ProbeSchedule: Latest Results
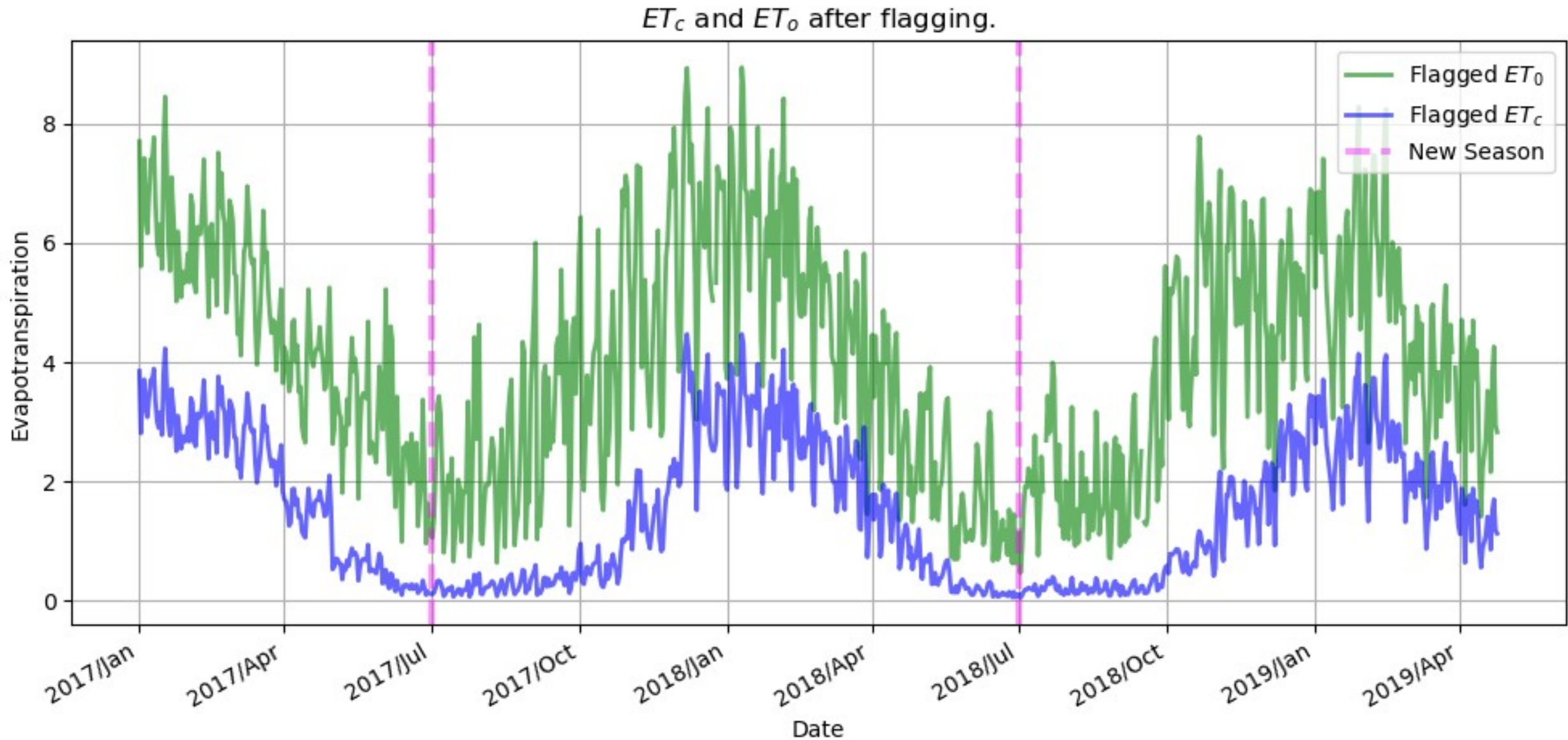
25 April 2019, 25/04/2019

**By *Henri Branken*
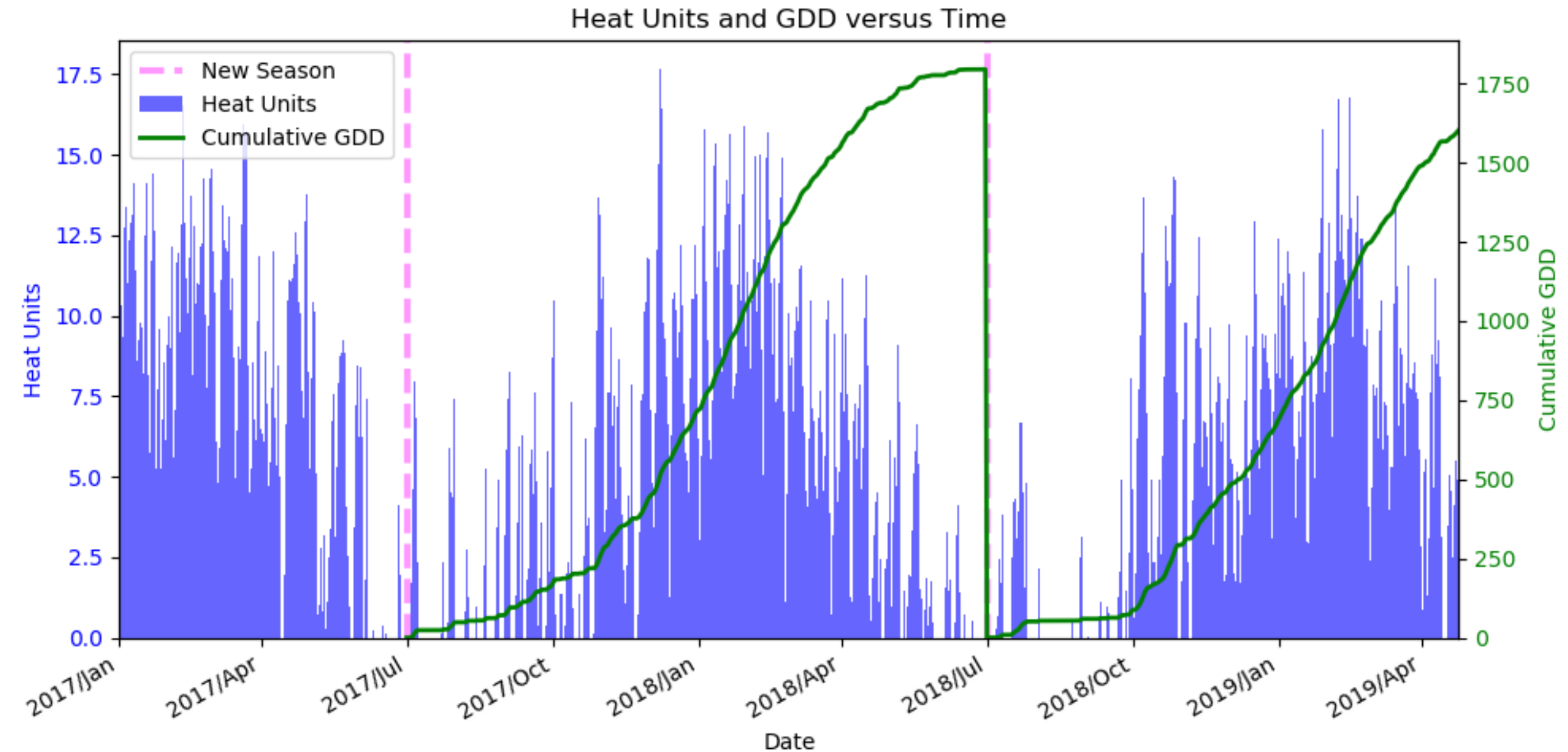MatogenAI**

Results of API Updates

ET_c and ET_o after flagging.

Results of API updates:

- As a result of eto and etc data having much better integrity, I **no longer flag consecutive etc entries that are identical.**

- I do, however, still flag for consecutive eto entries that are identical (more on this in a later slide).

