SAR PCA-based segmentation for hydraulic patterns identification

Henry J.-B.[1], Matgen P.[2], Tholey N.[1], Pfister L.[2], de Fraipont P.[1] and Hoffmann L.[2]

[1] Service Régional de Traitement d'Image et de Télédétection
Parc d'Innovation – Bld Sébastien Brant – BP 10413
F-67412 Illkirch Cedex

[2] Centre de Recherche Public-Gabriel Lippmann
Environment and Biotechnologies Research Unit (CREBS)
162a, avenue de la Faïencerie
L-1511 Luxembourg

Abstract — **SAR time series are used to identify hydraulic patterns within the alluvial plain of the Alzette River, Grand-Duchy of Luxembourg. Relationships between backscattering coefficient and hydro-meteorological variables are established under different land use conditions, at different scales. These relations with water table depths and rainfall are used to interpret the results of a PC analysis performed on 6 ERS images. The use of the k-means classification algorithm enables the identification of different hydraulic/hydrological behaviour patterns within the alluvial plain and the maximum flood extent.**

Index Terms—**Backscattering coefficient, rainfall, water table, SAR image time series, PCA, hydraulic/hydrological patterns classification.**

INTRODUCTION

The 3-day acquisition phase of ERS-1 has offered valuable synthetic aperture radar (SAR) time series well suited for hydrological and hydraulic studies. An important part of the Alzette river basin, located in the Grand Duchy of Luxembourg, was mapped during this period, from late 1993 to early 1994.

The main issue in time series analysis is to find a way to highlight significant changes. Principal component analysis (PCA) has often been used with optical multispectral data. This transformation leads to a multidimensional dataset, in which axis variables are uncorrelated: the higher the rank of the component, the lower the variance or the information content [1]. It was previously shown that stationary information is concentrated in the first components, using either optical [2] or radar data [3]. When PCA is applied for change detection purposes, the information on changes is rejected to the last components. Consequently, it might be difficult to make the difference between noise and change, especially with radar data [4]. It has been shown by ref. [4] that using change related images, such as ratio or difference, improves the quality of the detection. Moreover, it is shown that the first PC's are dominated by change information, when changing areas represent an important part of the source image [4].

Such recommendations are successfully applied by ref. [5] on a time series of ERS-1 intensity data. They illustrate the suitability of change related images for PC-driven change identification. Applied to bio-physical parameter retrieval, PC analysis is able to highlight different behaviours at the catchment scale and may help to identify source areas [6]. Ref. [6] shows the influence of topography on the first PC, whereas the second axis is more controlled by soil moisture patterns resulting from drainage patterns.

This paper presents an application of this catchment scale approach to smaller scales, i.e at the scale of the alluvial and flood plains. The test site, a part of the Alzette river basin, Grand Duchy of Luxembourg, was mapped to this end from late 1993 to early 1994 in the framework of the TecSpIn project. The TecSpIn project, realized in a close collaboration between SERTIT (France) and CREBS-CRPGL (Luxembourg), is supported by the French Space Agency (CNES) and the Luxembourg Ministry of Research and is centred on the use of space techniques for flood management. The general aim is to improve the predictability of two major issues in flood forecasting, i.e. the saturation degree of the basin (infiltration capacity) and the flood extent predictability. A better understanding of hydrological processes is targeted through an evaluation of the catchment's storage capacity. In a first step, the specific hydraulic patterns, drainage and storage capacities within the alluvial plain are identified in order to improve the knowledge of contributing areas during flood genesis. In a second phase, a complete space borne solution usable on similar catchments is evaluated, by establishing relationships between hydro-meteorological and remotely sensed variables.

Corresponding author : jb@sertit.u-strasbg.fr

TEST SITE AND DATABASE

In a first step, an important database is built up within the framework of the TecSpIn project over the Alzette river basin, where numerous studies have already been dealing with the interactions between physiographical and hydro-meteorological parameters [7], [8].

