ELSEVIER

Journal of Hydrology 247 (2001) 116–132

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Slope scale variation of flow patterns in soil profiles

E. Zehe[a,*], H. Flühler[b]

[a]*Institute of Water Resources Planning, Hydraulic & Rural Engineering, University of Karlsruhe, 76128 Karlsruhe, Germany*
[b]*Institute of Terrestrial Ecology, Soil Physics, Grabenstrasse 3, CH-8952, Schlieren, Switzerland*

Received 10 May 2000; revised 20 December 2000; accepted 5 March 2001

Abstract

Tracer patterns in soil profiles observed in ten transport experiments using a dye and Bromide are analysed to investigate the spatial distribution of the susceptibility for preferential flow on the slope scale. Flow patterns are characterised by interval scaled parameters that describe the plot scale variation of the vertical transport. This measure is used for a cluster analysis to discriminate similar flow patterns. A discriminant analysis of the obtained clusters shows whether the membership of a flow pattern to a cluster may be explained by independent quantities that characterise the conditions of the corresponding transport experiment and relative position of the field site on the slope. The results suggest that sites located at the bottom of the slopes, especially sites that are located near the brook, have a higher susceptibility for preferential flow than those in the upper parts of a hill. This is a characteristics of soil formation of hill slope soil catenas, which usually leads to deeper, biologically more active, and often, finer textured soils in the hill foot sector. This in turn strongly affects the transport regime of these soils.

Keywords: Preferential flow; Scaling; Pattern; Statistical analysis; Experiments

1. Introduction

The question whether an infiltration event produces preferential or matrix flow is of major importance for the fate of agrochemicals especially pesticides in natural soils (Rao et al., 1974; Jury et al., 1986; Flury et al., 1994, 1995; Flury, 1996; Stamm et al., 1998; Mohanty et al., 1998; Zehe and Flühler, 2001). To answer this question one has to clarify which properties determine the susceptibility of a soil for preferential flow, which is the primary manifestation of the plot scale variability of infiltration. Blöschl (1996) and Dooge (1986) pointed out that different types of variability—randomness or order—may dominate the spatial variability of hydrological processes. Seyfried and Wilcox (1996) reported that on different spatial scales the variability of a process is determined by the spatial distribution of different key parameters. These key parameters strongly reflect ecosystem characteristics and the scale of interest (Klemes, 1983). The aim of the present analysis is to identify key parameters that dominate the spatial variability of infiltration, i.e. parameters that favour or suppress preferential flow on the plot and the slope scale. We propose a method to quantify the variability of infiltration and tracer transport on the slope scale based on a cluster- and a discriminant analysis of a spatially distributed sample of plot scale flow patterns. These flow patterns result from transport experiments using the fairly mobile dye Brilliant Blue (BB) and Bromide (Br), that were carried out at ten different field sites in the Weiherbach valley (Germany). The Weiherbach

* Corresponding author.
E-mail address: erwin.zehe@bau.verm.uni-karlsruhe.de (E. Zehe).

0022-1694/01/$ - see front matter
PII: S0022-1694(01)00371-7

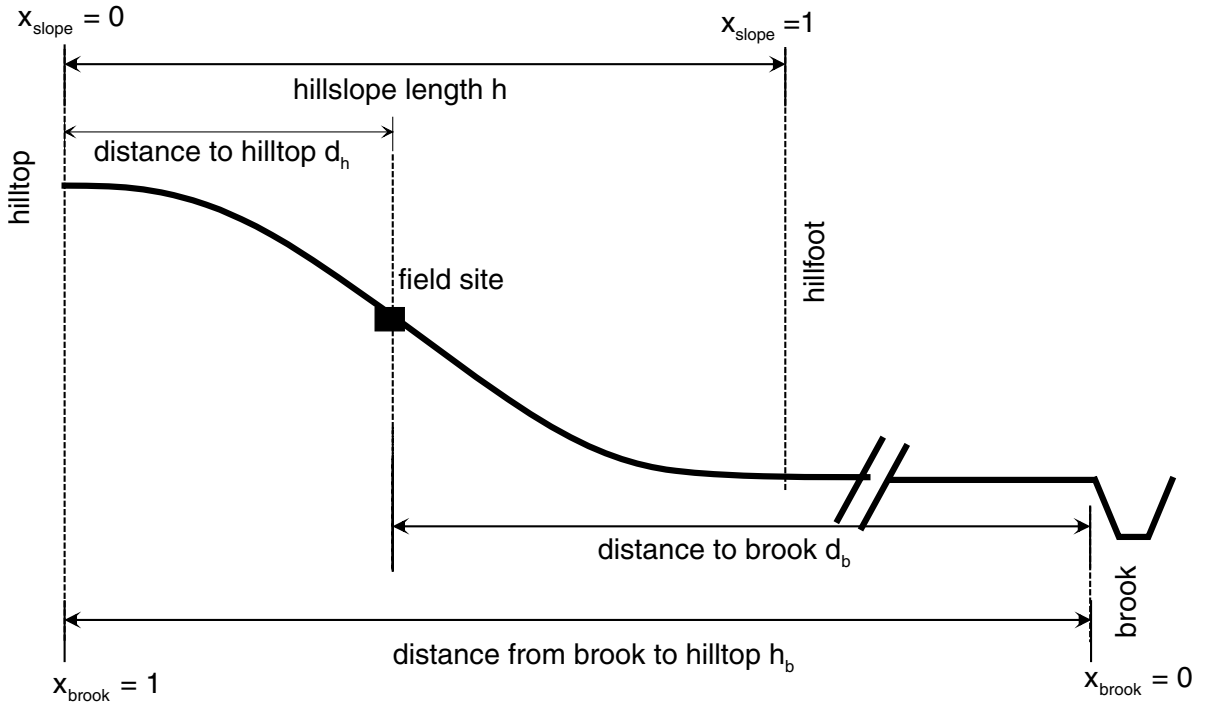

Fig. 1. Scheme how the parameters x_{slope} and x_{brook} were computed: $x_{slope} = d_h/h$, $x_{brook} = db/hb$. Thus, the foot of a hillslope in the meadows of the brook $x_{slope} = 1 - x_{brook}$.

catchment, setup of the plot scale irrigation experiments as well as the data collection is described in Zehe and Flühler (2001).

1.1. Infiltration-patterns on different scales in the Weiherbach valley

Of course, the spatial distribution of key parameters may change with time making them conditional to a particular time scale of interest. In this study we focus on an event time-scale with a characteristic time of 1 h–1 d (Blöschl, 1996). In this time frame, morphological parameters may be regarded as time invariant. On the plot scale with a characteristic length of 1–10 m the spatial variability of precipitation may be neglected as well. Hence, the spatial infiltration pattern is determined by the spatial variation of factors that favour or suppress preferential flow in the upper soil horizon. As these distributions may experimentally only be quantified in a statistical sense, the plot scale variability of infiltration appears to be random with the degree of preferential flow being its prime manifestation. Any quantitative measure that captures this degree of preferential flow may, therefore, be used as a key parameter in the sense explained above.

The prime factor causing preferential pathways in the soils of the Weiherbach valley are earthworms such as *Lumbricus terrestris*. This species prefers moist, but drained soils, with relatively moderate moisture and temperature fluctuations (Ehrmann, 1996). The typical hill slope catena in the Weiherbach valley are Calcaric Regosols (FAO/UNESCO, 1988; Pararendzina) or Luvisols (FAO/UNESCO, 1988; Parabraunerde) in the top or the mid-slope sectors and Colluvisols at the hill foot (FAO/UNESCO, 1988; Kolluvium).

