Models – Repositories of Knowledge
(Proceedings ModelCARE2011 held at Leipzig, Germany, in September 2011) (IAHS Publ. 355, 2012).

Simultaneous estimation of groundwater recharge rates, associated zone structures, and hydraulic conductivity values using fuzzy c-means clustering and harmony search optimisation algorithm: a case study of the Tahtali watershed

M. TAMER AYVAZ[1] & ALPER ELÇI[2]

1 Department of Civil Engineering, Pamukkale University, 20070 Denizli, Turkey
2 Department of Environmental Engineering, Dokuz Eylül University, 35160 Buca-Izmir, Turkey
alper.elci@deu.edu.tr

Abstract The aim of this study is to present a linked simulation–optimisation model to estimate the groundwater recharge rates, their associated zone structures, and hydraulic conductivity values for regional, steady-state groundwater flow models. For the zone structure estimation problem the fuzzy c-means clustering (FCM) method was used. The association of zone structures with the spatial distribution of groundwater recharge rates was then accomplished using an optimisation approach where the heuristic harmony search (HS) algorithm was used. Since the solution was obtained by a heuristic algorithm, the optimisation process was able to use a non-specific initial solution, i.e. an initial solution that does not have to be close to the final solution. The HS-based optimisation model determines the shape of zone structures, their corresponding recharge rates and hydraulic conductivity values by minimizing the root mean square error ($\Re$) between simulated and observed head values at observation wells and springs, respectively. To determine the best recharge zone structure, the identification procedure starts with computation of one zone and systematically increased the zone number until the optimum zone structure is identified. Subsequently, the performance of the proposed simulation–optimisation model was evaluated on the Tahtali watershed (Izmir, Turkey), an urban watershed for which a seasonal steady-state groundwater flow model was developed for a previous study. The results of our study demonstrated that the proposed simulation–optimisation model is an effective way to calibrate the groundwater flow models for the cases where tangible information about the groundwater recharge distribution does not exist.

Key words groundwater recharge; zone structure estimation; optimisation; harmony search

INTRODUCTION

The spatial distributions of groundwater recharge rates and hydraulic conductivities are key properties of groundwater flow models. For occasions when field data or measurements for these parameters are absent and cannot be obtained during the timeframe given for the modelling job, numerical estimation methods can be implemented. It is the objective of this study to propose a procedure to estimate groundwater recharge rates with the associated zone structure and hydraulic conductivities for steady-state groundwater flow models. The proposed procedure involves the adaptation of individual algorithms that were applied in the past in hydrology/hydrogeology. However, they were implemented separately and not as a combination of algorithms, as done in this study to identify hydraulic conductivities and recharge values. Here, the harmony search algorithm is used in combination with the fuzzy c-means clustering algorithm to determine zone structures and values for hydraulic conductivity and recharge. The applicability of the entire procedure was demonstrated on the semi-urban Tahtali watershed in Izmir-Turkey, which is a key component of Izmir's water supply system. The Tahtali dam reservoir (38°08′N; 27°06′E) is located 40 km south of Izmir and meets about 36% of the city's total water demand. The watershed of the reservoir has an area of 550 km^2 and is a sub-watershed of the larger K. Menderes River watershed (Fig. 1). Elçi *et al.* (2010) previously presented results of a seasonal, steady-state groundwater flow model, for which model parameters were obtained with the parameter estimation code, PEST (Doherty, 2004). Therefore, another objective of this study is to compare parameters obtained with the proposed procedure to previously obtained ones.