The Alzette river basin

The Alzette river originates in France, approximately 4 km south of the French-Luxembourg border. The river basin covers an area of 1172 km^2 and the valley accommodates almost 2/3 of the Grand Duchy of Luxembourg's population, as well as an important part of the industrial infrastructure. Since the occurrence of three major flooding events in 1993 and 1995, public authorities and the Public Research Centre - Gabriel Lippmann have built up a complementary stream- and raingauge network covering the Alzette basin. In the Alzette floodplain, the observation network has been completed with piezographs monitoring groundwater levels since 1986. The study area (fig. 1) is restricted to two flood prone sectors, upstream and downstream of Luxembourg City. The topography of the floodplain is characterised by a natural sandstone bottleneck, located near Luxembourg City. Upstream of the bottleneck, the valley is up to 2.5 km wide, while in the Luxembourg sandstone formation, the valley is only 75 meters wide. Downstream of the bottleneck the valley widens again. Alluvial deposit depths vary between 4 and 8 meters. Sand, gravel, as well as marls and clay alternate in the alluvial deposits.

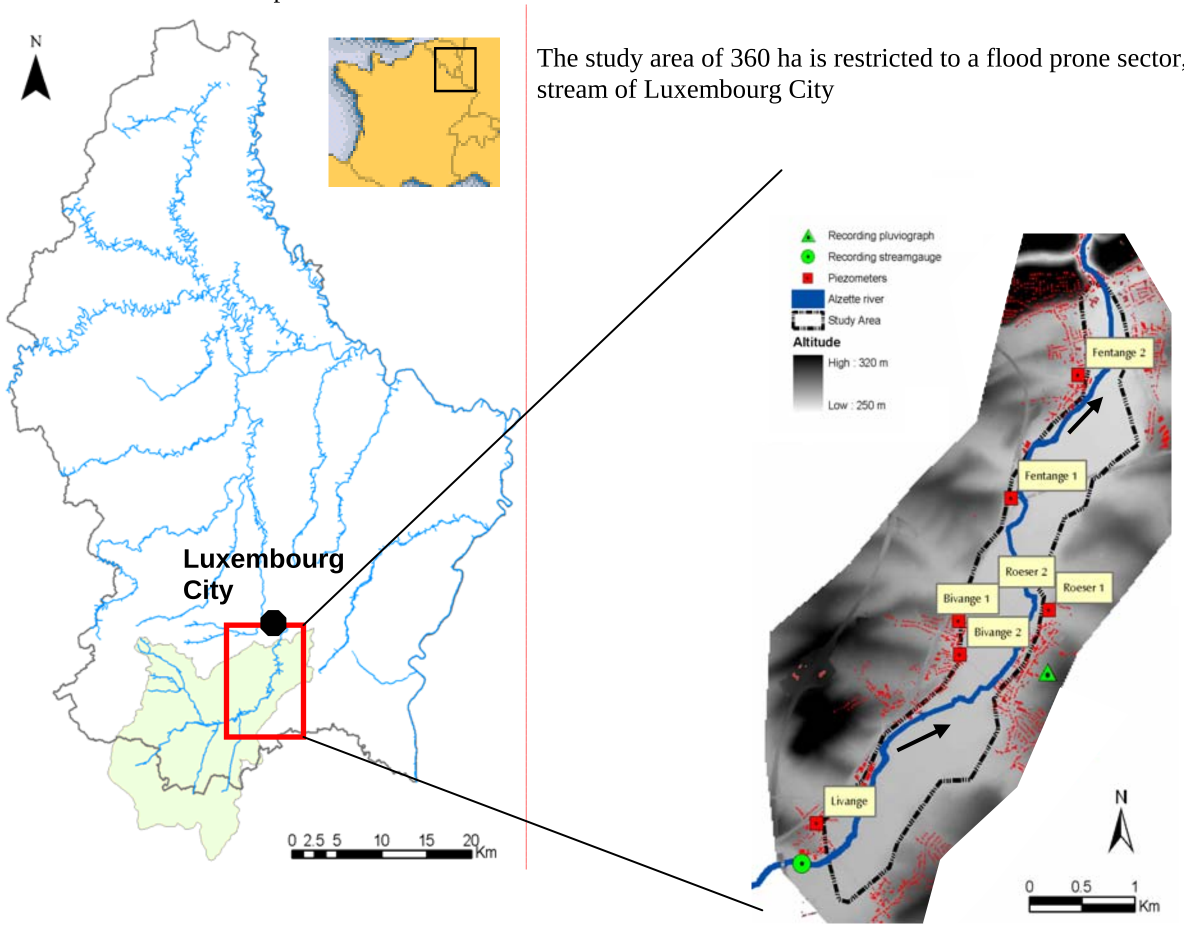


Fig. 1. DTM of the Alzette alluvial plain and water table measurement stations

Database constitution

The database is comprised of ERS SAR images on the one hand, and hydro-meteorological data (rainfall, runoff, piezometric levels, water budget) on the other hand. Radar data are first radiometrically calibrated [9], and then, orthorectified using a 75m resolution DTM. The EO database is composed of 13 ERS-1 and 2 images, acquired on descending pass, with 9 of them during the ERS-1 *Ice Phase* from 20/11/1993 to 23/02/1994, covering two important floods that occurred at the end of December 1993 and in early January 1994. The next 3 images were acquired before, during and after the flood of late December 1999.

