**General remarks:**

* In all cases, the dependent variable was Tm (at pH 5 I guess). pH mid transitions may be more relevant ?
* Extra data needed ? If we extrapolate to longer loop lengths, which models are more likely to collapse ?
* C run length is defined as length of an individual run (4 for example) or total number of C - multiplied by 4 as there are 4 runs ? (should be 16). Note : runs may be of unequal length; we started to collect data on CCC...CCCC...CCC..CCCC for example.

C run length is the length of the individual run which is constant for C per sequence in the c

**Approach A / Slide 1 :**

* What are these optimal parameters ?
* C positive / T negative ? (shouldn’t it be G negative instead ?!)
* In the end, to compare i-Hunter with G4Hunter, the best might to normalize the scores within the same range (between 0 and 1 or -1 and +1). For G4Hunter, the best score is +4 for a pure GGGGGG... run.

**Approach B:**

* Briefly explain what a gradient boosting machine is (this is a complete black box to me)
* OK with the features chosen – but the last one (total length) is dependent on the others ? (should be equal to C x 4 + T1 + T2 + T3 ?)
* Slide 4: I don’t understand the method... but probably not a big problem !
* Slide 5: correlation seems really nice (actually better than Tm vs pHT)

**Approach C:**

* Much easier for me to understand the principle
* For the proposed equation, I will have to visually test it to grasp its implications.

**Conclusions:**

* All fits seem better than “reasonable” actually. I am positively impressed and very happy.
* Most important feature is C-track : makes complete sense to me.
* T3 is more important: this was unexpected to me ! Can you provide examples in which T3 has a more profound impact than T1 for example ?
* For the transparent analytical equation, I suppose it can be improved over time as more data is added ?