There are different techniques to obtain 3D structures of macromolecular assemblies from electron micrographs: electron tomography with single-axis, dual-axes or conical tilting, random conical reconstruction, angular reconstitution, and orthogonal tilting reconstruction. The techniques including tilting are the most suitable for investigating structures of heterogeneous samples. However, most tilting techniques leave out areas of missing data in the 3D reconstructions because the tilt angle range is limited in the electron microscope. Combining volumes representing the same structure can achieve a volume with no missing data or much less missing data. However, missing data present an obstacle to analyzing variations between 3D volumes since they cause artifacts in the 3D reconstructions. Especially artifacts caused by missing data in different orientations can be easily misinterpreted as structural differences.

Here we present an algorithm, Probabilistic Principal Component Analysis using Expectation maximization algorithm (PPCA-EM), which is adopted from Tipping and Bishop's framework, and which can perform principal component analysis (PCA) on a set of 3D volumes with arbitrary missing data. Like traditional PCA, PPCA-EM not only reduces the dimensionality of the data which reduces the complexity of any subsequent classification, but also reduces the noise, thus increasing the robustness of the classification. Unlike traditional PCA, PPCA-EM estimates the missing observations as well. This becomes highly valuable when no clearly defined classes exist but the data shows continuous variations that prevent commonly used averaging techniques from being applied.

The algorithm has been tested extensively on simulated 3D volume sets containing four slightly different conformations of a macromolecule. Different sets were created by varying the signal-to-noise ratio (SNR), the percentage of missing data and the number of volumes in a set. The tests have shown the strength and the limitation of the algorithm.