

지하 불균질 예측 향상을 위한 마르코프 체인 몬테 카를로 히스토리 매칭 기법 개발

정진아 · 박은규*

경북대학교 지질학과

A Development of Markov Chain Monte Carlo History Matching Technique for Subsurface Characterization

Jina Jeong · Eungyu Park*

Department of Geology, Kyungpook National University, Daegu, Korea

ABSTRACT

In the present study, we develop two history matching techniques based on Markov chain Monte Carlo method where radial basis function and Gaussian distribution generated by unconditional geostatistical simulation are employed as the random walk transition kernels. The Bayesian inverse methods for aquifer characterization as the developed models can be effectively applied to the condition even when the targeted information such as hydraulic conductivity is absent and there are transient hydraulic head records due to imposed stress at observation wells. The model which uses unconditional simulation as random walk transition kernel has advantage in that spatial statistics can be directly associated with the predictions. The model using radial basis function network shares the same advantages as the model with unconditional simulation, yet the radial basis function network based the model does not require external geostatistical techniques. Also, by employing radial basis function as transition kernel, multi-scale nested structures can be rigorously addressed. In the validations of the developed models, the overall predictabilities of both models are sound by showing high correlation coefficient between the reference and the predicted. In terms of the model performance, the model with radial basis function network has higher error reduction rate and computational efficiency than with unconditional geostatistical simulation.

Key words : Bayesian inversion, History matching, Subsurface characterization, Markov chain Monte Carlo

1. 서 론

지하 매질 분포의 불균질성은 지중 유체의 거동 및 오염물질 확산에 있어 예측의 불확실성을 증폭시키는 원인이 되며 이에 따라 보다 정확한 지하 매질 예측 기법의 개발이 요구된다. 현실적으로 지하와 관련된 대부분의 연구에서는 직접 측정자료 및 간접자료의 양이 크게 제한된다. 따라서 지하 매질 분포 예측에는 많은 불확실성이 개입하게 되며 이러한 불확실성을 축소하고 정확한 예측을 얻어낼 수 있는 기법의 개발이 지속적으로 요구되고 있으며 국내외적으로 이러한 기법들의 개발이 진행되고 있다 (Bertino et al., 2002; Deutsch and Tran, 2002; Reichle

et al., 2002; Chen and Zhang, 2006; Mariethoz et al., 2009; Chen et al., 2009; Fu and Gomez-Hernandez, 2009; Mariethoz et al., 2010; Zhou et al., 2011, Shin et al., 2010).

지하로부터 얻어지는 자료에 기초하여 지하매질 분포를 얻기 위한 정량적 방법에는 크게 지구통계 기법(geostatistical method)에 근거한 예측과 지중에 가해지는 특정 스트레스에 대한 지하 유체의 반응을 역으로 해석하는 역산 기법(inversion method)이 있다. 지구통계 기법은 대상으로 하는 부지의 수 개 포인트에서 직·간접 자료를 수집하고 이들의 공간적 정규성에 기초하여 지하매질의 분포를 예측하는 통계학적 기법이며 예측하고자 하는 물성에

*Corresponding author : egpark@knu.ac.kr

Received : 2015. 2. 24 Reviewed : 2015. 3. 28 Accepted : 2015. 3. 31

Discussion until : 2015. 8. 31

대한 직접정보(hard information) 및 예측하고자 하는 물성의 특성을 간접적으로 지시하는 간접정보(soft information)을 모두 이용할 수 있다(Pesti et al., 1993; Hyndman et al., 1994; Ezzedine et al., 1999; Yoon et al., 2001; He et al., 2002; Kretz et al., 2004; Will et al., 2005; Jeong et al., 2014). 현재 이용되고 있는 지구통계 기법으로는 크리깅(kriging)으로 대표화되는 선형 지구통계 모델과 비선형 모델이 있으며 이들에 기초한 추계론적 모사기법들이 있다(Deutsch and Journel, 1998).

지구통계 기법을 이용하는데 있어 중요한 이슈로는 유체 흐름에 있어서의 포인트 자료 대표성과(서포트 불일치, support inconstancy) 한정된 자료를 통하여 얻어진 공간 통계의 정규성(비정규성, non-stationarity)을 들 수 있다. 포인트 자료의 경우 연구 대상 도메인에 비하여 매우 작은 규모에서 측정된 정보이며 도메인 전반의 유체흐름을 설명하기 위해서는 보다 큰 규모에서의 측정이 요구된다. 일반적인 지하수 자료 취득의 경우 포인트 자료는 슬러그 테스트(slug test)나 푸쉬-풀 테스트(push-pull test) 등 매우 작은 규모에의 대표성을 얻을 수 있는 시험들을 통하여 얻어지게 되며, 지구통계 내삽 기법을 통하여 이들을 규모확장(upscaling) 할 경우 도메인 전반의 유체흐름을 설명하기 어려운 경우가 일반적이다. 또한 연구 대상 도메인에서 포인트 자료는 일부 관정을 통하여 얻어지며 그 수가 적거나 일부에 편중될 경우 도메인 전체의 지구통계 특성을 반영하지 않을 수 있으며 이에 따라 비정규성 문제에 의한 예측의 실패 내지 부정확성 문제가 발생할 수 있다.

지중 유체의 거동에 기초하여 지하매질의 분포를 예측하는 역산기법은 대수적 역산에 기초한 방법론과 통계적 추론(혹은 기계학습)에 기초하는 방법으로 나눌 수 있다. 일반적으로 수리현상에 관계하는 지하 물성에 대한 특성화 목적으로 실시하는 대수적 역산 기법은 대수적 유체거동 모델과 지구통계 모델 또는 대수적 유체거동 모델과 관찰 및 모델예측 사이의 공분산 관계를 베이저안 프레임워크(Bayesian framework) 하에서 중합하여 역산 해를 얻어낸다. 통계적 추론 기법에 기초하는 역산기법은 일반적으로 매질 분포 확률인 사전확률과 매질분포에 기초한 측정 자료의 확률을 의미하는 우도함수를 이용하여 사후확률을 반복적 과정을 통하여 얻어내는 몬테카를로 마르코프 연쇄(Monte Carlo Markov chain, MCMC)에 기초한다.

이들 기법을 이용하는데 있어 중요한 이슈로는 (1) 역산 모델이 일반적으로 요구하는 많은 정향모델링(forward

modeling)에 기인하는 연산 효율성과 (2) 역산모델의 제한인자(constraint)로 이용되는 필요 입력 파라미터의 효율성을 들 수 있다. 일반적으로 역산모델은 반복적 연산을 통하여 해를 얻어내며, 이러한 반복적 연산과정에는 많은 수의 정향모델링이 필요하다. 일반적으로 선형성이 높고 규모가 크지 않은 도메인의 정향모델링 경우, 연산에 소요되는 전산 자원이 적으나 다중유체 거동과 같이 비선형성이 높거나 도메인의 크기가 큰 도메인의 경우 많은 전산 자원을 필요로 하게 된다. 따라서 작게는 수 백에서 수 십만번에 걸친 반복 정향모델링을 실시할 경우 연산에 소요되는 시간은 역산 모델의 비효율성을 초래하게 된다.

역산모델 입력 파라미터의 경우 실제 야외에서 얻기 매우 어려운 특성이 있다. 특히 지구통계 파라미터의 경우 앞서 언급하였던 비정규성으로 인하여 야외에서 얻어지는 값에 대한 신뢰성을 얼마나 두어야 할지 결정하기 어려우며 얻고자 하는 수리적 물성치의 경우에도 그 범위를 결정하기 매우 어렵다.

본 연구에서는 지구통계적 예측의 한계를 극복하고자 하였으며 연구 대상으로 하는 도메인 규모에서 발생하는 수리현상에 기초하여 도메인 내 지중의 수리적 매질특성을 특성화하는데 목표가 있다. 따라서 도메인 내에서 실시하는 전규모적 수리시험에 따른 지하수위 변화 등이 지중 수리적 파라미터의 제한요소가 될 수 있다. 또한 본 연구에서는 대상으로 하는 도메인에 대하여 매우 적은 정보만이 가용하다고 가정하고 연구를 진행하였다. 실질적으로 지하의 물성분포를 규명하고자 하는 프로젝트의 경우, 시작 당시 지하 물성분포에 대한 어떤 가정도 예측 결과를 편향되게 할 가능성이 크며 이에 따라 초기 가정을 최소화하여 예측을 실시하는 것이 바람직하다. 실제 지구통계 분야에서 다중포인트 공간통계(multi-point statistics) 기법이 종종 사용되고 있으나, 이러한 기법의 적용을 위해서는 도메인의 지질분포 및 구조에 대한 정보를 매우 상세하게 알고 있음을 가정하며 이에 상응하는 트레이닝 이미지를 이용하여 예측을 시행하게 된다(Strebelle, 2000, 2002; Deutsch and Tran, 2002; Hu and Chugunova, 2008). 이러한 경우 예측결과는 트레이닝 이미지에 종속되게 되어 예측의 편향성을 증대시킬 수 있으며 충분한 정보가 부재할 경우 실제 도메인의 물성분포 특성과 전혀 다른 예측을 하게 된다는 한계성이 있다.

