MADA Data Analysis Project

Monica Chan

2021-11-04

# 1 Summary/Abstract

Applying MADA learned techniques to a RT-PCR dataset to determine the effectiveness of sample enhancement strategies for environmental surface sampling of low bio-burden viruses.

# 2 Introduction

## 2.1 Background

The global Coronavirus Disease 2019 (COVID-19) pandemic continues to persist due to the timing of an available vaccine, varying adherence to COVID-19 public health safety recommendations, and the emergence of SARS-CoV-2 (SC2) variants. Environmental reservoirs are a reported, indirect transmission route for viral pathogens. Although many viruses have a low infectious dose, they can be difficult to detect due to the low bioburden in the environment. The ability to detect viral pathogens is critical to identifying hidden reservoirs which can inform infection control. Culture-based methods can be labor and time intensive. The use and limitations of commercial “target capture” (TC) technologies are varied compared to molecular approaches utilizing automated extraction. The objective of this study is improve outbreak response through rapid pathogen detection by evaluating a TC technology’s ability to capture SC2 from complex matrices.

### 2.1.1 Questions/Hypotheses to be addressed

There are several major aims of this:

1. Do the sample enhancement strategies work?
2. How well do they work compared to each other?
3. Are there any environmental variables recorded that have a positive impact on detection?
4. Can this data be modeled to back calculate a sample’s original concentration based on the Ct values recovered?

Testing several sample enhancement strategies (Nanoparticles and Bait & Capture) for molecular detection of heat-inactivated SARS-CoV-2 from stainless-steel surfaces. The primary aim of this work is to determine if any of the enhancement strategies increase the concentration of RNA copies in a sample that will express itself as a lower Ct value from the Rt-PCR output. Lower Ct values equate to more copies being present in the sample meaning less time needed to confirm presence.

In the case the sample enhancements work, we would then compare how well they worked against an automated extraction method. The degree of “wellness” will be determined by how low the Ct value results are produced after a sample has going through a specific method.

In addition to the comparison of enhancement strategies the effects of the tool used for sampling (macrofoam or knit-fiber), time held, and other environmental variables will be explored any significant effects, with interest in positive impacts to detection.

## 2.2 Methods and Results

### 2.2.1 Experiment setup and conditions

Using established standard methods for environmental sampling with a self-contained polyurethane swab from stainless steel surface, a TC approach (NTP; Nanotrap® Magnetic Virus Particles, Ceres Nanosciences) was compared to an automated extraction method (MX48; Maxwell® RSC Viral Total Nucleic Acid Purification Kit, Promega) for recovery of heat-inactivated SC2 from the polyurethane swab as detected by rRT-PCR

## 2.3 Data aquisition

Reports per each rRT-PCR run were generated and quality checked on their respective machines before transfer to a Teams repository. Summary data was selected from the reports and added to additional environmental data gathered on based on the extraction date of the sample.

Completed run data is then collated into one file and read into R./A composited data set was compiled and saved as a csv and used as the base data for the project

### 2.3.1 Data import and cleaning

processing script.R was modified to do initial cleaning and preparation of the data set for further exploration.

Data was imported and cleaned. Data was then split between the two different experimental types (P1 and P2).

