

**A classification of spores by support vectors
based on an analysis of their ornament spatial distribution –
An application to Emsian miospores from Saudi Arabia**

**[Classification de spores assistée par vecteurs de support
basée sur l'analyse de la distribution spatiale des ornements –
Une application à des miospores emsiennes d'Arabie Saoudite]**

Pierre BREUER¹

Godefroid DISLAIRE²

John FILATOFF³

Eric PICARD⁴

Philippe STEEMANS⁵

Citation: BREUER P., DISLAIRE G., FILATOFF J., PIRARD E. & STEEMANS P. (2007).- A classification of spores by support vectors based on an analysis of their ornament spatial distribution – An application to Emsian miospores from Saudi Arabia. *In*: STEEMANS P. & JAVAUX E. (eds.), Recent Advances in Palynology.- [Carnets de Géologie / Notebooks on Geology](#), Brest, Memoir 2007/01, Abstract 02 (CG2007_M01/02)

Key Words: Phylogeny; miospores; image analysis; sum and difference histograms; support vector classification

Mots-Clefs : Phylogénie ; miospores ; analyse d'images ; histogrammes des sommes et différences ; classification de vecteurs à support

Introduction – Geological setting

Continuous morphological intergradations exist between two trilete spore taxa from the Jauf Formation (Early Devonian) of the Widyan and Tabuk basins (northwestern Saudi Arabia). The alternation of siliciclastics and carbonates in this unit has been used to subdivide it into five members: from bottom to top, the Sha'iba,

Qasr, Subbat, Hammamiyat and Murayr members. The Jauf Formation in northwestern Saudi Arabia was deposited in a nearshore environment (AL-HAJRI *et alii*, 1999; AL-HAJRI & OWENS, 2000).

The latest study of miospores (BREUER *et alii*, 2005a, in press) suggests that the Jauf Formation is late Pragian to Emsian in age.

¹ Laboratoire de Paléobotanique, Paléopalynologie et Micropaléontologie, Université de Liège, Allée du 6 août, B18, Sart-Tilman, 4000 Liège (Belgium)
piet79@yahoo.fr

² Secteur de Géorressources Minérales et Imagerie Géologique, Université de Liège, Chemin des chevreuils 1, B52, Sart-Tilman, 4000 Liège (Belgium)
godefroid.dislaire@gmail.com

³ Geological Survey of Western Australia, Department of Industry and Resources, 100 Plain Street, East Perth, 6004 (Australia)
john.filatoff@aramco.com

⁴ Secteur de Géorressources Minérales et Imagerie Géologique, Université de Liège, Chemin des chevreuils 1, B52, Sart-Tilman, 4000 Liège (Belgium)
eric.pirard@ulg.ac.be

⁵ Laboratoire de Paléobotanique, Paléopalynologie et Micropaléontologie, Université de Liège, Allée du 6 août, B18, Sart-Tilman, 4000 Liège (Belgium)
p.steemans@ulg.ac.be

Manuscript online since March 22, 2007

Additional biostratigraphic evidence is provided by other fossil groups collected in outcrop (e.g. BOUCOT *et alii*, 1989; FOREY *et alii*, 1992). Among them, trilobites and conodonts indicate that the uppermost Sha'iba and Qasr members (lower Jauf Formation) are Pragian-early Emsian in age and brachiopods suggest that the Hammamiyat Member (upper Jauf Formation) is late Emsian in age.

The palynological material of this report comes from boreholes previously studied (see BREUER *et alii*, 2005a, in press). They are in two discrete areas about 350 km apart. Two of them (BAQA-1 and BAQA-2) are near Baq'a in the Widyan Basin, while JNDL-4 is near Domat Al-Jandal in the Tabuk Basin. The stratigraphic levels encountered in each locality overlap, and the correlations easily established using lithologic and wireline logs are confirmed by palynological data reported by BREUER *et alii* (2005a, in press). BAQA-1 and BAQA-2 cover an interval from the Sha'iba Member, through the Qasr and Subbat members to the lowermost Hammamiyat. The succession at JNDL-4 represents the upper part of the Subbat and the Hammamiyat Member.

Samples from BAQA-1, BAQA-2, and JNDL-4 were prepared in the Palynological Research Facility of the University of Sheffield. For this study additional slides from BAQA-1 were processed in the Laboratory 'Paléobotanique, Paléopalynologie et Micropaléontologie' of the University of Liège. All samples were prepared using standard palynological acid maceration techniques. A vast majority of the samples were productive and contain well-preserved organic matter. All material is housed in the Centre for Palynology of the University, Department of Animal and Plant Sciences, University of Sheffield and in the collections of the laboratory of 'Paléobotanique, Palynologie et Micropaléontologie', University of Liège.

