

**Hyper-Dimensional Categorical Time Series EDA:
Exploratory Data Analyses on Inventor2001
Evolutionary Strategy Parental Trajectories**

Christopher D. Cuppan
George Mason University

Visual Exploratory Data Analysis (EDA) on the genetic history (i.e., gene trajectories) of an evolutionary algorithm (EA) may prove valuable to both understanding and nurturing the performance of the EA as well as provide insights into the potential performance of other similar EA classes. These gene trajectories may be viewed as time series. However, EA genotypic trajectories often reflect nominally-scaled state transitions (i.e., categorical data). Such categorical (i.e., nominally-scaled) time series data are not effectively visualized and modeled using methods developed for ordinal data. Often, arbitrary mappings of categorical data to ordinal values (e.g., integer coding) maintain undesirable interpretation consequences. Such is the case with the gene trajectories produced by the Inventor2001 Innovative Design Tool, whose data was analyzed herein. Inventor2001 applies an evolutionary strategy (ES) algorithm toward the problem of wind bracing design structures in tall buildings. Statistical and Computer Science literature are limited on the subject of categorical data time series. This is a paradox given the huge amounts of WWW-based data being collected that are within the categorical data time series domain. Work done at Virginia Tech (1993-97) by Ribler, et al developed a theoretical framework for visual categorical data time series analyses. A portion of that body of theory and techniques was applied to the problem of analyzing the Inventor2001 gene trajectories. Preliminary results suggest that the visual EDA techniques will be valuable in identifying the specific design hypothesis components (i.e., gene values) that are the primary contributors to design fitness as well as provide a graphical means of depicting the EA's convergence properties.