This can be found in processing\_code/ProcessingScript.RMD

### 2.3.2 Data exploration

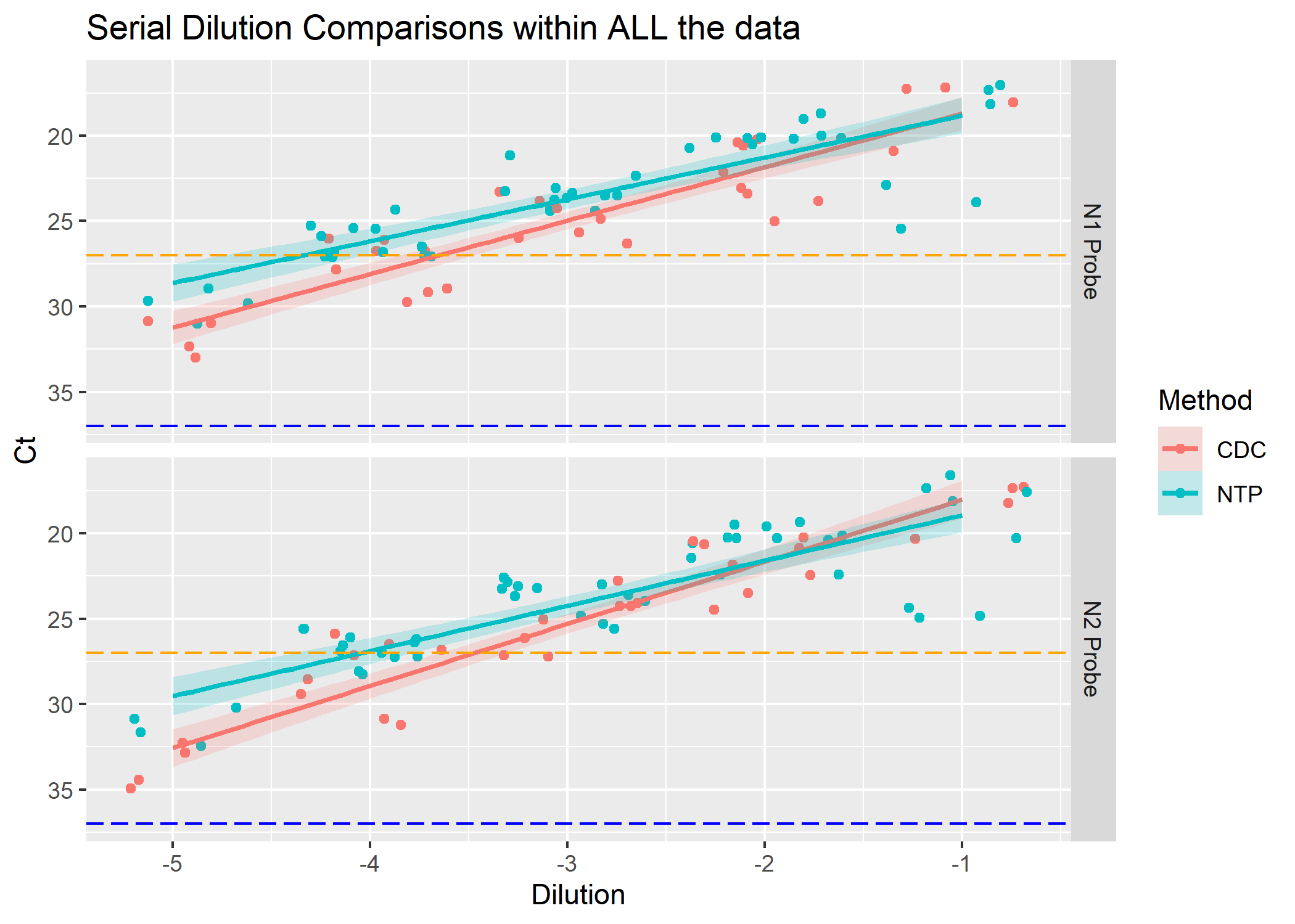
#### 2.3.2.1 All Data

Overview - split between part 1 and part 2

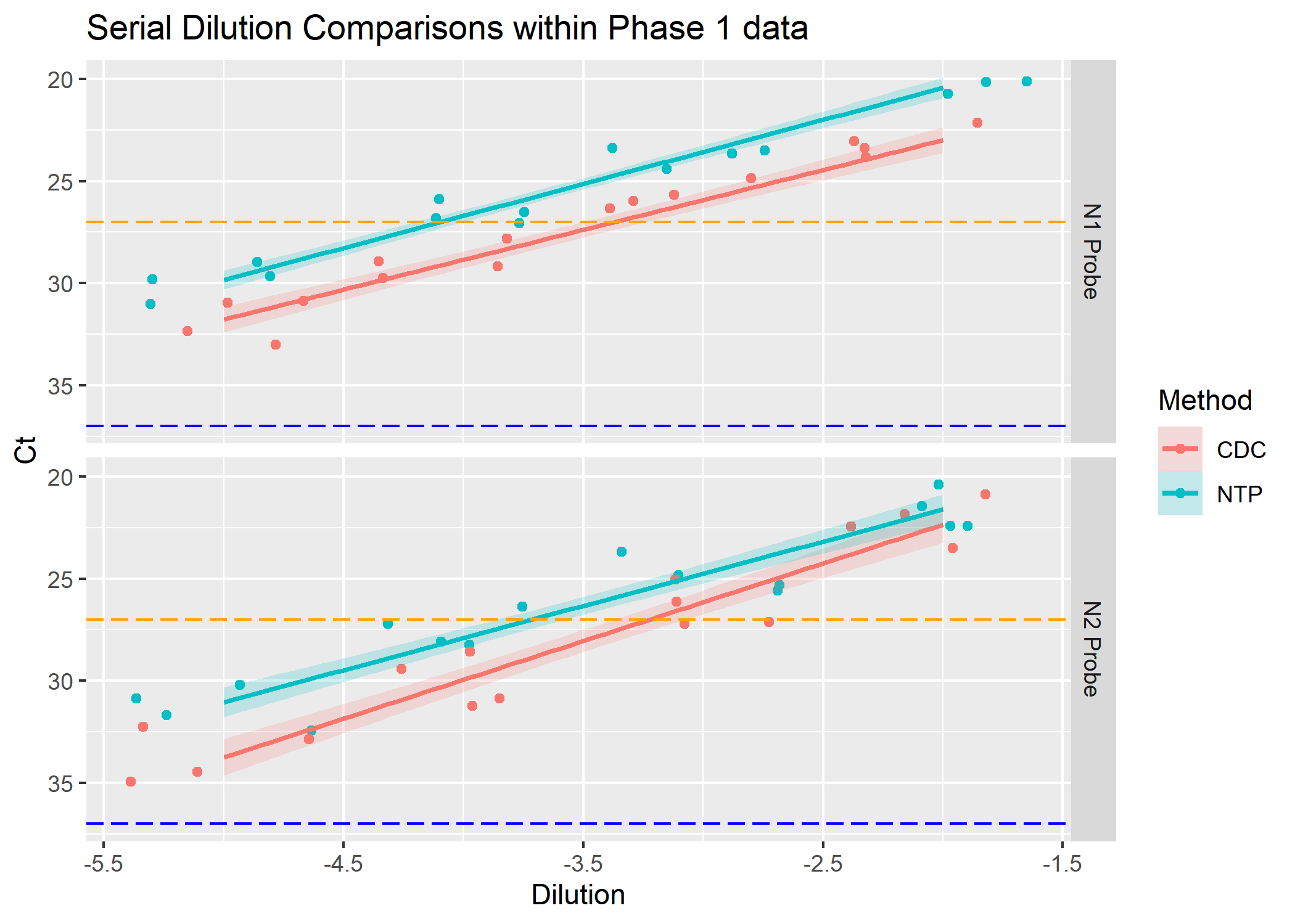
##### 2.3.2.1.1 Serial Dilutions confirmation

