Using Bayesian inference to bridge the groundwater age divide: a lightweight data driven technique

"Sucking the marrow from data"

<u>Matt Dumont</u>, Connor Cleary, Richard McDowell, Zeb Etheridge

National SCIENCE Challenges

OUR LAND AND WATER

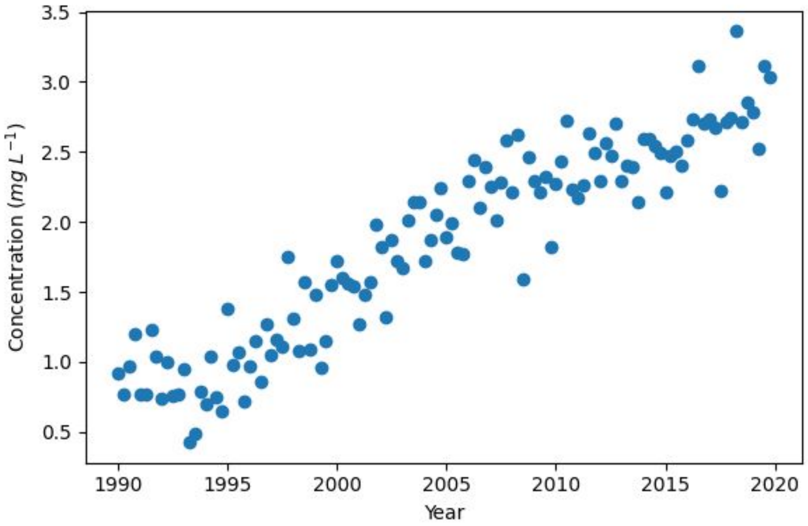
Toitū te Whenua, Toiora te Wai

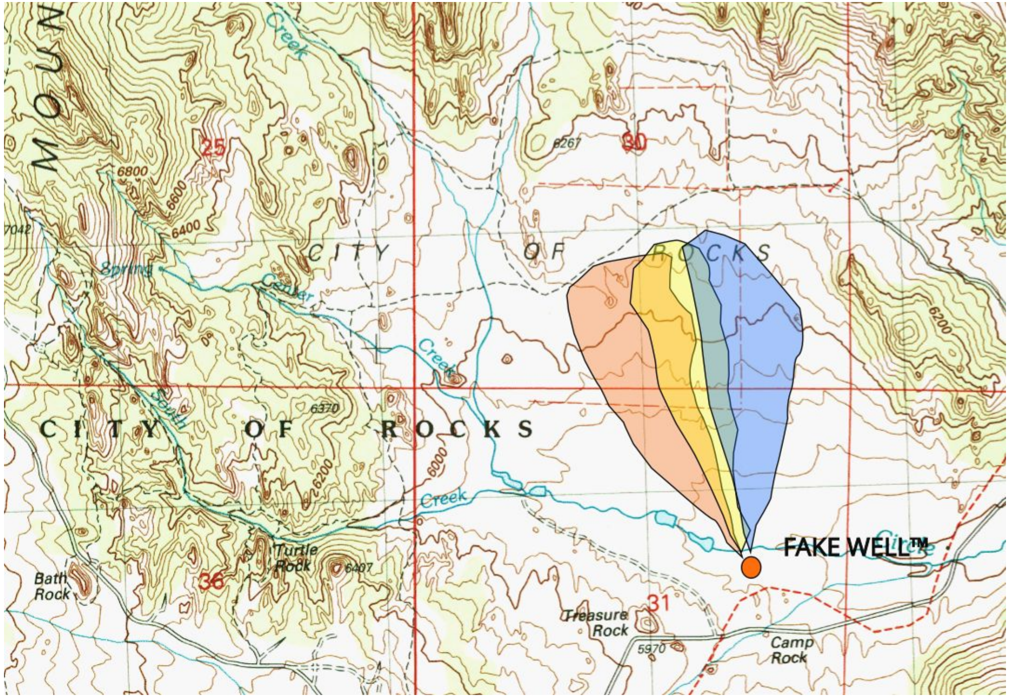
KSL Kōmanawa Solutions Ltd

A Thought Experiment

- Imagine a NO_3 time series for a Well

- We don't know the exact source area, but we have some suspicions.





- What can this well tell us about leaching concentrations in the source zone.

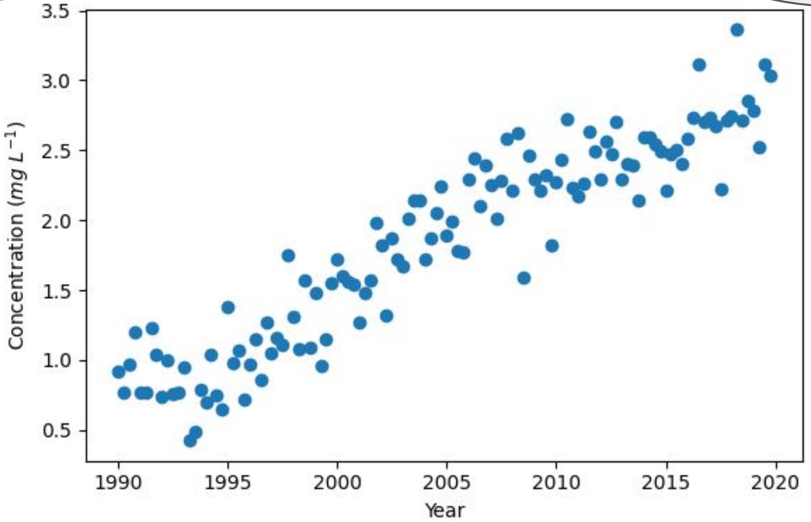
A disagreement

Left-hand Matt*

Source = 10 mg/L Now!

Right-hand matt*

Source = 5 mg/L NOW!



Who is correct (*right is problematic*):

A) Right handed Matt
B) Left handed Matt
C) No earthly way to know

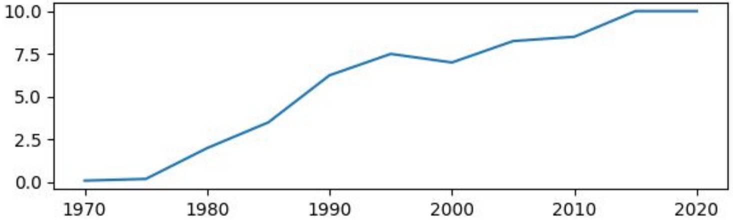
* A Chirality joke for the 1.5 hardcore chemists in the audience

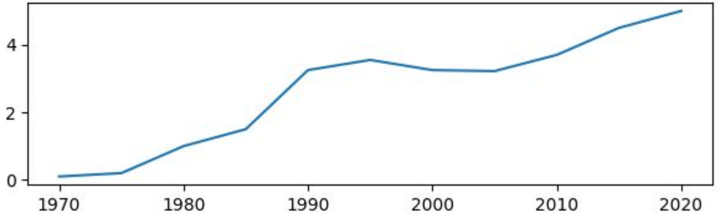
A disagreement

Left-hand Matt*

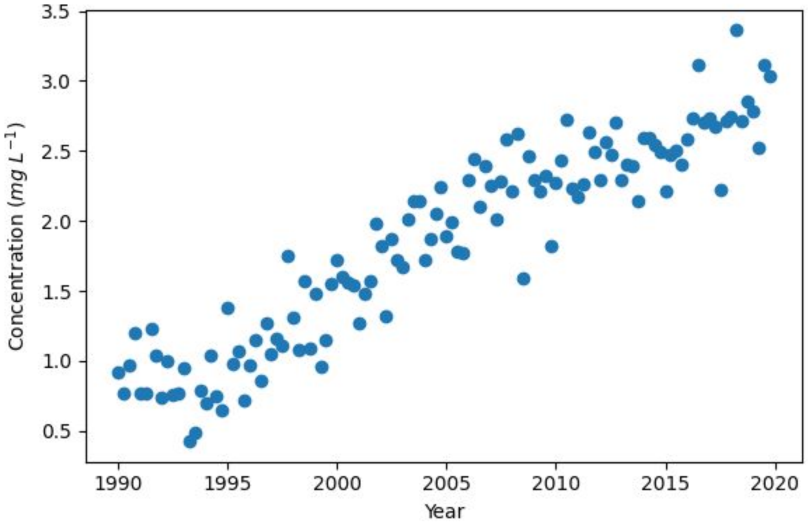
Now with Lag!©™
MRT = 25 y, P1 = 0.7 and Pathways!

