Carpathian Journal of Earth and Environmental Sciences, February 2014, Vol. 9, No. 1, p. 5 - 10

STREAM FLOW QUANTILE REGRESSION MODELLING USING LAND USE INFORMATION IN THE SPERCHIOS RIVER BASIN (CENTRAL GREECE)

Fotios MARIS, Kyriaki KITIKIDOU, Spyridon PAPARRIZOS & Simeon POTOURIDIS
Department of Forestry & Management of the Environment and Natural Resources, Democritus University of Thrace, GR-68200 Orestiada, Greece. Corresponding author: Kyriaki Kitikidou <kkitikid@fmenr.duth.gr>

Abstract: The estimation of flow in river basins is frequently required in hydrological practice and is of great economic significance. This study uses streamflow and catchment characteristic data, including land use information, from the Sperchios river basin (Central Greece) and applies quantile regression modelling. From 100% of the observations, 20% were selected randomly and used for independent testing of the quantile regression. The test observations and observations used for the model development were found to have similar characteristics. The quantile regression method generally provides accurate flow estimates; however, these techniques involve a chance of large errors under particular conditions, and provisions should be made accordingly.

Keywords: flow estimation; land use; modelling; quantile regression.

1. INTRODUCTION

Flow estimation in river basins is a common problem in hydrologic practice. There are several methods generally adopted for this task, including the Probabilistic Rational Method (I. E. Aust., 1997), the Quantile Regression Technique (Rahman, 2005, Haddad, 2006), Artificial Neural Networks (Glezakos et al., 2009; Iliadis et al., 2011) and others.

Precipitation regression models integrating statistical and GIS techniques have become widespread and common (Brown & Comrie, 2002). In recent years, geographic and topographic factors have been integrated into the modelling of precipitation (Johansson & Chen, 2005). Authors have used weighting functions to incorporate gauge data of neighbouring topographic facets for regressions (Daly et al., 2002), while others have developed models relating climate to site position and elevation (Goodale et al., 1998)

It is more accurate to estimate runoff on the basis of geophysical characteristics, as presented by Willmott et al., (1985), Dickinson et al. (1993) and De Smedt et al., (2000). Due to the complexity of hydrological processes and basin characteristics, physically based distributed models using GIS techniques are becoming popular. Naden (1992) and Troch et al., (1994) presented spatially distributed hydrologic rainfall-runoff models, including hillslope overland flow and channel flow routing. Muzik (1996) presented a method for surface runoff routing using a GIS-based distributed unit hydrograph. Schumann & Funke (1996) applied a two-dimensional instantaneous unit hydrograph within a GIS framework. Maidment et al., (1996) presented an elaborate grid-based model, where the spatially distributed unit hydrograph is derived from a time-invariant flow velocity field. Olivera & Maidment (1999) proposed a method for routing spatially distributed excess precipitation over a watershed using response functions derived from a digital terrain model. Some of these physically based distributed models have obtained worldwide recognition, including Topmodel (Beven & Kirkby, 1979; Beven, 1991) and SHE (Abbott et al., 1986; Refsgaard & Storm, 1995).

The main focus of this paper is to present a physically based distributed hydrological model that uses detailed basin characteristics to predict flow using quantile regression.

