

Article

Parameterization of Pilot Point Methodology for Supplementing Sparse Transmissivity Data

Aditya Kapoor *  and Deepak Kashyap

Department of Civil Engineering, Indian Institute of Technology (IIT) Ropar, Nangal Road, Rupnagar 140001, India; deepakkashyap@iitrpr.ac.in

* Correspondence: adityain2003@gmail.com or aditya.kapoor@iitrpr.ac.in

Abstract: Pilot point methodology (PPM) permits estimation of transmissivity at unsampled pilot points by solving the hydraulic head based inverse problem. Especially relevant to areas with sparse transmissivity data, the methodology supplements the limited field data. Presented herein is an approach for estimating parameters of PPM honoring the objectives of refinement of the transmissivity (T) interpolation and the model calibration. The parameters are the locations and number of pilot transmissivity points. The location parameter is estimated by defining a qualifying matrix Q comprising weighted sum of the hydraulic head-sensitivity and the kriging variance fields. Whereas the former component of Q promotes the model calibration, the latter one leads to improved T interpolation by locating pilot points in un-sampled tracts. Further, a three-stage methodology is proposed for an objective determination of the number of pilot points. It is based upon sequential upgradation of the Variogram as the pilot points are added to the data base, ensuring its convergence with the head-based optimal Variogram. The model has been illustrated by applying it to Satluj-Beas interbasin wherein the pumping test data is not only sparse, but also unevenly distributed.

Keywords: pilot point method; groundwater modeling; inverse problem; kriging; calibration with sparse data



Citation: Kapoor, A.; Kashyap, D. Parameterization of Pilot Point Methodology for Supplementing Sparse Transmissivity Data. *Water* **2021**, *13*, 2082. <https://doi.org/10.3390/w13152082>

Academic Editor: Maurizio Polemio

Received: 15 June 2021

Accepted: 27 July 2021

Published: 30 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Distributed models are nowadays routinely employed for regional planning of groundwater development. The capability of a flow model to reliably predict the planning state variables depends upon its numerical efficiency and accuracy of several inputs like aquifer parameters, boundary conditions, and so on [1]. An important input in this regard is the transmissivity (T) field which is known to influence the long-term model predictions significantly. These fields are usually generated by interpolating cell transmissivities from the sampled T values that are typically obtained through the pumping tests. The interpolation is mostly based upon kriging [2–4], finite element linked basis functions, spline function, and polynomial approximation [5–7]. However, such T fields may not be credible on account of the sparseness of sampled T values and hence resulting models may lack predictive capability. Conducting an adequate number of pumping tests may not always be possible due to their high cost and difficult logistics.

The problem of T data deficit can be overcome by adopting a multivariate interpolator like cokriging that simultaneously invokes sampled values of T along with those of a secondary variable that is well-correlated to T and is more abundantly available [3,8,9]. This cokriging application has been reported with varying secondary variables like hydraulic head [10,11], specific capacity [12,13], and electrical conductivity [14].

An alternative approach is to overcome the T data deficit by head based inverse problem (IP) solutions [15,16]. Such an approach leads to T fields that are optimally consistent with the historical heads and the flow equation. The simulator component of the IP is joined to the interpolator to permit identification of T values at the chosen nodal points

that may be the pumping test sites or un-sampled locations termed as pilot points [17–21]. The solution typically involves optimizing (minimizing) the sum of squares of the residues (*SSR*) in respect of the observed and model-simulated hydraulic head fields treating the unknown T values as the decision variables [17]. The minimization of *SSR* is conducted at a stipulated target time that is usually ahead of the initial condition time by several simulation time steps. This batch optimization is usually necessary because of the limited resolution of the data in time domain [22]. This solution though essentially implying optimal reproduction of the historic hydraulic head fields, is assumed to lead to optimal model predictive capability. The reported studies have generally used kriging as the T interpolator employing the sampled T values [17,23–26]. The other IP based approach for parameter estimation is to predict (model) the parameter values at the sampled locations and assimilate the sampled values into the predicted ones by ensembled Kalman filters (EnKF) [27–29]. These numerical filters employ the covariance matrices of the prediction and the sample, and facilitate correction (termed as Kalman gain) of the modeled values.

Solving the inverse problem at un-sampled pilot points provides additional T values that may promote the credibility of the interpolated T fields and improve the model predictive capability. The pilot point methodology (PPM) has two parameters viz. the number of pilot points and their spatial coordinates (locations). These parameters may be estimated addressing the two underlying motivations viz. containment of the interpolation errors and secondly improvement of the model calibration. Apparently, the number of pilot points needs to be optimized. Whereas too few pilot points may defeat the purpose, an excessive number may lead to over-parameterization [23]. Ramarao et al. [17] suggested that number of pilot points can be driven by either the desired efficacy (threshold *SSR*) of the calibration or a user-defined maximum number. The next question is how to decide their locations. Although placing the pilot points anywhere in the flow domain would improve the interpolation and enhance the model calibration, additional criteria are required for optimizing the advantage accrued. The current practice in this regard is to add the pilot points one at a time, choosing the location with an objective of maximizing the potential reduction of *SSR*. This is accomplished by placing a pilot point at the location of the highest sensitivity [30] of hydraulic head field to the local transmissivity. This approach may, however, lead to placement of pilot points in already well-sampled tracts. This aspect is amply illustrated in the field illustration presented subsequently.

Other reported subjective criteria are placement of pre-stipulated number of the pilot points in areas of high hydraulic gradients [31], uniformly distributed [25,26,32–35], in areas of lower monitoring density [20], or at arbitrary locations [23]. These heuristic approaches may compromise the potential gain from the pilot point data with respect to the envisaged objectives.

Ideally speaking, the PPM parameters (number of pilot points and their locations) should be governed by the requirement of reducing the interpolation error in nodal T values—which in turn should improve the calibration. The improved calibration is expected to lead to improved model predictive capability which is the desired outcome of the entire exercise. However, in practice, reduction of the interpolation error may not unconditionally lead to better calibration due to modeling assumptions and data errors and spatial distribution of sensitivity. The present practice in respect of the PPM parameters is oriented solely towards enhancing the calibration with almost no thought given to the interpolation aspect. This is implemented by ensuring optimal reproduction of observed hydraulic head fields corresponding to the *historic* forcings. The consequent optimization of the calibration is assumed to be conducive for the optimal model predictive capability. However, in case the T field is not interpolated accurately, the predictions corresponding to a *different* set of forcings may not be realistic in spite of optimal reproduction of the historic hydraulic head fields. This implies that unless the T field is credibly generated, optimal calibration may not translate into optimal model predictive capability. Therefore, for ensuring improved model prediction capability, both the requirements viz. improvement of interpolation and calibration need to be considered compositely while determining the PPM parameters.

