Weight Forecasting Models

Jamie Williams - March 2021

Models

- Model #1: 7-day rolling average, Markov approximation
  - Model #1.0: simple linear regression (MLE, not Bayesian)
  - Model #1.1: Bayesian probabilistic model, but doesn't model predictor uncertainty
  - Model #1.2: Bayesian probabilistic model, includes explicit treatment of measurement uncertainty of predictors (calories, steps)
- Model #2: day-to-day changes (no rolling average)
- Model #3: for energy expenditure, use activity minutes instead of steps: minutes of 'very active', 'fairly active', 'lightly active', 'sedentary'.

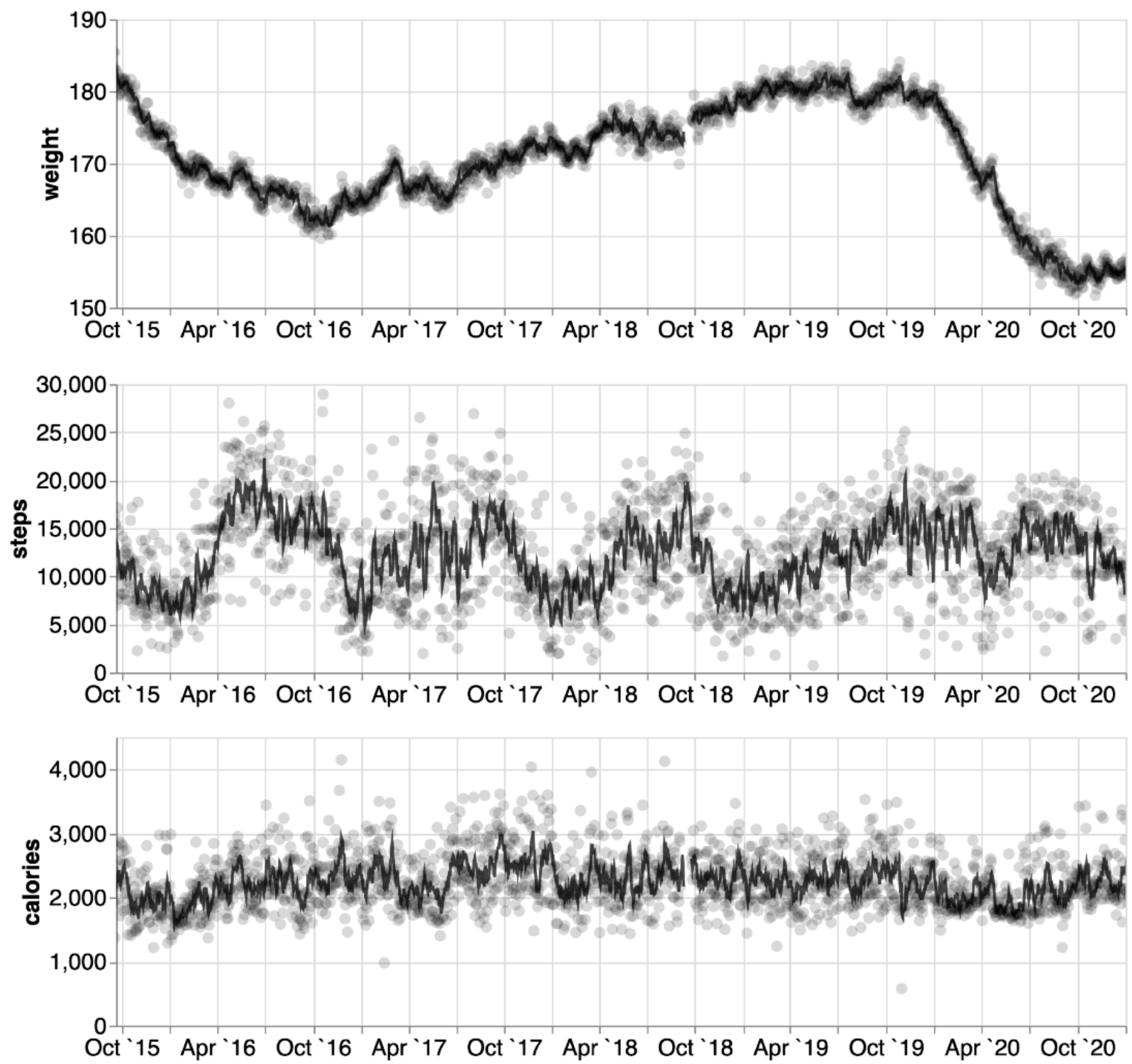
⋮

Model #1.0

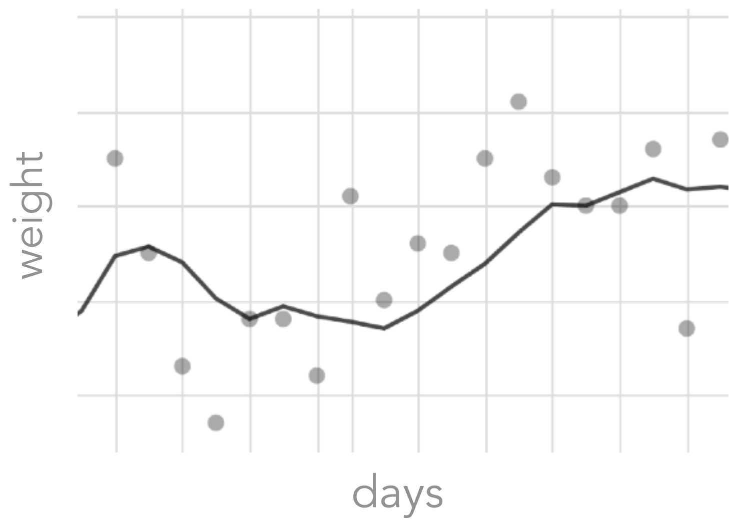
(implemented)