As reported by Schmaland (1996), this distribution of soil types is closely linked to the spatial distribution of macroporousity. He measured the macropore volumes v_{mak} at 12 field sites in the Weiherbach valley, on plots of 0.5-m^2 size, averaged over a depth of down to 2 m. Schmaland (1996) observed systematically higher values of v_{mac} in Colluvisols

Table 1
Site, date of irrigation, soil type, landuse, normalised position at the slope x_{slope} (low values of $x_{slope} \in [0,1]$ indicate that the plot is located at the hill top), the normalised distance to the Weiherbach brook x_{brook} (low values of $x_{brook} \in [0,1]$ indicate that the plot is located close to the brook), cumulated irrigation I_c, average irrigation rate I, initial water content θ and relative saturation S_{rel} of the upper soil horizon

Site	Date	Soil type	Land use	I_c (mm)	I (mm/h)	θ (m^3/m^3)	S_{rel} (−)[a]	x_{brook} (−)	x_{slope} (−)
1	7/96	Luvisol	Wheat	23.8 ± 2.4	10.8 ± 1.1	0.179 ± 0.01	0.22	0.85	0.95
2	6/96	Colluvisol	Corn	21.9 ± 1.8	9.1 ± 0.8	0.171 ± 0.01	0.19	0.99	0.94
3	6/96	Calc. Regosol	Grassland	23.4 ± 2.3	10.7 ± 1.1	0.180 ± 0.01	0.32	0.99	0.79
4	6/96	Luvisol	Wheat	22.2 ± 2.4	10.2 ± 1.1	0.168 ± 0.01	0.18	0.88	0.66
5	6/96	Calc. Regosol	Corn	22.8 ± 2.5	10.7 ± 1.2	0.237 ± 0.01	0.47	0.63	0.13
6	6/96	Colluvisol	Wheat	22.8 ± 2.5	10.9 ± 1.2	0.253 ± 0.01	0.45	0.82	0.93
7	7/96	Calc. Regosol	Corn	22.9 ± 2.5	10.2 ± 1.1	0.159 ± 0.01	0.26	0.21	0.04
8	6/96	Luvisol	Corn	21.8 ± 1.5	10.9 ± 0.8	0.205 ± 0.01	0.29	0.10	0.34
9	7/96	Colluvisol	Wheat	22.3 ± 1.8	9.7 ± 0.8	0.232 ± 0.01	0.38	0.75	0.90
10	11/97	Colluvisol	Mustard	25.3 ± 2.3	11.0 ± 1.0	0.278 ± 0.01	0.53	0.01	0.99

[a] The following residual water and saturated water contents were uses: Colluvisol and Luvisol $\theta_s = 0.43$, $\theta_r = 0.11$, Calcaric Regosol $\theta_s = 0.46$, $\theta_r = 0.06$ (from Delbrück, 1997)

than in Calcaric Regosols or Luvisols. This trend may be quantified by the relative position of the site on the slope $x_{slope} \in [0,1]$, indicating whether a field site is located at the hill top ($x_{slope} = 0$) or at the hill foot ($x_{slope} = 1$), and its relative distance to the brook x_{brook} in the Weiherbach valley, measuring whether the plot is located it at maximum ($x_{brook} = 1$) or at minimum distance from the brook ($x_{brook} = 0$, see scheme in Fig. 1). As Colluvisols are mostly located at hill foots, they may be found at points with large values of $x_{slope.}$ At locations with low x_{brook} values one may find Colluvisol with the most even moisture and temperature regime, due to the small distance from the brook. The observed macropore volumes v_{mak} at the 12 field sites show indeed a positive correlation of $r = 0.79$ with x_{slope}, indicating higher volumes at the hill foot, and a negative correlation of $r = -0.84$ with x_{brook}, indicating highest macropore volumes at sites close to the brook. As the macroporousity determines to some degree the susceptibility of a soil for preferential flow, this order in its spatial distribution may cause order in the variability of infiltration on the slope scale.

1.2. Field experiments

In spring 1996 and fall 1997, ten tracer experiments were carried out on plots of $1.4 \times 1.4\ m^2$ size at ten different locations in the Weiherbach-valley. Details concerning the preparation and irrigation of the plots as well as the measurement of experimental conditions are given in Zehe and Flühler (2001). Table 1 lists the date of the experiments, soil type and land use at the field sites as well as data that are necessary for the intended discriminant analysis: the normalised position of a plot at the hill slope x_{slope}, its normalised distance from the brook x_{brook}, the cumulated irrigation I_C, the average irrigation rate I and the initial water content θ in the upper 15 cm of the soil. As we intended to study the influence of different soil types, their macroporousity and the initial water content on infiltration, we tried to achieve, that the ten plots were irrigated at the same rate and with the same amount of tracer solution. The coefficients of variation of I_c and I in between the irrigation experiments are 0.06 and 0.05, respectively, which is of similar magnitude as the variation of I_c and I on the individual plots, given in Table 1. Additionally the relative saturation S_{rel} of the upper soil layer was computed and listed in Table 1 using:

$$S_{rel} = \frac{\theta - \theta r}{\theta_s - \theta_r}$$

The used values of the residual and the saturated water content θ_r and θ_s are given below Table 1.

One day after irrigation, two vertical soil profiles were prepared at each site. The dye pattern at the surface of each profile was photographed. Then soil samples were extracted from each visible stained grid cell of a regular grid of $1 \times 1\ m^2$ size, subdivided with

rubber strings into 0.1×0.1 m^2 cells, that was placed onto the vertical profile, as well as 0.1 m below the leading edge of the dye pattern. These samples were analysed for their Br-content in the laboratory using HPLC. Further details characterising the data collection as well as the determination of Br content of the soil samples are given in Zehe and Flühler (2001).

Furthermore the number width and depth of macropores were determined on separate horizontal profiles at sites 1, 5 and 10. In accordance with the study of Schmaland (1996) we observed the lowest macroporousity at site 5, which is located in a Calcaric Regosol at a hill top. A clearly higher macroporousity was found at site 10, in a Colluvisol close to the brook, the highest value was found at site 1, which is also located in a Colluvisol. Further details concerning the determination of the macroporousity at these sites are given in Zehe and Flühler (2001).

2. Theory

We intend to test whether the above mentioned slope scale variance structure of macroporousity had a distinct influence on the results of the plot scale transport experiments. A suitable way to quantify to which extent the occurrence of preferential flow patterns is determined by the interplay of morphological and event driven factors, is

- to quantify the similarity of observed flow patterns by means of cluster analysis; and
- to explain these similarities by the experimental conditions of the tracer experiments such as initial moisture content or macroporousity of the soil by means of a discriminant analysis.

2.1. Cluster analysis

Cluster analysis is a method to subdivide a database into groups of similar objects (Kaufman and Rousseeuw, 1992; Anderberg, 1973). As similarity of objects is defined according to their distances in a parameter space, they must be represented by interval scaled parameters or properties. In the present context the objects are the flow patterns resulting from the above described tracer experiments. In general clustering methods are divided into hierarchical and partitioning techniques.

2.1.1. Hierarchical methods

An agglomerative hierarchical method starts with the finest partition where each object forms a cluster on its own, computes the initial distance matrix, combines the two 'closest' clusters forming a new one, computes the new distance matrix, joins the closest clusters and so on until all objects are joined into a single cluster. The main difference between individual hierarchical algorithms is the way to compute distances between clusters (Fahrmeir, 1984). In the present study the Wards method (Kaufman and Rousseeuw, 1992) was chosen, because the distance between clusters is defined in such a way that the combination of the two closest clusters leads to a minimum increase of a the within-group-ssp-matrix (**s**um of **s**quares and **p**roducts) **W**, and therefore leads to very homogenous groups [Eq. (1)].

The two most efficient criteria to determine the number of groups or clusters in a data set (Milligan and Cooper, 1985) are the pseudo-F, which is proportional to ratio of the traces of the between-groups-ssp-matrix **B** [Eq. (1)] and the within-groups-ssp-matrix **W**, and the R^2, the part of the total variance which is explained by the subdivision. Both parameters, the pseudo-F and the R^2, should be large for a optimal number of groups, which means highest similarity within and highest dissimilarity between the groups.

$$\mathbf{W} = \sum_{k=1}^{g} \sum_{n=1}^{N_k} (\mathbf{x}_{kn} - \overline{\mathbf{x}_k})(\mathbf{x}_{kn} - \overline{\mathbf{x}_k})' \qquad \mathbf{B} = \sum_{k=1}^{g} (\overline{\mathbf{x}_k} - \overline{\mathbf{x}})(\overline{\mathbf{x}_k} - \overline{\mathbf{x}})' \tag{1}$$

The data set consists of $N = \Sigma_k N_k$ objects e.g. flow patterns. Each object is characterised by a m-dimensional property vector, $\mathbf{x}_{kn}$ is the m-dimensional property vector of the n-th object in group k, where $\overline{\mathbf{x}}$ is the overall mean of the property vectors and $\overline{\mathbf{x}}_K$ is the average property vector in group k, also called group or cluster centroid. The group consists of N_k objects and N_g is the number of groups.