Right-hand matt*





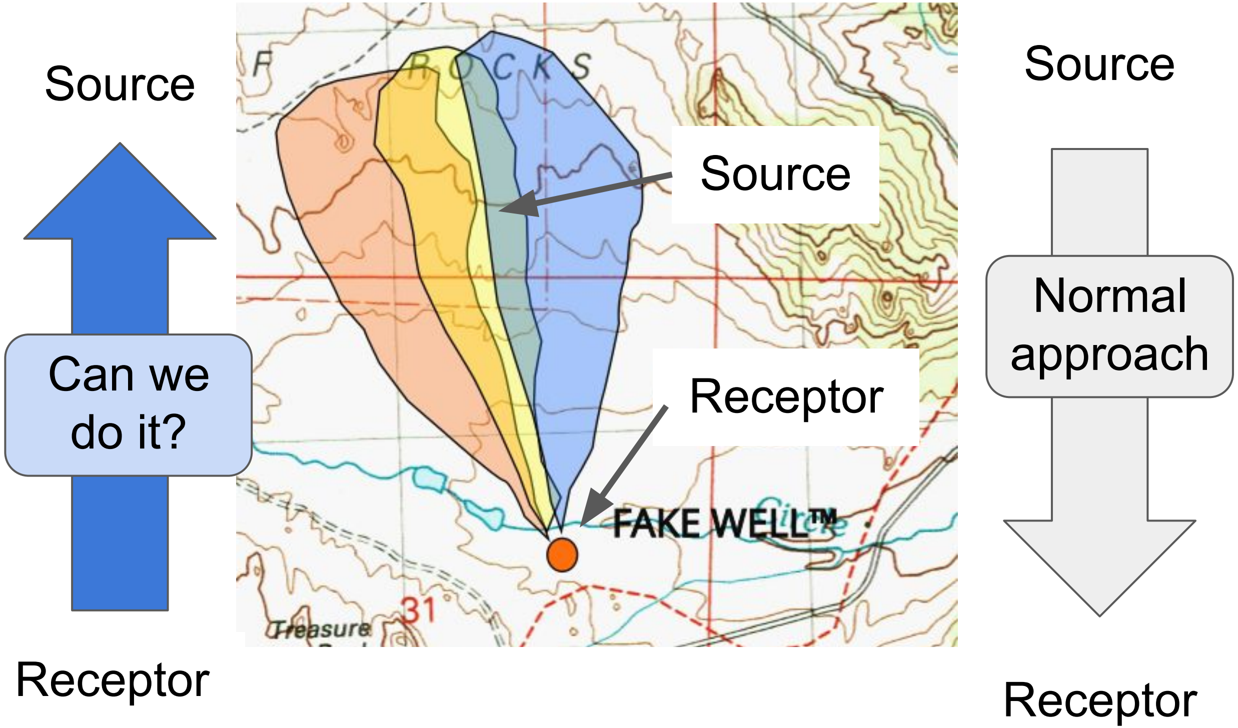
Who is correct (*right is problematic*):
A) Right handed Matt
B) Left handed Matt
C) No earthly way to know



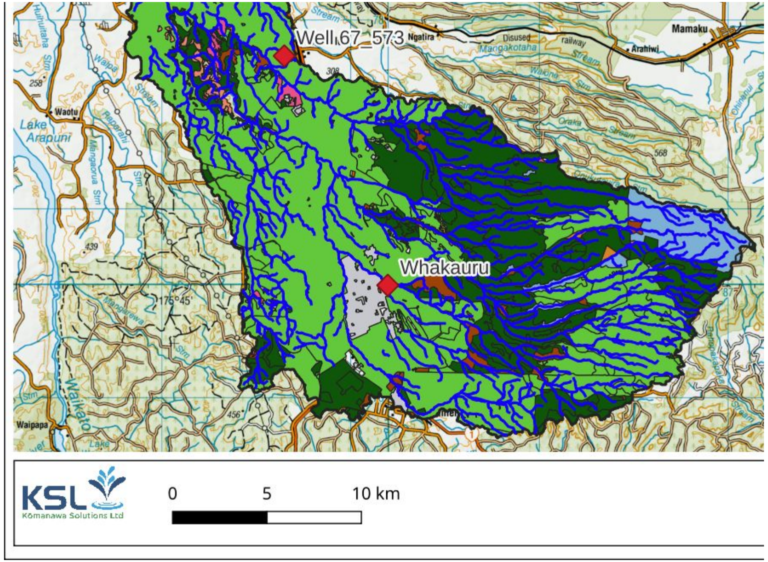
Can understand the source from the receptor?

Goal:

- Quantify the historical source concentration
- Predict future receptor concentrations

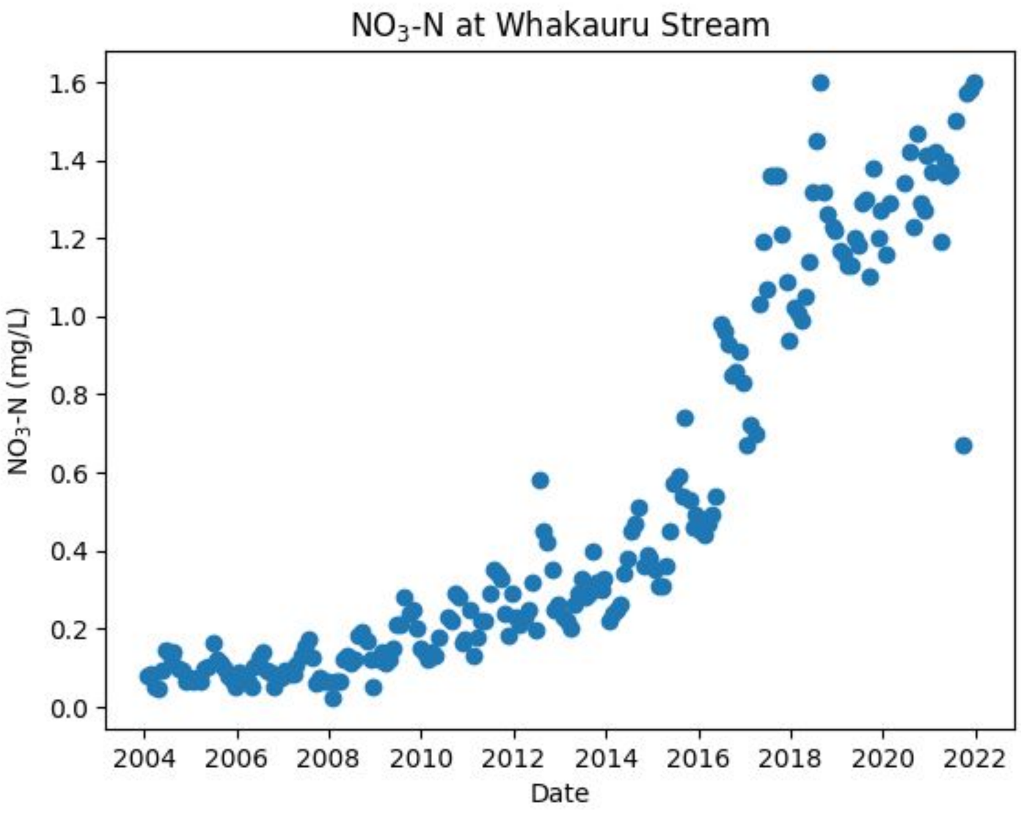


How? Let's set the scene



- Baseflow dominated hydrology
- History of intensification in some catchments, e.g. major forest to dairy conversion in Whakauru catchment in 2008-2009