최소한의 정보만을 이용하여 예측을 수행하는 경우 활용될 수 있는 최첨단 기법(state of the art technique)으로 메트로폴리스 알고리즘(Metropolis algorithm)이 고려될 수 있다(Efendiev et al., 2009; Fu and Gomez-Hernandez,

2009; Mariethoz et al., 2010). 해당 기법은 현장에서 실시된 수리시험에 근거한 각 관측공에서의 압력 변화에 기초하여 매질 물성분포를 예측하며 이를 히스토리 매칭 (history matching)이라 한다(Kitanidis, 1995; Chen and Zhang, 2006; Cardiff and Kitanidis, 2009; Murakami et al., 2010; Rubin et al., 2010; Zhou et al., 2012; Kitanidis, 2015). 메트로폴리스 알고리즘은 예측을 위하여 예측하고자 하는 값의 포인트 측정치가 반드시 필요하지는 않으며 도메인 전반의 공간 통계적 정보 역시 반드시 요구되지 않는다. 그러나 현장 측정치나 공간 통계 정보가 가용할 경우 이를 활용할 수 있으며 이러한 정보들이 보다 많이 이용되었을 경우 예측 결과가 광역 최저치 (global minima)를 찾아낼 가능성이 높아진다. 그러나 본 연구에서는 현장의 정보가 매우 제한적이라는 현실적인 상황을 고려하여 도메인 내에서 실시된 양수시험 및 각 관측점에서 측정되어진 시간에 따른 수위변화, 즉 히스토리 커브만이 모델의 입력인자로 이용되었다. 따라서 예측 결과는 경우에 따라 국소 최저치(local minima)나 광역 최저치를 찾게 되며 이에 따른 기법의 견고성(robustness)에 대해서는 예측 결과에 따라 추후 논의되게 될 것이다.

전반적으로 본 연구에서 다루게 될 내용은 다음과 같다. 양역법(explicit method)인 메트로폴리스 알고리즘에 대하여 이론적 배경을 소개하고 본 연구에서 제안하는 독창적 메트로폴리스 알고리즘 개발 과정에 대한 내용을 다룰 것이다. 또한 양역법에 의한 예측을 가상의 예를 통하여 기법의 장점 및 단점에 대하여 분석할 예정이다.

2. 메트로폴리스 알고리즘

MCMC란 마르코프 연쇄에 기초하는 추계론적 샘플링 기법으로 반복적 모사의 입력인자인 모델 파라미터가 마르코프 연쇄과정을 따라 재현되는 몬테카를로 샘플링을 의미한다. 메트로폴리스 알고리즘은 랜덤워크에 기초하는 MCMC 기법의 일종으로 확률분포를 직접 샘플링하기 어려운 경우 추계론적 샘플링을 통하여 실제 확률분포를 근사하는 방법이며, 특히 일반적인 샘플링 기법(i.e. important sampling 및 rejection sampling)과는 달리 고차원의 파라미터 공간에서 효율적으로 적용될 수 있다는 장점을 지니고 있다(Oliver et al., 1997). 메트로폴리스 알고리즘은 최초 Metropolis et al.(1953)에 의해 제안되었으며 강력한 확률분포 근사 및 최적화 문제 해결능으로 인하여 매우 넓은 분야에서 다양하게 활용되고 있다.

메트로폴리스 알고리즘을 통한 최적화는 종종 Fig. 1과

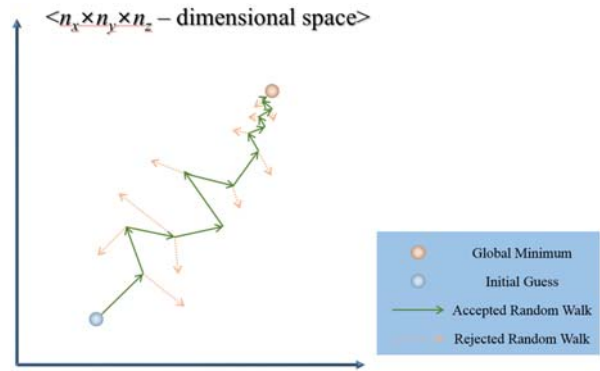


Fig. 1. Schematic diagram of optimization through Metropolis algorithm.

같은 개념도로 설명된다. 우선 메트로폴리스 알고리즘을 통한 최적화는 일반적으로 매우 높은 차원에서 고려되며 대표적으로 격자망이 이용되는 수치해석을 통한 모델 파라미터 최적화 문제에서 빈번하게 사용된다. 만약, 3차원 직교좌표계 공간에서 격자망이 $n_x \times n_y \times n_z$ 개로 이루어져 있다고 할 때, 각 격자에서의 상태값(e.g. 수리전도도)은 타 격자의 값과 독립적으로 존재한다고 할 수 있으며, 따라서, 비록 공간적으로는 3차원에 해당하나 파라미터 공간에서는 $n_x \times n_y \times n_z$ 차원으로 생각할 수 있으며 Fig. 1은 $n_x \times n_y \times n_z$ 차원 공간상에서의 랜덤워크를 보여주고 있다. 하나의 상태에서 다른 상태로의 랜덤워크는 채택 (accept) 또는 기각(reject) 될 수 있으며 랜덤워크의 채택과 기각에는 두 상태 간의 우도함수비가 이용된다. 하나의 상태에서 다른 상태로의 랜덤워크에 우도함수가 이용된다는 관점에서 메트로폴리스 알고리즘은 전형적인 베이저안 분석기법이라 할 수 있으며 이를 수식으로 표현하면 다음과 같다.

$$p(\mathbf{x}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{d})} \Rightarrow p(\mathbf{x}|\mathbf{d}) \approx p(\mathbf{d}|\mathbf{x})p(\mathbf{x}) \quad (1)$$

여기서 \mathbf{x} 는 $n_x \times n_y \times n_z$ 차원의 상태벡터를 의미하며 \mathbf{d} 는 $n_{obs} \times n_t$ 차원의 관찰행렬을 의미한다. 위의 식 (1)은 관찰에 기초한 상태의 사후확률은 상태에 기초한 사전확률과 상태를 상정한 관찰의 우도 곱으로 표현할 수 있다는 것으로 해석할 수 있다. 여기서 관찰이란 n_{obs} 개의 관측공에서 n_t 개의 유한한 시간 동안 측정된 히스토리 자료를 의미한다. 또한 마르코프 프로세스(Markov process)에 의해

$$\lim_{n \rightarrow \infty} p(X_n = \mathbf{x} | X_0 = \mathbf{x}_0) = \pi(\mathbf{x}) \quad (2)$$

와 같이 쓸 수 있으며, 즉 초기 예측 \mathbf{x}_0 와는 상관없이 n 이 무한대에 도달할 경우 \mathbf{x} 로의 n 스텝 전이확률은 상태 벡터 \mathbf{x} 의 주변확률($\pi(\mathbf{x})$, marginal probability)에 도달하

게 된다. 따라서 MCMC 기법에 의해 초기 상태벡터로부터 우도에 기초한 사후 확률을 얻고 사후 확률을 사전확률로 하여 이러한 과정을 반복적으로 실시할 경우 최적해 인근에 도달할 수 있다.

그러나 지중 물성분포를 예측하기 위한 문제에 완전히 무작위적인 랜덤워크를 이용하여 메트로폴리스 알고리즘을 시행할 경우 최적해에 매우 느리게 수렴되며 그 결과 역시 국소 최저치(local minima)에 도달할 위험성이 있다. 이에 따라 무작위적인 랜덤워크가 아닌 공간통계적 상관성에 기초한 전이 커널 함수를 이용하여 랜덤워크를 발생시키는 방안이 고려할 수 있다. 여기서 이용되는 공간통계적 상관성은 실제 필드가 지닌 공간통계적 상관성을 이용하는 것이 가장 바람직하나 실제 필드의 공간통계 정보를 모른다 하더라도 임의적 설정을 통하여 적용 가능하다. 다만 공간통계 특성에 대해서는 가정이 필요하며, 본 연구에서는 연속적 지하물성이 가우시안 과정(Gaussian process)를 따른다는 기존 연구의 일반적 가정을 이용하였다(Kitanidis, 1995).

만약 지하물성을 예측하기 위한 메트로폴리스 알고리즘의 랜덤워크가 공간적 상관성에 기초한 가우시안 과정을 따른다고 가정할 경우 이를 설명할 수 있는 공간통계 모델이 요구된다. 이에 는 대표적으로 일반적인 지구통계 예측에서 활용되는 공간 공분산(spatial covariance) 모델 혹은 세미-베리오그램(semi-variogram) 모델이 있다. 세미-베리언스는 수학적으로

$$\gamma(\mathbf{h}) = \frac{1}{2} E[(X(\mathbf{u}+\mathbf{h}) - X(\mathbf{u}))^2] \quad (3)$$

과 같이 쓸 수 있으며 여기서 \mathbf{u} 는 위치벡터 그리고 \mathbf{h} 는 래그벡터를 의미한다. 일반적인 세미-베리오그램 모델은 일정 래그 범위 내에서 점차적으로 증가하는 경향을 갖는다. 이는 실제 지중 매질들의 공간적 상관성을 지시하며, 상관거리(correlation length)라 불리는 일정 거리 내에서는 상관성이 거리에 따라 점차 감소하고 상관거리 이상에서는 더 이상 상관성이 존재하지 않음을 의미한다.

