Problem 1 (UFO sighting durations)

The National UFO Reporting Center (NUFORC) is an organization in the United States that investigates UFO sightings and/or alien contacts. NUFORC has been in continuous operation since 1974 when it was founded by Robert J. Gribble. It has catalogued almost 90,000 reported UFO sightings over its history, most of which were in the United States.

The UFO_scrubbed.RData file contains data on 64,506 UFO sightings over the last century. The data contain lon/lat coordinates, state, shape of the UFO, the date and time as well as the duration (in seconds) of the sighting. Your goal in this problem is to assess the spatial structure (if any!) of sighting durations.

UFO sighting durations are highly skewed, so analyze ln(duration). Note that to estimate or fit any semivariograms, the size of the data is preclusive – you will need to sub-sample fewer locations to do estimation.

(a) Provide a plot of all sighting locations with the US map overlaid – what do you notice generally about locations of sightings?

Choose five states other than Colorado. For each of these six states, including CO,

(b) Make heat maps of log-duration of sightings using the quilt.plot function in the fields package and comment on visual patterns.

(c) Generate and plot binned-semivariograms for each of these states – do any have particularly strong evidence of correlation of sighting durations?

(d) Using OLS or WLS, estimate a relevant spatial model for each of the six states, and report parameter estimates. Calculate the nugget-to-sill ratio and comment on its size for each state.

For the last two parts you don’t need to do any calculations, but carefully comment and respond to the following questions.

(d) If you wanted an automated procedure to detect which state has the strongest signal of spatial dependence, how would you do it?

(e) You want to see a UFO, but also want to maximize the duration of the experience. How would you use these data to find a “best” location?
Problem 2 (Covariance and prediction)

Suppose \( X(t), t \in \mathbb{R} \) is a mean zero isotropic Gaussian process with covariance function \( C(r) \). We observe \( X \) at two locations, \( t = -1 \) and \( t = 1 \), and want to predict \( X \) at \( t = 0 \). Set up a predictor as follows:

\[
\hat{X}(0) = w_1 X(-1) + w_2 X(1).
\]

(a) Derive optimal weights \( w_1 \) and \( w_2 \) that minimize \( \mathbb{E}(X(0) - \hat{X}(0))^2 \) (where, note, \( X(t) \) is random at \( t = -1, 0, 1 \)).

(b) Find the exact distribution of the random variable \( \hat{X}(0) \).

(c) Specialize your answer from part (a) for the exponential case \( C(r) = e^{-|r|} \).

Problem 3 (Bochner’s Theorem; graduate students only)

A function \( C(r) \) is a positive definite function on \( \mathbb{R} \) (and thus is a covariance function for some stochastic process) if and only if

\[
C(r) = \int_{\mathbb{R}} f(\omega) \exp(i\omega r) d\omega
\]

for some nonnegative integrable function \( f \geq 0 \) (\( i \) is the imaginary number). In this case, \( f(\omega) \) is known as the spectral density for \( C \), and we have an inversion formula

\[
f(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} C(r) \exp(-i\omega r) dr.
\]

For real-valued processes (i.e., the ones we care about), the inversion formula is simpler:

\[
f(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} C(r) \cos(\omega r) dr.
\]

Using this inversion formula, determine which of the following functions are positive definite on \( \mathbb{R} \) (that is, determine which of the following has a nonnegative spectral density \( f \)):

(a) \( e^{-|r|} \cos r \).

(b) \( (1 + r^2)^{-1} \).

(c) \( (1 - r^2)_+ \).

(Note that the full Bochner’s Theorem doesn’t require that a density exist with respect to the Lebesgue measure, and applies to positive definite functions in arbitrary dimensions.)