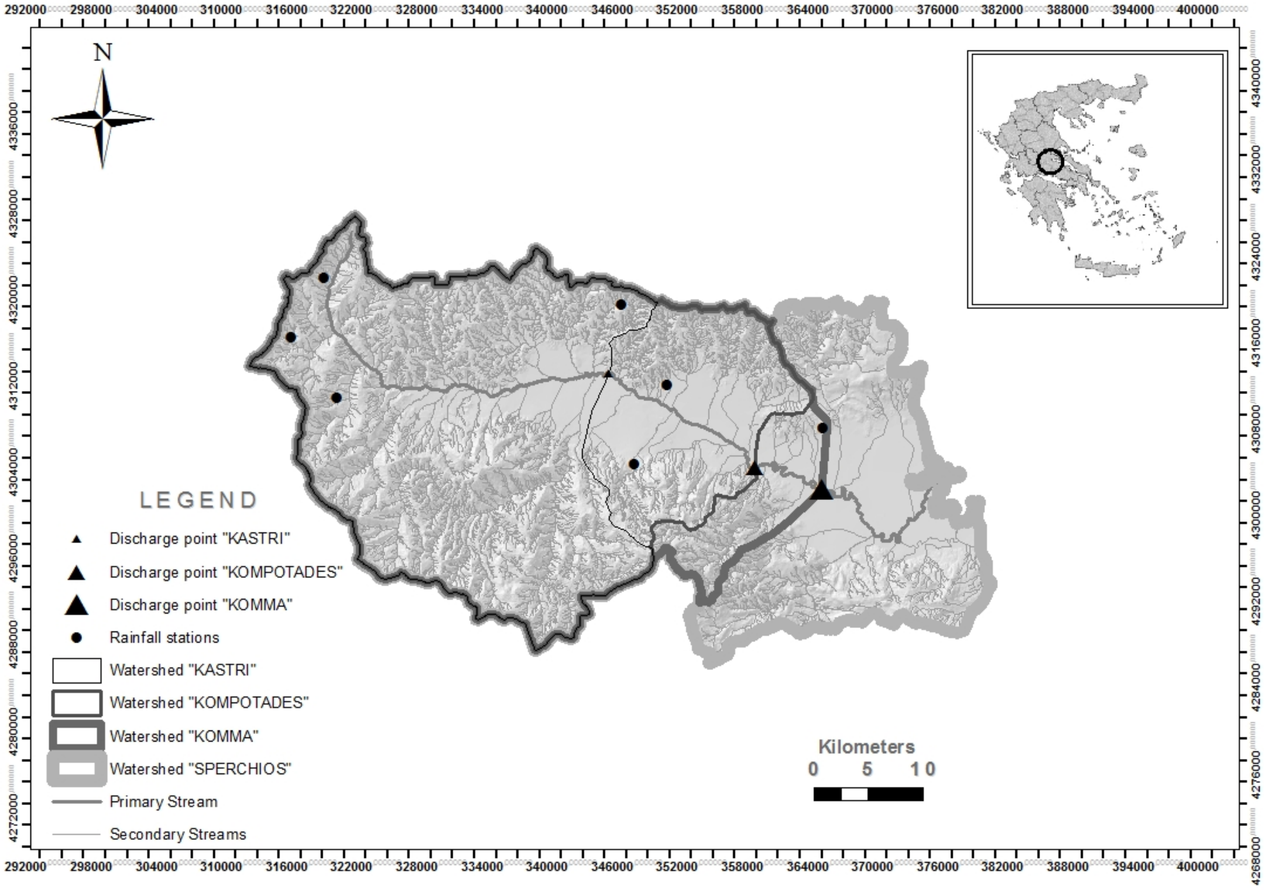

Figure 1. Sperchios river basin map.

2. MATERIALS AND METHODS

2.1 Study area

This study was conducted in the Sperchios River basin, located in Central Greece (Fig. 1).

Table 1. Discharge point descriptions.

Name	Kastri	Komma	Kompotades
Area (km^2)	870.55	1282.55	1170.99
Perimeter (km)	155.90	181.69	174.42
Mean slope (%)	15.13	13.71	13.91
Mean altitude (m)	795.65	705.96	705.00
Roundness index	0.45	0.49	0.48
Length of main stream (km)	45.89	72.99	62.27
Length of secondary streams (km)	2600.69	3694.57	3426.44

The basic characteristics of the three discharge points are given in tables 1 and 2. Coordinates of the raingauge stations are given in table 2, according to the Greek Geodetic Reference System 1987 (GGRS87), which is a geodetic system commonly used in Greece. GRS80 (Geodetic Reference System 1980) is a geodetic reference system consisting of a global reference ellipsoid and a gravity field model. Although GGRS87 uses the GRS80 ellipsoid, the origin is shifted relative to the GRS80 geocenter, so that the ellipsoidal surface is best fitted for Greece (Delikaroglou, 2008). The specified coordinates are measured in meters.