If the objects are represented by vectors consisting of parameters of different physical dimensions such as temperature and length, some authors favour the

standardisation of the data to assign equal weights to all these parameters (Steinhausen and Langer, 1977; Späth, 1975). On the other hand, different parameter ranges may reflect natural weights and a standardisation may destroy this natural structure in a sample (Kaufmann and Rousseeuw, 1992).

2.1.2. Partitioning methods

The goal of a partitioning clustering method is to improve a subdivision of a sample into a fixed number of groups with respect to a quality criterion, such as Wilks' lambda or the trace of the within-group-sspmatrix **W**, by exchanging objects between groups. A partitioning cluster analysis starts typically with an initial guess of the cluster seeds, defined as the cluster centroids of the initial partition. Various partitioning algorithms differ with respect to the time when the group centroids are updated during the exchanging procedure (Steinhausen and Langer, 1977). As the total counting of all possible partitions is very time consuming, partitioning algorithms use optimisation methods such as the *Branch and Bound algorithm* to find an optimal partition. As an optimisation method may only lead to a local optimum of a quality criterion, it should always be checked whether the subdivision of data set depends on the choice of the seeds or the sequence of the objects.

2.2. Discriminant analysis

The object of a discriminant analysis is to find out whether a given classification of objects into several groups may be explained by linear combinations—discriminant functions—of independent objects properties that were not used for the classification. In the present context the a priori subdivision of objects is a classification of similar flow patterns into groups, which corresponds to a partition of field sites where a similar type of infiltration was observed. The independent properties are the initial water content, the cumulated irrigation I_C, the irrigation intensity I of the transport experiments and the dimensionless parameters x_{brook} and x_{slope} as surrogates for the macroporousity at the field sites. These five independent properties will in the following be referred to as experimental conditions of the transport experiments. The dependent variable is group membership k, which results from the cluster analysis. During the discriminant analysis a flow pattern is therefore represented by the vector (k, θ, I_C, I, x_{brook}, x_{slope}).

2.2.1. Linear discriminant analysis

The goal of a linear discriminant analysis is to find linear combinations $\mathbf{a}_i \cdot \mathbf{x}_j$ of the p-dimensional property-vectors $\mathbf{x}_j$ in such a way, that the computed discriminant vectors $\mathbf{y}_i$ lead to a maximum separation of the objects with respect to the group membership k (Fahrmeir, 1984; Schuchard-Ficher, 1980). The p-dimensional vector **a** contains the 'discriminant coefficients' and may be obtained by maximisation of the following quadratic function $Q(\mathbf{a})$ (Steinhausen and Langer, 1977; Fahrmeir, 1984), in the present case $p = 5$ because of the number of experimental conditions:

$$Q(\mathbf{a}) = \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}} = \max \rightarrow \frac{d}{d\mathbf{a}}\left(\frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}}\right) = 0$$

$$\rightarrow \mathbf{W}^{-1}\mathbf{B}\mathbf{a} = \lambda\mathbf{a} \tag{2}$$

As the dimension f of matrix $\mathbf{W}^{-1}\mathbf{B}$ is the minimum out of p and the number of groups minus one, the number of the eigenvalues λ_i and, therefore, of discriminant functions is limited to f. The relative part r_{pi} of the variance of the discriminant values that may be explained exclusively by a single experimental condition x_i, is obtained using the standardised discriminant coefficients as follows:

$$r_{pi} = \sum_n e_n s_{ni}$$

e_n is the ratio of the eigenvalue λ_n to the sum of all eigenvalues and s_{ni} is the ratio of the standardised discriminant coefficient $\tilde{a}_{ni}$to their sum, which is the product of the discriminant coefficient a_{ni} and the standard deviation of the variable x_i in the data set.

$$e_n = \frac{\lambda_n}{\sum_{i=1}^{r}\lambda_i}, \qquad s_{ni} = \frac{|\tilde{a}_{ni}|}{\sum_{j=1}^{p}|\tilde{a}_{nj}|}, \qquad \forall n = 1,\ldots,f \tag{3}$$

2.2.2. Quality criteria

There are different quality criteria to judge the results of a discriminant analysis. The Wilks-lambda $\Lambda \in [0,1]$ is a multivariate test quantity to check whether the means of the discriminant values $\langle y_i \rangle$ in

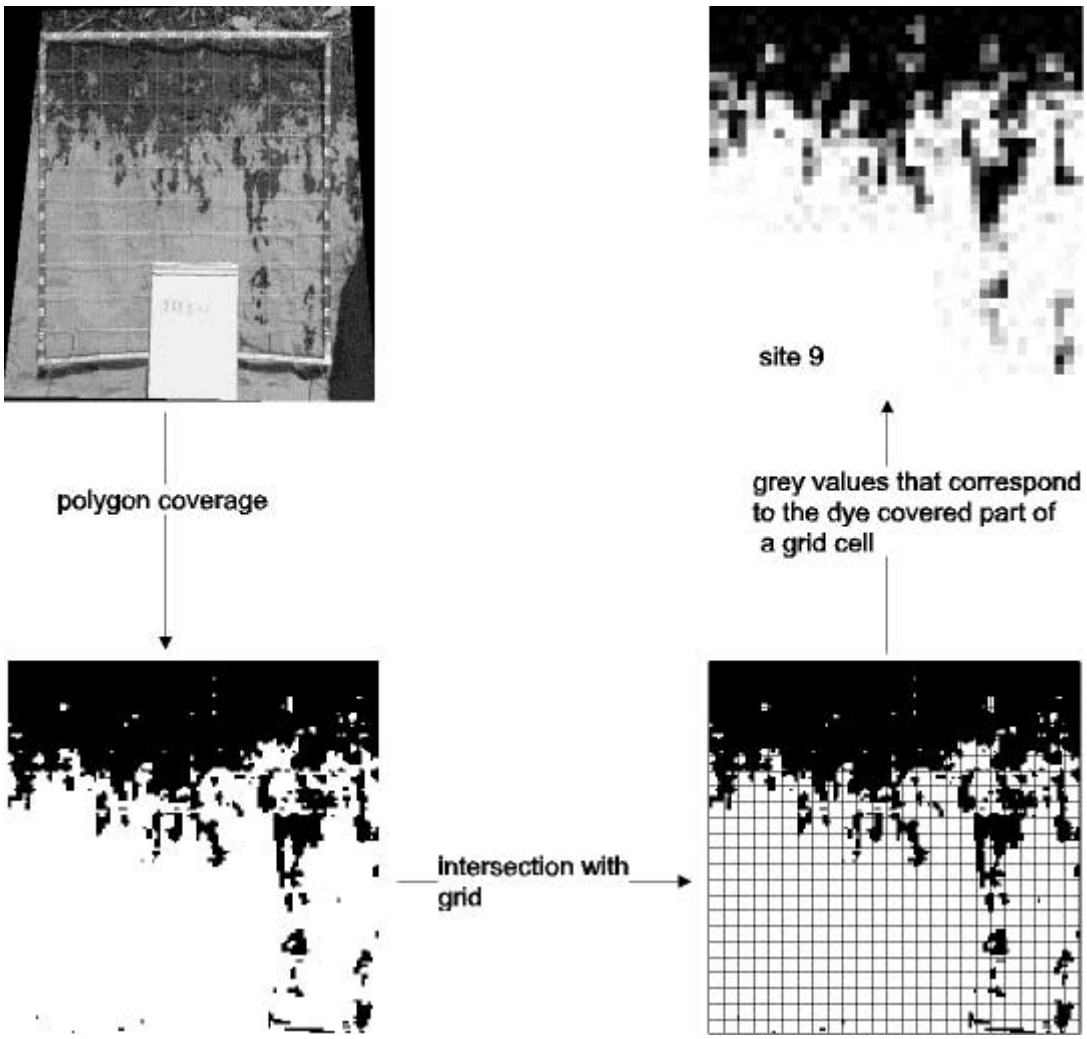

Fig. 2. Transformation of a dye pattern into a grey scale matrix in four steps: the edges of the dye flow pattern were digitised and cleaned into a polygon coverage, the coverage was intersected with a regular grid, the dye covered part of the grid cells were computed and stored into a matrix.

the groups do significantly differ:

$$\mathrm{H}_0 : \langle y_1 \rangle = \langle y_2 \rangle = \ldots = \left\langle y_{\mathrm{g}} \right\rangle \text{ vs } \mathrm{H}_1 : \langle y_i \rangle \neq \left\langle y_j \right\rangle \text{ for at least one pair } i \neq j \ (^{**})$$

The smaller the Λ the higher is the significance of the subdivision. With Λ one may compute a χ^2-distributed test property, to look up the probability $1 - \alpha$ that hypothesis H_1 in (**) is correct:

$$\chi^2 = -\left(N - \frac{p - g}{2} - 1\right) \ln(\Lambda) \qquad DF = p\,(g - 1) \tag{4}$$

$p = 5$ is the number of independent experimental conditions of the tracer experiments, g is the number of groups and $N = 10$ the number of field sites in the data set and DF are the degrees of freedom.