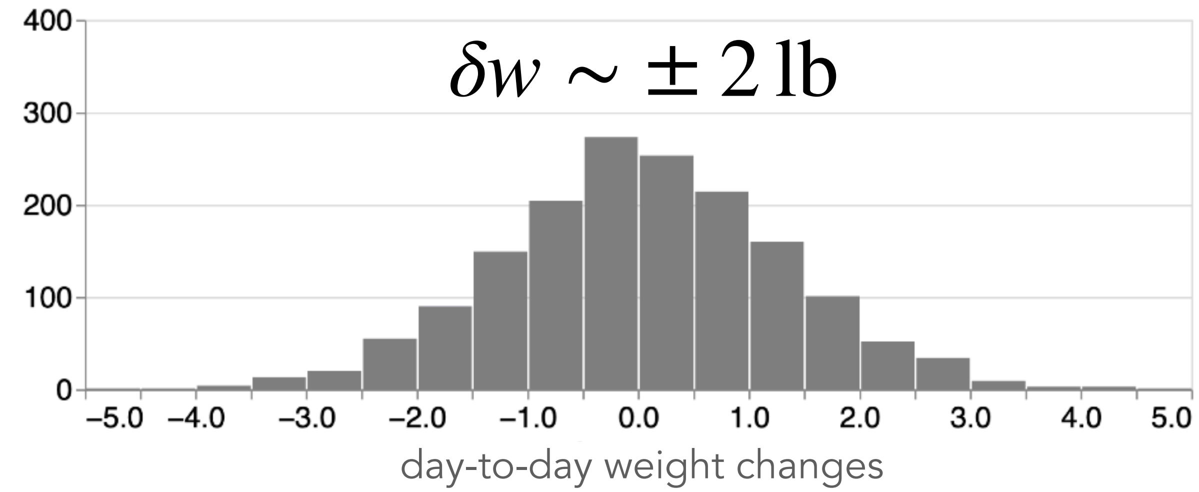
Train a simple regression model

I have over five years of daily measurements.



