Geophysical Research Abstracts, Vol. 9, 02484, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-02484

Nonlinear principal component analysis of noisy data

W. Hsieh (1)

(1) University of British Columbia, Canada (whsieh@eos.ubc.ca / Fax: 1-604-822-6088)

With noisy data, having plentiful samples eliminates overfitting in nonlinear regression, but not in nonlinear principal component analysis (NLPCA). To overcome this problem in NLPCA, a new information criterion (IC) is proposed for selecting the best model among multiple models with different complexity and regularization (i.e. weight penalty). This IC gauges the inconsistency between the nonlinear principal components u and $\tilde{u}$ for every data point $\mathbf{x}$ and its nearest neighbour $\tilde{\mathbf{x}}$ via the negative correlation between u and $\tilde{u}$, which tends to increase with overfitted solutions. Tests were performed using autoassociative neural networks for NLPCA on synthetic and real climate data (tropical Pacific sea surface temperatures and equatorial stratospheric winds), with the IC performing well in model selection and in deciding between an open curve or a closed curved solution.