Homework 14 – Small bacteria and ocean warming

Heterotrophic bacteria comprise a large portion of total biomass in marine ecosystems, and flow cytometric sampling is able to distinguish two clusters of bacteria based on their nucleic acid content, 'high nucleic acid' (HNA) and 'low nucleic acid' (LNA). LNA bacteria tend to be smaller in terms of cell size, and tend to be dominated by the widespread SAR11 clade. Moran et al. (2015) used a ten year, monthly time series from the southern Bay of Biscay to consider seasonal and longer-term patterns in the abundance, biomass, and cell size of these two groups of bacteria. They were particularly interested in whether patterns of abundance and cell size may be driven by temperature. The attached dataset includes a variety of measurements taken at monthly resolution. We will focus on 'LNA ab uml' = LNA abundance in cells per mL, 'LNA bv' = LNA mean cell size in units of cubic microns, 'LNA B' = LNA total biomass in units of µg C per L, and '%LNA BB' = percent of total bacterial biomass in the LNA fraction. Other important variables are year, month, date, day of year, and 'temp 5 m E2' = temperature at 5 meters depth. Make sure to convert the date column to a date format – the package lubridate is a useful one for this purpose.

1. Begin by exploring seasonal patterns in LNA abundance, cell size, biomass, and percent biomass, as well as temperature. I.e., make exploratory plots of these variables that focus on seasonality. What have you learned so far?

2. Now make new exploratory plots of the same five variables, focusing on longer-term trends. How do you interpret these plots?

3. Let's create statistical models to test for long-term trends, as well as relationships with temperature. Because the sampling is monthly, with only a couple gaps, we can treat the time series as one measured at discrete intervals. To keep track of the sample order you'll need to construct a new column which numbers each sample in the order they were sampled. I.e., the column should look like 1, 2, 3, 4, etc., where the numbers correspond to the month in the time series. You can use the year and month columns to construct this.

   Use linear models to test for long term (linear) trends in the five variables: LNA abundance, LNA cell size, LNA biomass, LNA percent biomass, and temperature. For each of these models, assess whether the residuals show evidence of temporal autocorrelation. If autocorrelation is present, add a residual autocorrelation component to your model. Assess whether the model successfully accounted for autocorrelation. What have you learned from these models? Does accounting for autocorrelation change the results?

4. Perform a similar set of analyses using temperature as the predictor, to test whether temperature can explain variation in the four LNA metrics, and whether autocorrelation needs to be accounted for when testing these relationships. Make

appropriate plots of the results and discuss the magnitude of the relationships. Considering the whole set of analyses, what are your interpretations of the dynamics and drivers of LNA bacteria in this system?