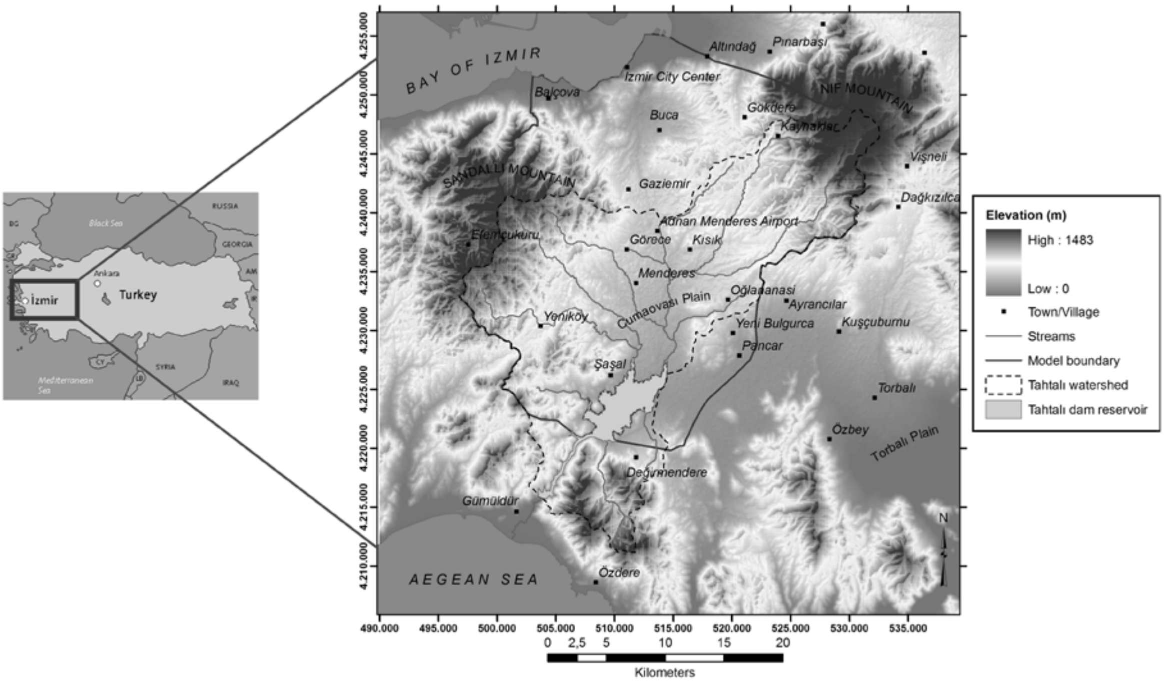

Fig. 1 General location map of application area showing the groundwater flow model boundaries and the Tahtali watershed.