세미-베리오그램 모델을 랜덤워크에 활용할 경우 프로포절 상태 발생을 위하여 각 단계 마다 추계론적 지구통계 모사를 실시하고 이를 해당 단계 상태벡터에 더하여 주는 방법이 있다. 이러한 방법은 기존에 개발된 지구통계 모사 소프트웨어를 활용할 수 있다는 점에서 매우 손쉽고 효과적인 방법이다. 그러나 이 같은 방법들은 공간적으로 특정 위치에서의 상태를 예측목표 상태에 가깝게 만들어 랜덤워크가 점차 최적치로 다가가도록 만드는 동

시에 특정 위치 이외의 위치에서는 반대로 상태가 예측 목표와 멀어지도록 만들 수 있다는 복잡성이 있어 전반적인 알고리즘의 수렴 속도를 늦출 수 있다. 이러한 문제점을 보완함과 동시에 공간적 상관성이 가우시안 과정을 따른다는 가정을 보존하기 위하여 전이 커널 함수로 공간 영역에서의 RBF(Radial Basis Function)를 이용할 수 있다. RBF는 상대적 거리(i.e. Euclidean norm)에만 기초하는 실수함수이며, 본 연구에서는 RBF가 공간 통계적인 특성을 반영할 수 있도록 공간 공분산이 반영된 마할라노비스(Mahalanobis) 거리를 가우시안 RBF의 거리값으로 활용할 수 있도록 다차원 공간 상에서의 섭동(perturbation)을 설명하는 마할라노비스-가우시안 RBF(식 (4)) 및 마할라노비스-지수 RBF(식 (5))를 다음과 같이 구성하였다.

$$\Delta \mathbf{x} = \exp[-(\mathbf{u} - \mathbf{u}_0)^T \Sigma^{-1} (\mathbf{u} - \mathbf{u}_0)] \quad (4)$$

$$\Delta \mathbf{x} = \exp\left[-\sqrt{(\mathbf{u} - \mathbf{u}_0)^T \Sigma^{-1} (\mathbf{u} - \mathbf{u}_0)}\right] \quad (5)$$

$$\Sigma = \mathbf{R}^T \Sigma_0 \mathbf{R} \quad (6)$$

$$\Sigma_0 = \begin{bmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & \lambda_z \end{bmatrix} \begin{bmatrix} \lambda_x & 0 & 0 \\ 0 & \lambda_y & 0 \\ 0 & 0 & \lambda_z \end{bmatrix} \quad (7)$$

$$\mathbf{R} = \mathbf{R}_x \mathbf{R}_y \mathbf{R}_z \quad (8)$$

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix}, \quad \mathbf{R}_y = \begin{bmatrix} \cos \theta_y & 0 & -\sin \theta_y \\ 0 & 1 & 0 \\ \sin \theta_y & 0 & \cos \theta_y \end{bmatrix},$$

$$\mathbf{R}_z = \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

여기서 $\Delta \mathbf{x}$ 는 랜덤워크에 의한 상태벡터의 전이(transition)를 의미하며, \mathbf{u}_0 와 \mathbf{u} 는 각각 원점 및 각 격자의 위치벡터, λ_x , λ_y , 및 λ_z 는 각각 x , y , 및 z 축을 따른 상관거리, θ_x , θ_y , 및 θ_z 는 각 축의 회전각을 의미한다. 따라서 각 랜덤워크 마다 \mathbf{u}_0 의 위치가 도메인 내에서 무작위로 결정되며 이에 의해 생성된 $\mathbf{x} + \Delta \mathbf{z}$ 가 프로포절 상태벡터로 제시된다. 이와 같이 전이 커널 함수를 구성할 경우 앞서 언급한 추계론적 지구통계 모사에 의한 프로포절 상태벡터에 비하여 불필요한 상태변화가 나타나지 않음으로 인하여 보다 최적해 인근에 빠르게 수렴될 수 있으며 추가적으로 프로포절 상태벡터의 연산에 소요되는 전산자원 역시 매우 적다.

각 단계에서 랜덤워크의 크기를 일정하게 할 경우 그 크기가 최적해 인근에서 최적해를 지나치지 않도록 작게 유지하여야 한다. 그러나 이와 같이 할 경우 너무 작은 랜덤워크로 인하여 최적해 수렴에 매우 많은 연산이 소요된다. 이에 따라 랜덤워크 크기를 최적해에 다가갈수록 점차 축소하는 방법론이 필요하며 담금질 기법(simulated annealing technique)의 냉각속도(cooling schedule)가 이러한 목적을 위하여 이용될 수 있다. 담금질 기법은 통계적 메타휴리스틱(metaheuristic) 기법의 일종으로 목적함수(objective function) 및 냉각속도에 의해 최적해 수렴이 결정된다. 여기서 냉각속도란 매 단계별 변화량을 의미하며 상태가 최적해에 다가갈수록 에너지의 감소를 가정하게 되며 볼츠만 분포(Boltzmann distribution)와 유사한 형태를 따라 점차 변화가 줄어들게 된다(Mariethoz et al., 2010).

지하물성 분포는 다중 규모의 공간통계 모델로 설명될 수 있다. 이는 규모가 서로 다른 다양한 지질현상에 의해 지하 매질분포가 결정됨을 의미한다. 이러한 지질현상에 의한 공간적 상관구조를 해석하기 위해서는 다중 공간구조 모델(multi-structural spatial model)이 랜덤워크의 전이 커널로 이용되어야 하며 세미-베리오그램 모델로 이러한 문제에 접근할 경우 다양한 공간구조 규모에 상응하는 추계론적 모사를 재현하여 이용하여야 하나 현존하는 대부분의 추계론적 모델의 비정규성 해석에는 한계가 있으며 이에 따라 다소간의 복잡성이 있다. 전이 커널로 마할라노비스-가우시안 RBF를 사용할 경우 다양한 지향적 공간구조의 재현이 손쉬우며, 각 재현은 한정된 공간에서만 정규성을 가지도록 구성할 수 있어 비정규성 문제를 보다 효율적으로 다룰 수 있다. 유한한 다중 규모의 공간구조로 해석된 지하 물성분포를 수학적으로 표현하면 다음과 같다.

$$A = \sum_{i=1}^N A_i = \sum_{i=1}^N \sum_{j=1}^{M(i)} \omega_j^{(i)} \phi(\|\mathbf{x} - \mathbf{x}_j^{(i)}\|; \Sigma_j^{(i)}) \quad (10)$$

여기서 N 은 서로 다른 규모의 공간구조 개수, M 은 각 규모의 공간구조를 만들기 위하여 필요한 RBF 재현 개수, ω_j 는 각 RBF 재현의 가중치(weight), ϕ 는 마할라노비스-가우시안 RBF, 및 $\Sigma_j^{(i)}$ 는 마할라노비스-가우시안 RBF에 이용되는 공분산행렬을 의미한다.

전이 커널로 마할라노비스-가우시안 RBF를 사용하는 추가적인 장점으로는 공간구조의 규모 및 지향성을 손쉽게 조절할 수 있다는 점이다. 위의 식 (10)에서 주어지는 바와 같이 재현마다 공분산 행렬(Σ)을 구성하는 상관거리

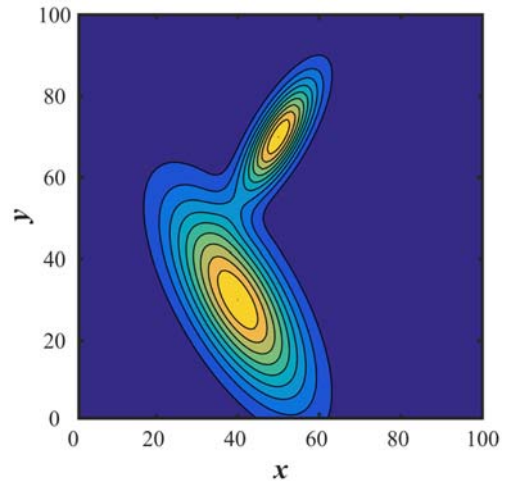


Fig. 2. A non-stationary spatial structure composed by two different Mahalanobis-Gaussian RBF realizations.

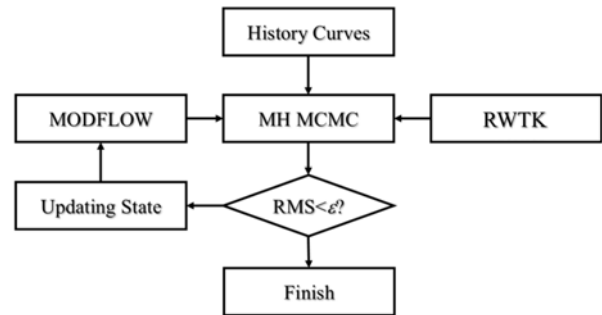


Fig. 3. Computational sequences used for subsurface hydraulic properties characterization based on history matching through Metropolis algorithm.

행렬(Σ_0) 및 회전행렬(\mathbf{R})을 조절함으로써 공간 구조의 형태를 쉽게 바꿀 수 있다. Fig. 2는 하나의 도메인에서 두 개의 서로 다른 공간구조를 갖는 RBF에 의해 구성된 재현을 보여주고 있다. 그림에서와 같이 마할라노비스-가우시안 RBF를 공간구조 구성에 이용할 경우 각 RBF 재현 시 공간구조를 서로 달리함으로써 도메인 전반의 비정규성을 쉽게 표현할 수 있다.

MCMC 히스토리 매칭을 이용한 지하물성 분포 예측의 개략적인 알고리즘을 도시화 하면 Fig. 3과 같다. 메트로폴리스 알고리즘을 이용한 히스토리 매칭은 MATLAB 코드로 구성하였으며 외부 프로그램인 MODFLOW2000 (Hill et al., 2000)을 연계하였다. 개발된 코드는 모듈화 되어있으므로 전이 커널로 사용되는 지구통계 모델 및 유체흐름 수치모델은 타 모델로 용이하게 교체될 수 있으며 이에 따라 다양한 유체의 지하흐름에 따른 물성예측이 가능하도록 하였다.

