

EES 3000 | MATLAB ENVIRONMENTAL DATA ACTIVITY

Module 1: Probability of Precipitation Events

INSTRUCTOR GUIDE AND RUBRIC

Purpose and prerequisites

This guide accompanies the separate EES 3000 student assignment and the `snotel.csv` dataset. The objective is to help students translate precipitation-event notation into logical arrays, verify empirical probabilities, and explain scientific meaning. Students need introductory probability and MATLAB vectors, indexing, scripts, and logical operators. Allow multiple class or lab sessions plus independent work; the precise duration can be adapted to the course.

Required student products. A documented MATLAB script (or live script), numerical results for Questions 1–7, event-identity checks, a complete truth table for Question 2, and a concise written explanation of the conditional, forecasting, and scientific findings.

Instructor preparation. Distribute the student handout, `snotel.csv`, and `EES3000_Module1_Student_Starter.m`. Keep `EES3000_Module1_Instructor_Solutions.m` separate. Verify that the data and code work together in your MATLAB version.

Implementation guide

Suggested teaching sequence

- Stage 1: Launch and check data.** Demonstrate column identification, missing-value checks, and wet-day indicators. Ask students to predict which stations may co-occur most often.
- Stage 2: Event algebra (Questions 1–2).** Use a small truth table before applying logical operators to all days; compare direct counting with probability identities.
- Stage 3: Dependence across sites (Questions 3–5).** Move from two-station conditional probabilities to three-station inclusion–exclusion and a four-event partition. Discuss what empirical dependence means physically.
- Stage 4: Transfer and synthesis (Questions 6–7).** Use the self-contained tornado-alert scenario to connect conditional probability to forecast decisions, then ask for a concise original finding from data.

Teaching tips and likely misconceptions

Distinguish zero rainfall from missing observations; check that all station comparisons use a common denominator. Students often swap $P(R | L)$ and $P(L | R)$; ask them to identify the conditioning subset explicitly. Use De Morgan's laws, $(R \cup L)^c = R^c \cap L^c$ and $(R \cap L)^c = R^c \cup L^c$, as independent checks. In Question 4, independence of R from L and separately from S does *not* imply independence of R from $L \cap S$; do not silently assume it. For Question 6, discuss why a low base rate can limit the positive predictive value even with high sensitivity. Ask students to set `rng(1)` if extending the activity to simulation.

Assessment rubric (100 points)

Evidence assessed	Points
Event rules, two-site probability identities and interpretation (Q1)	15
Correct classification and compound-event estimate (Q2)	15
Conditional calculations and physical reasoning (Q3)	15
Three-site inclusion–exclusion and assumption critique (Q4)	15
Complete partition and direct-count validation (Q5)	15
Bayes calculation, base-rate reasoning, and edge cases (Q6)	15
Defensible original observation with quantitative support (Q7)	10
Total	100

Across all questions: Award credit for transparent code, correct mathematical reasoning, independent verification where requested, and scientifically meaningful explanations. Collect the script and short report; review one or two event definitions and conditional denominators during class discussion.

Instructor code checks (not a student answer key)

```
% Examples of consistency checks after loading R, L, S:
assert(abs(mean(R|L) - (mean(R)+mean(L)-mean(R&L)))<1e-12)
assert(abs(mean(~(R|L)) - mean(~R & ~L))<1e-12)
assert(abs(mean(R|L|S) - (1-mean(~R & ~L & ~S)))<1e-12)
```

Source and redistribution checklist

Required: student handout (PDF or editable L^AT_EX), `snotel.csv`, and student starter MATLAB script.

Recommended: instructor notes and a separately distributed, tested solution script. Before public distribution, verify the dataset's reuse permissions and document the original station records, time span, and download source. The supplied assignment identifies the dataset and columns but does not provide complete provenance. Add the course statistics textbook's author, title, edition, and publication year once confirmed; these details were not included in the supplied materials.

Background references: USDA NRCS SNOTEL program, <https://www.nrcs.usda.gov/>; MathWorks

documentation on logical arrays and probability distributions, <https://www.mathworks.com/help/>. For conceptual background, see a standard introductory probability or environmental statistics text selected by the instructor.

Answer guidance and validation

Q1: $P(R \cup L) = P(R) + P(L) - P(R \cap L)$ and $P(R \cap L) = 1 - P(R^c \cup L^c)$. The exclusive-or event means exactly one of the two stations is wet.

Q2: $E_1 = (R^c \cap L^c) \cup S$: the satisfying binary patterns $[R, L, S]$ are 000, 001, 011, 101, 111. Verify by applying the logical expression to all eight possibilities.

Q3: $P(R | L) = P(R \cap L)/P(L)$ and $P(L | R) = P(R \cap L)/P(R)$. The denominators differ; students must interpret the conditioning population.

Q4: $P(R \cup L \cup S) = P(R) + P(L) + P(S) - P(R \cap L) - P(R \cap S) - P(L \cap S) + P(R \cap L \cap S)$. Under the stated separate independence assumptions, $\alpha = 1 - P(L) - P(S)$; do not infer independence of R and $L \cap S$.

Q5: The four Brighton/Ben Lomond combinations are mutually exclusive and exhaustive. Summing $P(S \cap E_i)$ must equal the directly counted $P(S)$.

Q6: For $f = sp/[sp + q(1 - p)]$, with p, q fixed and s increasing to 1, the maximum is $p/[p + q(1 - p)]$ when the denominator is positive. The condition $f = 1$ is $q(1 - p) = 0$ with $sp > 0$; explicitly discuss the zero-alert edge case.

Q7: Accept any correctly defined and reproducible probability observation with a numerical estimate and a stated limitation. An example is Snowbird-only wet days, $P(S \cap R^c \cap L^c)$.

Testing: The companion instructor MATLAB script contains complete computations and `assert` checks for event identities. It should be executed against the redistributed dataset before teaching; the numerical answers depend on that file and its missing-data treatment.

Rubric application

Use the point totals in the rubric below. Within each question, assess (i) a correct event definition and method, (ii) transparent, executable MATLAB implementation, (iii) accurate calculations and requested validation, and (iv) a scientifically defensible explanation. Provide targeted feedback for incorrect denominators, misclassified outcomes, or unexamined independence assumptions.