

《原著》

## 히스토리매칭 기법을 이용한 비모수 지구통계 모사 예측성능 향상 예비연구

정진아 · 프라딕 포디알 · 박은규\*

경북대학교 지질학과

### A Preliminary Study of Enhanced Predictability of Non-Parametric Geostatistical Simulation through History Matching Technique

Jina Jeong · Pradeep Paudyal · Eungyu Park\*

Department of Geology, Kyungpook National University, Daegu, Korea

#### ABSTRACT

In the present study, an enhanced subsurface prediction algorithm based on a non-parametric geostatistical model and a history matching technique through Gibbs sampler is developed and the iterative prediction improvement procedure is proposed. The developed model is applied to a simple two-dimensional synthetic case where domain is composed of three different hydrogeologic media with  $500\text{ m} \times 40\text{ m}$  scale. In the application, it is assumed that there are 4 independent pumping tests performed at different vertical interval and the history curves are acquired through numerical modeling. With two hypothetical borehole information and pumping test data, the proposed prediction model is applied iteratively and continuous improvements of the predictions with reduced uncertainties of the media distribution are observed. From the results and the qualitative/quantitative analysis, it is concluded that the proposed model is good for the subsurface prediction improvements where the history data is available as a supportive information. Once the proposed model be a matured technique, it is believed that the model can be applied to many groundwater, geothermal, gas and oil problems with conventional fluid flow simulators. However, the overall development is still in its preliminary step and further considerations needs to be incorporated to be a viable and practical prediction technique including multi-dimensional verifications, global optimization, etc. which have not been resolved in the present study.

**Key words :** Geostatistical simulation, History matching, Groundwater, Subsurface prediction, Gibbs sampler

#### 1. 서 론

지하공간 특성화 자료는 지하를 대상으로 하는 모든 산업에 있어 예측되는 결과의 불확실성을 제한하는 중요한 역할을 한다. 뿐만 아니라 특성화 자료는 개념모델 수립, 설계, 및 위해성 평가 등 과정 전반에 걸쳐 활용되므로 지하매질 분포에 대한 건전한 예측은 과업의 성공적인 수행에 기초가 된다. 최근 국내에서는 이산화탄소 지중저장 대상 분지 및 고준위 방사성 폐기물 처분부지 선정 등이 주요 이슈가 되고 있다. 이러한 사업들의 대상은 공통적으로 지하공간이며 협소한 국토면적을 고려할 때 지하를 대상으로 한 범국가적 사업은 추후에도 지속될 것으로 예

측된다. 또한 1900년 후반부터 인지되기 시작한 전지구적 기후변화에 따라 가뭄을 포함한 기상의 다양한 변동성이 지속되고 있다. 이에 따라 지하수 자원의 가치가 재고되면서 지하수의 효율적인 관리 및 보전이 중요한 화두가 되고 있다. 지역적 규모에서 지하수 자원의 효과적인 개발, 관리, 및 보전을 위해서는 마찬가지로 지하 공간에 대한 특성화가 밑바탕이 되어야 한다.

일반적으로 지하에는 서로 다른 물리화학적 특성을 지닌 다양한 매질이 불균질하게 분포되어 있으며 다양한 규모의 불균질성(heterogeneity)은 대표적 지하 유체인 지하수의 흐름이나 화학적 성상 및 확산에 중대한 영향을 미친다. 따라서 지하 매질의 불균질성을 적절하게 규명하는

\*Corresponding author : [egpark@knu.ac.kr](mailto:egpark@knu.ac.kr)

원고접수일 : 2012. 7. 27 심사일 : 2012. 9. 3 게재승인일 : 2012. 9. 3  
질의 및 토의 : 2012. 12. 31 까지

일은 지하를 대상으로 하는 다양한 과업에 있어 성공적 수행 여부를 결정한다고 하여도 과언이 아니다. 지하매질의 불균질성을 규명하기 위한 노력은 조사 혹은 탐사 기술의 개발 및 현장적용 기술 개발 등의 하드웨어적인 부분과 조사 및 탐사 등을 통하여 얻어진 자료를 해석하여 신뢰성 있는 결과를 만들어 내는 소프트웨어적인 부분으로 대별된다. 소프트웨어적 기법들은 하드웨어적 기법을 보완하여 지하매질의 불균질성 예측을 개선하고 정량화하는 역할을 한다. 그러나 아직까지 국내에서는 하드웨어적 기법들에 기초한 정성적인 규명들이 주를 이루고 있으며 소프트웨어적 기법들의 중요성이 간과되고 있는 현실이다.

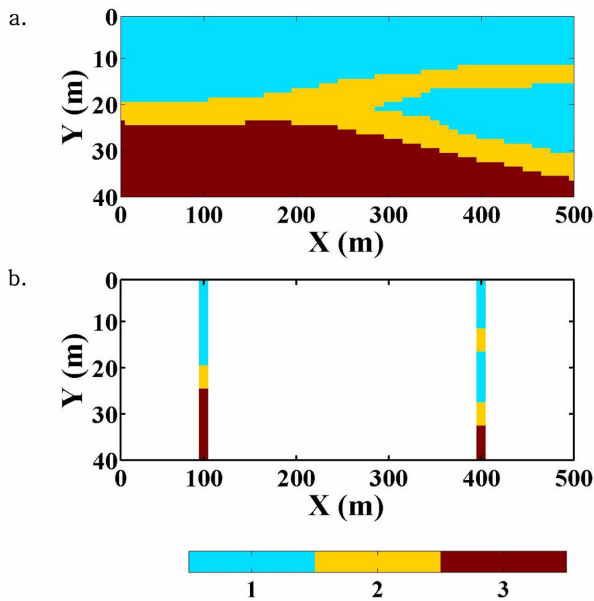
최근 30여 년 동안 정량적 지하매질 예측에 이용되고 있는 대표적 방법론은 지구물리학적 역산해석과 지구통계 모사기법이다. 지하수 분야에 있어서도 역산해석이나 지구통계 기법 혹은 두 가지 기법을 병합한 수리특성 예측 정량화가 이루어져왔다. 1980년대 이후 예측의 신뢰성을 향상시키는 방법으로 다양한 자료를 중첩하여 해석하는 방법들이 지속적으로 개발되어 왔으며 이들 방법들은 대수층의 투수성을 예측하기 위하여 현장으로부터 측정된 투수성 자료와 수리 또는 압력수두(Kitanidis, 1995; Yeh et al., 1996), 탄성과 주시(Doyen, 1988; Hyndman et al., 1994), 전기비저항(Daily et al., 1992; Pesti et al., 1993; Ezzedine et al., 1999) 등 투수성을 간접적으로 지시하는 자료를 연계하여 예측을 시행하였다. 또한 지구통계 모사기법에서는 야외에서 측정된 자료들이 가져올 수 있는 예측의 불확실성을 다루기 위하여 추계론적 방법론을 채택하고 있으며 이에 는 대표적으로 SGSim (Sequential Gaussian SIMulation) 및 SISim(Sequential Indicator SIMulation) 등이 있다(Deutsch and Journel, 1998). 이렇게 다양한 방법들을 통하여 얻어진 정량적 예측결과는 궁극적으로 모델링 예측 또는 평가에 활용된다. 그러나 실측된 수리수두와 모사수두 간의 차이를 목적 함수로 두지 않는 대부분의 예측결과들은 지하수 모델링을 위한 입력자료로 활용되기에 어려움이 있으며 추가적인 보정(calibration) 작업을 요구한다. 또한 비제약(unconstrained) 지구통계 재현(realization)에 근거한 몬테카를로(Monte Carlo) 모사를 통해 추계론적 분석을 실시할 경우 모사 결과가 과도한 불확실성을 포함한다는 문제점을 지닌다.

Konikow and Bredehoeft(1992)는 모델링의 예측능을 시험할 수 있는 유일한 도구가 히스토리매칭(history matching)임을 강조한 바 있다. 히스토리매칭 기법은 석유공학 분야에서 널리 이용되고 있으며 과거 유전의 생산정

자료에 기초하여 미래의 생산량 예측에 이용되는 최적화 기법이다(Jung and Choe, 2009; Han et al., 2010). 지하수 분야에서는 석유공학 분야와 달리 히스토리매칭을 통한 자동 최적화 예측 사례가 적으나 일부 연구에서 퇴적암 대수층(McKenna and Poeter, 1995) 및 결정질암 대수층(Day-Lewis et al., 2000)에 대하여 유사기법이 적용된 바 있으며, 그 외에도 trial-and-error 방법을 이용한 실측 대비 모사수위의 보정은 일상적으로 이루어지고 있다(Rogers, 1992).

