Distributed Hydrological Modelling using Land Use Classification Automatically Derived from Satellite Imagery

Neil Robertson, Roland Burckhard, Andres de Moran and Tak Chan
Heriot-Watt University, Edinburgh, UK

1. INTRODUCTION

Zurich is the largest urban area in Switzerland and lies on the banks of the Zürichsee (Lake Zurich), out of which flows the river Limmat. It is built along the shores of the lake and the river Limmat, within a long straight valley bordered by two ranges of hills, Hönggerberg and Zürichberg-Adlisberg to the North, Üetliberg to the South. This geography was created by glacial processes, the hills being moraine deposits from the glacier that excavated the valley. The original vegetation that covered these hills was a mixed deciduous-perennial forest, some of which is still present, despite the peri-urban environment. As the hill slopes fall towards the Limmat valley floor, the forest has been gradually replaced with built-up areas as the city expanded.

Many small watercourses that form in these hills drain into the Limmat, other minor watercourses or Lake Zurich itself. Where they pass through urban areas, these streams are frequently forced into rigid channels or culverts (Scherrer & Schelble, 2007). As a result, some of these streams flood the surrounding areas during storm events. So far, runoff hydrographs for these streams were determined using lumped models, the hydrodynamic sewer network model BaSYS-GVM and a hydrological response units methodology created by Scherrer and Naef (2003). In this paper, an approach based on the distributed rainfall runoff model Wetspa Extension and automatic surface detection is described in an attempt to refine the runoff hydrograph at the critical locations in the river network.

Distributed rainfall-runoff models are currently an important area of research in hydrology, due to the greater precision they are thought to offer over lumped models, where a larger area, usually a sub-catchment, is assigned a single set of hydrological parameters (slope, imperviousness, land use, etc.). Distributed models, where hydrological parameters are calculated in very small areas, such as square cells (raster models) or very small sub-catchments, were made computationally possible by the advent of Geographical Information Systems. A variety of mathematical, physical and computational approaches have been used to model the interconnected sections present in these systems: The surface channels, the sub-surface water bodies, and the sewer systems in urban areas. Feedback between these sections constitutes one of the major problems in the modelling process, as do the complex physical processes governing water infiltration. This study has considered the specific case of a small catchment within an urban environment.

Amongst the necessary spatial data, land coverage is one of the most important types, since it plays a decisive role in establishing the runoff coefficient. However, this data is difficult to collect on-site, and can change frequently and significantly. This has stimulated research into procedures for automatically calculating land cover categories from satellite images. A variety of different remotely sensed data, such as infrared radiation, and many different techniques, such as wavelets, texture, or shape detection, have been applied in the literature.

This paper is organised as follows. First we summarise our dataset (Section 1.2), then prior work in the field of both image analysis and hydrological modelling (Section 2). We then present our approach to segmenting images into labelled land use areas in Section 3. The theoretical and practical aspects of our hydrological model are presented in Section 4. In Section 5 we show results of the automatically derived land use data used as input to the hydrological model. Finally we conclude and discuss potential avenues for future work in this area.

1.2 Data

The example of Zurich was chosen because the different departments of the City and the Canton are in possession of good quality GIS data:

Geomatik + Vermessung Stadt Zürich (Office for Geomatics and Surveying, City of Zürich)
- Surface use (ESRI Shapefile)
- Contour lines of the catchments (AutoCAD DWG and DXF)
- Orthophotographs (Georeferenced TIFF)
- Digital Elevation models (ESRI GRID)

Entsorgung + Recycling Zürich (Disposal + Recycling Utility Zürich)
- Urban drainage network (Oracle Dump for GeoMedia)
- Recorded flood events (ESRI Shapefile)
- Various reports, studies and results of hydraulic calculations

Amt für Raumordnung und Vermessung, Kanton Zürich (Office for Regional Planning and Surveying, Canton of Zurich)
- Groundwater (ESRI Shapefile)
- Soil map of the Canton of Zurich (ESRI Shapefile)
- Surface water network (ESRI Shapefile)

MeteoSwiss (Swiss Federal Office of Meteorology and Climatology)
- Meteorological data

2. RELATED WORK

2.1 Image Segmentation

Image segmentation is an important topic in computer vision and consequently a variety of methods have been applied to the problem ranging from the application of filter banks to agent-based machine learning techniques. Robertson developed an off-line reflective architecture which learns from a corpus of hand-labelled data (Robertson, 1999). Wavelets have been used to speed up the process of classification (Kim, 2003). Although not applied specifically to aerial images Varma and Zisserman developed an improved method for texture analysis based on the statistics of images, rather than the more popular use of banks of filters (Varma, 2005). Texture is clearly a significant image feature enabling segmentation, as well as colour. Colour-based segmentation is quite popular. It has been shown that RGB gives reasonable performance in a non-parametric scheme but that colour spaces which separate the luminance and chrominance components perform better over a range of imaging conditions (Janssen, 2008). The features colour and texture have been combined but not addressing the complexities of urban data (Duboisson-Jolly, 1998).