Table 2. Rain gauge station coordinates.

Discharge points		Stations						
		Timphristos	Neochori	Trilofo	Pitsiota	Zilephto	Ipati	Lamia
Kastri	Long.	319,509.57	315,128.87	345,728.34	318,201.90	-	-	-
	Lat.	4,309,543.16	4,315,427.67	4,318,435.32	4,320,789.12	-	-	-
Komma	Long.	319,509.57	315,128.87	345,728.34	318,201.90	349,608.86	347,121.90	364,389.66
	Lat.	4,309,543.16	4,315,427.67	4,318,435.32	4,320,789.12	4,310,940.68	4,303,479.81	4,306,936.53
Kompotades	Long.	319,509.57	315,128.87	345,728.34	318,201.90	349,608.86	347,121.90	-
	Lat.	4,309,543.16	4,315,427.67	4,318,435.32	4,320,789.12	4,310,940.68	4,303,479.81	-

Table 3 and figure 2 were generated based on GIS land use digital data and the CORINE land cover (CLC) 2000 database (CLC, 2000).

2.2 Quantile Regression Technique

The United States Geological Survey (USGS)-proposed Quantile Regression Technique (QRT) was applied at seven gauged catchments of three discharge points in the Sperchios river basin. The flow quantiles are estimated from recorded streamflow data, which are then regressed against relevant catchment characteristic variables that direct the flow generation process (Benson, 1962; Cruff & Rantz, 1965; Riggs, 1973). The quantile regression method is expressed as follows:

$$Q_Y = aB^b C^c D^d \dots$$

where Q_Y is the flow magnitude with a Y year average recurrence interval (flow quantile); B, C, D, … are catchment characteristic variables (predictors); and a, b, c, d, … are regression coefficients.

The independent variables considered in the model are the discharge points' areas (km^2), perimeters (km), mean slopes (%), mean altitudes (m), roundness indexes, lengths of main and secondary streams (km), land use classifications (km^2) and mean annual precipitations (mm) for the years 1981-1982.

The output of the model is the yearly flow of the river (m^3/sec). That the response function can be obtained using standard GIS techniques is highly convenient.

Table 3. Land use area (km^2) and percentages of the discharge points.

Land use (CORINE 2000 classification)	Kastri		Komma		Kompotades	
	Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
Broad-leaved forest (311)	156.51	17.98	161.14	12.56	159.91	13.66
Transitional woodland-shrub (324)	151.93	17.45	207.71	16.20	182.94	15.62
Mixed forest (313)	111.54	12.81	130.23	10.15	115.35	9.85
Land principally occupied by agriculture, with significant areas of natural vegetation (243)	261.60	30.05	443.88	34.61	401.38	34.28
Coniferous forest (312)	93.94	10.79	132.48	10.33	121.53	10.38
Natural grasslands (321)	20.92	2.40	32.62	2.54	29.45	2.51
Sclerophyllous vegetation (323)	73.98	8.50	174.36	13.59	160.31	13.69