Lineage

Palynologists have discussed only rarely the phylogenetic evolution of miospores in Palaeozoic sediments (e.g. VAN DER ZWAN, 1979; MARSHALL, 1996; MAZIANE *et alii*, 2002; BREUER *et alii*, 2005b). Some authors have demonstrated a continuous morphological intergradation among some dispersed spores that previously were attributed to discrete species. The morphological signal presented by miospores may not reflect only biological evolution, because it may be influenced locally by other parameters such as state of preservation, sedimentary sorting (JAGER, 2004) and/or reworking (BREUER *et alii*, 2005b).

A possible phylogenetic lineage including several morphotypes (Fig. 1), one of which has been described in BREUER *et alii* (in press), is proposed here. The two end-members of this lineage are well differentiated; however all intermediary forms are present in the assemblages. The morphological variation concerns distal ornamentation. This ornamentation involves changes in the shape of small cones and their spatial distribution that vary progressively between the two end-members of the lineage. In the simplest morphotype ornaments are evenly distributed on the distal surface. In the intermediary morphotypes they combine progressively to form a pseudo-reticulum, the walls of which are constituted by the discrete ornaments. In the most complex morphotypes, ornaments merge to form elongated ridges which describe a completely closed reticulum. Thus a progressive change in ornamentation occurs: from the simplest spores (ornaments constituted by evenly distributed small cones) to the most complex ones (ornaments forming a reticulum).

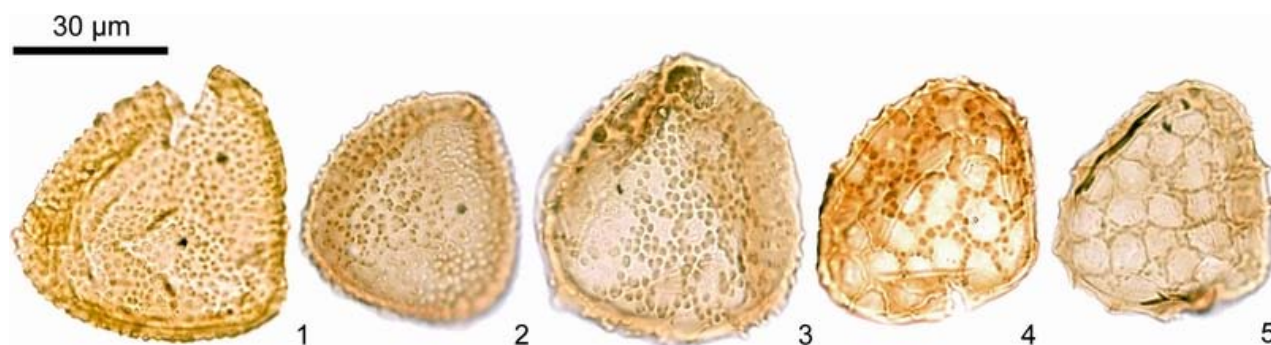


Figure 1: Microphotographs of the different morphotypes of the lineage.

Biostratigraphy

From a stratigraphic point of view, the range of the lineage is restricted to the Jauf Formation. Specimens occur first in the upper part of the Sha'iba Member and disappear in the upper part of the Hammamiyat Member. The spore assemblages from the upper part of the Sha'iba Member are considered to be in the PoW Oppel Zone of STREEL *et alii* (1987), based

on the general characteristics of the assemblages and the presence of typical taxa such as *Brochotriletes foveolatus* NAUMOVA, 1953, *Clivosispora verrucata* MCGREGOR, 1973, *Dictyotriletes emsiensis* (ALLEN) MCGREGOR, 1973, *D. subgranifer* MCGREGOR, 1973, and *Verrucosisporites polygonalis* LANNINGER, 1968 (Fig. 2) (BREUER *et alii*, 2005a, in press). The presence of *D. subgranifer* may indicate that they represent to the uppermost interval of the

zone (Su Interval Zone) of the PoW Opper Zone. The PoW Opper Zone is of Pragian-earliest Emsian age, with the Su Interval Zone encompassing the latest Pragian-earliest Emsian. In the Qasr Member, the stratigraphically important spores *Emphanisporites schultzi* MCGREGOR, 1973, and ? *Knoxisporites riondae* CRAMER et DÍEZ, 1975 (Fig. 2) make their first appearance. Similar spore assemblages are recovered throughout the Qasr Member and lower part of the Subbat Member. These assemblages are typical of those of Emsian age elsewhere and probably

