gasmet_data-conversion_2023

jesse puka-beals

2023-04-21

Overview

Gasmet measurements are recorded using calcmet software

This software generates .SPE files that autosaves into a folder named by date. For example folder 20221027 is for the most recent sampling on October 27th. Each time someone takes measurements on a new day, the data from the last sampling date is compressed. For example, there was once a folder called 20221021 but it was compressed into "SAMPLE20221021_1" on october 27th when new samples were recorded.

Our objective is to get all of the data into a single data frame that has the date, the experiment, the field, the plot and the predicted gas concentrations (i.e. CO2 in ppm).

Scope

only 2023 data

Workflow steps

1. convert the cabinet files (a compressed file format) into an unzipped and uncompressed format
2. load the SPE files into calcmet and have calcmet predict the gas concentrations of interest and output those predictions into large files
3. combine those large files of predicted values into a single file
4. tidy that large file
5. done

convert cabinet files

There was only one cabinet file…

load SPE files into calcmet and generate outputs

There were only 1922 files for 2023 as of 21Apr2023, so we only need one batch.

Need this fileEncoding arg to use read.delim()…

```
read.delim(
  "2023data_20Apr2023.TXT",
  fileEncoding="ISO-8859-1") -> all_batches
```

tidy large file

Now we have one large file!

Time to clean up some column names

```
dat <- all_batches
# View(dat)

dat1 <- dat %>%
  dplyr::select(Date,Time,SpectrumFile,Carbon.dioxide.CO2,
                Nitrous.oxide.N2O,Methane.CH4,
                Ammonia.NH3,Cell.temperature) %>%
  rename(date=Date,
         time=Time,
         co2=Carbon.dioxide.CO2,
         spectrumfile=SpectrumFile,
         n2o=Nitrous.oxide.N2O,
         ch4=Methane.CH4,
         nh3=Ammonia.NH3,
         temp=Cell.temperature) %>%
  rowid_to_column()

# View(dat1)
```

Now we want to split out that Spectrum file name

The root for all these batches is the same so this will be super easy in the instances where the naming convention was followed

```
library(stringr)

# so all these files are from the same folder so this will
dat1$spectrumfile[1]
```

```
## [1] "D:\\Shared drives\\SEECRS (Gutknecht) lab\\Data\\Gasmet data\\2023\\spe_compiled\\R_ncas_i02_202
```

```
str_sub(dat1$spectrumfile,76,-5)[1]
```

```
## [1] "R_ncas_i02_2023-04-13_2405_00000"
```

```
dat1 %>%
  mutate(filename = str_sub(spectrumfile,76,-5),
         .before=date) -> dat2
```

```
dat2$filename[1]
```

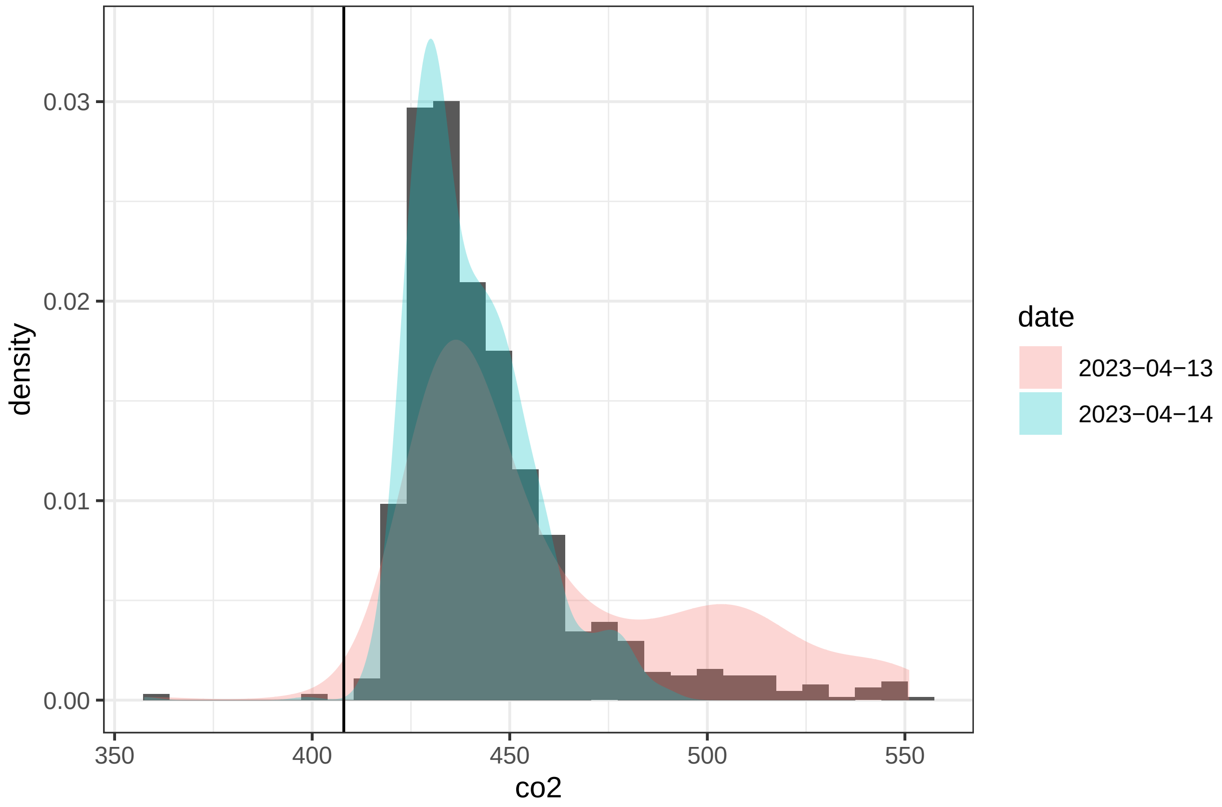
```
## [1] "R_ncas_i02_2023-04-13_2405_00000"
```

```
str_split(dat2$filename,
          pattern = "_",
          simplify = T) %>%
  as.data.frame() %>%
  dplyr::rename(experiment = V1,
                site = V2,
                date_check = V3,
                plot = V4,
                calcmet_chronological_tally = V5) %>%
  cbind(dat2) -> dat3

# There are some measurements with an extra "R_".
# According to Angela, this was done automatically by calcmet

# When I filter CO2 values must be greater than 100, all these "R_" value
# disappear, suggesting this is just junk data, but it messes up the above code
# so below I tidy up the column names to prep for export

dat3 %>%
  filter(co2>100) %>%
  dplyr::select(-V6) -> dat4

# dat4 %>%
  # write.csv("gasmet_2023_ncas.csv",row.names=F)
```