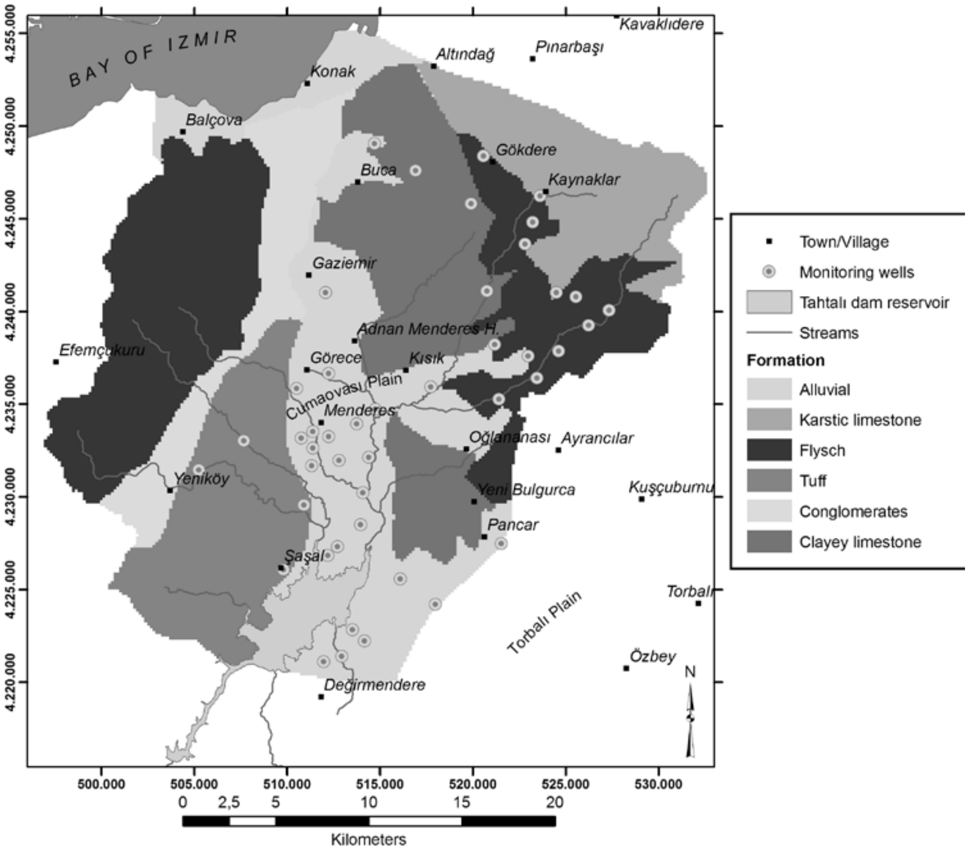

Fig. 2 Geological map of the study area and locations of the groundwater level monitoring wells. Formations also represent the hydraulic conductivity structure used in the groundwater model. (K_1: alluvial, K_2: karstic limestone, K_3: flysch, K_4: tuff, K_5: conglomerates, K_6: clayey limestone).

MODEL DEVELOPMENT

Optimisation Model: Harmony Search Algorithm (HS)

HS, first proposed by Geem *et al.* (2001), is a heuristic optimisation algorithm which gets its basis from musical processes. It is well known that the purpose of the musical processes is to seek a musically pleasing harmony through making several improvisations (Yang, 2009). Although HS is a newly proposed optimisation algorithm, it has been applied to many different problems including water-related applications, structural design, information technology, transport related problems, thermal and energy problems, and many other applications. The HS algorithm has similarities to evolution strategies and other heuristic algorithms. Unlike most other algorithms HS does not mimic processes encountered in nature. The state-of-the-art in the structure of HS algorithm, an overview of its applications and developments, and comparisons to other algorithms can be found in Ingram & Zhang (2009) and Geem (2010). The mathematical statement of HS is as follows:

Let HMS be the harmony memory size, N be the number of decision variables, $\mathbf{x}^i = \left\{ x_j^i \right\} \Big|_{i=1}^{\mathrm{HMS}} \Big|_{j=1}^{N}$ be the solution vectors, and $\mathbf{x}' = \left\{ x'_j \right\} \Big|_{j=1}^{N}$ be the newly generated solution vector. Using these parameters, an optimisation problem can be solved based on the following scheme:

1. Initialization of HM: Generate initial solution vectors as many as HMS, $\mathbf{x}^1 \cdots \mathbf{x}^{\mathrm{HMS}}$.
2. Generate a new solution vector $\mathbf{x}'$ for each x'_j:
   – with probability HMCR select x'_j from memory, $x'_j = x_j^{Rnd[1,HMS]}$
3. Pitch adjustment: For each x'_j:
   – with probability PAR change x'_j as, $x'_j = x'_j \pm bw \times Rnd(0,1)$.
   – with probability (1 − PAR) do nothing.
   – with probability (1 − HMCR) select a new random value from the possible range.
4. If $\mathbf{x}'$ is better than the worst $\mathbf{x}^i$ in harmony memory, replace $\mathbf{x}^i$ with $\mathbf{x}'$.
5. Repeat steps 2–5 until the given termination criterion is satisfied.

As can be seen from the computational scheme given above, HS requires some solution parameters which are Harmony Memory Size (HMS), Harmony Memory Considering Rate (HMCR), Pitch Adjusting Rate (PAR), and distance bandwidth, $bw = \left(x_j^{\max} - x_j^{\min} \right) / \delta$, where $x_j^{\min}$ and $x_j^{\max}$ are lower and upper bounds of the *j*th decision variable, and δ is a predefined segment which is used as 100 in this study. Note that the HM is a matrix where decision variables and corresponding objective function values are stored. The HMCR is the probability of selecting any harmony from HM. If HMCR is selected too low, only few elite harmonies are selected and the algorithm can converge too slowly. On the other hand, if HMCR is selected too high, the pitches in HM are mostly used and other possibilities are not explored well (Yang, 2009).

If the generated decision variable is selected from the HM, an evaluation for the requirement of pitch adjustment is necessary. This evaluation is performed using the PAR parameter which is the probability of making pitch adjusting. Pitch adjusting is a process that is analogous to taking the slightly neighbour value based on the predefined bandwidth (bw). The pitch adjusting process is similar to the mutation operator in a genetic algorithm, which maintains the diversity of population (Geem *et al.*, 2001). Based on the experience of the authors, HMS = 10, HMCR = 0.95, and PAR = 0.50 are appropriate values to solve many optimisation problems dealing with groundwater modelling (Ayvaz, 2009, 2010).