Day-to-day weight fluctuations

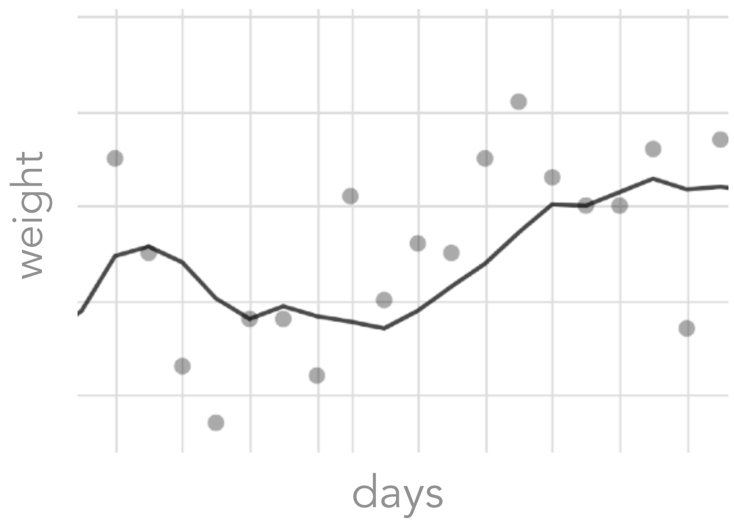


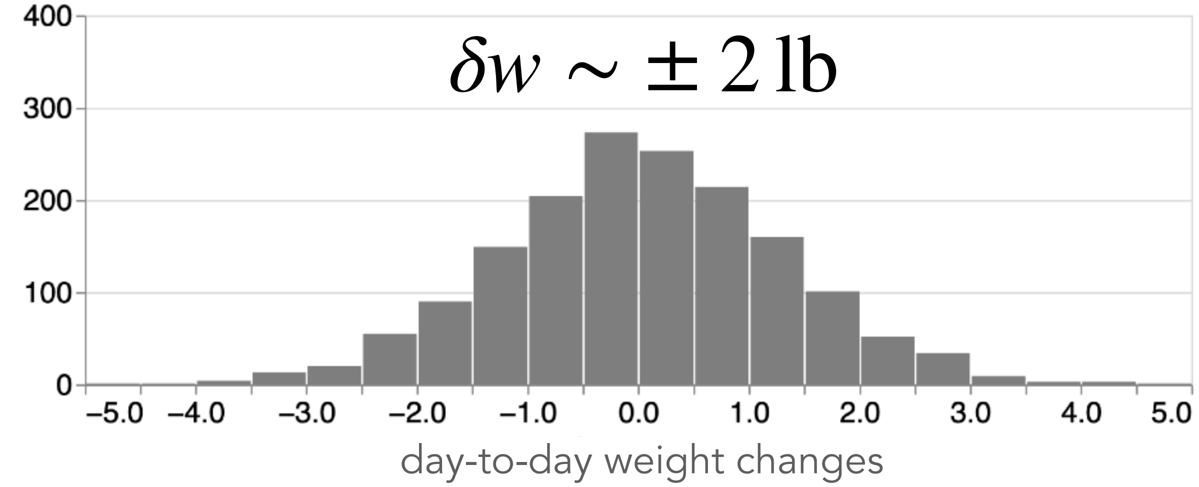


Throughout the day, my total weight can **vary** due to various processes:

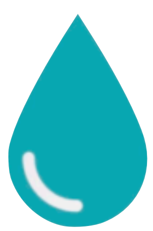
- drinking
- eating
- exercising / sweating
- urinating & defecating
- breathing

Day-to-day weight fluctuations





Think of it as **two coupled systems** varying over **different time scales**:



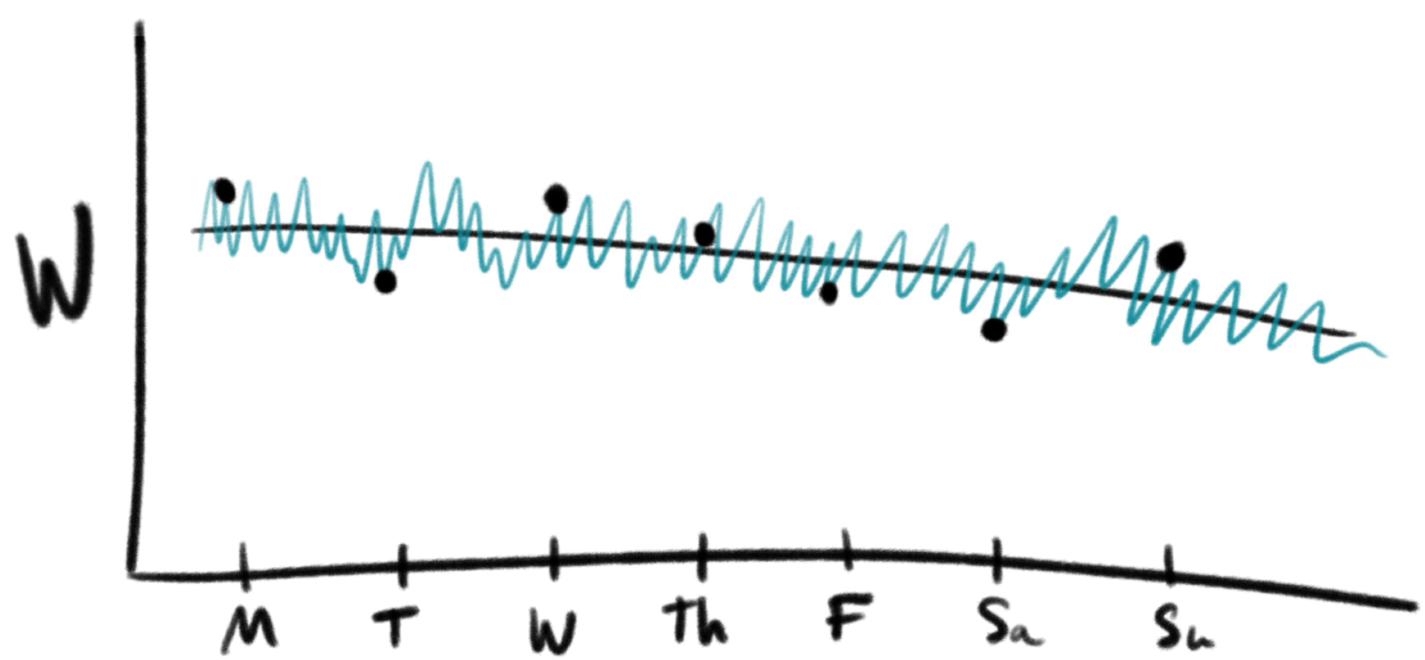
‘wet’ mass

fast / hourly

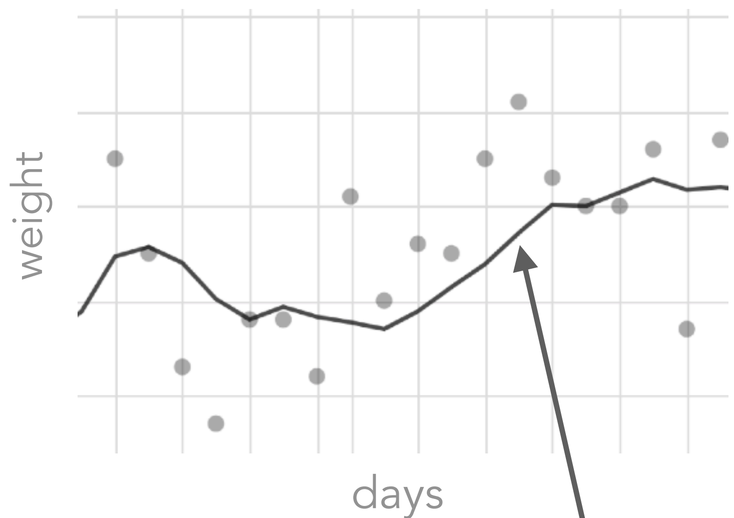
‘dry’ mass

slow / weekly

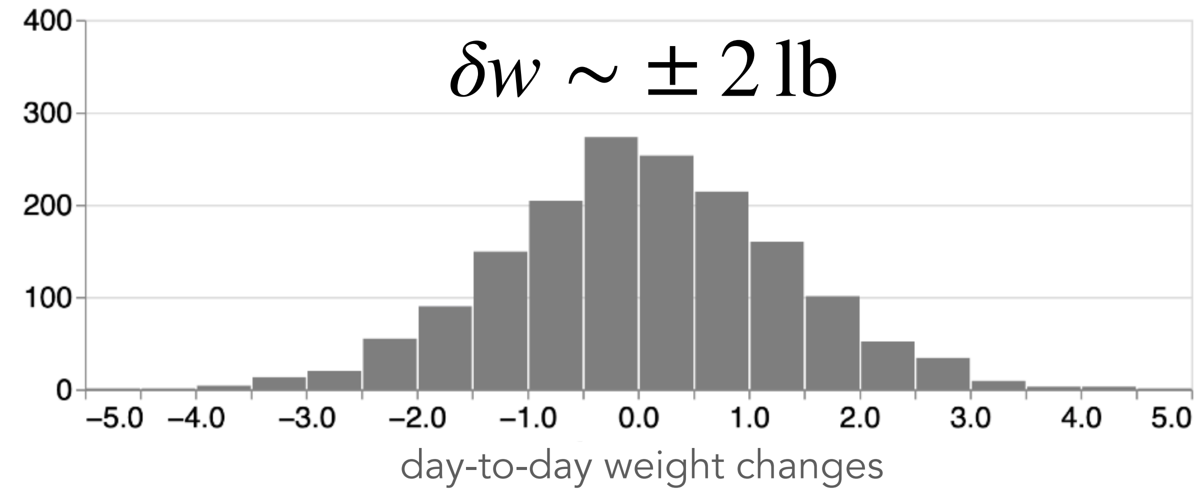
bone, muscle, fat, organs

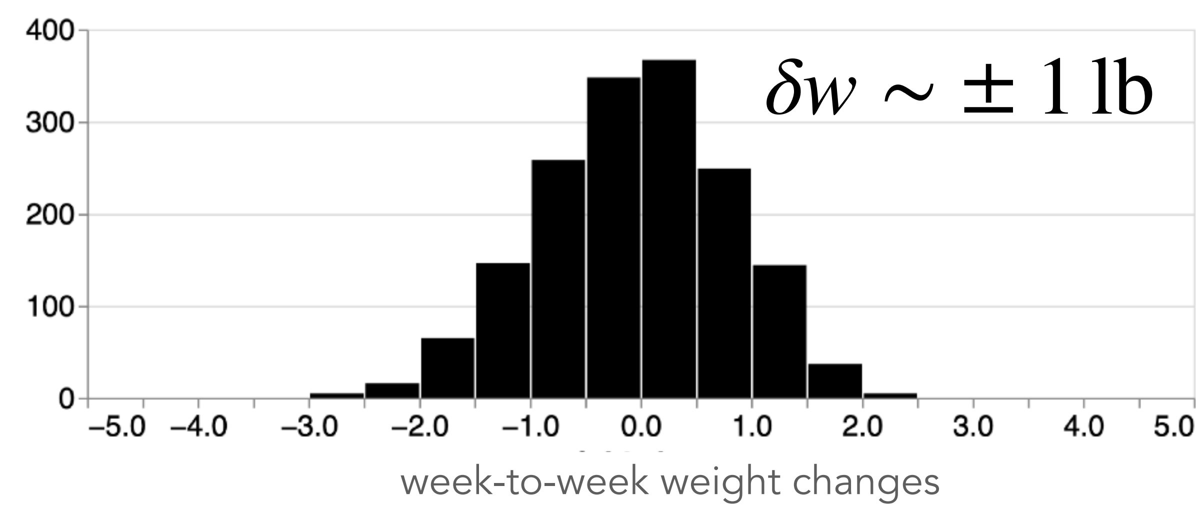


