

CHAPTER OVERVIEW

11: Finding Structure in Data

One of the more intriguing aspects of chemometrics is the ability to discover and extract information from a large data set that appears, at first glance, to lack any defined order. And yet, it is likely that there are determinate factors that explain the data. Consider a data set that consists of the daily concentration of NOX—the combined amounts of NO₂ and of NO in the air expressed as µg/m³—in samples of urban air. Although a plot of the concentration of NOX as a function of time likely appears noisy, we can easily identify variables that might affect the daily measurements:

- temperature: we need more energy on colder days, which increases the use of fuels that generate NOX emissions
- day of the week: perhaps more traffic on work days than on weekends
- atmospheric conditions: strong winds may disperse NOX emissions and stagnation may concentrate NOX emissions
- location of air samplers: samplers at busy intersections may give different results from samplers located in city parks

The chemometric methods introduced in this chapter—cluster analysis, principal component analysis, and multivariate linear regression—provide ways to probe the underlying factors that provide structure to our data.

[11.1: What Do We Mean By Structure?](#)

[11.2: Cluster Analysis](#)

[11.3: Principal Component Analysis](#)

[11.4: Multivariate Linear Regression](#)

[11.5: Using R for a Cluster Analysis](#)

[11.6: Using R for a Principal Component Analysis](#)

[11.7: Using R for a Multivariate Linear Regression](#)

[11.8: Exercises](#)

This page titled [11: Finding Structure in Data](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [David Harvey](#).