Mariethoz et al.(2010)은 마르코프 연쇄 몬테카를로(Markov chain Monte-Carlo, MCMC) 기법의 일종인 Metropolis-Hastings 샘플링 알고리즘을 이용하여 모수와 비모수 예측 모두에 준용될 수 있는 베이지안(Bayesian) 역산해석 방법을 소개한 바 있다. 그러나 이 모델은 반복적인 샘플링 및 평가를 수행하여야 하므로 많은 전산자원이 필요하다는 문제점이 있다. 이 모델을 제외한 현재까지 개발된 대부분의 히스토리매칭 기법들은 지하물성의 연속적 변화를 가정하여 모수 또는 연속변수를 기반으로 하고 있다. 그러나 지하의 물성들은 지질에 종속되며 이러한 지질은 단속적인 특성을 보인다. 또한 지하매질은 지향적 비정규성(directional non-stationarity)을 보이는 것이 일반적이다(Park, 2010). 즉, 어느 한 방향으로 진행할 때의 지질매질 간 상호전이(mutual transition)와 그 반대 방향으로 진행할 때의 전이는 다른 경우가 보다 일반적이다(Carle and Fogg, 1996). 따라서 일반적으로 다중해(non-unique solution) 문제를 겪는 히스토리매칭 예측의 제한인자로 세미베리오그램을 이용하는 기존 지구통계 모사기법을 적용할 경우 한계성이 나타날 수밖에 없다. 이러한 기존 지구통계 모델의 대안으로 마르코비안 전이확률(Markovian transition probability)을 이용하는 모델의 적용이 보다 합리적일 수 있다. Park(2010)은 다차원 공간에서 지질이나 토양 속성과 같은 비모수의 공간적 예측을 위하여 마르코비안 전이우도함수를 개발하고 이를 지향적 비정규성을 지니는 2차원 및 3차원 공간 예측에 적용한 바 있다.

본 연구의 최종적인 목적은 MCMC 기법의 일종인 깁스 샘플러(Gibbs sampler)에 근거하여 히스토리 매칭 기법을 통한 비모수 지구통계 예측성 향상 모델을 제안하는데 있다. 일반적으로 깁스 샘플링은 Metropolis-Hastings 샘플링에 비하여 직접적이고 빠른 방법론으로 알려져 있다. 모델 개발을 위하여 가상의 상황을 상정하여 마르코비언 우도함수 기반 지구통계 모델을 통한 재현을 만들어 내고 이를 이용한 부정류 지하수 유동모델링이 수행될 것



**Fig. 1.** (a) A synthetic true field with dimension of 500 m × 40 m, and (b) collected borehole data at x = 100 and 400 m, composed of three different subsurface media (silt = 1; coarse sand = 2; and fine sand = 3).

이다. 또한 연구의 최종적인 결과물로 히스토리 매칭 기법을 이용한 비모수 지구통계 예측기법을 제안할 것이다. 전반적인 연구의 초기 단계로서 제안 모델의 제한성 및 향후 개선 방향에 대하여 추가적으로 논의할 것이다.

## 2. 실험방법

### 2.1. 가상의 대수층 및 양수시험 설정

본 연구에서는 개발된 모델의 예측능 및 불확실성에 대한 정량적인 평가를 위하여 가상의 자료를 생성하여 이용하였다. 가상의 자료는 Fig. 1a와 같이 수평 방향으로 500 m 그리고 수직 방향으로 40 m의 디멘전을 가지는 2차원 공간을 공간적인 상관성 규모가 최소 10:1 이상임을 가정하여 10 m × 1 m 크기의 셀로 균등하게 차분한 50 × 40개의 셀로 이루어진 도메인이며 세 가지의 서로 다른 지질로 구성되어 있는 것으로 가정하였다. 전반적으로 도메인에 해당하는 대수층은 상하부 불투수층에 의하여 피압상태에 놓여있으며, 이를 구성하는 세 가지 지질은 미고결 상태의 세립사질(매질1), 분급이 잘된 자갈(매질2), 및 분급이 잘된 조립질 모래(매질3)를 가정하였다. 따라서 매질2로 이루어진 층이 지하수의 주요 투수성 구조 역할을 한다. 그림에서 매질1 층은 도메인의 우측에서 사질 층과 교호하며 나타나는 양상을 보여주며 매질3의 매질은

도메인의 하부를 차지한다. 보다 사실적인 상황을 구성하기 위하여 본 연구에서는 도메인의 좌측과 우측 각각 x = 100 m 및 x = 400 m 지점에 도메인 전체를 관통하는 가상의 보어홀을 이용한 자료 수집을 가정하였으며 이로부터 추출된 매질의 자료는 Fig. 1의 (b)에서 보는 바와 같다. 또한 추출된 코어로부터 각 매질의 수리전도도 값인 0.001(매질1), 10(매질2), 및 0.1(매질3) cm/s의 값을 얻을 수 있는 것으로 가정하였으며 이에 따라 매질의 수리전도도에 대한 불확실성은 없는 것으로 간주하였다 (Bear, 1972). 또한 각 매질의 비저유계수는  $1 \times 10^{-6}$ (매질1),  $1 \times 10^{-4}$ (매질2), 및  $1 \times 10^{-5}$ (매질3)  $m^{-1}$ 로 가정하였다 (Heath, 1983).

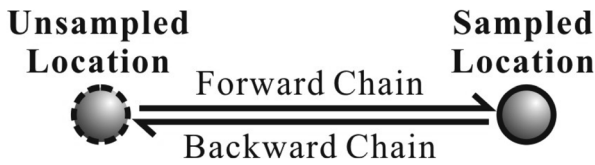
또한 정밀한 대수층 매질 및 주요 투수성 구조의 공간적 분포 규명을 위하여 x = 100 m 위치의 보어홀에서 상부로부터 10 m 구간마다 상하부를 패커로 고립시킨 후 수두를 50 m로 유지시키면서 x = 400 m 위치의 투수성 구조가 확인된 심도 15 및 30 m에서 매 24시간 마다 총 10일 동안 수위를 측정하였다. 이러한 상황을 모의하기 위하여 유한차분 지하수 유동 소프트웨어인 MODFLOW-2000(Hill et al., 2000)이 이용되었다.

### 2.2. 지구통계 예측 모델

초기예측의 생성 및 히스토리매칭 결과와의 중합을 통한 예측에는 Park(2010)에 의해 개발된 GCMC(generalized coupled Markov chain) 모델의 개량 모델이 이용되었다. 본 연구에서는 기존의 모델(Park, 2010)의 예측을 사전확률로 우도함수를 추가하여 최종적으로 개선된 사후확률이 얻어지도록 수정하였다. GCMC 모델은 지시자 기반 비모수 모델로 공간자료가 지니는 마르코비안 특성에 기초하여 매질 간의 전이확률(transition probability) 및 우도함수(likelihood function)를 이용하여 다차원 공간을 예측한다. 또한 앞서 밝힌 바와 같이 기존의 세미-베리오그램(semi-variogram) 및 크리깅(kriging) 기반 모델들에서 다룰 수 없었던 지향적 비정규성을 예측에 이용할 수 있으며 공간적으로 비정규적인 매질의 분포 예측에 적합하다. 1차원에서의 우도함수는

$$\pi_i(S_j) = \frac{\sqrt{1 \cdot p_{ji}^{(d1)} \cdot 2 \cdot p_{ij}^{(d1)} \cdot m_i \cdot m_j}}{m'_i}, \quad i = 1, \dots, n, \quad (1)$$

와 같이 주어지고, 여기서  $\pi_i(S_j)$ 는 인근 관측지점(sampled location)에서의 매질이  $S_j$ 일 경우 예측지점(unsampled location)에서의 매질이  $S_i$ 로 주어질 우도이며  $^1p_{ji}$ , 및  $^2p_{ij}$ 는 각각 관측지점에서 예측지점으로의  $j \rightarrow i$



**Fig. 2.** A schematic diagram of one-dimensional conditioning of sampled data where in forward chain, the probability of sampled data unsampled conditioned on unsampled data is acquired and, in backward chain, the probability of unsampled data conditioned on sampled data is pursued.