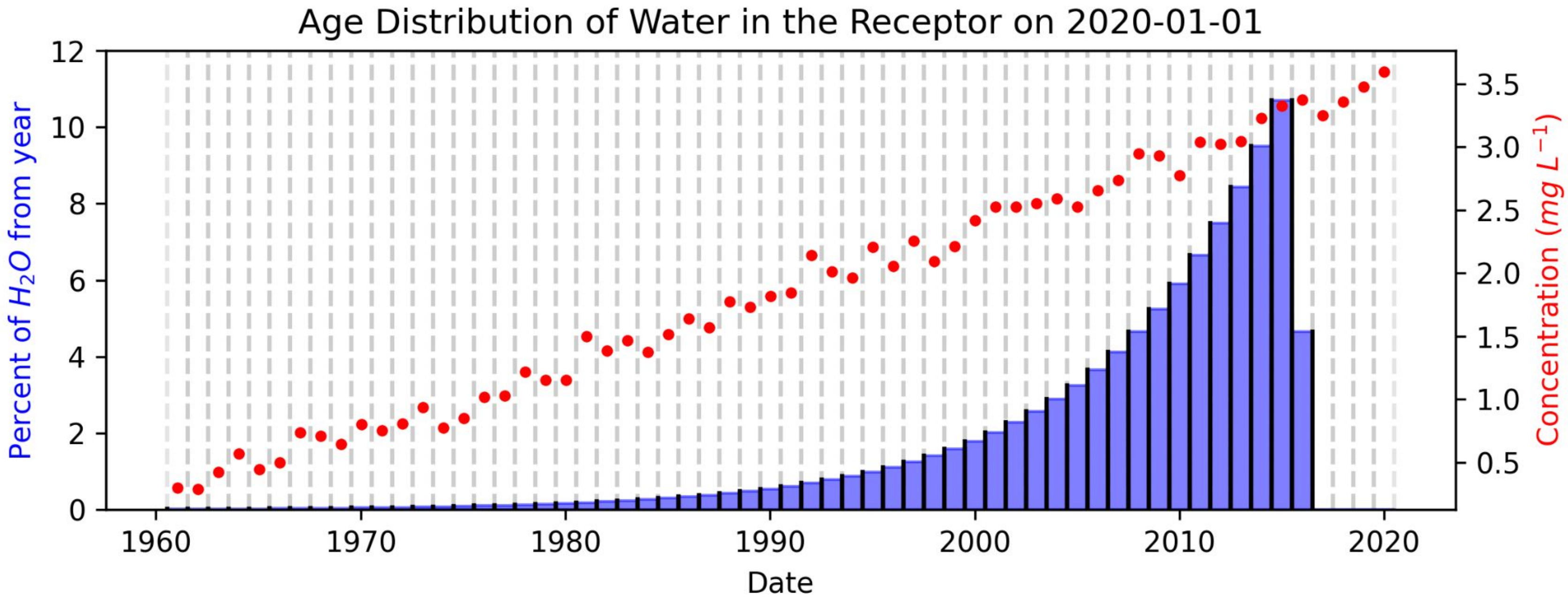
MRT = 12, F_P1 = 0.7



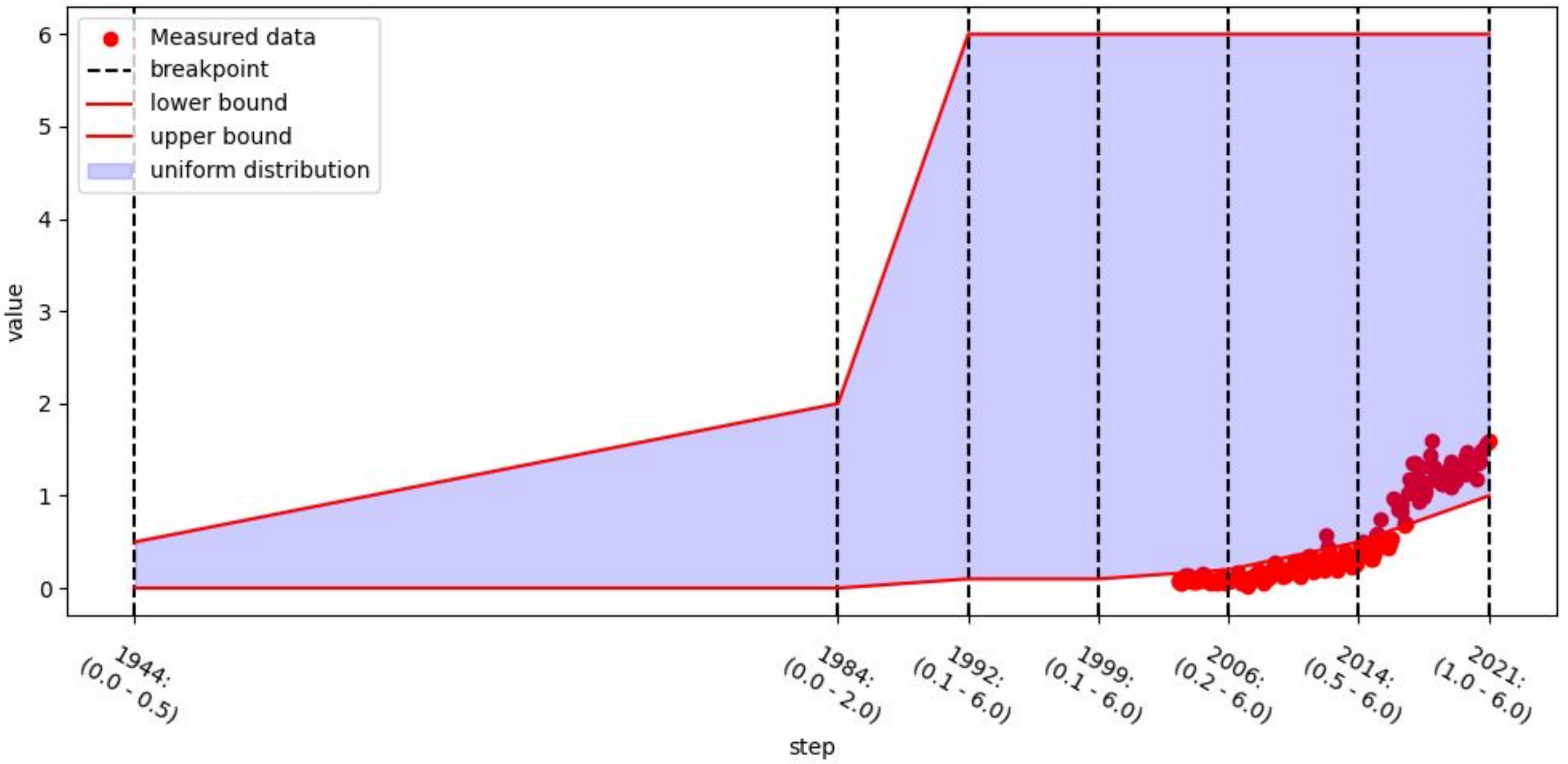
How? How do we link source concentrations to the receptor?

Age Model: Exponential Piston Flow - fit from Tritium
MRT = 12, F_P1 = 0.7

$$Conc_n = \sum_{a=0}^{p} Conc_a \times age\ percent_a$$

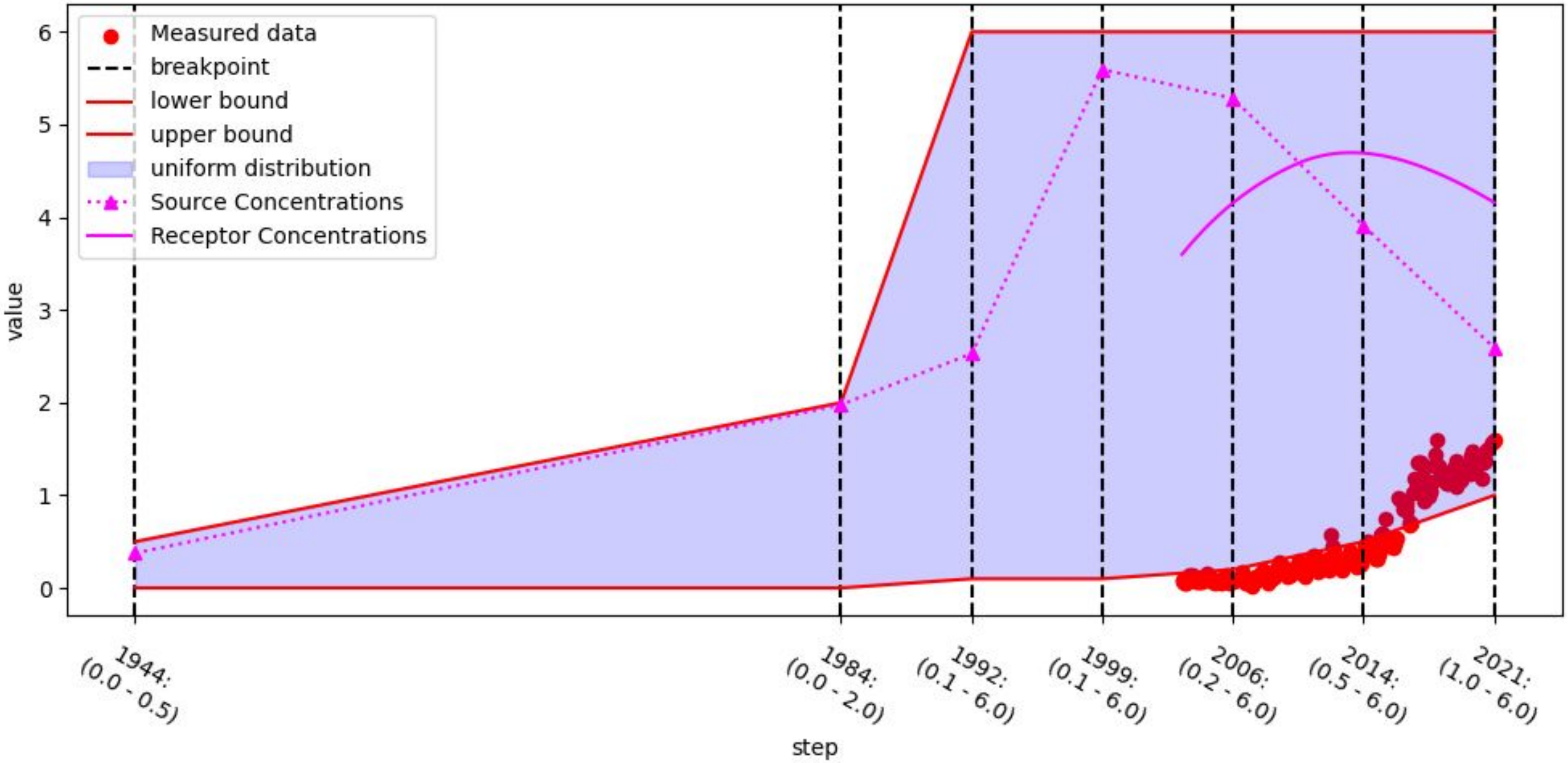


How? Consider all reasonable source concentrations (**The Prior**)



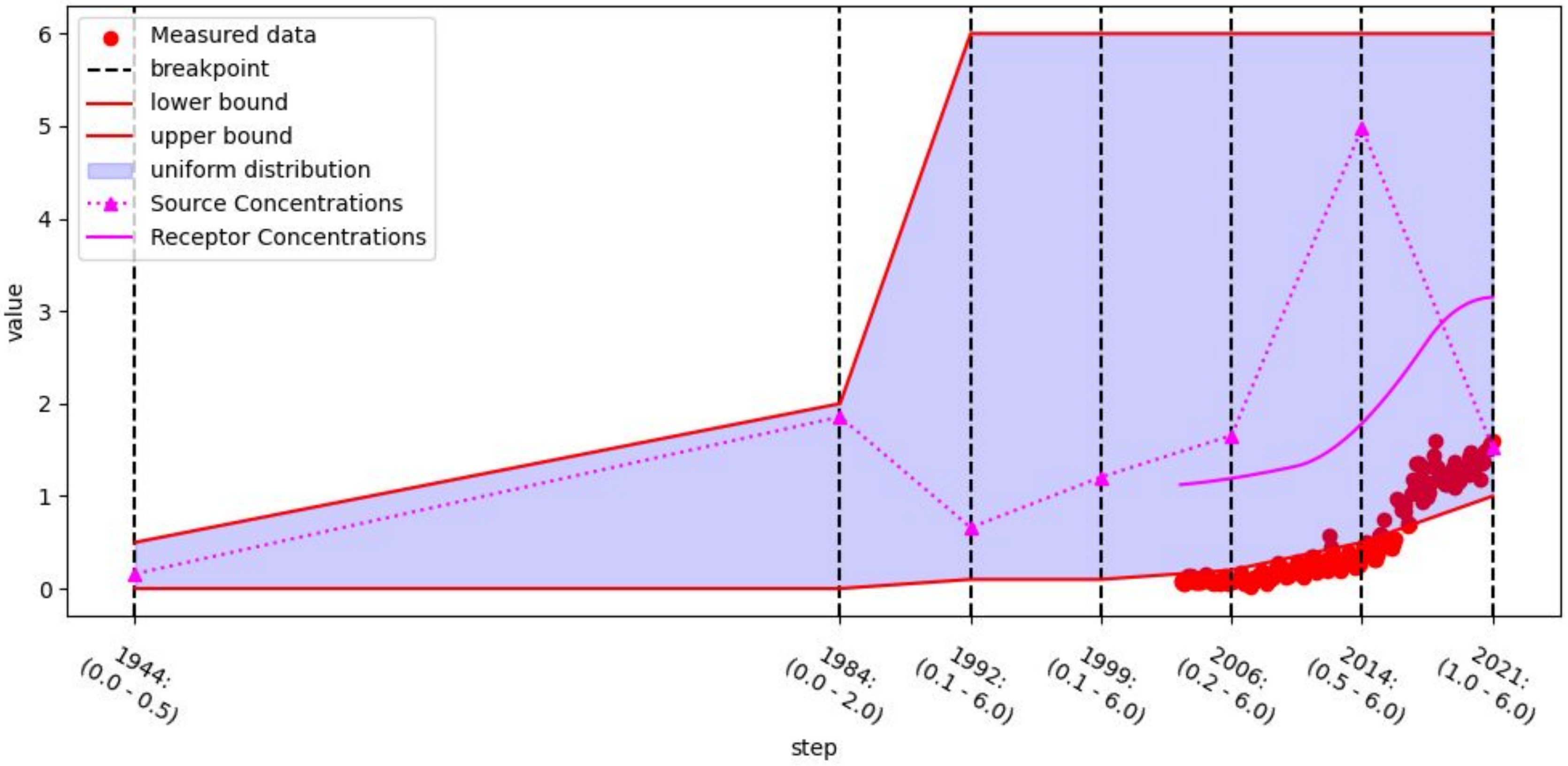
How? Let's select 1 and try it out

NOPE!



