

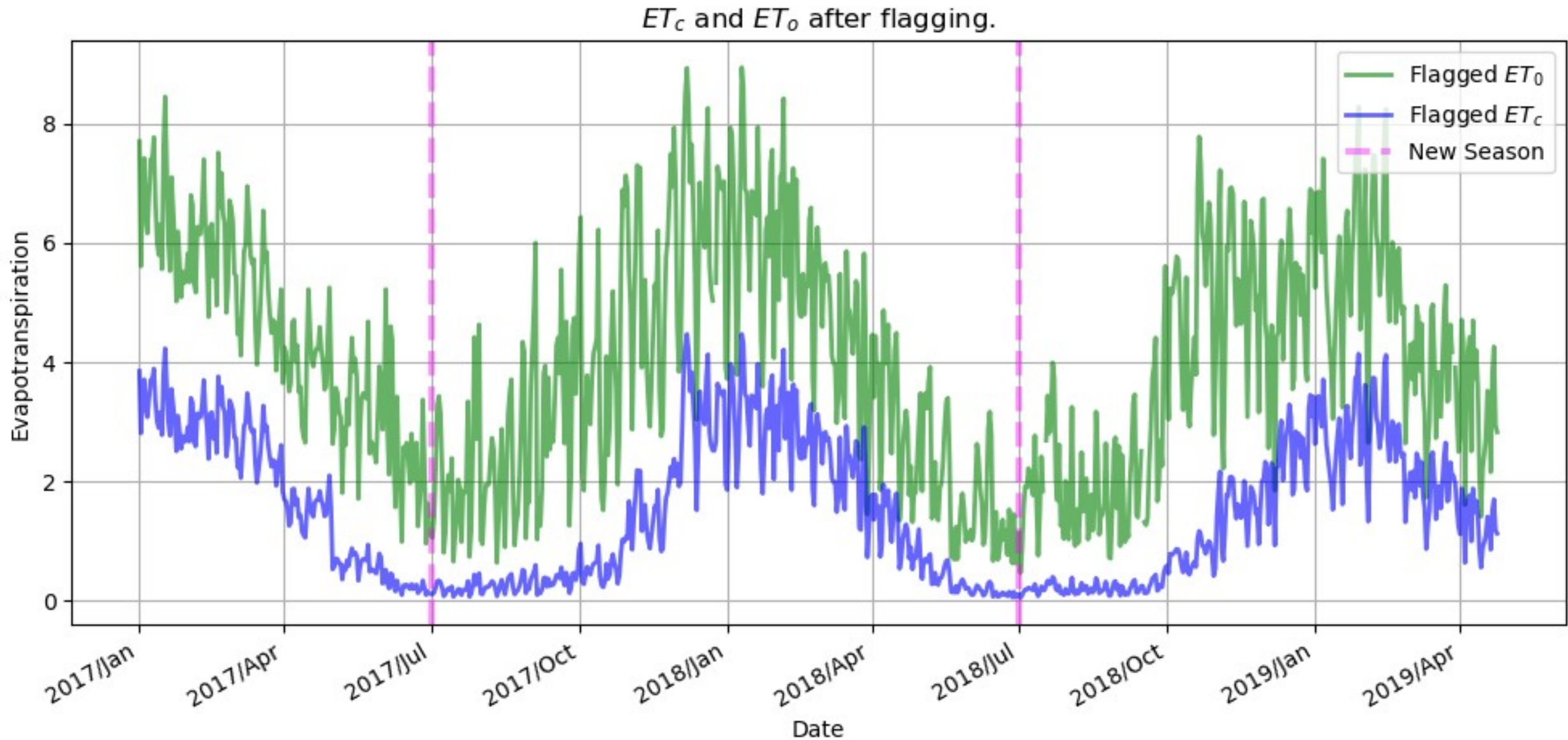
# ProbeSchedule: Latest Results



**25 April 2019, 25/04/2019**

**By *Henri Branken*  
MatogenAI**

# Results of API Updates

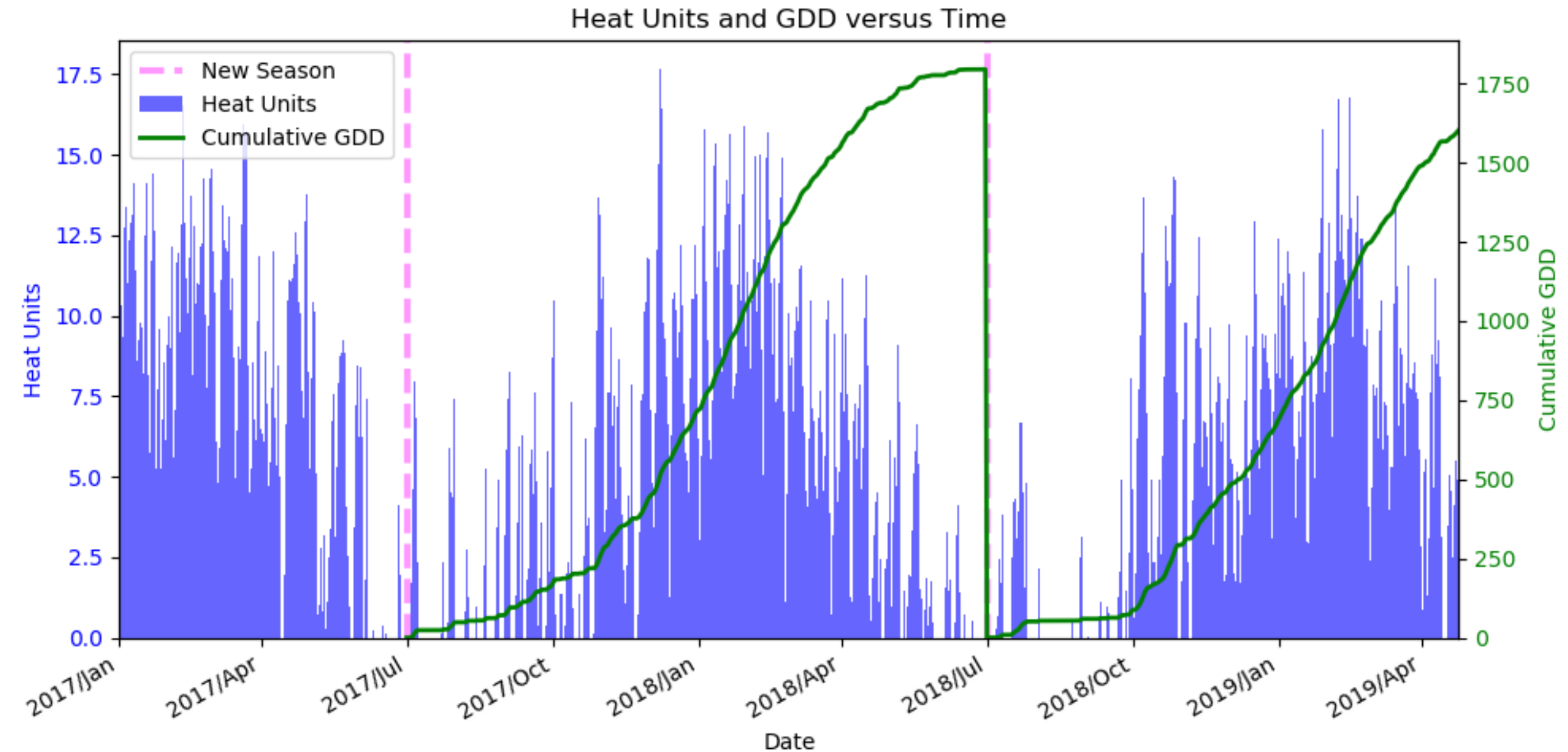


# Results of API updates:

- 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).