Since hydro-meteorological variables are point measurements, a surfacing operation is performed on rainfall datasets to get a spatial evaluation of rain distribution, combined to a time integration over 2, 4, 7 and 10 days. The water table is also measured on several sites, but no probing results were obtained on surfacing. The final layer of the database is constituted of historical flood extents, obtained with radar image analysis, hydraulic modelling, aerial photography and ground survey.

RELATING BACKSCATTERING COEFFICIENT TO HYDRO-METEOROLOGICAL VARIABLES

The backscattering coefficient (σ_0) analysis over the 1993-94 series first shows the inadequacies of 3 images, because of bad weather conditions: negative temperatures, presence of snow. These conditions imply very low σ_0 values, mainly due to the penetration depths differences caused by freeze and snow water content. Therefore, these data are not considered in the quantitative study.

Different hydraulic situations are identified for each chosen acquisition, depending on flow measurements or simulations (tab. 1). The biggest flooded areas are observed during river overflows and during drainage in the following days. However, some areas known as resurgence prone areas, are systematically flooded during winter months, resulting in low backscattering coefficient (σ_0) values fairly independent of any spatio-temporal dynamic of groundwater elevation.

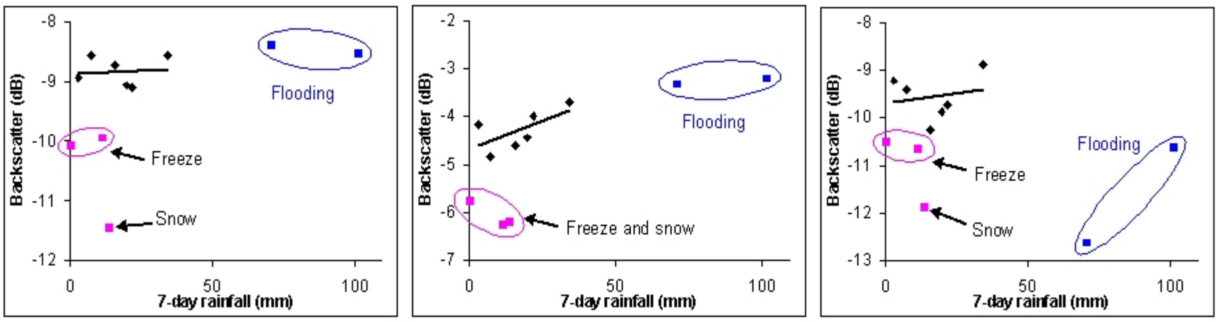
Tab. 1. 1993-94 radar database for quantitative analysis

Date	Flood stage
25/12/93	River overflow
06/01/94	River overflow
12/01/94	Generalised resurgence
02/02/94	Generalised resurgence
08/02/94	Local resurgence
23/02/94	Dry

Calibrated ERS SAR data enable the quantitative study of the spatio-temporal variations of σ_0, in parallel with rainfall, water table levels and other environmental parameters.