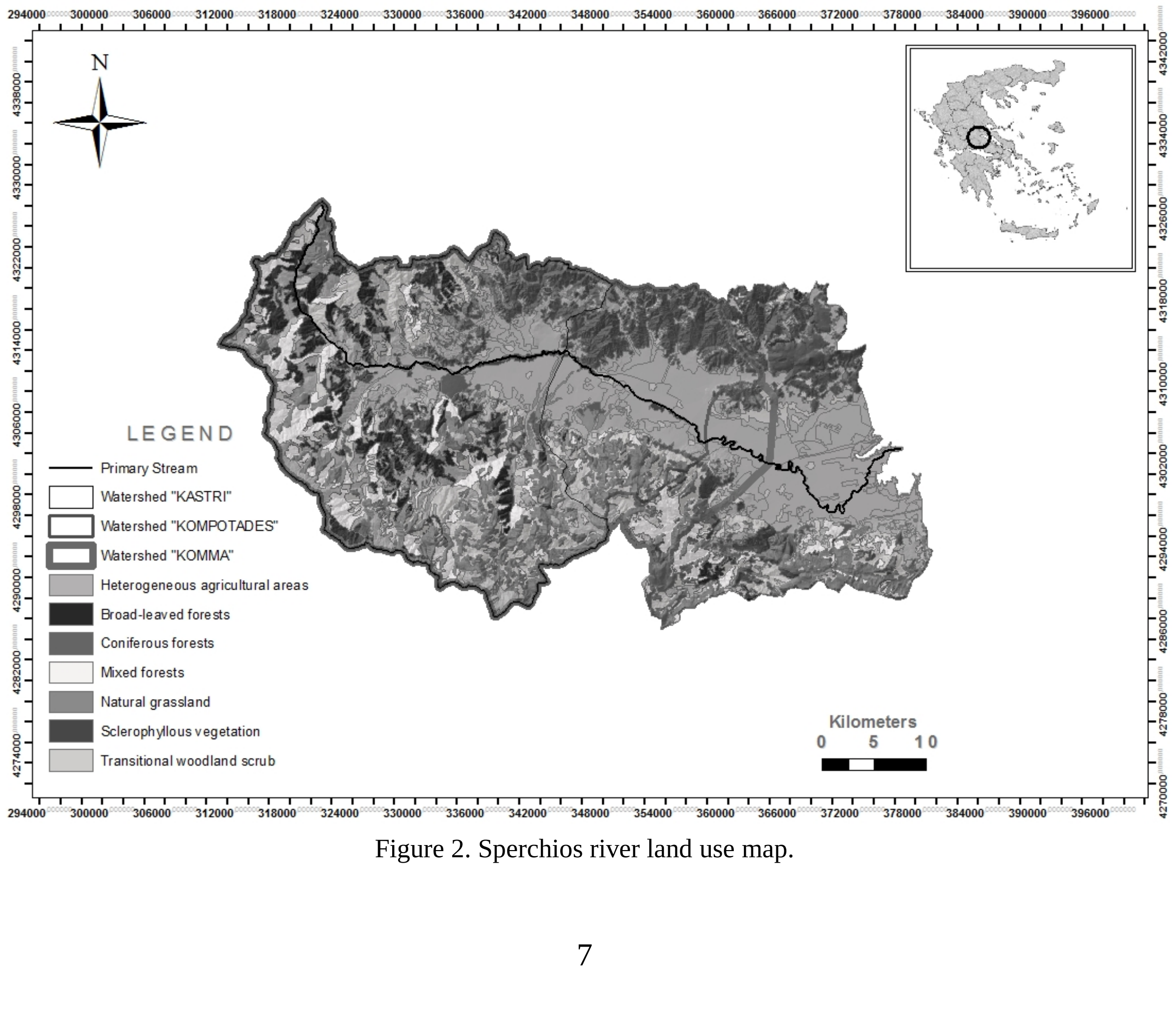

Figure 2. Sperchios river land use map.

The analyses were performed using GIS ArcView 9.3 (van Lammeren et al., 2009) and an extension command SPSS_QUANTREG that allows the performance of quantile regressions using the R package QUANTREGSPSS 19.0 (IBM, 2012).

3. RESULTS

Each of the flow quantiles was regressed against the 11 predictor variables using the SPSS extension command QUANTREG. A number of alternative models were developed for each of the quantiles. The model that showed the highest coefficient of determination (R^2) and the lowest standard error of estimate (SE) and that satisfied the model assumptions most closely (as discussed below) was selected.

The regression coefficients in the prediction equation were found to be significantly different from zero (at a significance level of 0.05). The value of R^2 is very high (0.915 for the fitting and 0.938 for the testing data), and the SEs are acceptable (8.45% and 6.19% of the mean observed flow quantile for the fitting and testing data, respectively).