Bar Graph Rectified

Heat Units and GDD versus Time
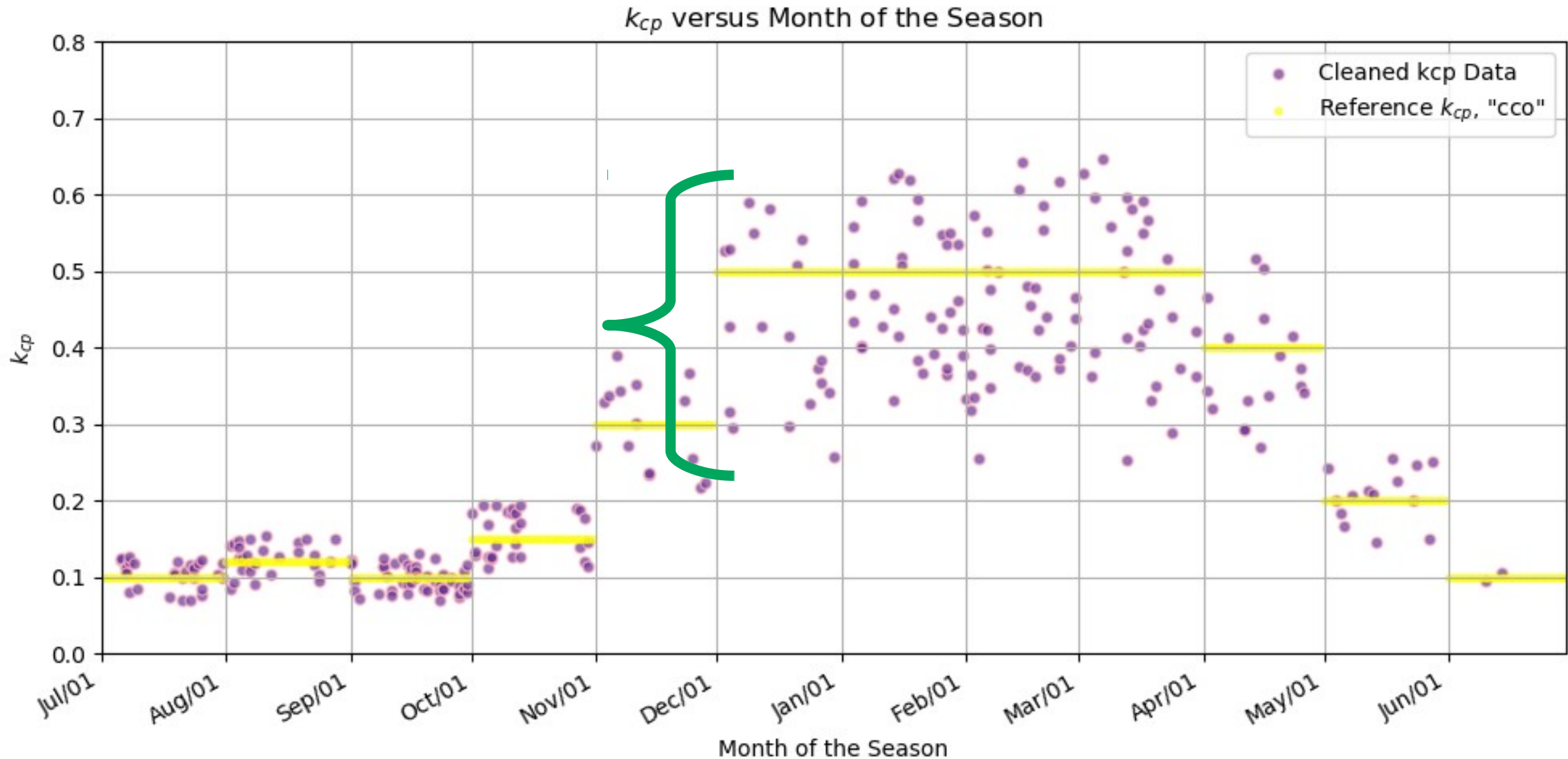
Fine-tuning etcp deviation

- By "*etcp deviation*" we mean the **percentage deviation** of **etcp away from** the expected crop water usage, **etc**.

- Setting the **maximum percentage deviation** above which we flag samples has a noticeable **effect on the final results**.

- In the following we compare results if **ETCP_PERC_DEVIATION** is **30%** and **50%**.

ETCP_PERC_DEVIATION = **30%**
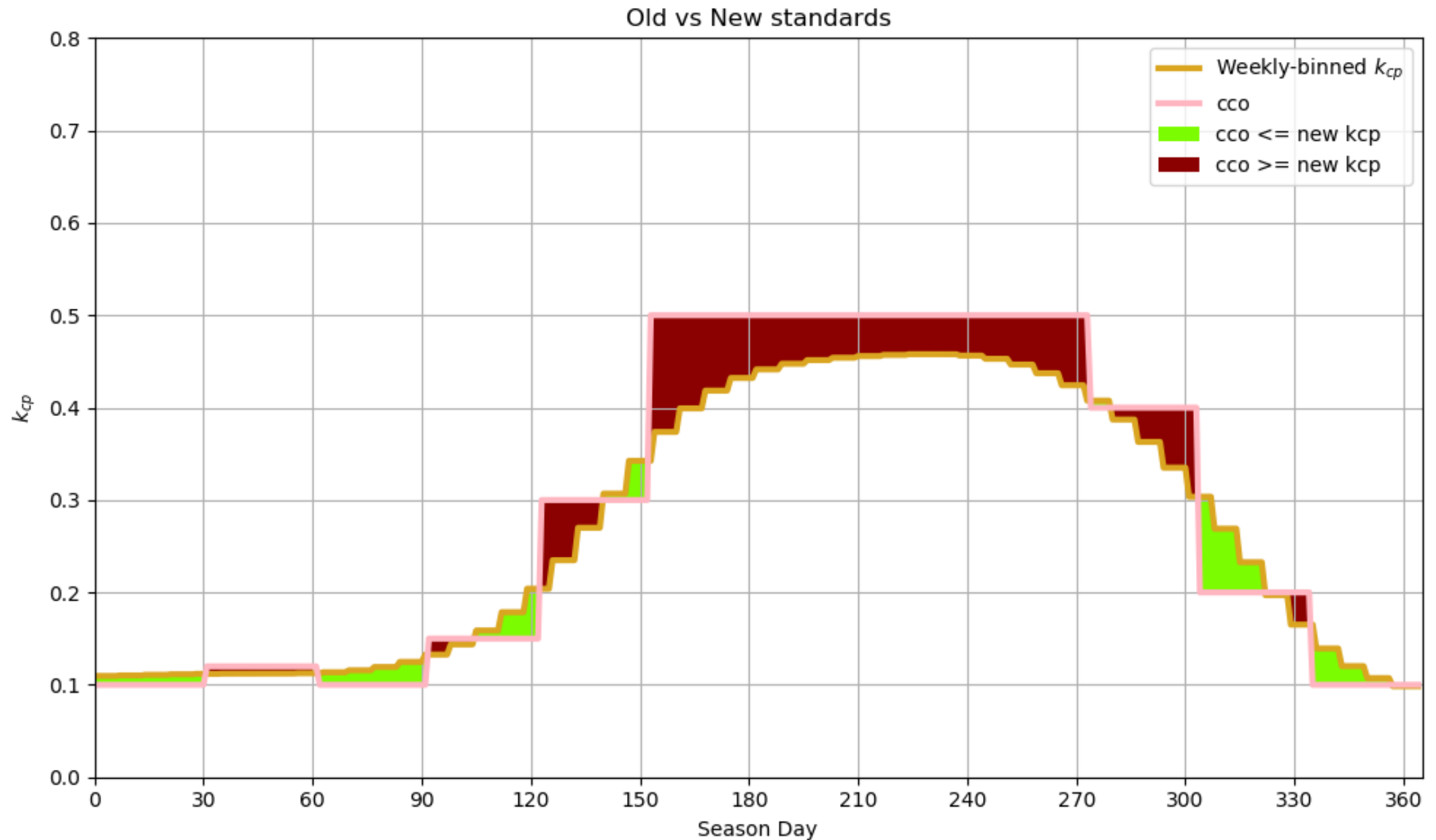
k_{cp} versus Month of the Season