The square of the canonical correlation coefficient

R_{ci} of a discriminant function y_i

$$R_{ci} = \sqrt{\frac{\lambda_i}{1 + \lambda_i}} \tag{5}$$

is a measure for the fraction of the variance of the discriminant values that may be explained by the group membership of the objects in the data set. It may be regarded as the analogue to the R^2 in a multiple linear regression. As individual discriminant function y_i are eigenvectors belonging to different eigenvalues λ_i, they are linear independent but not orthogonal, i.e. the corresponding values of R_c^2 do not sum to 1. Different discriminant functions may reflect the influence of different factors on the partition of the objects of interests, i.e. the partition of similar flow patterns into groups (Fahrmeir, 1984).

2.2.3. Classification of objects

Using the discriminant functions one may predict the expected group membership of an object as a function of its properties as follows:

- compute the discriminant score y as a function of object properties;
- compute the distance between y and the average discriminant values $\langle y_i \rangle$ of each group i;
- classify the object into the group j with the smallest Euclidean distance between y and $\langle y_j \rangle$.

To check the capability of a discriminant function for object classification, the data set is usually subdivided into a calibration subset, which is used to compute the discriminant functions, and a test subset of objects. If the data set is small like in the present case, the 'jack-knife-method' is used: the discriminant function is determined with a subset that contains all but one of the objects, the omitted object is classified and the estimated value of the group membership is compared to the correct value. The repetition of this procedure for all objects gives unbiased results for the estimated expected error rate (Fahrmeir, 1984).

3. Material and methods

3.1. Transformation of dye coverage into grey values

The data base for the present study consists of dye and Br flow patterns obtained from the ten plot scale transport experiments described in Zehe and Flühler (2001). The dye flow patterns were transformed into dye-coverage distributions (Fig. 2) as follows: In the first step, the contrast between dye pattern and background soil was improved by the setting pixels of the Brilliant Blue colour spectrum (72–252) to 255, using the program *Microsoft Picture Publisher 5.0*. Pixel values outside this range were set to 55. Hence, the resulting pictures contain binary information about 'stained' or 'non-stained' areas. Using the GIS *Arcinfo* the edges of the stained areas were digitised and 'cleaned' into a 'polygon coverage' that holds the attribute 255 for stained and 0 for non-stained polygons. This 'polygon coverage' was intersected with a rectangular grid of 2.5 cm^2 cell size, the dye covered parts of the grid cells were computed and stored into a grey-value matrix. The grey values range from 0 to 255 spanning the spectrum from no to full dye coverage. Thus, the dye pattern on a 1 by 1 m cross section of a single vertical soil profiles is in the following represented by a grey value matrix consisting of 40 by 40 grid cells.

3.2. Representation parameters for flow patterns

The choice of appropriate interval scaled parameters that preserve enough information about the main structures of the flow patterns such as flow fingers is essential for the success of the cluster analysis. We define a Br flow pattern as a matrix of concentration values resulting from the soil sampling at the two vertical soil profiles at each site. Due to the geometry of the sampling grid, such a concentration matrix consists of 20 columns and ten rows, each cell is representing a square of 0.1 by 0.1 m. Similar a dye pattern is defined as matrix of grey values, reflecting the dye coverage distribution on the surface of the two vertical soil profiles excavated at each site. Hence, such a grey value matrix consists of 80 columns and 40 rows. We assume statistical homogeneity of the spatial distributions of initial soil water content, macropore volume and saturated hydraulic conductivity in the upper soil horizon at the plots. Thus, the distribution of the Br centre of mass in the columns of the Br flow pattern as well as the distribution of the 'grey value' centre of mass in the columns of a grey value matrix may be regarded as statistical homogenous too.

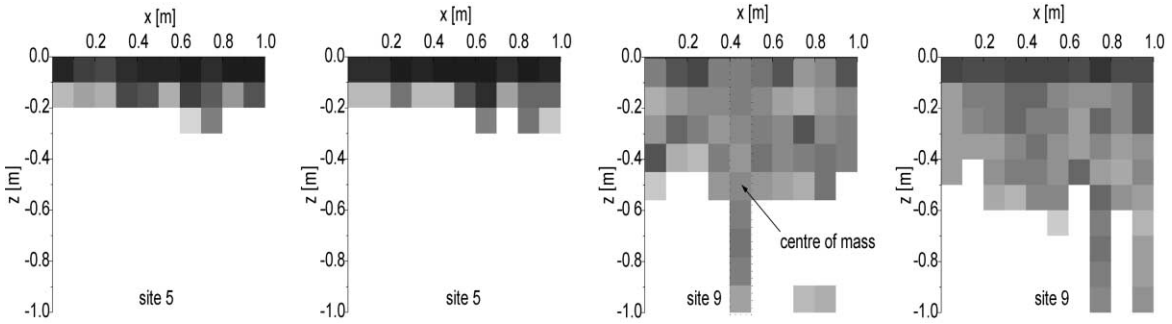

Fig. 3. Two replicated profile cuts exhibiting the Br distribution in a matrix flow pattern (site 5) and in a preferential flow pattern (site 9).

In a first step, the centres of mass were computed for the 20 grids column of a Br flow pattern as well as for the 80 grid columns of the corresponding grey value pattern, as follows:

$$z_{m,j} = \frac{\sum_{i=1}^{N} \rho_i C_{ij}(\mathbf{z}_{i-1} + \Delta z/2)\Delta z^3}{\sum_{i=1}^{N} \rho_i C_{ij} \Delta z^3}$$

z_{mj} is the centre of mass computed for grid column j of a Br or a grey value matrix, C_{ij} is either the grey value of the grid cell, which is dimensionless ranging from 0 to 255, or the residual concentration of Br in the soil (g/kg) in the cell in row i and column j, N is the number of columns (20 or 80) of the matrix, Δz is the size of the grid cells (10 or 2.5 cm) and ρ_i the bulk density.

The results are a distribution of Br centre of mass related to a Br flow pattern as well as a distribution of the grey value centre of mass related to the dye flow pattern observed at a field site. In the following these distributions will be referred to as z_m distribution of a Br or a dye flow pattern, were z_m indicates the depths of the centres of mass.

The z_m distributions of a Br and the corresponding dye flow pattern observed a field site are characterised by computing their mean $\langle z_m \rangle$, standard deviation σ and skewness s [Eq. (6)]. The first two parameters determine the average transport depth and the variability of the centre of mass. The skewness s of the z_m-distribution indicates whether the distribution has an

Table 2
Parameters characterising the distribution of the depth of the tracer centre of mass, computed to represent the dye and Br flow patterns for a cluster analysis

Site	Brilliant Blue			Br		
	$\langle z_m \rangle$ (m)	σ (m)	s (m^3)	$\langle z_m \rangle$ (m)	σ (m)	s (m^3)
1	0.054	0.019	3.5×10^{-6}	0.070	0.015	3.0×10^{-6}
2	0.064	0.015	7.4×10^{-7}	0.070	0.017	4.5×10^{-6}
3	0.070	0.026	2.1×10^{-5}	0.109	0.029	3.5×10^{-5}
4	0.052	0.015	3.1×10^{-6}	0.072	0.010	5.2×10^{-7}
5	0.044	0.020	7.4×10^{-6}	0.062	0.012	1.6×10^{-6}
6	0.051	0.030	3.5×10^{-5}	0.110	0.040	6.4×10^{-5}
7	0.035	0.015	8.1×10^{-6}	0.066	0.008	3.1×10^{-7}
8	0.045	0.013	7.1×10^{-7}	0.069	0.016	1.3×10^{-6}
9	0.171	0.054	2.0×10^{-4}	0.180	0.065	3.0×10^{-4}
10	0.200	0.109	6.8×10^{-4}	0.173	0.072	-1.9×10^{-5}

