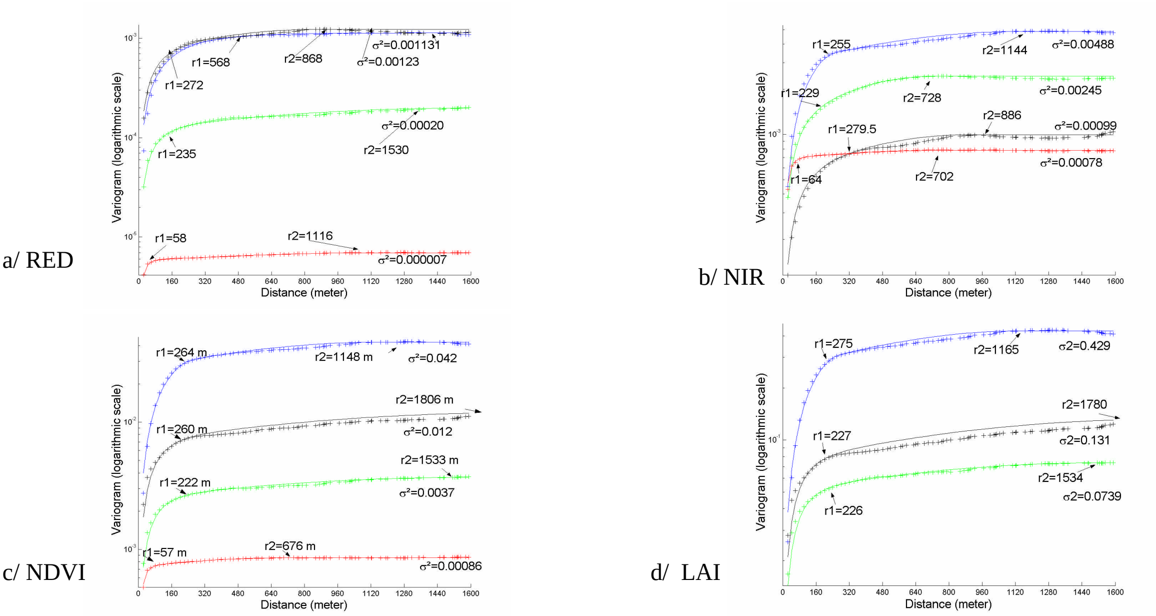

Figure 2: Experimental (crosses) and theoretical variograms (line) for the 4 landscapes studied: Alpilles (red), Puechabon (black), Nezer (green), Counami (red)