ETCP_PERC_DEVIATION = **30%**

Old vs New standards

ETCP_PERC_DEVIATION = **30%**

k_{cp} versus GDD

ETCP_PERC_DEVIATION = **50%**

k_{cp} versus Month of the Season

ETCP_PERC_DEVIATION = **50%**

Old vs New standards

ETCP_PERC_DEVIATION = **50%**

k_{cp} versus GDD
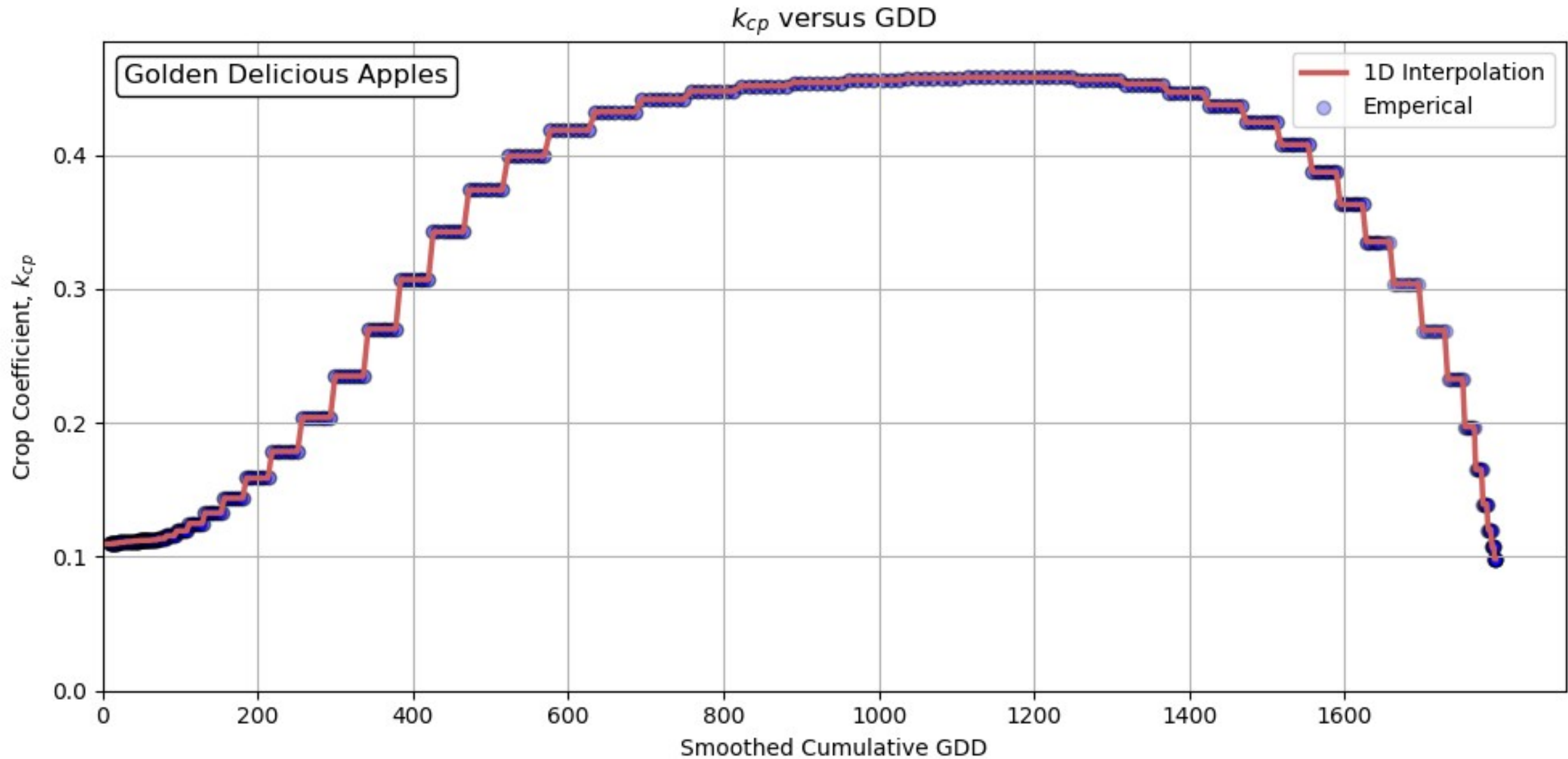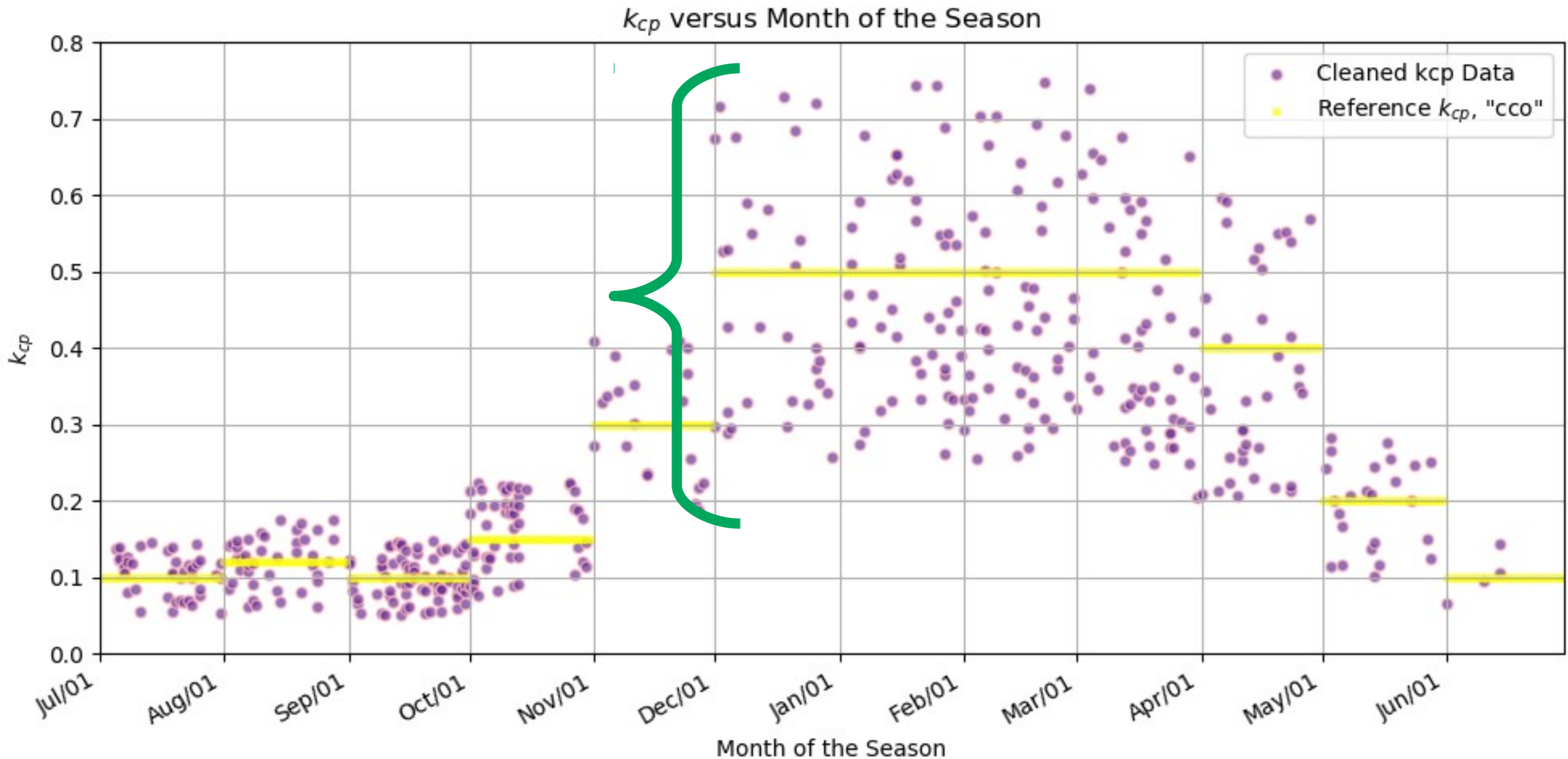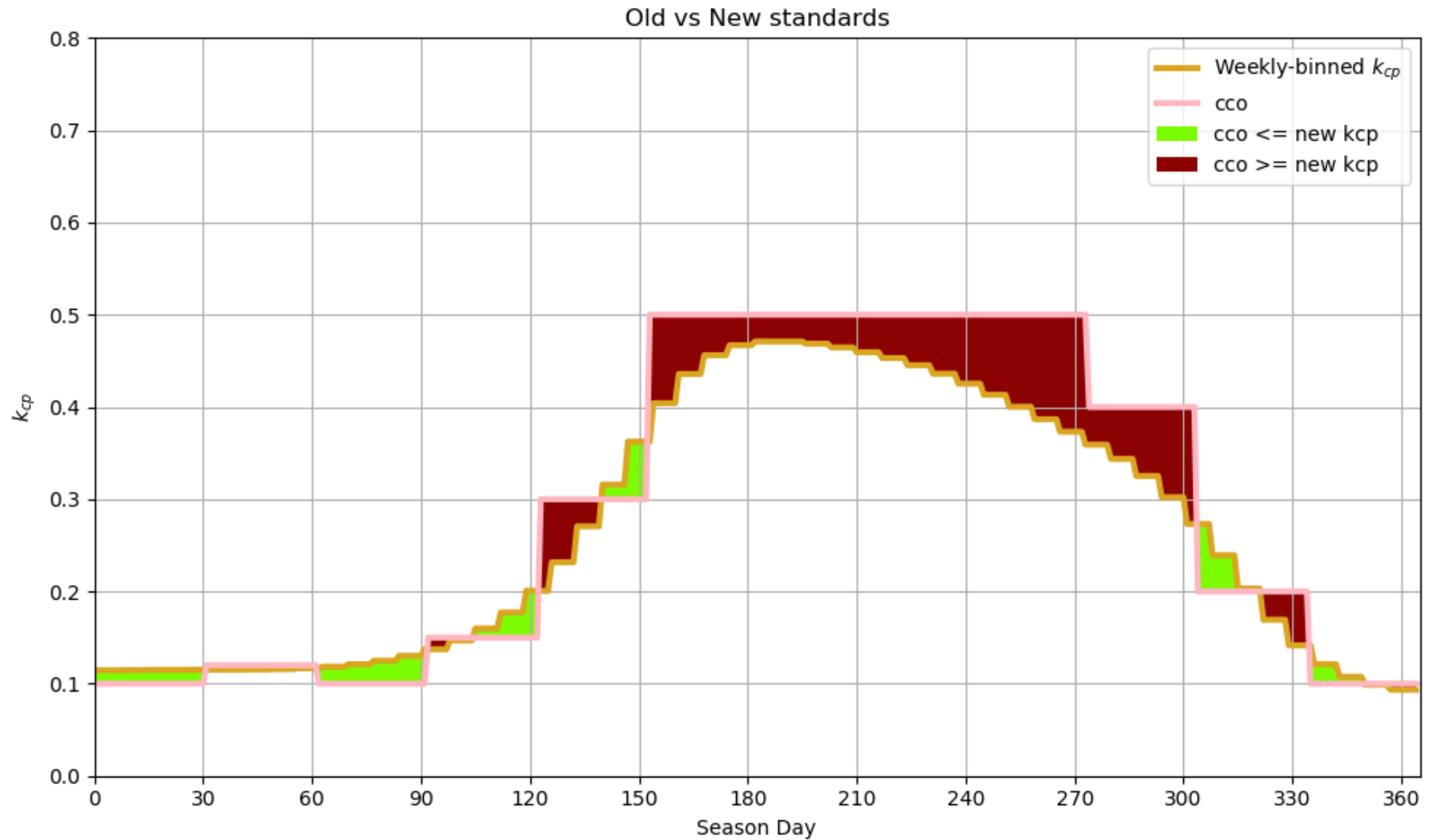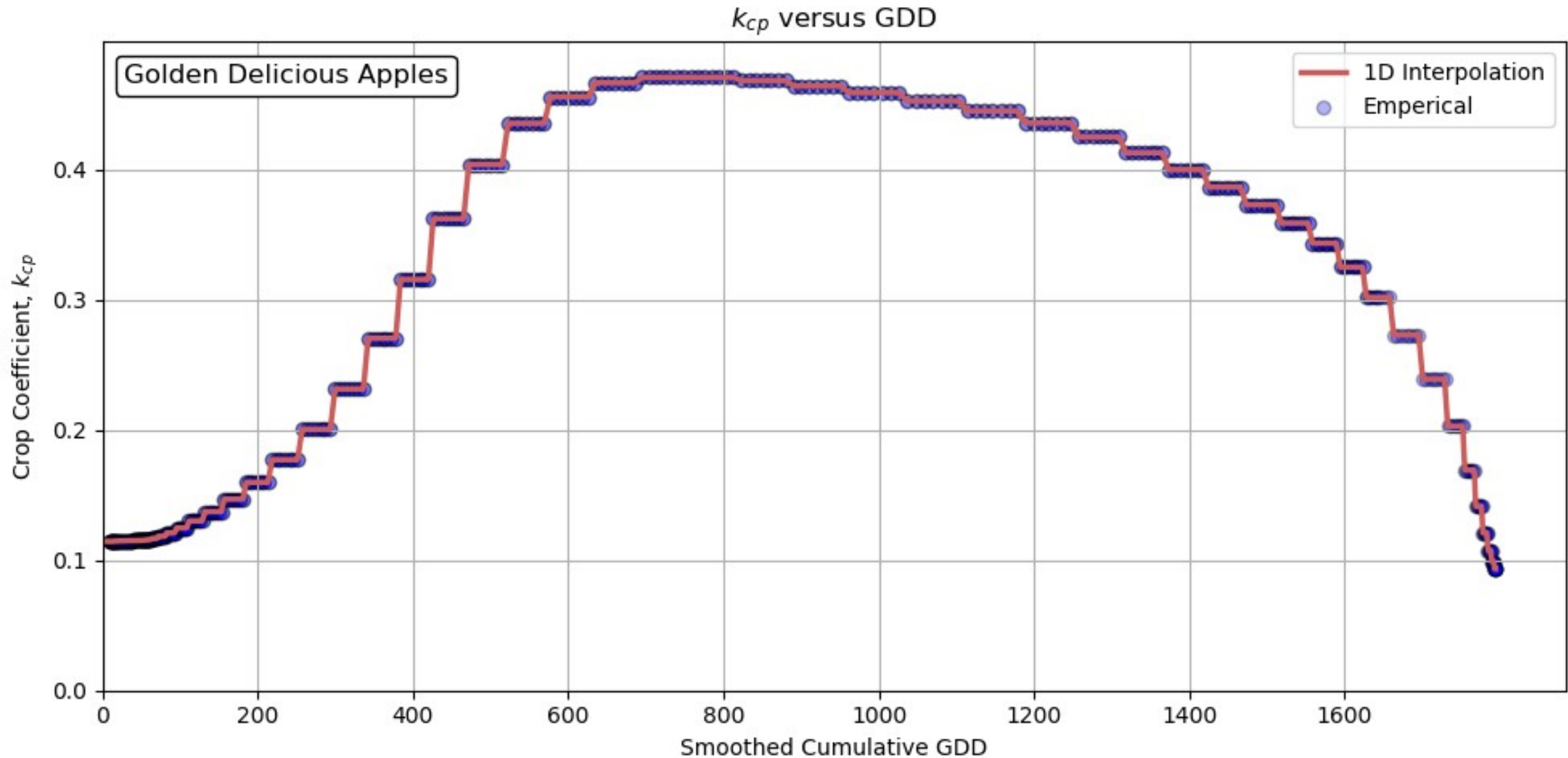
Left: **30%**. Right: **50%**.

```
+----------------------------------------------------------+      +----------------------------------------------------------+
|              Report for the entire SET of probes:        |      |              Report for the entire SET of probes:        |
+----------------------------------------------------------+      +----------------------------------------------------------+
| Rain perturbing etcp            |    12.09% |    714/5908 |      | Rain perturbing etcp            |    12.09% |    714/5908 |
| Software simulation             |    10.05% |    594/5908 |      | Software simulation             |    10.05% |    594/5908 |
| `Possible` irrigation           |     7.13% |    421/5908 |      | `Possible` irrigation           |     7.13% |    421/5908 |
| Null profile value              |    12.36% |    730/5908 |      | Null profile value              |    12.36% |    730/5908 |
| Profile data blip               |     0.91% |     54/5908 |      | Profile data blip               |     0.91% |     54/5908 |
| Large profile dip               |     1.51% |     89/5908 |      | Large profile dip               |     1.51% |     89/5908 |
| Etcp is positive                |    12.81% |    757/5908 |      | Etcp is positive                |    12.81% |    757/5908 |
| Etcp larger than allowed maximum|     0.00% |      0/5908 |      | Etcp larger than allowed maximum|     0.00% |      0/5908 |