The merger of pilot point transmissivity values with the sampled data has implications in terms of additional data points for interpolation and also modification of the interpolation tools like Variogram in the context of kriging. The latter one has generally been ignored in the reported studies [17,24].

The PPM can be used for overcoming the deficit of other model parameters and forcing functions such as hydraulic conductivity and recharge [20,25,36]. It has also been used to generate optimal fields of vadose zone parameters [37,38].

The present study postulates a methodology for an objective determination of the PPM parameters (number of pilot points and locations) addressing the twin expectations from the pilot points viz. improved T interpolation and the calibration. Contrary to the current practice, the methodology allows placement of pilot points even in low hydraulic head-sensitivity regions to ensure fulfillment of the former expectation. Its field application has been illustrated by applying it to Satluj-Beas interbasin (India).

2. Methodology

Notwithstanding the current practice of locating the pilot points from only the view point of model calibration, the proposed methodology simultaneously accounts for both the requirements viz. ensuring improved T interpolation and the calibration. The requirement of the improved model prediction capability is honored by partly following the current practice of employing the hydraulic head-sensitivity field (S). The interpolation improvement aspect has been accounted for by conducting the interpolation employing kriging which provides spatial distribution of kriging variance (E) [2,3,39] that indexes the interpolation error. The proposed model meets the two requirements compositely by invoking a new spatially distributed qualifying score (Q) that is obtained by taking a weighted mean of normalized E and S fields. Normalization becomes necessary because of widely varying orders of magnitude of E and S . Pilot point is placed at the location of highest Q and the corresponding pilot point transmissivity is estimated by solution of SSR based inverse problem. This location strategy may require placement of the pilot points at low S locations which may compromise the SSR reduction but may also reduce E significantly in the bargain. Using this procedure, pilot points are added sequentially one at a time, and the Variogram is updated (using the enhanced T data base) for subsequent usage. Further a three-stage procedure is proposed for an objective determination of the number of pilot points. Details of the proposed methodology are presented in the following sections.

2.1. Determination of Pilot Point Location

Spatial coordinates and the transmissivity value of a new pilot point are determined by employing the updated data base comprising the sampled T values and the transmissivity values of the pilot points already added (if any). The introduction of new pilot point is completed in the following steps.

2.1.1. Kriging

The updated data base is subjected to kriging analysis for generating the fields of transmissivity and the kriging variance. This is accomplished by interpolating T and computing the corresponding kriging variance at the nodes of a grid superposed over the study area. Kriging analysis is performed on a modified set (T_D) comprising de-trended log-transformed T values to honor the assumption of stationarity and normal distribution [2] respectively. Discrete values of the semivariance (γ) of (T_D) for varying Euclidean distance (h) are obtained by “pairing” the values with varying h . Subsequently parameters of the chosen theoretical model are estimated by least-squares criterion. The most-commonly used theoretical models are Gaussian, exponential and hole-effect models [2]. An important parameter imbibed in all models is the Sill (σ^2) depicting the asymptotic value of γ . The resulting model facilitates interpolation of T_D and the corresponding kriging variance values at the nodal points as follows:

$$T_{D0} = \sum_{i=1}^m \lambda_i T_{Di} \quad (1)$$

$$E_o(x, y) = \sigma^2 - \sum_{i=1}^m \lambda_i (T_{Di}) \quad (2)$$

where T_{D_o} , E_o = interpolated de-trended log-transformed transmissivity value and kriging variance at the node under reckoning, (T_{Di}) = sampled and the pilot point T values after log transformation and detrending, m = number of T_{Di} data, (λ_i) = weight associated with T_{Di} , σ^2 = sill of the T Variogram. The weights are derived from the criterion of the minimization of variance of estimation error and ensuring unbiasedness of the interpolation [2]. These two criteria jointly lead to a determinate set of normal equations [3] in terms of (γ) . The corresponding interpolated T value is subsequently computed by adding the trend (ξ) at the interpolation location and exponentiation as follows.

$$T = 10^{T_{D_o} + \xi} \quad (3)$$

2.1.2. Flow Modeling

The generated transmissivity field and other associated data of forcings, boundary conditions, and storage parameters are fed to a flow model of the study area to simulate the corresponding head field (h_c) at a chosen target time or at a steady state. Employing the sampled head field (h_s) at the target time or steady state, the sum of squares of the residues (SSR) is computed as follows.

$$SSR = \sum_{i=1}^{na} (h_{ci} - h_{si})^2 \quad (4)$$

where, na = number of active nodes, h_{si} = sampled head at i^{th} active node, and h_{ci} = simulated head at i^{th} active node. Further, the hydraulic head sensitivity matrix \mathbf{S} is determined as follows.

Step 1: The sampled T data is kriged to generate a T field that is fed to the flow model to simulate the head field and hence SSR as per Equation (4).

Step 2: Considering a general i^{th} active node in the grid, the kriged T field is perturbed by adding ΔT to the transmissivity of this node and kriging it together with the sampled T data set. This perturbed T field is fed to the flow model to simulate the modified head field and hence the modified SSR . The corresponding change in SSR ($\Delta SSR = \text{new } SSR - \text{original } SSR$) is employed to compute the head-sensitivity at i^{th} node (S_i) as follows.

$$S_i = \frac{\Delta SSR}{\Delta T} \quad (5)$$

This step is repeated for all active nodes. This provides the matrix of nodal head-sensitivities \mathbf{S} .