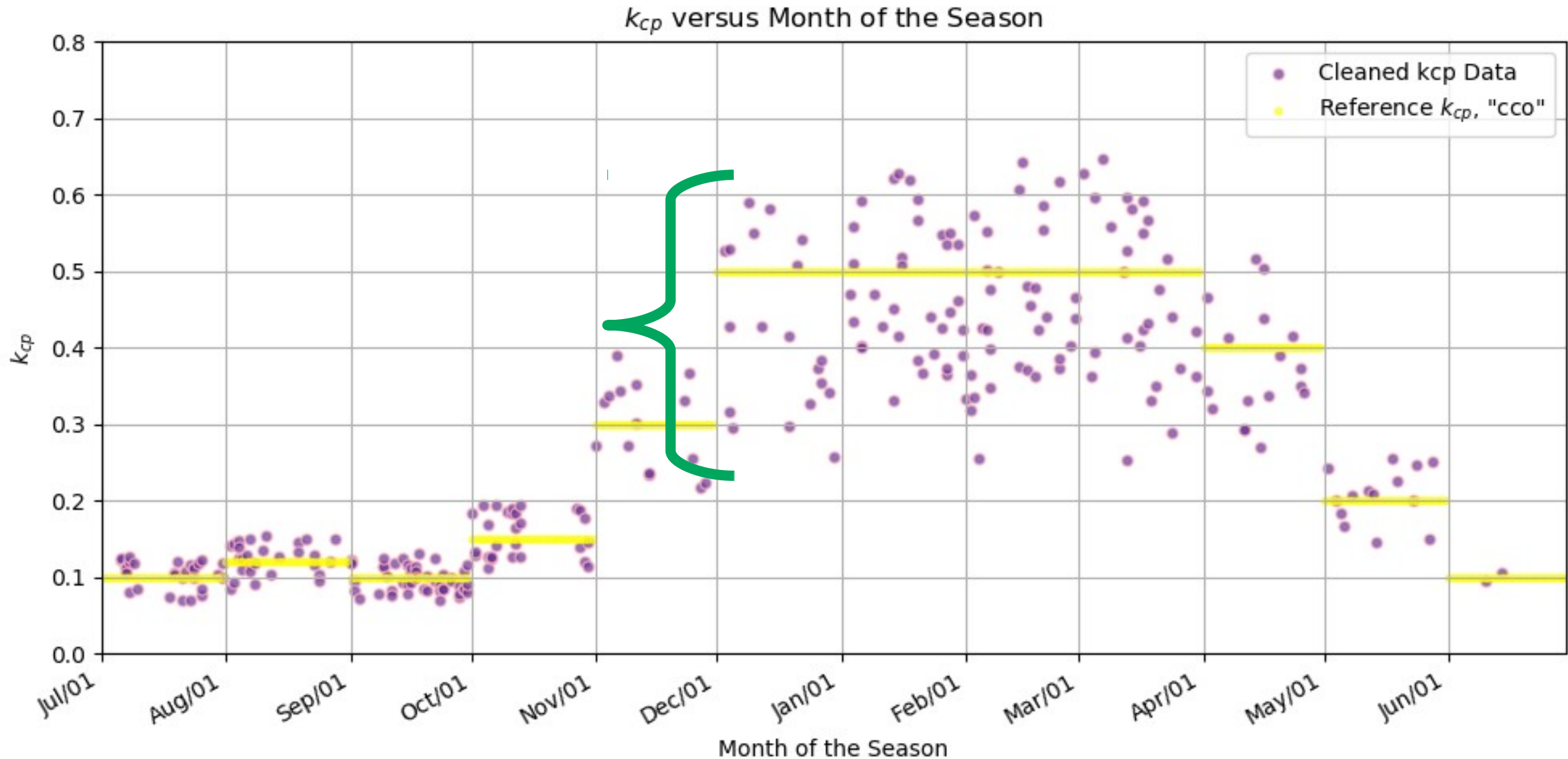
# Bar Graph Rectified



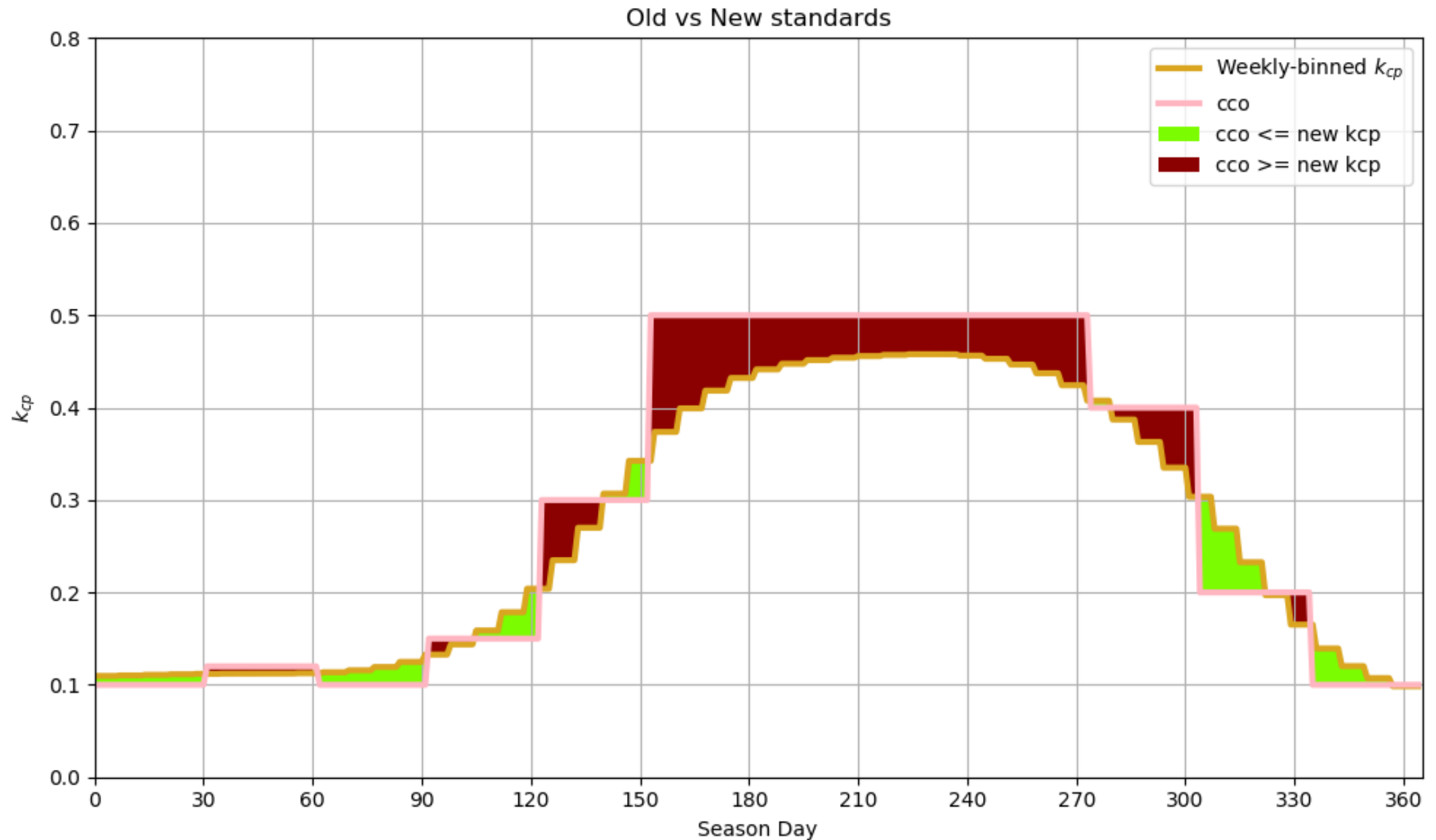
# Fine-tuning etcp deviation

- By "etcp deviation" we mean the **percentage deviation** of **etcp away from** the expected crop water usage, **etc.**
- **Changing this ETCP\_PERC\_DEVIATION** has a noticeable **effect on the final results.**
- In the following we compare results if **ETCP\_PERC\_DEVIATION** is **30%** and **50%**.