| Etcp suspiciously low  ───────> |     6.96% |    411/5908 |      | Etcp suspiciously low  ───────> |     5.13% |    303/5908 |
| Etcp suspiciously high ───────> |    33.26% |   1965/5908 |      | Etcp suspiciously high ───────> |    31.77% |   1877/5908 |
| Ratio of etcp over eto exceeds KCP_MAX | 0.00% |  0/5908 |      | Ratio of etcp over eto exceeds KCP_MAX | 0.00% |  0/5908 |
| kcp deviation too big           |     0.17% |     10/5908 |      | kcp deviation too big           |     0.36% |     21/5908 |
| kcp is NaN                      |    95.04% |   5615/5908 |      | kcp is NaN                      |    91.91% |   5430/5908 |
| Faulty or Missing Heat Units    |     0.00% |      0/5908 |      | Faulty or Missing Heat Units    |     0.00% |      0/5908 |
| Imputed eto                     |     0.00% |      0/5908 |      | Imputed eto                     |     0.00% |      0/5908 |
| Stuck or faulty eto             |    18.20% |   1075/5908 |      | Stuck or faulty eto             |    18.20% |   1075/5908 |
| eto exceeds ETO_MAX             |     0.00% |      0/5908 |      | eto exceeds ETO_MAX             |     0.00% |      0/5908 |
| Faulty etc                      |     0.34% |     20/5908 |      | Faulty etc                      |     0.34% |     20/5908 |
| etc exceeds ETCP_MAX            |     0.00% |      0/5908 |      | etc exceeds ETCP_MAX            |     0.00% |      0/5908 |
+----------------------------------------------------------+      +----------------------------------------------------------+
| Only 4.96% of data, that is 293/5908 samples, are useful for the probe SET. |   | Only 8.09% of data, that is 478/5908 samples, are useful for the probe SET. |
+----------------------------------------------------------+      +----------------------------------------------------------+
```