2.1.3. Placement Criteria for a Pilot Point

A qualification field \mathbf{Q} is postulated herein whose elements depict the relative fitness of various candidate nodes for hosting the pilot point. Formulation of the qualifying matrix would apparently depend upon the motivation behind introduction of the pilot point. If, in accordance with the current practice, the motivation is to optimize the model calibration by minimizing the potential (SSR), the qualifying matrix shall be \mathbf{S} , i.e., $\mathbf{Q} = \mathbf{S}$. This placement criterion has been herein termed Strategy 1. However, as has been discussed earlier, this may not improve the model prediction capability in spite of an improved calibration, if the corresponding interpolation errors in the T field are large. Therefore, the other criterion may be to place the pilot point with an objective of maximizing the reduction of interpolation errors. From this viewpoint the kriging variance matrix \mathbf{E} may be adopted as the qualifying matrix, i.e., $\mathbf{Q} = \mathbf{E}$ which has been termed Strategy 2., However, doing so may not lead to any significant improvement in model calibration. This may preempt any improvement in the prediction capability. The proposed methodology (termed as Strategy 3) in this regard is to account for the objectives of improving the calibration as well as containment

of the interpolation errors compositely. This is expected to lead to an improved prediction capability. This approach is implemented by stipulating the qualifying matrix as follows:

$$\mathbf{Q} = \alpha \mathbf{E} + (1 - \alpha)n\mathbf{S} \quad (6)$$

$$n = O(E)/O(S) \quad (7)$$

where $O(E)$ and $O(S)$ = orders of magnitude of E and S , respectively, n = a normalizing factor, and α ($0 < \alpha < 1$), $(1 - \alpha)$ = relative weights assigned to the interpolation refinement and model calibration aspects. Spatial point where the qualifying matrix \mathbf{Q} has the largest value is deemed to be the pilot point location. Terms S and E may have different orders of magnitudes and the factor n facilitates their normalization necessary for a fair representation of the twin requirements viz. improvement of the calibration and interpolation. Parameter α provides the modeler an opportunity to prioritize among the two requirements. The interpolation aspect gets a larger priority with a higher value of α (>0.5). This may be desirable when modeling inputs (historic forcing function, boundary conditions and so on) are known to be less reliable than the T values. On the other hand, a lower value of α (<0.5) prioritizes improvement of the calibration and may be adopted in the case of relatively more credible modeling data. Equal weights ($=0.5$) may be assigned if the two data sets are known to have errors of similar magnitudes or if no information about the relative errors is available.

It is relevant to point out that some of the pilot points placed as per qualification matrix \mathbf{Q} may occupy a low sensitivity zone, resulting in a lower reduction in SSR in comparison to that in Strategy 1.

2.1.4. Estimating Pilot Point Transmissivity

The transmissivity at the chosen pilot point location is estimated by solving a head based inverse problem treating the unknown pilot point transmissivity (T_p) as the decision variable. This involves minimization of (SSR) (Equation (4)) with respect to (T_p). In case the pilot point is located in the low S region, the optimization may not lead to substantial reduction of SSR .

2.2. Determination of the Number of Pilot Points

The current database is updated by adding to it the newly estimated pilot point transmissivity value. Then, the three stage procedure discussed earlier is followed to check the adequacy of the T network. In the first stage, an experimental Variogram is generated using the updated data base (T_U) and its parameters (θ_U) are estimated. In the second stage, the parameters (θ_*) of the optimal variogram are estimated by minimizing SSR (Equation (4)) with respect to (θ_*). The solution aims at estimating the optimal Variogram parameters that produce a transmissivity field capable of optimal reproduction of the sampled head field. It may be visualized that with the addition of pilot points, the sets (θ_U) and (θ_*) (obtained in stage 1 and 2 respectively) would tend to merge. A convergence between the two indicates that the interpolational and the head prediction requirements are simultaneously fulfilled. In the third stage, this convergence is investigated. A satisfactory convergence is deemed to imply that the current network is adequate and no additional pilot point is necessary.

3. Application

The proposed model has been illustrated by applying it to Satluj-Beas interbasin that encompasses an area of 8040 km² and is bounded by rivers Satluj and Beas, and Shivalik mountains (Figure 1a). Agriculture is the major occupation in the area with irrigational demands met almost exclusively through groundwater development. This has resulted in rapidly declining water table. The underlying aquifer is composed of alluvium deposits consisting dominant sand strata with intermittent clay layers. Since the clay layers are nearly horizontal and spatially intermittent, the formation may be deemed to be a single horizontally isotropic unconfined aquifer. As per reported data, its T varies from 540 to

1830 m²/day in the north-east to south-west direction [40,41]. A few well log sites and a lithological section along this direction are presented in Figure 1b. Lithological profile along this section is presented in Figure 1c.

The sampled data base (T_S) comprises only fifteen pumping test derived T values (Figure 1a). Further, most of the sampling locations cluster in the N-E tract that displays high S on account of proximity to the mountainous boundary (Figure 2a). An indiscriminate placement of pilot points invoking the prevalent S -based criterion would further increase the data density in this tract. Therefore, a study aimed at introducing pilot points in accordance with the proposed model was considered relevant.

3.1. Data Base

The necessary data of sampled T , historical head fields, recharge, abstraction, and boundary conditions were gathered from different state and central water resources agencies. Pumping test derived data was available at 15 locations (Figure 1a). Water table data consists of observations from 91 wells over a period of three years extending from June 2014 to October 2016. This time interval covers three monsoon and two dry seasons. For the same period, the river stage data from six locations (three on Satluj, two on Beas and one at the confluence) (Figure 1a) were collected. The study area comprises 29 administrative blocks (Figure 1a) and block wise groundwater recharge, abstraction, and specific yield data was derived from published reports [40].

