4/6/23 fs / regression thinking

recap so far,

- actually sampling is computationally expensive
- seems to be reasonable support (see moments.pdf) for an approach of modeling kill 1 (at least) as $N()$, and then it becomes a problem of predicting μ, σ from S, N.
- this is a (probably) predictable but highly nonlinear relationship.
  → a linear model is not useful b/c it does not capture things w/ high enough fidelity to make any kind of reasonable test.
  seems like a job for a ml algo & some description.

- so far I have made a rudimentary try w/ random forest regression, which to my (subjective) eye is better than the lm but still, I don't think, close enough to be useful for inference. In particular it has biased failures @ high/low hill vals and w/ some weird structure over SxN space.

  some of that may be // the biases in the training / indeed whole sample set. To my understanding RF isn't great at out of sample predictin. Which, may be an issue if what is the gain if we have to sample w/ very high density? unclear from where things stand now.

  so I guess I see 2 possible next routes...

a) see if RF improves w/ denser sampling

OR

b) try a method w/ better ~learning & oosi to my understanding the place to start there would be keras. which is way easier to use direct from python than via reticulate.