Relation between rainfall and σ_0

At a regional scale (upstream or downstream of Luxembourg City), different relationships are observed between mean7-day surface rainfall and mean σ_0 [10], on different land use types (fig. 2).



(a) Vegetated areas (b) Bare soils (c) Flood extent area

Fig. 2.Relationships between mean σ_0 and 7-day rainfall for different land use

On these different parcels, when flooding occurs, a saturation is observed on bare soils, whereas the global trend is not disturbed on vegetated areas. On the opposite, over flooded areas, the apparition of water on the soil surface makes the signal drastically drop, due to specular reflection.

Within the floodplain, the time series analysis of different radar images highlights different hydraulic patterns with respect to how the backscattering coefficient varies with the preceding day's rainfall amounts. Time-series analysis of signal return enables the location of areas with a notable drainage deficit. A more detailed insight proves that these areas lie in natural sinks with no direct superficial connection to the river network or in depressions with blocked flow-paths (*e.g.* due to road crossings), which implies a low drainage capacity. This phenomenon can be explained by the morphological structure of the upper Alzette valley. More generally, the morphological configuration of the Alzette plain makes this area behave as a bathtub (Fig. 1).

Relation between water table levels and σ_0

σ_0 measurements are also performed on homogeneously flooded parcels, affected by overflow or resurgence. On those flooded by resurgence, the link between temporal evolutions of water table gauges and σ_0 appears to be an interesting way for quantifying the radar power loss due to the specular reflection of water surface. The related increase in both σ_0 and water table elevation is clearly shown in fig.3. A drastic decrease in σ_0 is observed for the three parcels once water pounds on the surface. The regression relation is still under investigation, and will be extended to the whole area. Despite a lack of *in situ* soil moisture measurements, it is assumed that water table depths and the general saturation level are closely related, especially in the lowest parts of the alluvial plain.

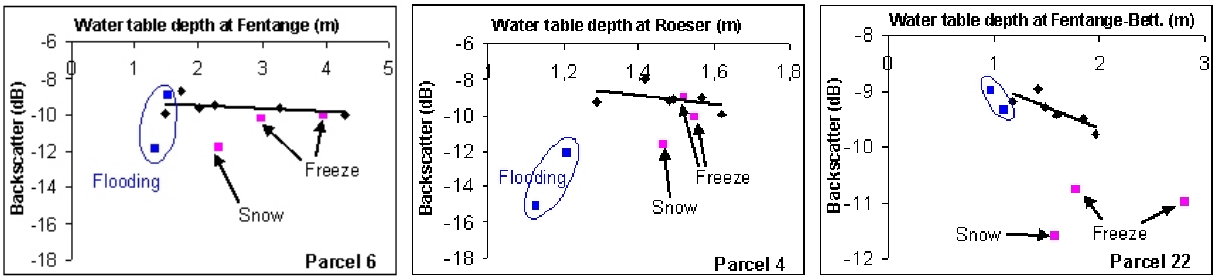


Fig. 3. Water table levels versus σ_0 variations

At an even smaller scale, it can be noticed that the brutal fall of σ_0 appears at different water table levels, again showing different hydraulic patterns over the floodplain. This suggests that it is possible to locate on a map the saturated cells for a given water table level recorded at the nearest piezometer. In order to automatically derive different behaviour classes, a Principal Component Analysis of a 6 image series is undertaken.

Relation between saturation index levels and σ_0

To avoid the problems related to the spatial variability of water table measurements, a synthetic Soil Saturation Index is proposed. It is funded on a Min-Max normalization of the variation range of water table depths over the October 1993 – April 1994 period. Thus, with *n* measurement stations of the water table depth *d*, the saturation index SSI for day *j* is obtained by:

$$SSI_j = \frac{1}{n}\sum_{i=1}^{n}\frac{d_{\max}^{i} - d_{j}^{i}}{d_{\max}^{i} - d_{\min}^{i}} \quad (1)$$

This transformation facilitates the comparison of the backscattering coefficient with a relevant indicator of the global saturation state of the alluvial plain. When the water table reaches its minimum depth of the year (January 1994 flood), the SSI will be 100 %, i.e. we assume that the soil is completely saturated. The SSI decreases linearly until the measured water table depth reaches its maximum value of the year (SSI = 0 %). Following this approach, each piezograph's range of values is individually taken into account and the sample of groundwater measurements becomes more homogeneous (fig. 4). Thus, on each day of satellite overpass, the SSI value computed at each site is more or less the same and the average SSI value can be considered as a valuable indicator of the overall soil saturation inside the study area.