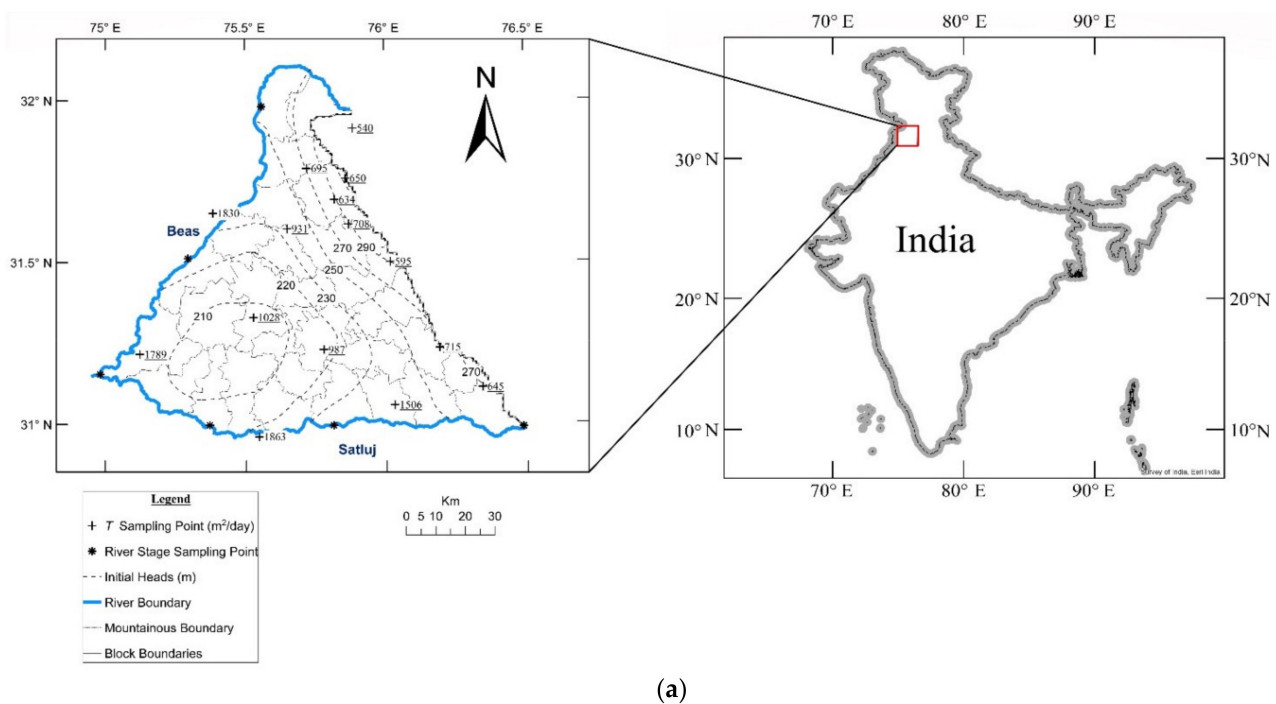


Figure 1. Cont.

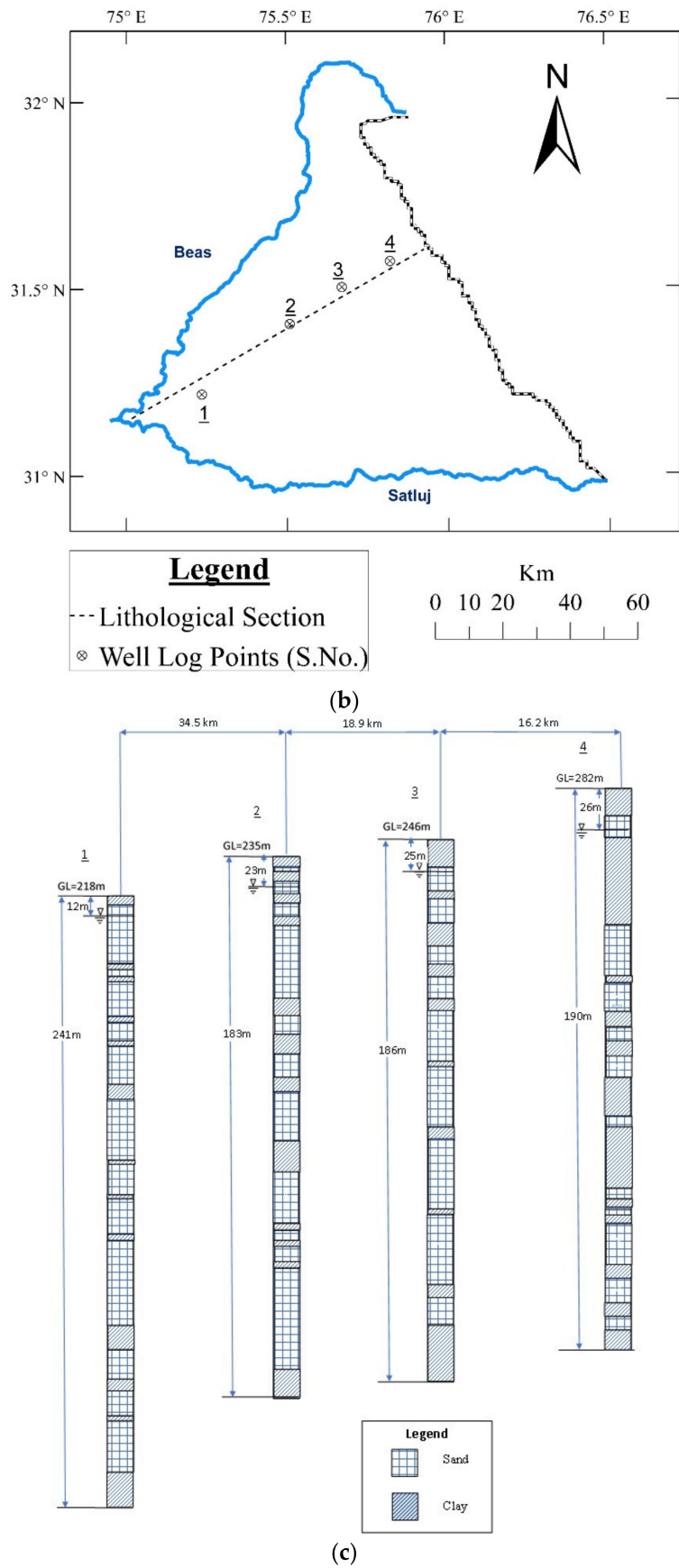


Figure 1. (a) Distribution of *T* sampling points in study area. (b) Lithological Section. (c) Lithological Profile along the section presented in (b).

