

Joo, Yongsung; Brumback, Babette; Lee, Keunbaik; Yun, Seong-Taek; Kim, Kyoung-Ho; Joo, Chaeman

Clustering of temporal profiles using a Bayesian logistic mixture model: analyzing groundwater level data to understand the characteristics of urban groundwater recharge. (English)

Zbl 1306.62300

J. Agric. Biol. Environ. Stat. 14, No. 3, 356-373 (2009).

Summary: The hydrogeologic conditions of groundwater can be examined by carefully studying the patterns of fluctuations in groundwater levels. These fluctuations are spatially and temporally influenced by many complicated factors, including rainfall, topography, land use, and hydraulic properties of soils and bedrock (i.e., aquifers). In this article we report a methodology based on the Bayesian logistic mixture model to simultaneously cluster profiles of groundwater level changes over time and estimate the relationships between the characteristics of each cluster and environmental variables. We apply the proposed method to analyze groundwater level profiles from 37 monitoring wells in Seoul, South Korea, and we find four clusters of wells. Using the estimated relationship between the clusters and the environmental variables, we discern the hydrogeologic conditions of each cluster, thus gaining insight into the recharge and subsurface flow of bedrock groundwater in an urban setting and the vulnerability of groundwater to the inflow of potential pollutants from ground surface.

MSC:

62P12 Applications of statistics to environmental and related topics

Keywords:

clustering of time course data; hydrogeology; model-based clustering

Software:

SemiPar

Full Text: [DOI](#)

References:

- [1] Basford, K. E., and McLachlan, G. J. (1985), "Likelihood Estimation With Normal Mixture Models," *Applied Statistics*, 34, 282–289. · [doi:10.2307/2347474](#)
- [2] Bockgard, K. (2004), "Groundwater Recharge in Crystalline Bedrock: Processes, Estimation, and Modelling," PhD thesis, Uppsala University, Sweden.
- [3] Brumback, B. A., Brumback, L. C., and Lindstrom, M. J. (2008), "Penalized Spline Models for Longitudinal Data," in *Longitudinal Data Analysis*, eds. G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, Boca Raton: Chapman & Hall, pp. 291–316.
- [4] Celeux, F., Forbes, F., Robert, C. P., and Titterton, D. M. (2006), "Deviance Information Criterion for Missing Data Models," *Bayesian Analysis*, 1, 651–674. · [Zbl 1331.62329](#) · [doi:10.1214/06-BA122](#)
- [5] Chae, G. T., Yun, S. T., Choi, B. Y., Yu, S. Y., Jo, H. Y., Mayer, B., Kim, Y. J., and Lee, J. Y. (2008), "Hydrochemistry of Urban Groundwater, Seoul, Korea: The Impact of Subway Tunnels on Groundwater Quality," *Journal of Contaminant Hydrology*, 101, 42–52. · [doi:10.1016/j.jconhyd.2008.07.008](#)
- [6] Fetter, C. W. (2000), *Applied Hydrogeology*, Upper Saddle River, NJ: Prentice-Hall.
- [7] James, G. M., and Sugar, C. A. (2003), "Clustering for Sparsely Sampled Functional Data," *Journal of the American Statistical Association*, 98, 397–408. · [Zbl 1041.62052](#) · [doi:10.1198/016214503000189](#)
- [8] Jiang, W., and Tanner, M. (1999), "Hierarchical Mixtures-of-Experts for Exponential Family Regression Models: Approximation and Maximum Likelihood Estimation," *The Annals of Statistics*, 27, 987–1011. · [Zbl 0957.62032](#) · [doi:10.1214/aos/1017939247](#)
- [9] Joo, Y., Lee, K., Yun, S., and Min, J. (2007), "Logistic Mixture of Multivariate Regressions for Analysis of Water Quality Impacted by Agrochemicals," *Environmetrics*, 18, 499–514. · [doi:10.1002/env.820](#)
- [10] Larocque, M., Mangin, A., Razack, M., and Banton, O. (1998), "Contribution of Correlation and Spectral Analyses to the Regional Study of a Large Karst Aquifer (Charente, France)," *Journal of Hydrology*, 205, 217–231. · [doi:10.1016/S0022-1694\(97\)00155-8](#)