Estimation of the groundwater recharge zone structure

The recharge zone structure of the model domain is determined using the fuzzy c-means (FCM) clustering algorithm (Bezdek, 1981). In FCM algorithm, fuzzy membership values are assigned to each data point, which is related to the relative distance of that point to the cluster centres. FCM provides a procedure to group the data points that populate some multidimensional space into a specific number of different clusters (Ayvaz, 2007). Although the FCM algorithm is extensively used in many pattern classification and image processing studies, to our knowledge there is no published application example for the groundwater recharge zone structure estimation problem. The mathematical statement of FCM, which is modified for the groundwater recharge zone structure estimation problem, can be summarised as follows:

Let n_x and n_y be the number of finite difference grid points of the MODFLOW model in *x* and *y* directions, respectively, $\mathbf{X} = \left\{ X_i \right\} \Big|_{i=1}^{n_x}$ and $\mathbf{Y} = \left\{ Y_j \right\} \Big|_{j=1}^{n_y}$ be the vectors that contain the locations of grid points in the *x* and *y* directions, respectively, and *c* be the number of clusters in which recharge rates are assumed to be homogeneous (hereafter the term "zone" is used instead of "cluster"). Zonation of the groundwater recharge distribution is performed by using the 3D fuzzy partition matrix $\mathbf{u} = \left[u_{ijk} \right] \Big|_{i=1}^{n_x} \Big|_{j=1}^{n_y} \Big|_{k=1}^{c}$ such that:

$$0 \le u_{ijk} \le 1 \; ; \qquad \sum_{k=1}^{c} u_{ijk} = 1 \; ; \qquad \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} u_{ijk} > 0 \tag{1}$$

where u_{ijk} represents the fuzzy membership value between the (*i,j*)th grid point and *k*th zone structure. Let $\hat{\mathbf{X}} = \left\{ \hat{X}_k \right\} \Big|_{k=1}^{c}$ and $\hat{\mathbf{Y}} = \left\{ \hat{Y}_k \right\} \Big|_{k=1}^{c}$ be the zone centres to be determined by the optimisation model in *x* and *y* directions, respectively. The elements of the fuzzy partition matrix are updated using the determined zone centres as follows:

$$u_{ijk} = \left(\sum_{t=1}^{c} \left(\frac{\left\| X_i - \hat{X}_k \right\|^2 + \left\| Y_j - \hat{Y}_k \right\|^2}{\left\| X_i - \hat{X}_t \right\|^2 + \left\| Y_j - \hat{Y}_t \right\|^2} \right)^{\frac{1}{\hat{m}-1}} \right)^{-1} \tag{2}$$

where $\left\| \cdot \right\|$ is the Euclidean norm, and $\hat{m}$ is the degree of fuzzification ($\hat{m} = 2$). It should be noted that if the calculated membership value of a grid point using equation (2) has a maximum value, then this grid point is assigned to this zone (Wang & Xue, 2002). By applying this procedure to all the finite difference grid points, the flow domain can be partitioned into *c* zones. After this partitioning process, homogeneous groundwater recharge rates are assigned to each zone by the optimisation model, and the aquifer's response is determined by performing a MODFLOW run.

Problem formulation and search procedure

The purpose of applying the proposed simulation–optimisation procedure to the Tahtali watershed model is to simultaneously estimate the groundwater recharge zone structure, associated recharge rates, and uniform hydraulic conductivity values within the six geological formations shown in Fig. 2. This problem can be formulated as an optimisation problem in which HS randomly generates the zone centres; FCM builds up the zone structures; and finally, randomly generated recharge rates and hydraulic conductivity values are assigned to the corresponding zone structures. Based on the errors for calculated hydraulic head values, zone centres, associated recharge rates, and hydraulic conductivity values are modified by the HS-based optimisation model. The objective of the optimisation model is to minimise the root mean squared error ($\Re$) between the simulated and observed hydraulic head values at the monitoring wells shown in Fig. 2. This problem can be mathematically stated as follows:

$$\min \; \Re^* \left(\Omega_c \right) = \Re \left(\Omega_c \right) + \sum_{k=1}^{c} P\left(\hat{X}_k , \hat{Y}_k \right) \tag{3}$$

$$\Re\left(\Omega_c\right)=\sqrt{\frac{1}{n_w}\sum_{i=1}^{n_w}\left(h_i\left(\Omega_c\right)-\tilde{h}_i\right)^2} \tag{4}$$