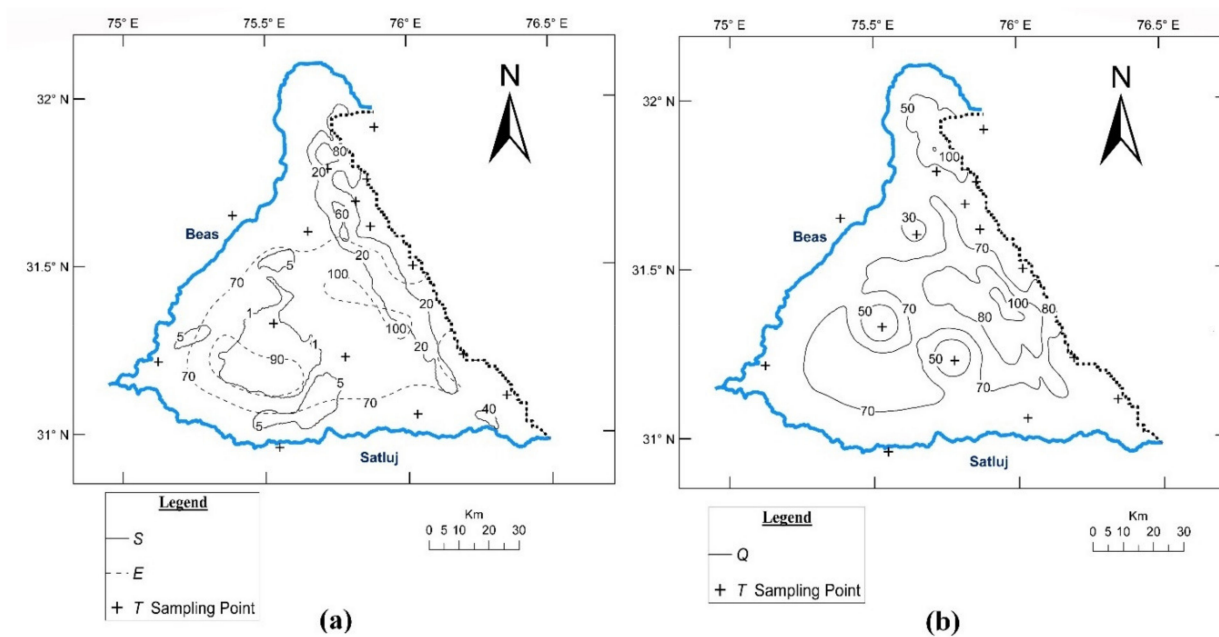


Figure 2. (a) Fields of S and E . (b) Field of Q .

3.2. Groundwater Flow Model

The groundwater flow model employed in the present illustration is based upon a numerical solution of the following differential equation governing two-dimensional horizontal flow in a confined aquifer.

$$\frac{\partial}{\partial x} \left(T_{xx} \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(T_{yy} \frac{\partial h}{\partial y} \right) - W = (S_c) \frac{\partial h}{\partial t} \quad (8)$$

where, x, y = orthogonal cartesian coordinates along principal permeability directions, t = time, h = head, T_{xx}, T_{yy} = transmissivities along x and y directions, S_c = storage coefficient, and W = sink term. Since there is no evidence of horizontal anisotropy in the study area, T_{xx} and T_{yy} were assumed to be identical. This confined flow equation has been invoked for modeling the unconfined flow assuming the temporal fluctuation of h to be small enough and the spatial gradients honoring Dupuit-Forchheimer assumptions. The adopted flow equation was solved to simulate the head fields at advancing discrete times using IADI based finite difference method. The flow model was externally linked to an optimizer based upon the SUMT algorithm [42] to facilitate the IP solutions.

Adopting a spatial step size of 1000 m, a finite difference grid comprising 8350 active nodes and 418 boundary nodes was superposed over the study area. The boundary conditions at the river boundary nodes were assigned as per the available river stage data. Neuman boundary condition was assigned along the mountainous boundary. The necessary flux rate was derived from the prevailing hydraulic gradients. The forcing function vector (\mathbf{W}) was derived from the available recharge and abstraction data. A SUMT based optimizer discussed in the model validation section was employed for this illustration exercise.

3.3. Determination of Pilot Point Parameters

Pilot points were added sequentially corresponding to three placement strategies viz. $\mathbf{Q} = \mathbf{S}$; $\mathbf{Q} = \mathbf{E}$; and $\mathbf{Q} = \alpha \mathbf{E} + (1 - \alpha)n\mathbf{S}$ serial numbered as 1, 2, and 3, respectively—strategy 3 being the proposed model. As a first step towards this end, the experimental and theoretical T variograms were developed. Gaussian model was selected as the variogram model with its parameters (σ^2 and L) were estimated as (7.6×10^{-3} and 22.3 km respectively) by

regressing the discrete point data of the experimental Variogram. These parameter values were employed to generate the E field.

The second step consisted of generating S and Q fields. It may be recalled that the hydraulic head fields were available at six discrete times over a period of three years extending from June 2014 to October 2016. Accordingly, treating hydraulic head field of June 2014 as the initial condition (Figure 1a), simulation was conducted until the target time October 2016. The performance measure (SSR) was computed in respect of the observed and simulated head fields at the target time. Elements of S were generated as discussed in Section 2.1.2 taking $\Delta T = 10 \text{ m}^2/\text{day}$. A value of 0.5 was assigned to the parameter α assuming the modeling input and the T data to be equally reliable. Further, noting that the orders of magnitude of E and S are 10^{-3} and 10^{-4} , respectively, a value of 10 was assigned to the factor n (refer Equation (7)). Q matrix was generated as per Equation (6) employing these values.

Third step consisted of placing the pilot points and estimating their T sequentially as per the three strategies until the desired convergence between the Variogram parameters is achieved.

3.4. Results

The computed fields of percentile S , E , and Q [$=1/2(E + 10S)$] are presented in Figure 2a,b. The convergence of two components (σ^2 and L) of θ_U and θ_* , for the three strategies, was achieved upon addition of 6, 5, and 5 pilot points, respectively. The convergence details with respect to strategy 3 (proposed model) are presented in Figure 3a,b. The generated pilot point locations corresponding to the three strategies and the corresponding T fields are presented in Figure 4a,b. It may be seen that while implementing strategy 2, all pilot points have fallen in the unsampled tracts—far away among themselves and from sampling locations. However, in strategy 1, the pilot points have fallen in already densely sampled tracts. In strategy 3 (proposed model), the pilot points have filled up the unsampled tracts, though not as effectively as strategy 2.

The metrics of improvement of the model calibration and of the interpolation for all three strategies are presented in Table 1. Strategy 0 therein depicts the baseline i.e., ‘no pilot point’ state. Metrics of the improvement in the model prediction capability includes the minimized SSR and its reduction in comparison to the baseline state. The improvement of the interpolation is indexed by A_{eff} representing the fractional area over which the kriging variance undergoes a reduction, and the mean E reduction (\bar{E}) therein.