Additional (more sophisticated) Logic when Refining the kcp Trendline

- For every iteration, verify whether the new smoothed **trendline better approximates** the **majority** of the probes' individual kcp datasets.
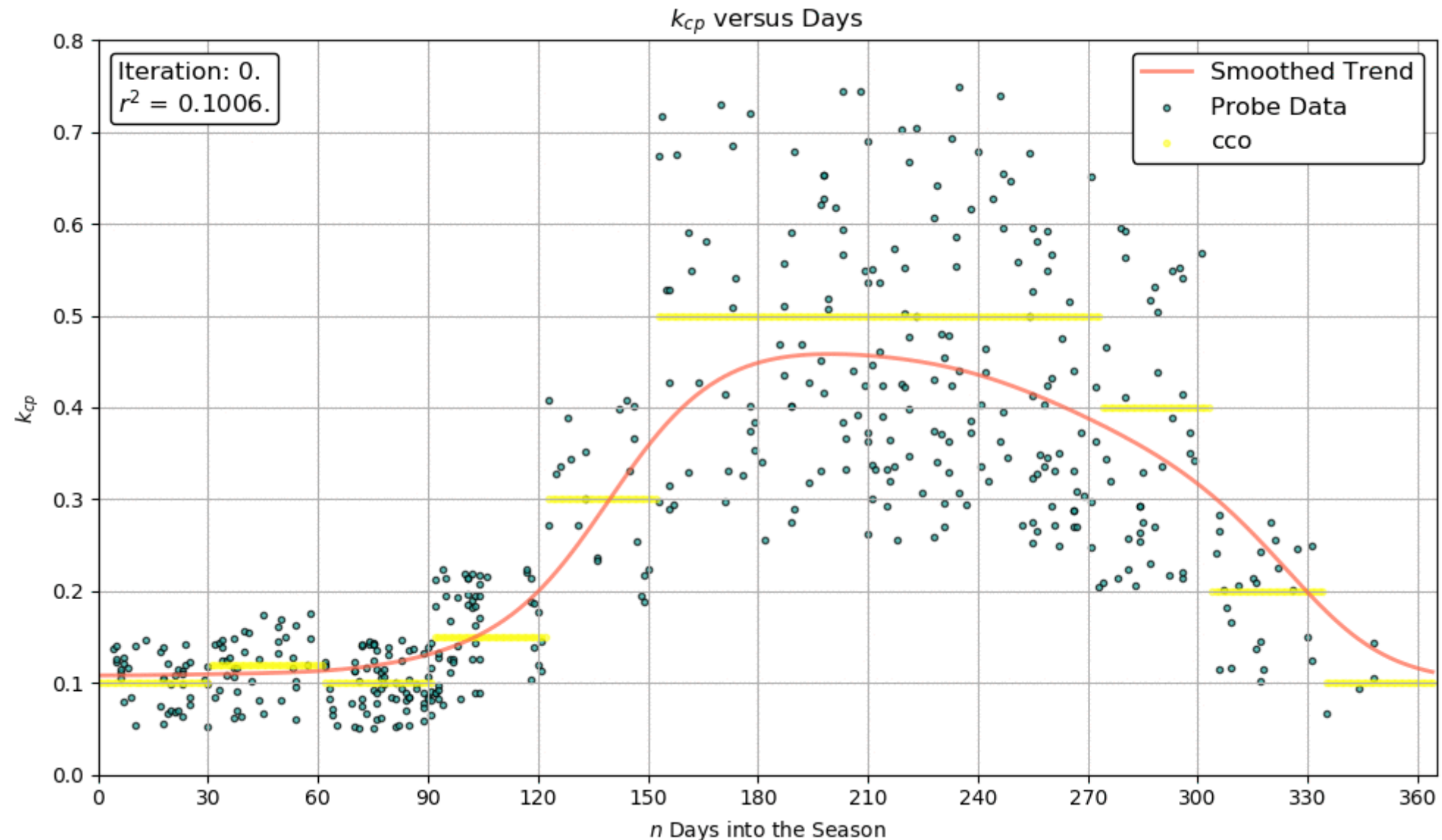
	Number of	Number of	Number of	
Iteration	healthy probes	Improvements	Deterioriations	n_imp >= n_det
[0, 1]	6	5	1	True
[1, 2]	5	5	0	True
[2, 3]	4	2	2	True
[3, 4]	3	2	1	True
[4, 5]	2	0	2	False