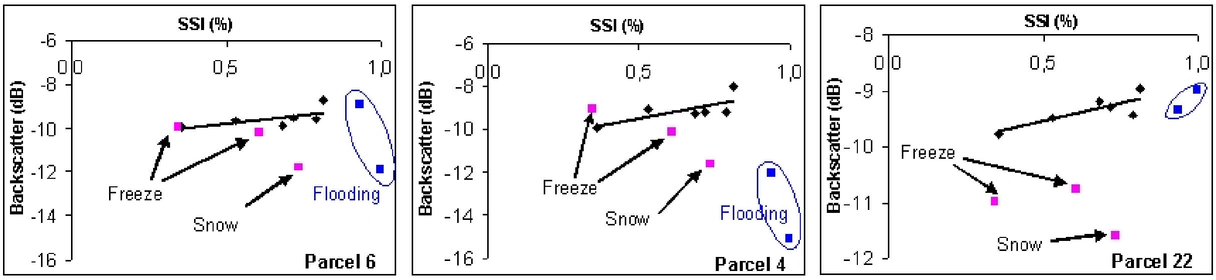


Fig. 4. Soil saturation index versus σ_0 variations

FLOOD PLAIN SEGMENTATION

To derive hydrological behaviour classes and outline the changes in backscattering due to soil moisture variations, it is important to process a sequence of images with homogeneous topography and stable land use, such as in the two study areas. In a PCA performed on an image series covering an entire catchment, the first and third principal components tend to be mainly influenced by spatial variations in topography and land use [6]. This is the reason why the floodplain is split into areas where homogeneous topography and land use allow investigating the changes in soil moisture in a sequence of radar images.

The PCA outlines very interesting spatially variable hydraulic patterns. As pointed out in the preceding sections, these variances are mainly due to different patterns with respect to changing water table levels and rainfall amounts in the preceding days. Water stored on the surface has a characteristic low backscattering coefficient. Thus, it can be observed that:

- The first PC is mainly influenced by the presence of water at the soil's surface, permitting an accurate mapping of the groundwater resurgence that remains visible during the entire investigation period. This characteristic of the first PC is validated with a ground survey of groundwater resurgence areas. To a lesser extent the first PC also allows mapping the flooded areas over the whole period, either by overflow or resurgence.
- The second PC is related to the temporal variability of flooded areas, discriminating resurgence from overflow. It pinpoints those areas of the floodplain where water appears at the soil surface only in very humid conditions. These areas help to assess antecedent soil moisture conditions.
- Finally, the third PC appears to be driven by soil drainage capacity, which is related both to presence of water and the nature of flooding. This interpretation is more ambiguous and still needs to be confirmed.

The combined use of the first and second PC enables a classification of the floodplain into behaviour pattern classes. These classes can provide a rapid assessment of the saturation degree by using updated radar data, and thus constituting a useful tool to flood forecasting applications. The relationship between SSI and different hydraulic/hydrological parameters (e.g. surface runoff coefficient, or the amount of rainfall necessary to reach a complete saturation of the basin) is currently under investigation. The first results illustrate the usefulness of radar data in flood forecasting [12].