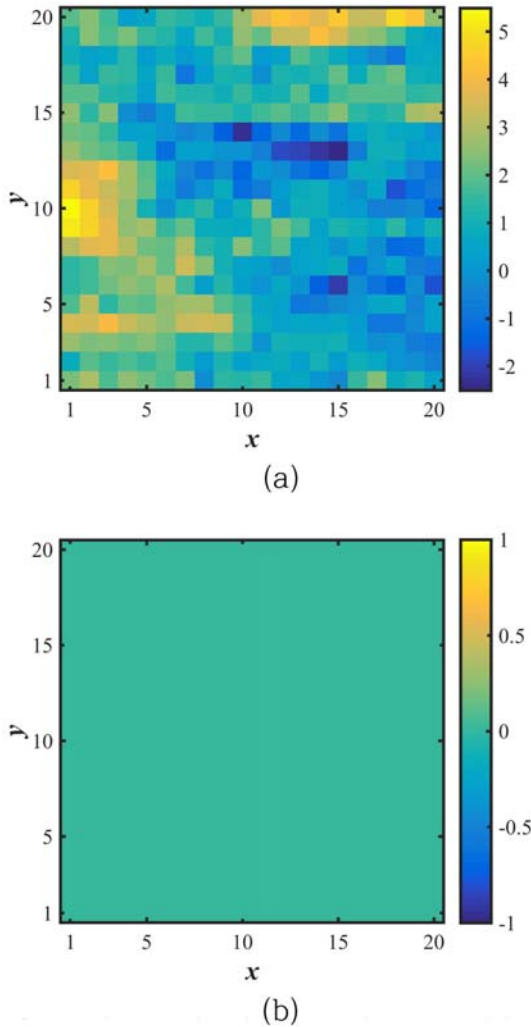


Fig. 4. (a) Synthetic logarithmic hydraulic conductivity domain and (b) initially guessed logarithmic hydraulic conductivity domain.

3. 실험 방법

3.1. 가상의 대수층 및 양수시험 설정

양역법에 의한 예측 결과를 가상의 예를 통하여 검증하기 위하여 x 및 y -방향으로 각각 20개의 격자($n_x = n_y = 20$; $\Delta x = \Delta y = 100$ m)로 차분된 총 400개 격자의 2차원 가상 도메인(2,000 m \times 2,000 m)을 아래의 Fig. 4a와 같이 설정하였다. 가상 도메인은 순차가우시안 지구통계 모사(sGsim, sequential Gaussian simulation; Deutsch and Journel, 1998)를 통해 획득하였으며 수리 전도도에 대한 기존 참조정보는 이용하지 않고 해당 공간에 대한 지구통계 특성만을 고려한 비조건부 모사(unconditional simulation)를 통하여 재현하였다. 본 연구에서는 작은 규모의 도메인(20 \times 20)을 설정함으로써 sGsim 기법을 통한 재현이

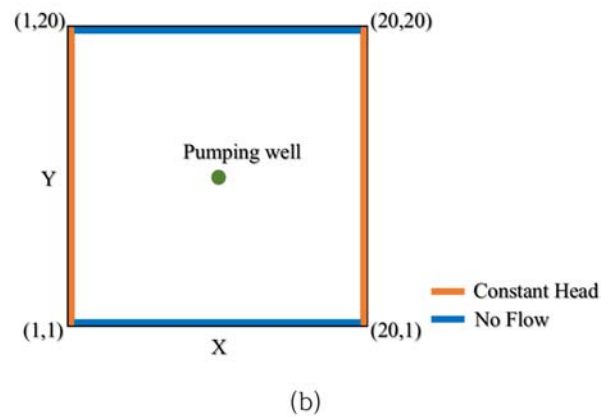
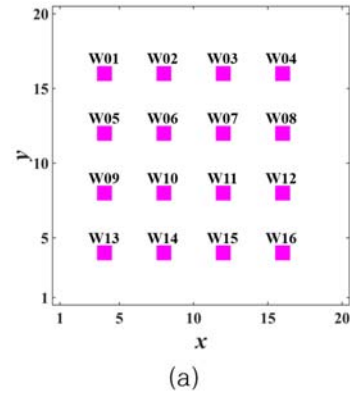


Fig. 5. (a) Location of 16 hydraulic head measurements for adjustment of predicted hydraulic parameters and (b) boundary condition used to simulate groundwater flow through the MODFLOW-2000.

정규성을 따르기에는 충분하지 않다. 이는 물성 분포의 지향성이 도메인 내에서 위치에 따라 서로 다른 형태를 보여줄을 의미한다. 또한 매우 작은 크기의 도메인 활용을 통해 개발 기법의 예측능 및 안정성 검증에 요구되는 불필요한 전산자원의 소모를 최소화하고자 하였다.

베이지안 기법에 근거한 자료중합을 위해 사용된 지하수위(hydraulic head) 시계열 자료는 Fig. 5a에 나타난 바와 같이 총 16개의 지점에서 관측 할 수 있도록 설정하였다. 본 연구의 목적은 개발된 기법의 예측능 및 예측의 안정성을 검증하는데 목적이 있으며 Fig. 5a와 같은 조밀한 관측점 배치는 예측의 불확실성을 최소화 하면서 이러한 목표를 달성하기 위한 방법론의 하나로, 현실적으로 Fig. 5a와 같이 많은 수의 관측점을 대수층에 설치하여 시간-수위강하 곡선을 얻어내는 경우가 매우 드물다.

자료중합을 위해 이용되어질 시계열 수위자료를 가상의 도메인 및 예측된 다중재현들에 대하여 획득하기 위하여 부정류 지하수 유동 모사를 실시하였다. 16개 지점에서

실제로 가정된 시간-수위강하 곡선과 예측된 수리전도도 재현에 대한 시계열자료는 연속방정식(continuity equation) 과 달시 법칙(Darcy's law)을 기반으로 하는 지하수 유동 방정식을 시스템 모델($F(A)$)의 수치연산을 통하여 획득하였으며 이는 다음과 같이 표현 될 수 있다.

$$\nabla \cdot (\mathbf{K} \nabla h) = S_s \frac{\partial h}{\partial t} \quad (11)$$

여기서 \mathbf{K} 는 수리전도도 [LT^{-1}], h 는 지하수위 [L], S_s 는 비저류계수 [L^{-1}], t 는 시간 [T]을 의미한다. 모사를 위해 유한 차분법(finite-difference method)을 이용하여 지하수 거동을 모사하는 MODFLOW-2000이 이용되었다. 모사가 이루어질 도메인들은 모두 상하부가 불투수층으로 피압되어 있는 대수층임을 가정하였으며 수직적인 대수층 물성의 변화는 없는 것으로 가정하여 2차원 모사를 실시하였다. 도메인의 경계조건으로 첫 번째 열 및 마지막 열(즉, $X = 1$ 및 $X = 20$ 인 위치)에 해당하는 격자에는 수위가 200 m로 고정된 고정수두경계(constant head boundary)를 설정하였으며 첫 번째 행 및 마지막 행(즉, $Y = 1$ 및 $Y = 20$ 인 위치)의 격자는 무흐름경계(no flow boundary)로 설정하였다(Fig. 5b). 초기수위를 200 m로 설정한 후 도메인의 중앙인 (10,10) 격자에 위치하고 있는 양수정에서의 수리수두가 2일 동안 100 m로 유지되는 부정류 모사(transient flow simulation)를 실시하였으며, 이 때 해당 기간을 20단계로 동일한 크기로 차분하여 수위분포를 획득하였다.

3.2. 메트로폴리스 알고리즘을 이용한 예측의 개선

3.2.1. 순차 가우시안 모사를 전이커널로 한 메트로폴리스 알고리즘

메트로폴리스 알고리즘의 상관-랜덤워크(correlated-random-walk)를 발생시키는 전이커널(transition kernel)로 추계론적 가우시안 과정 및 세미-베리오그램을 선정하여 도메인 내 수리전도도의 공간적 분포 예측에 활용하였으며 예측 과정은 다음과 같다.

1. 실제 현장의 16개 관측공에서의 지하수위 시계열자료 획득(h^*)
2. 수리적 물성치 도메인에 대한 초기 예측 실시(\mathbf{x}_0)
 - a. 초기 예측 도메인에 대한 지하수 유동모사 실시 후 획득한 지하수위 시계열 자료와 실제 지하수위 시계열 자료 간의 평균제곱근 오차 계산
3. 가우시안 분포의 상관-랜덤워크 획득하고 해당 가중치(w , 식 (10))를 곱하여 $\Delta \mathbf{x}$ 를 산정

4. 3번의 결과를 이전 단계의 예측 물성치(\mathbf{x})에 추가하여 사후 예측 실시($\mathbf{x}^* = \mathbf{x} + \Delta \mathbf{x}$)
5. 4번에서의 사후 예측에 대한 지하수위 시계열 자료 획득($h^{(n)}$)
6. 메트로폴리스 법칙에 따라 사후 예측에 대한 선택/기각 판정을 위한 우도비($L(\mathbf{x}^*) / L(\mathbf{x})$) 계산(즉, $p(\mathbf{x} \rightarrow \mathbf{x}^*) = \min [1, L(\mathbf{x}^*) / L(\mathbf{x})]$)
7. 지정된 기준(허용오차 또는 랜덤워크 반복 횟수 기준)에 만족할 때 까지 3번에서 7번 반복