$$\frac{\partial}{\partial x}\left(\bar{K}h\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(\bar{K}h\frac{\partial h}{\partial y}\right)=W-\bar{R} \tag{5}$$

$$P\left(\hat{X}_k,\hat{Y}_k\right)=\begin{cases} 0 & \text{if } (\hat{X}_k,\hat{Y}_k) \text{ is in active cell} \\ \lambda\kappa & \text{if } (\hat{X}_k,\hat{Y}_k) \text{ is in inactive cell} \end{cases} \tag{6}$$

$$\Omega_1 \rightarrow \Omega_2 \rightarrow \cdots \rightarrow \Omega_c \tag{7}$$

where Ω_c is the solution of the problem with c recharge zones, $\Re(\Omega_c)$ is the root mean square error for the solution of Ω_c, $\Re^*(\Omega_c)$ the penalised objective function for the solution of Ω_c, $h_i(\Omega_c)$ is the simulated hydraulic head value at observation well i for the solution of Ω_c, $\tilde{h}_i$ is the observed hydraulic head value at observation well i, n_w is the number of observation wells ($n_w = 51$), h is the hydraulic head over the flow domain, W is the sinks/source term due to pumping, $\bar{R}$ is the set of groundwater recharge rates to be estimated such that $\bar{R}\in\left\{R_1,R_2,\cdots,R_c\right\}$, $\bar{K}$ is the set of hydraulic conductivities to be estimated such that $\bar{K}\in\left\{K_1,K_2,\cdots,K_6\right\}$, $P\left(\hat{X}_k,\hat{Y}_k\right)$ is the penalty function depending on the locations of zone centres, λ is the penalty parameter, and κ is the nearest distance to the model boundary in terms of the row and column numbers of the finite difference grid (Fig. 3). As can be seen from equations (3)–(7), the groundwater flow process enters the problem in equation (5) for unknown $\bar{R}$ and $\bar{K}$ distributions. These distributions are determined by the optimisation model and passed on to MODFLOW to obtain the solution for groundwater flow in the study area.

It should be noted that the reason for using the penalty function given in equation (6) is the irregular shape of the modelling domain. All the grid cells outside the model boundary are specified as inactive cells, which are shown as the dark shaded area in Fig. 3. Although these cells appear in the finite difference grid structure of the MODFLOW model, they are excluded from the numerical solution. Therefore, these inactive cells must be also excluded from the search space

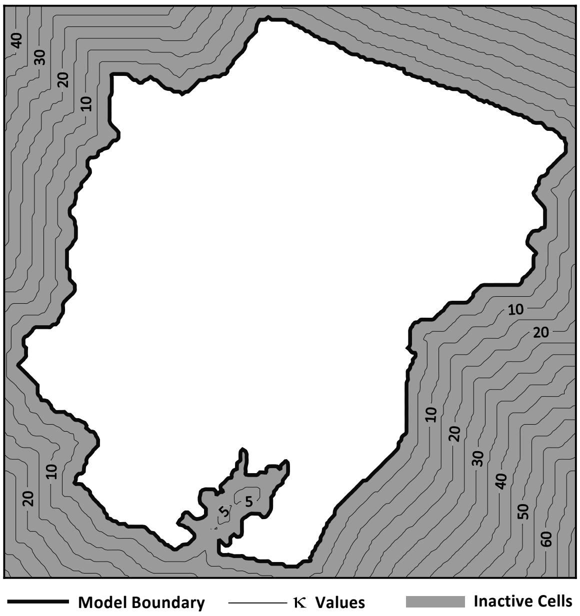

Fig. 3 The model domain and κ values used for the penalty function given in equation (6).

of the zone centres. Equation (6) states that if a zone centre is located in an inactive cell, the calculated objective function value is penalized with $\lambda\kappa$. The values of κ increase as the zone centres move away from the model boundaries. The value of λ is mostly arbitrary and problem dependent. Our trials show that $\lambda = 100$ can be used for the implementation of the penalty function given in equation (6). The decision variables of the optimisation model are the locations of the zone centres, associated recharge values for each zone, and uniform hydraulic conductivity values for each of the six pre-defined geological zones.