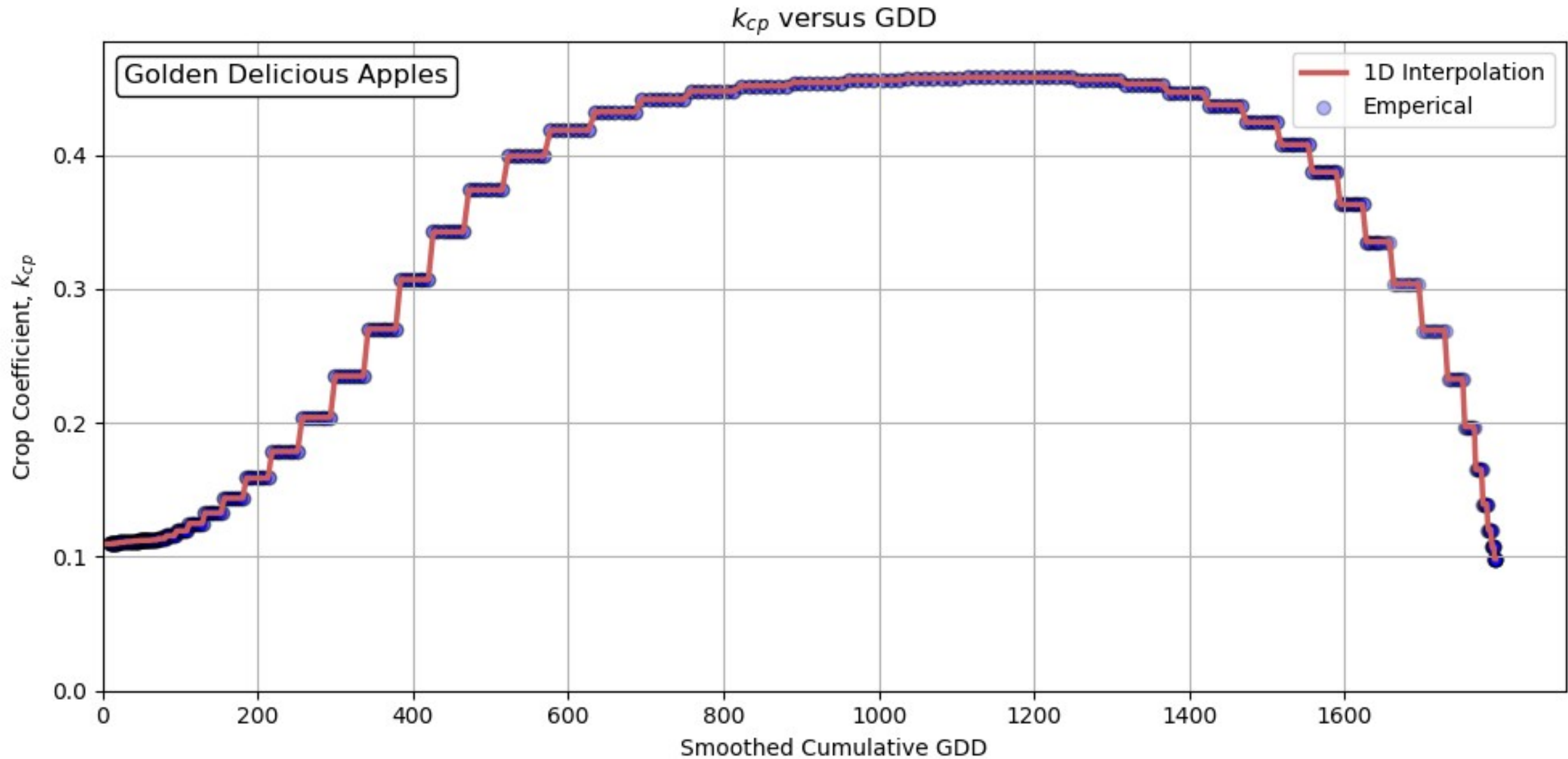
# ETCP\_PERC\_DEVIATION = 30%



# ETCP\_PERC\_DEVIATION = 30%

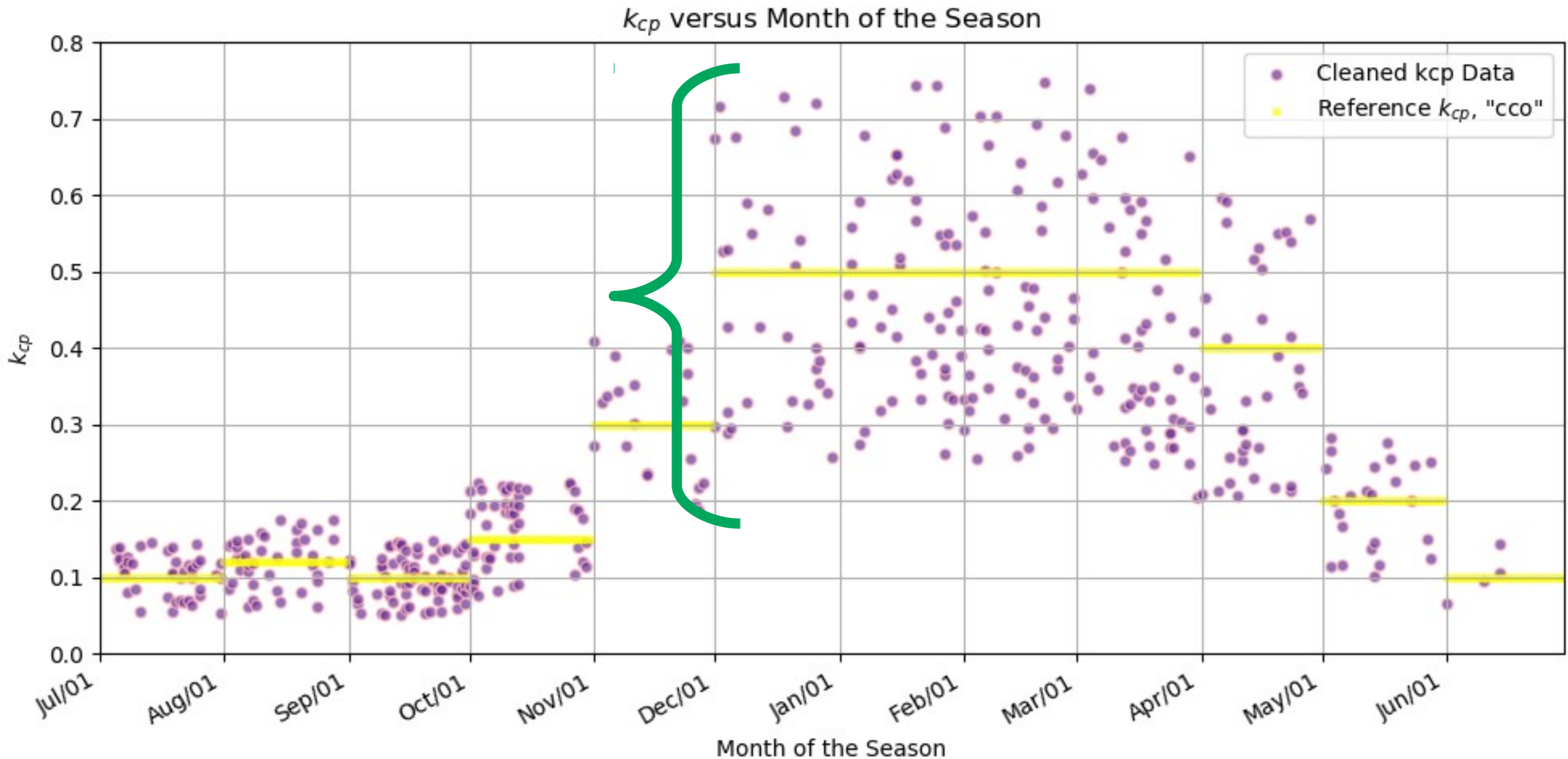


# ETCP\_PERC\_DEVIATION = 30%

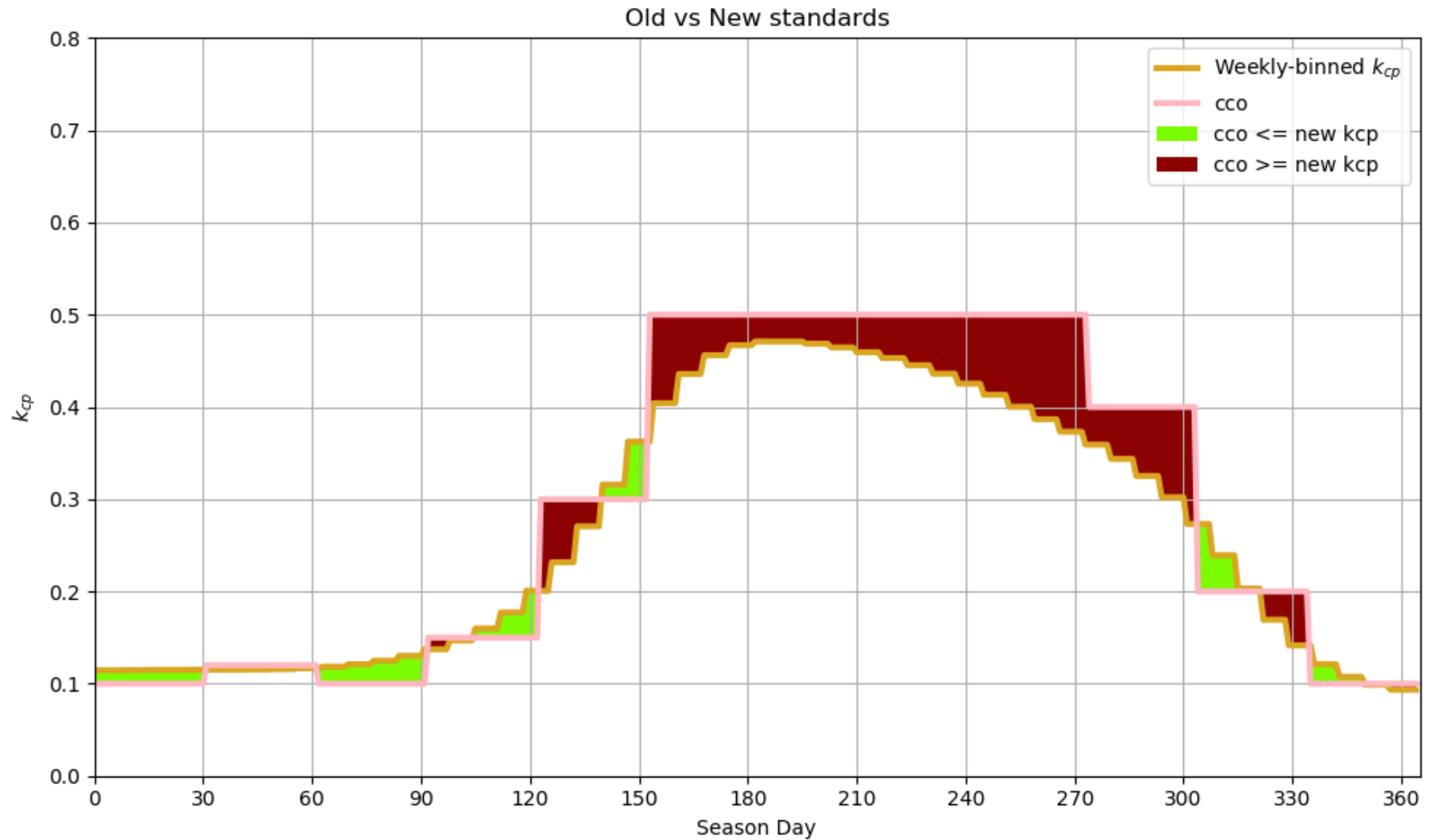




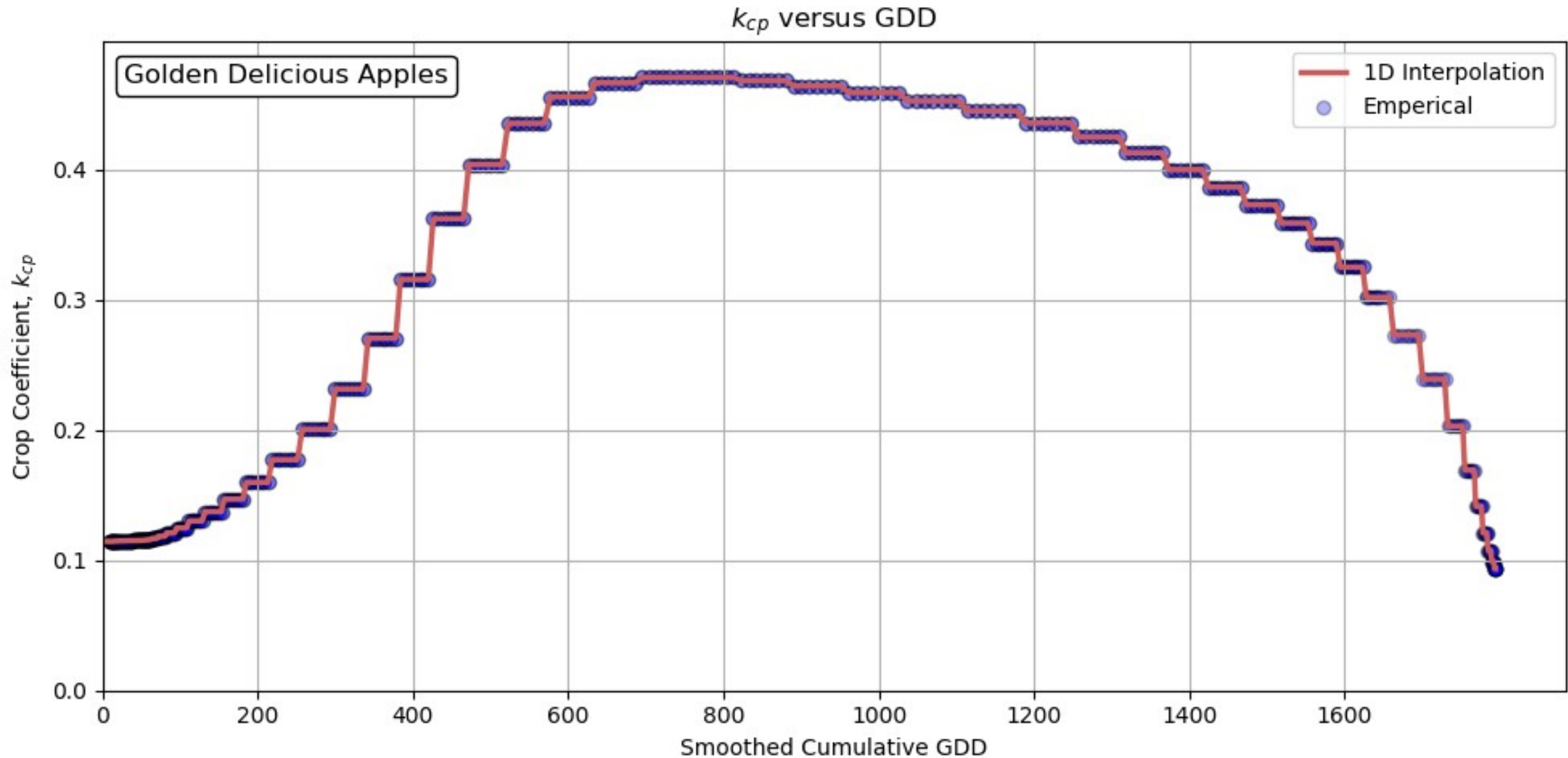
# ETCP\_PERC\_DEVIATION = 50%



# ETCP\_PERC\_DEVIATION = 50%



# ETCP\_PERC\_DEVIATION = **50%**



# Left: 30%. Right: 50%.

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

Report for the entire SET of probes:		