과정 3번에서 매 랜덤워크 단계에서 모든 차원의 파라미터를 섭동시키는 전이커널을 획득하기 위해 추계론적 비조건부 순차가우시안 지구통계 모사(i.e. sGsim)를 이용하였으며, 이는 과정 2에서 발생된 총 10개의 초기 예측 도메인에 각각 적용되었다. 이 때, 초기예측은 조건부 지구통계 모사를 이용하여 특정 공간 통계적 분포를 가지도록 획득 될 수 있을 뿐만 아니라 Fig. 4b와 같이 모든 격자에 대하여 동일한 값을 가지도록 설정될 수 있다. 각 초기예측에 대하여 예측의 개선은 16개의 관측공 지점에서의 예측된 지하수위 자료와 실제 지하수위 간의 평균제곱근 오차(root-mean-square deviation)가 허용 오차에 도달하거나 또는 총 1000회의 수치모델 수행이 이루어 질 때까지 반복 수행되었다. 예측의 개선이 이루어짐에 따라 실제 지하수위 시계열 자료와 예측 지하수위 시계열 자료 간의 평균제곱근 오차가 감소하고 점차 랜덤워크의 크기가 감소하게 되며, 이러한 감소는 최적해 인근에서의 랜덤워크가 최적해를 지나치지 않도록 하는 역할을 한다. 이러한 랜덤워크의 점진적인 축소는 담금질 기법(simulated annealing)의 냉각 스케줄(cooling schedule) 함수와 유사한 형태로 모델에 포함시킬 수 있으며, 본 연구에서는 랜덤워크의 크기를 감소시키는 인수(w)로 아래의 함수가 과정 3에 이용되었다.

$$w = a \times \exp\left(-k \times \frac{R_{initial} - R_{n-1}}{R_{initial}}\right) \quad (12)$$

여기서 a 와 k 는 각각 초기 랜덤워크의 크기와 담금질 기법의 냉각속도를 조절하기 위하여 경험적으로 결정되어지는 상수이며, 본 연구에서는 a 및 k 를 각각 1 및 10으로 설정하여 예측을 실시하였다. 위의 식에서 $R_{initial}$ 는 초기 수리전도도 예측을 통해 만들어진 지하수위 시계열 자료와 실제 지하수위 시계열 자료간의 평균제곱근 오차를 의미하며, R_{n-1} 은 $n-1$ 단계에서 예측된 수리전도도 분포를 이용하여 제작된 시계열 자료와 실제 관측 지하수위

간의 평균제곱근 오차를 의미한다. 평균제곱근 오차는 아래의 식을 통하여 계산된다.

$$R_n = \sqrt{\frac{1}{N_i} \sum_{i=1}^{N_{obs}} \sum_{j=1}^{N_i} (h_{ij}^* - h_{ij}^{(n)})^2} \quad (13)$$

여기서, N_{obs} 는 관측지점의 개수로 앞서 언급한 바와 같이 총 16개의 관측지점(Fig. 5a)을 활용하였으며, N_i 는 시스템 모델에서의 시간단계의 개수로 20개 시간단계가 이용되었고, h^* 와 $h^{(n)}$ 은 각각 실제 각 관측점에서 얻어진 지하수위자료 및 n 단계 수리전도도 예측을 통하여 제작된 각 관측점에서의 시간-수위강하 자료를 의미한다.

본 연구의 과정 6번에서 이용된 우도함수 비는 다음과 같이 계산되었으며

$$\frac{L(\mathbf{x}^*)}{L(\mathbf{x})} = \exp\left[-\frac{\mathbf{R}(\mathbf{x}^*) - \mathbf{R}(\mathbf{x})}{\gamma}\right] \quad (14)$$

여기서 적절한 γ 의 설정은 최적해로의 수렴을 용이하게 하며, 해당 기법은 지구통계 모델을 이용하여 랜덤워크를 발생시키기 때문에 이용된 지구통계 모델의 예측 성능에 영향을 받는 단점이 존재한다.

3.2.2. 방사상 기반함수를 전이커널로 한 메트로폴리스 알고리즘

다음으로 방사상 기반함수(radial basis function, RBF)를 메트로폴리스 알고리즘의 랜덤워크 전이커널로 사용하는 히스토리 매칭 기법을 통하여 수리전도도 분포를 예측하였으며 이의 과정은 다음과 같다.

1. 실제 현장의 16개 관측공에서의 지하수위 시계열자료 획득(h^*)
2. 이용될 방사상 기반함수의 개수 결정
 - a. 공분산 행렬(식 (7)) 및 예측구조의 회전형렬(식 (9)) 결정
3. 수리적 물성치 도메인에 대한 초기 예측 실시(\mathbf{x}_0)
 - a. 초기 예측 도메인에 대한 지하수 유동모사 실시 후 획득한 지하수위 시계열 자료와 실제 지하수위 시계열 자료 간의 평균제곱근 오차 계산
4. 방사상 기반 함수의 중심좌표 선정(즉, \mathbf{u}_0)
5. 방사상 기반 함수를 이용한 상관-랜덤워크 획득하고 해당 가중치(w , 식 (10))를 곱하여 $\Delta\mathbf{x}$ 를 산정
6. 5번의 결과를 이전 단계의 예측 물성치(\mathbf{x})에 추가하여 사후 예측 실시($\mathbf{x}_p^* = \mathbf{x} + \Delta\mathbf{x}$ 또는 $\mathbf{x}_n^* = \mathbf{x} - \Delta\mathbf{x}$)

7. 6번에서의 사후 예측에 대한 지하수위 시계열 자료 획득($h_p^{(n)}$ 와 $h_n^{(n)}$)
8. 우도비($L(\mathbf{x}_p^*)/L(\mathbf{x})$ 와 $L(\mathbf{x}_n^*)/L(\mathbf{x})$) 계산 및 보다 좋은 예측능의 사후 예측($L(\mathbf{x}^*)$) 선택(즉, $L(\mathbf{x}_p^*) > L(\mathbf{x}_n^*) \rightarrow L(\mathbf{x}^*) = L(\mathbf{x}_p^*)$; $L(\mathbf{x}_p^*) < L(\mathbf{x}_n^*) \rightarrow L(\mathbf{x}^*) = L(\mathbf{x}_n^*)$)
9. 메트로폴리스 법칙에 따라 8번에서 결정된 사후 예측에 대한 선택/기각 판정($p(\mathbf{x} \rightarrow \mathbf{x}^*) = \min [1, L(\mathbf{x}^*)/L(\mathbf{x})]$)
10. 지정된 기준(허용오차 또는 랜덤워크 반복 횟수 기준)에 만족할 때 까지 4번에서 9번 반복

본 연구에서는 과정 2번에서 이용될 방사상 기반함수로 아래와 같이 총 5가지 종류의 공간 공분산 값을 가지는 것으로 설정하였으며 이들을 단계적 및 순환적으로 하나씩 이용하여 매 단계마다 갱신을 실시하였다. 본 연구에서 예측을 위하여 이용된 총 5 가지의 방사상 기반함수는 각각 다음의 공분산을 이용한다.

$$\begin{aligned} \Sigma_1 &= \begin{bmatrix} \sqrt{10}^2 & 0 \\ 0 & \sqrt{4}^2 \end{bmatrix}, & \Sigma_2 &= \begin{bmatrix} \sqrt{5}^2 & 0 \\ 0 & \sqrt{2}^2 \end{bmatrix}, & \Sigma_3 &= \begin{bmatrix} \sqrt{2.5}^2 & 0 \\ 0 & \sqrt{1}^2 \end{bmatrix}, \\ \Sigma_4 &= \begin{bmatrix} \sqrt{1.25}^2 & 0 \\ 0 & \sqrt{0.5}^2 \end{bmatrix}, & \Sigma_5 &= \begin{bmatrix} \sqrt{0.2}^2 & 0 \\ 0 & \sqrt{0.2}^2 \end{bmatrix} \end{aligned} \quad (15)$$

여기서 x 및 y 방향을 따른 상관규모의 비는 $\Sigma_1, \Sigma_2, \Sigma_3$ 및 Σ_4 에 대하여 본 연구에서 가정된 도메인의 공간적 상관거리 이방성을 고려하여 각각 5:2의 비를 유지하도록 설정하였으며 Σ_5 에 대해서는 규모가 작아질수록 점차 지향성이 미약해질 것으로 가정하여 1:1의 상관규모를 이용하였다. 이와 같이 다양한 공간적 공분산 구조를 활용하는 이유에 대해서는 실제 수리전도도의 공간적인 분포가 하나의 공간통계에 의해서는 설명되기 어려우며 서로 다른 규모를 가진 공간구조의 조합으로 설명하여야 보다 합리적인 예측이 이루어질 수 있기 때문이다. 공간구조의 종류는 비교적 쉽게 보다 다양화 할 수 있으나 본 연구에서는 총 5개의 공간구조 모델로도 전체 예측이 충분히 이루어질 수 있을 것으로 가정하고 예측을 실시하였다. 이와 같은 다양한 크기의 공분산을 이용하여 무작위로 선택된 위치(\mathbf{u}_0)에서 마할라노비스-가우시안 함수를 발생시키고 이를 이전 단계에서 예측된 수리전도도 도메인에 더하여 예측의 섭동을 발생시킨 후, 지하수유동 수치모사를 이

용한 관측지점에서의 시간-수위강하 곡선 예측을 통하여 프로포절 수리전도도 분포에 대한 상대우도를 계산하는 과정을 거친다. 그 결과에 따라 프로포절 수리전도도 분포가 더 높은 상대우도를 가질 경우 랜덤워크가 채택되며 그렇지 않을 경우 기각되는 메트로폴리스 알고리즘을 통한 최적화 과정을 따른다.