belong to either the AB or to the lower FD Opper Zones. These assemblages are constrained above by the first appearance of *Rhabdosporites minutus* TIWARI et SCHAARSCHMIDT, 1975 (Fig. 2) in the upper part of the Subbat Member. Its occurrence marks the base of the FD Opper Zone (Min Interval Zone). The spore assemblages of the Hammamiyat Member are essentially similar throughout and also belong to the Min Interval Zone. That signifies that these assemblages are mid Emsian in age (BREUER *et alii*, 2005a, in press).

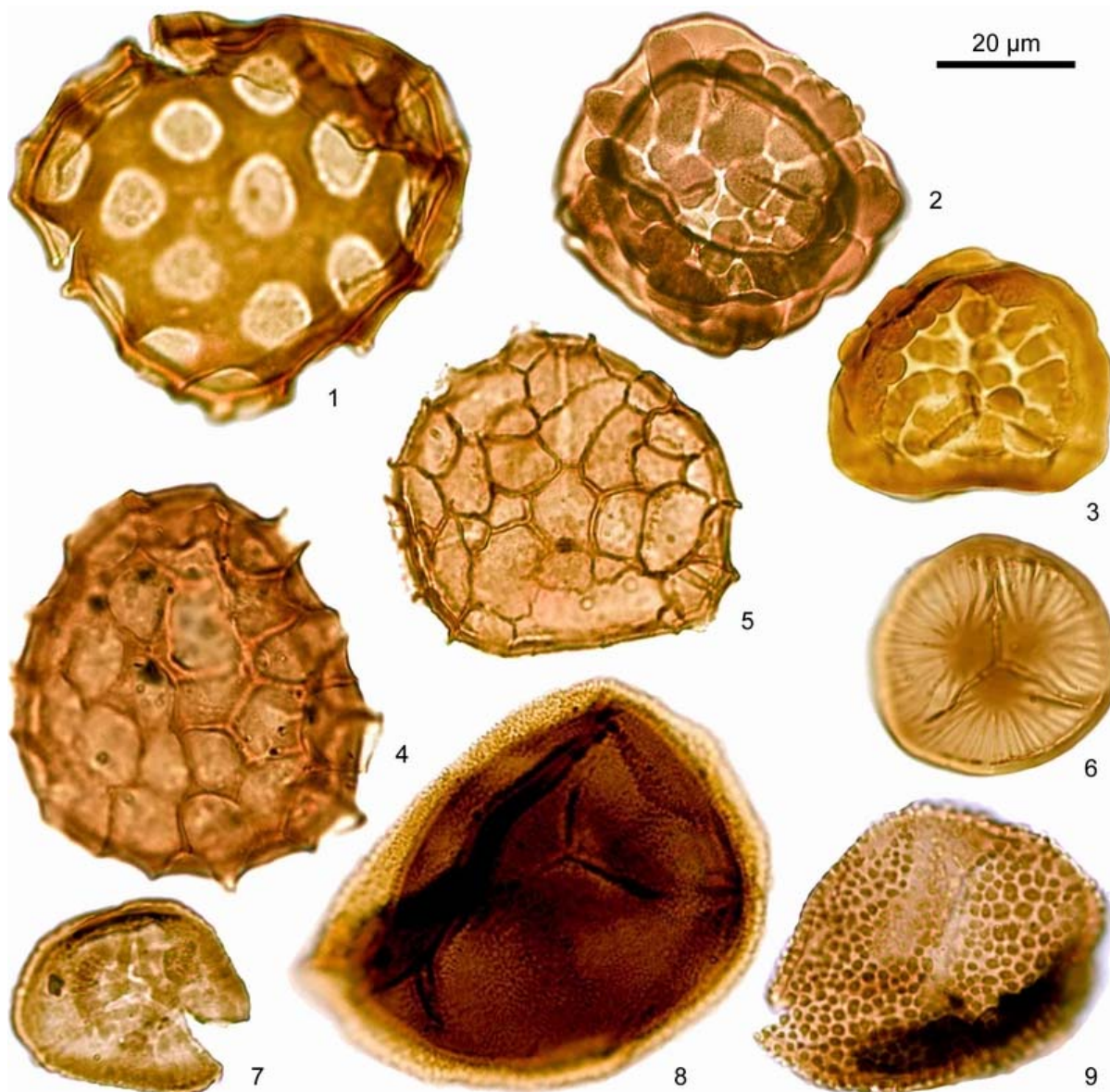


Figure 2: Microphotographs of characteristic miospores from the studied boreholes.