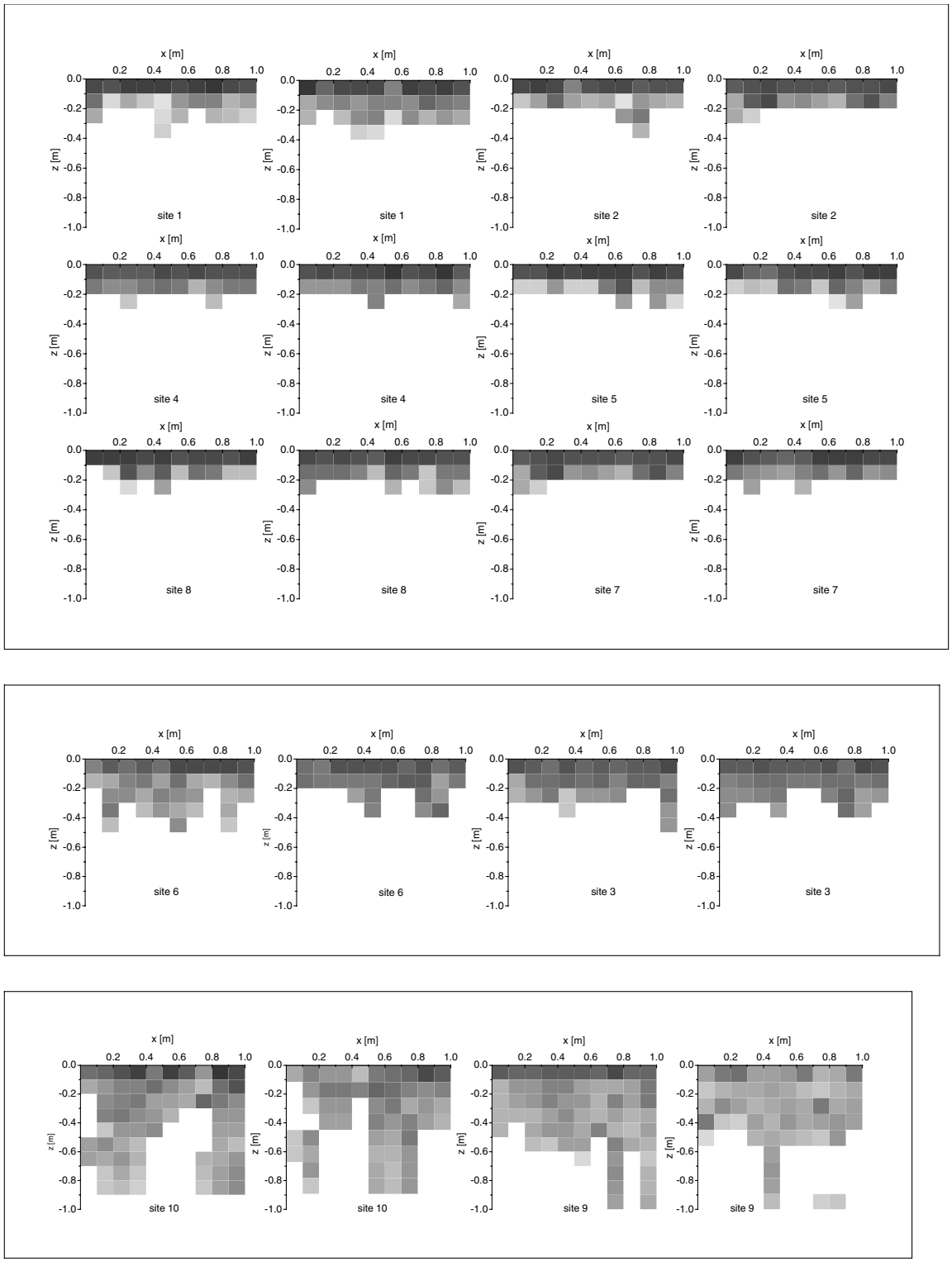

Fig. 4. Subdivision of similar Br flow patterns into groups by means of cluster analysis, the number of groups that satisfied the constraint of $R^2 > 90\%$ is 3.

asymmetric tendency to large vertical transport distances such as flow fingers.

$$\langle z_m \rangle = \frac{1}{N} \sum_{j=i}^{N} z_j^{\mathrm{m}}$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} (z_j^m - \langle z_m \rangle)^2} \qquad (6)$$

$$s = \frac{1}{N-1} \sum_{j=1}^{N} (z_j^m - \langle z_m \rangle)^3$$

Fig. 3 shows Br flow patterns from field site 5 and 9 that stand for a matrix and a preferential flow dominated regime. Table 2 shows the parameters of the corresponding z_{m}-distributions for the entire sample of dye- and Br patterns. As expected, the skewness s of the z_{m}-distribution of the preferential Br flow pattern at site 9 is approximately 100 times larger than the corresponding value of the advancing matrix flow front at site 5.

3.3. Cluster analysis of flow patterns

The subdivision of flow patterns was done with the program package *SAS 6.02* in two steps. First, a hierarchical Ward cluster analysis was performed to minimise the number of groups N_{g} of flow patterns with the constraint of $R^2 \geq 90\%$. The data were not standardised, because the physical dimensions of the parameters of the z_{m}-distribution are regarded as natural weights. In a second step the subdivision was improved using a partitioning algorithm. In order to check whether the result of the partitioning cluster analysis depends on the initial partition, it was repeated four times using different initial seeds.

3.4. Discriminant analysis of the subdivisions

Next, a discriminant analysis was performed to check to which extent a subdivision of flow patterns may be explained by the experimental conditions. Each flow pattern was represented by a vector (k, θ, I_{C}, I, x_{brook}, x_{slope}). The group membership k indicates into which group the flow pattern was classified by the cluster analysis and is the dependent variable. The initial soil water content θ, the irrigation rate I, the cumulated irrigation I_{C}, the normalised distance to the outlet ditch x_{brook} (as a correlate for macroporosity) and the normalised position of a field site at a hill slope x_{slope} are the independent variables (Table 1). The capability of the discriminant functions to classify a flow pattern as a function of the corresponding experimental conditions was checked with the jackknife-method.

4. Results

4.1. Subdivision of similar flow patterns into groups by cluster analysis

The minimal number of groups which satisfy the constraint $R^2 \geq 90\%$ is $N_{\mathrm{g}} = 3$ for the Br flow patterns and $N_{\mathrm{g}} = 2$ for the dye flow patterns. The result of the partitioning cluster analysis did not depend neither on the choice of the seeds nor on the sequence of the flow patterns in the data file. Figs. 4 and 5 show the Br and dye flow patterns sorted by the group membership k. The cluster analysis of both types of flow patterns leads to the same subdivision except for field sites 3 and 6. The Br and dye flow patterns from field sites 9 and 10, both located close to the Weiherbach brook, form a group of pronounced preferential flow patterns. The Br flow patterns from sites 3 and 6 form a single group where a less preferential flow occurred. Group 1 contains sites where 'infiltration fronts' of Br and of the dye Brilliant Blue were observed. The main difference between the subdivisions of dye and Br flow patterns is that the dye flow patterns from field sites 3 and 6 do not form a separate group like the corresponding Br patterns. They belong to group 1 of non-preferential flow patterns.

Fig. 6 shows the parameter $\langle z_{\mathrm{m}} \rangle$ plotted vs σ and $\langle z_{\mathrm{m}} \rangle$ plotted vs s for the subdivisions of both types of flow patterns. The separation between the individual groups is obvious in both cases. The group centroids are the group averages of the parameters $\langle z_{\mathrm{m}} \rangle$, σ and s. They represent an infiltration process class. Comparing the values of the centroids of corresponding groups of flow patterns, i.e. group 1 of dye and group 1 of Br flow patterns, furthermore group 2 of dye and group 3 of Br flow patterns, the differences are small with respect to the measurement errors. Hence, except for field sites 3 and 6, the cluster analysis of Br and dye flow patterns lead to the