4. 결과 및 고찰

4.1. 메트로폴리스 알고리즘을 적용한 수리전도도 도메인의 예측

4.1.1. 순차 가우스 모사를 전이커널로 한 메트로폴리스 알고리즘

Fig. 6은 총 16개의 관측점에서 얻어진 실제 지하수위 시계열 자료 및 최종 예측된 지하수위 부정류 변화를 보여준다. 대부분의 관측점에서 두 히스토리커브가 유사한 양상으로 변화하는 것으로부터 순차 가우스 모사를 랜덤 워크 전이커널로 이용한 메트로폴리스 알고리즘이 효과적인 예측능을 보여주는 것으로 판단된다. 특히 양수정이 위치한 도메인의 중앙부의 관측공에서의 히스토리 커브가 매우 유사하게 나타나며 이를 중심으로 멀어질수록 다소간의 오차가 발생하는 것을 확인 할 수 있다.

총 10개의 단일재현 중 무작위로 선택한 4개의 단일 예측결과는 Fig. 7과 같다. 해당 그림을 통하여 각 예측

들의 전반적인 수리전도도 분포양상이 유사하며 특히, 양수정이 위치한 도메인 중앙부를 중심으로 분포하는 관측정 주변에서의 수리전도도 분포양상이 모든 단일 재현에서 보다 유사하게 나타나는 것을 확인 할 수 있다. 그러나 도메인의 경계부에서의 수리전도도 분포는 각 재현마다 상호 차이가 있으며 이는 양수에 의한 수두의 변화가 경계부에 미치는 영향이 미미하기 때문일 것으로 판단된다. 모든 관측정에서의 합산 평균제곱근 오차는 랜덤워크 반복에 따라 지수적인 감소 양상을 보여주며(Fig. 8) 임의로 선택된 4개의 단일 재현에서 모두 유사한 변화양상을 보여주고 있으며 랜덤워크가 1000회 반복 된 후의 단일 재현들의 평균 RMSE는 1.9450 m로 나타났다.

Fig. 9는 최종적으로 예측된 총 10개의 수리전도도 도메인의 평균적 분포를 보여준다. 실제 수리전도도 도메인(Fig. 4a)과 비교하였을 때 전반적인 분포의 고저는 유사한 양상을 보인다. 그러나 Fig. 9a의 분포의 경우 각 단일 재현 예측 결과들의 평균에 해당하기 때문에 평균화 효과(averaging effect)에 의해서 수리전도도 값의 분포 범위가 다소 협소하게 나타나며 구조적인 분포 역시 실제 구조와 달리 부드러운 형태로 나타난다. 실제 수리전도도 값과 예측된 평균 수리전도도 값들 간의 산점도(Fig. 9b)를 살펴보면 평균화 효과에 의한 분포 범위 협소 양상이 뚜렷이 나타나고 있으나 전반적으로 일대일 대응 라인에 모든 값들이 분포하고 있는 것을 확인 할 수 있다. 이

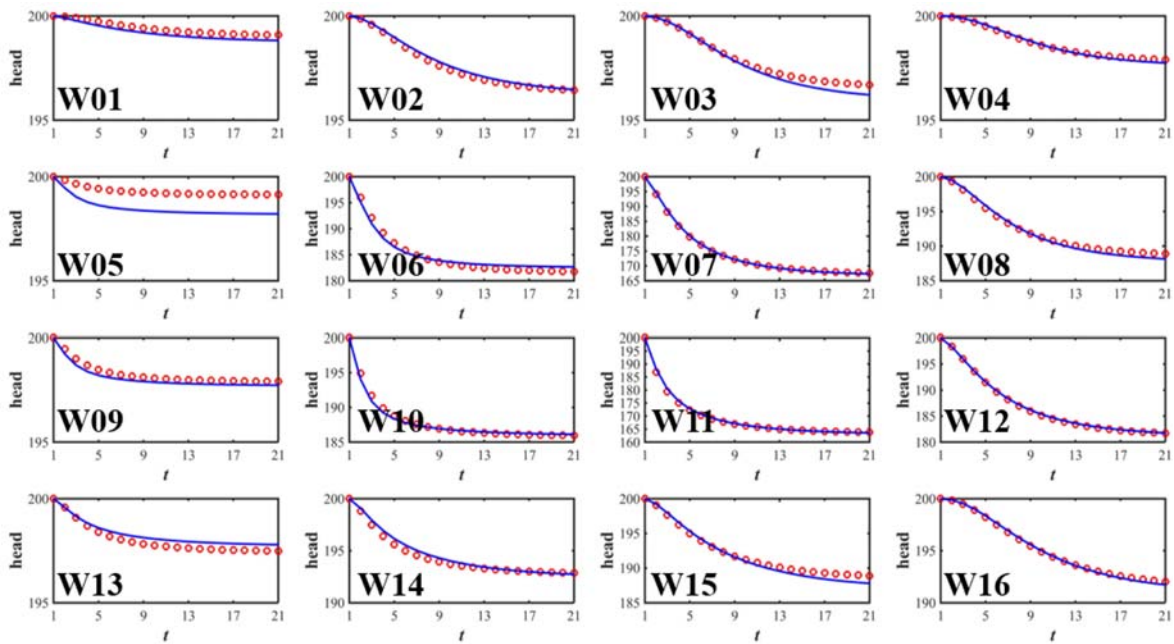


Fig. 6. Difference between true history curve (red dot) and predicted history curves (blue line) on each observed location in application of the Metropolis algorithm using uncondition simulation as random walk transition kernel.

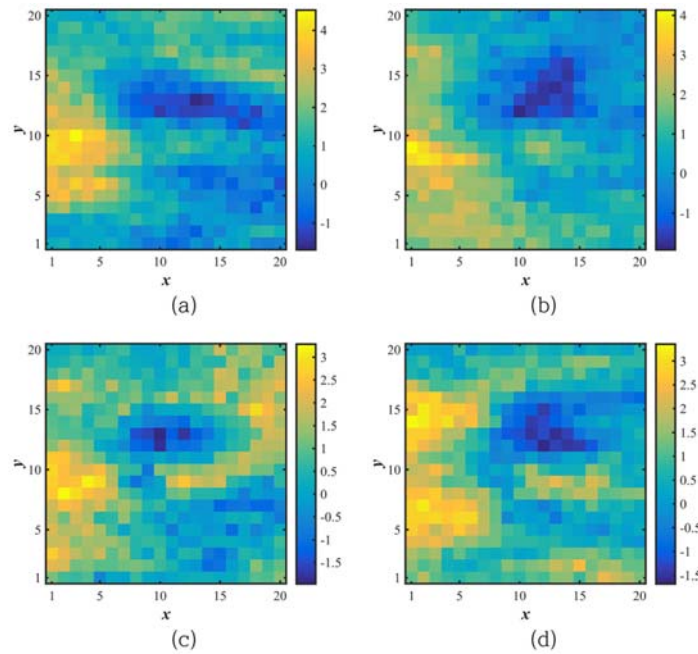


Fig. 7. Predicted single realizations obtained through Metropolis algorithm using uncondition simulation as random walk transition kernel.

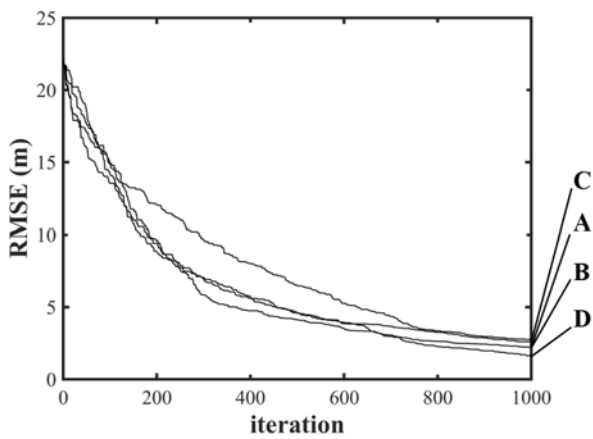


Fig. 8. Change of RMSE (root-mean-square deviation) between true history curves and predicted history curves obtained through Metropolis algorithm using uncondition simulation as random walk transition kernel.

두 값들 간의 상관계수는 0.6956으로 다소 높은 상관성을 보여주고 있으며 이를 통해 해당 알고리즘이 물성분포 예측에 효과적임을 확인 할 수 있다.

Fig. 4b에서 보는 바와 같이 초기 수리전도도를 예측함에 있어 어떤 정보도 이용되지 않았음에도 불구하고 이와 같이 실제와 유사한 예측 결과를 보여주는 것으로 보아 해당 기법은 초기 예측의 중요도가 낮은 것을 확인 할 수 있다. 이렇게 초기 정보의 부재를 가정하는 것은 실제 상황에 보다 부합하는 것으로 메트로폴리스 알고리즘에서

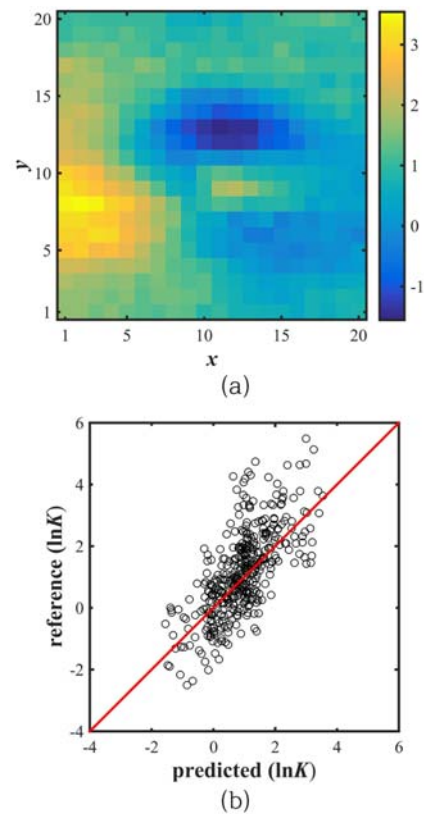


Fig. 9. (a) Predicted ensemble result obtained through Metropolis algorithm using uncondition simulation as random walk transition kernel and (b) a scatter plot between real and predicted logarithmic hydraulic conductivities (black dots) and one-to-one line (red line) whose correlation coefficient is 0.6956.