7day rolling avg, week-to-week changes



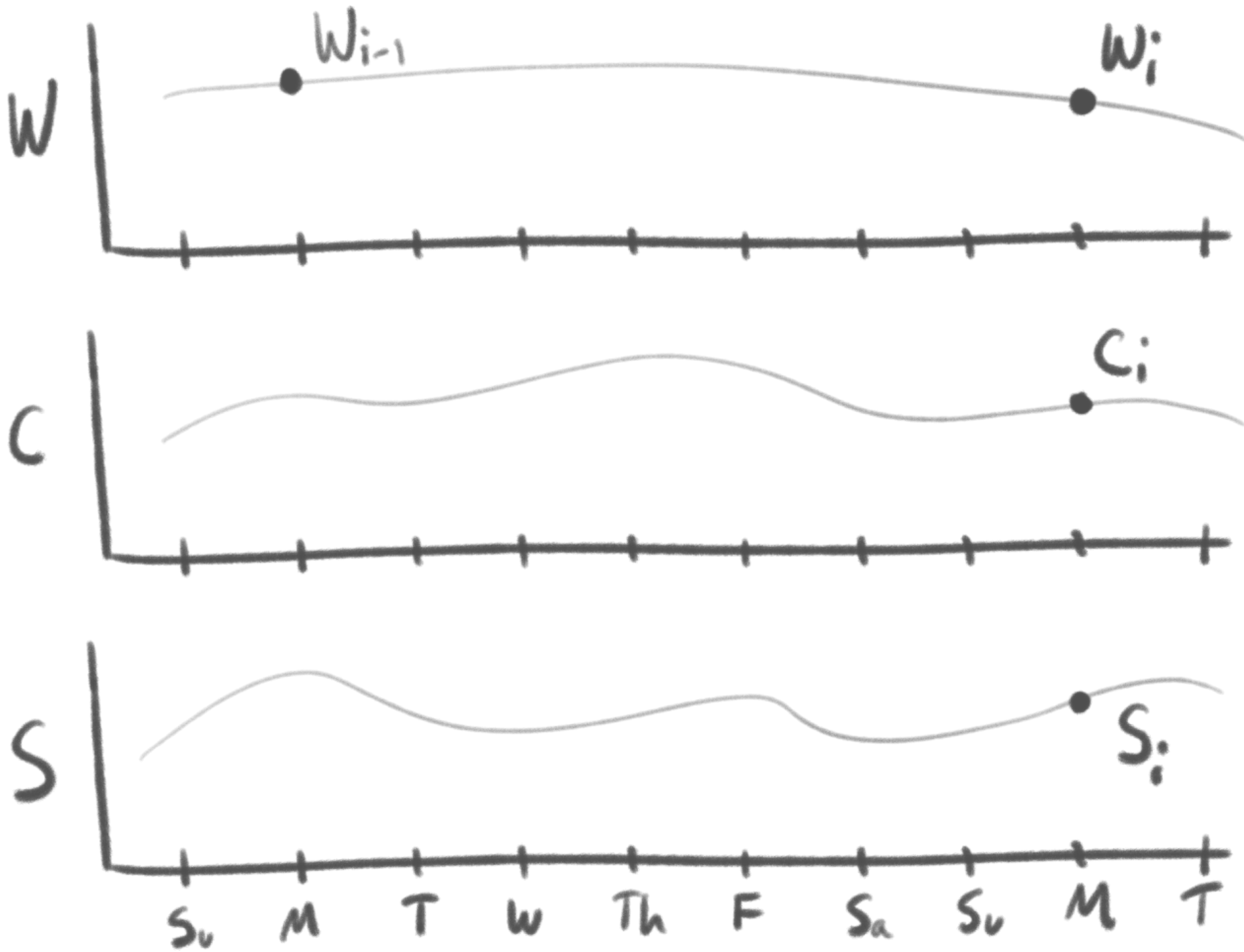
I decided to work with a **7 day rolling average** and consider <u>week-to-week</u> variations.





7day rolling avg, week-to-week changes

~first order Markov approx.