Although the proposed simulation–optimisation model may solve the problem based on the solution scheme given in equations (3)–(7), this mathematical formulation is only valid for cases where the number of zones (c) is known *a priori*. However, recharge zone structures, their numbers, and the associated recharge rates are unknown for most cases. Therefore, it is necessary to determine the number of zones such that the eventually identified zone structure optimally represents the field data. With this purpose, the zone structure estimation problem starts with one zone, and then, systematically increases the zone number until the best solution is obtained.

Furthermore, each successive solution for different zone numbers requires three additional decision variables (one is for recharge rate and two for zone centre coordinates). However, when the number of decision variables increases, there is a greater chance of producing local optimum solutions due to the increased dimension of solution space (Huang & Mayer, 1997). For such cases, the final value of the objective function may increase, although $\Re \rightarrow 0$, while $c \rightarrow \infty$ (Ayvaz, 2007). Therefore, final identified parameter values, zone structures, and objective function values are evaluated altogether to decide which successive zone structure best represents field conditions.

Our trial runs indicate that the value of the objective function does not improve significantly after about 15 000 iterations of HS. Therefore, the maximum number of iterations is set to 20 000. Completing 20 000 iterations of HS takes about 12 h on a workstation with an Intel Xeon 3.07 GHz processor and 6 GB RAM.

IDENTIFICATION RESULTS

Figure 4 shows the identified groundwater recharge zone structures for the solutions of Ω_2 to Ω_6 and the recharge zone structure originally used by Elçi *et al.* (2010). As can be seen from Fig. 4, centres of the identified zones remain inside the flow domain by virtue of the penalty function implementation. This result also implies that the final objective function values do not include any penalty term (i.e. $\Re^* = \Re$).

A summary of the identified hydraulic conductivity values, recharge rates, and final $\Re^*$ values for solutions Ω_1 to Ω_6, and the results by Elçi *et al.* (2010) are given in Table 1. The simulation–optimisation model calculates hydraulic conductivity values that are comparable between all solutions. On the other hand, the identified recharge rates are all different because the zone structures for recharge evolve during the optimisation, while the zone structure for hydraulic conductivity is fixed. Regarding final $\Re^*$ values after 20 000 iterations (where one iteration corresponds to a single MODFLOW run), it can be observed that the largest $\Re^*$ value (16.18 m) is obtained for Ω_1 where it is assumed that the flow domain takes a uniform recharge with a rate of 2.65×10^{-4} m/d. For this solution, the number of decision variables is seven, six for conductivities and one for uniform recharge rate. After this solution, the value of $\Re^*$ decreases as the solution approaches Ω_4, and increases again for Ω_5 and Ω_6. As mentioned earlier, theoretically the increase in the zone numbers should result in the decrease in the corresponding $\Re^*$ values. Therefore, by considering the identified parameter values, zone structures, and the final $\Re^*$ values, the four-zone structure (Ω_4) is selected as the best zone structure for the estimation problem discussed here (Fig. 4(c)).

In the previous modelling study by Elçi *et al.* (2010), the calibration of the same groundwater flow model was performed by adjusting the recharge rates and hydraulic conductivity values using the PEST parameter estimation code, while keeping the recharge rate for the zone representing the