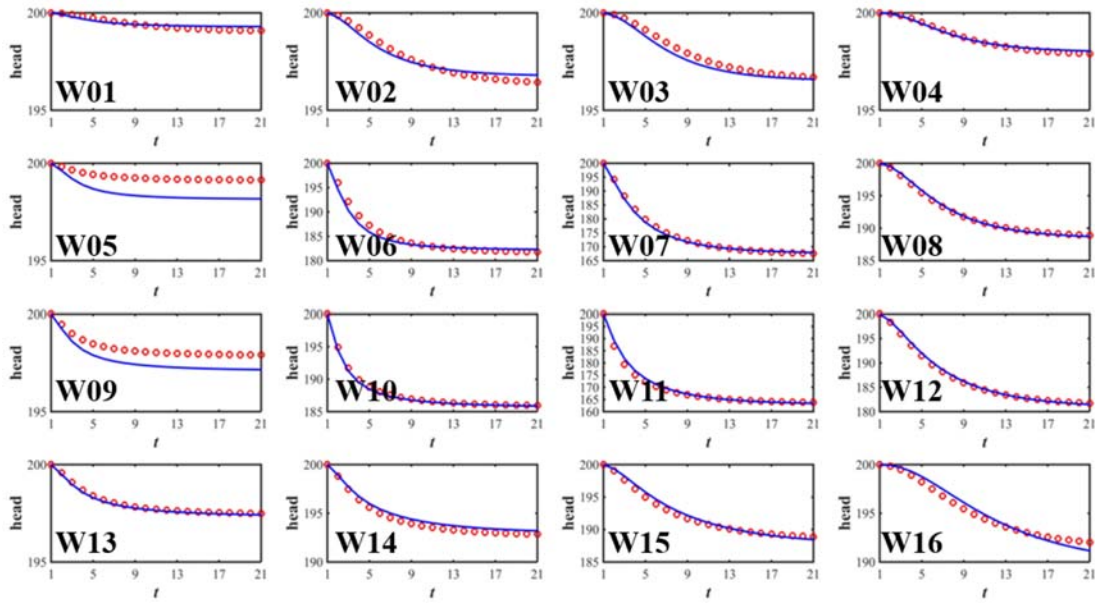


Fig. 10. Difference between true history curve (red dot) and predicted history curves (blue line) on each observed location in application of the Metropolis algorithm using radial basis function network as random walk transition kernel.

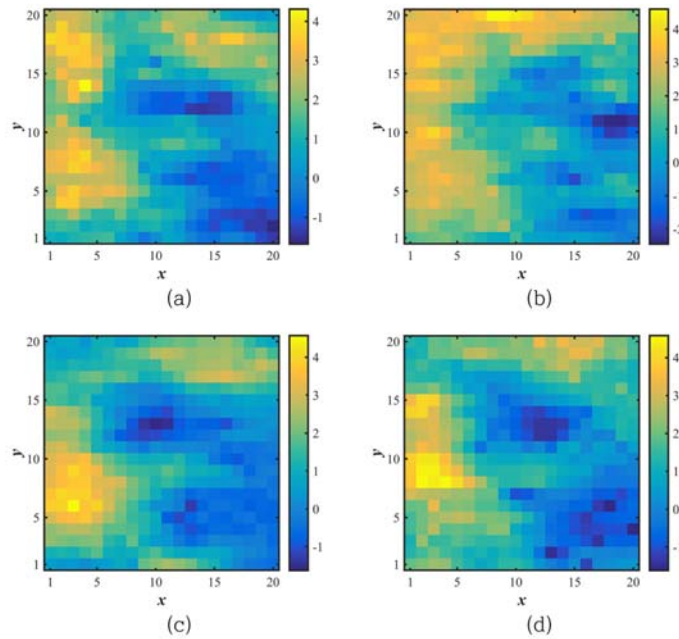


Fig. 11. Predicted single realizations obtained through Metropolis algorithm using radial basis function network as random walk transition kernel.

는 이러한 전처리 과정을 요구하지 않음으로 인하여 메트로폴리스 알고리즘을 이용한 예측이 효율적이라 할 수 있다.

4.1.2. 방사상 기반함수를 전이커널로 한 메트로폴리스 알고리즘

총 16개의 관측점에서 얻어진 실제 지하수위 시계열 자

료와 방사상 기반함수를 전이커널로 한 메트로폴리스 알고리즘에 의해 최종적으로 예측된 지하수위 부정류 변화는 Fig. 10과 같다. 앞선 순차 가우시 모사 전이커널을 이용한 예측결과와 동일하게 두 히스토리커브가 전반적으로 일치하는 양상을 보이며 특히 도메인의 중앙부의 관측공에서의 히스토리 커브가 매우 유사하게 나타나는 것을

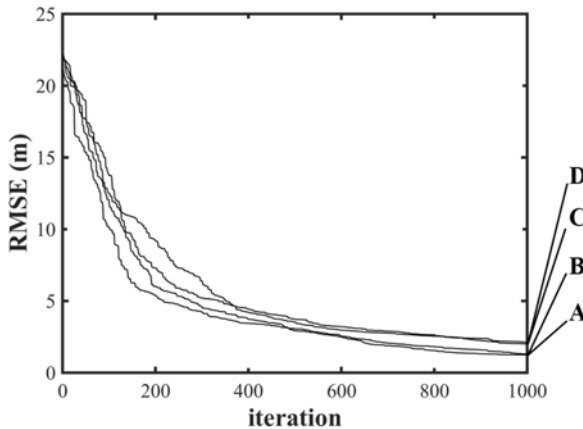


Fig. 12. Change of RMSE (root-mean-square deviation) between true history curves and predicted history curves obtained through Metropolis algorithm using radial basis function network as random walk transition kernel.

확인 할 수 있으며 이는 물성 예측에 있어 관측된 히스토리 커브가 유효하게 반영되었음을 의미한다.

Fig. 11은 총 10개의 단일재현 중 무작위로 선택한 4개의 단일 예측결과를 보여준다. 각 예측들은 앞선 순차 가우시안 모사를 전이커널로 한 예측 결과(Fig. 7)과 달리 모든 단일 재현들의 분포 범위 및 양상이 유사하게 나타나고 있다. 이는 예측 재현들이 이전 예측 결과 보다 실제 광역 최저치인 Fig. 4a에 보다 근접하게 예측되었음을 의미한다. 이러한 광역 최저치 근접성은 랜덤워크가 1000회 반복된 후의 단일재현들의 평균 RMSE가 1.6989 m로 이전 기법을 통해 산정된 평균 RMSE인 1.9450 m보다 낮음을 통해서도 확인 할 수 있다(Fig. 8과 Fig. 12). 또한 Fig. 8 및 Fig. 12에서 200 반복횟수 이하의 초기 변화 양상을 비교하였을 때 순차 가우시안 함수를 랜덤워크 전이커널로 이용한 메트로폴리스 알고리즘에 비하여 방사상 기반 함수를 이용한 경우가 더 빠른 감소율을 보임을 알 수 있다. 이는 전자 기법의 경우 이전 채택 수리전도도 분포에 전역적 섭동을 발생시킴으로써 도메인 전반에 대한 예측의 개선이 지역적 예측 저하를 수반할 수 있기 때문인 것으로 판단된다. 반면 방사상 기반함수를 이용한 예측은 전역적 섭동 함수 임에도 불구하고 지역적으로만 유효한 섭동을 발생시킨 후 예측개선이 이루어지기 때문에 지역적 예측저하의 개입 여지가 낮아 초기의 오차 감소율이 증가하는 것으로 판단된다.

최종적으로 예측된 총 10개의 수리전도도 도메인의 평균 분포는 Fig. 13a와 같다. Fig. 4a의 실제 수리전도도 분포 및 Fig. 9a의 순차 가우시안 전이커널을 이용한 예측 결과와 비교하였을 때 수리전도도 값들의 전반적인 분

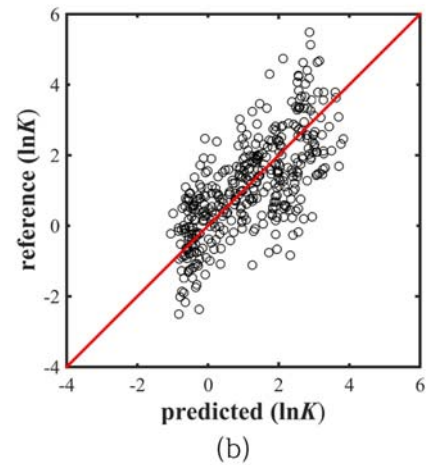
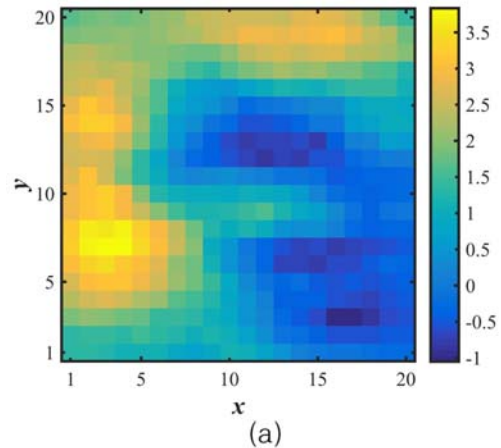


Fig. 13. (a) Predicted ensemble result obtained through Metropolis algorithm using radial basis function network as random walk transition kernel and (b) a scatter plot between real and predicted logarithmic hydraulic conductivities (black dots) and one-to-one line (red line) whose correlation coefficient is 0.7004.