전이확률 및 예측지점에서 관측지점으로의  $i \rightarrow j$  전이확률을 의미하며 전이확률행렬(transition probability matrix),

$${}^n \mathbf{p}^{(r)} = \begin{bmatrix} {}^n p_{1,1}^{(r)} & {}^n p_{1,2}^{(r)} & \cdots & {}^n p_{1,n-1}^{(r)} & {}^n p_{1,n}^{(r)} \\ {}^n p_{2,1}^{(r)} & {}^n p_{2,2}^{(r)} & \cdots & {}^n p_{2,n-1}^{(r)} & {}^n p_{2,n}^{(r)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ {}^n p_{n-1,1}^{(r)} & {}^n p_{n-1,2}^{(r)} & \cdots & {}^n p_{n-1,n-1}^{(r)} & {}^n p_{n-1,n}^{(r)} \\ {}^n p_{n,1}^{(r)} & {}^n p_{n,2}^{(r)} & \cdots & {}^n p_{n,n-1}^{(r)} & {}^n p_{n,n}^{(r)} \end{bmatrix} \quad (2)$$

에서  $(j, i)$  및  $(i, j)$  성분을 통해 얻을 수 있다. 또한  $m_i$  및  $m_j$ 는 각각  $S_i$  및  $S_j$ 의 주변확률(marginal probability)을 의미하며,  $m'_i$ 는 관측지점에서  $S_j$ 가 발견되기 이전의  $S_i$  주변확률을 의미한다. 1차원에서 두 방향으로의 매질 간 전이가 조건부 독립일 경우 두 체인은 동일한 가중치인 1/2를 갖게 되므로 식은 위와 같이 제곱근으로 표현된다. 이러한 두 방향으로의 독립적 전이를 그림으로 표현한 것은 아래의 Fig. 2와 같다.

### 2.3. 히스토리매칭을 통한 대수층 매질분포 예측의 향상

보어홀로부터 관측된 지하매질 자료와 심도별 양수시험 자료를 이용한 대수층 매질분포 고도화 예측의 전략으로 본 연구에서 제안하는 방법론은 다음과 같다. 우선 초기 예측으로 보어홀 자료에만 근거하여 대수층 매질의 분포를 예측하는 다중 재현을 실시한다. Fig 1a에서 알 수 있듯이 대수층 매질은 수직 방향에 있어 상부의 매질과 하부의 매질이 상응하지 않는 지향적 비정규성을 보이고 있다. 이러한 상황을 적절히 모사하기 위해서는 지향적 비정규성을 다룰 수 있는 모델이 필요하다. 이에 따라 본 연구에서는 GCMC(generalized coupled Markov chain, Park, 2010)를 주요 지구통계 예측도구로 이용하였으며 보어홀 자료(Fig. 1b)에만 근거하는 총 100개의 단일재현(single realization)이 만들어졌다. 본 연구에서 제안하는 기법의 기본적인 가정은 제한된 정보로부터 생성된 초기 예측의 스펙트럼이 실제 매질분포에 대한 정보를 반영한

다는 것이다. 또한 초기예측이 지니는 불확실성은 추가적인 정보의 획득을 통하여 제한될 수 있다는 것이다. 따라서 초기예측 단계의 재현들은 가능한 한 큰 자유도(degree of freedom)를 가지는 것이 바람직하며 이를 위하여 미관측 지점에 대한 지구통계 예측 알고리즘에는 최소한의 조건부자료(conditioning information)가 이용되었다.

앞서 밝힌 바와 같이 초기예측의 스펙트럼은 대수층 자료를 입력 자료로 활용함으로 인하여 실제 매질의 특성을 반영할 수 있으나 정보의 불충분으로 인하여 예측의 오차 내지 불확실성을 포함하게 된다. 예측의 불확실성을 제한하는 방법으로 히스토리매칭 기법이 활용될 수 있다. 이를 위하여 앞서 생성된 100개의 재현에 대해 MODFLOW를 이용하여 동일한 조건의 양수시험 모사를 실시하고 그 결과를 바탕으로 각 단일재현에 대한 채택 및 기각 판정을 실시하였다. 본 연구에서는 관측점 심도 15 및 30 m에서 측정된 실제 수위변화와 모사된 수위변화 간 RMSE(root mean square error)를 기준으로 상위 10%에 해당하는 재현을 채택하였으며 두 지점에서 공통적으로 채택된 재현을 포함하여 총 18개의 재현이 선정되었다. 본 연구에서 이용된 RMSE( $\bar{e}$ )는 아래의 식과 같으며,

$$\bar{e} = \sqrt{\frac{1}{n_{tst}} \sum_{i=1}^{n_{tst}} (h_i^* - h_i)^2} \quad (3)$$

여기서  $n_{tst}$ 는 시간 차분의 수,  $h_i^*$ 는 실제 매질 분포로부터 얻어진  $i$ 번째 단계에서의 수위,  $h_i$ 는 재현된 매질 분포로부터 얻어진  $i$ 번째 수위를 의미한다.

실제, 예측의 결과가 국소최저치(local minima)가 아닌 광역적최저치(global minima)가 되기 위해서는 일정 확률로 하위 90%에 해당하는 재현을 채택하는 것이 바람직하나 본 연구에서는 매질 분포의 단순함을 고려하여 국소최저치가 존재하지 않는다는 가정 하에서 상위의 재현만을 채택하였다. 그러나 매질의 분포가 보다 복잡할 경우 국소최저치가 다수 존재할 수 있으며 그러한 경우에 대해서는 향후 연구에서 추가적으로 다룰 예정이다. 또한 본 연구에서 이용되는 깃스 샘플링 기법은 국소최저치 문제에 대하여 안정적인 것으로 알려져 있다(Sheng et al., 2003). 채택된 재현들은 초기 생성된 재현들 중 실제 대수층 매질의 분포 상황을 보다 적절하게 반영하는 것으로 가정할 수 있으며 이러한 재현들의 조합은 지구통계 모사를 이용하여 예측의 질을 향상시키는데 이용될 수 있다. 본 연구에서는 이러한 점에 착안하여 채택된 재현을 이용한 앙상블 확률 분포도(ensemble probability distribution)를 제작하고 이를 지구통계 모델의 보조적 입력자료로 활

용하였다. 연구에 이용된 기존 지구통계모델과의 중합을 위하여 아래의 식과 같이 깁스 표집기(Gibbs sampler)를 이용한 베이지안 업데이트 기술을 활용하였다(Geman and Geman, 1984; Besag, 1986; Doyen, 1994; Ezzedine et al., 1999).

$$\pi' = C f(\rho|Z) \pi \quad (4)$$

위의 식 (1)에서  $\pi'$  은 사후확률(posterior probability),  $C$ 는 정규화 상수(normalize constant),  $\pi$ 는 사전확률(prior probability), 그리고  $f(\rho|Z)$ 는 각 매질의 앙상블 확률에 근거한 우도함수(likelihood function)이다. 따라서 식 (1)에 의해 업데이트된 사후확률은 사전확률에 비하여 보다 많은 정보를 포함하게 된다.

이와 같이 적합한 재현을 채택하고 이를 앙상블 확률로 구성하여 예측의 질을 향상시키며 다시 향상된 재현을 통하여 적합한 재현을 채택하여 업데이트하는 과정은 더 이상 추가적인 예측의 개선이 이루어지지 않는 상황까지 지속적으로 반복 수행된다.

이러한 반복적인 과정을 통하여 비모수 지구통계 예측능의 향상이 이루어지는 과정을 정리하면 다음과 같다.

- 1) 보어홀로부터 획득한 비모수자료 입력
  - 2) 지구통계 모사 실행
  - 3) 다중의 재현 생성
  - 4) 다중의 재현을 이용한 지하수 유동 모사 실시
  - 5) 실제 현장으로부터 얻어진 양수시험 자료와 재현으로부터 얻어진 시험자료의 비교 및 히스토리 매칭 실시
  - 6) RMSE를 기준으로 상위 10%의 재현을 선택
  - 7) 재현들의 평균적인 RMSE가 목적하는 오차한계 이내일 경우 프로세스를 종료하며 그렇지 않을 경우 상위 10%의 재현으로 앙상블 확률도를 제작하여 지구통계 모사에 입력하여 2)~7) 과정을 반복수행
- 또한 이를 개념적 과정도로 표현한 것은 Fig. 3과 같다.