The selected regression equation was checked against the least squares assumptions (Norusis, 2000). The normal cumulative probability plot did not show a significant departure from a straight line, indicating that the residuals were approximately normally distributed (Fig. 3).

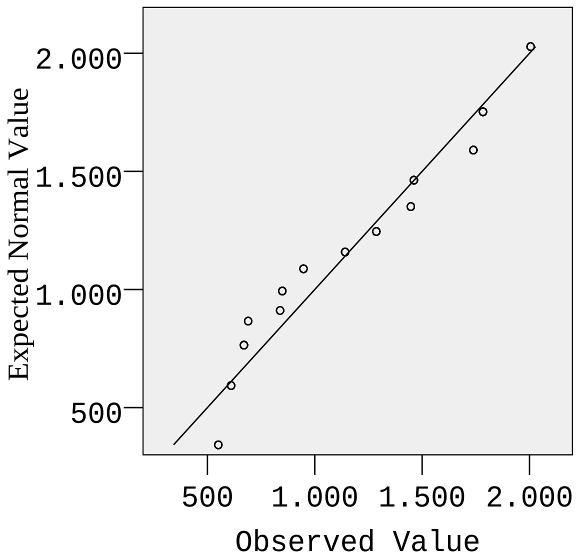

Figure 3. Normal quantile-quantile plot of residuals for *Q*.

The plots of the standardised residuals against the standardised predicted values did not show any systematic patterns between the predicted values and the residuals (Fig. 4). These results indicate that the assumptions of the linear model and the homogeneity of variance were largely satisfied for the prediction equations.

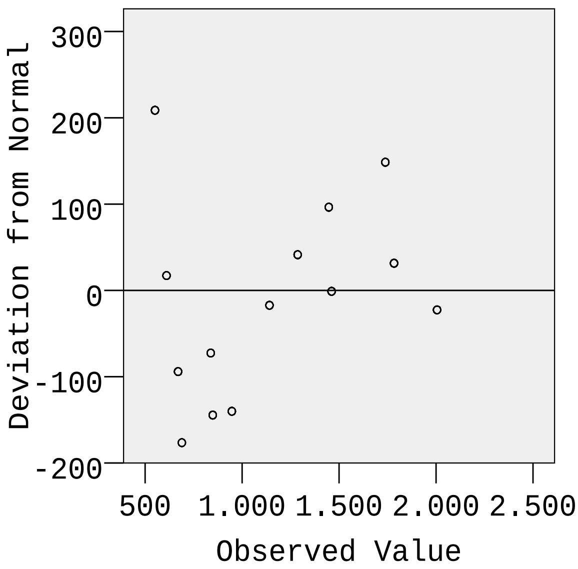

Figure 4. Plot of standardised residual and standardised predicted values for *Q*.

The value of the Durbin-Watson statistic was found to be within the range 1.55-1.65, which is close to 2; thus, the residuals are not highly correlated. The values of the Durbin-Watson statistic range from 0 to 4, and a value of 2 indicates the absence of any correlation. Neither outliers nor extreme data points were found.

The selected prediction equation is:

$$Q = 2{,}388e\text{-}010 \times Year^{-0.045} \times Area^{-0.132} \times Perimeter^{0.265} \times Mean\ Slope^{0.573} \times Mean\ Altitude^{0.322} \times Roundness\ Index^{-8.166} \times Main\ Stream\ Lenght^{3.929} \times Secondary\ Streams\ Length^{-0.128} \times Land\ Use^{0.009} \times Land\ Use\ Area^{0.001} \times Mean\ Yearly\ Precipitation^{0.002}$$

A comparison of the calculated and observed discharges is presented in figure 5. A reasonable agreement between the model results and the observed data is seen. Because the model is clearly overparameterised, we do not intend to optimise its performance by calibration. In addition, hydrological observations are never completely free from error. Moreover, due to the natural variability of hydrologic processes and the complexity of basin characteristics, it is evident that mathematical models will never reach perfection. Hence, we accept the results as they are.