Probe Information
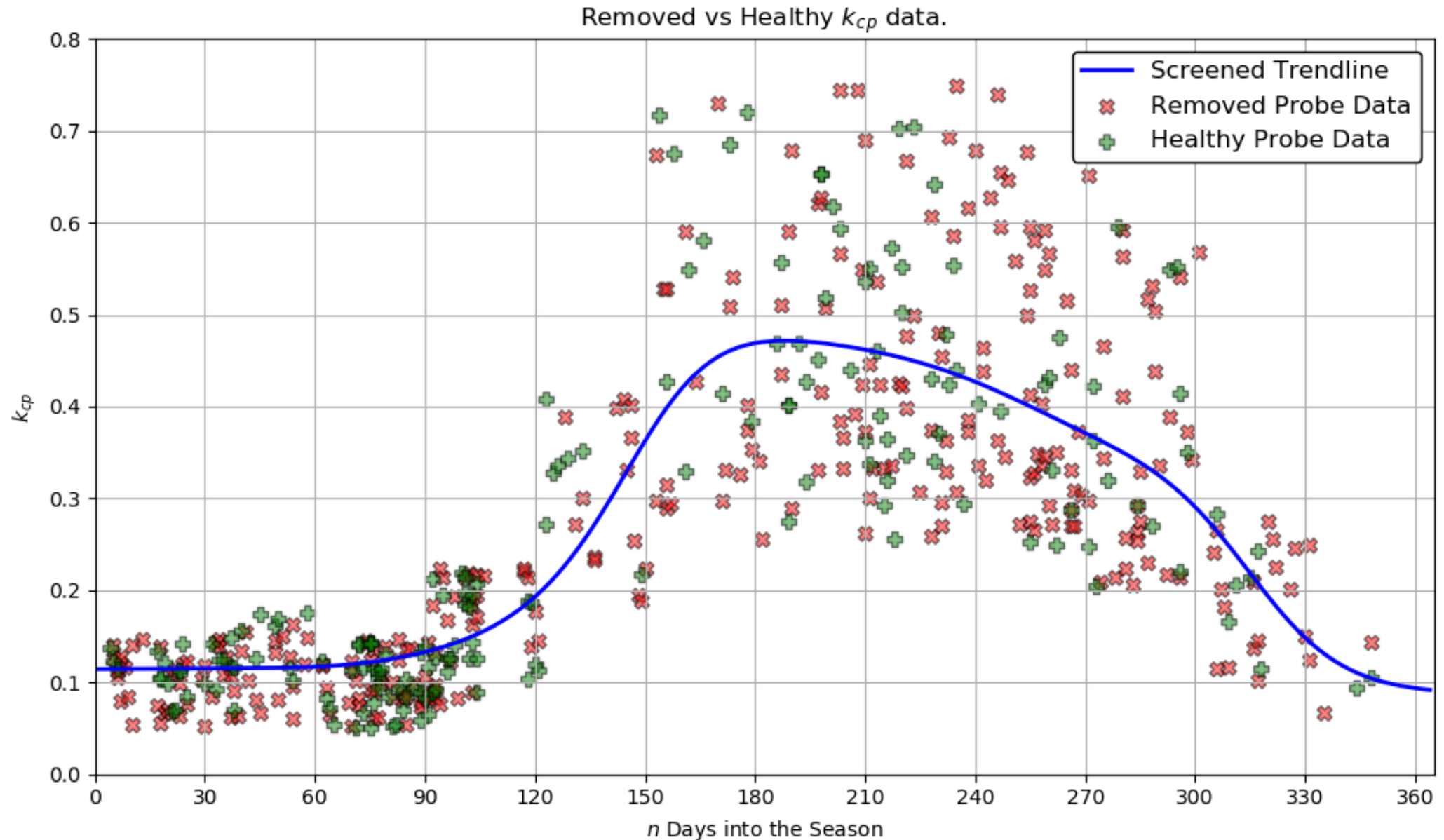
Complementary table to the previous one:

- Notice how we succeeded in removing 4 "bad" Probes from the dataset.
- Additionally, we also require the r-square between the trendline and scatter plot to decrease in order to proceed with a probe removal.
- Keep in mind a smaller r-square value means a "better" trendline fit.

```
+----------+--------------+--------------+--------------+------------------+-------------+
|                              Screening Report                                         |
+----------+--------------+--------------+--------------+------------------+-------------+
|          |  n scatter   |     cco      |   scatter    |      delta       |  probe_id   |
|  iter    |   points     |  r-squared   |  r-squared   |  scatter r-sqr   |  removed    |
+----------+--------------+--------------+--------------+------------------+-------------+
|      0 |          478 |    0.0518844 |    0.1005873 |              -   |          -  |
|      1 |          424 |    0.0492342 |    0.0992430 |       -0.0013444 |  ----> P-370|
|      2 |          384 |    0.0463118 |    0.0970331 |       -0.0022099 |  ----> P-392|
|      3 |          315 |    0.0448074 |    0.0954229 |       -0.0016102 |  ----> P-372|
|      4 |          179 |    0.0521010 |    0.0920653 |       -0.0033576 |  ----> P-384|
+----------+--------------+--------------+--------------+------------------+-------------+
```

.gif Animation showing progression of trendline (Notice Upper-Left Annotations)

k_{cp} versus Days

Iteration: 0.
$r^2 = 0.1006.$

Legend:
— Smoothed Trend
○ Probe Data
● cco

k_{cp} (vertical axis)

n Days into the Season (horizontal axis)

Comparing **Removed** versus **Remaining** datapoints.

Removed vs Healthy k_{cp} data.

Animation of remaining three "Healthy" Probes:

k_{cp} versus Days

ID: P-371.
Old $r^2 = 0.1017$.
New $r^2 = 0.0998$.
Improvement: True.