2.2 Hydrological Modelling and Flood Risk Assessment

Various studies have been carried out in the Zurich area to assess existing risks from flooding and other natural hazards. Regarding hydrology, an important study has been the Runoff Process Map of the Canton of Zurich (Abflussprozesskarte des Kantons Zürich), authored by Margreth and Naef in 2006 at the request of the Office for Waste, Water, Energy and Air of the Canton of Zurich (AWEL) (Cited in Scherrer & Schelble, 2007). The Runoff Process Map divides unbuilt or unaltered surfaces of the canton into hydrological response units according to the overland and subsurface runoff processes that dominate in each one, following a methodology created by Scherrer and Naef (2003). The specific problems faced by the city of Zurich as an urban area have made it necessary to continue the research with more detail, especially regarding small streams within the city. Scherrer & Schelble (2007) group the sixteen stream catchments within the city boundaries according to their geomorphological, geological and pedological characteristics, and divide their surface into hydrological response units using the results contained in the Runoff Process Map. Fig. 1 shows the proportion of the runoff response for the catchment discussed in this paper:

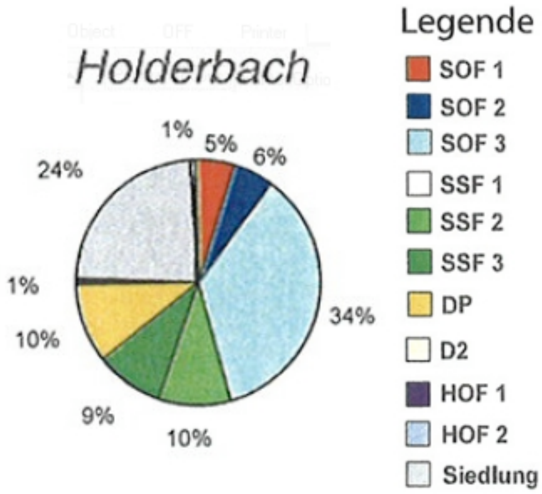


Figure 1: *Proportion of runoff response for each catchment (Scherrer & Schelble, 2007)*

In Figure 1, the different types of runoff response are as follows.:

- SOF: Saturated overland flow
- SSF: Sub-surface flow
- HOF: Hortonian overland flow
- DP: Deep percolation
- D: Runoff in drained areas
- Siedlung: Residential areas

A description of each type can be found in Scherrer & Naef (2003). The numbers refer to the speed of the response: 1 for a fast response, 2 for a slightly delayed response, 3 for a greatly delayed response (Scherrer & Naef, 2003). The Holderbach catchment has a high flood risk, combining a fast runoff response with a high proportion of built-up areas, and occasionally areas with a steep slope, such as deep valleys.

3. AUTOMATIC IMAGE SEGMENTATION

Tiled satellite orthophotos (in TIFF format) from the dataset previously described above were used as basic data, and a series of calculations were performed on them in order to classify each pixel or group of pixels as a certain land type. In order to calibrate the algorithm, segmented orthophotos were used. These images had been manually divided into polygons according to the different land use types using already available surface use data. Single tiles will be used as an example here for illustrating the process. The process was divided into three stages performed offline using Matlab. Fig. 2 shows the process schematic. As can be seen, the first stage and second stage are performed in succession, whereas the third stage is independent of the previous outcomes. When the three stages have been computed, the results are combined probabilistically, and the best outcome is chosen (i.e. we have a maximum likelihood scheme).