Summary table scatter plot of SDs based on dilution factors

knitr::include\_graphics("../../results/ALL-SD-Plot.png")

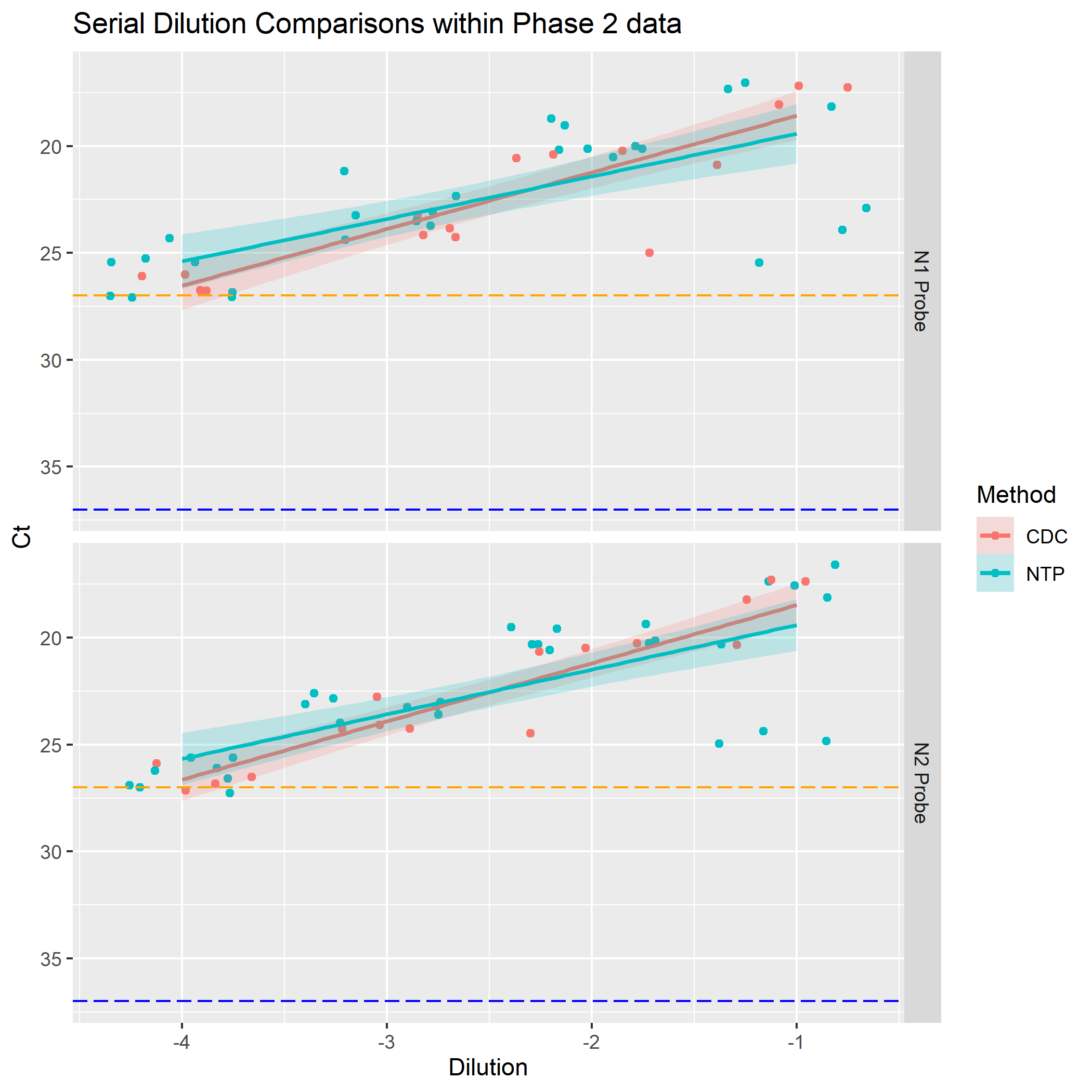
 #### Phase 1

knitr::include\_graphics("../../results/P1-SD-Plot.png")