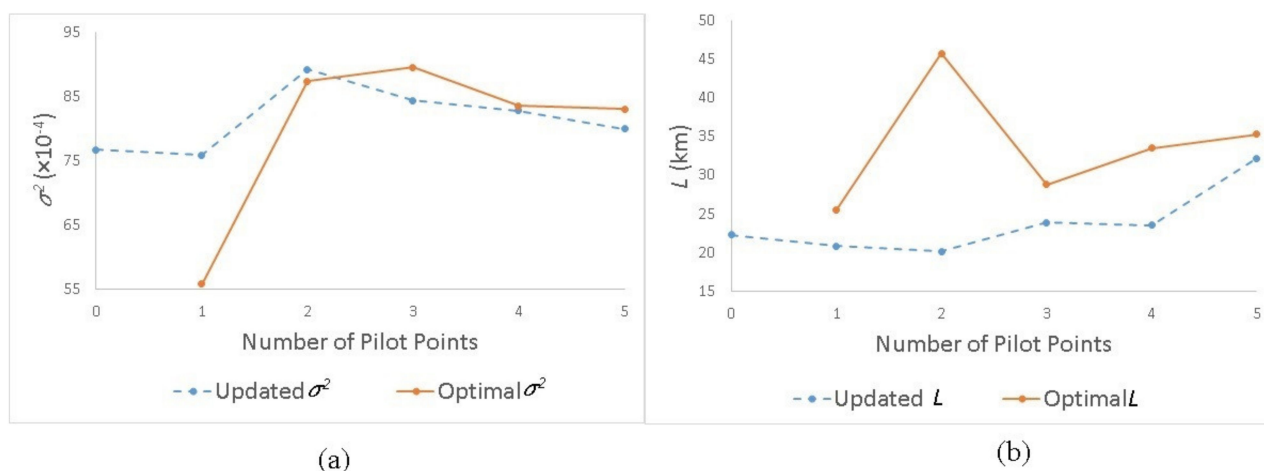


Figure 3. Convergence between the two components of θ_U and θ_* : (a) σ^2 and (b) L .

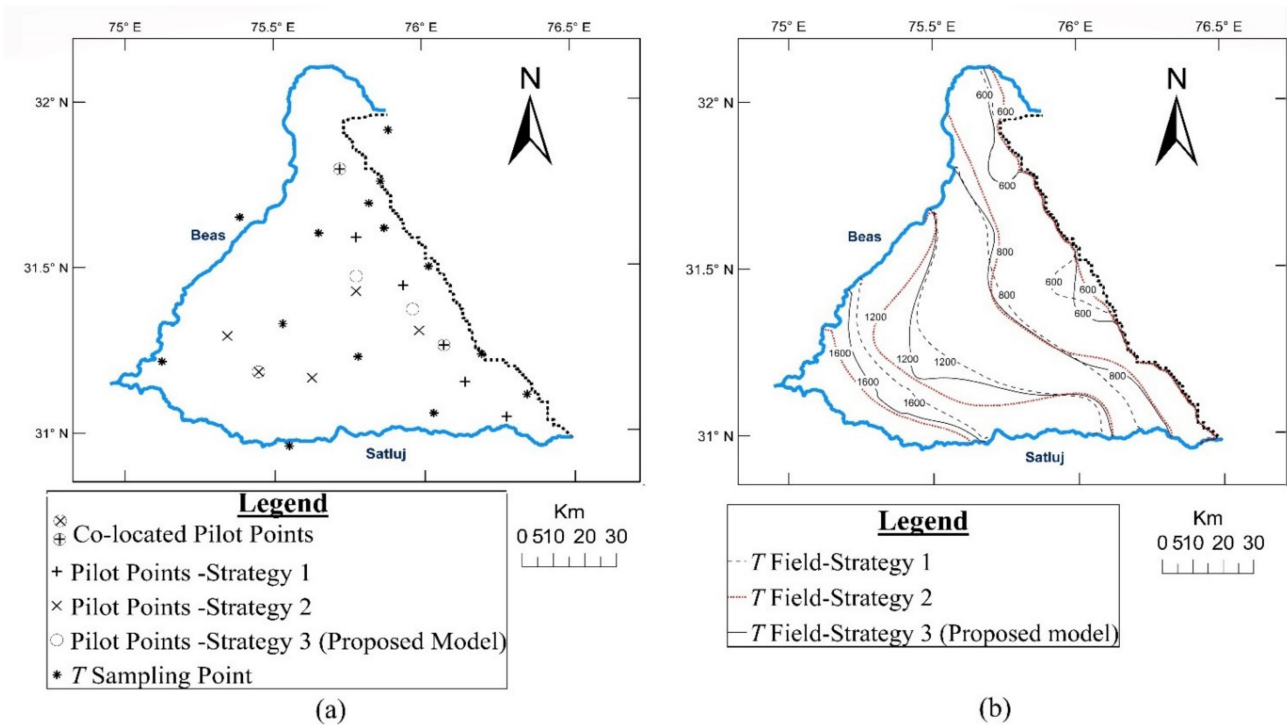


Figure 4. (a) Strategy wise pilot point placement. (b) T fields resulting from the three pilot point placement strategies.

Table 1. Strategy wise pilot point performance metrics.

Strategy	N	SSR (m ²)	ΔSSR (m ²)	$\bar{E} (\times 10^{-3})$	A_{eff} (Fraction)
0	0	1931	-	-	-
1	6	1678	253	0.65	0.28
2	5	1811	120	0.81	0.43
3	5	1713	218	0.74	0.38

4. Discussion

It is interesting to compare the results from strategies 1 and 2. It may be recalled that the strategy1 ($Q = S$) depicts the prevalent practice of the pilot point placement. As already discussed, this strategy focuses primarily on minimizing SSR with no consideration to containment of the interpolation error E . On the other hand, strategy 2 ($Q = E$) aims at minimizing the interpolation error (E) without any thought to minimization of SSR. As such, implementation of strategy 1 leads to far larger reduction of SSR (253 m²) with relatively moderate reduction of E ($\bar{E} = 0.65 \times 10^{-3}$) (Table 1) occurring over a small fraction (0.28) of the area. Reduction of E is in fact incidental since it is not accounted for in the placement criterion. It may also be seen that due the lack of focus on the interpolation aspect, the pilot points have not fallen in the sparse tracts of the sampled data network leading to clustering of data points (Figure 4a) in the north-eastern region, and defeating an important objective of providing pilot points in unsampled tracts. Thus, this strategy, though leading to much-reduced SSR, poses a concern about the interpolation aspects. This, in turn, may inhibit improvement of the model prediction capability notwithstanding the robust reduction of SSR.

The results from strategy 2 are just the reverse—leading to relatively large reduction of E ($\bar{E} = 0.81 \times 10^{-3}$) over larger fractional area (0.43) with only a minimal and incidental reduction of SSR (120 m²). Being focused on the interpolation aspect, it has also led to pilot point points generally filling up the unsampled tracts (Figure 4a). It may be seen that the pilot points have generally been located in southwest region which has been poorly represented in the sampled data base (T_S). Further, since the pilot points are located in the

low S region, the reduction in SSR is small. This poses a concern about the compromised model prediction capability.

The proposed strategy (strategy 3) focuses on both the aspects viz. improvement of calibration and containment of interpolation errors. Accordingly, it provides compositely the reduction of SSR and of E . Whereas the reduction of SSR is significant, being close enough to what is obtained in strategy 1, E reduction and its areal extent is in the proximity of strategy 2 metrics. Also, the large unsampled tracts get filled up—though not as well as with strategy 2.