Rain perturbing etcp	12.09%	714/5908
Software simulation	10.05%	594/5908
`Possible` irrigation	7.13%	421/5908
Null profile value	12.36%	730/5908
Profile data blip	0.91%	54/5908
Large profile dip	1.51%	89/5908
Etcp is positive	12.81%	757/5908
Etcp larger than allowed maximum	0.00%	0/5908
Etcp suspiciously low	5.13%	303/5908
Etcp suspiciously high	31.77%	1877/5908
Ratio of etcp over eto exceeds KCP_MAX	0.00%	0/5908
kcp deviation too big	0.36%	21/5908
kcp is NaN	91.91%	5430/5908
Faulty or Missing Heat Units	0.00%	0/5908
Imputed eto	0.00%	0/5908
Stuck or faulty eto	18.20%	1075/5908
eto exceeds ETO_MAX	0.00%	0/5908
Faulty etc	0.34%	20/5908
etc exceeds ETCP_MAX	0.00%	0/5908
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.

Probe Information				
Iteration	Number of healthy probes	Number of Improvements	Number of Deteriorations	$n\_imp \geq n\_det$
[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

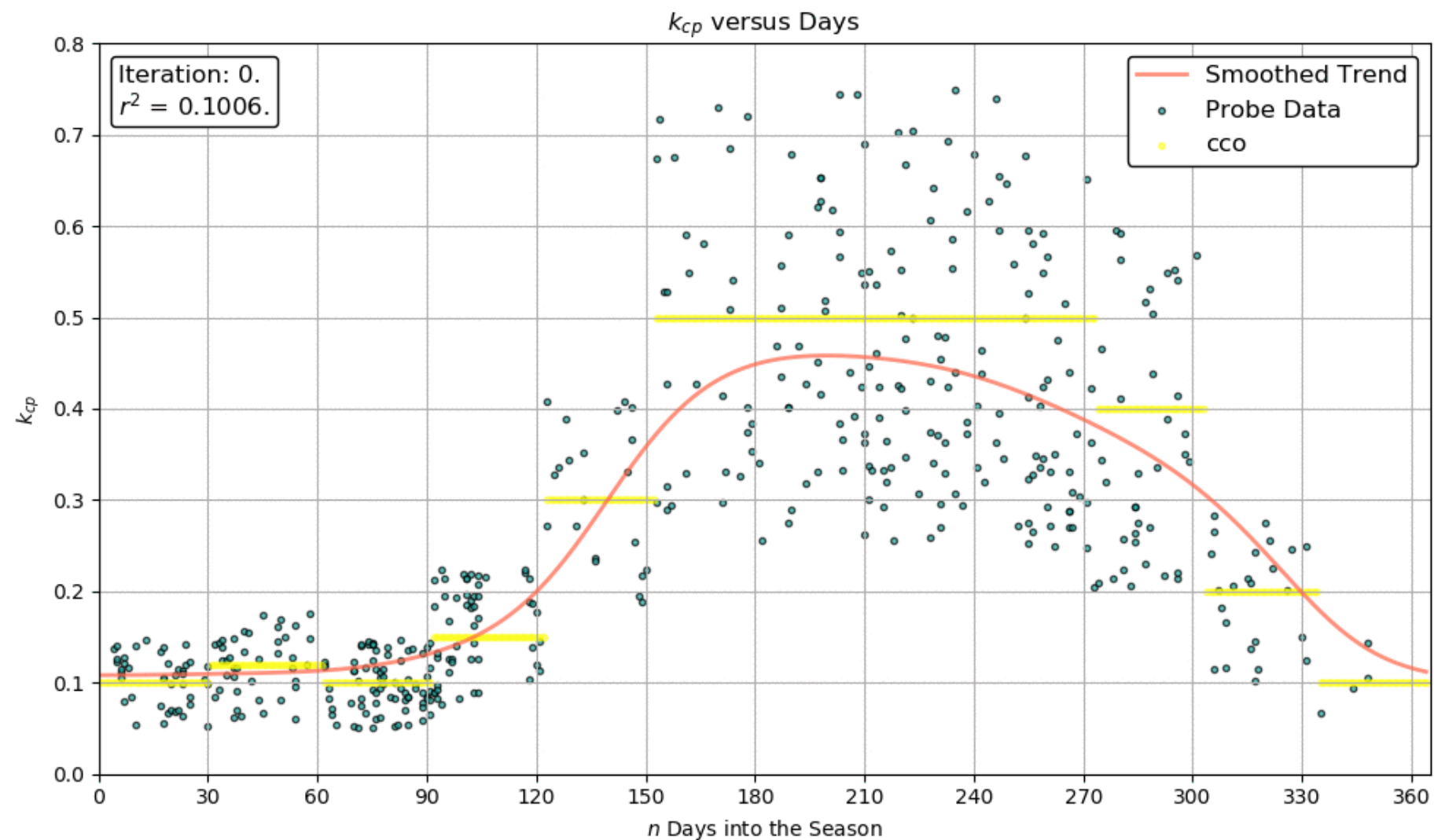
# Complementary table to the previous one:

- 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.

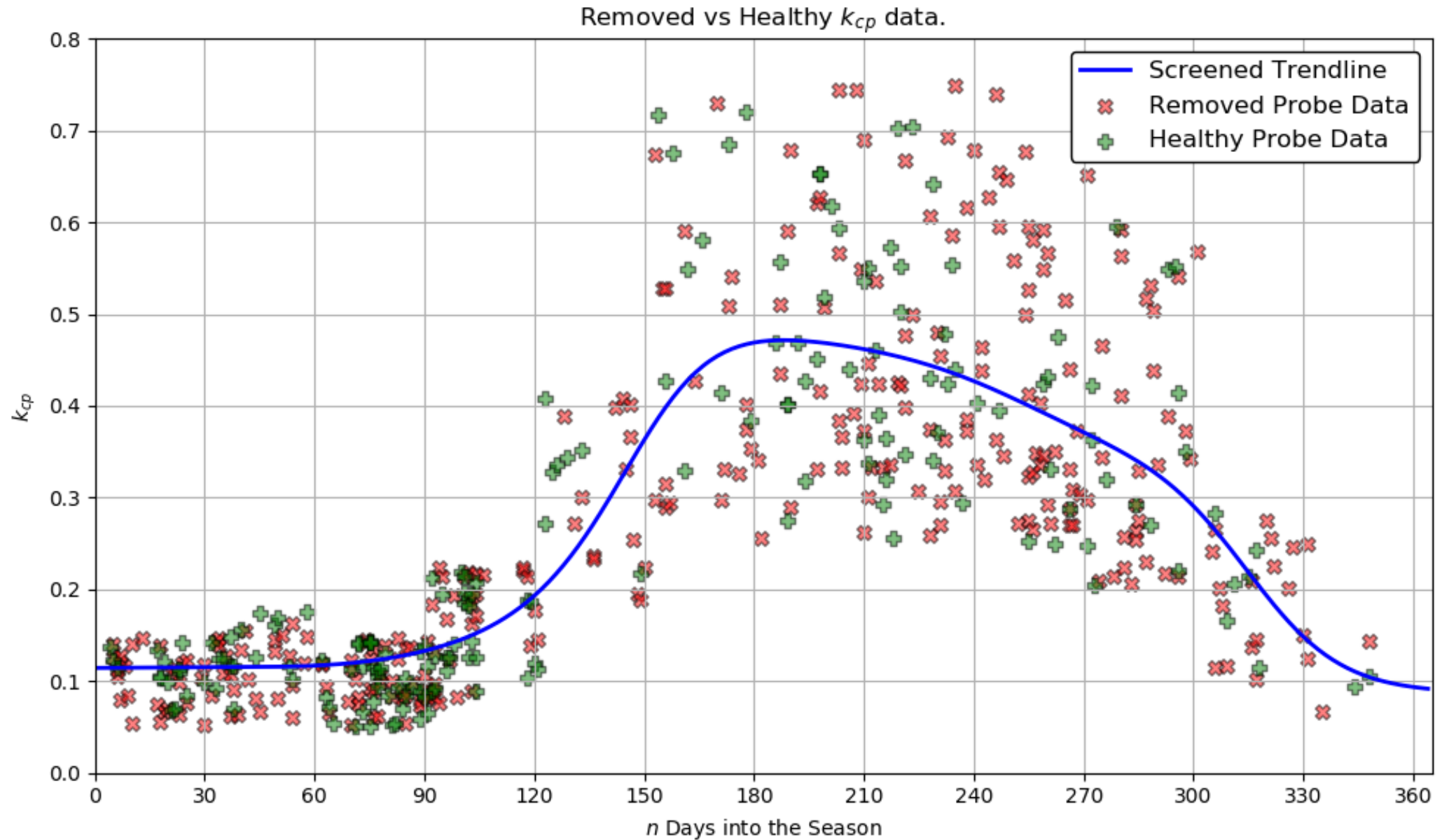
Screening Report						
iter	n scatter points	cco r-squared	scatter r-squared	delta scatter r-sqr	probe_id	removed
0	478	0.0518844	0.1005873	-	-	
1	424	0.0492342	0.0992430	-0.0013444	→ P-370	
2	384	0.0463118	0.0970331	-0.0022099	→ P-392	
3	315	0.0448074	0.0954229	-0.0016102	→ P-372	
4	179	0.0521010	0.0920653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