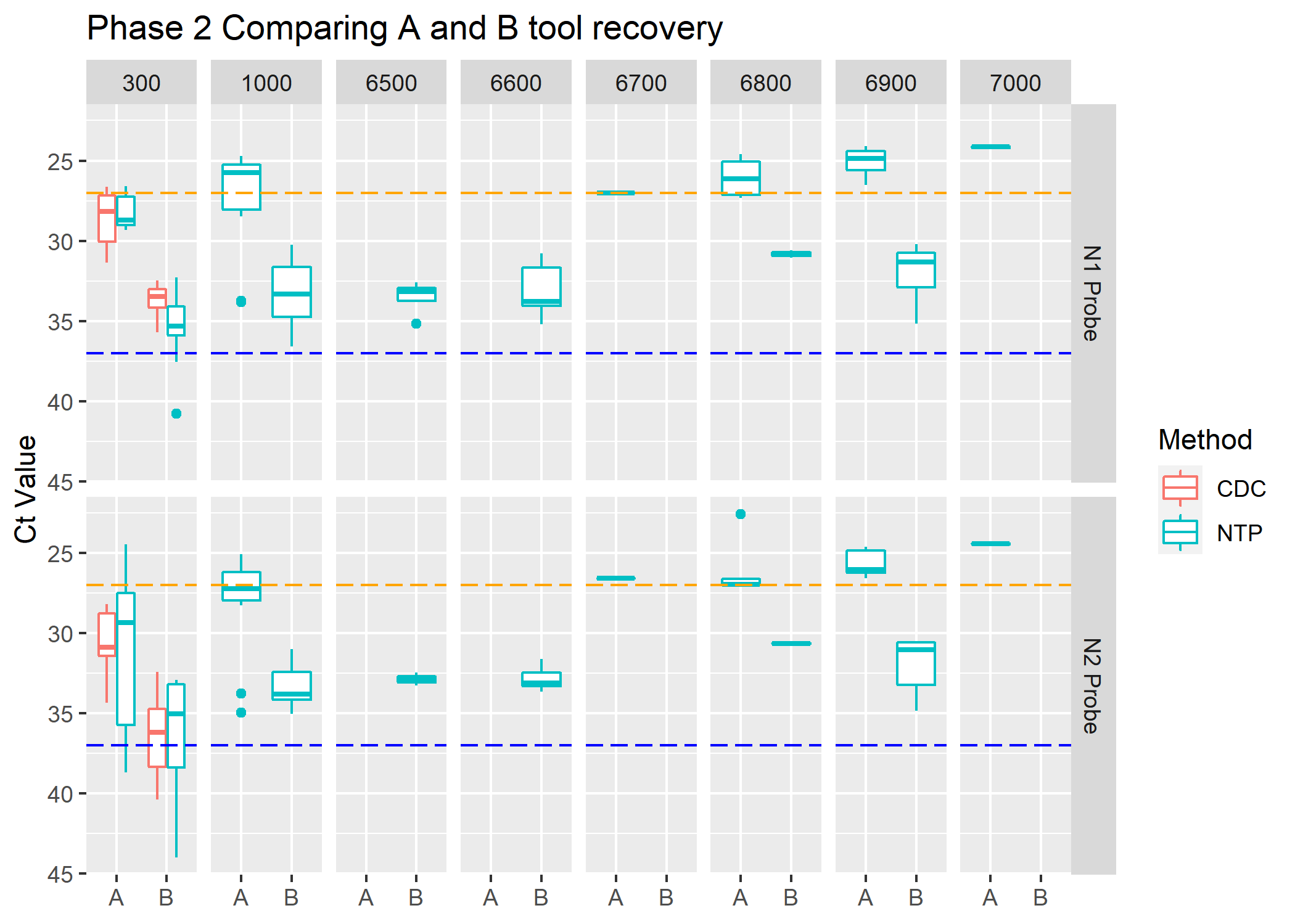


### 2.3.3 Phase 2

knitr::include\_graphics("../../results/P2-SD-Plot.png")