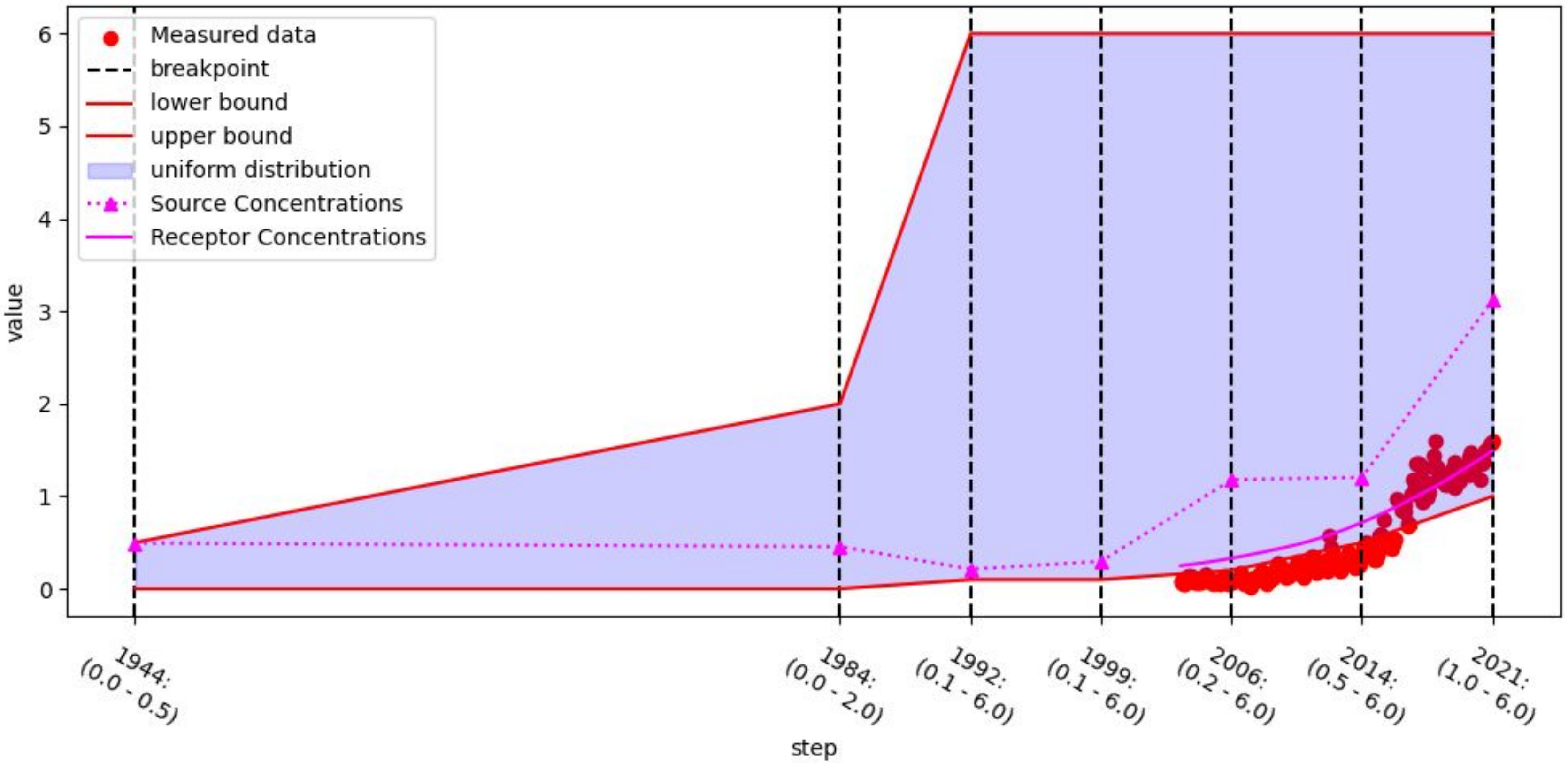
How? Let's select another and try it out

Better shape, but NOPE!



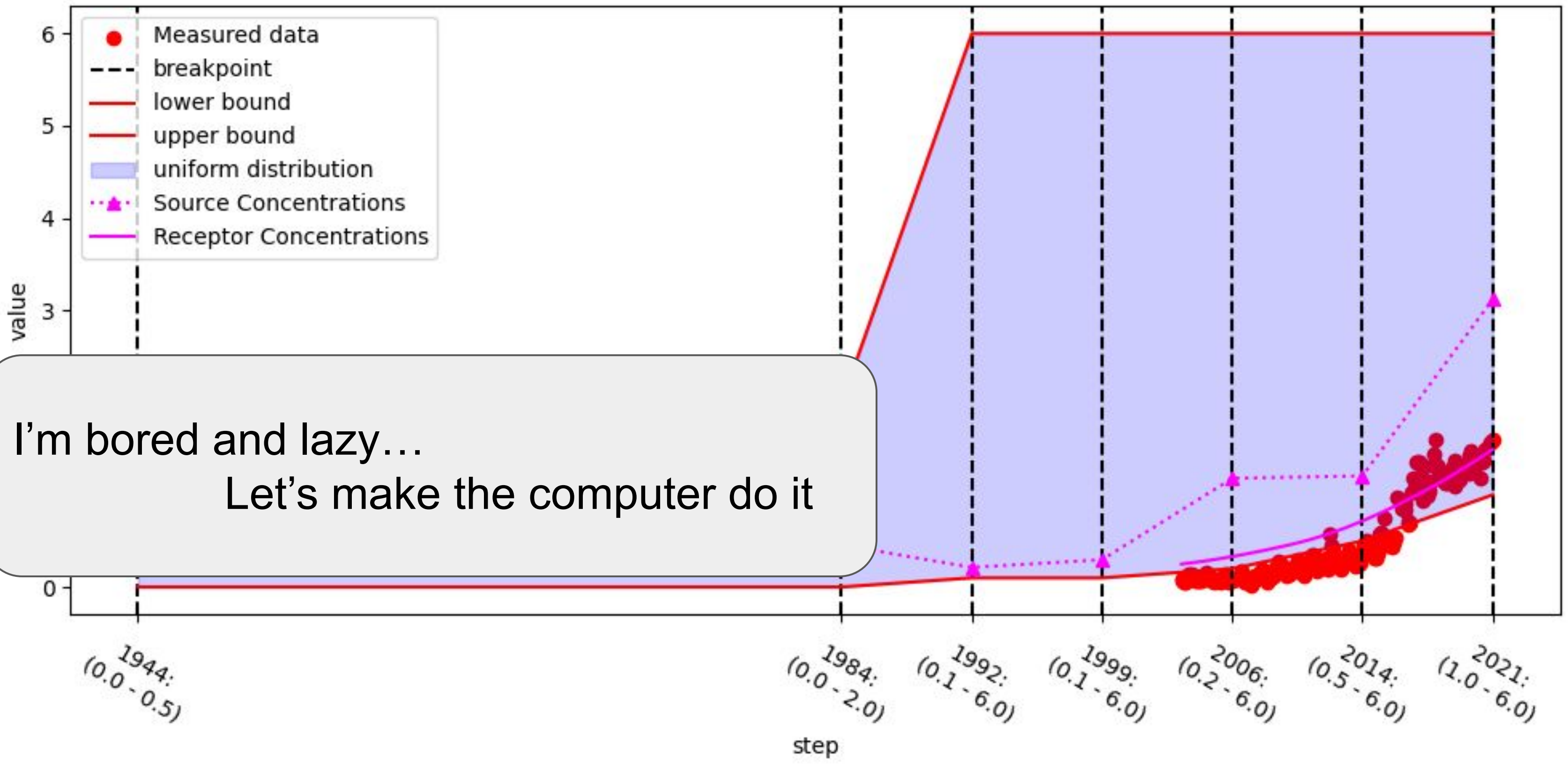
How? Let's select another and try it out

Not terrible!