W_i 7d avg weight

W_{i-1} 7d avg weight 1 week ago

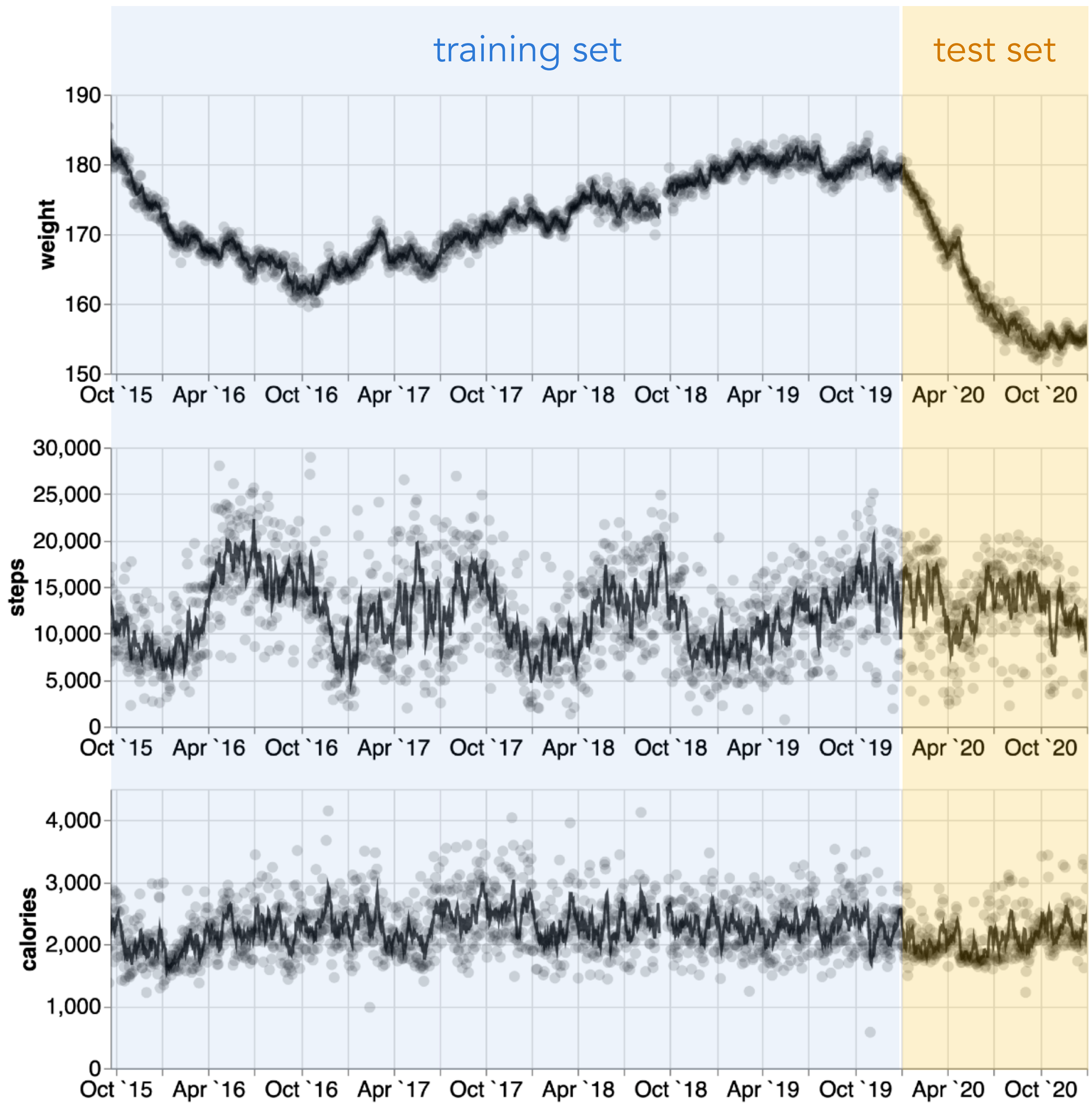
C_i 7d avg calories

S_i 7d avg steps

The Dataset

predictors			target
w_{i-1}	c_i	s_i	w_i
#	#	#	#
#	#	#	#
#	#	#	#

Train / test split



Regression fit to the data

$$W_i = a_0 + a_w W_{i-1} + a_c C_i + a_s S_i$$

Rearrange to get into the form of an energy balance equation:

$$\underbrace{W_i - W_{i-1}}_{\text{weight change in a week}} = \Delta W_i = a_c[\underbrace{C_i}_{\text{calories } \textbf{in}} - \underbrace{\alpha_s S_i}_{\text{calories } \textbf{out} \text{ due to steps}} - \underbrace{(\alpha_0 + \alpha_w W_{i-1})}_{\text{calories } \textbf{out} \text{ due to BMR}}]$$

