Generative Topographic Mapping (GTM) is a dimensionality reduction method, and the probabilistic counterpart of Kohonen maps. Each \( i \)-th molecule in \( N \)-dimensional initial space is projected into the \( k \)-th node of 2D latent space with a probability \( R_{ik} \), so that each compound is represented both by a mean position (a point) on a 2D map, and a probability distribution \( R_i \), which may be used for predictions of activity (property) of new compounds.

Here, we suggest several different GTM-based definitions of applicability domain (AD) of both regression and classification models. This concerns the approaches involving: (i) a likelihood threshold, (ii) relative population of nodes, (iii) class entropy, and, (iv) ratio of classes’ probabilities. These approaches are demonstrated for regression models for stability constants of ligand-metal complexes, and GTM-based classification models \([1]\) for Biopharmaceutics Drug Disposition Classification System (BDDCS) \([2]\) and inhibitors of P-glycoprotein 1 (Pgp), an ATP-dependent efflux pump.

![Figure 1. Graphical interpretation of the applicability domain for GTM classification models.](image)

CPF = 1, BA= 0.84 , coverage = 100 %  
CPF = 4, BA= 0.90, coverage = 78 %