How? Let's select another and try it out

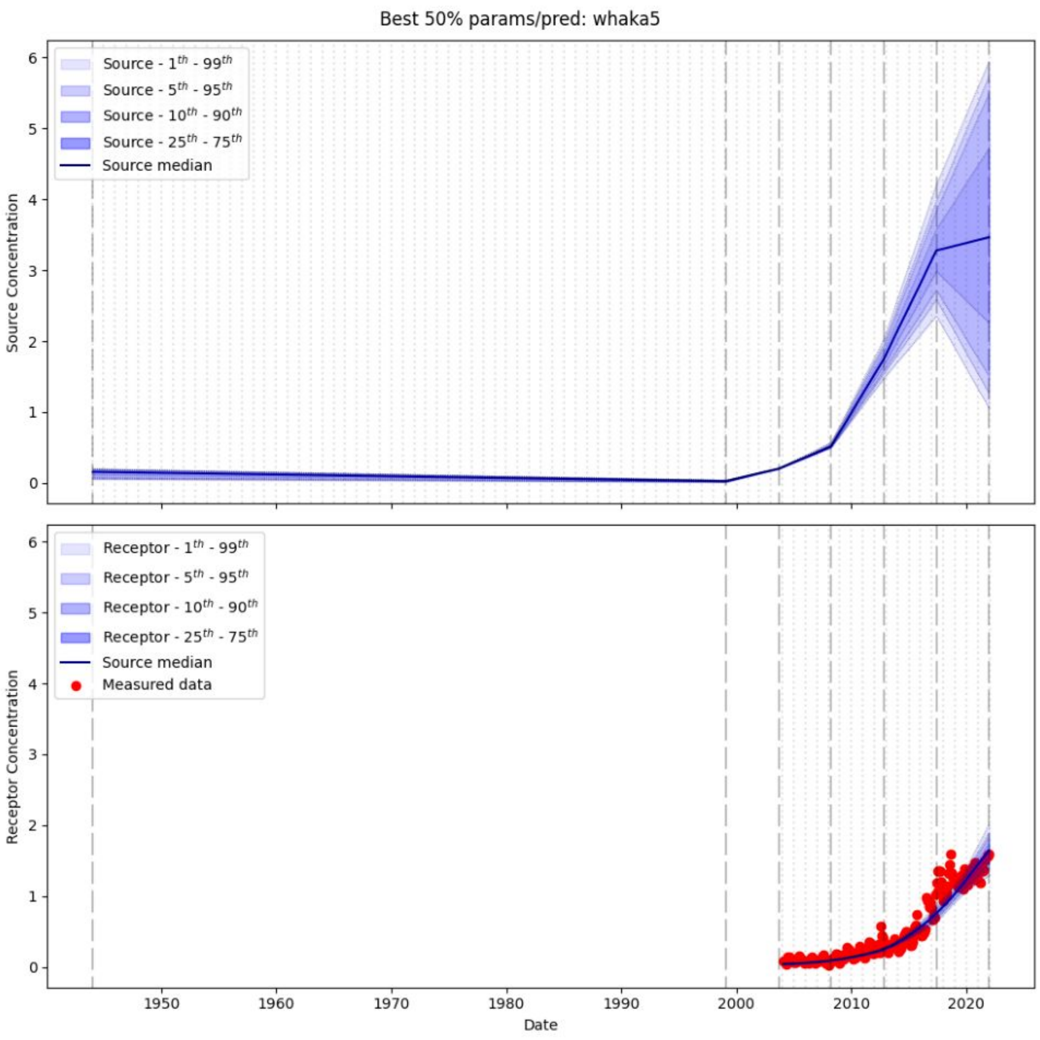
Not terrible!



I'm bored and lazy…

Let's make the computer do it

… add some
Bayesian magic
➡
THE POSTERIOR

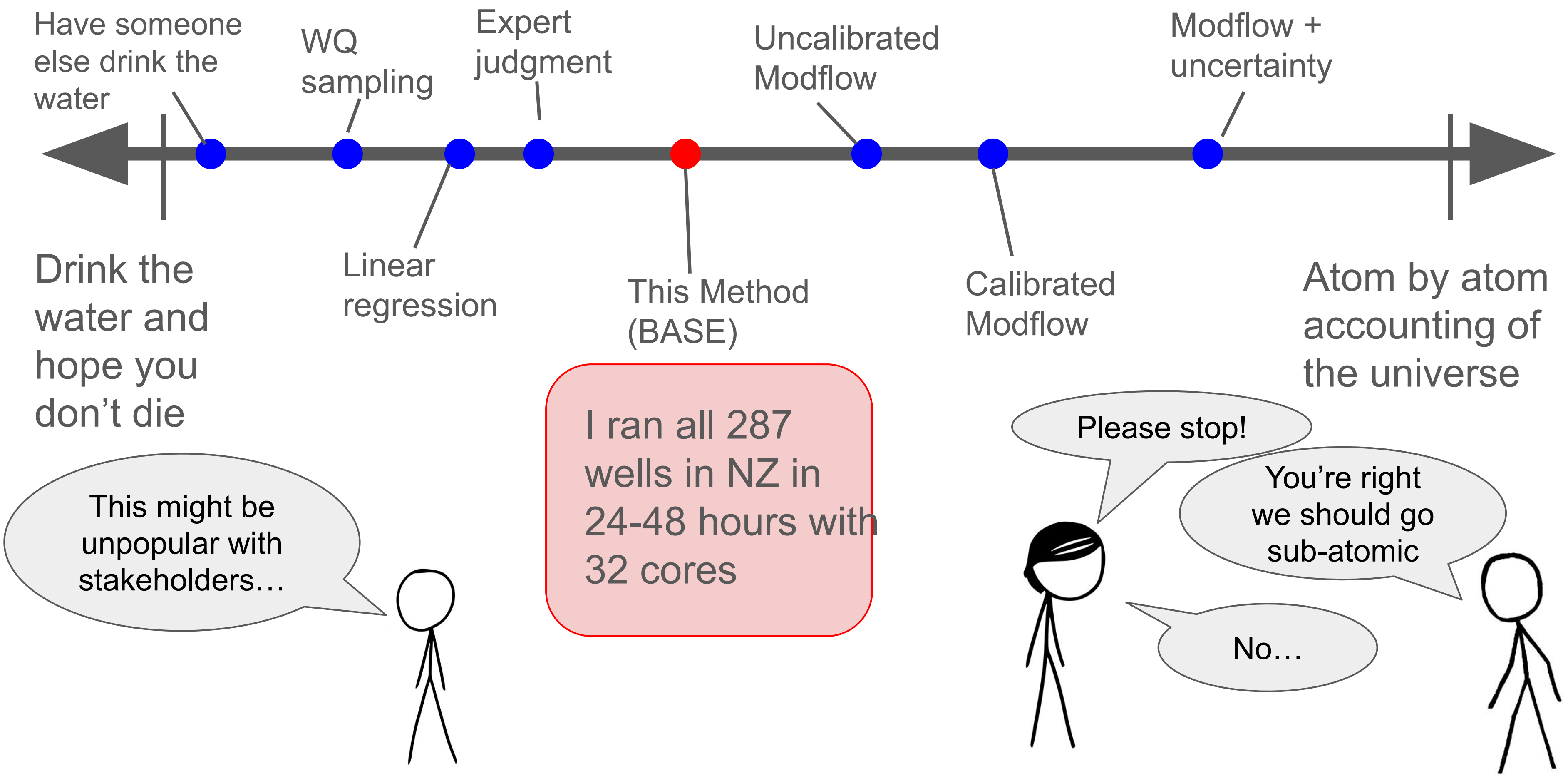


Let’s state the method assumptions

1. The source area remains static over time → *you don’t need to know where it is just that it’s not changing*
2. The age **distribution** is accurate
3. The age distribution does not change with time
4. The prior distribution includes the true source concentration

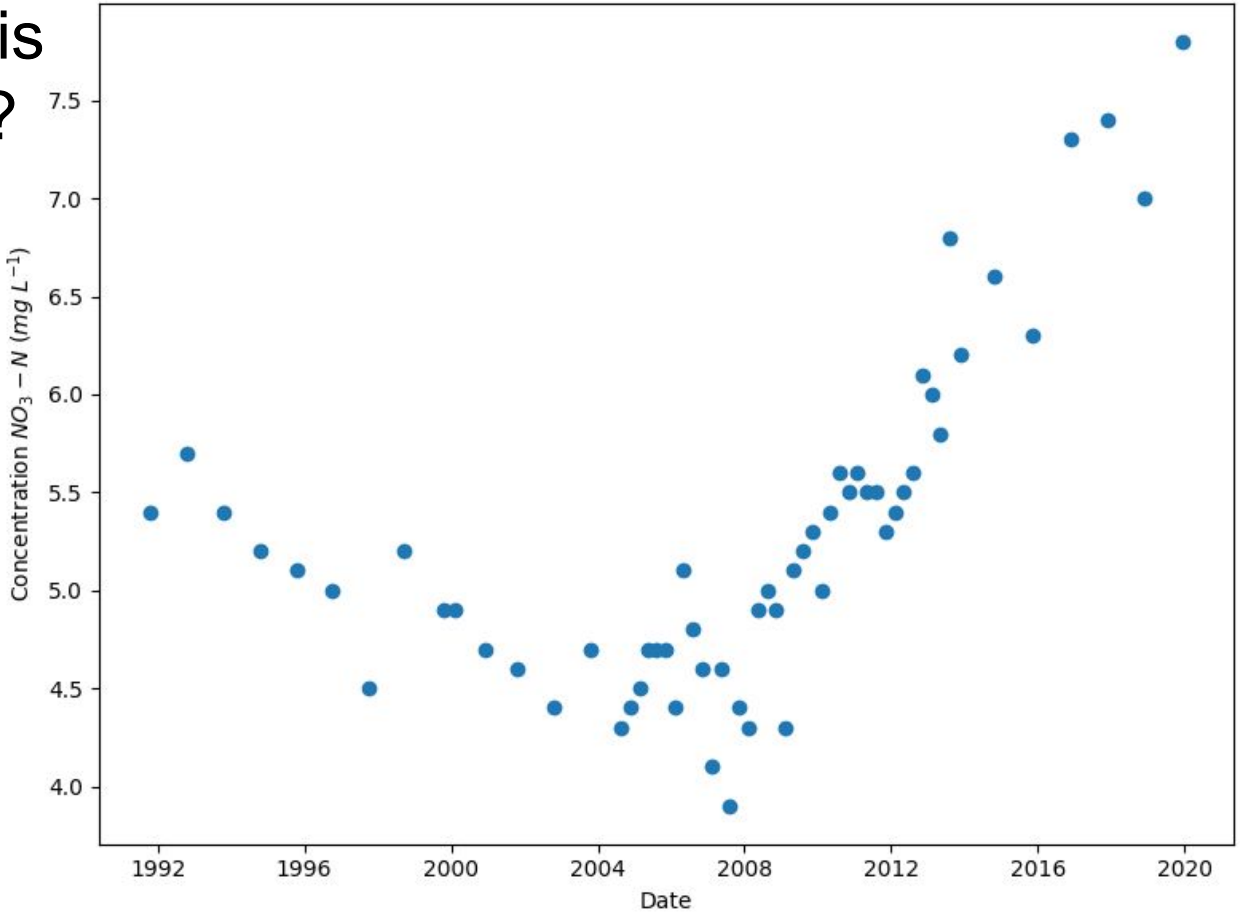
It’s data driven and does not rely on other upstream models (like Overseer)