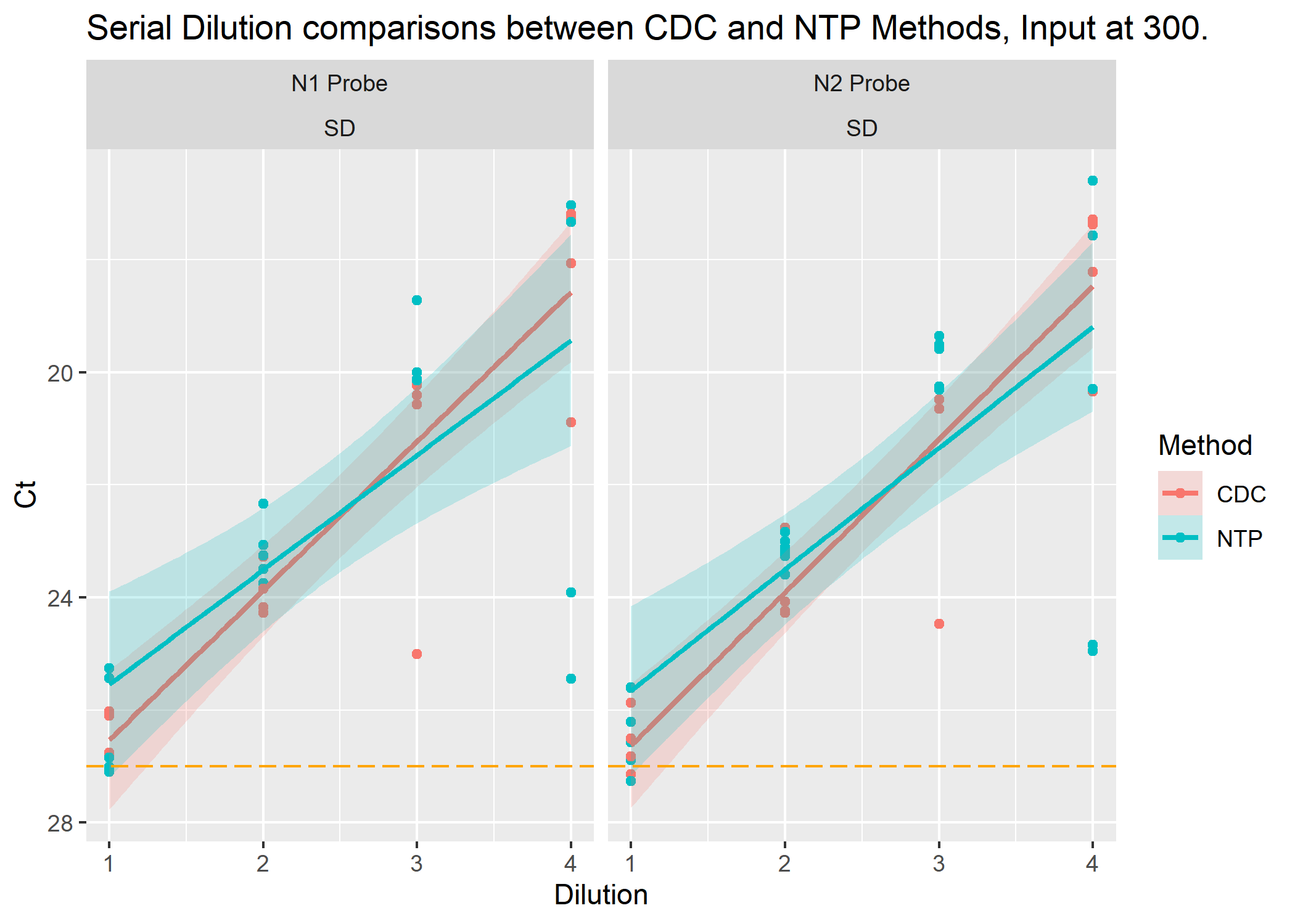
 #### Phase 2 Comparing A and B Tool Recovery

knitr::include\_graphics("../../results/ASK-P2P1-Tools.png")

 Tool A has an average Ct recovery higher than B. This makes sense as the surfaces A recovered from is 2 logs higher (-1) than that of B (-3). A log drop is ~3Ct values we can see from the graphic that it trends as predicted. Suggests that

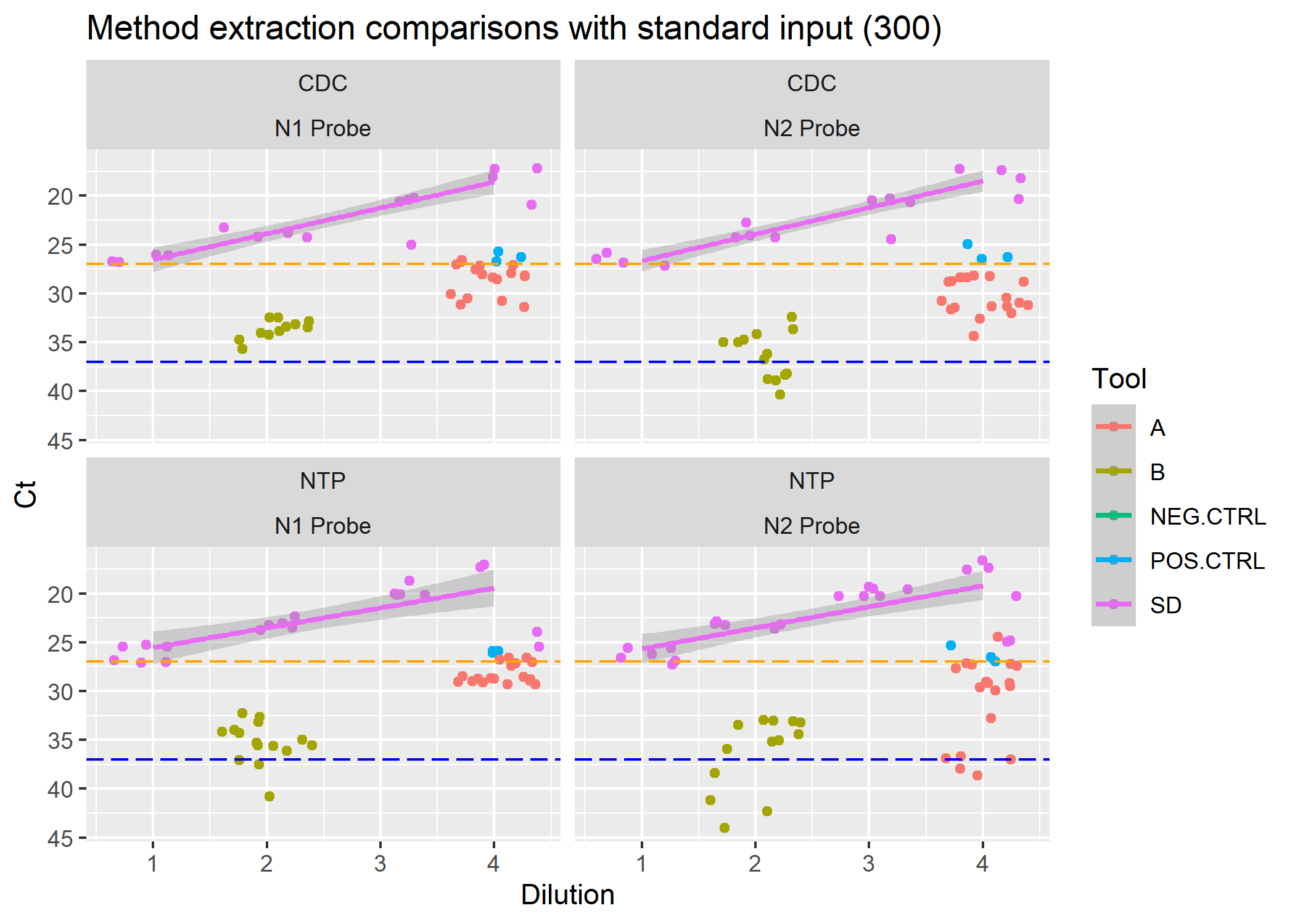
#### 2.3.3.1 Phase 2 Comparing Method Serial Dilutions (CDC vs NTP) with standard input of 300uL