### 3. 결과 및 고찰

제안된 모델의 적용을 위하여 Fig. 1b의 보어홀 추출 자료에 기초한 초기예측 지구통계 모사를 실시하였다. 모사에는 Park(2010)이 개발한 마르코비안 비모수 지구통계 모델인 GCMC3D가 이용되었다. 모사를 위하여 이용된 매질 간의 전이확률 자료는  $x=100$  및  $400$  m의 보어홀 자료를 입력자료로 하여 얻어진 보로노이 다이어그램(Voronoi diagram)으로부터 추출되었다. 보로노이 다이어그램은 유클리디안 공간을 거리에 따라 분할하는 단순한

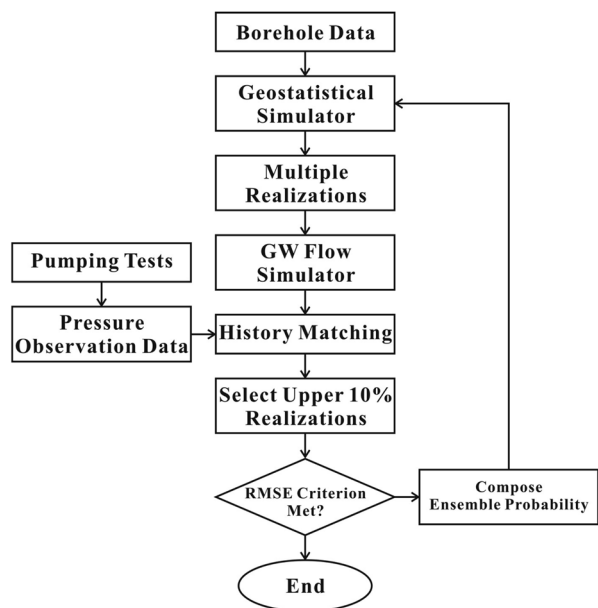


Fig. 3. Proposed iterative procedure of improving geostatistical predictions based on pumping tests and history matching scheme.

방법이다. 산정된 평균적 수평 상관거리는 각각 매질1 210.2 m, 매질2 128.1 m, 그리고 매질3 125.0 m이다. 또한 수직적 상관거리는 상부에서 하부 방향으로 매질1 53.0 m, 매질2 12.0 m, 그리고 매질3 24.4 m이며 하부에서 상부 방향으로 매질1 19.5 m, 매질2 12.0 m, 및 매질3 580.7 m로 나타나 수직 방향으로의 지향적 비정규성이 있음을 보여준다. 매질1, 매질2, 및 매질3 각각의 주변확률은 51.2, 18.8, 및 30%로 산정되었으며 GCMC3D 모사에 활용된 전이확률행렬은 Table 1에 주어진 바와 같다.

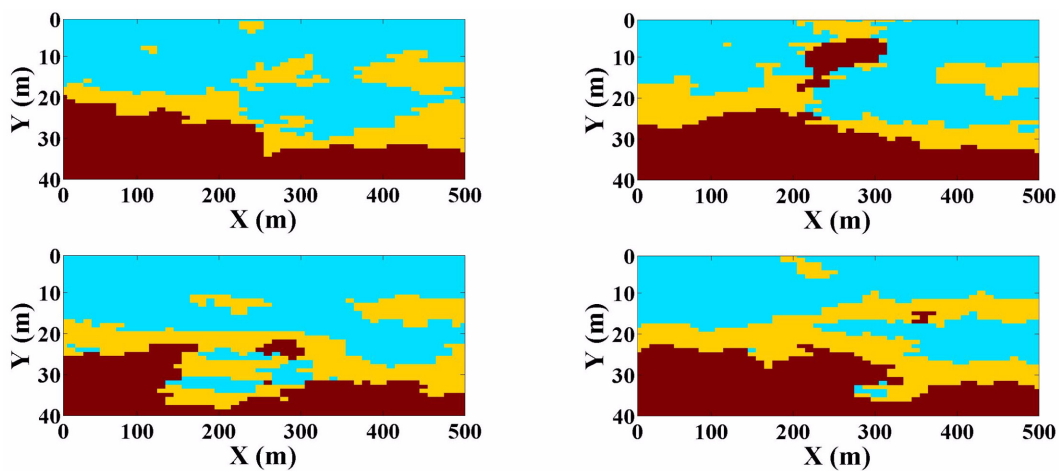
Fig. 4는 총 100개의 초기예측 중 무작위로 선택된 4개의 예측결과를 보여주며 Fig. 5은 100개의 초기예측을 중첩하여 출현 빈도로부터 얻어진 각 매질의 앙상블 확률 분포를 보여준다. 그림에서 붉은색은 높은 확률, 즉 100개의 재현 중 다수의 재현이 해당 매질을 예측하였음을 의미하며, 이와 반대로 파란색은 낮은 확률을 의미한다. Fig. 4에서와 같이 초기예측에서는 주요 투수성 구조인 조립 사질 매질의 연결성이 Fig. 1a와 달리 도메인의 중앙부에서 단절되는 양상으로 주로 나타나며 이러한 특성은 Fig. 5의 앙상블 확률분포에서도 잘 나타난다. 특히, 입력 자료의 위치로부터 원격에 위치하는 도메인의 중앙부의 경우 Fig. 5에서 보이는 바와 같이 앙상블 확률의 분산이 크게 나타나며 이는 이 부근에서 예측의 불확실성이 크다는 것을 의미한다. 반대로 입력한 자료 인근인  $x=100$  및  $400$  m 주변에서는 확률의 분산이 상대적으로 낮게 산정되었으며 이는 낮은 불확실성을 의미한다.

**Table 1.** Transition probabilities acquired from Voronoi diagram based on borehole data at  $x = 100$  and  $400$  m

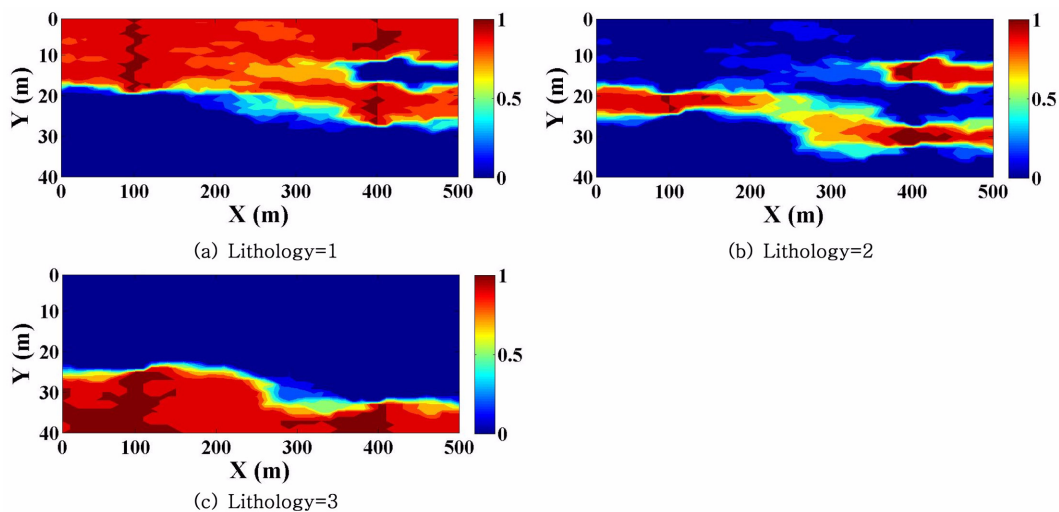
	Positive X-direction TPM				Positive Y-direction TPM		
	Lith. = 1	Lith. = 2	Lith. = 3		Lith. = 1	Lith. = 2	Lith. = 3
Lith. = 1	0.993	0.006	0.001	Lith. = 1	0.925	0.074	0.001
Lith. = 2	0.016	0.981	0.003	Lith. = 2	0.069	0.796	0.135
Lith. = 3	0.007	0.01	0.983	Lith. = 3	0.002	0.002	0.996

	Negative X-direction TPM				Negative Y-direction TPM		
	Lith. = 1	Lith. = 2	Lith. = 3		Lith. = 1	Lith. = 2	Lith. = 3
Lith. = 1	0.99	0.006	0.004	Lith. = 1	0.973	0.026	0.001
Lith. = 2	0.016	0.968	0.016	Lith. = 2	0.201	0.796	0.003
Lith. = 3	0.001	0.002	0.997	Lith. = 3	0.001	0.085	0.914



**Fig. 4.** Randomly selected 4 single realizations solely based on conditioning borehole data at  $x = 100$  and  $400$  m.