1. *Brochotriletes foveolatus*. Borehole BAQA-1, sample & slide 345.5', EFC H54/4.
2. *Clivosipora verrucata*. Borehole BAQA-1, sample & slide 395.2', F47/1.
3. *Clivosipora verrucata*. Borehole JNDL-4, sample & slide 87.2', F34/1.
4. *Dictyotriletes emsiensis*. Borehole BAQA-2, sample & slide 56.0', X46.
5. *Dictyotriletes subgranifer*. Borehole BAQA-1, sample & slide 366.9', O31.
6. *Emphanisporites schultzi*. Borehole BAQA-1, sample & slide 395.2', G50.
7. ? *Knoxisporites riondae*. Borehole BAQA-1, sample & slide 366.9', E27/4.
8. *Rhabdosporites minutus*. Borehole JNDL-3, sample & slide 368.8', H45/1.
9. *Verrucosiporites polygonalis*. Borehole BAQA-1, sample & slide 371.1', R25/4.

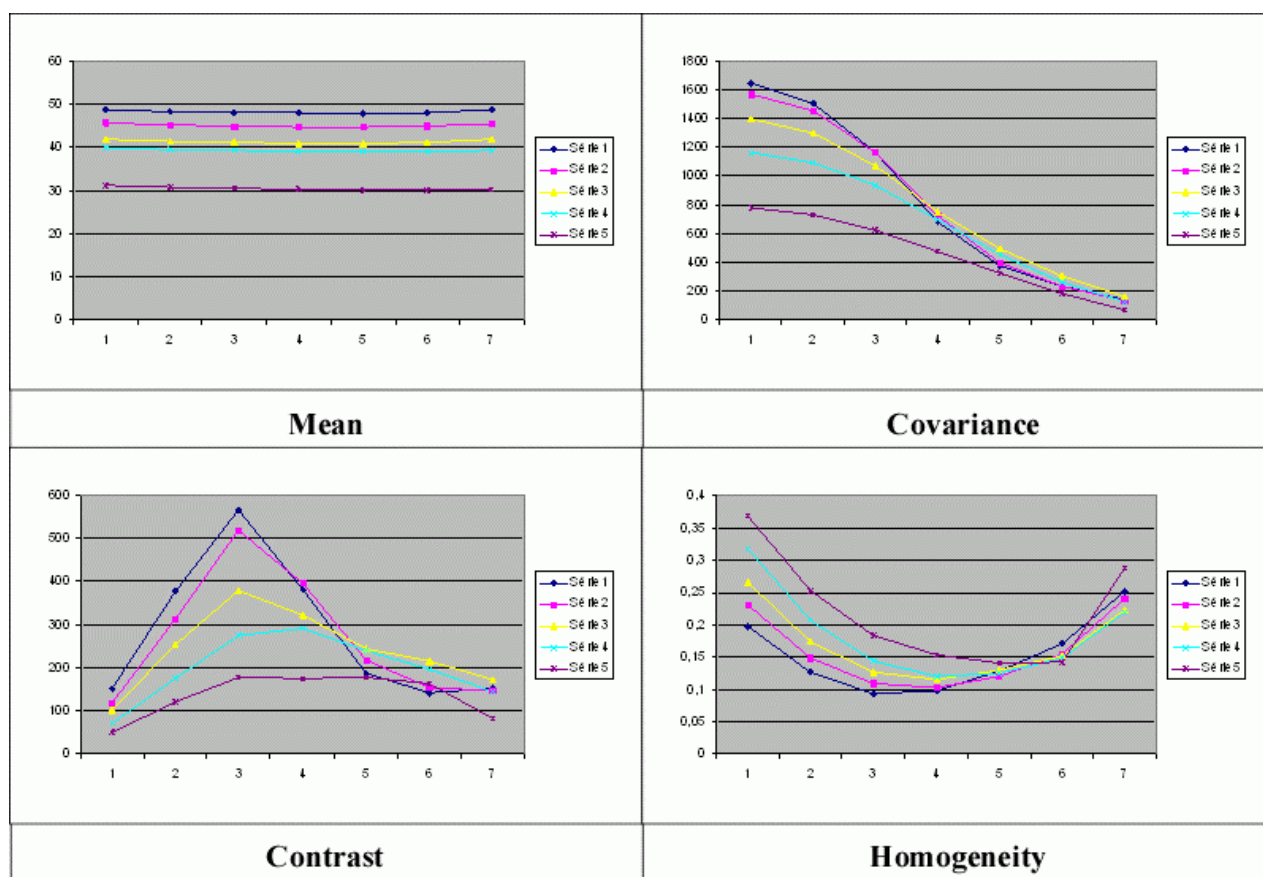


Figure 3: Mean, contrast, covariance and homogeneity for the 5 classes and for 7 decreasing scales.