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

For the .pdf viewers, take a look at **progression.gif**



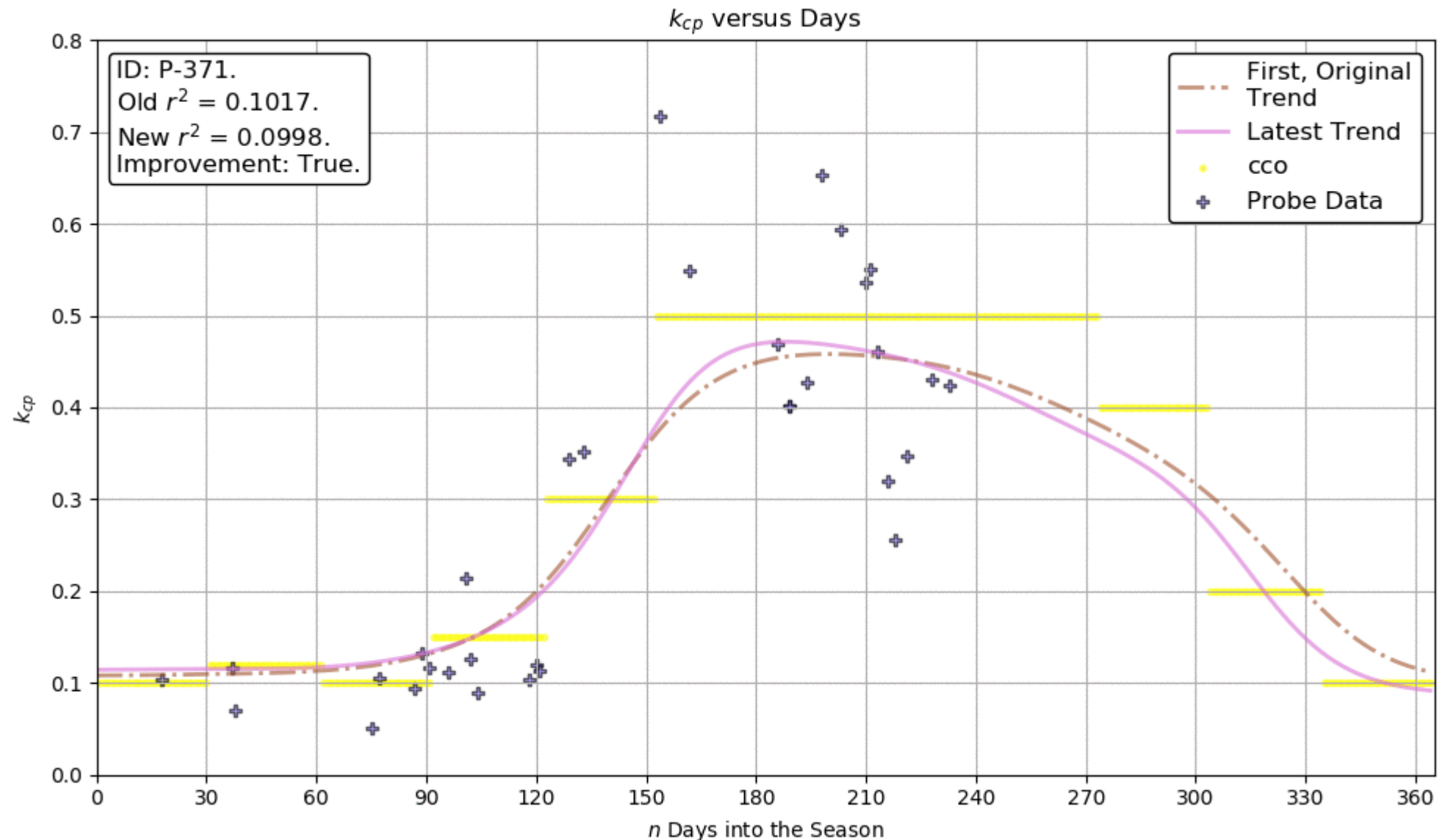
# Comparing **Removed** versus **Remaining** datapoints.





# .gif Animation of remaining three "Healthy" Probes:

For the .pdf viewers, take a look at **healthy.gif**



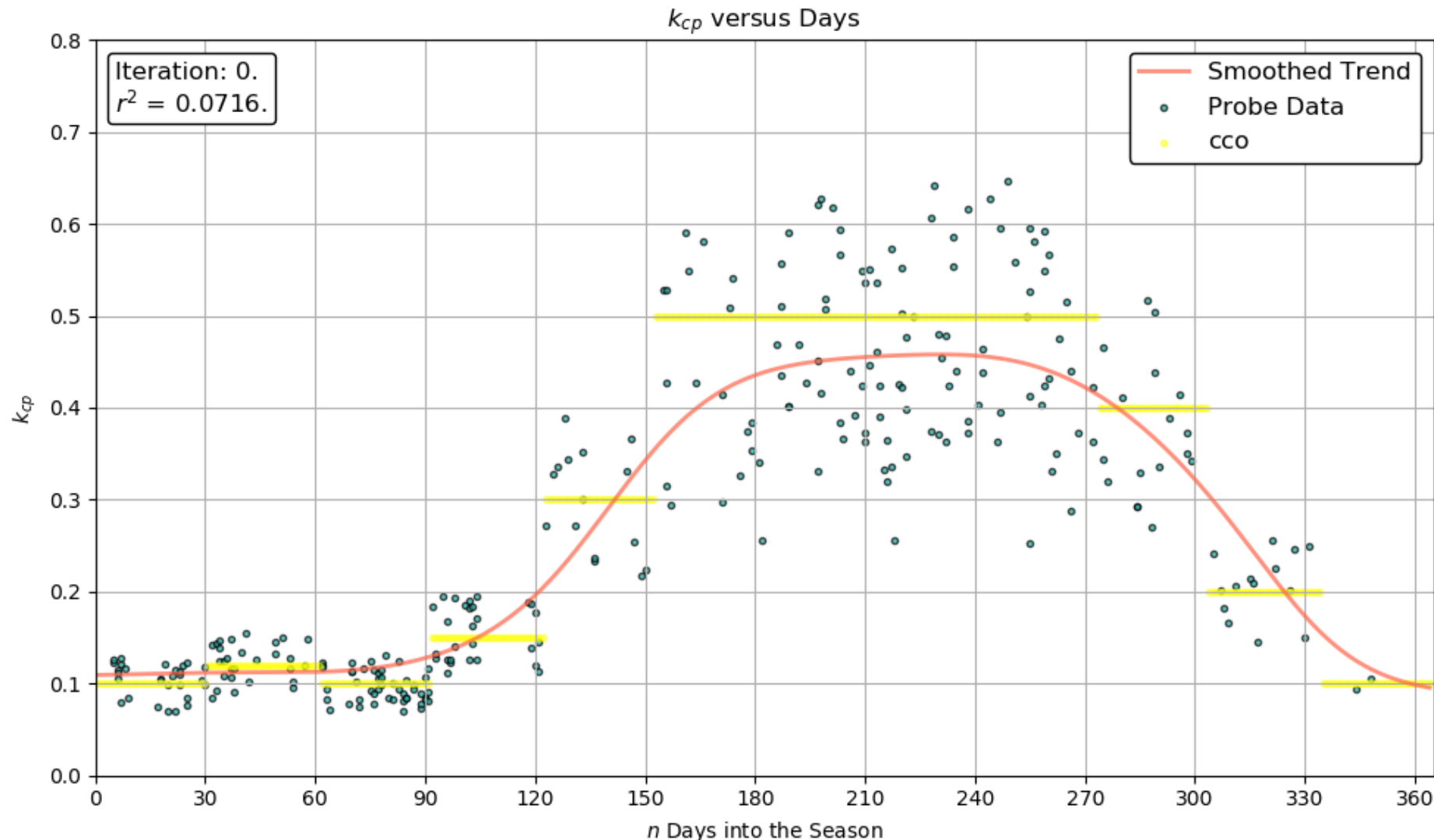
Previous 5 slides correspond to  
ETCP\_PERC\_DEVIATION = **50%**

- 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.