knitr::include\_graphics("../../results/ASK-P2P2-Methods.png")

 Serial dilution data between CDC and NTP. Noting that all SD are in the confirmed realm of under 27 Ct Value CDC method is better than NTP at -1 concentration of the SD. the -2 shows little difference between method recovery Ct values. NTP appears to remain better than CDC at -3 and -4 levels. Accounting for the SE looks like both methods are the same until the -2 dilution. Might be significant difference at the -4 dilution.

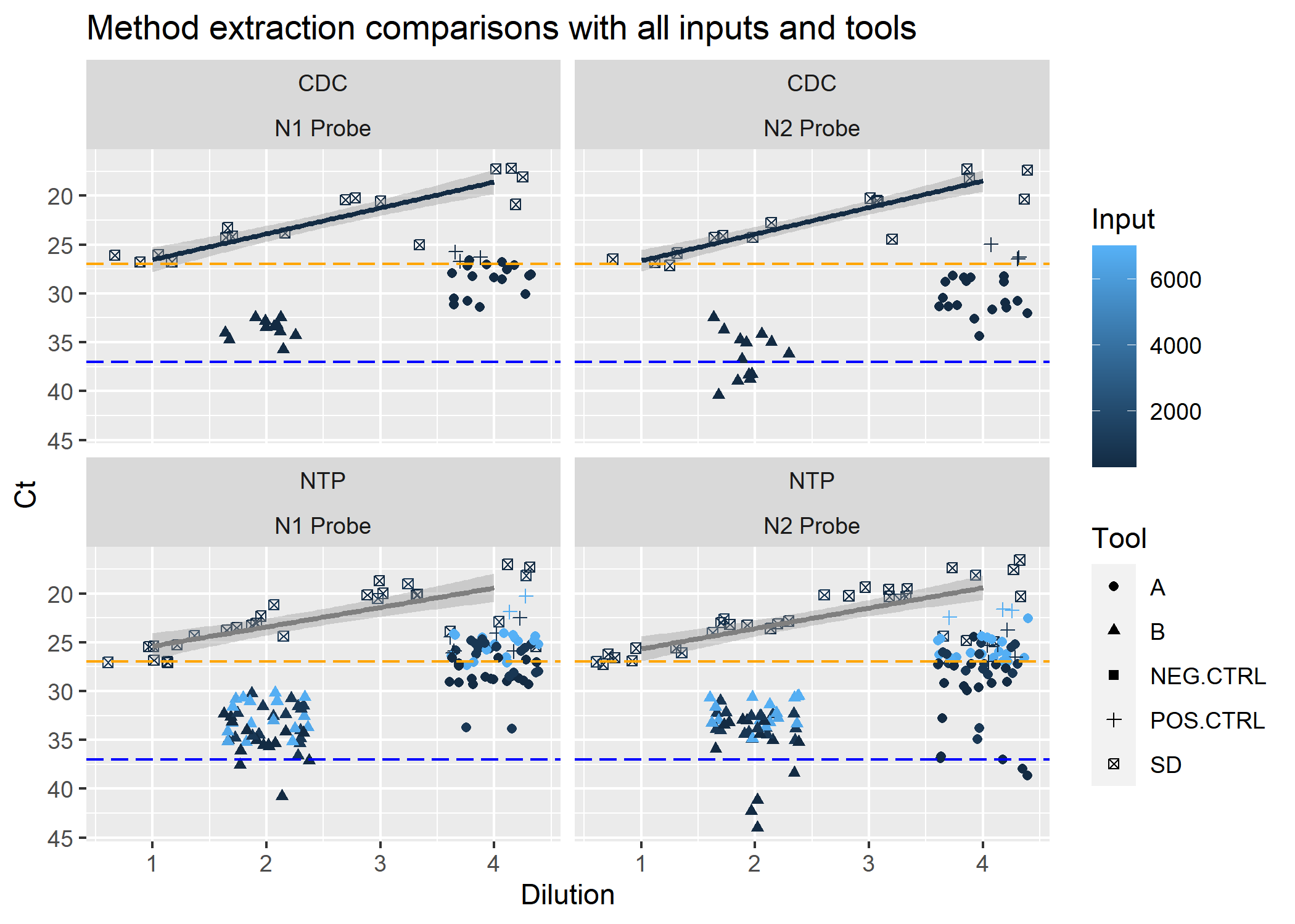
#### 2.3.3.2 Phase 2 Comparing Methods with standard input of 300uL