So just how complicated is this? *(not perfectly to scale)*



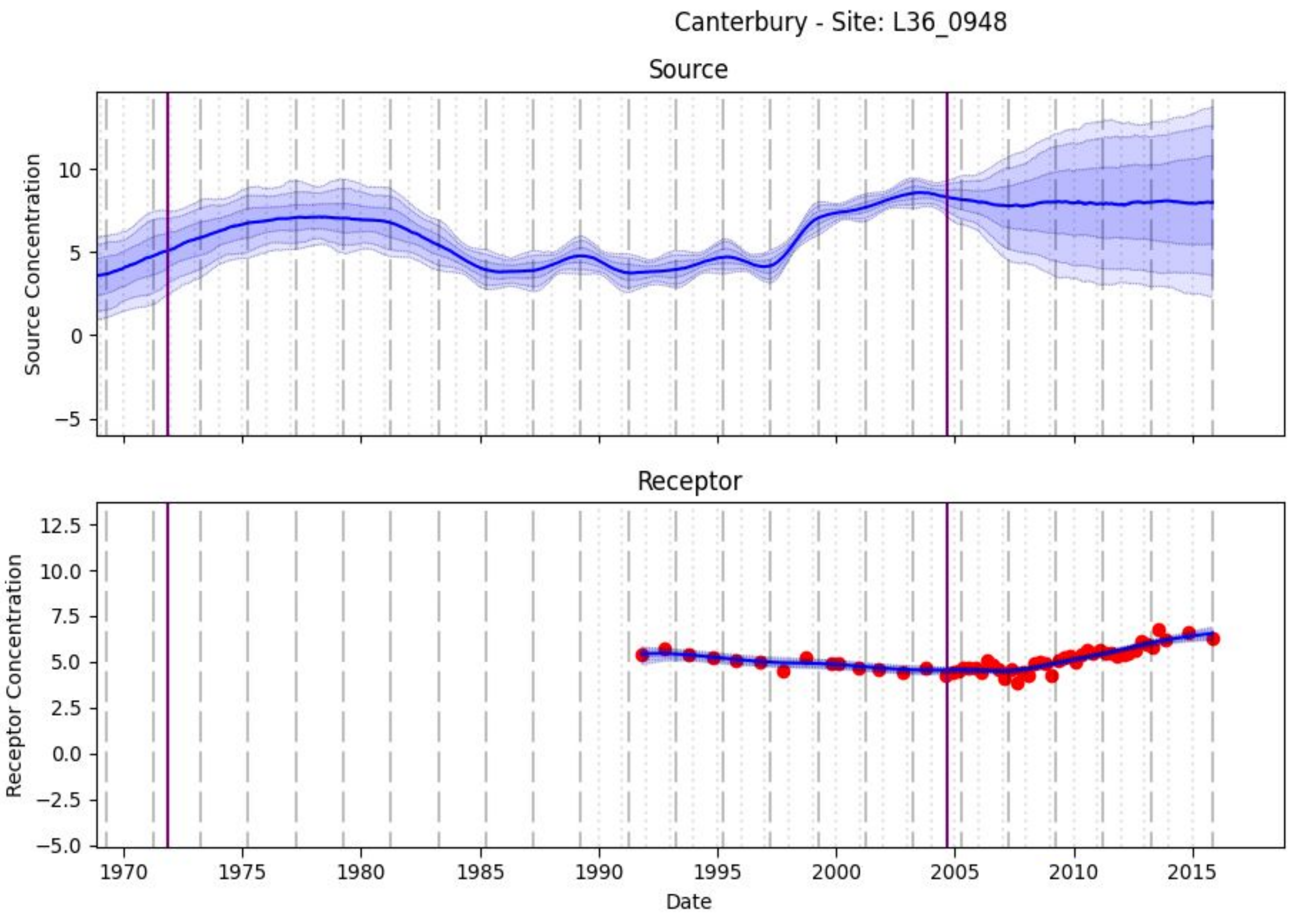
Example Problem 1
→What the @#$% is happening here?

<u>Stats</u>
Site = L36/0948
Region = Canterbury
MRT = 18.5
F_p1 = 0.5



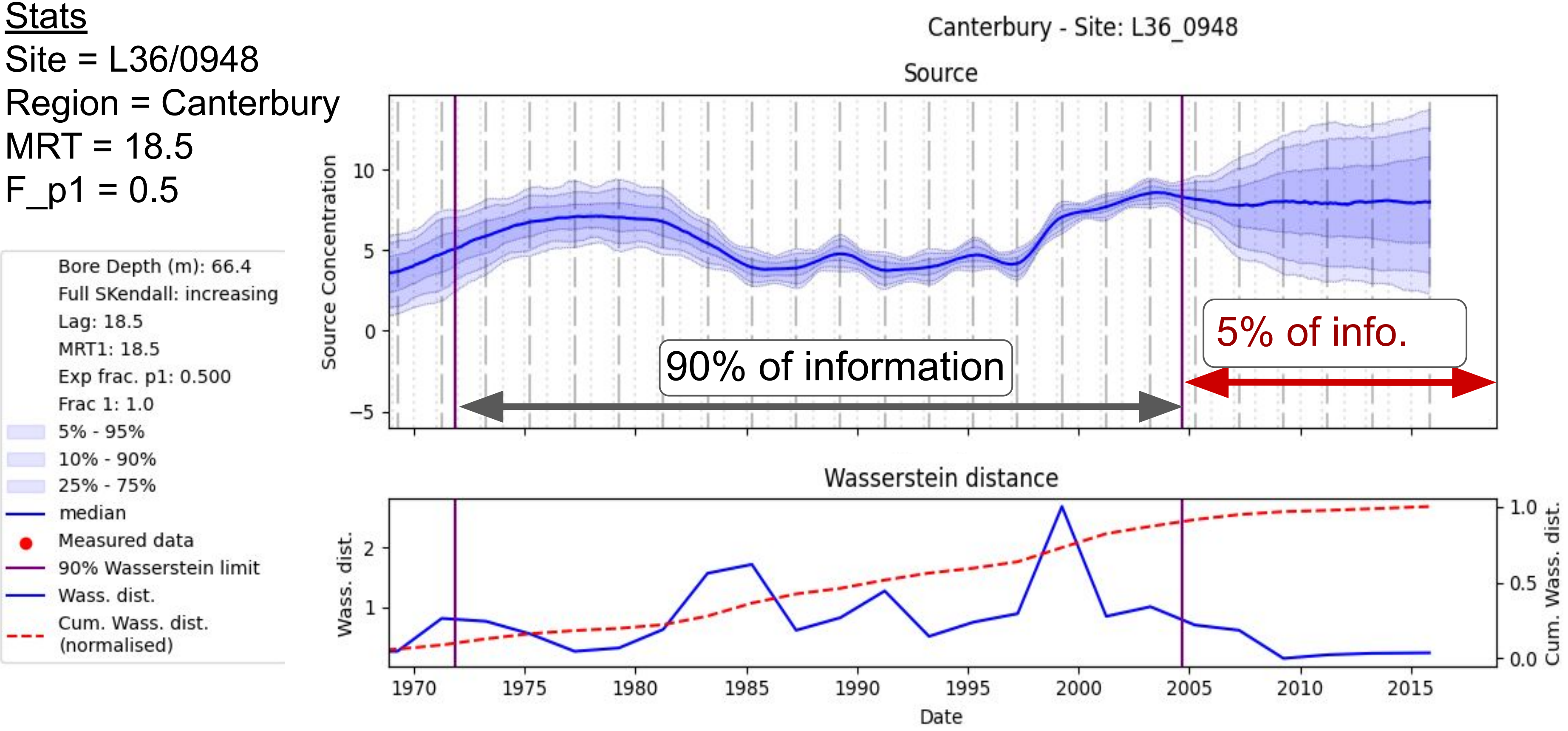
Example Problem 1→What the @#$% is happening here?