Image analysis

The spore classification is based on the spatial distribution of the ornaments (from a simple pattern to a complex organization). Initially, we expect texture image analysis to provide a tool to gauge and thus to quantify morphological evolution and, eventually, to make automatic classification feasible. Both structural – extraction of texture elements and analysis of placement rules – and statistical methods – spatial statistics – are of interest in this case. Here we discuss only spatial statistics employing Sum and Difference Histograms.

Material

About 400 specimens of the miospore lineage defined here were found in more than 60 palynological slides but only half of the material was used for the image analysis. Excluded from the set are laterally compressed specimens and partial spore fragments.

A data set of region of interest (ROI) of 256x256 pixels was extracted from the original images in order to test the Spatial Vector Classification on SDH features (Fig. 3).

Sum and Difference Histograms (SDH)

Sum and Difference Histograms (SDH) are based on Co-occurrence Matrices (CoM) and are intended to make these matrices usable. Actually, CoM $P[d][z1][z2]$ are defined, for G grey level images, as the joint probability that a pair of point satisfying the dipole $d = (dx, dy)$ will have grey level values of $z1$ and $z2$. They display the image information as a co-occurrence of pixel pairs but although better-organized provide too much data. SDH (Unser, 1986) are often preferred to CoM and reduce the amount of data and computing time required. By writing $p(z, z')$ the probability of the pair of values z and z' at positions satisfying d , we define $p\Sigma(z + z')$ and $p\Delta(z - z')$ the probability of a sum and difference value for a given dipole.

HARALICK descriptors (HARALICK *et alii*, 1973) – mean, variance, contrast, energy, entropy and homogeneity – used to characterize a CoM have translations for SDH and results have shown that SDH are at least as efficient as co-occurrence matrices for classify textures.

$$\begin{aligned}
 \text{Mean} &= \sum_x \sum_{x'} z \cdot p(z, z') = \frac{1}{2} \sum_i i \cdot p_z(i) \\
 \text{Variance} &= \sum_x \sum_{x'} (z - \mu)^2 p(z, z') = \frac{1}{2} \left[\sum_i (i - 2\mu)^2 p_z(i) + \sum_j j^2 p_\Delta(j) \right] \\
 \text{Covariance} &= \sum_x \sum_{x'} (z - \mu)(z' - \mu) p(z, z') = \frac{1}{2} \left[\sum_i (i - 2\mu)^2 p_z(i) - \sum_j j^2 p_\Delta(j) \right] \\
 \text{Contrast} &= \sum_x \sum_{x'} (z - z')^2 p(z, z') = \sum_j j^2 \cdot p_\Delta(j) \\
 \text{Energy} &= \sum_x \sum_{x'} (p(z, z'))^2 \approx \sum_i p_z^2(i) + \sum_j p_\Delta^2(j) \\
 \text{Entropy} &= \sum_x \sum_{x'} -p(z, z') \log(p(z, z')) \approx -\sum_i p_z(i) \log(p_z(i)) - \sum_j p_\Delta(j) \log(p_\Delta(j)) \\
 \text{Homogeneity} &= \sum_x \sum_{x'} \frac{1}{1 + (z - z')^2} p(z, z') = \sum_j \frac{1}{1 + j^2} p_\Delta(j)
 \end{aligned}$$

This gives 7 values for the 65,535 pairs available in a 256x256 pixel image. Thereby, by keeping only one mean value by distance d and by selecting the distance as power of 2, this reduces data as 7 by 7 values when considering 7 scales ($d=1, 2, 4, 8, 16, 32, 64$) (Fig. 3).

Weighted Laplacian of Gaussian for ornament extraction

Classification based on Haralick features for the ROI data set gave results damaged by a large mean square error. Thereby, as only

ornaments are of interest in the classification, we first pre-processed the images in order to remove the background. Scale-Space theory (Lindeberg, 1994) advocated weighted Laplacian of Gaussian (wLoG) on increasing scales to capture blobs (the ornaments) independently of their sizes. An absolute value of the convolution of the ROIs by the wLoGs has been chosen to be independent of the fact that ornaments can appear black or white depending on the orientation of their relief (Fig. 4).