knitr::include\_graphics("../../results/ASK-P2P3-Methods0300.png")



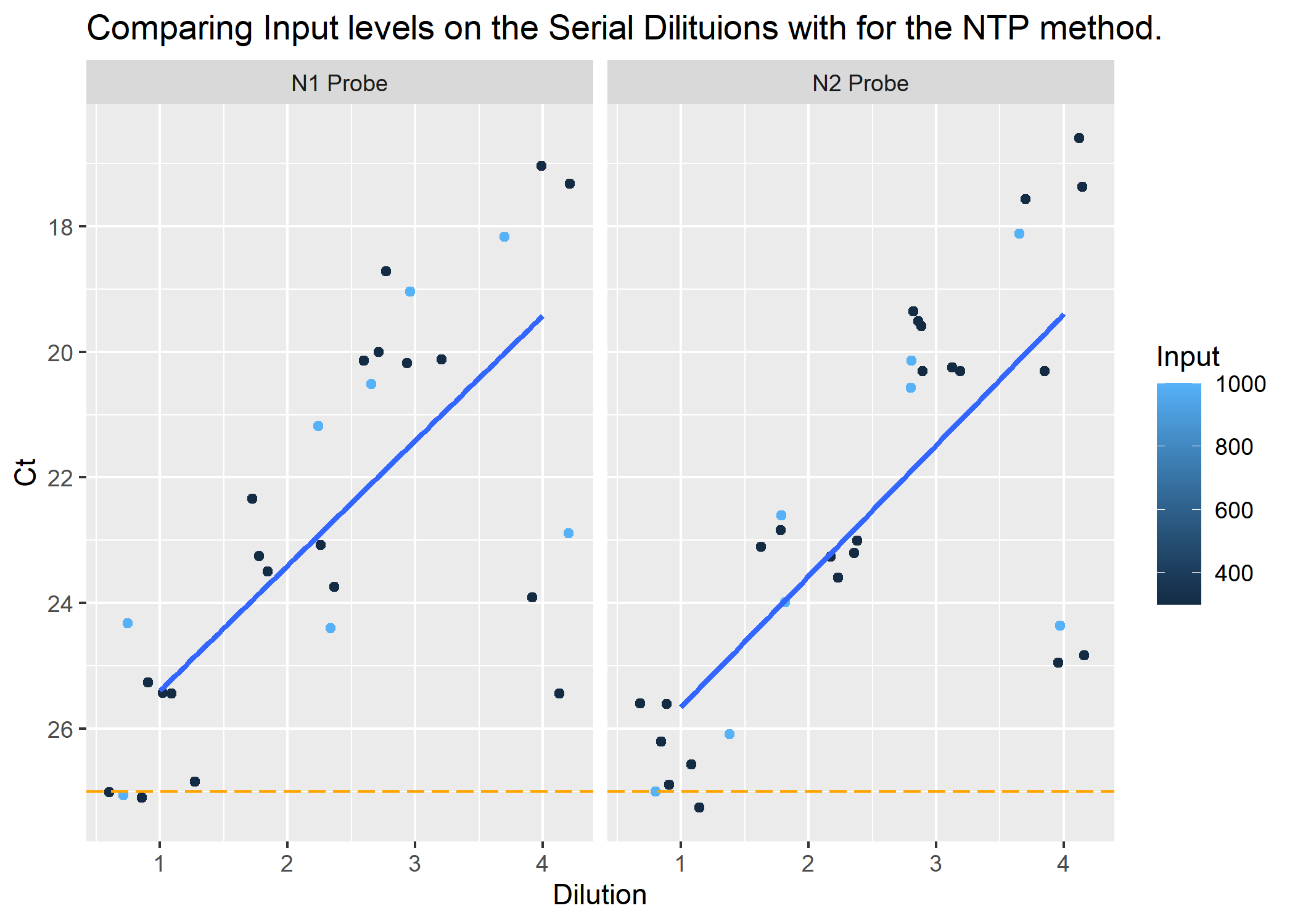
#### 2.3.3.3 Phase 2 Method extraction comparisons with all inputs and tools

knitr::include\_graphics("../../results/ASK-P2P4-MethodsALL.png")



