

EES 3000 | MATLAB ENVIRONMENTAL DATA ACTIVITY

Module 1: Probability of Precipitation Events

STUDENT ASSIGNMENT

Activity overview

Scientific question. How often does precipitation occur at different mountain stations, and what does precipitation at one station tell us about another?

Learning outcomes. After this activity, students can (1) construct daily event indicators in MATLAB; (2) compute and verify compound probabilities using event algebra and De Morgan's laws; (3) distinguish conditional probabilities; (4) apply total probability and Bayes' theorem; and (5) communicate a defensible environmental interpretation.

Prerequisites and materials. Basic MATLAB scripts, vectors, indexing, logical operators, and introductory probability; MATLAB and the supplied `snotel.csv` dataset.

Submit. One annotated MATLAB script (`.m` or `.mlx`) and a short report containing probability estimates, checks, and written interpretations for Questions 1–7.

Data preparation

The six columns of `snotel.csv` are year, month, day, and daily precipitation (mm) at Brighton, Ben Lomond Peak, and Snowbird, respectively. Define R , L , and S as precipitation occurring at these three sites. For this activity, a wet day has precipitation strictly greater than 0 mm. Examine missing data first; use the same set of valid daily records when comparing events across sites. Report the number of records used.

```
D = readmatrix('snotel.csv');
valid = all(isfinite(D(:,4:6)),2);
P = D(valid,4:6);
R = P(:,1) > 0; L = P(:,2) > 0; S = P(:,3) > 0;
n = size(P,1);
% Complete the calculations and add interpretations below.
```

Questions: from individual events to scientific interpretation

- 1. Union, intersection, and complements (15 points).** (a) Calculate $P(R \cup L)$ and $P(R \cap L)$ using the addition or multiplication rule; verify each through a complementary-event identity. (b) Calculate $P((R \cup L) \setminus (R \cap L))$ and explain the corresponding weather condition.

- 2. Classifying three-station outcomes (15 points).** Let $E_1 = (R^c \cap L^c) \cup S$. (a) Estimate $P(E_1)$. (b) List all eight binary station outcomes, in the order [Brighton, Ben Lomond Peak, Snowbird], and identify which satisfy E_1 .
- 3. Conditional probability (15 points).** (a) Calculate and compare $P(R | L)$ and $P(L | R)$. (b) Explain which station's wet-day observation is the more reliable indicator of precipitation at both stations, and give a plausible physical interpretation without claiming causation.
- 4. Any-site precipitation (15 points).** Let $E_2 = R \cup L \cup S$. (a) Calculate $P(E_2)$ using inclusion-exclusion and verify by direct counting. (b) Determine α in

$$P(E_2) = P(R)\alpha + P(L \cup S) + P(R \cap L \cap S).$$

Assuming R is independent of both L and S , derive α and state clearly whether pairwise independence is sufficient to simplify the three-way intersection.

- 5. Law of total probability (15 points).** Partition days into the four mutually exclusive Brighton/Ben Lomond outcomes $R \cap L$, $R \cap L^c$, $R^c \cap L$, and $R^c \cap L^c$. Calculate $P(S) = \sum_{i=1}^4 P(S \cap E_i)$ and check against direct counting.
- 6. Bayes' theorem and forecast reliability (15 points).** Let B denote an actual tornado and A a positive tornado alert. Define sensitivity $s = P(A | B)$, false-alarm probability $q = P(A | B^c)$, and event base rate $p = P(B)$. Use

$$f = P(B | A) = \frac{sp}{sp + q(1 - p)}.$$

- (a) With q and p fixed, explore how much f can increase as s rises to 1; explain why the answer depends on the base rate. (b) Under what conditions does $f = 1$? Address the requirement that positive alerts must have nonzero probability.
- 7. Scientific synthesis (10 points).** State one scientifically interesting, data-supported probability observation. Include the event definition, numerical estimate, and one sentence about its limitations.

Before you submit

1. Confirm the number of complete daily records and define all events consistently.
2. Show the probability calculations, at least one independent check where requested, and clearly labeled numerical results.
3. Explain conditional denominators and provide written interpretations for Questions 3, 6, and 7.
4. Submit your documented MATLAB script and short report; include a note on missing values and the observation period if known.

Files and reference information

Required files: `snotel.csv` and `EES3000_Module1_Student_Starter.m`. Place both in the MATLAB working folder. The supplied SNOTEL file is a course dataset; consult the instructor for the source and observation period. Background: USDA NRCS SNOTEL program (<https://www.nrcs.usda.gov/>); MathWorks MATLAB documentation (<https://www.mathworks.com/help/matlab/>).