Session 7: distributions (Lab Report 1)

Lise Montefiore and Natalie Nelson, PhD; Biological & Agricultural Engineering, NCSU

03/02/2018

Background

Your engineering firm was hired 5 years ago to consult the City of Raleigh on their plans to implement green infrastructure throughout the city. In particular, they are interested in replacing several impervious parking lots with pervious pavement (Figure 1A). Pervious concrete is a mixture of coarse aggregete, cement, water, and little to no sand. Its porous structure allows water to infiltrate to the underlaying soils (Figure 1B), thereby reducing surface water runoff.

The City of Raleigh has specifically asked your firm to help them determine the supplier they should choose for the city-wide pervious pavement project. Depending on pervious concrete mix-designs, pervious concrete can have different properties (coarse aggreagate size, permeability rate, etc.). Over time, pervious concrete can be clogged by coarse and fine particles (transported by wind, car tires, etc.), decreasing its infitration properties. Additionally, the pervious concrete can begin to break down over time.

The City of Raleigh is interested in choosing a pervious pavement that will resist clogging and degradation. When they hired your firm 5 years ago, they asked you to test five commercial pervious pavement (P1, P2, P3, P4, P5) materials. You installed test lots for each of the pervious pavement types. Now, 5 years later, you are assessing the pavements for clogging and degradation. To do this, you used a shop vacuum to pull all of the particles out of the pavement. You then used a water jet to push any remaining particles out of the pavement, and collected them with the shop vac. Therefore, the dry + wet collections captured the total particles contained in the pavement. You determined the distribution of particle sizes from the "dry" and "wet" collection methods using sieve analysis (Figure 1C).

You have identified that, when pervious pavement degrades, it produces particles that are greater than 5.15 mm in size. In contrast, particles that are smaller than 5.15 mm come from the surrounding environment, and are those that ultimately clog the pervious pavement - the smaller the particles, the more likely they are to aggregate and form clogs. You have determined that the pervious pavement you recommend to the City of Raleigh should have no more than 20% of particles above 5.15 mm; pavements that are found to have more than 20% of particles above 5.15 mm degrade too quickly. Of the pavements that meet the criteria of having less than 20% of particles > 5.15 mm, you will ultimately choose the pervious pavement that minimizes clogging potential.