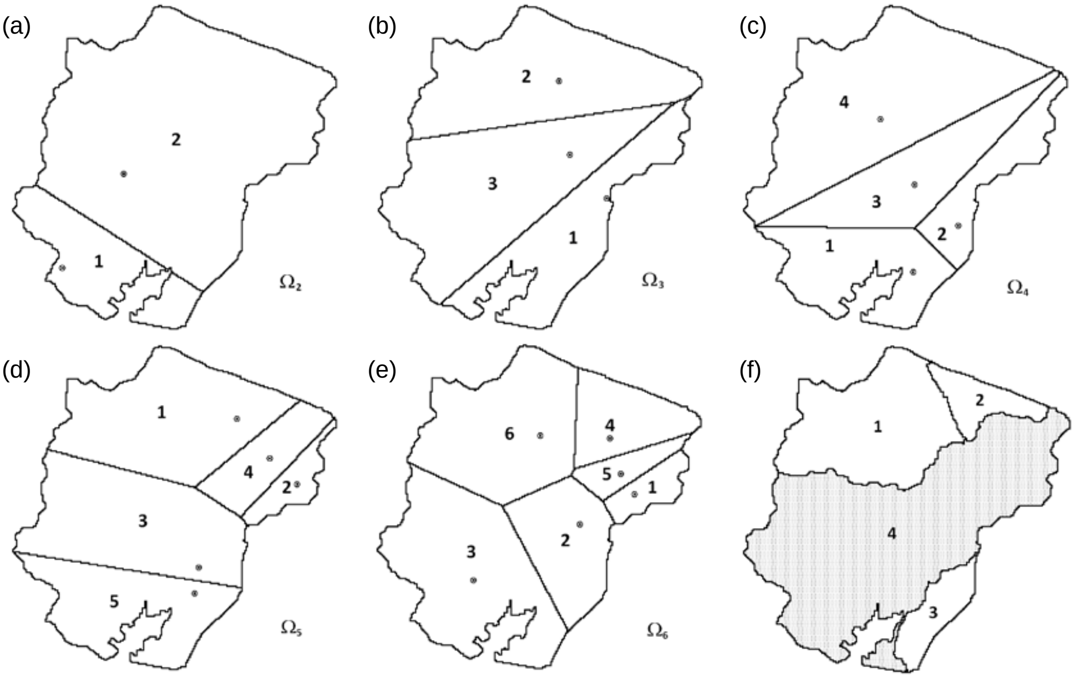

Fig. 4 Comparison of the identified zone structures, (a)–(e): for Ω_2 to Ω_6 (small circles correspond to zone centres); (f): the zone structure used by Elçi *et al.* (2010) (shaded area represents the Tahtali watershed).

Table 1 Summary of the identified hydraulic conductivity values, recharge rates and final $\Re^*$ values.

Identified Parameters		Parameter ranges	Solutions						Elçi *et al.* (2010)
			Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	
Hydraulic Conductivities (m/d)	K_1	1~100	20.81	22.12	20.16	25.87	26.35	24.38	7.06[b]
	K_2	0.01~10	0.01	0.01	0.01	0.01	0.01	0.28	0.01[b]
	K_3	0.001~1	1.00	1.00	0.54	0.91	0.95	0.65	0.30[b]
	K_4	0.1~10	5.74	6.07	7.49	7.57	6.99	6.46	7.09[b]
	K_5	0.05~5	5.00	5.00	5.00	4.98	4.38	4.47	1.91[b]
	K_6	0.05~5	3.58	3.62	2.61	2.56	1.64	3.06	1.35[b]
Groundwater Recharge Rates (m/d)	R_1	1.00E–10~1.00E–02	2.65E–04	1.05E–03	1.00E–10	9.53E–04	4.02E–05	3.23E–06	6.27E–05[b]
	R_2	1.00E–10~1.00E–02	–	2.66E–04	1.38E–04	2.13E–05	1.00E–10	6.56E–04	1.27E–04[b]
	R_3	1.00E–10~1.00E–02	–	–	6.21E–04	9.05E–04	5.78E–04	6.97E–04	5.00E–04[b]
	R_4	1.00E–10~1.00E–02	–	–	–	1.33E–04	6.31E–04	1.13E–04	9.02E–05[c]
	R_5	1.00E–10~1.00E–02	–	–	–	–	8.70E–04	1.03E–03	–
	R_6	1.00E–10~1.00E–02	–	–	–	–	–	7.55E–05	–
Final $\Re^*$ value		(m)	16.18	15.97	12.96	11.90	12.55	13.30	16.40[a]

[a] This value equals to $\Re$ and does not include the penalty function in equation (7).

[b] These values were calculated using the PEST model for fixed recharge and conductivity zone structures

[c] This value was calculated based on a external transient precipitation–runoff model

Tahtali watershed constant at a value that was obtained by an independent precipitation–runoff model. For that study the hydraulic conductivity zone structure was based on the geology of the study area and the four-zone recharge zone structure was manually created based on land use/land cover and lithology information. Comparison of results obtained by Elçi *et al.* (2010) with the results for Ω_4 shows that optimised hydraulic conductivity values in this study are in the same order of magnitude, except for zone 2 (K_2). However, this is not the case for recharge rates, as they are different for both studies. This difference can be explained by the different outcome of zone structures in both studies. It had to be assumed by Elçi *et al.*, that the recharge rate for the entire watershed (zone 4 in Fig. 4(f)) is uniform since the precipitation–runoff model was a lumped model. In the current model, however, this part of the model domain was split into other zones, each allowed to have different recharge rates. Elçi *et al.* (2010) obtained a final $\Re$ value of 16.40 for the four-zone structure given in Fig. 4(f), which indicates a less optimised solution compared to the $\Re^*$ value (11.90) of the Ω_4 solution given in Fig. 4(c). As can be seen from these results, the final $\Re^*$ value decreases by 27% through the use of the simulation–optimisation procedure when compared to Elçi *et al.* (2010). Based on the error evaluation, it can be concluded that the groundwater flow model is improved with the proposed procedure.