Data vis

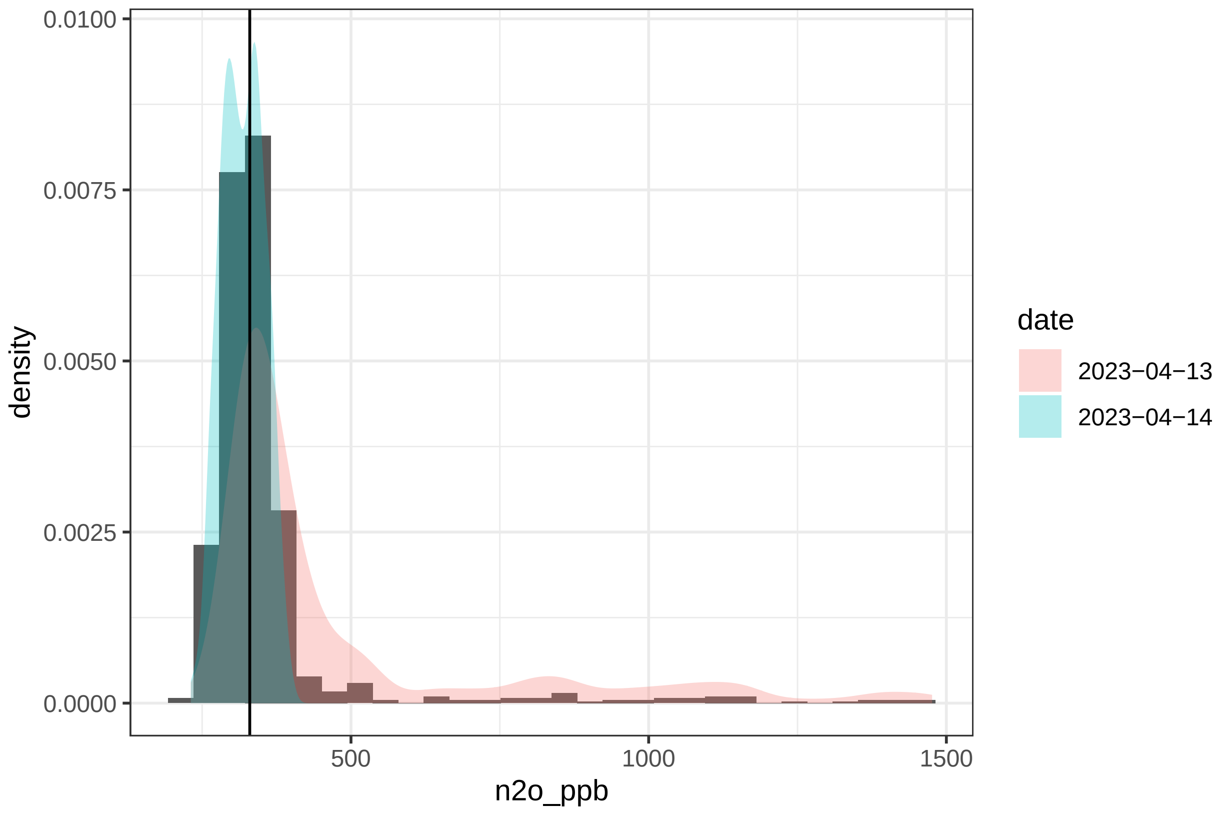
Time to enjoy the fruits of our labor

```
dat4 %>%
  filter(co2>350) %>%
  # glimpse()
  ggplot(aes(co2)) +
  stat_bin(aes(y=after_stat(density))) +
  geom_density(aes(fill=date),
               alpha=.3,
               col=NA) +
  theme_bw() +
  geom_vline(
    xintercept = 408
  ) +
  labs(
    caption = "2017 atmospheric CO2 was ~408 parts per million"
  )
```



2017 atmospheric CO2 was ~408 parts per million

```r
dat4 %>%
  filter(co2>350) %>%
  mutate(
    n2o_ppb = n2o*1000
  ) %>%
  # glimpse()
  ggplot(aes(n2o_ppb)) +
  stat_bin(aes(y=after_stat(density))) +
  geom_density(aes(fill=date),
               alpha=.3,
               col=NA) +
  geom_vline(
    xintercept = 330
  ) +
  theme_bw() +
  labs(
    caption = "2017 atmospheric N2o was ~330 parts per billion"
  )
```



2017 atmospheric N2o was ~330 parts per billion