A quick test drive of randomForest on a
smallish set (200 training) does OK but
seems to have some degree of biased error.
High var explained (97%).
compared to a (v. rudimentary) linear model,
there's less (obvious) structured error.

→ IDK what would happen predicting far out of
sample.

Looking at some 1:1 plots

hill 1 from PS

1:1

hill 1 from $N(\mu, \sigma)$ w/ pars from RF

some are great and some are systematically
not.

→ Is that something that would be improved with a) more values or b) a more nuanced algorithm? (e.g. NN instead of RF?)