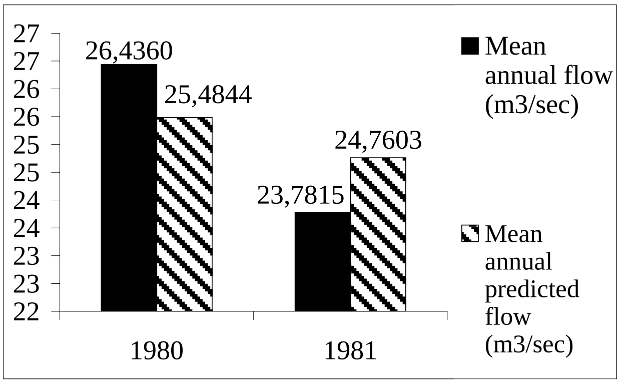

Figure 5. Observed and predicted discharges.

4. DISCUSSION

This study applies Quantile Regression for flow estimation in a Central Greece river basin. The following conclusions can be drawn from this study:

- The Quantile Regression method provides accurate flow estimates.
- There is approximately an 8% chance that the error in flow estimates will exceed 100%. The users of these techniques should be aware of these large errors and make the necessary provisions accordingly.

The developed model is a physically based distributed hydrological model for predicting the hydrologic behaviour in a river basin. The generation of flow depends on rain intensity, which depends on slope and land use. The runoff is subsequently routed through the basin along flow paths determined by the topography using a diffusive wave transfer model that leads to response functions between any start and end point, depending on the slope, flow velocity and dissipation characteristics along the flow lines. The model uses detailed basin characteristics, and the calculations are performed with standard GIS tools. Thus, the model is especially useful for analysing the effects of topography and land-use on the hydrologic behaviour of a river basin. Finally, such information can enable us to determine the influence of changes in land use on the hydrological behaviour of the river basin.

REFERENCES

Abbott, M., Bathurst, J., Cunge, P., O'Connell, P., & Rasmussen, J., 1986. *An introduction to the European Hydrological System – Système Hydrologique Européen, 'SHE', 2. Structure of a Physically-based, Distributed Modelling System*. J. Hydrology 87, 61-77.

Benson, M., 1962. *Evolution of methods for evaluating the occurrence of floods*. U.S. Geol. Surv. Water Supply Pap., 1580-A, 30pp.

Beven, K., 1991. *Spatially distributed modeling: Conceptual approach to runoff prediction*. In: Bowles, D., O'Connell, P. (eds), Recent Advances in the Modelling of Hydrologic Systems, Kluwer Academic, pp. 373-387.

Beven, K., & Kirkby, M., 1979. *A physically-based variable contributing area model of basin hydrology*. Hydrol. Sci. Bull., 24(1), 43-69.

Brown, D., & Comrie, A., 2002. *Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA*. Climate Research 22, 115-128.

CORINE land cover, 2000. *CLC2000 seamless vector database*. <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-seamless-vector-database/>

Cruff, R., & Rantz, S., 1965. *A comparison of methods used in flood frequency studies for coastal basins in California*. U.S. Geol. Surv. Water Supply Pap., 1580-E, 56pp.

Daly, C., Gibson, W., Taylor, G., Johnson, G., & Pasteris, P., 2002. *A knowledge-based approach to the statistical mapping of climate*. Climate Research 22(1), 99-113.

Delikaroglou, D., 2008. *The Hellenic Positioning System (HEPOS) and its foreseeable implications on the Spatial Data Infrastructure in Greece*. Tech. Chron. Sci. J. TCG I(1-2), 95–103.

De Smedt, F., Yongbo, L., & Gebremeskel, S., 2000. *Hydrologic modelling on a catchment scale using GIS and remote sensed land use information*. In: C. Brebbia (ed.), Risk Analysis II, WIT press, Southampton, Boston, pp. 295-304.