The second PPM parameter viz. number of pilot points has been determined by assessing the convergence between θ_U and θ_* . Introduction of five pilot points has resulted in satisfactory convergence. Addition of anymore pilot points may result in over parameterization. As discussed earlier, a satisfactory convergence ensures simultaneous fulfilment of interpolational and the calibration requirements. It is evident from Figure 3a,b that the variogram sill parameter (σ^2) derived from the enhanced T data T_U ($N = 5$) converged to a higher value than its initial value derived from the sampled T data T_S ($N = 0$). Higher sill denotes a larger variance in T_U than in T_S which in-turn indicates that the optimal T field based upon T_U exhibits a larger heterogeneity as compared to the one obtained from T_S . The variogram range parameter (L) attains a larger converged value with T_U as compared to the initial value derived from T_S . Larger range indicates that the influence of a T point in kriging procedure extends to a much larger distance than what is indicated by T_S . The application of the proposed methodology has led to a continuous field that is hydrogeologically realistic in alluvial aquifers.

The proposed methodology overcomes the shortcomings present in the prevalent PPM parameterization practices. Further, it permits stage wise updating of the variogram employing the current set of generated pilot point values along with sampled T data. The methodology illustrated herein for supplementing sparse T data can also be used to overcome deficit of other aquifer and vadose zone parameters.

5. Conclusions

1. The prevalent hydraulic head-sensitivity based practice for implementation of the pilot point methodology is oriented towards the refinement of the flow model calibration with almost no thought given to the interpolation aspect. This may jeopardize the model prediction capability apart from failing to populate the un-sampled tracts by numerically generated pilot points.
2. The proposed parameterization of this methodology provides objective criteria for placement of the plot points as well determining their number.
3. The proposed methodology ensures, apart from a better reproduction of sampled hydraulic head fields, filling up of the vacant tracts by the pilot points and hence better interpolation.
4. The proposed criterion for the number of pilot points ensures that the interpolated transmissivity field is optimally conducive for the hydraulic head simulations.
5. The application of the proposed methodology has supplemented optimally the sparse and clustered T data in Satluj-Beas interbasin. The resulting set of T values would facilitate credible model predictions since they simultaneously permit optimal interpolation and calibration. The methodology can be similarly applied to other areas for overcoming deficit of data of T and other parameters.

Author Contributions: Conceptualization, A.K. and D.K.; Methodology, A.K. and D.K.; Software, A.K. and D.K.; Supervision, D.K.; Visualization, D.K.; Writing—original draft, A.K. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: The research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used during the study were provided by a third party (head data, river stage data, transmissivity data, blockwise recharge abstraction and specific yield data). Direct request for these materials may be made to the provider as indicated in the Acknowledgments. Codes generated or used during the study are proprietary in nature (computer code for kriging, optimization and numerical groundwater flow model).

Acknowledgments: The data used in the present study was gathered from Water Resources Organization, Punjab; Department of Agriculture, Punjab; and Central Groundwater Board, Chandigarh. Authors are thankful to Chief Engineer K.S. Takshi, Neeraj Pandit and Tejdeep Singh from these organizations for their help and meaningful discussions. Authors are also grateful to the anonymous reviewers and editorial board for their very thoughtful and constructive suggestions which have improved the quality of the paper significantly.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

IP	Inverse Problem
SSR	Sum of Squares of Residues
PPM	Pilot Point Methodology
EnKF	Ensembled Kalman Filters

References

- Carrera, J.; Neumann, S.P. Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 3. Application to Synthetic and Field Data. *Water Resour. Res.* **1986**, *22*, 228–242. [CrossRef]
- Kitanidis, P.K. *Introduction to Geostatistics: Applications in Hydrogeology*, 1st ed.; Cambridge University Press: Cambridge, UK, 1997.
- Wackernagel, H. *Multivariate Geostatistics*, 3rd ed.; Springer: Berlin, Germany, 2003.
- Yeh, W.W. Optimization methods for groundwater modeling and management. *Hydrogeol. J.* **2015**, *23*, 1051–1065. [CrossRef]
- Yao, L.; Huo, Z.; Feng, S.; Mao, X.; Kang, S.; Chen, J.; Xu, J.; Steenhuis, T.S. Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China. *Environ. Earth Sci.* **2014**, *71*, 1911–1924. [CrossRef]
- Chapra, S.C.; Canale, R.P. *Numerical Methods for Engineers*, 6th ed.; McGraw-Hill Higher Education: New York, NY, USA, 2010.
- Adhikary, S.K.; Muttill, N.; Yilmaz, A.G. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrol. Process.* **2017**, *31*, 2143–2161. [CrossRef]
- Isaaks, H.E.; Srivastava, R.M. *Applied Geostatistic*, 1st ed.; Oxford University Press: New York, NY, USA, 1989.
- Oliver, M.A.; Webster, R. *Basic Steps in Geostatistics: Variogram and Kriging*; Springer International Publishing: New York, NY, USA, 2015.
- Kitanidis, P.K.; Vomvoris, E.G. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resour. Res.* **1983**, *19*, 677–690. [CrossRef]
- Yeh, T.C.J.; Gutjahr, A.L.; Jin, M. An iterative cokriging-like technique for ground-water flow modeling. *Groundwater* **1995**, *33*, 33–41. [CrossRef]
- Aboufirassi, M.; Mariño, M.A. Cokriging of aquifer transmissivities from field measurements of transmissivity and specific capacity. *J. Int. Assoc. Math. Geol.* **1984**, *16*, 19–35. [CrossRef]
- Ahmed, S.; de Marsily, G. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resour. Res.* **1987**, *23*, 1717–1737. [CrossRef]
- Ahmed, S.; de Marsily, G.; Talbot, A. Combined use of hydraulic and electrical properties of an aquifer in a geostatistical estimation of transmissivity. *Groundwater* **1988**, *26*, 78–86. [CrossRef]
- Gannett, M.W.; Wagner, B.J.; Lite, K.E. Groundwater Simulation and Management Models for the Upper Klamath basin, Oregon and California. United States Geological Survey Scientific Investigations Report 2012–5062. 2012. Available online: <https://pubs.er.usgs.gov/publication/sir20125062> (accessed on 20 June 2020).
- Davis, K.W.; Putnam, L.D.; LaBelle, A.R. Conceptual and Numerical Models of Groundwater Flow in the Ogallala and Arikaree Aquifers, Pine Ridge Indian Reservation Area, South Dakota, Water Years 1980–2009 United States Geological Survey Scientific Investigations Report 2014–5241. 2015. Available online: <https://pubs.usgs.gov/sir/2014/5241/pdf/sir2014-5241.pdf> (accessed on 20 June 2020).
- RamaRao, B.S.; LaVenue, A.M.; de Marsily, G.; Marietta, M.G. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. *Water Resour. Res.* **1995**, *31*, 475–493. [CrossRef]
- Alcolea, A.; Carrera, J.; Medina, A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Adv. Water Resour.* **2006**, *29*, 1678–1689. [CrossRef]