포양상은 유사하게 나타난다. 실제 수리전도도 값과 예측된 평균 수리전도도 값들 간의 산점도를 도시한 결과(Fig. 13b) 전반적으로 일대일 대응 라인에 모든 값들이 분포하고 있으며 이들 간의 상관계수는 0.7004로 이전 예측 결과 보다 근소한 차이로 높은 경향을 보이거나 매우 유사함을 알 수 있다.

5. 결 론

본 연구에서는 수리시험을 통해 획득한 지하수위 변화 자료에 기초하여 도메인 내 수리적 매질분포를 특성화하는 방안으로 두 가지 메트로폴리스 알고리즘이 제안되었다. 이러한 베이지안 역산기법은 예측하고자 하는 도메인의 매질 정보가 완전히 부재하거나 충분하지 않을 경우에

도 지하수위자료만을 이용하여 예측을 수행할 수 있는 효과적인 기법으로 예측을 위해 가용할 자료가 매우 제한적인 실질적인 상황에 적용하여 지하물성 분포를 예측할 수 있다는데 의의가 있다.

본 연구에서는 메트로폴리스 알고리즘의 랜덤워크로 이용되는 전이커널로 순차 가우시안 모사 및 방사상 기반함수가 이용되었으며 이를 이용한 모델의 예측능을 2차원 가상 도메인을 통해 검증하였으며 이는 x 및 y -방향으로 각각 20개로 차분된 총 400개 격자의 수리물성 도메인이며 순차 가우시안 모사를 이용하여 설정되었다. 앞서 언급된 두 가지 전이커널을 이용하여 예측된 결과와의 비교 및 분석이 이루어졌으며 그 결과 두 가지 경우 모두 지하수위 자료가 추가정보로 이용됨에 따라 수리적 물성치 분포의 예측능이 단계적으로 개선됨을 확인할 수 있었다. 특히, 설정된 해당 도메인의 크기가 매우 작아 물성치의 전체적 분포 양상이 비정규적 특성을 보임에도 불구하고 제안된 기법을 이용한 예측능의 향상 효과가 뚜렷하게 나타남을 확인하였다. 특히, 방사상 기반함수를 랜덤 워크 전이커널로 이용하여 예측을 실시하는 경우 지역적 섭동만을 이용하여 예측을 단계적으로 개선시킴으로써 순차 가우시안 함수를 이용하는 경우 보다 광역 최저치로의 보다 빠른 수렴을 도모할 수 있는 것으로 나타났다. 또한 순차 가우시안 함수를 이용하는 메트로폴리스 알고리즘이 정규성이 강제된 지구통계모델을 이용하여 도메인 전반에 걸친 섭동을 발생시켜야 한다는 점에서 비정규적 물성 분포 특성을 예측하는데 제한점이 있을 것으로 판단된다. 그러나 방사상 기반 함수를 이용할 경우 이러한 지구통계모델이 포함하고 있는 한계성을 극복할 수 있을 뿐만 아니라 외부적인 지구통계모사과정을 요구하지 않음으로 인하여 전산자원이 효율적으로 활용될 수 있을 것으로 예상된다.

그러나 본 연구에서 보여준 결과는 하나의 레퍼런스 수리전도도 분포를 이용하여 얻어진 결과로 해당 연구에서 제안된 메트로폴리스 알고리즘이 타 기법에 비하여 상대적으로 우수하다는 결론을 얻기 위해서는 추가적인 연구가 필요하다. 이를 위하여 향후 이루어질 연구에서는 3차원 도메인 적용을 통한 검증과 다양한 정규성 및 비정규성, 그리고 다양한 정도의 불균질성을 갖는 레퍼런스 수리전도도 분포를 재현하고 이를 이용하여 종합적인 결론을 통해 검증을 실시할 예정이다.

사 사

본 연구는 환경부의 이산화탄소 저장 환경관리기술개발

사업에서 지원받았습니다.

References

- Bertino, L., Evensen, G., and Wackernagel, H., 2002, Combinin geostatistics and Kalman filtering for data assimilation in an estuarine system, *Institute of Physics Publishing*, **18**, 1-23.
- Cardiff, M. and Kitanidis, P.K., 2009, Bayesian inversion for facies detection: an extensible level set framework, *Water Resour. Res.*, **45**(10), W10416, doi:10.1029/2008WR007675.
- Chen, Y. and Zhang, D., 2006, Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Resour.*, **29**, 1107-1122.
- Chen, Y., Oliver, D.S., and Zhang, D., 2009, Data assimilation for nonlinear problems by ensemble Kalman filter with reparameterization, *J. Pet. Sci. Eng.*, **66**, 1-14.
- Deutsch, C.V. and Journel, A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide, 2nd edition, Oxford University Press., New York, 369p.
- Deutsch, C.V. and Tran, T.T., 2002, FLUVSIM:a program for object-based stochastic modeling of fluvial depositional systems, *Comput. Geosci.*, **28**, 525-535.
- Efendiev, Y., Datta-Gupta, A., Ma, X., and Mallick, B., 2009, Efficient sampling techniques for uncertainty quantification in history matching using nonlinear error models and ensemble level upscaling techniques, *Water Resour. Res.*, **45**(11), W11414, doi:10.1029/2008WR007039.
- 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.
- Fu, J. and Gomez-Hernandez, J.J., 2009, A blocking Markov chain monte carlo method for inverse stochastic hydrogeological modeling, *Math. Geosci.*, **41**, 105-128.
- He, Z., Yoon, S., and Datta-Gupta, A., 2002, Streamline-based production data integration with gravity and changing field conditions, *SPE J.*, **7**(4), 423-436, doi:10.2118/81208-PA.
- 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.
- Hu, L.Y. and Chuginova, T., 2008, Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review, *Water Resour. Res.*, **44**(11), W11413.
- Hyndman, D.W., Harris, J.M., and Gorelick, S.M., 1994, Cou-

- pled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, **30**(7), 1965-1977.
- Jeong, J., Park, E., Han, W.S., and Kim, K.-Y., 2014, A novel data assimilation methodology for predicting lithology based on sequence labeling algorithms, *J. Geophys. Res. Solid Earth*, **119**, 7503-7520.
- Kitanidis, P.K., 1995, Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, **31**(10), 2411-2419.
- Kitanidis, P.K., 2015, Compressed state Kalman filter for large systems, *Adv. Water Resour.*, **76**, 120-126.
- Kretz, V., Ravalec-Dupin, M.L., and Roggero, F., 2004, An integrated reservoir characterization study matching production data and 4D seismic, *SPE J.*, **7**(2), 116-122, doi:10.2118/88033-PA.
- Mariethoz, G., Renard, P., Cornaton, F., and Jaquet, O., 2009, Truncated plurigaussian simulations to characterize aquifer heterogeneity, *Ground Water*, **47**(1), 13-24.
- 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.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E., 1953, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087-1091.
- Murakami, H., Chen, X., Hahn, M.S., Liu, Y., Rockhold, M.L., Vermeul, V.R., Zachara, J.M., and Rubin, Y., 2010, Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area, *Hydrol. Earth Syst. Sci.*, **14**, 1989-2001, doi:10.5194/hess-14-1989-2010.
- Oliver, D.S., Cunha, L.B., and Reynolds, A.C., 1997, Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math. Geosci.*, **29**(1), 61-91.
- Pesti, G., Bogardi, I., Kelly, W.E., and Kalinski, R., 1993, Cokriging of geoelectric and well data to define aquifer properties, *Ground water*, **31**(6), 905-912, doi:10.1111/j.1745-6584.1993.tb00863.x.
- Reichle, R.H., McLaughlin, D.B., and Entekhabi, D., 2002, Hydrologic data assimilation with the ensemble Kalman filter, *Am. Meteorol. Soc.*, **130**, 103-114.
- Rubin, Y., Chen, X., Murakami, H., and Hahn, M., 2010, A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, *Water Resour. Res.*, **46**(10), W10523, doi:10.1029/2009WR008799.
- Shin, Y., Jeong, H., and Choe, J., 2010, Reservoir characterization using an EnKF and a non-parametric approach for highly non-Gaussian permeability fields, *Energy Sources, Part A: Recovery, Util. Environ. Effectes*, **32**(16), 1569-1578.
- Strebelle, S., 2000, Sequential simulation drawing structures from training images, Ph.D. thesis, Stanford University, USA, 187p.
- Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.*, **34**(1), 1-22.
- Will, R., Archer, R., and Dershowitz, B., 2005, Integration of seismic anisotropy and reservoir-performance data for characterization of naturally fractured reservoirs using discrete-feature-network models, *SPE J.*, **8**(2), 132-142, doi:10.2118/84412-PA.
- Yoon, S., Malallah, A.H., Datta-Gupta, A., Vasco, D.W., and Behrens, R.A., 2001, A multiscale approach to production-data integration using streamline models, *SPE J.*, **6**(2), 182-192, doi:10.2118/71313-PA.
- Zhou, H., Gomez-Hernandez, J.J., Franssen, H.-J.H., and Li, L., 2011, An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Adv. Water Resour.*, **34**, 844-864.
- Zhou, H., Li, L., Franssen, H.-J.H., and Gomez-Hernandez, J.J., 2012, Patter recognition in a bimodal aquifer using the normal-score ensemble Kalman filter, *Math. Geosci.*, **44**, 169-185.