<u>Stats</u>
Site = L36/0948
Region = Canterbury
MRT = 18.5
F_p1 = 0.5

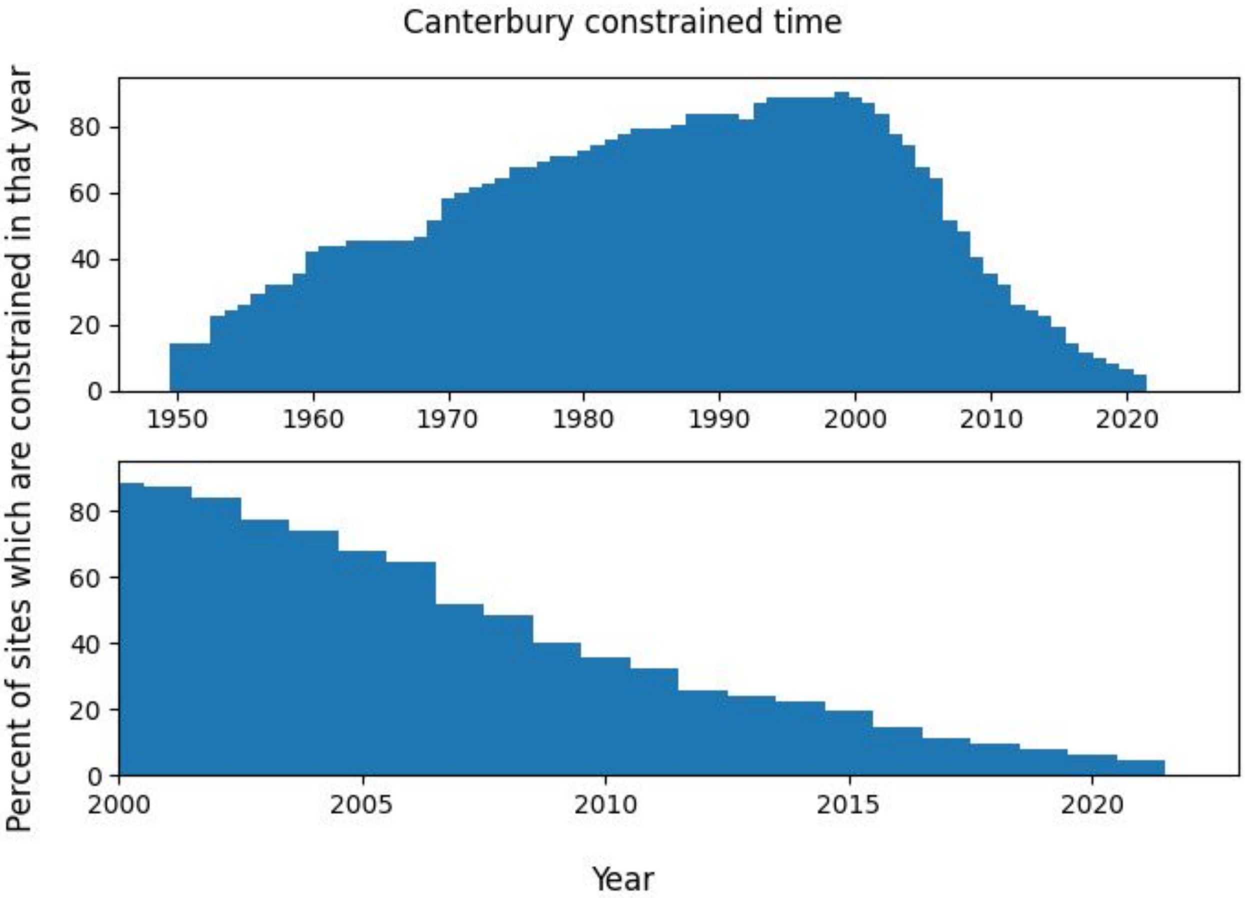


Example Problem 1 →What the @#$% is happening here?

<u>Stats</u>
Site = L36/0948
Region = Canterbury
MRT = 18.5
F_p1 = 0.5



Now let's go for something a bit more *Majestic*

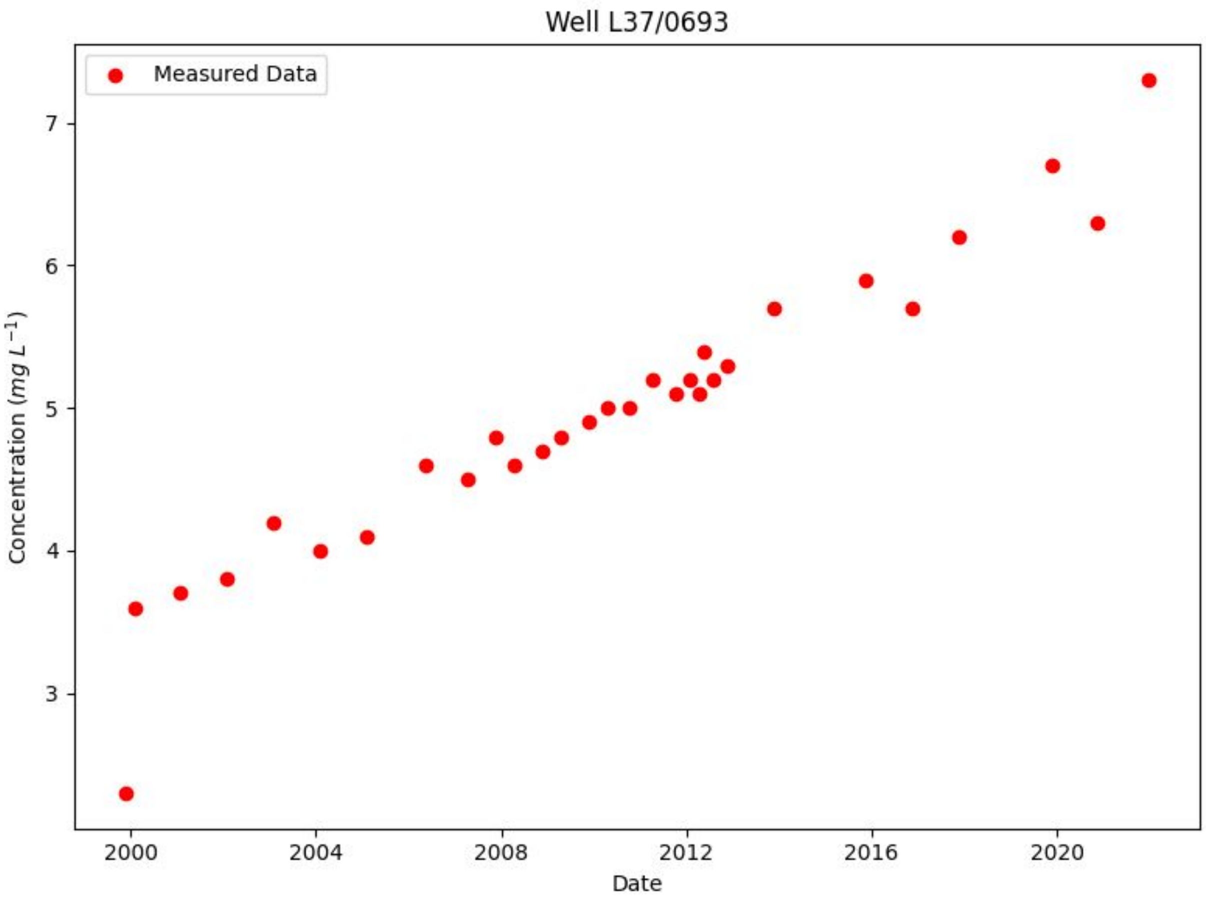


When are we sampling?

Example Problem 2 →Where are we going?

What will steady state NO_3-N be?

Stats
Site = L37/0693
Region = Canterbury
MRT = 46
F_p1 = 0.5

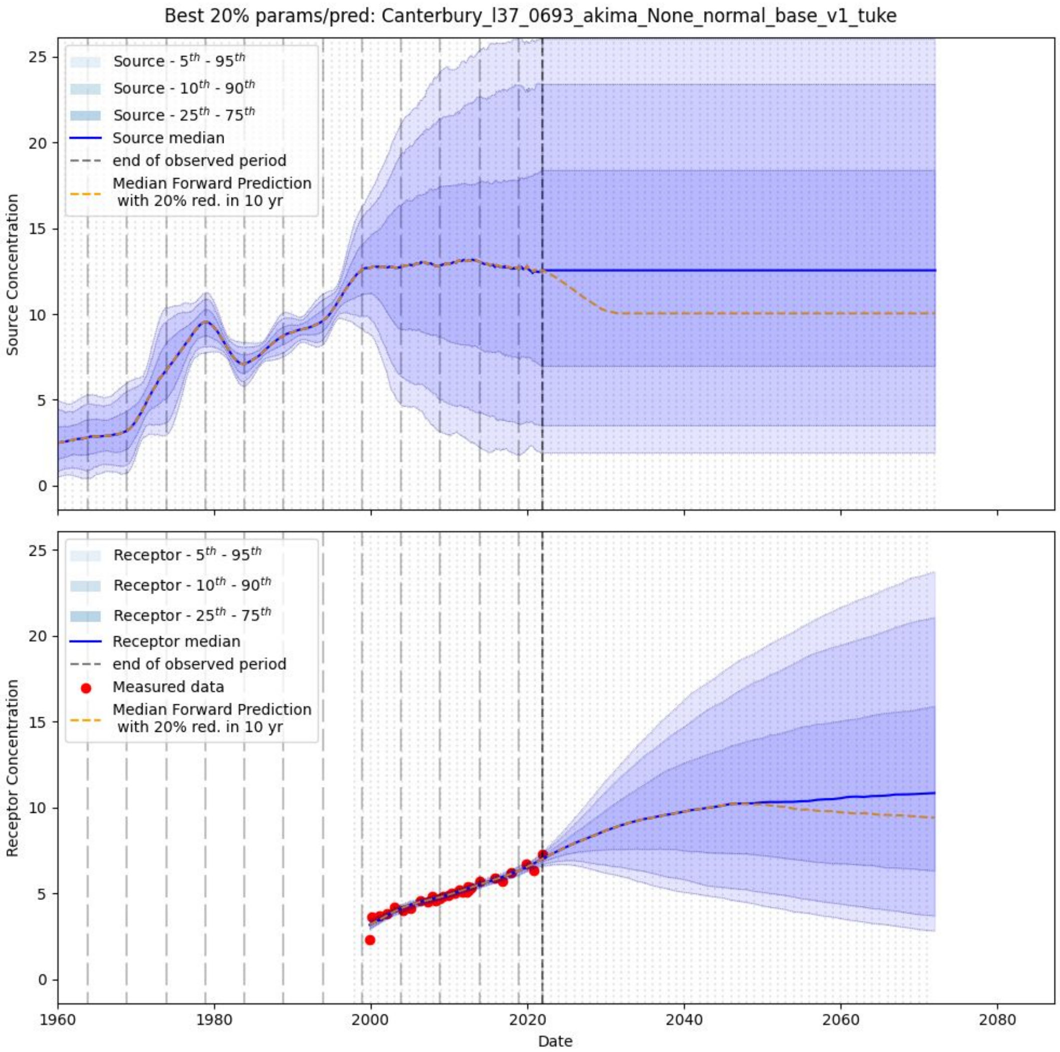


Example Problem 2
→Where are we going?

