Chapter 2

STATISTICAL ANALYSIS OF SPATIAL STRUCTURE IN MICROBIAL COMMUNITIES

Overview of methods and approaches

Rima B. Franklin¹ and Aaron L. Mills²
¹Department of Biology, Virginia Commonwealth University, Richmond, VA 23284 USA;
²Laboratory of Microbial Ecology, University of Virginia, Charlottesville, VA 22904-4123 USA

Abstract: This chapter provides a review of the basic statistical techniques used to detect and quantify spatial structure in ecological data as they can be applied to the analysis of microbial communities. It also discusses the general implications of spatial structure in data analysis, including the inappropriate use of parametric statistical tests with spatially autocorrelated data, and suggests possible alternative procedures. Methods discussed include geostatistics and variogram analysis, kriging, correlograms, Mantel and partial Mantel tests, and time-series analysis.

Keywords: spatial structure, microbial communities, statistical analysis, autocorrelation, geostatistics, kriging, scale, spatial autocorrelation

1. INTRODUCTION

Most ecological theories and models acknowledge that elements close to one another in space (or time) are likely to be more similar, as they are influenced by the same generating processes, the same energy inputs, or a connected physical environment. However, the classical statistical procedures employed to analyze these phenomena usually consider the biological organisms, and their controlling variables, to be distributed in a random or uniform way, neglecting the natural spatial structure. While the importance of spatial structure in experimental design and statistical analyses is generally accepted, most ecologists do not fully consider it when designing a sampling scheme or evaluating data. As scientists have become more aware of the
importance of the spatial components of the phenomena they study, and as the number of statistical and computational tools available for quantifying these processes has increased, explicit considerations of spatial structure in microbial ecology studies has become more common.

In general, studies of spatial structure begin with exploratory analyses, followed by the application of techniques aimed at detecting and characterizing spatial patterns. Statistically, this entails testing data against the null hypothesis that there is no pattern in the data. In situations where it is determined that the data are patterned, analyses that allow one to distinguish competing agents of pattern can be applied. This latter stage involves posing alternative models for the pattern, and comparing these models against each other to find the most likely explanation for the observed structure. Depending on the nature of the data and the patterning agents involved, this stage of analysis can take several directions. The purpose of this chapter is to provide an overview of some of the more commonly available methods for detecting and characterizing spatial structure in ecological data, and to discuss their application in the analysis of microbial communities. Sampling strategies for the environment must also be considered, along with procedures for hypothesis testing that incorporate spatial structure. Most of the chapter focuses on the application of geostatistical approaches, including variograms and kriging, to the analysis of spatial autocorrelation. A brief overview of other methods is included, as are references for those seeking more information. In addition, the reader is advised to consult Chapter 8 for more detailed discussion of advanced multivariate methods and spectral analysis.

2. MOTIVATION FOR STUDYING SPATIAL SCALE

Research into the spatial distribution of microorganisms and microbial communities has many ecological and environmental applications, which have been discussed in detail as part of Chapter 1. In addition, the existence of spatial structure has important consequences for scientists considering other aspects of microbial ecology, even those who are not interested in spatial structure per se. For example, a better understanding of spatial variation and scale-dependent patterns is important for the design of field experiments and for the correct application of statistical hypothesis tests. These issues have been extensively reviewed in both the statistical and ecological literature (Bonham and Reich, 1999; Dale and Fortin, 2002; Hoosbeek et al., 1998; Legendre et al., 2004; Sokal et al., 1993; van Es and van Es, 1993), and will be briefly discussed here.