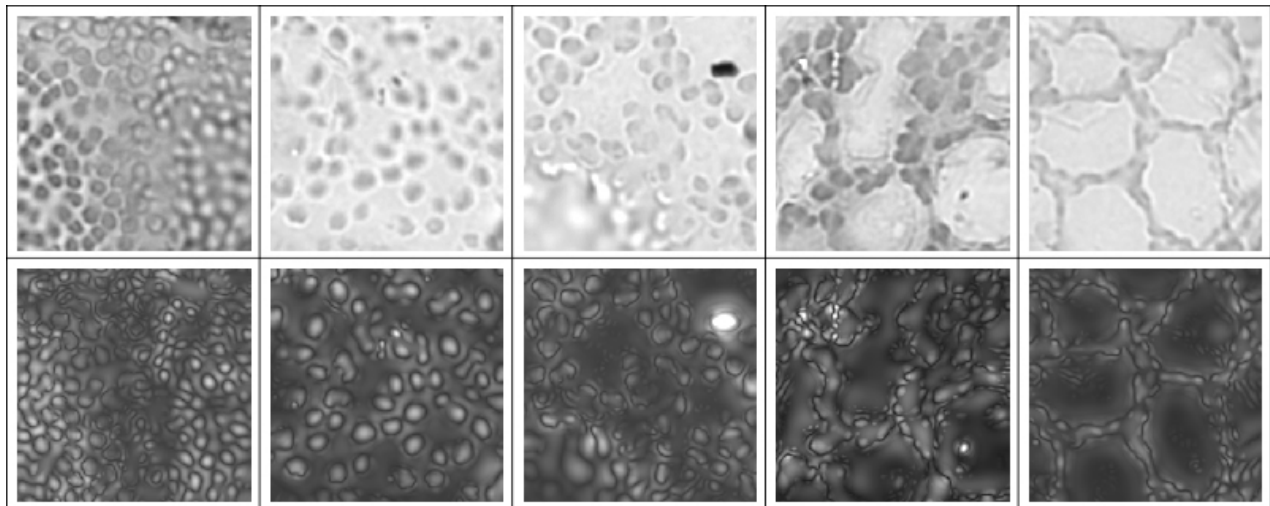


Figure 4: Regions of interest illustrating the 5 classes and their wLoG pre-processed version where background lighting is removed and ornaments highlighted.

Pattern classification

We used Support Vector Machine (CHANG & LIN, 2001; CORTES & VAPNIK, 1995; DUDA *et alii*, 2001; HASTIE *et alii*, 2001) to test a classification based on HARALICK descriptors on the original ROIs and on their wLoG version. A

classification task involves training and testing data. We used respectively 30, 30, 30, 40, 15 images for the training and 30, 20, 30, 30, 10 for the testing of the corresponding classes 1, 2, 3, 4, 5. Each instance in the training set contains one 'target value' – class label – and several 'attributes' – descriptors -. The goal of

SVM is to produce a model which predicts the target value of data instances in the testing set which are given only the attributes.

We selected the following attributes: the HARALICK descriptors for the scale 2 (7 attributes), the Contrast for the scale 3 and 4 and the Homogeneity for the scale 4 and 6 in order to exploit the differential scale dependency of these two last descriptors.

The model used is the nu Support Vector Classification with a linear kernel type.

The classification of the ROI gives an accuracy of 56% with a mean square error of 1.2 when the classification of the pre-processed ROI gives accuracy of 61% with a mean square error of 0.46 (Fig. 5).