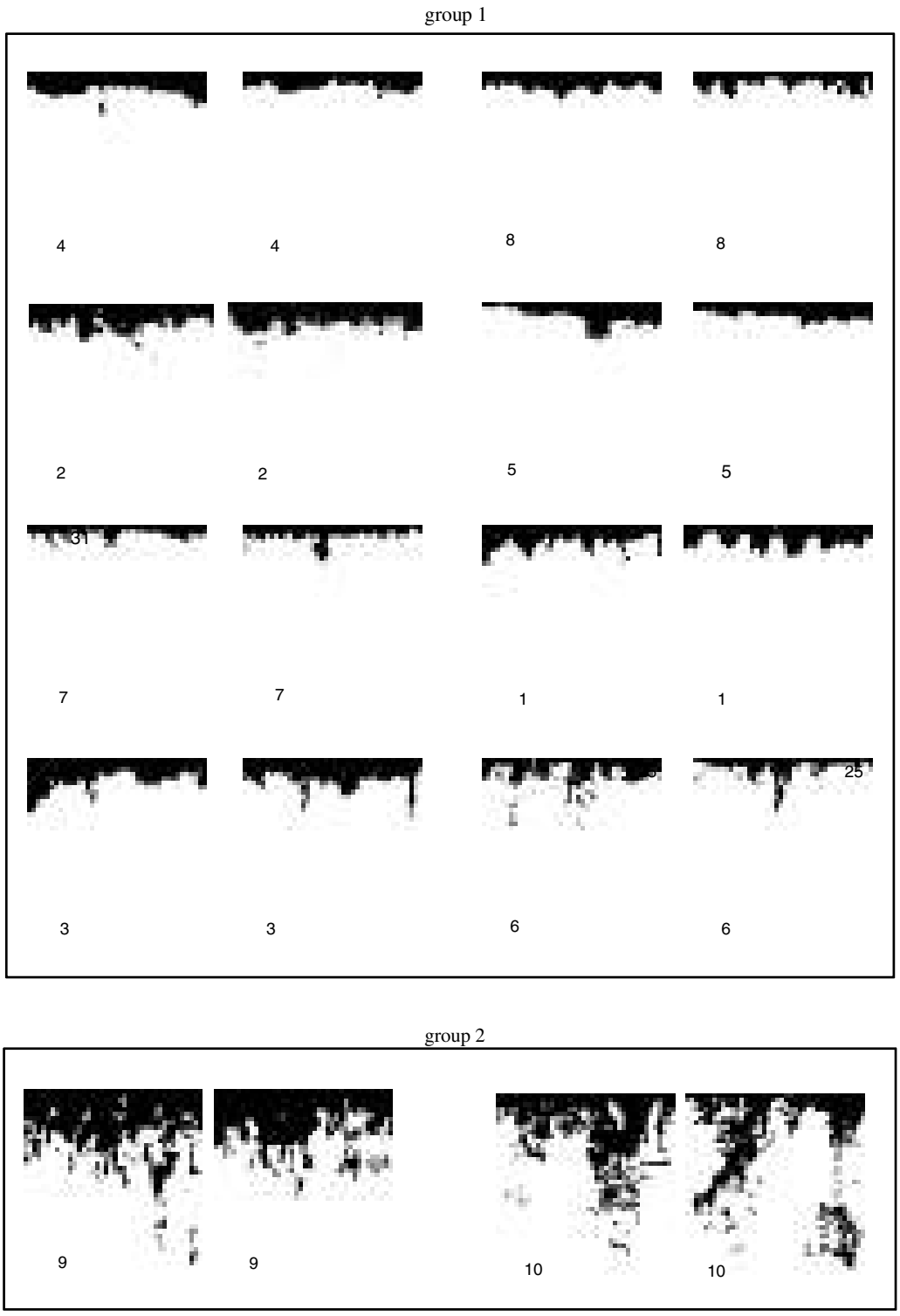

Fig. 5. Subdivision of similar dye flow patterns into groups by means of cluster analysis, the number of groups that satisfied the constraint of $R^2 > 90\%$ is 2.

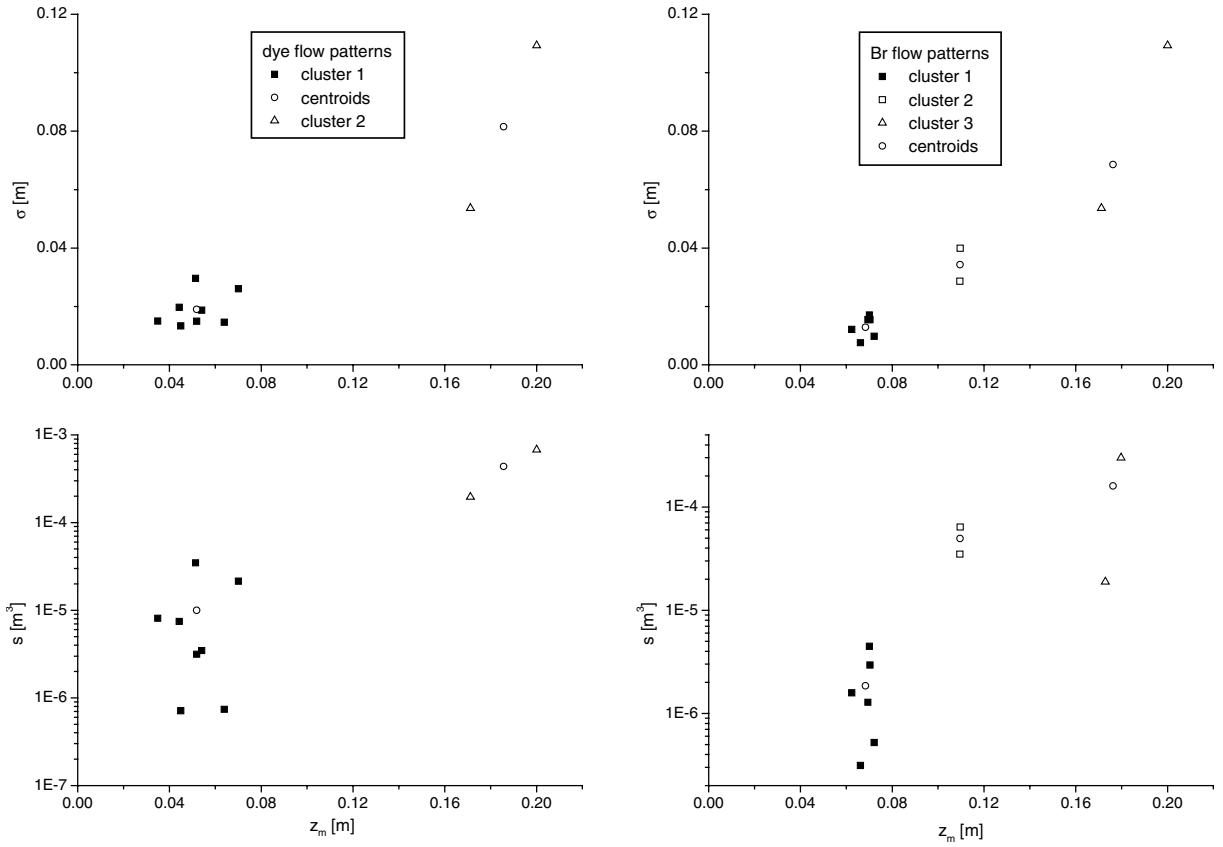

Fig. 6. $\langle z_m \rangle$ plotted vs σ and $\langle z_m \rangle$ plotted vs s for the subdivisions of both types of flow patterns. The separation between the individual groups is obvious in both cases.

same subdivisions with similar infiltration types and corresponding group centroids of the same magnitude.

4.2. Discriminant analysis of the subdivisions

As the number of groups of similar Br flow patterns is $N_g = 3$, two discriminant functions may be computed. In case of the dye flow patterns the number of discriminant functions is limited to one. Table 3 shows different quality criteria for the discriminant analysis of the subdivisions based on Br and on dye flow patterns: the Wilks-lambda Λ, the eigenvalues λ_i to which the discriminant function belongs to, the relative portion e_i of the eigenvalues and the squared canonical correlation coefficients R_c^2 of the discriminant functions y_{Br1}, y_{Br2} and y_{BB}.

Although Λ is small for both subdivisions, the small number of field sites leads to levels of significance $1 - \alpha$ of 84 and 89%. Nevertheless, the values of R_c^2 indicate that in case of the dye flow patterns 84% of the variance of the discriminant values may be explained by the group membership of the sites. In case of the Br flow patterns 94 % of the variance of the discriminant values may be explained using the first discriminant function y_{Br1}. Using the second discriminant function still 32% of the variance may be explained. The relative weight of the discriminant function y_i is indicated by the value of e_i which is 0.81 for y_{Br1} and 0.19 for y_{Br2}.