$a_c \sim 0.002$ weekly pounds per calorie

→ **±500 calories** per day to gain/lose **1lb** per week

$\alpha_s \sim 0.024$ calories per step

→ **10K steps** burns about **240 calories**

$\alpha_0 \sim 616$ calories

$\alpha_w \sim 8$ calories per lb

→ at **160lbs**, burn about **1900 calories**

Forecasting by iterating discrete equations

initial weight

$$W_1 = a_0 + a_w W_0 + a_c C_1 + a_s S_1$$

$$W_2 = a_0 + a_w W_1 + a_c C_2 + a_s S_2$$

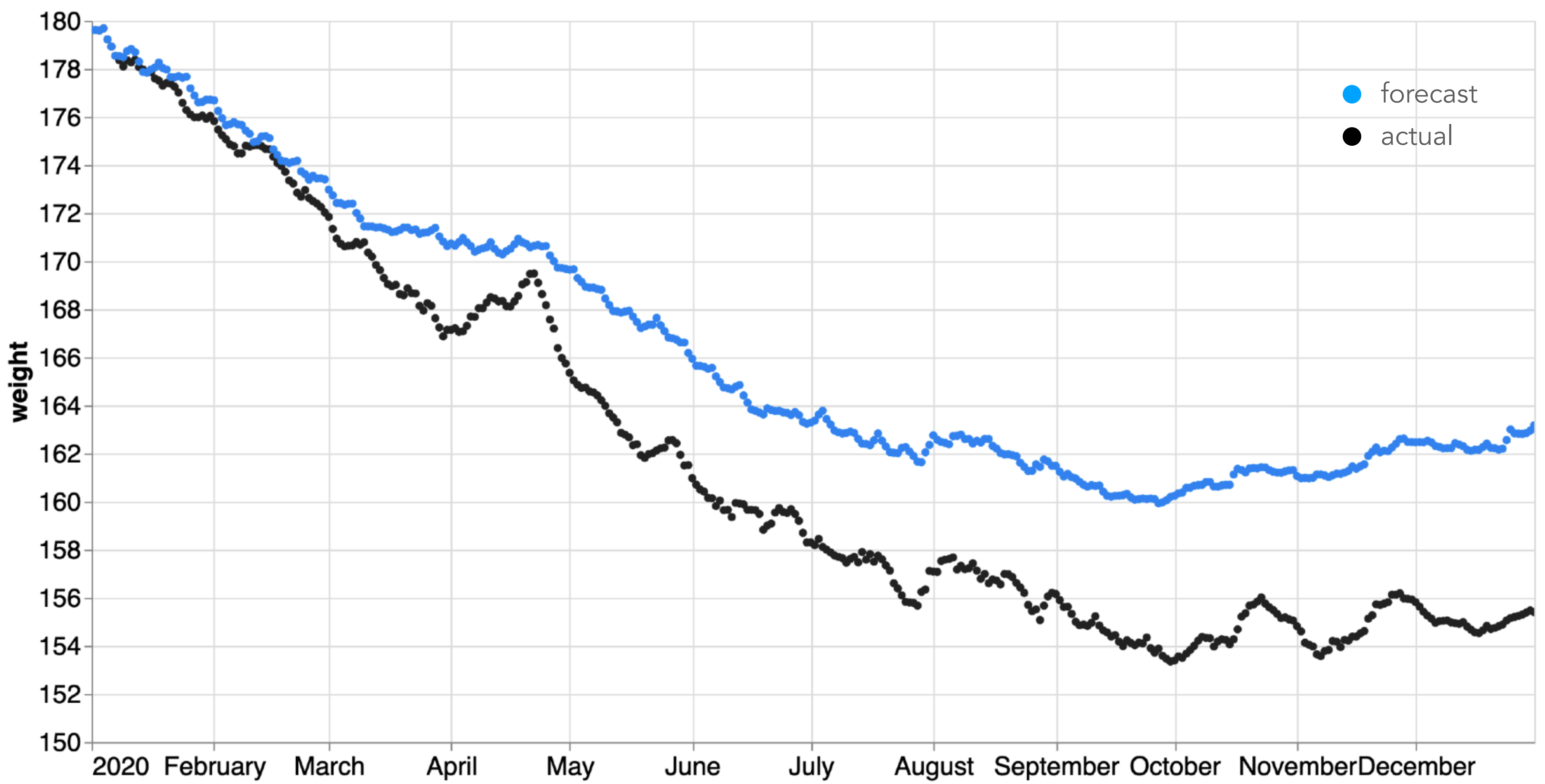
$$\vdots$$

n steps $=$ n weeks

Forecasting by iterating discrete equations

i	W_{i-1}	C_i	S_i	W_{pred}
1	$\#_0$	$\#_1$	$\#_1$	$\#_1$
2	$\#_1$	$\#_2$	$\#_2$	$\#_2$
3	$\#_2$	$\#_3$	$\#_3$	$\#_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Forecast deviates over longer periods



Model #1.1

(proposal)

Model 1.1: Bayesian prob. model, no explicit predictor uncertainty

$$\Delta W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\sigma \sim \text{Expoential}(\lambda)$$