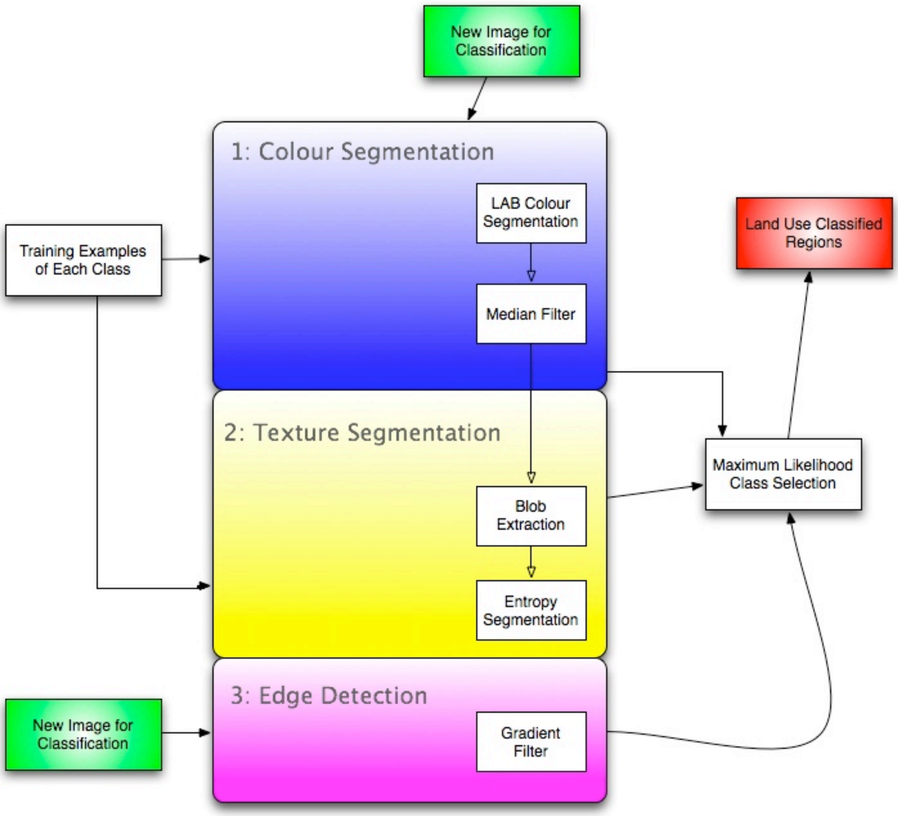


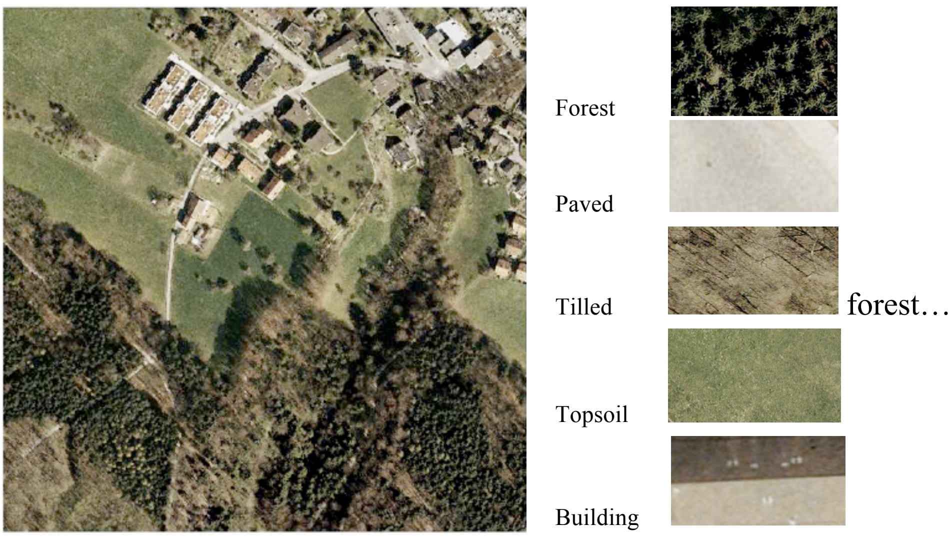
Figure 2: S*chematic of the aerial orthophoto segmentation process*

3.1 First stage: colour similarity

Each orthophoto tile contained several layers with different resolutions, therefore it was necessary to extract one of the layers, with a resolution of 4000x4000 pixels. This layer contained four bands (red, green, blue and an extra empty band). The supplied images had been segmented according to surface use data supplied by the city of Zurich, using a specific classification containing 34 types. These original types were reclassified into five simple land uses, as shown in *Table 1*. Clearly this is a potential source of error since the hydrological model makes use of some land-use categories which are impossible to detect from the images alone. We therefore generalise the categories using the labelling in the official land-use document. Fig. 3 shows an example of input tile and surface use samples. Both the original image and the segmented image were converted from RGB colour space into LAB colour space (Leon, 2006), and then the square Euclidean distance between the original and segmented *a* and *b* values were calculated. Each pixel in the original image was then assigned the land use type for which the distance was minimum. Brightness values were computed for categories 1, 2 and 3, and pixels exceeding a maximum brightness level were reclassified as category 0, since they will not be classified until the later steps. That is, buildings and paved areas have ambiguous colour and are classified more accurately via entropy and gradients (steps 2 and 3). An image filter is applied to the results of the first stage, in order to remove noise (small groups of pixels which are incorrect) (Russ, 1999). The result of the first stage segmentation is shown in Fig. 4(a).

Detailed	Simplified
Building.AgricultureForestryGardening	5-Building
Building.Annexe	
Building.Commerce	
Building.HotelRestaurant	
Building.IndustryCommerce	
Building.PublicAdministration	
Building.Residential	
Building.Transportation	
paved.other_paved.Athletics	4-Paved
paved.other_paved.BuildingPeriphery	
paved.other_paved.Carpark	
paved.other_paved.other_paved	
paved.Pavement	
paved.Pond	
paved.Railways	
paved.Road_Path.Agriculture	
paved.Road_Path.CycleLane_Pedestrian	
paved.Road_Path.Forest	
paved.Road_Path.Road	
paved.TrafficIsland	
tilled.Forest	2,3-Tilled
tilled.other_tilled	
topsoil.Field_Meadow_Pasture	1-Topsoil
topsoil.IntensiveCultivation.other_IntensiveCultiviation	
topsoil.Moor	
topsoil.other_topsoil.embankment	
topsoil.other_topsoil.other_topsoil	
topsoil.other_topsoil.TrafficIsland	
topsoil.Parkland.	
topsoil.Parkland.Cemetary	
topsoil.Parkland.Park	
topsoil.Parkland.Sports	
SurfaceWater.Flowing	(not included)
SurfaceWater.Still	
(not applicable)	0-Unclassified