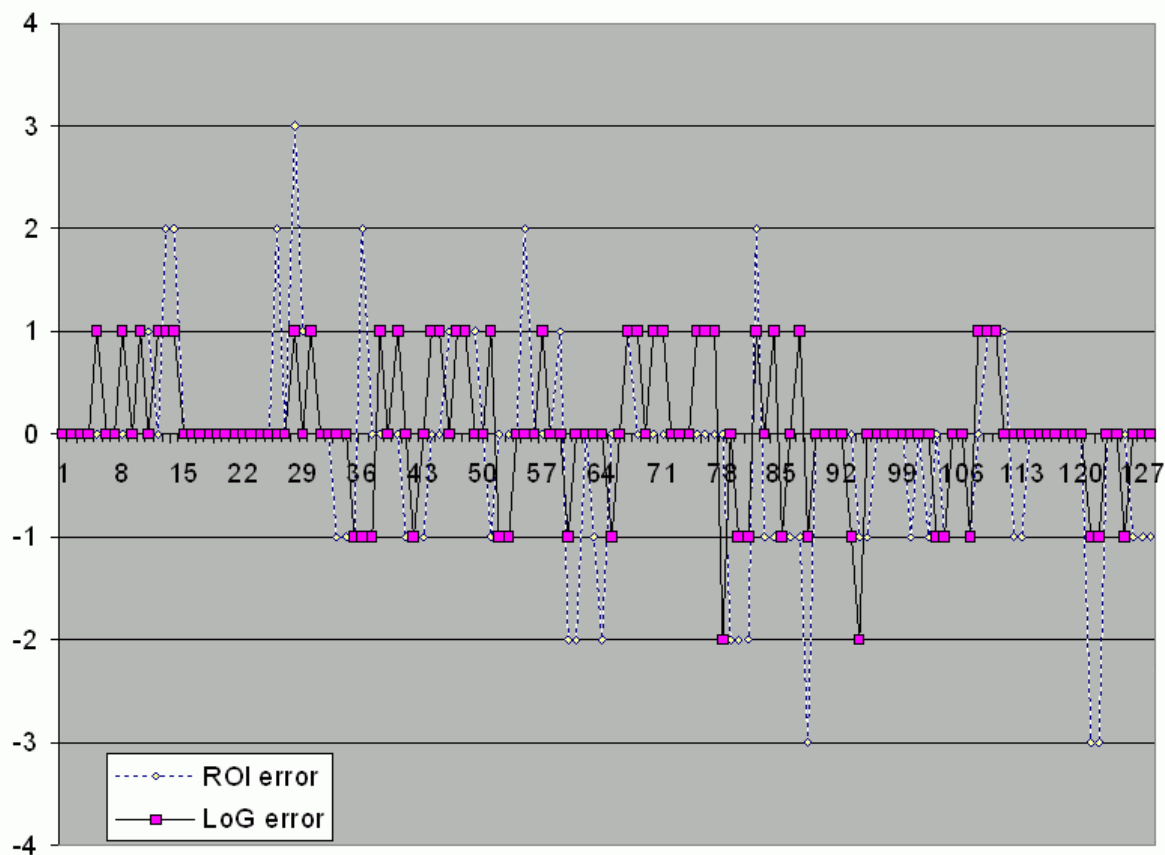


Figure 5: Classification error for the ROIs and the wLoG convolved ROIs.

Discussion

The accuracy of classification increases only from 56% to 61% when the ROI is pre-processed with the wLoG. But what is important is the reduction of the mean square error that balances the significance of the wrong classification. In fact, it drops so strongly that when a spore is misplaced it is reported only as being in a neighbouring class. In other words, if we accept the error of misplacing a spore in an adjacent class, the 'accuracy' of the classification increases from 85% to 98%.

In fact, the *a priori* human classification plays an important role as regards judgments of the accuracy and usefulness of computer-assisted classification. As classification by humans inevitably includes such types of error, we could adjudge this one test as equally subject to error but anticipate that a better-tuned training set would give better results.

Conclusion

Continuous morphological intergradations between two miospore taxa have been found in an Early Devonian miospore assemblage from Saudi Arabia. On the distal surface of these spores the ornaments and their organization show a gradual evolution in complexity between two end-members. All the intermediary forms co-occur in the assemblages. Thus there is a progressive evolution in the organization of the ornamentation ranging from the simplest spores to the most complex ones. This illustrates once again that miospore taxonomy is artificial because the two end-members of this lineage have been assigned to discrete genera.

Statistical texture image analysis provides a gauging tool to quantify morphological evolution and seems to allow assisted automatic classification. In this test case classification of spores by image analysis was

judged accurate in but 61% of the cases but rose to 98% if a misidentification to an adjacent class was accepted.

Acknowledgments

We wish to express our gratitude to management of Saudi Aramco for permission to publish this paper. We acknowledge M. GIRALDO-MEZZATESTA (Liège) for the preparation of palynological slides. Thanks are also expressed to Y. GUEDON (Montpellier, France) for the review of the paper. P. BREUER is supported by a F.R.I.A. grant.

