Please submit the codes (python/ jupyter notebook) and a pdf with the plotting results and corresponding comments. For the non-coding exercise, pdf of typed or written report is accepted.

Submit the required files to
Blackboard/Course Content/Tutorials/Environmental Data Analysis/L11_Linear_Model

Section 1: Coding Exercise

1.1

Use the dataset in `linedata01.txt`. Write a python scripts to least-squares fit polynomials of degree 2, 3 and 4 to the data (i.e. $\cancel{y = ax + b}$; $ax^2 + bx + c$; ...) . Make plots that show the observed and predicted data. Display the value of each coefficient and its 95% confidence limits. Comment on your results.

1.2

Modify the tutorial python script, section [4.8 (eda04_11)], to try to achieve a better fit to the Black Rock Forest temperature dataset.

a) Add additional periods of $\frac{T_y}{2}$ and $\frac{T_y}{3}$, where T_y is the period of 1 year, in an attempt to better capture the shape of the annual variation. Plot the results.
b) In addition to the periods in part A, add additional daily periods of $T_d, \frac{T_d}{2}$ and $\frac{T_d}{3}$, where T_d is the period of 1 day. Plot the results.
c) How much does the total error change in the cases?
[Hint: If the error goes up, your code has a bug!]
d) How much do the slope, m_2, and its confidence intervals change?

1.3

Suppose that a person wants to determine the weight, m_j, of M = 40 objects by weighting the first, and then weighing the rest in pairs: the first plus the second, the second plus the third, the third plus the fourth, and so on…

a) What is the corresponding data kernel, **G**? What is the percentage of non-zero elements in **G**?
b) Write a python script that creates this data kernel and computes the covariance matrix, $\boldsymbol{C}_m$, assuming that the observations are uncorrelated and have a variance, $\sigma_d^2 = 1\ kg^2$
c) Make a plot of σ_{mj} as a function of the object number, j, and comment on the results.

Section 2: Non-coding Exercise

1.4

Consider the equation $d_i = m_1 e^{-m_2 t_i}$.

a) Can this equation be arranged in the linear form: **d = Gm**?
b) Show that the equation can be *linearized* into the form, **d' = G'm'**, where the primes represent new, transformed variables. [Hint: Ex L08.3]
c) What would have to be true about the measurement error in order to justify solving this linearized problem by least square?
[Hint: What is the assumption of using least square method and what is the probability density function of the **d'** ?]
[Comment: Notwithstanding your answer, this problem is often is often solved with least squares in a *let's hope-for-the-best* mode.]

1.5 *[Optional for undergraduate (ESSC4510) students]*

a) What is the relationship between the elements of the matrix, $\boldsymbol{G^T G}$, and the columns, $c^{(j)}$, of **G**?
b) Under what circumstance is $\boldsymbol{G^T G}$ a diagonal matrix?
c) What is the form of the covariance matrix in this case?
d) What is the form least-squares solution in this case? Is it harder or easier to compute than the case where $\boldsymbol{G^T G}$ is not diagonal?
e) Examine the straight-line case in this context?