

Markov state modeling

Climate evolution on the geological time scale

At the link

<https://drive.google.com/file/d/1nMr4XdOXo3Xfs5zAJPczm7h0fXVwUsU1>

you can download a time series with the value of temperature as a function of time in 96 positions equally distributed on Earth. The time series has been generated with the model described in [Fraedrich K, et al 2005 The Planet Simulator: Towards a user friendly model. Meteorologische Zeitschrift 14, 299–304.] and covers a time span in which the Earth enters a few times in a Ice age.

Build a Markov State model from this time series, following, for example, this pipeline:

1. find the microstates by a cluster analysis performed with k-means or k-medoids. Optionally, perform a preliminary dimensional reduction using PCA
2. choose a time lag τ , and estimate the transition matrix between each pair of clusters $\pi_{\alpha,\beta} = P(\beta, \tau | \alpha, 0)$
3. By analyzing the spectrum choose an appropriate number of Markov states. Find the markov states by inspecting the sign of the eigenvectors.
4. Visualize the markov states in the space of the two first principal components, or, alternatively, in the coordinates obtained by Isomap
5. (optional): discuss the stability of the results with respect to the metaparameters of the approach, in particular the time lag τ and the number of clusters.
6. (optional): Build a multivariate logistic regression model that predicts the Markov state from the original variables.

The metastable states of a peptide in water solution

From the link

<https://github.com/alexdepremia/Unsupervised-Learning-Datasets.git>

you can download the file `high_variance_distances.dat.gz`. It corresponds to a time series of the value of distances between pair of atoms belonging to different residues in the peptide of sequence Arg-Phe-Phe-Glu-Ala. The chosen distances are all those that during the dynamics vary significantly (variance larger than 5 \AA^2). This peptide is a fragment of α -synuclein, and studying its conformational dynamics is important to understand Alzheimer disease.

Build a Markov State model from this time series, following the same pipeline described above.