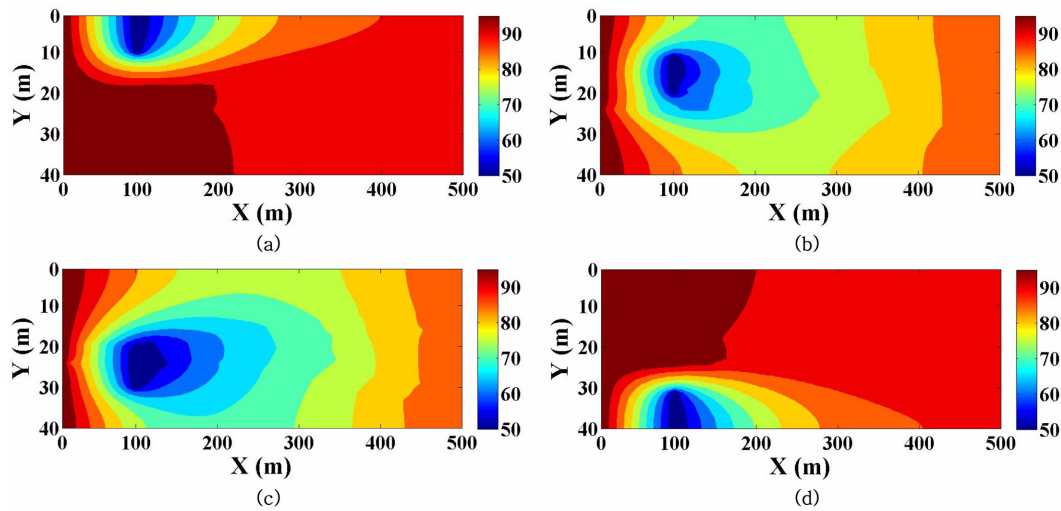


**Fig. 5.** Ensemble probability distributions of 3 different subsurface media built from 100 single realizations conditioned only on 2 borehole data at  $x = 100$  and  $400$  m.

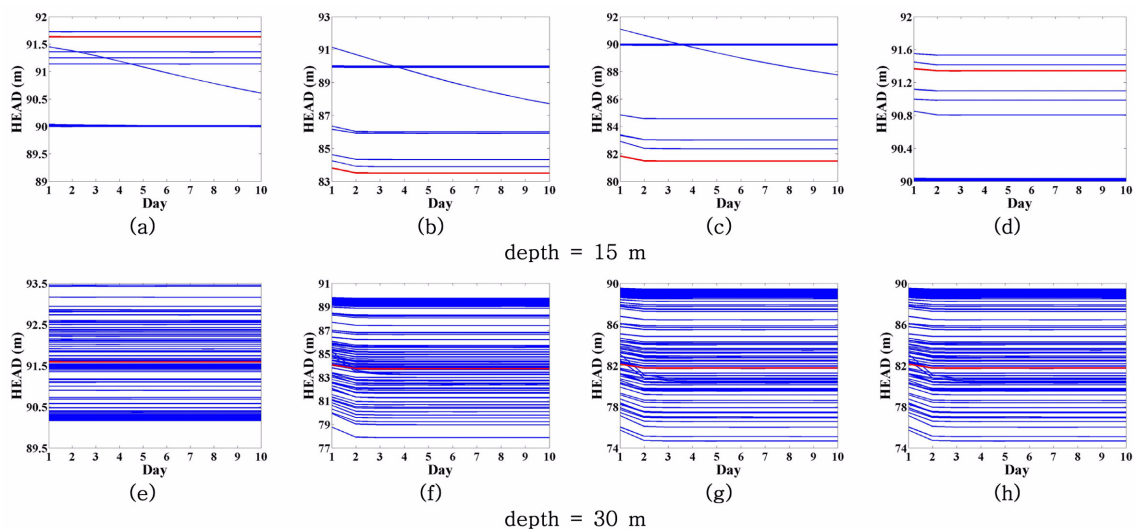
얻어진 각각의 재현들은 MODFLOW 입력자료 중 수리전도도 및 비저유계수의 분포를 구성하는데 활용되었다.

또한 MODFLOW의 고정수두경계 패키지를 이용하여 도메인 좌측  $x = 0$  m에 해당하는 경계는 100 m 수두가 지





**Fig. 6.** Hydraulic head distributions based on synthetic true map (Fig. 1a) at  $t = 10$  days since the pumping at 4 different vertical interval (a) 0~10 m (test 1), (b) 10~20 m (test 2), (c) 20~30 m (test 3), and (d) 30~40 m (test 4), where pumping well is located at  $x = 100$  m.

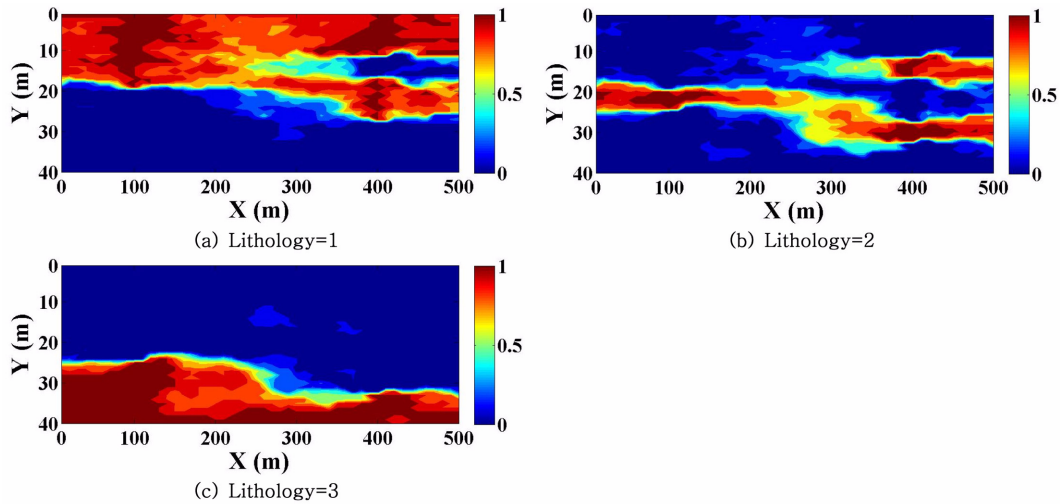


**Fig. 7.** Comparisons of true history curves and the curves at  $x = 400$  m from groundwater flow simulations of 100 single realizations of (a) test 1 at obs. depth  $y = 15$  m, (b) test 2 at obs. depth  $y = 15$  m, (c) test 3 at obs. depth  $y = 15$  m, (d) test 4 at obs. depth  $y = 15$  m, (e) test 1 at obs. depth  $y = 30$  m, (f) test 2 at obs. depth  $y = 30$  m, (g) test 3 at obs. depth  $y = 30$  m, and (h) test 4 at obs. depth  $y = 30$  m.

