**Plan for data cleaning:**

**Greenhouse gas data: (this data comes into the pipeline pre-cleaned because we need to process it in SoilFluxPro, another software, first)**

**File names: Node 1\_cleaned\_021722.xlsx; Node 2\_cleaned\_021722.xlsx; Node 3\_cleaned\_021722.xlsx; Node 4\_cleaned\_021722.xlsx**

-Import one node at a time

-Consolidate columns to include Date/Timestamp, N2O flux, N2O concentration, CO2 flux, CO2 concentration

-Convert Date/Time from UTC to Central Time (*does not account for Daylight Savings*)

-Make an additional column with only Date

-Pull together all nodes into one file (have a column that corresponds to Node)

-Look at summary plots of each node

-Remove any outliers we might’ve missed

**Soil oxygen data (IN CENTRAL TIME):**

**File name: O2 temp and VWC data\_all nodes\_June21-Feb22.xlsx**

-Import one node at a time

-Consolidate columns to include data from all five sensors at 5cm and at 15cm (10 columns of sensor data), and Date/Timestamp

-Remove NAs, NaN

-Calculate averages across all five sensors at each depth and consolidate columns further (one mean column for 5cm, one for 15cm, and one for Date/Timestamp)

-Look at summary plots

-Remove all values greater than 22

-Remove all values less than 0

-Remove specific outliers

**Soil VWC data (IN CENTRAL TIME):**

**File name: O2 temp and VWC data\_all nodes\_June21-Feb22.xlsx**

-Import one node at a time

-Consolidate columns to include data from all five sensors at 5cm and at 15cm (10 columns of sensor data), and Date/Timestamp

-Remove NAs, NaN

-Calculate averages across all five sensors at each depth and consolidate columns further (one mean column for 5cm, one for 15cm, and one for Date/Timestamp)

-Look at summary plots

-Remove outliers (there are no obvious ranges that are or are not acceptable)

**Soil temperature data (IN CENTRAL TIME):**

**File name: O2 temp and VWC data\_all nodes\_June21-Feb22.xlsx**

-Import one node at a time

-Consolidate columns to include data from all five sensors at 5cm and at 15cm (10 columns of sensor data), and Date/Timestamp

-Remove NAs, NaN

-Calculate averages across all five sensors at each depth and consolidate columns further (one mean column for 5cm, one for 15cm, and one for Date/Timestamp)

-Look at summary plots

-Remove outliers (there are no obvious ranges that are or are not acceptable)

**Atmospheric pressure, relative humidity, air temperature (IN CENTRAL TIME):**

**File name: Eddyflux\_air temp\_pressure\_RH\_.Oct20-March22.xlsx**

**-This data is irrespective of node (eddy flux chambers correspond to the whole site)**

**-This dataset includes air temperature, atmospheric pressure, and relative humidity**

**-Consolidate date and time columns into one column**

**- Consolidate columns to include Date/Timestamp, air temperature (column AN), atmospheric pressure (column AO), and % relative humidity (column AX).**

**Atmospheric pressure:**

-Convert atmospheric pressure from Pa to KPa (divide by 1000)

-Remove all -9999

-Remove all NA or NaNs

-Look at summary plots

-Remove outliers (only include pressure values that range from 90 to 110 kPa)

**Relative humidity:**

-Remove all -9999

-Remove all NA or NaNs

-Look at summary plots

-Remove outliers (there are no obvious ranges that are or are not acceptable)

**Air temperature:**

-Remove all -9999

-Remove all NA or NaNs

-Convert from Kelvins to Celsius (subtract 273.15)

-Look at summary plots

-Remove outliers (there are no obvious ranges that are or are not acceptable)

**Match O2 data to atmospheric pressure, relative humidity, air temperature:**

-Match the Date/Timestamp for O2 at 5cm and O2 at 15cm for each node to the Date/Timestamp corresponding to atmospheric pressure, relative humidity, and air temperature (the Date/Timestamp will be the same for atmospheric pressure, relative humidity, and air temperature for each sampling point).