Bibliographic references

- AL-HAJRI S.A., FILATOFF J., WENDER L.E. & NORTON A.K. (1999).- Stratigraphy and operational palynology of the Devonian System in Saudi Arabia.- *GeoArabia*, Bahrain, vol. 4, p. 53-68.
- AL-HAJRI S.A. & OWENS B. (eds.) (2000).- Stratigraphic palynology of the Paleozoic of Saudi Arabia.- *GeoArabia Special Publication*, vol. 1, Gulf PetroLink, Bahrain, 231 p.
- BOUCOT B., McCLURE H.A., ALVAREZ F., ROSS J.P., TAYLOR D.W., STRUVE W., SAVAGE N.N. & TURNER S. (1989).- New Devonian fossils from Saudi Arabia and their biogeographical affinities.- *Senckenbergiana Lethaea*, Frankfurt/Main, vol. 69, p. 535-597.
- BREUER P., AL-GHAZI A., AL-RUWAILI M., HIGGS K.T., STEEMANS P. & WELLMAN C.H. (in press).- Early to Middle Devonian miospores from northern Saudi Arabia.- *Revue de Micropaléontologie*, Paris.
- BREUER P., AL-GHAZI A., FILATOFF J., HIGGS K.T., STEEMANS P. & WELLMAN C.H. (2005a).- Stratigraphic palynology of Devonian boreholes from northern Saudi Arabia. In: STEEMANS P. & JAVAUX E. (eds.), Pre-Cambrian to Palaeozoic Palaeopalynology and Palaeobotany.- *Carnets de Géologie / Notebooks on Geology*, Brest, Memoir 2005/02, Abstract 01, p. 3-9.
- BREUER P., STRICANNE L. & STEEMANS P. (2005b).- Morphometric analysis of proposed evolutionary lineages of Early Devonian land plant spores.- *Geological Magazine*, Cambridge, vol. 142, p. 241-253.
- CHANG C.-C. & LIN C.-J. (2001).- LIBSVM: a library for support vector machines.- Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- CORTES C. & VAPNIK V. (1995).- Support-Vector Networks.- *Machine Learning*, Dordrecht, vol. 20, p. 273-297.
- DUDA R.O., HART P.E. & STORK D.G. (2001).- Pattern Classification. 2nd Ed. Wiley, New York.
- FOREY P.L., YOUNG V.T. & McCLURE H.A. (1992).- Lower Devonian fishes from Saudi Arabia.- *Bulletin of the British Museum of Natural History (Geology)*, London, vol. 48, p. 25-43.
- HARALICK R.M., SHANMUGAM K. & DINSTEN I. (1973).- Textural features for image classification.- *IEEE Transactions on Systems, Man and Cybernetics*, Los Alamitos, vol. SMC-3, N° 6, p. 610-621.
- HASTIE T., TIBSHIRANI R. & FRIEDMAN J. (2001).- The elements of statistical learning: Data mining, inference and prediction.- Springer, New York, 536 p.
- JÄGER H. (2004).- Facies dependence of spore assemblage and new data on sedimentary influence on spore taphonomy.- *Review of Palaeobotany and Palynology*, Amsterdam, vol. 130, p. 121-140.
- LINDBERG T. (1994).- Scale-Space Theory in Computer Vision.- Kluwer Academic Publishers, Dordrecht, Netherlands, 440 p.
- MARSHALL J.E.A. (1996).- *Rhabdosporites langii*, *Geminospora lemurata* and *Contagisporites optivus*: an origin for heterospory in the progymnosperms.- *Review of Palaeobotany and Palynology*, Amsterdam, vol. 93, p. 159-189.
- MAZIANE N., HIGGS K.T. & STREEL M. (2002).- Biometry and paleoenvironment of *Retispora lepidophyta* (KEDO) PLAYFORD 1976 and associated miospores in the latest Famennian nearshore marine facies, eastern Ardenne (Belgium).- *Review of Palaeobotany and Palynology*, Amsterdam, vol. 118, p. 211-226.
- STREEL, HIGGS K., LOBOZIAK S., RIEGEL W. & STEEMANS P. (1987).- Spore stratigraphy and correlation with faunas and floras in the type marine Devonian of the Ardenne-Rhenish regions.- *Review of Palaeobotany and Palynology*, Amsterdam, vol. 50, p. 211-229.
- UNSER M. (1986).- Sum and difference histograms for texture classification.- *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Los Alamitos, vol. PAMI-8, N° 1, pp. 118-125.
- VAN DER ZWAN C.J. (1979).- Aspects of Late Devonian and Early Carboniferous palynology of southern Ireland. I. The *Cyrtospora cristifer* Morphon.- *Review of Palaeobotany and Palynology*, Amsterdam, vol. 28, p. 1-20.