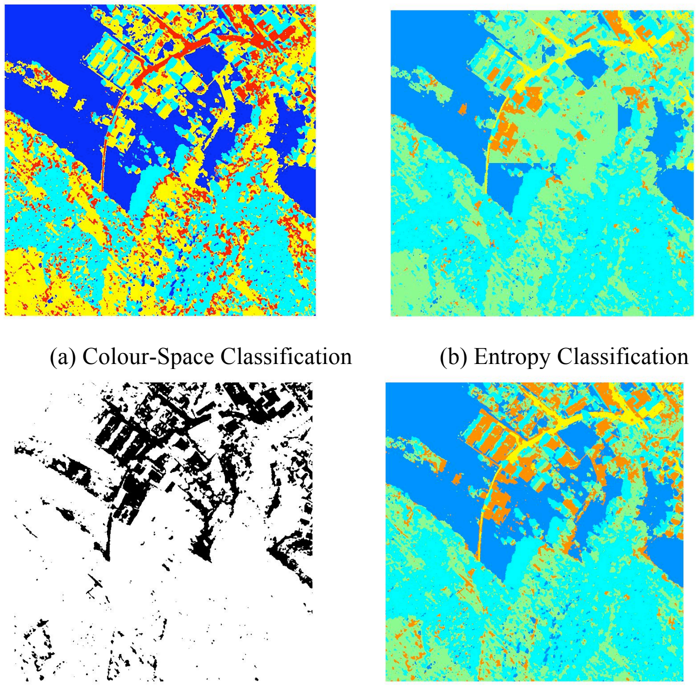
Table 1: *Reclassification of original land use categories*



(a) One tile

(b) Examples of the training classes

Figure 3: *Example input image and samples from the class training dataset.*



(a) Colour-Space Classification (b) Entropy Classification

(c) Gradient Filter (d) Final result

Figure 4: *The outputs of each stage of the segmentation process using the input tile (see Fig. 3(a)).*

3.2 Second stage: texture classification via entropy

Contiguous pixels belonging to the same land use type are grouped into blobs (Haralick, 1992). Image texture is computed for each blob and for the segmented sample image, and these textures are compared. The measure of texture is entropy, which is a metric for the “disorderdness” of the distribution of pixel intensity values. This is given by $E = -\sum p_i \log p_i$ where the probabilities, p, are computed from the image histograms (Kurz, 2006). Textures are measured in each RGB image band separately, and then differences are calculated between each blob and each of the five land use categories, one band at a time. In this way, the blob is assigned a rating per band, and the three band ratings are then added up to obtain the final blob rating. Each blob is assigned the land use type that has the minimum texture difference to the segmented sample of that type. The process from the previous step is applied to RGB values, but considering ranges of values. This results in each blob being assigned a second land use value, that can coincide with the one obtained in step 3 or not. The texture difference percentage and the colour difference percentage for each land use type are multiplied, and the maximum resulting percentage is taken as the land use type. The result of the second stage is shown in Fig. 4(b).

3.3 Third stage: Gradient (edge) detection

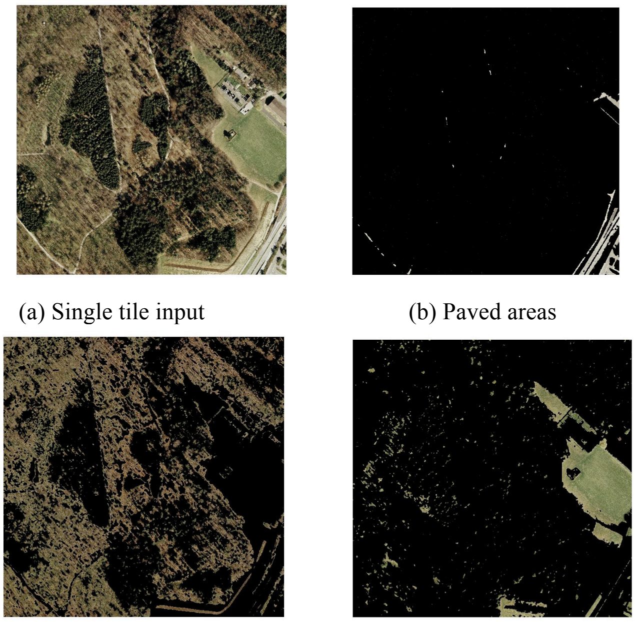
A Sobel edge detection function is required to distinguish between forest and tilled areas on one hand and paved and building areas on the other. This shows improved performance over Prewitt and Roberts. This function measures the gradient of colour intensity values from one pixel to the 8 contiguous ones (Svoboda 2007). Fig. 4(c) shows the result of this operation. Areas identified as forest or tilled surfaces are shown in white. Blobs corresponding to land uses detected in the other two stages (topsoil and paved) are subtracted from the output image, so that the new results are only tilled and building areas. The final result of the third stage segmentation only includes buildings. Fig. 4(d) shows the result of this process.

3.4 Assessment of the accuracy of the segmentation

The accuracy of the three segmentation stages was assessed by comparing the land use category for each pixel in the processed images then calculating the percentage of accurate land use detection in each category (derived from hand-labelled images). Blobs for each land use category from the stage result having the highest detection percentage are combined into the final resulting image. Table 2 shows the final segmentation result as a percentage showing that, where there is an equal distribution of all classes, an 84% classification rate is achieved. An illustrative example is shown in Fig. 5. In practice where there are many buildings the result could be around 65%.