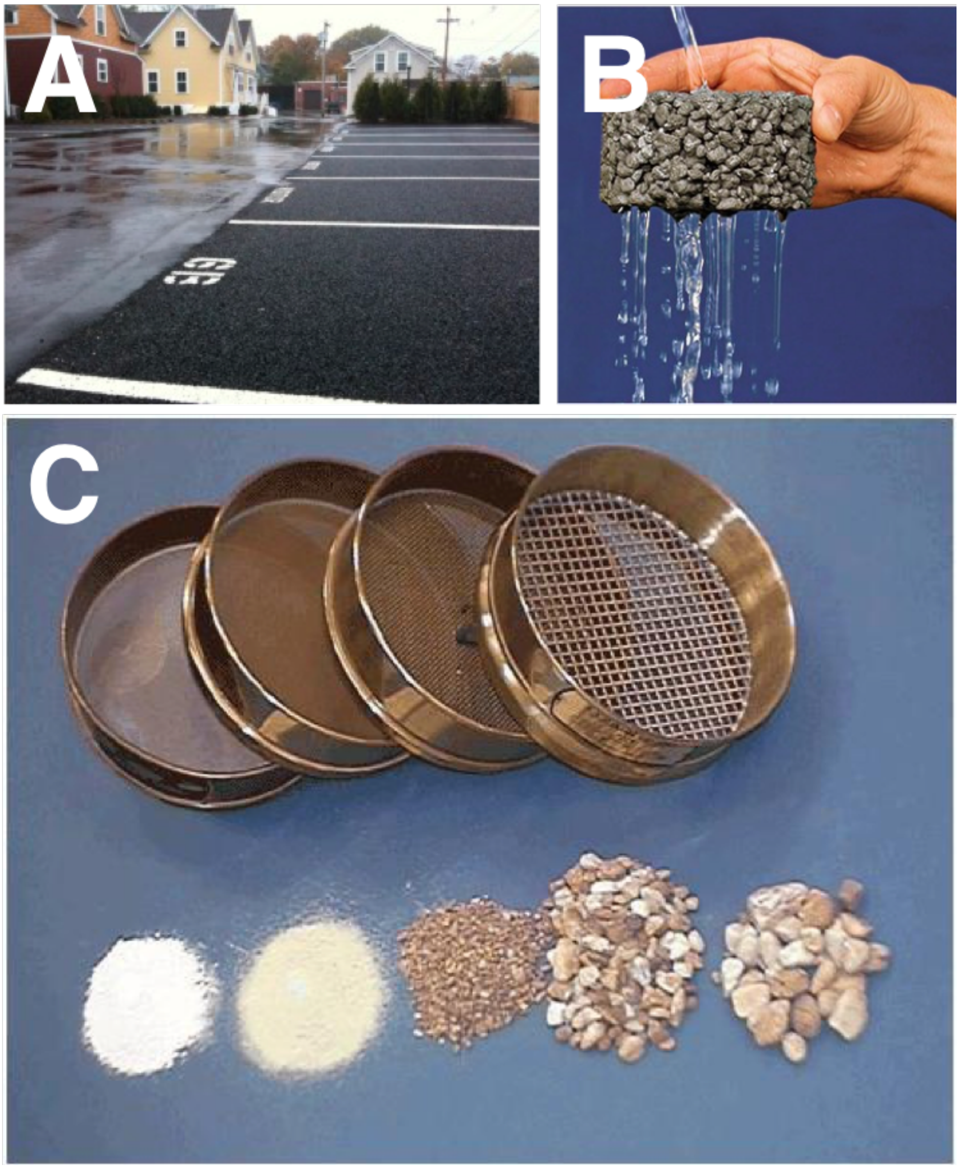

Figure 1. (A) An impervious roadway shown next to pervious parking spaces. Note that water collects on the impervious concrete, while it does not on the pervious pavement (credit: Green Building Elements); (B) A slice of pervious pavement, which demonstrates its capacity to convey water through the material (credit: VDC Green); (C) Sieves used in a sieve analysis (credit: Basic Civil Engineering).

Particle data

Your data are included in two data files. The first file is "Dry-particle-mass.csv", which includes the masses of particles collected after the "dry" collection for each of the 5 pervious pavement types. Each row corresponds to the mass of particles associated with a particle size (mm). The second file is "Dry-wet-particle-mass.csv", which includes the total masses of particles from the "dry" and "wet" collections combined. **Note that these data are cumulative.** The masses give you an indication of the relative proportion of a given particle size class. You are ultimately interested in the sizes, not the masses.

Analysis objectives

All analysis should be performed in R. It is suggested that these objectives be addressed in the order they are listed.

- Determine which pervious pavement minimizes degradation and clogging. Support your answers with cumulative distributions. Plot the cumulative distributions on a log x-scale (see section below for details on how to code this). The distributions for the 5 pervious pavement types should all be plotted in the same panel, or should be multi-panel via facetting - do **not** submit 5 individual graphs for the 5 distributions.
- Determine what effect the "dry" and "dry+wet" collections had in terms of collecting particles. How do the distributions differ? What explains the differences? Support your answers with cumulative distributions (the same specifications made in the prior bullet also apply here).
- Once you choose the pervious pavement type you will recommend to the City of Raleigh, determine what the mean and standard deviation of the particle sizes are **if you assume that the data are normally distributed**. Remember, you can calculate these measures from the cumulative data based on properties of the normal distribution.
- Determine if the particle sizes associated with your recommended pervious pavement type follow a normal or lognormal distribution using the Q-Q Correlation method. Report the correlation coefficients associated with these tests, and support your answers with Q-Q plots. Do you think the Q-Q plots support the results of the Q-Q coefficient test?

New functions

Note that none of these functions require any new packages to be downloaded.

- Log x-axis: add `coord_trans(x='log10')` to your ggplot code.
- Take the natural log of your data (e.g., x): `log(x)`
- Removing "Inf" values: when calculating the Q-Q correlation, you cannot have any "Inf" (for "infinity") values in your data. The functions can't perform calculations with Inf (similar to how they can't perform calculations with NAs). To remove the Inf values, use `slice()`. Specifically, provide `slice()` with the rows you want to keep. For example, if you want to only keep rows 1 to 5, you would run `d <- d %>% slice(1:5)`.
- Correlation: to calculate the correlation between two data series (e.g., x and y), use `cor(x,y)`.

Report

The primary objective of your report is to summarize this study, detail your methods, and provide recommendations to the City of Raleigh on the pervious pavement type they should choose for their city-wide green infrastructure initiative. Although you can work together to perform the analysis, the report **must be in your own words** and all material should be original. Any reports with evidence of plagiarism will receive no credit.

Be sure to address each of the analysis objectives in your report. Your report should be structured as follows:

- Title page - title (create your own title), name, date
- Executive abstract - summarize the report in 4-6 sentences; be sure to include your main results and recommendations, as this is the most important information included in abstracts.
- Introduction - provide all background information needed to understand the contents of the report. Be sure to *paraphrase* the writeup included here; do not copy directly.

- Methods - explain the statistical methods you used, both based on the theory and the specific application at hand. Demonstrate that you understand how and why you used the statistical principles (cumulative distributions, characteristics of normal and lognormal distributions, determining the mean and standard deviation from a cumulative distribution assuming the data are normal, Q-Q correlation test) in this case study. Include equations when relevant. **You do not have to detail the steps you took in R.**
- Results - state the results objectively. Explain what the figures and numbers you've produced from this analysis are telling you.
- Recommendations - offer recommendations to the City of Raleigh as to which pervious pavement type they should choose, and why.
- Appendices - include your R code (should be neatly organized and professional) and the Q-Q correlation table.

Additional notes on formatting:

- All figures and tables should be captioned, numbered, and referred to in the text (e.g., "The cumulative distribution of the dry collection (Figure 1) demonstrates that…")
- Table captions are placed *above* the table. Figure captions are placed *below* the figure.
- Equations should be centered and included on their own lines.