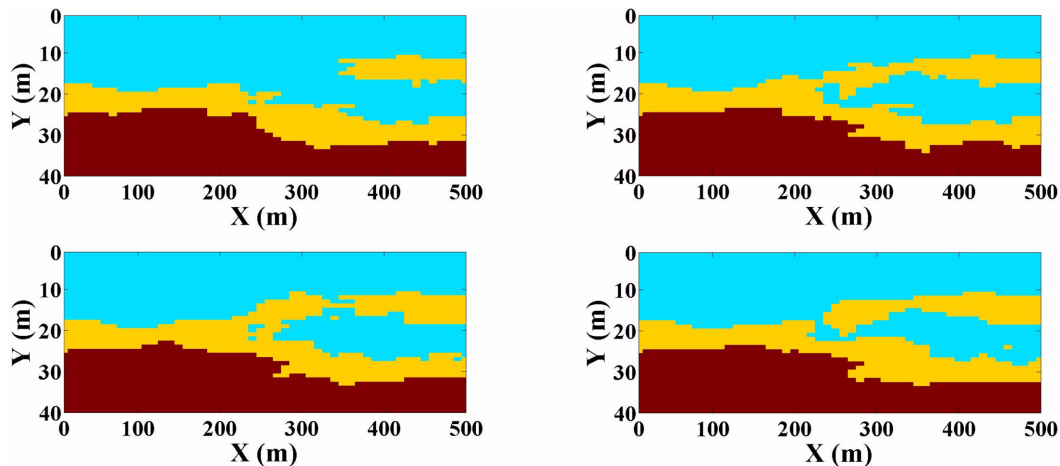
속적으로 유지되는 경계, 우측  $x = 500$  m 경계는 90 m의 경계로 설정하였으며 상하부  $y = 0$  및 40 m 경계는 특정 조건을 설정하지 않았으므로 공통적으로 불투수경계로 설정되었다. 양수정의 위치인  $x = 100$  m에 대해서는 0~10(양수시험 1), 10~20 (양수시험 2), 20~30 (양수시험 3), 및 30~40 m(양수시험 4) 구간에서 총 4번의 구간 고립 양수시험을 순차적으로 시행하는 것을 모의하기 위하여 양수시험 1~4에 해당하는 고정수두경계 패키지 파일을 구성하고 해당하는 구간의 양수시험 목표수두를 50 m로 설정하였다. 모사에 이용된 초기수두는 실제 매질분포(Fig.

1a) 및 각 재현에 근거하여 양수가 없는 상황을 가정한 정류(steady-state) 모사로부터 얻어졌다. 초기수위 모사를 제외한 모든 모사는 부정류(transient) 방식을 통하여 실시되었으며 이에 이용된 시간 차분은 1일로 설정되어 총 10일 동안의 모사가 이루어졌다. Fig. 6의 수위분포는 실제 매질분포를 가정하여 양수시험 1~4에 해당하는 시험 이후 10일의 지하수위 분포를 보여준다. 그림에서 양수시험 구간의 심도에 따라 관측점 위치( $x = 400$  m)에서의 수위가 서로 다르게 나타나는 것을 확인할 수 있다.

Fig. 7은 Fig. 1a에서 보이는 실제 대수층 매질 분포와



**Fig. 8.** Ensemble probability distributions of 3 different subsurface media built from selected 10 single realizations based on history matching RMSE.



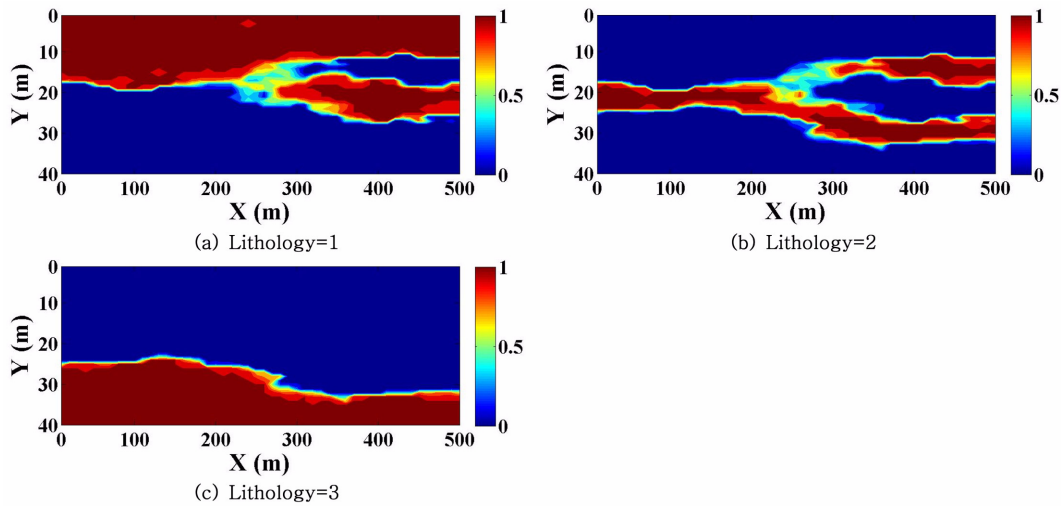
**Fig. 9.** Randomly selected 4 single realizations based on conditioning borehole data at  $x=100$  and  $400$  m and ensemble probability from 1st round history matching scheme.

100개의 초기예측을 이용하여 제작된 히스토리 곡선을 보여준다. 그림에서 붉은색 실선은 실제 대수층 매질에 근거하여 모사된 수위 변화 곡선이며 파란색 실선은 지구통계 예측을 통하여 재현된 매질 분포에 근거하여 모사된 수위 변화 곡선이다. 그림에서와 같이 지구통계 재현을 이용하여 예측된 수위는 매우 큰 분산을 보이고 있으며 특히 관측정 15 m 심도에서의 히스토리 커브는 매우 큰 과예측 또는 저예측(over- or underestimation) 양상 및 왜도(skewness)를 보이는 것으로 분석되었다. 이렇게 100개의 초기 매질예측이 보여주는 큰 분산 및 분포의 편중은 지구통계 예측만으로 얻어진 지하수위 예측이 큰 불확실성을 포함하며 예측 결과를 보다 실용적으로 활용하기 위해서는 추가적인 정보를 통한 불확실성의 축소가 필요함

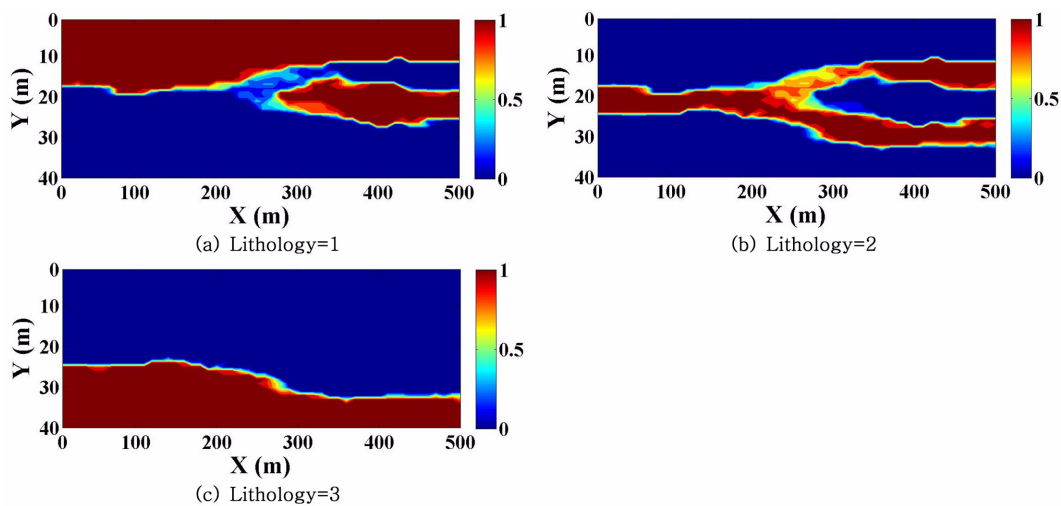
을 보여준다.

앞서 밝힌 바와 같이 초기재현으로부터 만들어진 히스토리 커브와 실제 히스토리 커브 간의 RMSE에 근거하여 낮은 오차를 보이는 재현이 보다 높은 우도를 보이는 것으로 가정하고 상위 10%의 재현들이 채택되었다. 이러한 채택 및 기각 방법에 대해서는 향후 연구에서 보다 심도 있게 다룰 예정이다. 이들을 이용한 앙상블 확률 분포는 Fig. 8과 같다. Fig. 8과 Fig. 5에서 주요 투수층인 조립사질 매질의 확률분포의 두드러진 차이점은, Fig. 8에서의 확률분포가 Fig. 5보다 상대적으로 개선된 매질의 연결성을 보여준다는 것이다. 이러한 결과는 100개의 단일재현에 근거한 지하수위의 부정류 모사결과 중 실제 매질분포와 유사한 재현들이 보다 실제 히스토리 커브와 유사한





**Fig. 10.** Ensemble probability distributions of 3 different subsurface media built from 100 single realizations conditioned on 2 borehole data at  $x = 100$  and  $400$  m and 1st round history matching ensemble probability.