Image \ Class	**Topsoil**	**Till**	**Paved**	**Building**	**Total**
Mainly Topsoil	63	85	56	60	**66**
Mainly Till	63	89	38	65	**84**
Mainly Urban	64	79	51	63	**62**
Even Distribution	80	93	41	64	**84**

Table 2: *Classification rates for the automatic method compared to hand-labelled ground truth computed over the entire test dataset. Note the confusion between Paved and Building classes.*



(a) Single tile input (b) Paved areas

(c) Tilled soil (d) Grassy areas

Figure 5: Examples of three of the classes extracted from a mostly rural section of the Zurich aerial image datset.

4. DISTRIBUTED HYDROLOGICAL MODELLING

4.1 Distributed Modelling

In distributed process models the area being studied is divided into a grid or network of elements, and calculations are carried out in each element, with all parameters discretised over the element grid (Skidmore, 2002). The elements can be linked to estimate water movement (surface and subsurface) between them. Elements are generally smaller than a subcatchment, or according to Vieux (Beven and Moore, 1993:9), "smaller than the scale of the process or system being modelled". Although spatial variability is taken into account from the start, the main disadvantages are that this capability may not be exploitable due to the lack of spatially distributed data, or to the uncertainty caused by a large number of parameters, and large amounts of data to be analysed (Beven & Moore, 1993:9).

The first stage in the evolution of distributed models, brought about by the availability of GIS, was the overlaying of soil, vegetation and topographical spatial data, resulting in the classification of the catchment surface into hydrological response units (HRUs). Runoff was then generated in each HRU, and the results were routed through the catchment, in a similar way to the Ross time-area method (Beven, 2001). This approach still uses the unit hydrograph concept, but in a spatially distributed form. Maidment et al (1996) describe a unit hydrograph model where each cell in a discretised catchment is assigned its own unit hydrograph, and then all the unit responses from all the cells are routed through the catchment to the outlet point.

Various researchers produced models based on explicit equations of all the surface and subsurface processes in a catchment (Beven, 2001). This idea still underpins most distributed hydrological models today. The current state of the art for distributed hydrological models is a continual development of ever more detailed fully explicit models. This brings the problem of parameter overabundance, and consequently very difficult calibration. There is the possibility that simpler models with simplified equations, or a degree of spatial lumping, offer more accurate results. A distinction has thus arisen between models that operate at a catchment, subcatchment, or smaller scale, and those operating at basin scale, reflecting the different types of data available in each case (Beven, 2001).

An example of the specific problems of urban hydrological modelling is described by Djokic and Maidment (in Beven & Moore, 1993:10), such as abrupt changes in drainage patterns caused by the artificially altered terrain and the presence of buildings and other obstacles, which will mean surface routing algorithms will have to be modified. The presence of a drainage network can significantly alter hydrographs and flow patterns. For these authors, this means recommending TIN (Triangulated Irregular Networks) elevation models over grid-based models for hydrological modelling, whereas Smith (1993) proposes an urban model that takes into account a raster representation of the street plan and location of the sewer nodes, and finds that grid models are adequate. Zech and Escarmelle (1999) propose an urban model which uses a modified DEM where buildings have been raised above the ground by a constant amount, whereas sewers have been modelled as trenches through an arbitrary elevation decrease.

4.2. Wetspa Extension

WetSpa Extension is a GIS-based distributed catchment model, developed at the Department of Hydrology and Hydraulic Engineering of Vrije Universiteit Brussel, Belgium. It is based on the earlier WetSpa model, designed for the prediction of Water and Energy Transfer between Soil, Plants and Atmosphere, at a regional or catchment scale, with a fixed daily timescale (Wang et al, 1996, Liu & De Smedt, 2004). Figure 7 shows a diagram of the hydrological processes that are considered in the original WetSpa model, which divides the area being studied into cells.

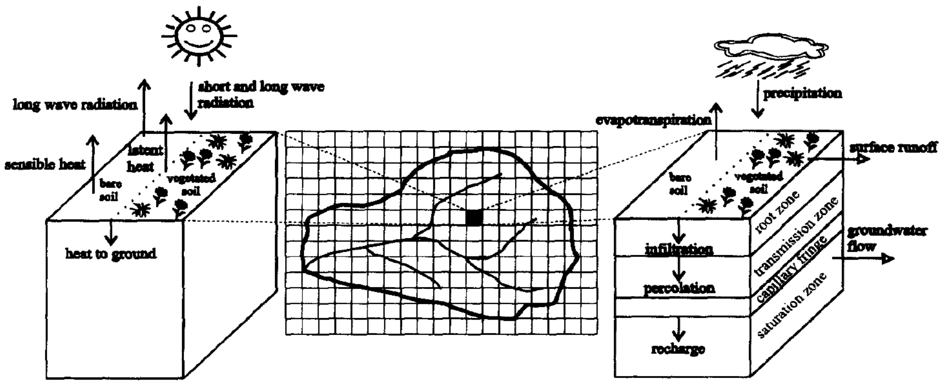


Figure 6: WetSpa model structure (from Wang et al, 1996)

Processes in the atmosphere, vegetation canopy, and within the soil in the root, transmission and saturation zones are all modelled for each cell, where a balance of water and energy is enforced. There is a simplified representation of water movement in the soil in the form of one-dimensional vertical flow. Overland flow follows the Hortonian model, with runoff generation governed by variable source area. A groundwater flow model is integrated using the two-dimensional Dupuit-Forchheimer horizontal flow equation. The water table position is determined through an explicit finite difference scheme for each cell and in each time step (Liu & De Smedt, 2004). WetSpa Extension is a development of WetSpa with the following additions or changes: Simulation time steps can be days, hours or minutes, in order to enable the model to be used for different hydrological modelling applications, such as long-term water resources studies, or short-term event-based flood simulations. The model has been implemented in ESRI ArcView 3.2, with the water balance and flow calculations themselves implemented as MS-DOS executable files programmed in FORTRAN (Liu & De Smedt, 2004).