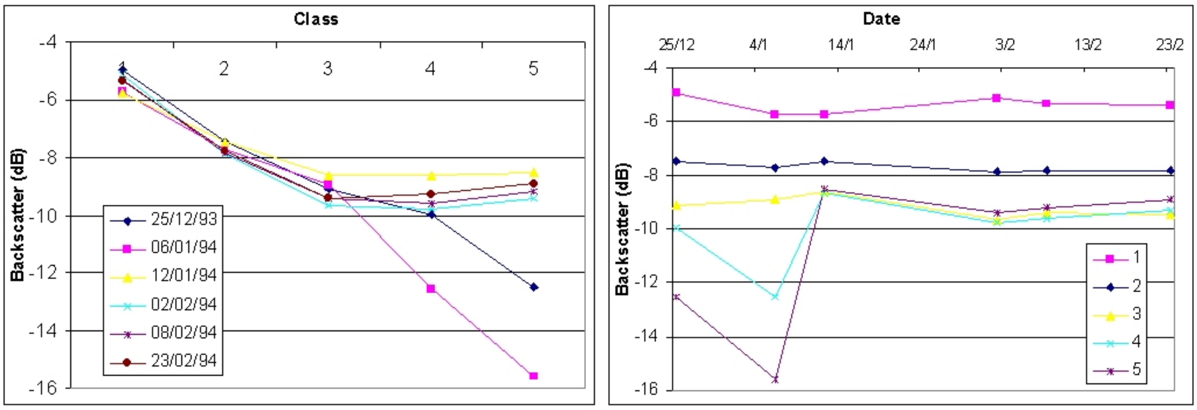
The correlation between groundwater depths measured at the nearest piezometric station and the backscattered signal varies depending on the different classes obtained from first and second PC combination. These classes can be used as interpretation keys of new images, to assess the saturation level.

FLOOD PLAIN SEGMENTATION

The k-means algorithm is used to classify the first PCs over two distinct scales: the alluvial plain and the maximum flood extent. In the first case, five classes are derived whereas only three appear to be relevant in the second case.

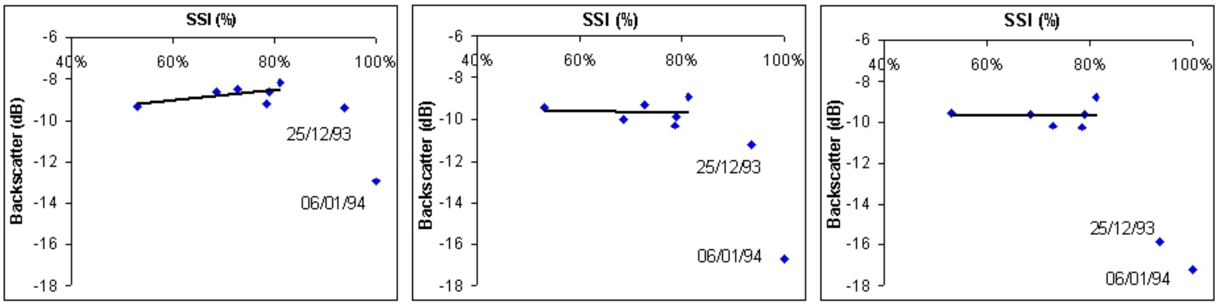
In the first case (segmentation of the alluvial plain), classes 4 and 5 produce a geographical coverage similar to the maximum flood extent. This observation is confirmed by the analysis of the radar signatures of each class and their evolution over time (fig. 5). The three other classes are less concerned by the flooding: classes 1 and 2 are entirely located out of the maximum extent, and class 3 comprises of both flooded and non-flooded areas.

When confronted with SSI evaluations, each radar signature highlights the dynamics of flooding in each class (fig. 6). For example, class 5 is flooded twice during the 1993-94 period. Thus, two distinct data points are characterised by very low signal return (low σ_0) and very high saturation indices (up to 100% on 6th of January).



(a) Classes signatures for each acquisition date (b) Evolution along time of radar signatures

Fig. 5. Behaviour classes identified at the alluvial plain scale



(a) Class 3 (b) Class 4 (c) Class 5

Fig. 6. Radar signatures of behaviour classes vs. soil saturation index

In the second case (segmentation of the maximum flood extent), each class seems to be characterized by the frequency of submersion during the study period. In this next step, the first results obtained at the alluvial plain scale, will be refined. Current analyses are focussed on statistical evaluation and validation of each linear regression, in the wetting up phase.

CONCLUSION

The TecSpIn bilateral research project has chosen the Alzette River, Luxembourg, because of the high vulnerability of its urbanized floodplain but also because of the advanced knowledge of its regime. The large database is composed of EO SAR data and many observed and simulated hydro-meteorological data.