Table 4 lists the relative portion r_{pi} of the variance of the discriminant values that may be explained by one of the independent parameters exclusively, computed with Eq. (3). For both subdivisions the parameters x_{brook} and x_{slope} explain together approximately 60% of the variance of the

Table 3
Quality criteria for the discriminant analysis of the subdivision of sites based on Br- and dye patterns

Discriminant function	Λ	λ_i	e_i	R_c^2	$1-\alpha$
y_{Br1}	0.04	14.60	0.81	0.94	0.84
y_{Br2}		0.47	0.19	0.32	–
y_{BB}	0.17	4.98	1.00	0.84	0.89

discriminant values. The initial soil water content θ ranks second explaining 21% of the variance. The amount of cumulated irrigation I_C and the average irrigation rate I reflect the influence of 'irrigation parameters' with approximately 19% of the variance of the discriminant values.

The discriminant functions in Eq. (7) predict the expected group membership of the flow patterns resulting of an additional tracer experiment using the corresponding experimental conditions.

$$\text{Bromid}\begin{cases} y_{BR1} = 8.45\cdot x_{brook} + 1.19\cdot x_{slope} - 0.33\cdot\theta + 0.32\cdot I + 0.62\cdot I_c \\ y_{BR2} = 0.18\cdot x_{brook} + 1.27\cdot x_{slope} - 0.29\cdot\theta + 1.39\cdot I - 0.60\cdot I_c \end{cases} \quad (7)$$

$$\text{Brilliant Blue}\{ y_{BB} = 5.56\cdot x_{brook} + 2.64\cdot x_{slope} - 0.25\cdot\theta + 1.53\cdot I. - 0.01\cdot I_c$$

Tables 5 and 6 show the results of the classification tests based on the jack-knife-method for both subdivisions. When the discriminant function is based on the Br data, one single misclassification occurs. Site 3 is classified into group 1 although it belongs to group 2. As group 2 does only contain two field sites, this corresponds to a conditional error rate of $\phi = 0.5$ (Fahrmeir, 1984). When the discriminant function is based on the dye tracer data, no classification error occurs.

Using the a priori probability p that a site belongs to a certain group, which in the present study is the reciprocal value of the number of groups $1/N_g$, the estimated expected error ϵ rate may be computed as (Fahrmeir, 1984):

$$\epsilon = \sum_i p_i \phi_i \quad (8)$$

Table 7 shows the number of groups for both subdivisions, the estimated expected error ϵ_{disc} determined with Eq. (8) and the error ϵ_{chance} that would occur, if the flow patterns were classified into the groups by chance. For both subdivisions the errors rates for the classification based on discriminant functions are lower than the values for a classification by chance.

Table 4
Relative portions r_{Br} and r_{BB} of the variance of discriminant values that may be explained by the a single parameter for both subdivisions

	Parameter	x_{brook}	x_{slope}	θ	I_C	I
Br	r_{Br}	0.462	0.125	0.212	0.096	0.105
Brilliant Blue	r_{BB}	0.418	0.192	0.212	0.175	0.003

Table 5
Result of the jack-knife-test with the discriminant function based on Br data

From group	Classified in group			ϕ
	1	2	3	
1	6	0	0	0
2	1	1	0	0.5
3	0	0	2	0

5. Discussion

5.1. Differences in the subdivisions of dye and Br flow patterns

Except for sites 3 and 6 the cluster analysis of the dye and Br flow patterns lead to the same subdivisions of field sites into groups. Group 1 contains in both cases 'non-preferential' flow patterns that could be characterised as 'infiltration fronts'. Group 2 of dye and group 3 of Br patterns consist of pronounced preferential flow patterns such as in case of site 10 and 9 where both tracers penetrated the subsoil. At these sites flow and transport 24 h after irrigation is

Table 6
Result of the jack-knife-test with the discriminant function based on the dye tracer data

From group	Classified in group		ϕ
	1	2	
1	6	0	0
2	0	2	0

still strongly dominated by the small scale variability of the velocity field, reflecting the macroporous heterogeneity of the soil.

The reason for the discrepancy between the classifications of Br and dye flow patterns from field sites 3 and 6 is that the average transport distance of $\langle z_m \rangle$ for Br is at both field sites 4–6 cm larger than the corresponding value for the dye Brilliant Blue (Table 8). At other field sites this difference is only 1–2 cm (Table 2). Obviously the dye was more strongly retarded relative to Br at sites 3 and 6 than at the other ones. This discrepancy between the classification of Br and dye flow patterns shows clearly, that the average transport distance of $\langle z_m \rangle$ is the parameter with the largest influence on the subdivision of flow patterns.

5.2. Part of the variation of flow patterns explained by different parameters

The results of the discriminant analysis of both subdivisions of flow patterns suggest the same hierarchy of parameters that influence the observed variability of flow patterns on the slope scale. By weighting the values given in Table 4 with the corresponding value of canonical correlation coefficients R_c^2 (0.94 and 0.84) we may compute the part of slope scale variation of the flow patterns that may be explained by the independent parameters. The relative measures x_{brook} and x_{slope} explain together approximately 50% of the observed flow pattern variation. The parameter x_{brook} is a more 'global' measure, characterising the position of a field site relative to the location of the brook (Fig. 1). The local position of a plot at hill slope is described by x_{slope}. Both parameters turned out to be appropriate surrogates for the observed distribution of macroporousity (Schmaland, 1996).

The initial soil water content is on the second rank, explaining approximately 18% of the observed slope scale variation of the flow patterns. The irrigation parameters I_c and I appear to be of less importance, with a part of variance explained by of approximately 17%. However, the coefficients of variation C_v of I_c and I in between the individual irrigation experiments are 0.06 and 0.05 (Table 9), i.e. differences are small. The initial water covers clearly a wider range, especially if expressed in terms of relative saturation S_{rel} with a coefficient of variation of 0.37 (Table 9). Due the values of C_v, the part of flow pattern variance explained by the initial water content appears to be more significant than the corresponding value explained by I and I_c. However, flow and transport in porous media are strongly non-linear, and we believe that the transition from matrix dominated flow to preferential flow is not a smooth, continuous process but a threshold process. From this point of view, also small changes in the experimental conditions may cause this transition.

Table 7
Estimated expected error rates for the classification based on the discriminant functions ϵ_{disc} [computed with Eq. (8)] and for a classification by chance ϵ_{chance}

	N_g	ϵ_{disc} (%)	ϵ_{chance} (%)
Br flow patterns	3	16.6	66.6
Dye flow patterns	2	0.0	50.0

Table 8
Comparison between the z_m-parameters of Br- and dye pattern at field sites 3 and 6

Site	$\langle z_m \rangle$ (m)	σ (m)	s (m^3)
6 BB[a]	0.051	0.030	3.5×10^{-5}
6 Br	0.110	0.040	6.4×10^{-5}
3 BB	0.070	0.026	2.1×10^{-5}
3 Br	0.109	0.029	3.5×10^{-5}

[a] Brilliant Blue.

Table 9
Coefficient of variation C_v computed for the parameters characterising the experimental conditions of the plot scale experiments

Parameter	x_{brook}	x_{slope}	q	S_{rel}	I_c	I
C_v	0.60	0.54	0.20	0.37	0.05	0.06

5.3. Identification of preconditions for preferential flow

The obtained results suggest a hierarchy of necessary preconditions for the occurrence of preferential flow events. A sufficient number of preferential pathways, in our case as sufficient number of earthworm burrows, that link the soil surface and the subsoil appears, in accordance with other studies (Villholth et al., 1998; Stamm et al., 1998), to be the most important one. The typical hill slope soil catena in the Weiherbach catchment determines, due to the habitat preferences of the earthworm *Lumbricus terrestris*, the spatial pattern of macroporousity on the slope scale at least to some extend. This in turn did influence the observed flow and transport: with one exception preferential flow events were exclusively found in Colluvisols, strong events at site 9 and 10 as well as an intermediate type at site 6. These intermediate type of preferential flow pattern was also observed at site 3, which is located at the mid slope sector in a Calcaric Regosol under grassland. However, it is known that grassland favours a high earthworm activity (Ehrmann, 1996), which probably caused a sufficiently high macroporousity at this site. Thus, it appears that there are two concurring influences to determine the spatial pattern of macroporousity and, therefore, to some extend the susceptibility of a location for preferential flow: the time invariant hill slope soil catena on the one hand which is disturbed by a time variant land use pattern on the other hand.

