Journal of Contaminant Hydrology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Contaminant Hydrology

journal homepage: www.elsevier.com/locate/jconhyd

A predictive estimation method for carbon dioxide transport by data-driven modeling with a physically-based data model

Jina Jeong[a], Eungyu Park[b,*], Weon Shik Han[c], Kue-Young Kim[a], Seong-Chun Jun[d], Sungwook Choung[e], Seong-Taek Yun[f], Junho Oh[b], Hyun-Jun Kim[g]

[a] Korea Institute of Geoscience and Mineral Resources, Daejeon, Republic of Korea
[b] Department of Geology, Kyungpook National University, Daegu, Republic of Korea
[c] Department of Earth System Sciences, Yonsei University, Seoul, Republic of Korea
[d] Geogreen 21, Inc., Seoul, Republic of Korea
[e] Division of Earth and Environmental Sciences, Korea Basic Science Institute, Choengju, Republic of Korea
[f] Earth and Environmental Sciences, Korea University, Seoul, Republic of Korea
[g] BK21 Plus Eco-Leader Education Center, Korea University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Data-driven model
Process-based model
Data model
CO_2 concentration
Early warning system (EWS)
Cyber-physical system (CPS)

ABSTRACT

In this study, a data-driven method for predicting CO_2 leaks and associated concentrations from geological CO_2 sequestration is developed. Several candidate models are compared based on their reproducibility and predictive capability for CO_2 concentration measurements from the Environment Impact Evaluation Test (EIT) site in Korea. Based on the data mining results, a one-dimensional solution of the advective–dispersive equation for steady flow (i.e., Ogata-Banks solution) is found to be most representative for the test data, and this model is adopted as the data model for the developed method. In the validation step, the method is applied to estimate future CO_2 concentrations with the reference estimation by the Ogata-Banks solution, where a part of earlier data is used as the training dataset. From the analysis, it is found that the ensemble mean of multiple estimations based on the developed method shows high prediction accuracy relative to the reference estimation. In addition, the majority of the data to be predicted are included in the proposed quantile interval, which suggests adequate representation of the uncertainty by the developed method. Therefore, the incorporation of a reasonable physically-based data model enhances the prediction capability of the data-driven model. The proposed method is not confined to estimations of CO_2 concentration and may be applied to various real-time monitoring data from subsurface sites to develop automated control, management or decision-making systems.

1. Introduction

Risk assessment provides important information regarding the potential hazards imposed by CO_2 leakage. During the operational phase of CO_2 storage, timely information from risk assessment enables decision-makers to prepare plans preemptively to mitigate or avoid such risks, and heavily relies on monitoring.

Because of the importance of monitoring techniques, the Intergovernmental Panel on Climate Change (IPCC) evaluated existing monitoring, measurement, and verification (MMV) techniques and suggested that site-specific MMV plans should be applied for individual geological storage sites (Metz et al., 2005). Therefore, several small-scale controlled CO_2 release tests have been performed to evaluate site-specific MMV technologies alongside commercial-scale CO_2 injection (Jones et al., 2015). Well-designed controlled release testing is important to improve monitoring and site-characterization techniques, which aid in the development of continuous monitoring and estimation tools for the operation period. In a controlled CO_2 release test conducted by the Zero Emissions Research and Technology (ZERT) project in Montana, USA, CO_2 was released from a line source buried at depths of 1.3–2.5 m in an agricultural field (Strazisar et al., 2009; Zheng et al., 2012). Soil CO_2 flux measurements, in conjunction with numerical simulations, indicated that CO_2 leakage patterns at the surface are primarily governed by both subsurface heterogeneity and the dynamic

Abbreviations: AIC, Akaike information criterion; ASGARD, Artificial Soil Gassing and Response Detection; CPS, cyber-physical system; DHM, data-driven hydrogeological model; EIT, Environment Impact evaluation Test; EWS, early warning system; GPR, Gaussian process regression; IPCC, Intergovernmental Panel on Climate Change; K-COSEM, Korea CO_2 Storage Environmental Management; MMV, monitoring, measurement and verification; PHM, process-based hydrogeological model; RSS, residual sum of squares; ZERT, Zero Emissions Research and Technology

* Corresponding author.
E-mail address: egpark@knu.ac.kr (E. Park).

http://dx.doi.org/10.1016/j.jconhyd.2017.09.011
Received 30 December 2016; Received in revised form 5 August 2017; Accepted 24 September 2017