- [11] Lee, J. Y., and Lee, K. K. (2000), "Use of Hydrologic Time Series Data for Identification of Recharge Mechanism in a Fractured Bedrock Aquifer System," *Journal of Hydrology*, 229, 190–201. · [doi:10.1016/S0022-1694\(00\)00158-X](https://doi.org/10.1016/S0022-1694(00)00158-X)
- [12] Ma, P., and Zhong, W. (2008), "Penalized Clustering of Large Scale Functional Data With Multiple Covariates," *Journal of the American Statistical Association*, 103, 625–636. · [Zbl 05564516](https://zbmath.org/journal/05564516) · [doi:10.1198/016214508000000247](https://doi.org/10.1198/016214508000000247)
- [13] Park, S. S., Kim, S. O., Yun, S. T., Chae, G. T., Yu, S. Y., Kim, S., and Kim, Y. (2005), "Effects of Land Use on the Spatial Distribution of Trace Metals and Volatile Organic Compounds in Urban Groundwater, Seoul, Korea," *Environmental Geology*, 48, 1116–1131. · [doi:10.1007/s00254-005-0053-8](https://doi.org/10.1007/s00254-005-0053-8)
- [14] Pfeiffer, R. M., Carroll, R. J., Wheeler, W., Whitby, D., and Mbulaiteye, S. (2007), "Combining Assays for Estimating Prevalence of Human Herpesvirus 8 Infection Using Multivariate Mixture Models," *Biostatistics*, 9, 137–151. · [Zbl 1274.62855](https://zbmath.org/journal/1274.62855) · [doi:10.1093/biostatistics/kxm018](https://doi.org/10.1093/biostatistics/kxm018)
- [15] Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components," *Journal of the Royal Statistical Society, Ser. B*, 59, 731–792. · [Zbl 0891.62020](https://zbmath.org/journal/0891.62020) · [doi:10.1111/1467-9868.00095](https://doi.org/10.1111/1467-9868.00095)
- [16] Rodhe, A., and Bockgard, N. (2006), "Groundwater Recharge in a Hard Rock Aquifer—a Conceptual Model Including Surface Loading Effects," *Journal of Hydrology*, 330, 389–401. · [doi:10.1016/j.jhydrol.2006.03.032](https://doi.org/10.1016/j.jhydrol.2006.03.032)
- [17] Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge, U.K.: Cambridge University Press.
- [18] SMG (Seoul Metropolitan Government) (1996), *Master Plan of Groundwater Management of Seoul*, Seoul, Korea: SMG (in Korean).
- [19] Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Ser. B*, 64, 583–640. · [Zbl 1067.62010](https://zbmath.org/journal/1067.62010) · [doi:10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353)
- [20] Stephens, M. (2000), "Dealing With Label-Switching in Mixture Models," *Journal of the Royal Statistical Society, Ser. B*, 62, 795–809. · [Zbl 0957.62020](https://zbmath.org/journal/0957.62020) · [doi:10.1111/1467-9868.00265](https://doi.org/10.1111/1467-9868.00265)
- [21] Turner, T. R. (2000), "Estimating the Propagation Rate of a Viral Infection of Potato Plants via Mixtures of Regressions," *Applied Statistics*, 49, 371–384. · [Zbl 0971.62076](https://zbmath.org/journal/0971.62076)
- [22] Yang, Y., Lerner, D. N., Barrett, M. H., and Tellam, J. H. (1999), "Quantification of Groundwater Recharge in the City of Nottingham, UK," *Environmental Geology*, 38, 183–198. · [doi:10.1007/s002540050414](https://doi.org/10.1007/s002540050414)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.