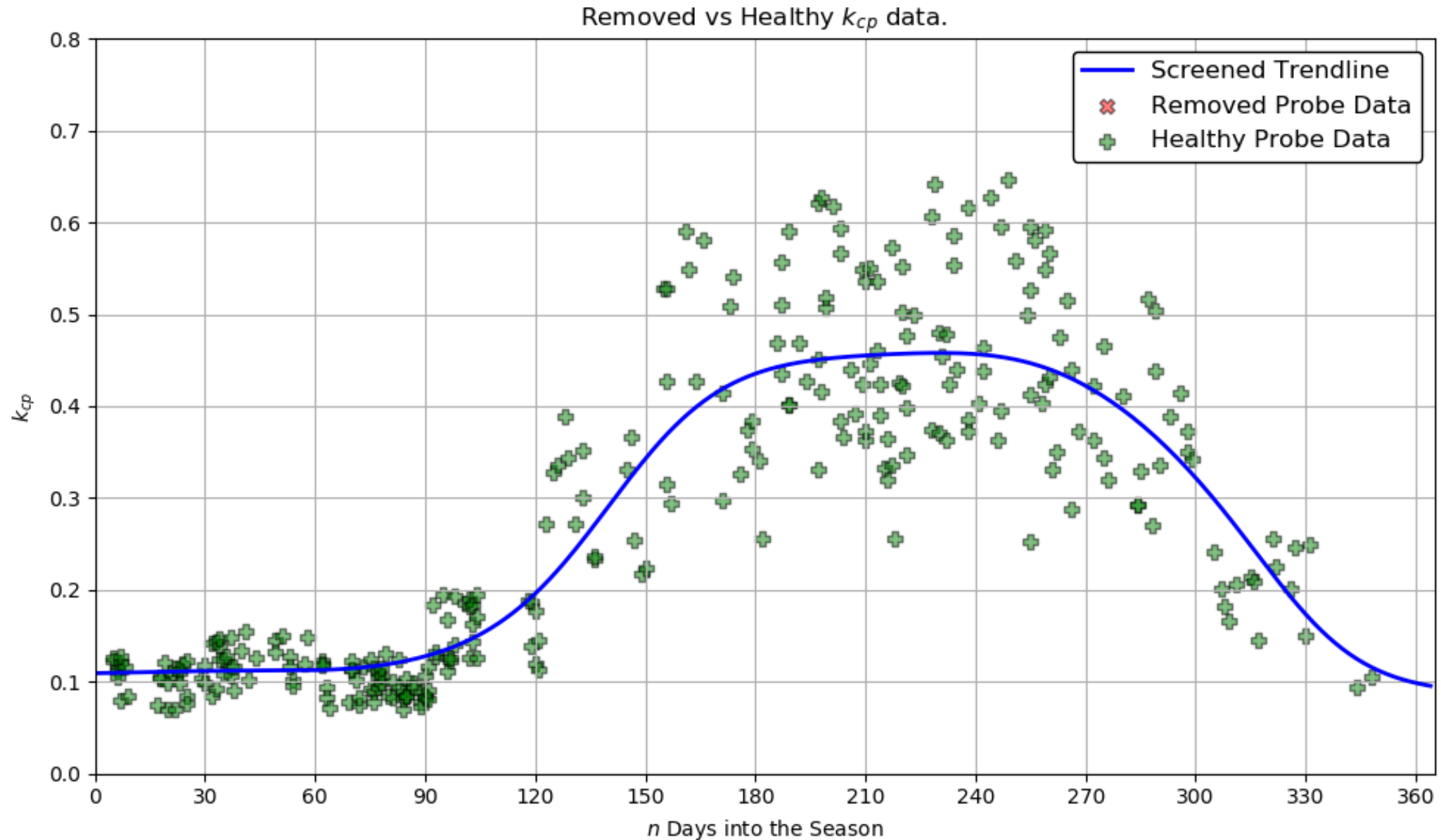
Probe Information				
Iteration	Number of healthy probes	Number of Improvements	Number of Deteriorations	n_imp >= n_det
[0, 1]	6	2	4	False

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.)



Comparing "**Removed**" versus **Remaining** datapoints for  $\text{ETCP\_PERC\_DEVIATION} = 30\%$ .

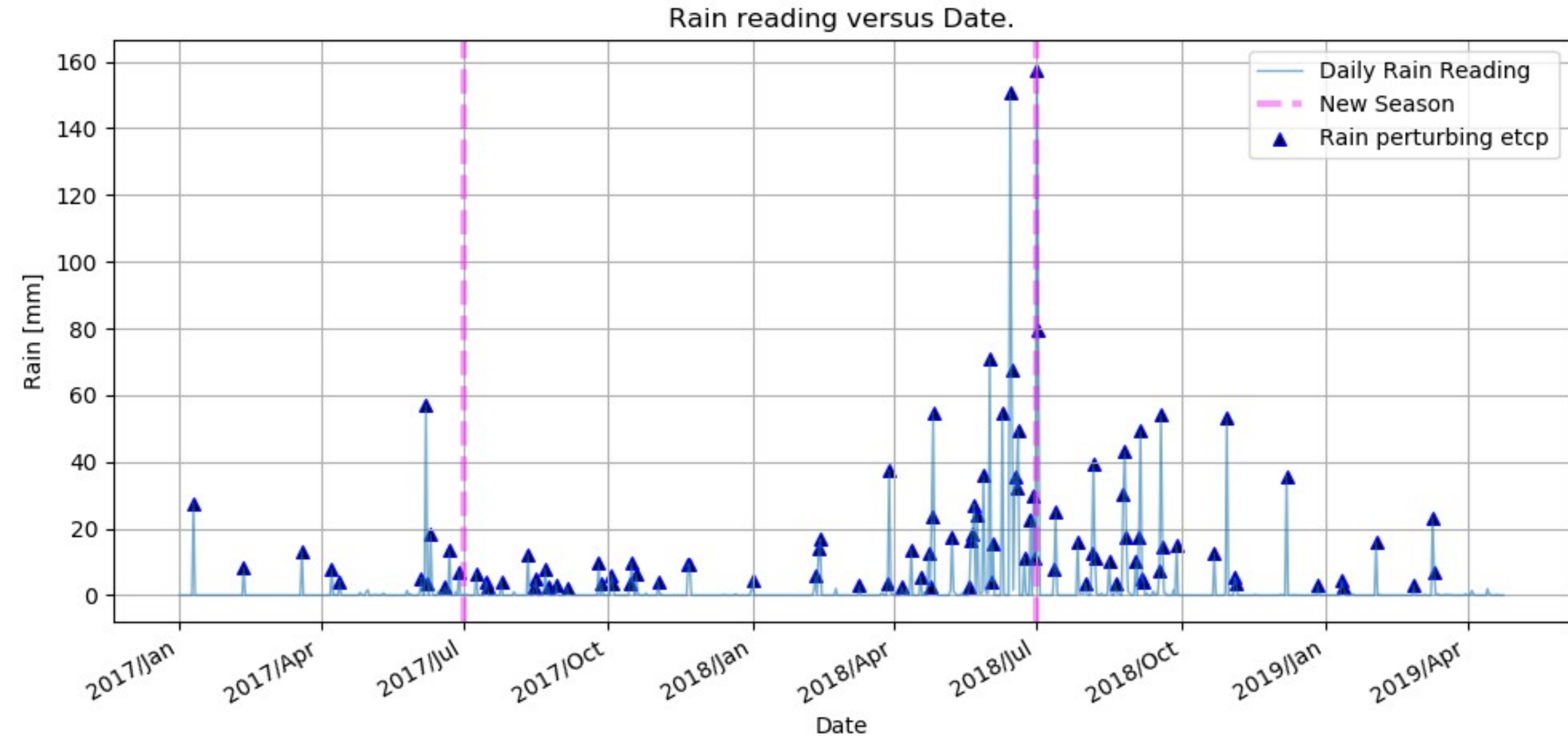




# Remaining Questions:

- @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):



# 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**