CONCLUSIONS

A coupled simulation–optimisation model is developed for the simultaneous estimation of groundwater recharge zone structure, their associated recharge rates and hydraulic conductivity values. The following conclusions can be drawn with respect to the performance of the model.

Model performance is expected to be better for cases with a higher number of more homogeneously distributed observation points and/or for cases where measured hydraulic conductivity and recharge rates are available. The number of iterations required to obtain a solution may appear significant when compared to iteration numbers seen in other calibration algorithms such as PEST. However, it should be noted that PEST is run to determine only the parameter values for a pre-defined zone structure, whereas our proposed model determines both zone structures and associated parameter values simultaneously. Also, to complete a single PEST iteration it is necessary to run MODFLOW multiple times. Therefore, the number of iterations needed by our proposed model and by PEST is not comparable. Furthermore, identified results for the same number of zones suggested that in the case of repeated runs, similar hydraulic head distributions can be obtained even if the identified zone structure and the associated recharge rates are different.

The applicability of the developed model is evaluated in a case study for the Tahtali watershed (Izmir-Turkey) and the estimation results are compared to previous modelling results for the same model domain that were obtained with a different optimisation approach. Comparison of the results indicates that the proposed model is an effective way to calibrate steady-state groundwater flow models, where tangible information about the groundwater recharge distribution does not exist.

Acknowledgements This study is based on the work supported by The Turkish Academy of Sciences (TÜBA) – The Young Scientists Award Programme (GEBIP). The first author would like to thank TÜBA for their support of this study.

REFERENCES

Ayvaz, M. T. (2007) Simultaneous determination of aquifer parameters and zone structures with fuzzy c-means clustering and meta-heuristic harmony search algorithm. *Adv. Water Resour.* 30, 2326–2338.

Ayvaz, M. T. (2009) Application of harmony search algorithm to the solution of groundwater management models. *Adv. Water Resour.* 32(6), 916–924.

Ayvaz, M. T. (2010) A linked simulation-optimization model for solving the unknown groundwater pollution source identification problems. *J. Cont. Hydrol.* 117(1–4), 46–59.

Bezdek, J. C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms. Advanced Applications in Pattern Recognition.* Plenum Press, New York, USA.

Doherty, J. (2004) PEST – Model independent parameter estimation. User Manual. Watermark Numerical Computing.

Elçi, A. Karadaş, D. & Fıstıkoğlu, O. (2010) The combined use of MODFLOW and precipitation-runoff modeling to simulate groundwater flow in a diffuse-pollution prone watershed. *Water Sci. Tech.* 62(1), 180–188.

Geem, Z. W., Kim, J. H. & Loganathan, G. V. (2001) A new heuristic optimization algorithm: harmony search. *Simulation* 76(2), 60–68.

Geem, Z. W. (2010) State-of-the-art in the structure of harmony search algorithm. In: *Recent Advances in Harmony Search Algorithm* (ed. by Z. W. Geem), Springer.

Huang, C. & Mayer, A. S. (1997) Pump-and-treat optimization using well locations and pumping rates as decision variables. *Water Resour. Res.* 33(5), 1001–1012.

Ingram, G. & Zhang, T. (2009) Overview of applications and developments in the harmony search algorithm. In: *Music-Inspired Harmony Search Algorithm: Theory and Applications* (ed. by Z. W. Geem), Springer.

Wang, H. & Xue, D. (2002) An intelligent zone-based delivery scheduling approach. *Comp. Industry* 48(2), 109–125.

Yang, X. S. (2009) Harmony search as a metaheuristic algorithm. In: *Music-Inspired Harmony Search Algorithm: Theory and Applications* (ed. by Z. W. Geem), Springer.