19. Doherty, J.E.; Fienen, M.N.; Hunt, R.J. Approaches to Highly Parameterized Inversion: Pilot-Point Theory, Guidelines, and Research Directions. United States Geological Survey Scientific Investigations Report 5168. 2010. Available online: <https://pubs.usgs.gov/sir/2010/5168/pdf/sir20105168.pdf> (accessed on 25 June 2020).
20. Doherty, J.E.; Hunt, R.J. *Approaches to Highly Parameterized Inversion: A Guide to Using PEST for Groundwater-Model Calibration*; United States Geological Survey: Reston, VA, USA, 2010. Available online: https://pubs.usgs.gov/sir/2010/5169/pdf/GWPEST_sir2010-5169.pdf (accessed on 15 April 2021).
21. Anderson, M.P.; Woessner, W.W.; Hunt, R.J. *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, 2nd ed.; Elsevier: San Diego, CA, USA, 2015.
22. Sun, A.Y.; Morris, A.; Mohanty, S. Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data. *Adv. Water Resour.* **2009**, *32*, 280–292. [[CrossRef](#)]
23. Certes, C.; de Marsily, G. Application of the pilot point method to the identification of aquifer transmissivities. *Adv. Water Resour.* **1991**, *14*, 284–300. [[CrossRef](#)]
24. LaVenue, A.M.; Pickens, J.F. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. *Water Resour. Res.* **1992**, *28*, 1543–1569. [[CrossRef](#)]
25. Woodward, J.R.S.; Wohling, T.; Stengler, R. Uncertainty in the modelling of spatial and temporal patterns of shallow groundwater flow paths: The role of geological and hydrological site information. *J. Hydrol.* **2016**, *534*, 680–694. [[CrossRef](#)]
26. Jazaei, F.; Waldron, B.; Schoefernacker, S.; Larsen, D. Application of Numerical Tools to Investigate a Leaky Aquitard beneath Urban Well Fields. *Water* **2019**, *11*, 5. [[CrossRef](#)]
27. Chen, X.; Hammond, G.E.; Murray, C.J.; Rockhold, M.L.; Vermeul, V.R.; Zachara, J.M. Application of ensemble-based data assimilation techniques for aquifer characterization using tracer data at Hanford 300 area. *Water Resour. Res.* **2013**, *49*, 7064–7076. [[CrossRef](#)]
28. Tang, Y.; Shen, Z.; Gao, Y. *An Introduction to Ensemble-Based Data Assimilation Method in the Earth Sciences, Nonlinear Systems-Design, Analysis, Estimation and Control*; Lee, D., Burg, T., Volos, C., Eds.; IntechOpen: Rijeka, Croatia, 2016.
29. Li, N.; Kinzelbach, W.; Li, H.; Li, W.; Chen, F. Improving parameter and state estimation of a hydrological model with the ensemble square root filter. *Adv. Water Resour.* **2021**, *147*, 103813. [[CrossRef](#)]
30. Sun, N.Z. *Inverse Problems in Groundwater Modeling*, 6th ed.; Springer Science & Business Media: Dordrecht, The Netherlands, 1999.
31. Marsily, G.; Lavedan, G.; Boucher, M.; Fasanino, G. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In *Geostatistics for Natural Resources Characterization*; Marechal, A., Journel, A.G., David, M., Eds.; Springer: Dordrecht, The Netherlands, 1984; Volume 2, pp. 831–849.
32. Tonkin, M.J.; Doherty, J. A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resour. Res.* **2005**, *41*, 1–16. [[CrossRef](#)]
33. Jardani, A.; Dupont, J.P.; Revil, A.; Massei, N.; Fournier, M.; Laignel, B. Geostatistical inverse modeling of the transmissivity field of a heterogeneous alluvial aquifer under tidal influence. *J. Hydrol.* **2012**, *472*, 287–300. [[CrossRef](#)]
34. Carniato, L.; Schoups, G.; deGiesen, N.; Seuntjens, P.; Bastiaens, L.; Sapion, H. Highly parameterized inversion of groundwater reactive transport for a complex field site. *J. Contam. Hydrol.* **2015**, *173*, 38–58. [[CrossRef](#)]
35. Chow, R.; Dugge, J.B.J.; Wohling, T.; Novak, W. Evaluating Subsurface Parameterization to Simulate Hyporheic Exchange: The Steinlach River Test Site. *Groundwater* **2019**, *58*, 93–109. [[CrossRef](#)]
36. Engeler, I.; Franssen, H.J.H.; Muller, R.; Stauffer, F. The importance of coupled modelling of variably saturated groundwater flow-heat transport for assessing river-aquifer interactions. *J. Hydrol.* **2011**, *397*, 295–305. [[CrossRef](#)]
37. Kowansli, M.B.; Finsterle, S.; Rubin, Y. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. *Adv. Water Resour.* **2004**, *27*, 583–599.
38. Casique, E.M.; Neuman, S.P.; Vesselinov, V.V. Maximum likelihood Bayesian averaging of airflow models in unsaturated fractured tuff using Occam and variance windows. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 863–880. [[CrossRef](#)]
39. Delhomme, J.P. Kriging in the hydrosociences. *Adv. Water Resour.* **1978**, *1*, 251–266. [[CrossRef](#)]
40. Central Ground Water Board. Report of the Working Group on Methodology for Assessment of Development Potential of Deeper Aquifers. Faridabad. 2009. Available online: <http://cgwb.gov.in/documents/gec97.pdf> (accessed on 15 June 2020).
41. Singh, A.; Panda, S.N.; Kumar, K.S.; Sharma, C.S. Artificial groundwater recharge zones mapping using remote sensing and GIS: A case study in Indian Punjab. *Environ. Manag.* **2013**, *52*, 61–71. [[CrossRef](#)] [[PubMed](#)]
42. Fiacco, A.V.; McCormick, G.P. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques (Reissue)*, 1st ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1990.