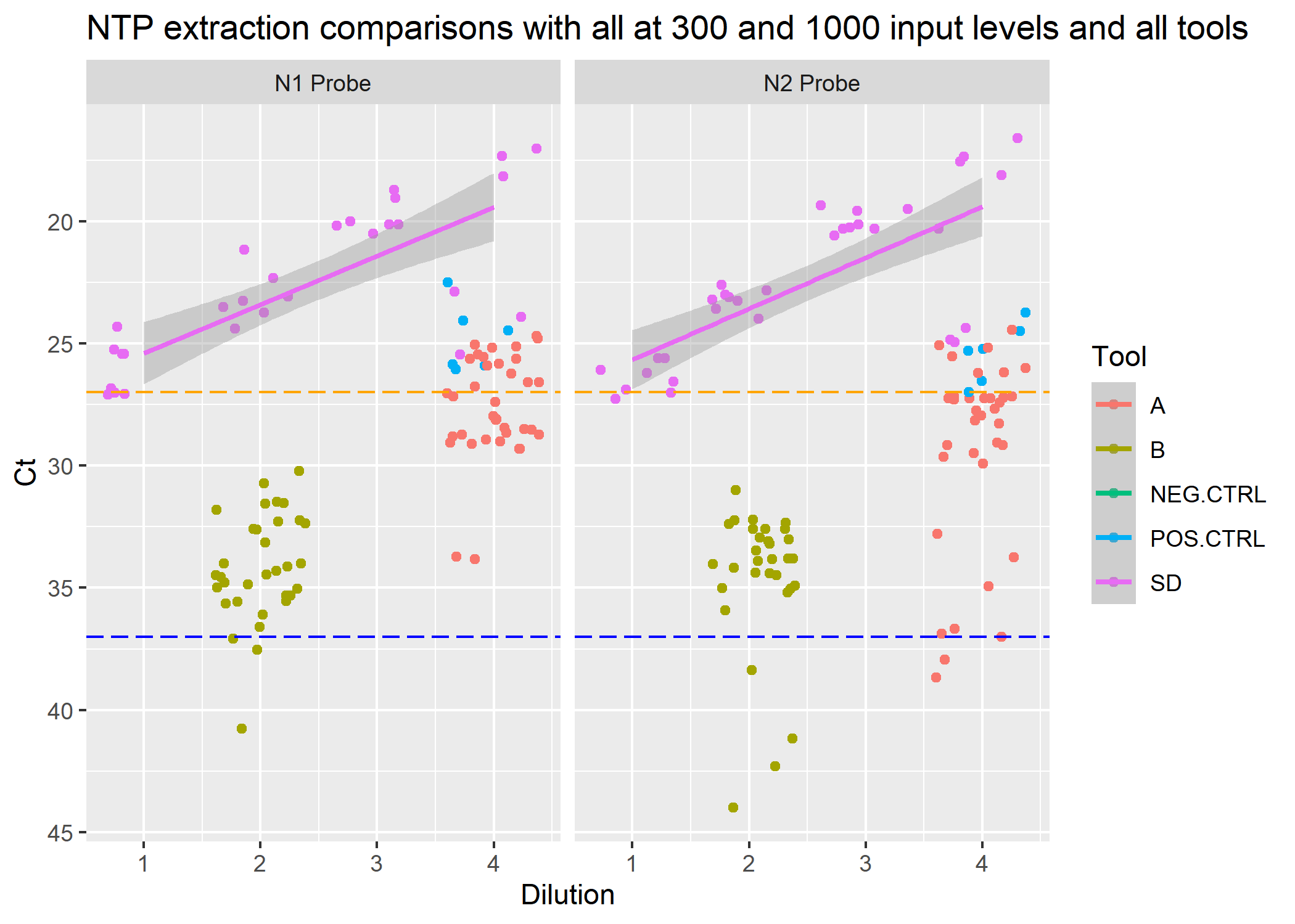
#### 2.3.3.4 Phase 2 Nanoparticles only - Compare detectors with dilution and input in Serial Dilutions

knitr::include\_graphics("../../results/ASK-P2P5-NTP-SD.png")



#### 2.3.3.5 Phase 2 Nanoparticles only - All tools at 300 and 1000ul input levels

knitr::include\_graphics("../../results/ASK-P2P6-NTP.png")



### 2.3.4 Data Analysis

### 2.3.5 Data Modeling

## 2.4 Exploratory analysis

Table ?? shows a table summarizing the data.

Figure ?? shows a scatterplot figure produced by one of the R scripts.

## 2.5 Full analysis

Example table ?? shows a table summarizing a linear model fit.

# 3 Discussion

## 3.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 3.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 3.3 Conclusions

*What are the main take-home messages?*

# 4 References