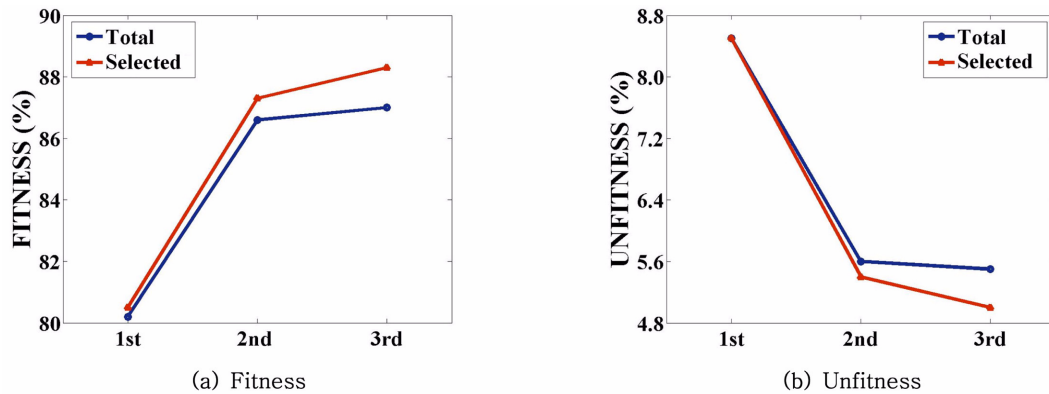
**Fig. 11.** Ensemble probability distributions of 3 different subsurface media built from 100 single realizations conditioned on 2 borehole data at  $x = 100$  and  $400$  m and 2nd round history matching ensemble probability.

지하수위 감소를 보여주어 상대적으로 낮은 RMSE를 보이기 때문인 것으로 판단된다. 그러나 여전히 관측정 위치 심도 15 m에 나타나는 투수성 구조는 양수정으로부터 연장된 투수성 구조와 단절되어 고립된 양상을 보이며 제안된 알고리즘을 통한 추가적 개선이 필요하다.

Fig. 9 및 10은 각각 채택된 재현들의 양상별 확률을 입력인자로 하여 예측된 100개의 재현 중 4개의 단일재현과, 100개의 재현에 기초하여 제작된 양상별 확률분포를 보여준다. 앞서의 결과와 달리 히스토리매칭 정보에 기초하여 개선된 예측의 주요 투수성 구조 양상별 확률분포는 Fig. 5와 달리 연결성이 상대적으로 뚜렷하게 나타나

며 상부의 투수성 구조 역시 양수정을 지나는 투수성 구조와 연결되는 양상을 보여준다. 이러한 예측의 개선은 앞서 실행한 초기예측 중 실제 히스토리 커브에 상응하는 지하수위 모사결과를 보여주는 재현들을 취사선택하여 양상별 확률을 구성하고 이를 두 번째 단계의 예측에서 입력인자로 활용하여 추가적인 정보(i.e. 양수시험자료)가 반영되었기 때문인 것으로 판단된다. 또한 이는 본 연구에서 제안한 히스토리 커브 정보를 통한 예측의 개선 모델이 유효하게 작동하고 있음을 의미한다.

본 연구에서는 위에서 언급한 모델 적용 루틴과 Fig. 3의 과정을 통하여 총 3회의 반복적 예측을 수행하였으며



**Fig. 12.** Changes of fitness (a) and unfitness (b) of the ensemble probability from total and the selected realizations with iteration number.

**Table 2.** Changes of fitness (a) and unfitness (b) of the ensemble probability from total and the selected realizations with iteration number, and the iteration representative values

	Total Realizations							
	Fitness (%)				Unfitness (%)			
	Lith. = 1	Lith. = 2	Lith. = 3	Average	Lith. = 1	Lith. = 2	Lith. = 3	Average
1st Iteration	87.27	65.31	87.95	80.2	10.97	12.02	2.59	8.5
2nd Iteration	92.04	75.76	91.91	86.6	6.7	8.01	2.21	5.6
3rd Iteration	92.14	77.18	91.74	87.0	6.55	8.02	1.89	5.5

	Selected Realizations							
	Fitness (%)				Unfitness (%)			
	Lith. = 1	Lith. = 2	Lith. = 3	Average	Lith. = 1	Lith. = 2	Lith. = 3	Average
1st Iteration	85.62	68.17	87.7	80.5	9.53	12.72	3.29	8.5
2nd Iteration	91.67	78.18	91.98	87.3	5.76	8.2	2.2	5.4
3rd Iteration	92.09	80.63	92.26	88.3	5.14	7.84	1.93	5.0

이에 따라 최종 3 단계에서 얻어진 앙상블 확률 분포는 Fig. 11와 같다. 그림에서 투수성 구조의 확률 분포가 Fig. 1a의 실제 구조와 상당히 유사해 졌음을 확인할 수 있다. 또한 최초예측에서 연결성이 거의 나타나지 않았던 도메인 우측 상부의 투수성 구조 역시 뚜렷하게 전체 투수성 구조에 연결되어 있는 양상이 관찰되어 히스토리매칭을 통한 양수시험 결과가 지구통계 예측에 적절하게 예측에 반영되었음을 확인할 수 있다. 따라서 긍정적인 측면에서 본 연구에서 제안하는 김스 표집기를 통한 히스토리매칭 및 지구통계 중합 모델이 예측의 질을 개선함을 알 수 있다.

보다 정량적인 방식으로 단계적인 예측능의 변화를 확인하기 위하여 Park(2010)이 제안한 fitness 및 unfitness 값의 추이를 살펴보았다.  $Fitness(F(\alpha))$  및  $unfitness(UF(\alpha))$ 에 이용된 식은 Park(2010)이 제안한 바와 같이

$$F(\alpha) = \frac{\sum_{i=1}^l \sum_{j=1}^m \delta_{\alpha}(i,j) E_{\alpha}(i,j)}{\sum_{i=1}^l \sum_{j=1}^m \delta_{\alpha}(i,j)}, \quad \alpha = 1, 2, 3, \text{ 및} \quad (5)$$

$$UF(\alpha) = \frac{\sum_{i=1}^l \sum_{j=1}^m \delta_{\alpha}(i,j) E_{\alpha}(i,j)}{\sum_{i=1}^l \sum_{j=1}^m \delta_{\alpha}(i,j)}, \quad \alpha = 1, 2, 3, \quad (6)$$

와 같다. 여기서  $\alpha$ 는 매질의 인덱스로 1은 매질1, 2는 매질2, 및 3은 매질3을 의미한다. 또한  $l$  및  $m$ 은 수평 및 수직 방향으로의 차분된 셀의 개수를 의미하며,  $E_{\alpha}(i,j)$ 는  $\alpha$ 에 해당하는 매질의 앙상블 확률 분포도에서  $(i, j)$  번째 셀의 값을 의미한다. 바이너리 값으로 주어지

는  $\delta_{\alpha}(i,j)$  는  $\alpha$ 에 해당하는 앙상블 확률이 0보다 클 경우 1 및 그렇지 않을 경우 0의 값을 생성한다. 이에 대한 여함수(complementary function)인  $\delta_{\alpha}(i,j)$  은 반대로  $\alpha$ 에 해당하지 않는 타 매질의 앙상블 확률이 0보다 클 경우 1, 그리고 그렇지 않을 경우 0의 값을 생성한다.

Fitness는 그 의미와 유사하게 앙상블 확률분포가 평균적으로 얼마나 실제 매질분포를 예측할 수 있는지의 포텐셜을 보여주며 unfitness는 앙상블 확률분포가 잘못된 예측을 할 포텐셜을 보여준다(Park, 2010). Fig. 12 및 Table 2는 재현들에 의한 단계별 예측능 혹은 해석능을 보여준다. 앞서 정성적인 분석과 마찬가지로 제안된 모델에 의한 예측의 질은 지속적으로 증가하는 양상을 보여준다. Fig. 12에서 fitness 값은 첫 번째 단계에서 세 매질 평균 80.2%를 보여주며 두 번째 및 세 번째 단계에서 각각 86.6% 및 87%를 보여준다. 상대적으로 unfitness 값은 지속적으로 감소하는 추세를 보여주며 첫 번째 단계에서 평균 8.5%, 그리고 두 번째 및 세 번째 단계에서 각각 5.6% 및 5.5%를 보여준다. RMSE 기준으로 상위 10%에 해당하는 재현을 선별하여 얻어진 fitness 및 unfitness의 경우에도 각각 80.5, 87.3, 및 88.3%의 단계적 증가와 8.5, 5.4, 및 5%의 단계적 하향 추세를 보여 예측의 질이 점차 개선됨을 보여준다. 이러한 결과들은 이전의 정성적 측면과 함께 정량적 측면에서 본 연구에서 제안하는 기법이 점진적 예측능 개선을 보여주며, 주요 수리구조에 대한 연결성을 예측하는데 있어 긍정적임을 보여준다.