The initial water content turned out to be the most important event driven parameter. The relative initial water saturation at sites 3, 6, 9 and 10, where preferential Br flow patterns were observed, ranges from 0.32 to 0.52 (Table 1). As mentioned in the introduction, the macroporousity was determined at sites 1, 5 and 10. Although the observed macroporousity at site 1 was considerably higher as at site 10, matrix dominated flow was observed, due to the lower initial water saturation of 0.22. This is evidence for the above stated assumption, that switching from matrix flow dominated infiltration to preferential dominated infiltration is a threshold process. The difference between the relative saturation at site 1 (no preferential flow) and site 3 (week preferential flow) is 0.1 (Table 1), the corresponding difference between the initial saturation at site 1 and site 9 (strong preferential flow) is 0.18. The cumulated irrigation and the irrigation rate at site 1 and site 3 are similar, at site 9 these values are even a little lower. Thus, during irrigation of those sites of sufficient high macroporousity with an approximately constant irrigation amount and rate of 22 and 10 mm/h the switching from matrix flow to preferential flow dominated infiltration occurred between an initial relative water saturation between 0.2 and 0.4.

6. Summary and conclusions

The presented combination of a plot scale flow pattern recognition and multivariate statistical analysis yields quantitative information about the slope scale variability of infiltration:

- Flow patterns are represented by parameters that describe the plot scale variation of the vertical transport distance of the tracers centre of mass z_m. Similar flow patterns are objectively classified into groups using a combination of hierarchical and partitioning cluster analysis. The group averaged representation parameters are associated with macroscopic types of infiltration on the plot scale.
- A discriminant analysis quantifies to which extent the group membership of a given field site can be explained by independent experimental conditions expressed as θ, I_C, I, x_{brook}, and x_{slope}. Their relative influence on the spatial structure of infiltration may be identified based on to their relative part in the sum of the standardised discriminant coefficients.
- The type of flow pattern expected for an additional experiment i.e. the group membership can be predicted with the discriminant functions on the base of the corresponding parameters of the experiment. The group averages of the parameters $\langle z_m \rangle$, σ and s, which describe the plot scale statistics of the depth of the centre of mass and give a quantitative measure of the spatial structure of the expected infiltration process in statistical terms.

Br, as well as dye flow patterns, turned out to be a suitable database for this analysis. However, the simple approach to transform the dye coverage of a

soil profile into a grey scale matrix bears systematic errors because areas of high and low colour intensity have the same weight. As the intensity of the dye coverage (often) decreases with increasing depth, this method tends to overestimate the grey values at the leading edge. The method suggested by Forrer et al. (2000) to determine the concentration distribution of Brilliant Blue in a dye flow pattern, would definitely improve the suitability of dye patterns for the presented method.

Approximately 50% of the variation of flow patterns observed at different field sites may be explained by the parameters x_{brook} and x_{lsope}, which were used as surrogates for macroporousity and soil type. The initial water content is on the second rank, explaining 18% of the variation of flow patterns. These results suggest a hierarchy of necessary preconditions that favour the occurrence of preferential flow on the plot scale, of which the most important one is a sufficient number of preferential pathways, in our case earthworm burrows, that link the soil surface and the subsoil. Further there is evidence that during irrigation of field sites of sufficient macroporousity, with an approximately constant amount and rate, matrix flow dominated and preferential flow dominated infiltration patterns occurred in a narrow range of initial water saturation from 0.2 to 0.4. Thus, the transition of matrix flow to preferential flow dominated infiltration on the plot scale appears to be threshold process.

On the slope scale there seem to be two concurring influences that determine the spatial pattern of macroporousity, the most important factor predisposing a site for preferential or matrix flow. The time invariant hill slope soil catena on the one hand: Colluvisols, especially those located close to Weiherbach brook, favour the activity of earthworms due to their higher water retention capacity and the more even water regime, which often leads to more and deeper macropores. On the other hand this time invariant pattern is superposed by the time dependent influence of the land use pattern.

We state that these conclusions could have been drawn, at least to some extend, by inspecting the flow patterns and the corresponding experimental conditions 'by eye', without the invented multivariate analysis. However, the successful treatment of a simple case is a prime precondition for applying the invented analysis to more complex cases, where key parameters are not that obvious as in the present study. The proposed method could be used to 'store' a wide empirical knowledge about event driven parameters combined with morphological parameters in discriminant functions and to use this plot scale information for predictions on a larger scale as described above. The main weakness of this approach is, common to all empirical procedures, the needed data base of flow patterns, that covers a representative range of soils, land use forms and related site and experimental conditions. It is unlikely that a single project or a single team can generate sufficient information. Our study could be used as a base line for proposing a protocol of minimum information for a flow pattern data base. We, therefore, suggest that flow pattern studies should be conducted and published providing sufficient information about the experimental and the soil conditions to define the related parameter vectors.

References

Anderberg, M.R., 1973. Cluster Analysis for Applications. Academic Press, New York 359 p.

Blöschl, G. 1996. Scale and Scaling in Hydrology. Wiener Mitteilungen, Wasser-Abwasser-Gewässer, Band 132, Habilitationsschrift, Technical University Wien.

Dooge, J.C.I., 1986. Looking for hydrologic laws. Water Resour. Res. 22 (9), 46S–58S.

Ehrmann, O., 1996. Regenwürmer in einigen südwestdeutschen Agrarlandschaften: Vorkommen. Entwicklung bei Nutzungsänderung und Auswirkung auf das Bodengefüge. Hohenheimer Bodenkundliche Hefte, Nr. 35.

Fahrmeir, L., 1984. Multivariate Statistische Verfahren. De Gruyter, Berlin 902 p..

Flury, M., 1996. Experimental evidence of transport of pesticides through field soils—a review. J. Environ. Qual. 25 (1), 25–45.

Flury, M., Flühler, H., Leuenberger, J., Jury, W.A., 1994. Susceptibility of soils to preferential flow of water: a field study. Water Resour. Res. 30 (7), 1945–1954.

Flury, M., Leuenberger, J., Studer, B., Flühler, H., 1995. Transport of anions and herbicides in a loamy and a sandy soil. Water Resour. Res. 31 (4), 823–835.

Forrer, I., Kastel, R., Papritz, A., Flühler, H., 2000. Quantifying dye tracers in soil profiles by image processing. Europ. J. Soil Sci. 51 (2), 1–10.

Jury, W.A., Elabd, L., Resketo, M., 1986. Field study of naporpamide movement through unsaturated soil. Water Resour. Res. 22 (5), 749–755.

Kaufman, L., Rousseeuw, P.J., 1992. Finding Groups in Data. Wiley, New York 342 p..

Klemes, V., 1983. Conceptualization and scale in hydrology. J. Hydrology 65, 1–23.

Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a dataset. Psychometrika 50 (2), 159–179.

Mohanty, B.P., Bowman, R.S., Hendrick, J.M.H., Simunek, J., van Genuchten, M.T., 1998. Preferential transport of nitrate in an intermittent-flood-irrigated field: model development and experimental evaluation. Water Resour. Res. 34 (2), 159–179.

Rao, P.S.C., Green, R.E., Balasubramanian, V., Kanehire, Y., 1974. Field study of solute movement in a highly aggregated oxisol with intermittent flooding. J. Environ. Qual. (3), 197–202.

Schmaland, G. (1996). Geologie des Kompartiments, Makroporen. In: Plate, E.J. (Hrsg.) Zwischenbericht 1996 Weiherbachprojekt. Institut für Hydrologie und Wasserwirtschaft, Universität Karlsruhe, 56–58.

Schuchard-Ficher, C., 1980. Multivariate Analysemethoden. Springer, Berlin, Heidelberg 346 p.

Seyfried, M.S., Wilcox, B.P., 1996. Scale and nature of spatial variability: field examples having implications for hydrologic modeling. Water Resour. Res. 31 (1), 173–184.

Späth, H., 1975. Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. Oldenbourg, München 217 p.

Stamm, C., Flühler, H., Gächter, R., Leuenberger, J., Wunderli, H., 1998. Rapid transport of phosphorus in drained grass land. J. Environ. Qual. 27, 515–522.

Steinhausen, D., Langer, K., 1977. Einführung in Methoden und Verfahren der statistischen Klassifikation. de Gruyter, Berlin 206 p.

Villholth, K., Jensen, K., Frederica, J., 1998. Flow and transport processes in a macroporous subsurface-drained glacial till soil—I: Field investigation. . J. Hydrol. 207, 98–120.

Zehe, E., Flühler, H., 2001. Preferential transport of Isoproturon on a plot and a field scale, tile-drained site. J. Hydrology 247, 100–115.