Based on the influence of environmental conditions on radar signal, coupling strategies are studied between σ_0 measurements and other variables (*e.g.* piezometric levels or rainfalls). Good relationships are observed under different land use conditions, at different scales, even if they still need to be improved. The development of a synthetic Soil Saturation Index allows a better use of water table depths information all over the alluvial plain.

With these strong observations on the relationships between radar signal and hydro-meteorological variables, a method based on Principal Components Analysis is proposed to take advantage of the information content of the ERS-1 time series. The first three components present satisfactory hydraulic interpretations. A classification algorithm helps to identify and characterise different areas, which can be considered as representative of the way the alluvial plain is globally functioning. Hydro-meteorological variables are then re-used to provide interpretation keys and ensure their reliability. First results are encouraging, but many developments are still necessary, especially in validating the regressions built for the description of the wetting up phase. They are currently investigated and many improvements are expected until the end of the project.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Robert Kipgen at the 'Services de la Gestion de l'Eau', Mr. Gilbert Schleich at the 'Service National de la Protection Civile' of the Grand-Duchy of Luxembourg for providing some of the data used in this study and Mr. Jean-Paul Abadie at the French Space Agency (CNES) for supporting the project.

REFERENCES

[1] R.C. Gonzalez and P. Wintz, "*Digital image processing*", Addison-Wesley, Reading, MA, 1977.

[2] J.A. Richards, "Thematic mapping from multitemporal image data using the Principal Components transformation", *Remote Sensing Environ.*, vol. 16, no. 1, pp. 35-46, 1984.

[3] G.M. Henebry, "Advantages of principal components analysis for land cover segmentation from SAR image series", *Proc. of the third ERS symposium*, Florence, 21-26 March, ESA SP-414, pp. 175-178, 1997.

[4] P. Gong, "Change detection using principal component analysis and fuzzy set theory", *Can. J. Remote Sensing*, vol. 19, pp. 22-29, 1993.

[5] Y. Moisan, M. Bernier and J.M. Dubois, "Détection des changments dans une série d'images ERS-1 multidates à l'aide de l'analyse en composantes principales", *Int. J. Remote Sensing*, vol. 20, no. 6, pp. 1149-1167, 1999.

[6] N. Verhoest, P. Troch, C. Paniconi and F. de Troch, "On the retrieval of bio-physical parameters from a multi-temporal series of ERS-SAR PRI images", *Water Resourc. Res.*, vol. 34, pp. 3235-3244, 1998.

[7] L. Pfister, J. Humbert, J. Iffly and L. Hoffmann, "Use of regionalized stormflow coefficients in view of hydro-climatological hazard mapping," *Hydrol. Sci. J.*, vol. 47, no. 3, pp. 479-491, 2002.

[8] L. Hoffmann, A. E. Idrissi, L. Pfister, B. Hingray, F. Guex, A. Musy, J. Humbert, G. Drogue and T. Leviandier, "Development of regionalized hydrological models in an area with short hydrological observation series," *River Research and Applications*, vol. 20, pp. 243-254, 2004.

[9] H. Laur, P. Bally, P. Meadows, J. Sanchez, B. Schaettler, E. Lopinto and D. Esteban, "ERS SAR calibration, derivation of the backscattering coefficient $\sigma 0$ in ESA ERS SAR PRI products," ESA/ESRIN, Issue 2, Rev. 5b, ES-TN-RS-PM-HL09, 1998.

[10] K. Fellah, Y. Besnus, S. Clandillon, A. Ettajani, C. Meyer and P. de Fraipont, "Multi-temporal SAR data in environmental studies: researching a quantitative approach for soil moisture retrieval," in *Proceedings of the First ERS-1 Pilot Project Workshop*. Toledo: ESA SP-365, June 1994, pp. 85-92, 1994.

[11] R. Romanowicz and K. Beven, "Dynamic real-time prediction of flood inundation probabilities," *Hydrol. Sci. J.*, vol. 43, no. 2, pp. 181-196, 1998.

[12] S. Franks, P. Gineste, K. Beven and P. Merot, "On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process," *Water Resourc. Res.*, vol. 34, pp. 787-797, 1998.