#### 4. 결 론

본 연구에서는 비모수 지구통계 모사의 예측능 향상 및 불확실성 저감을 위하여 히스토리매칭 기법을 활용하는 방안으로 MCMC 기법의 일종인 깃스 샘플러를 이용하는 모델을 제안하고 모델의 적용을 위한 알고리즘을 개발하였다. 개발된 모델은 두 개의 보어홀 정보와 이 중 하나의 보어홀을 양수정으로 한 지하수 양수시험 모의 자료에 적용되었다. 매질의 분포 예측을 위해서는 Park(2010)이 개발한 비모수 지구통계 모사 모델인 GCMC가 이용되었다. 제안 모델의 적용 및 정성정량적 분석 결과 추가적인 정보인 히스토리 매칭자료를 이용한 지구통계 모사를 통하여 지하 매질의 공간적 분포 예측이 단계적으로 개선됨을 알 수 있었다. 특히 주요 투수성 구조의 분포 예측에 있어 제안된 모델의 개선 효과가 매우 뚜렷함을 확인하였다.

기존 히스토리 매칭에 기반하는 대부분의 모델들은 모

수 혹은 연속변수 기반이므로 지하 매질에서 일반적으로 나타나는 급격한 물리화학적 특성의 변화 등을 예측함에 있어 다양한 제한성이 따른다. 이러한 제한성을 극복하기 위해서는 비모수, 다중-가우시안(multi-gaussian), 혹은 지시자 기반 예측이 이루어져야 하나 아직까지 해당 분야의 기술 개발은 활발하지 않은 상황이다. 따라서 본 연구에서 제안하는 비모수 기반 히스토리 매칭 기술은 하드자료인 보어홀 자료와 소프트자료인 히스토리 커브 자료가 중첩되어 단속적이고 비정규적인 지하매질의 분포 예측을 지속적으로 개선할 수 있다는 측면에서 중요성을 갖는다. 또한 본 제안 모델은 GCMC 모델(Park, 2010)을 주요 지구통계 예측도구로 이용함으로써 일반적으로 지향적 비정규성을 보이는 지질매질 분포에 보다 적합하다. 따라서 본 연구에서 제안하는 모델의 기술이 성숙될 경우 다양한 실제적 상황에 적용하여 지하 불균질성 파악 및 현장 특성화에 이용될 수 있을 것이다. 또한 지하수의 개발이나 정화, 유가스전 개발, 지열개발 등 지하로부터 얻어지는 하드정보와 소프트 정보가 동시에 존재할 경우 제한 없이 적용되어 예측능 개선에 활용될 수 있을 것으로 판단된다.

그러나 아직까지 제안 모델은 전반적 개발 단계에 있어 초기 단계에 머물러 있으며 향후 지속적인 개선이 필요하다. 특히, 본 연구에서의 적용 사례는 매우 간단한 2차원 상황을 고려하였으며 수리적인 경계조건 등이 지나치게 간소화 되어 있다. 이러한 측면은 향후 3차원 상황 및 보다 실제적인 현장 특성을 고려한 테스트를 통하여 지속적으로 검증 및 발전되어야 할 것이다. 그러나 복잡한 3차원 문제들이 공통적으로 갖는 매우 느린 수치연산 속도는 제안 모델의 적용에 있어 제한점이 될 것으로 판단된다. 또한 아직까지 개발의 초기 단계로 이 연구에서 얻어진 결과가 비유일성해(non-unique solution) 문제가 다수 나타나는 지하수 유동 문제에 있어 광역적 최적화(global optimization)를 달성하였는지에 대해서는 여전히 의문이며 모델의 예측 결과가 국소 최저치(local minima) 문제를 해결할 수 있도록 지속적인 개선을 시도되어야 할 것이다. 이러한 개선을 위하여 다양한 최적화 기법들이 도입될 수 있을 것으로 판단된다.

#### 사 사

이 논문은 2010년도 정부(교육과학기술부)의 지원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010-0009900).

## 참 고 문 헌

- Bear, J., 1972, Dynamics of fluids in porous media, American Elsevier Pub. Co., New York, 764p.
- Besag, J., 1986, On the Statistical analysis of dirty pictures, *J. R. Statist. Soc.*, **48**(3), 259-302.
- Carle, S.F. and Fogg, G.E., 1996, Transition probability-based indicator geostatistics, *Math. Geol.*, **28**, 453-477.
- Daily, W., Ramirez, A., LaBrecque, D. et al, 1992, Electrical resistivity tomography of vadose water movement, *Water Resour. Res.*, **28**(5), 1429-1992.
- Day-Lewis, F.D., 2000, Identifying fracture-zone geometry using simulated annealing and hydraulic-connection data, *Water Resour. Res.*, **36**(7), 1707-1721.
- Deutsch, C.V. and Journel, A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide, 2nd edn., Oxford University Press., New York, 369 p.
- Doyen, P.M., 1988, Porosity from seismic data: A geostatistical Approach, *Geophys.*, **53**(10), 1263-1275.
- Doyen, P.M., Pasaila, D.E., and Strandenes, S., 1994, Bayesian sequential indicator simulation of channel sands from 3-D seismic data in the Oseberg field, Norwegian North Sea, SPE 69th Annual Technical Conference and Exhibition, SPE, New Orleans, 25-28.
- Ezzedine, S., Rubin, Y., and Chen, J., 1999, Bayesian method for hydrogeological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund site, *Water Resour. Res.*, **35**(9), 2671-2683.
- Geman, S. and Geman, D., 1984, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721-741.
- Han, Y., Park, C., and Kang, J., 2010, Multi-objective History Matching for Estimating the Individualized Performance of Multiple Production Wells, *The Korean society for geosystem engineering*, **47**(5), 660-667.
- Heath, R.C., 1983, Basic Ground-Water Hydrology, U.S. Geological Survey, Reston, Virginia, 81p.
- Hill, M.C., Harbaugh, A.W., Banta, E.R., McDonald, M.G., MODFLOW-2000, 2000, The U.S. geological survey modular ground-water model-user guide to modularization concepts and the ground-water flow process, U.S. Geological Survey, Reston, Virginia, 113p.
- Hyndman, D.W., Harris, J.M., and Gorelick, S.M., 1994, Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, **30**(7), 1965-1977.
- Jung, S. and Choe, J., 2009, Improvement of Numerical Stability in History-matching Using Streamline Assisted Ensemble Kalman Filter, *The Korean society for geosystem engineering*, **46**(4), 453-465.
- Kitanidis, P.K., 1995, Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, **31**(10), 2411-2419.
- Konikow L.F. and Bredehoeft, J.D., 1992, Ground-water models cannot be validated, *Adv. Water Resour. Res.*, **15**, 75-83.
- Mariethoz, G., Renard, P., and Caers, J., 2010, Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resour. Res.*, **46**, W11530, doi:10.1029/2010WR009274.
- McKenna, S.A. and Poeter, E.P., 1995, Field example of data fusion in site characterization, *Water Resour. Res.*, **31**(12), 3229-3240.
- Park, E., 2010, A multidimensional generalized coupled Markov chain model for surface and subsurface characterization, *Water Resour. Res.*, **46**, W11509, doi:10.1029/2009WR008355.
- Pesti, G., Bogardi, I., Kelly, W.E. et al., 1993, Cokriging of geoelectric and well data to define aquifer properties, *Groundwater*, **31**(6), 905-912.
- Sheng, Q., Moreau, Y., and De Moor, B., 2003, Biclustering microarray data by Gibbs sampling, *Bioinformatics*, **19**, 196-205, doi:10.1093/bioinformatics/btg1078
- Rogers, L., 1992, History matching to determine the retardation of PCE in ground water, *Ground Water*, **30**(1), 50-60.
- Yeh, T.-C.J., Jin, M., and Hanna, S., 1996, An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, **32**(1), 85-92.