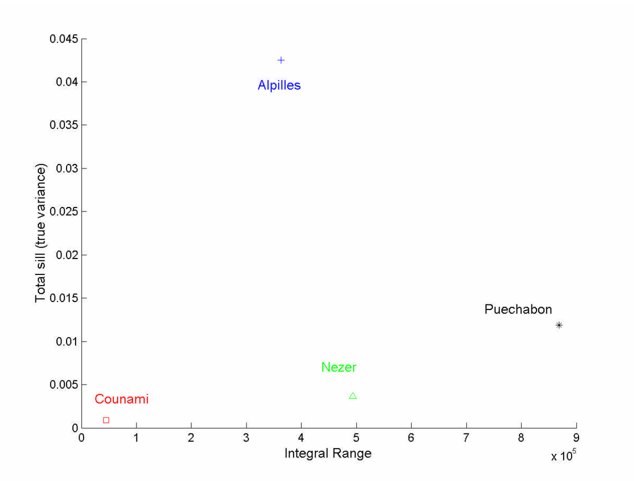

Figure 3: Integral Range versus total sill (σ²) for NDVI varial

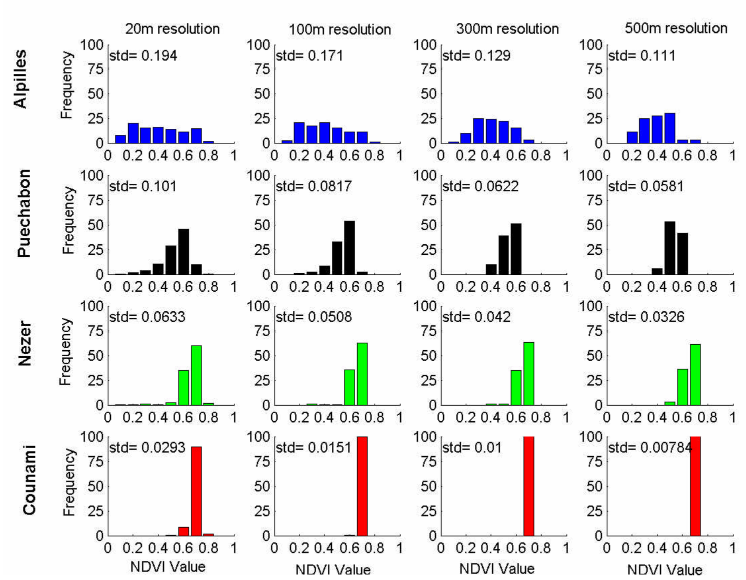

Figure 4: NDVI distribution for different sites at different resolutions

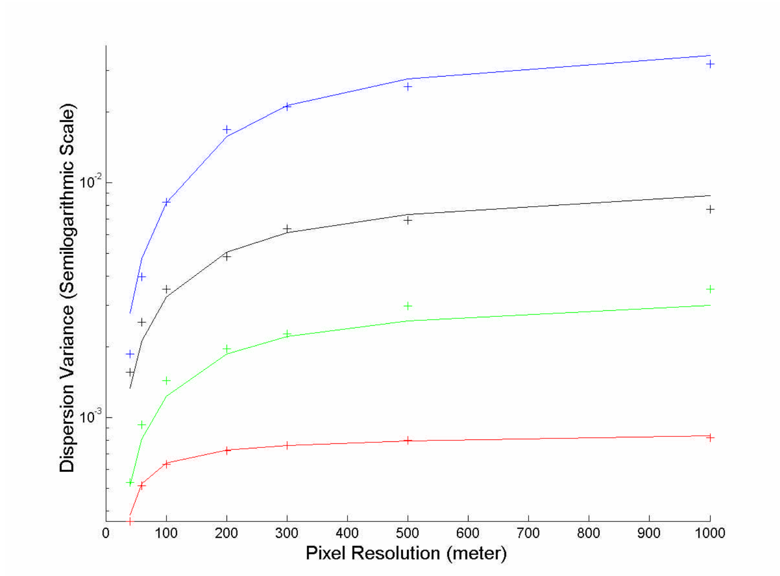

Figure 5: Dispersion Variance versus resolution

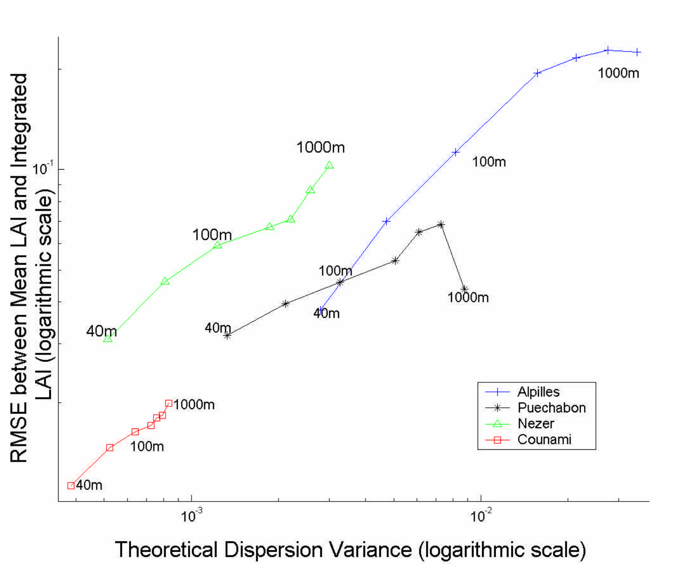

Figure6: RMSE between integrated LAI and mean LAI versus dispersion variance .