First, Original Trend
Latest Trend
cco
Probe Data

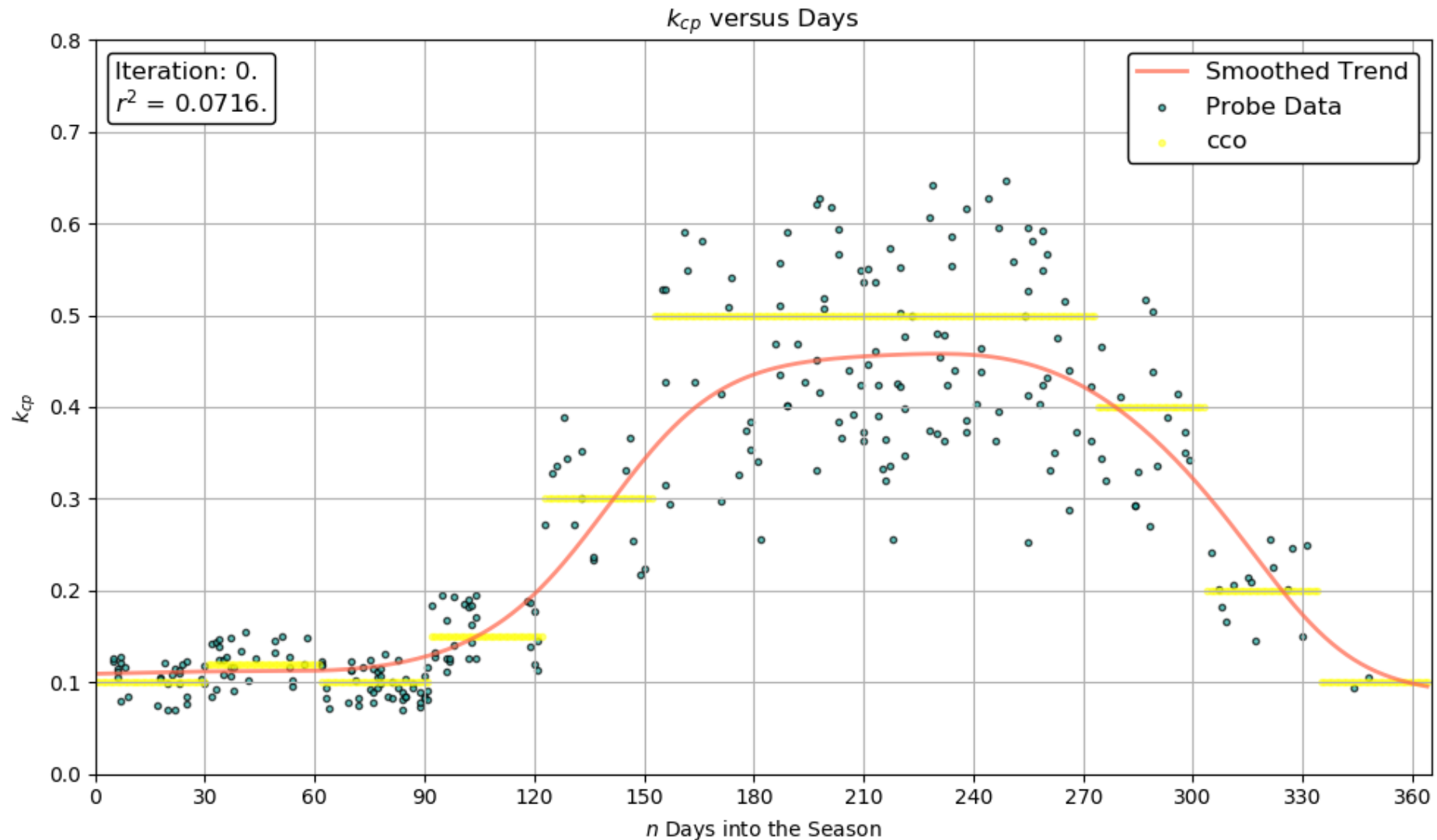k_{cp}

n Days into the Season
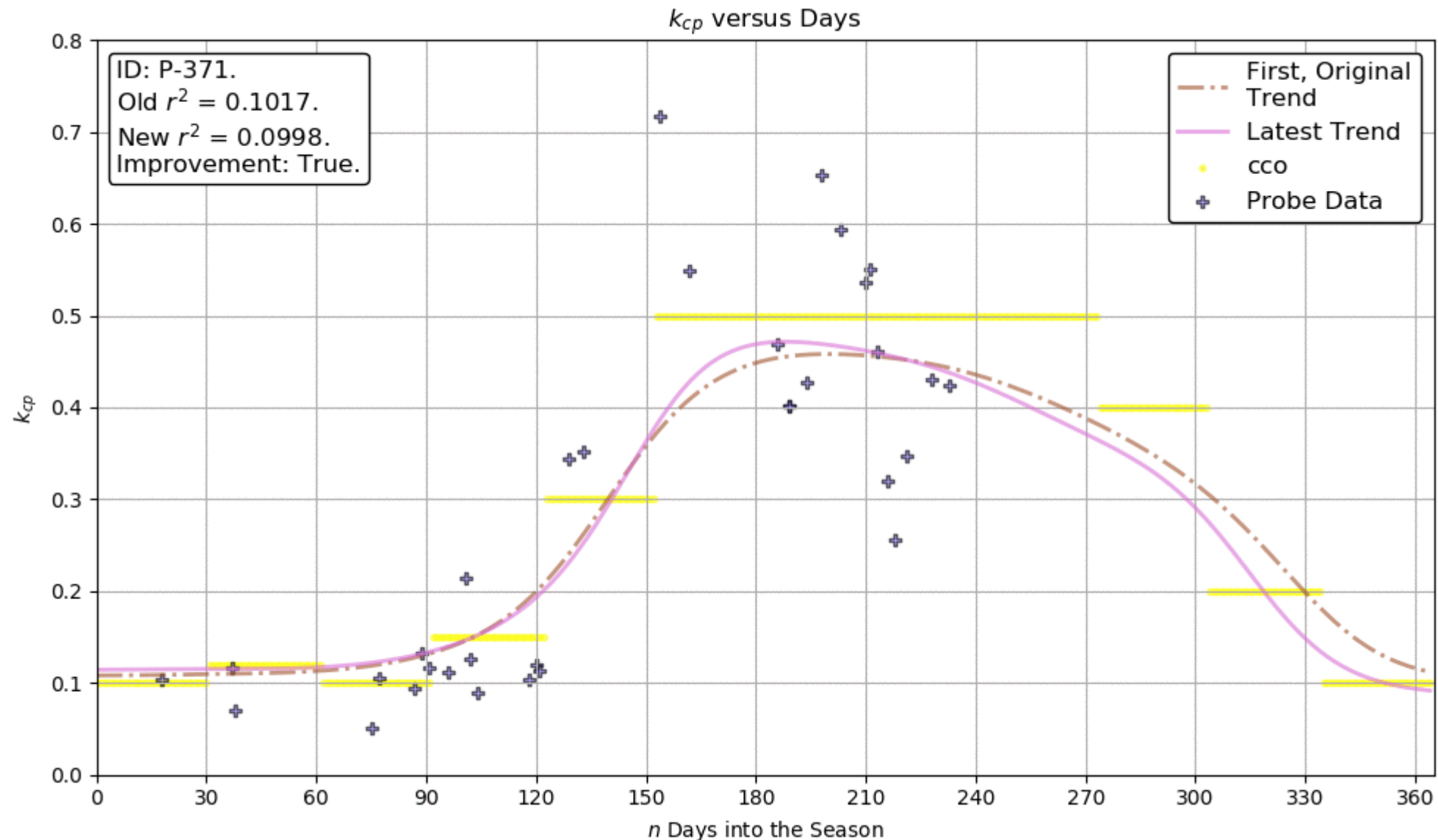
Previous 5 slides correspond to ETCP_PERC_DEVIATION = **50%**

- Results vary for **different values** of ETCP_PERC_DEVIATION.

- Consider the following table when ETCP_PERC_DEVIATION = **30%.**

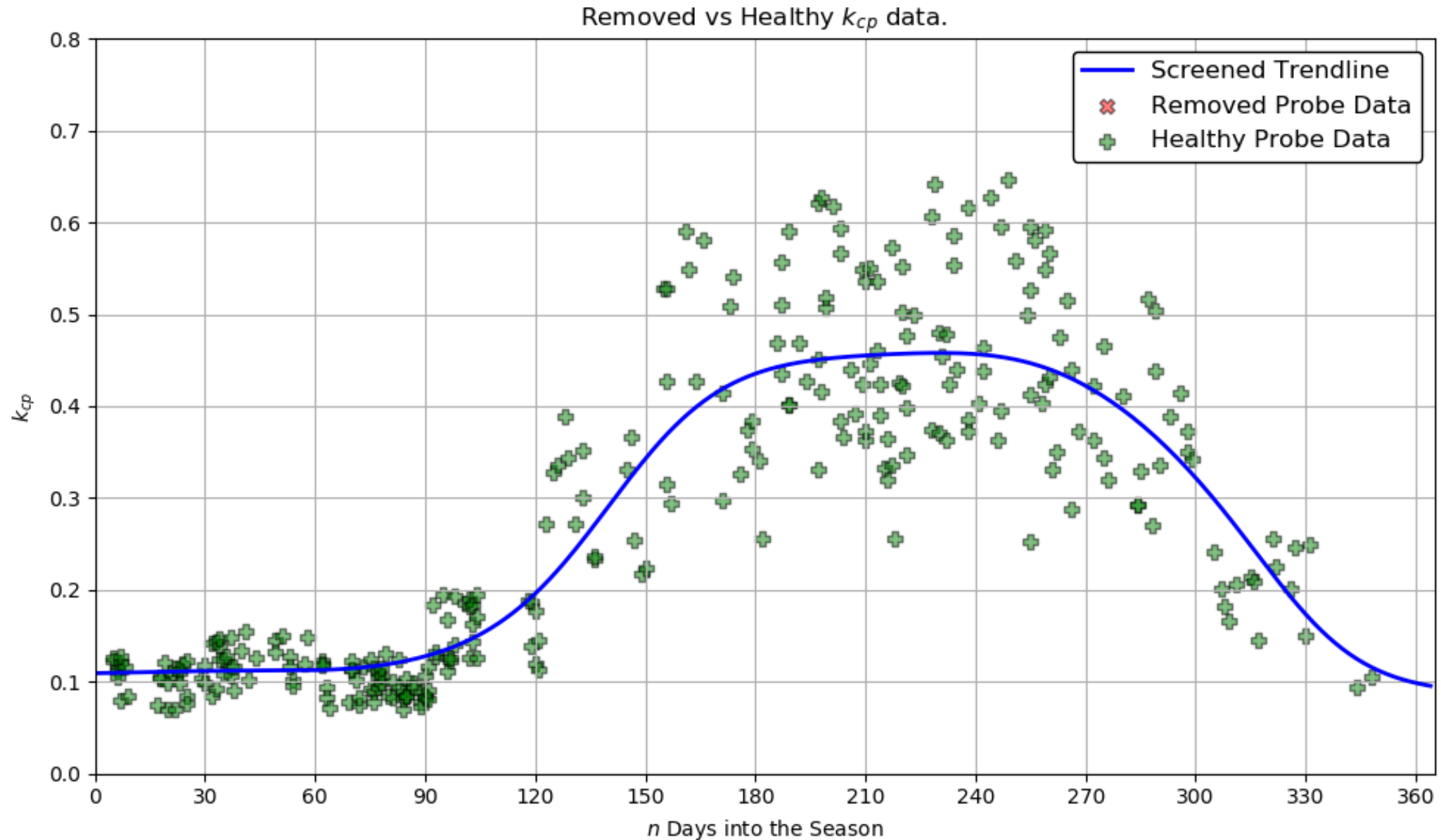- Notice that we cannot even remove 1 probe.

```
+-------------------------------------------------------------------------------------+
|                                  Probe Information                                   |
+-------------+------------------+------------------+------------------+---------------+
|             |    Number of     |    Number of     |    Number of     |               |
| Iteration   | healthy probes   | Improvements     | Deteriorations   | n_imp >= n_det|
+-------------+------------------+------------------+------------------+---------------+
|  [0, 1]     |        6         |        2         |        4         |    False      |
+-------------+------------------+------------------+------------------+---------------+
```

