

Applicability Domain Based on K-Nearest Neighbors Approach in the QSPR Model

Jong Hyuk Lee¹, Seung Joon Kim¹, Sung Kwang Lee^{1,*}

^{1,*}Department of Chemistry, Hannam University, Daejeon, Korea, leesk@hnu.kr

Abstract: *As the interest in use of QSPR models has been growing rapidly, it is quite evident from their decision –making use for environmental regulation such as European REACH and K-REACH legislation. The reliability of the prediction from QSPR models is also a matter of growing concern. Thus, it is essential process to define the applicability domain in QSPR model for better reliability of unknown compound. In this study k nearest neighbors approach for applicability domain evaluation, all distance of all the training sample from their k nearest neighbors are calculated and used to define a unique threshold to 95th percentile accumulated distance decide if a test sample is inside or outside the model's applicability domain. The Results on the selected case study defined an applicability domain with a positive impact on model statistics retaining maximum possible samples that were reliably predicted.*

Keywords: *Applicability domain, QSPR, k nearest neighbours*