Dickinson, R., Henderson-Sellers, A., & Kennedy, P., 1993. *Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as coupled to the NCAR community climate model*. NCAR Technical Note, National Center for Atmospheric Research, Boulder, Colorado, pp. 40-41.

Glezakos, T., Tsiligiridis, T., Iliadis, L., Yialouris, C., Maris, F., & Ferentinos, K., 2009. *Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds*. Neurocomputing 73, 49–59.

Goodale, C., Aber, J., & Ollinger, S., 1998. *Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model*. Climate Research 10, 35-49.

Haddad, K., Rahman, A., & Weinmann, E., 2006. *Design Flood Estimation in Ungauged Catchments by Quantile Regression Technique: Ordinary Least Squares and Generalised Least Squares Compared*. In: 30th Hydrology & Water Resources Symposium: Past, Present & Future. Sandy Bay, Tas., 110-115.

IBM, 2012. *IBM® SPSS® Supplementary Modules - Statistical Tools*. <https://www.ibm.com/developerworks/mydevelo

perworks/files/app/collection/bbe88aaf-f3cd-466a-83fb-592d48eecb1c>

Iliadis, L., Maris, F., & Tachos, S., 2011. *Soft computing techniques toward modeling the water supplies of Cyprus.* Neural Networks 24, 836-841.

Institution of Engineers, Australia (I. E. Aust.). 1997. *Australian Rainfall and Runoff: A Guide to Flood Estimation*. Book IV, I. E. Aust., Canberra. 374 p.

Johansson, B., & Chen, D., 2005. *Estimation of areal precipitation for runoff modeling using wind data: A case study in Sweden*. Climate Research 29(1), 53-61.

Maidment, D., Olivera, J., Calver, A., Eatherall, A. & Fraczek, W., 1996. *A unit hydrograph derived from a spatially distributed velocity field.* Hydrologic Processes 10, 831-844.

Muzik, A., 1996. *GIS-derived distributed unit hydrograph*. In: Kovar, K., Nachtnebel, H. (eds), HydroGIS ’96, Vienna, IAHS Publ. No. 235, pp. 453-460.

Naden, P., 1992. *Spatial variability in flood estimation for large catchment: The exploitation of channel network structure*. J. Hydrol. Sci. 37, 53-71.

Norusis, M., 2000. SPSS for Windows. SPSS Inc. Chicago, Illinois.

Olivera, F., & Maidment, D., 1999. *Geographic information systems (GIS)-based spatially distributed model for runoff routing*. Water Res. Research 35(4), 1155-1164.

Rahman, A., 2005. *A quantile regression technique to estimate design floods for ungauged catchments in south-east Australia*. Australian J. Water Resources 9(1), 81-89.

Refsgaard, J. & Storm, B., 1995. *MIKE SHE. Computer Models of Watershed Hydrology*. In: Singh V. (ed), Water Resources Publications, Colorado, USA, pp. 809-846.

Riggs, H., 1973. *Regional analysis of streamflow characteristics*. U.S. Geol. Surv. Water Resour. Invest. Tech., Book 4, 15pp.

Schumann, A., & Funke, R., 1996. *GIS-based components for rainfall runoff models*. In: Kovar, K., Nachtnebel, H. (eds), HydroGIS ’96, Vienna, IAHS Publ. No. 235, pp. 477-484.

Troch, P., Smith, J., Wood, E., & de Troch, F., 1994. Hydrologic controls of large floods in a small basin, J. Hydrology 156, 285-309.

Van Lammeren, R., Weitkamp, G., Verhagen, S., & Ruijten, J., 2009. *Manual Geo-visualization assignments GRS-60312 ArcGIS 9.3 (ArcMap & ArcScene)*. Wageningen UR-CGI.

Willmott, C., Rowe, C. & Mintz, Y., 1985. *Climatology of the terrestrial seasonal water cycle*. J. Climatology 5, 589-606.

Received at: 14. 01. 2013
Revised at: 27. 10. 2013
Accepted for publication: 14. 11. 2013
Published online at: 18. 11. 2013