5. RESULTS

5.1. Comparison of land use classifications

The raster coverage that resulted from the automatic land use classification is shown in Fig. 7 (b). For comparison, the original land use information shapefile supplied by the city of Zurich is shown in Fig. 7 (a). The land use classifications used by the City of Zurich, and the simplified classification implemented in the automatic classification program were different from the categories used by WetSpa Extension. Therefore, it was necessary to reclassify both rasters before modelling could be performed. The resulting rasters are shown in Fig. 8, which also shows a comparison of the two reclassified rasters, with the differing pixels marked in red. As can be seen, there is a majority of correctly classified pixels but the errors are due to the image processing (previously discussed) and the re-classification process.

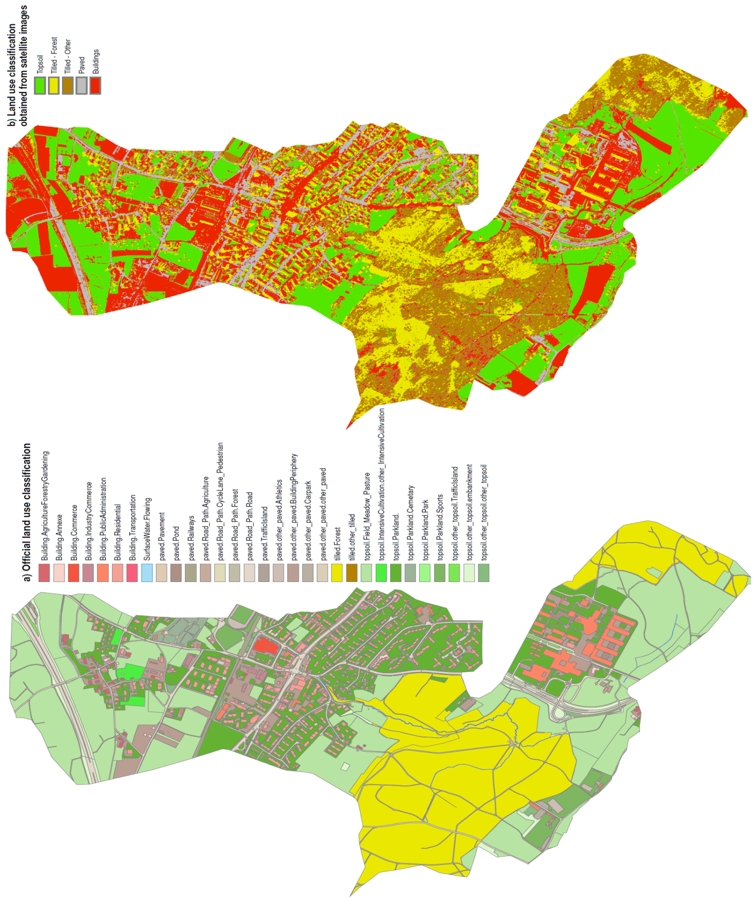


Figure 7: *Original land use classification compared to automatically derived classification*

5.2. Comparison of hydrological modelling results

Fig. 9 shows the result of modelling the Holderbach catchment in WetSpa Extension using the automatically derived land use categories together with the rest of the necessary data, and the synthetic T10-60 rainfall event, which is also shown, and Fig. 10 shows the results for the same catchment when the original, hand-labelled land use categories are used instead.