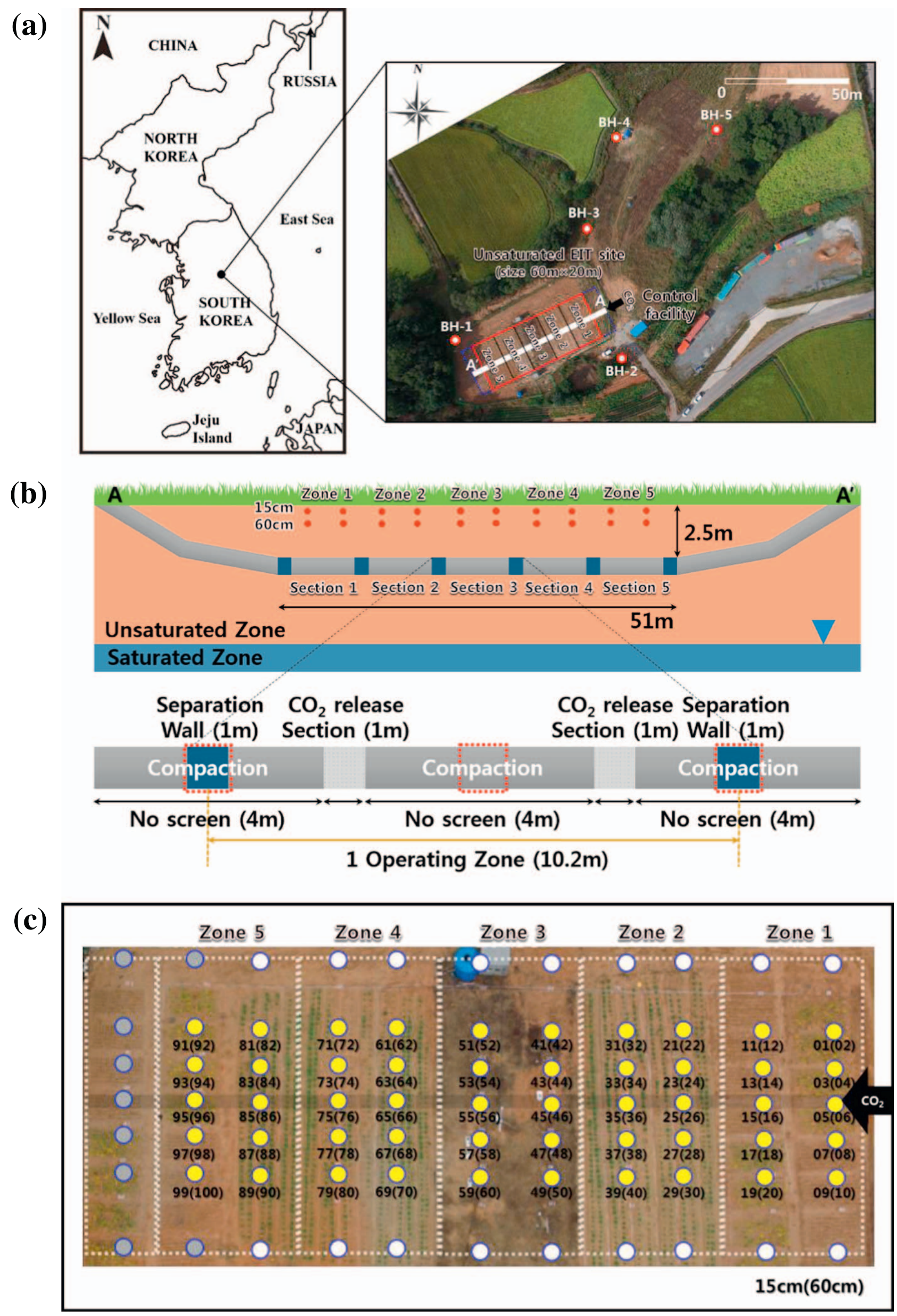

Fig. 1. (a) Location and an aerial photograph of the study area and the controlled CO_2 release test facility, (b) a schematic cross-sectional view of the installed horizontal well for CO_2 injection and the detailed design of the wellbore, and (c) the location of the monitoring points and their identification numbers.

fluctuation of the groundwater table (Lewicki et al., 2007). In an example from the Artificial Soil Gassing and Response Detection (ASGARD) project in Nottingham, UK, continuous monitoring of CO_2 concentration and flux revealed distinct damage to both soil microbiology and vegetation growth caused by CO_2 release (Smith et al., 2013). Similarly, a controlled CO_2 release experiment was performed in the Ginninderra site (NSW, Australia) by Geoscience Australia and the CO2CRC during 2012–2013 (Feitz et al., 2014). Based on > 10 different near-surface monitoring techniques, it was found that climatic conditions, groundwater fluctuation, and soil moisture substantially affected subsurface CO_2 migration. In these previous studies, a range of suitable monitoring technologies was identified. Nevertheless, uncertainties regarding measurements must be addressed when these technologies are applied for risk assessment and forecasting of CO_2 leakage.

Operational risk assessment, which is based on a framework of a real-time monitoring, includes, as the principal components, one or more predictive estimation methods. Estimation methods can be categorized into (1) process-based and (2) data-driven models. Process-based hydrogeological models (PHMs) generally adopt the mathematical form of a partial differential equation (PDE), which is solved using either analytical or numerical techniques. Strictly following physical laws, the PHMs directly solve conservation equations for fluid flow (e.g., Darcy's law) or solute transport (e.g., Fick's law) with several constitutive equations (e.g., equations of states, relative permeability, and capillary pressure) for cases with interdependent parameters. In terms of risk assessment of CO_2 leakage, PHMs are employed in the interpretation of CO_2 leakage potential associated with geological or anthropogenic conduits, such as faults and abandoned wells (Nordbotten et al., 2005; Stauffer et al., 2009; Oldenburg et al., 2010;