$$\mu_i = b_c C_i - b_s S_i - (b_0 + b_w W_{i-1})$$

$$b_0 \sim \text{Normal}(\mu_0, \sigma_0)$$

$$b_w \sim \text{Normal}(\mu_w, \sigma_w)$$

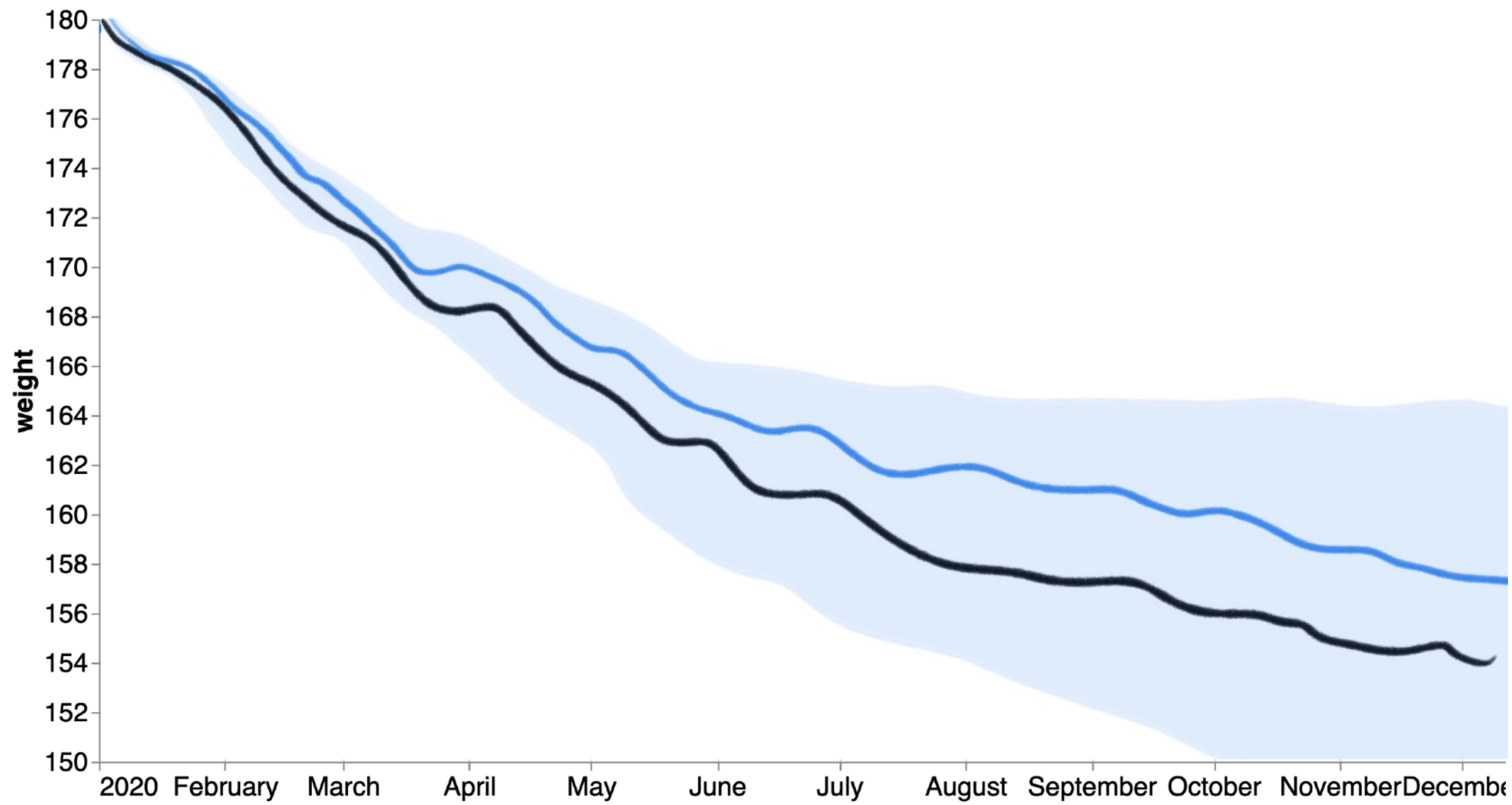
$$b_c \sim \text{Normal}(\mu_c, \sigma_c)$$

$$b_s \sim \text{Normal}(\mu_s, \sigma_s)$$

Assuming sensible choice of prior parameters are chosen based on prior predictive sim.

Forecast Uncertainty

With the Bayesian approach, we can get the forecast uncertainty via repeated trajectory simulations. I have an idea how one would generate these trajectories, but I'm unsure how to go about it in Stan.



Model #1.2

(proposal)

Model 1.2: Bayesian prob. model, treat predictor uncertainty explicitly

In this model, we try to explicitly treat the measurement uncertainty of calories and steps.

$$\Delta W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\sigma \sim \text{Expoential}(\lambda)$$

$$\mu_i = b_c C_i - b_s S_i - (b_0 + b_w W_{i-1})$$

$$b_0 \sim \text{Normal}(\mu_0, \sigma_0)$$

$$b_w \sim \text{Normal}(\mu_w, \sigma_w)$$

$$b_c \sim \text{Normal}(\mu_c, \sigma_c)$$

$$b_s \sim \text{Normal}(\mu_s, \sigma_s)$$

$$C_i \sim \text{Normal}(\mu_C, \sigma_C)$$

$$S_i \sim \text{Normal}(\mu_S, \sigma_S)$$

$$W_{i-1} \sim \text{Normal}(\mu_W, \sigma_W)$$

Model 1.2: Bayesian prob. model, treat predictor uncertainty explicitly

Treating measurement uncertainty is discussed in **Statistical Rethinking in Ch. 15.1**, and also in the Stan documentation:

https://mc-stan.org/docs/2_18/stan-users-guide/bayesian-measurement-error-model.html

It resembles the noisy scale weight model we've covered in class (and in the Pyro docs). But, I'm not sure the appropriate way to estimate measurement noise here.

```
data {
  int<lower=0> N;       // number of cases
  vector[N] x;          // predictor (covariate)
  vector[N] y;          // outcome (variate)
}
parameters {
  real alpha;           // intercept
  real beta;            // slope
  real<lower=0> sigma;  // outcome noise
}
model {
  y ~ normal(alpha + beta * x, sigma);
  alpha ~ normal(0, 10);
  beta ~ normal(0, 10);
  sigma ~ cauchy(0, 5);
}
```

→

```
data {
  ...
  real x_meas[N];     // measurement of x
  real<lower=0> tau;  // measurement noise
}
parameters {
  real x[N];          // unknown true value
  real mu_x;          // prior location
  real sigma_x;       // prior scale
  ...
}
model {
  x ~ normal(mu_x, sigma_x);  // prior
  x_meas ~ normal(x, tau);    // measurement model
  y ~ normal(alpha + beta * x, sigma);
  ...
}
```