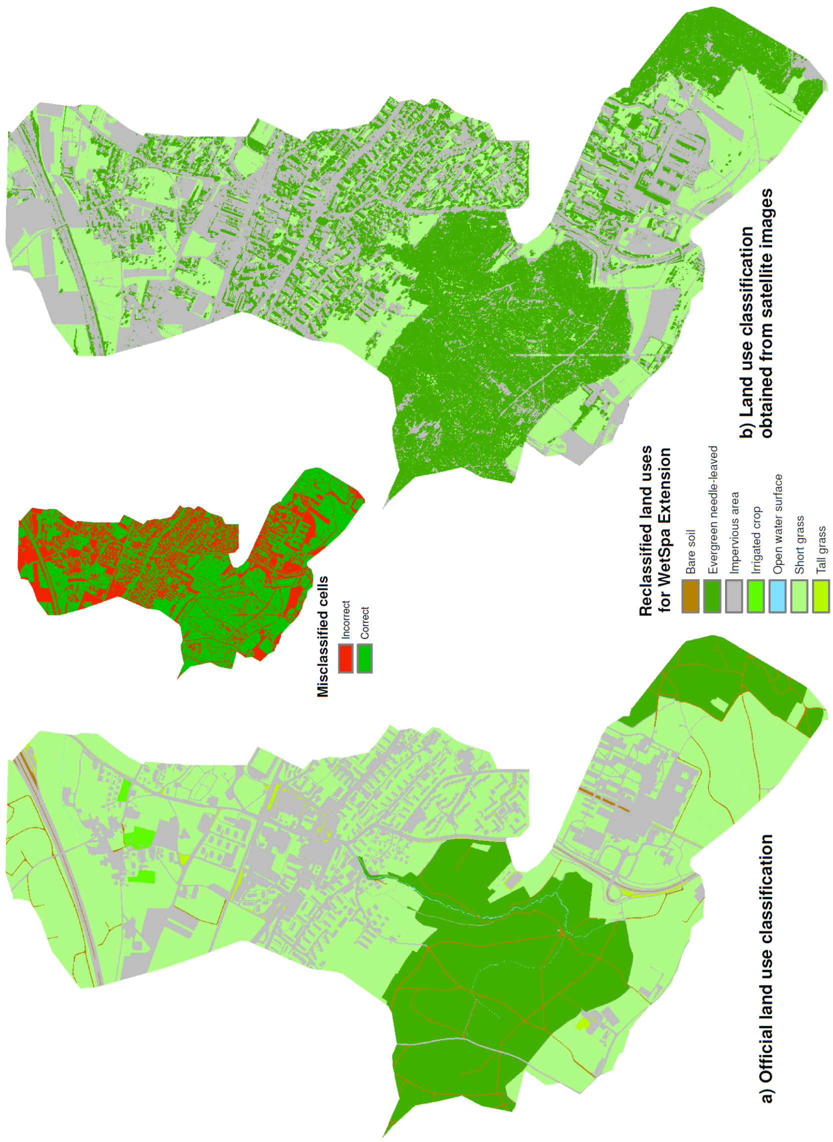


Figure 8: *Original and automatically derived land classifications, reclassified for use in WetSpa Extension*

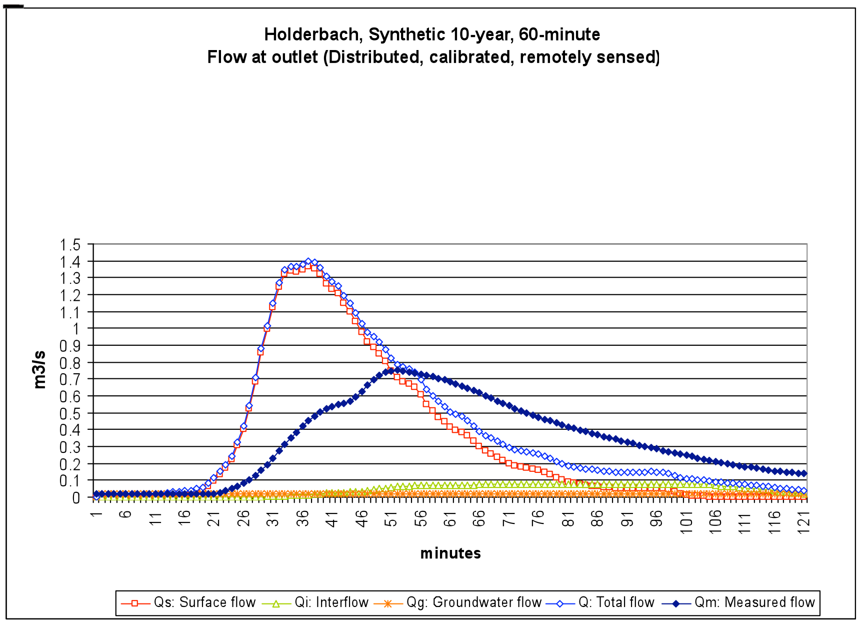


Figure 9: *Modelling results: Distributed, calibrated, using automatic classification*

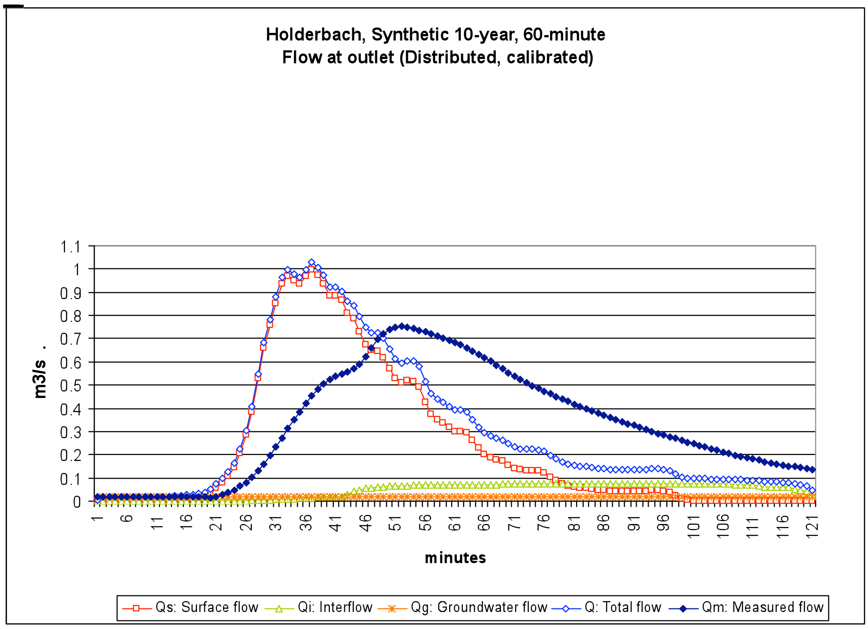


Figure 10: *Modelling results: Distributed, calibrated, using original classification*