No .gif animation this time, because logic check is **not satisfied** at the first iteration.

Therefore we end up where we originally started.
(There is no change since the initial smoothed trendline.)

k_{cp} versus Days

Iteration: 0.
$r^2 = 0.0716.$

— Smoothed Trend
• Probe Data
• cco

k_{cp}

n Days into the Season

Comparing **"Removed"** versus **Remaining** datapoints for ETCP_PERC_DEVIATION = 30%.

Removed vs Healthy k_{cp} data.

Remaining Questions:

- @Jac, are you still satisfied with **RAIN_THRESHOLD = 2mm**? (In other words, flag all samples for which rain > 2mm. See the next slide.)

- Clearly, the results are sensitive depending on the value specified for **ETCP_PERC_DEVIATION**. Do you have a **preferred value in the [30%; 50%] range** of values?

- What is an optimal **API_START_DATE**? Currently, the API_START_DATE is set to **2017-01-01**.

- I have coded the **API_END_DATE** to be equal to **yesterday's** date. (I assume this is acceptable?)

Flagged Rain Events (Ouplaas, Kromfontein):

Rain reading versus Date.
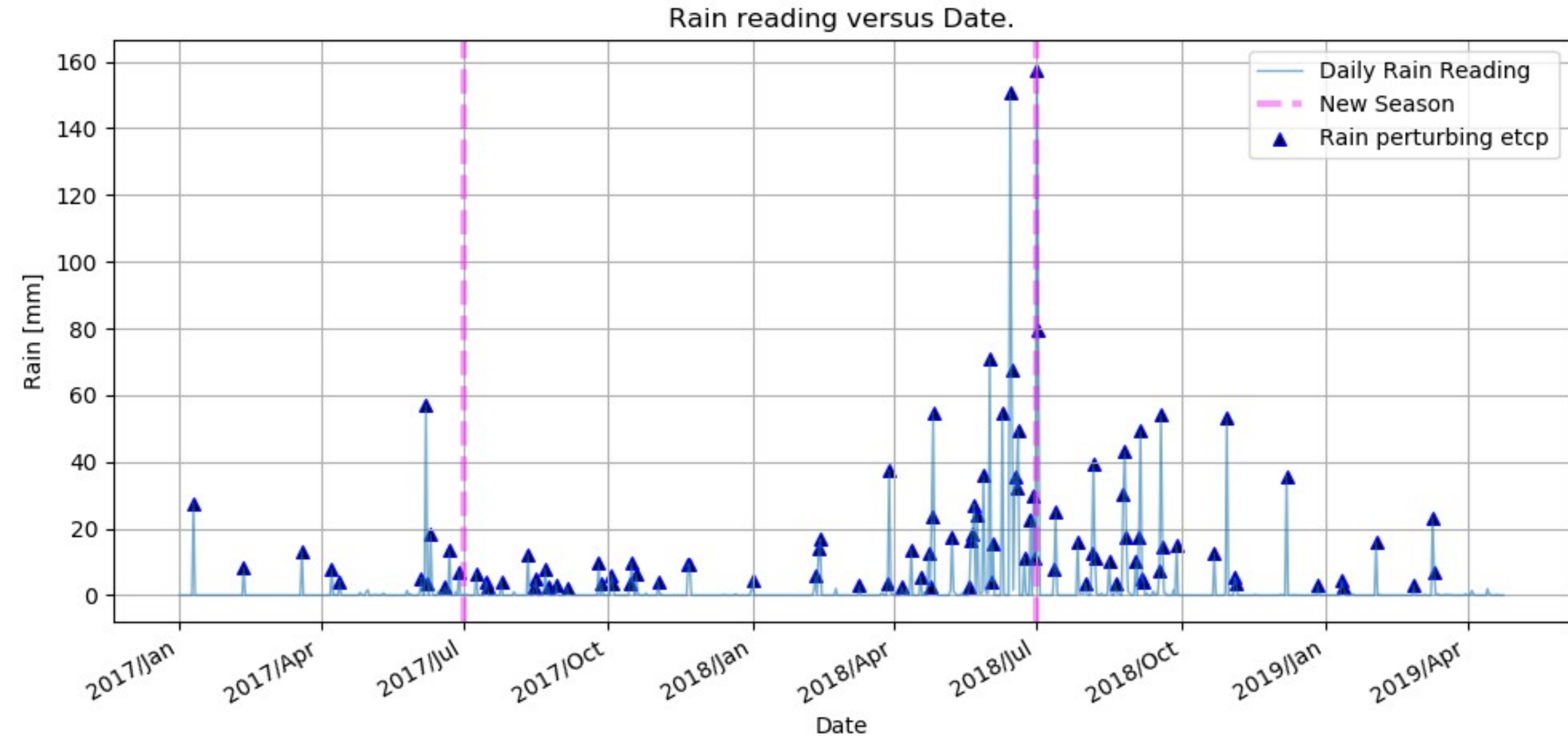
Remaining Questions:

- I make the assumption that Jac wants to apply this software to any set of probes **satisfying the following condition:**

  - All the probes in the probe set (for a given cultivar) come from **one** hemisphere only, which can be either the Northern or Southern hemisphere. Therefore, the individual probe datasets have the "*same season starting date*" for the given cultivar.

  - By extension of the above, the probes in the probe set **cannot** be a combination of Southern hemisphere and Northern hemisphere probes… In such a case the season starting date would differ for different probes, and the current coding would have to be extensively modified…

- @Jac, please comment on the above.

To keep in mind:

- The following **parameters** will vary depending on the specific **cultivar** we are considering:
    - **ETO_MAX** (which is an educated guess)
    - **KCP_MAX** (which is an educated guess)
    - **TEMPERATURE_BASE** (educated guess or some theoretical value?)
    - **BEGINNING_SEASON_MONTH** (will depend whether we are working with Northern or Southern hemisphere probes)
- It is possible for **consecutive eto entries** that are identical to still be legitimate eto entries.
- Despite the fact given above, I still flag any consecutive identical eto entries. I therefore *err on the side of caution*, rather than to *err on the side of carelessness.*

Time required for execution of scripts:

- The total time required (starting from fetching the raw data via the API) to execute the scripts for the current set of **7 Probes** on **Henri's Laptop** is:

  – **2 minutes and 15 seconds.**

Thank You