What will steady state NO_3-N be?

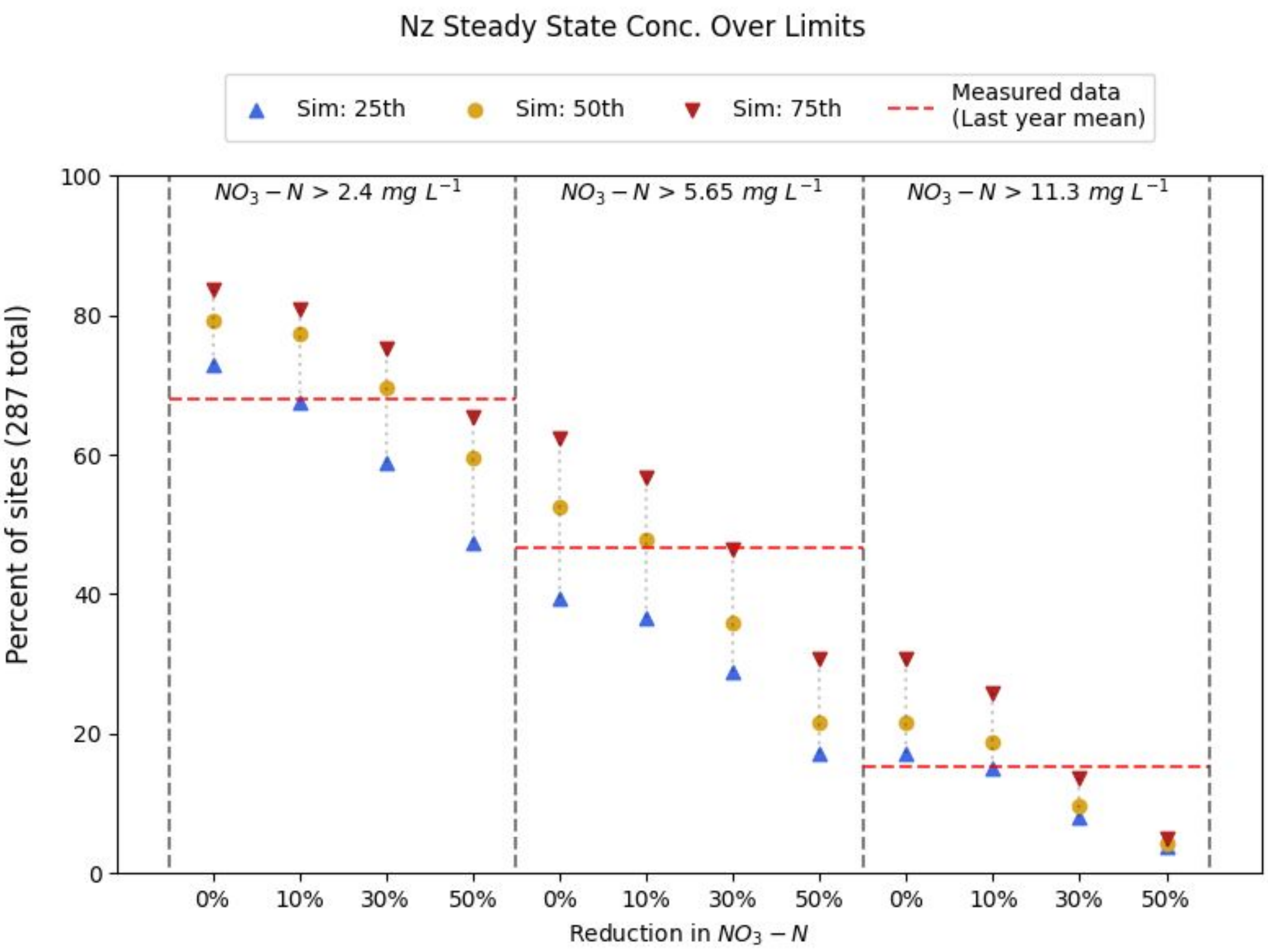
If we assume minimal changes: ***median of 12.9 mg/l***

<u>Stats</u>
Site = L37/0693
Region = Canterbury
MRT = 46
F_p1 = 0.5



Now let's go for something a bit more *majestic*

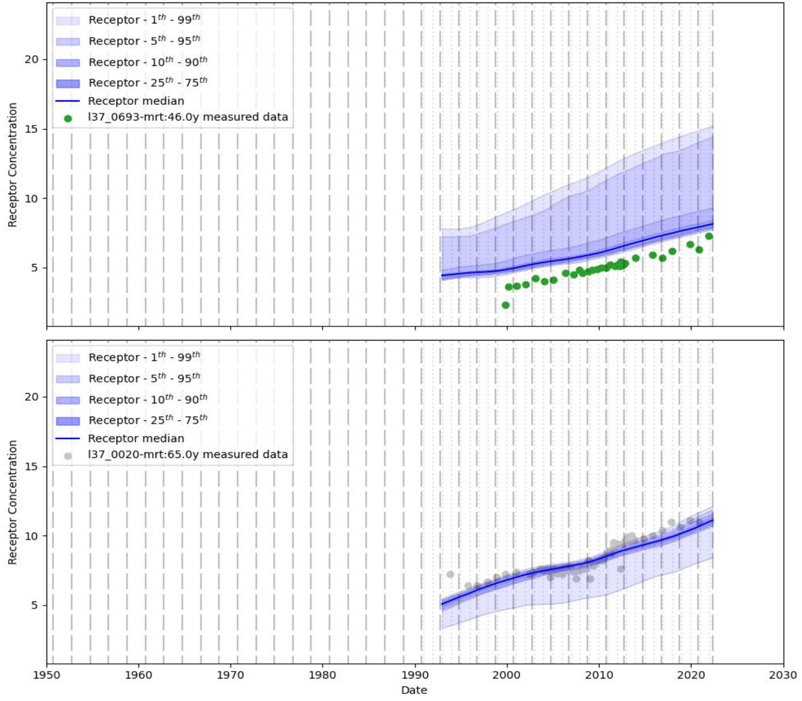
What will steady state NO_3-N be?

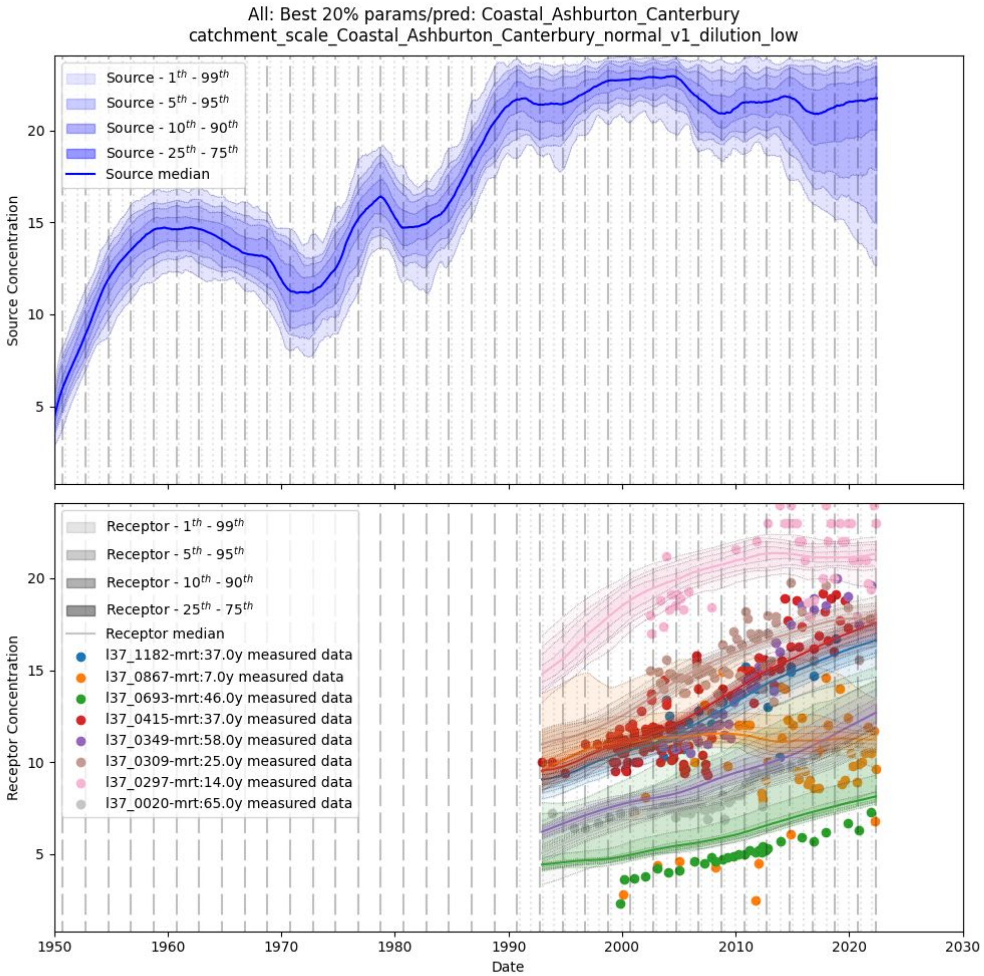


Conclusions

1. Our BASE technique is can allow the user to:
   a. Estimate the historical concentration of source zone (leaching area).
   b. Give an estimate of the likely long term (steady state) concentration of a well.
   c. Predict how changes in the source zone will affect concentrations in the well.
2. BASE is data driven and does not rely on upstream models like Overseer.
3. It is quick to run and is much more lightweight than other models (e.g., Modflow).
4. It can provide an independent parallel line of evidence for forward modelling.
5. Other information can be brought into the prior to test whether the prior is consistent with the observed data.
6. Open source python package (soon): https://github.com/Komanawa-Solutions-Ltd/komanawa-BASE
7. We are in the process of writing a paper on the technique.

Example Problem 3
→Are these wells showing the same thing?





Example Problem 3
→Are these wells showing the same thing?

Ashburton