6. DISCUSSION and CONCLUSION

Regarding the automatic classification of land use, the algorithm represents a first approach to the problem but serves to highlight the challenges involved in this approach. The divergence between real and detected land use is also be caused by the reclassification that was necessary to use the data within WetSpa Extension, with its specific land use classification. This classification places great importance on the type of vegetation. These details are lacking in the classification used by the City of Zurich, and also could not be distinguished using the applied automatic classification methodology. To perform this reclassification, simplifications were made, such as classifying all woodland areas as containing a specific type of tree. However given that vegetation was detected at 90% accuracy this explains the relative success of the technique. Considering the image processing aspect alone, the main source of error is the misclassification of paved areas as buildings. Although this does not seem to cause difficulty for the hydrological model (due to similar permeability) some improvements could be made using shape as a feature i.e. buildings are predominantly textureless rectangles when viewed from above. Some minor misclassifications also occur due to shadow. There are established techniques which could potentially be applied to remove shadow but further work is necessary to explore their applicability in this domain (Finlayson 2006).

The data used to model the catchment results in similar hydrographs to those obtained using the correct data, despite the proportion of misclassified cells present in the automatically classified land use data. In fact, as can be seen in the figures, in some cases the hydrograph obtained with the automatically classified land uses is more similar to the comparison 'measured' flow than the original hydrograph. This is an indication of the fact that the model is not very sensitive to changes in soil permeability, and other factors. However, the 'measured' flows used for comparison were derived from a different, simpler model than the one used here, and therefore are not necessarily more accurate than the results obtained for this study.

BIBLIOGRAPHY

Beven, K.J., 2001. Rainfall-runoff modelling - The primer 1º ed., Chichister, West Sussex: John Wiley & Sons Ltd.

Beven, K.J. & Moore (Eds.), I., 1993. Terrain analysis and distributed modelling in hydrology, Chichister, West Sussex: John Wiley & Sons Ltd.

Dubuisson-Jolly, M-P. and Gupta A., Color and texture fusion: application to aerial image segmentation and GIS updating, Siemens Corporate Research, 1998

Finlayson, G.D., Hordley, S.D, Cheng Lu, Drew, M.S., On the removal of shadows from images, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006

Janssen, Nils and Robertson, Neil, On the detection of low-resolution skin regions in surveillance images, IEEE Int. Wkshp Visual Surveillance, Marseille, France, 2008.

Kim, Byung-Gyu; Shim, Jae-Ick; Park, Dong-Jo, Fast image segmentation based on multi-resolution analysis and wavelets, Pattern Recognition Letters 24 (2003) 2995–3006.

Kurz, Ludwik and Benteftifa, M. Hafed, Analysis of Variance in Statistical Image Processing, Cambridge University Press, 1997, pp. 128-136

León, Katherine; Mery, Domingo; Pedreschi, Franco and León, Jorge, 2006, Color measurement in L*a*b* units from RGB digital images, USACH

Liu, Y. & De Smedt, F., 2004. WetSpa Extension, a GIS-based Hydrologic Model for Flood Prediction and Watershed Management: Documentation and user manual.

Maidment, D.R. et al., 1996. Unit hydrograph derived from a spatially distributed velocity field. Hydrological Processes, 10, 831-844.

Robertson, P.; Brady, J.M., Adaptive image analysis for aerial surveillance, Intelligent Systems and their Applications, IEEE Volume 14, Issue 3, May/Jun 1999 pp.30 – 36

Russ, John C., The Image Processing HandBook, Third Edition, CRC Press LLC, 1999, pp. 1, 49-60, 174-191

Scherrer, S. & Schelble, G., 2007. Gefahrenkartierung Hochwasser für die Stadt Zürich- Beurteilung von 16 Einzugsgebieten nach der Abflussreaktion.

Scherrer, S. & Naef, F., 2003. A decision scheme to indicate dominant hydrological flow processes on temperate grassland. Hydrological Processes, 17(2), 391-401.

Skidmore (Ed.), A., 2002. Environmental modelling with GIS and remote sensing, London: Taylor & Francis.

Smith, M.B., 1993. A GIS-based distributed parameter hydrologic model for urban areas. Hydrological Processes, 7, 45-61.

Svoboda, Tomas; Kybic, Jan and Hlavac, Valclav, Image processing, analysis and machine vision: a Matlab companion, Paperback, 2007

Varma, M; Zisserman, A, A Statistical Approach to Texture Classification from Single Images, International Journal of Computer Vision, Volume 62 , Issue 1-2 (April-May 2005), Special Issue on Texture Analysis and Synthesis, Pages: 61 – 81, 2005

Wang, Z., Batelaan, O. & De Smedt, F., 1996. A distributed model for water and energy transfer between soil, plants and atmosphere (WetSpa). Physics and Chemistry of The Earth, 21(3), 189-193.

Zech, Y. & Escarmelle, A., 1999. Use of high-resolution geographical databases for rainfall-runoff relation in urbanised areas. Water Science and Technology, 39(9), 87-94.