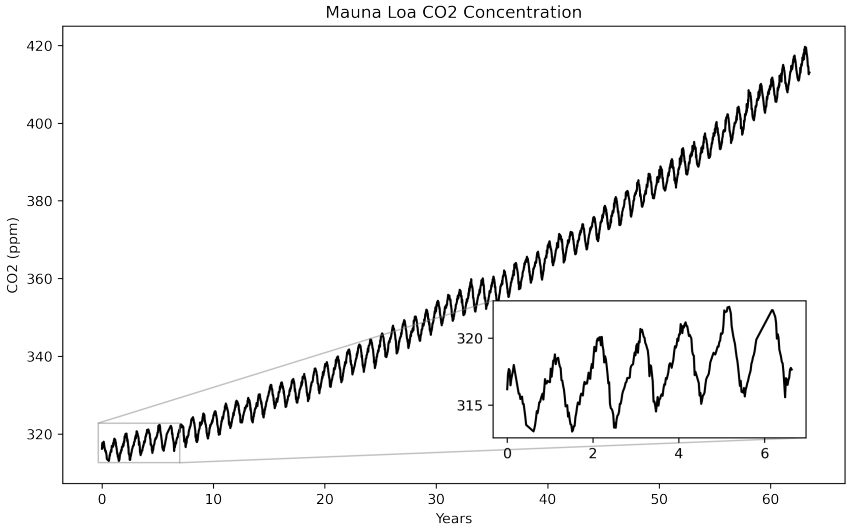
# Bayesian Models for Mauna Loa

Andre Vacha

16/12/2021

## Introduction

In this project, we will model the trusted Mauna Loa Observatory data set, a time series of weekly atmospheric C02 concentration data collected since 1958.

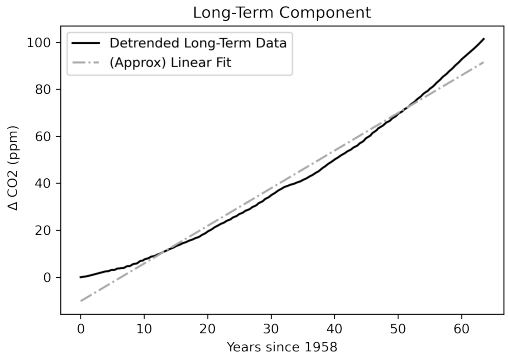



This figure, known as the Keeling curve, is far from innocuous: it suggests a rising trend in CO2 decoupled from the rhythms of seasonality, and is the quantitative basis for anthropogenic climate change. Equipped with this data alone and a novice's understanding of climatology, we will attempt to forecast 60 additional years of CO2 data.

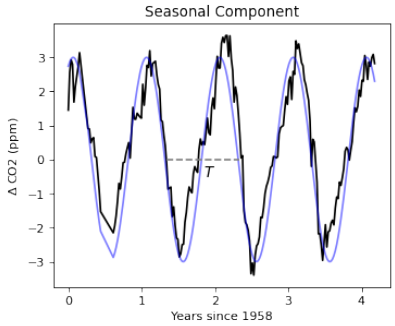
## Exploratory Analysis

To build modeling intuitions, we will first decompose the data into a series of components. To first make the time series more amenable to analysis, we perform a simple transforma-

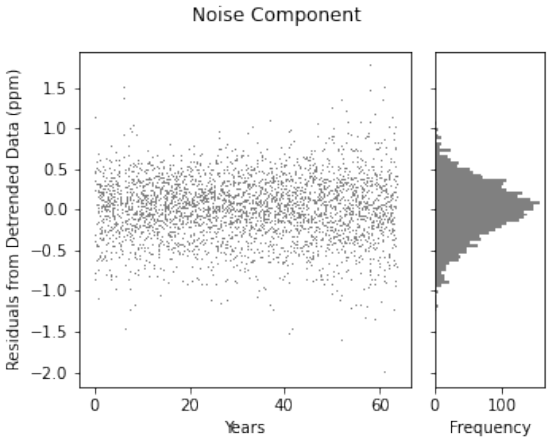
tion: the timestamps are converted to a continuous time 'Years' variable, measuring the number of years since 1958.




Detrended with a Savgol filter and anchored against a simple least-squares linear fit, the data seems to follow a non-linear growth pattern. The characteristic bow shape of exponential growth seems to be suggested, though there is a degree of medium-term variation also at play. In the interest of simplicity, we will neglect a potential medium-term component.




Netting out the long-term component, we can see a strongly periodic seasonal component. Traced with sine-wave (blue) oscillations, it appears to be sinusoidal with a period of 1 year (see annotated 'T') and amplitude of about 3 ppm. These parametric observations will become useful when we define hyperpriors in a later section. This suffices as a characteristic time scale because there is an underlying annual carbon fertilization cycle in terrestrial plants: in winter, lower insolation triggers CO2 increases, and in summer, higher insolation triggers CO2 decreases.



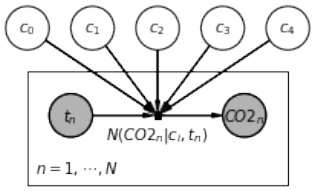

Finally, we extract a fine-grained noise component. In modeling noise, we need to check that noise is uncorrelated with other observed variables.

The noise is strikingly homoscedastic across the time series, and Gaussian-distributed along the C02 marginal with mean 0. In other words, this is a kind of idealized white noise: ocean winds creating a flux in CO2 readings, light day-to-day miscalibrations, and so on.

# 1 Modeling

We will construct two models: a linear-cosine model and an exponential-cosine model. The former can be thought of as the counterfactual or "middle child" model, while the latter can be thought of as the improved model. We will not go into Bayes Factor analysis, but the limitations of the first model will be discussed in its section.

1. Linear Cosine Model



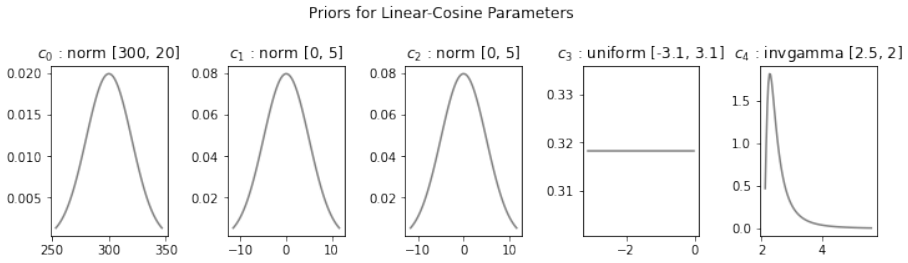

The time variable $t$ and CO2 variable are observed (shaded in grey), while the parameters are unobserved (shaded white).

$$p(\mathbf{CO2}|\theta, \mathbf{t}) = \prod_{t=1}^{n} p(CO2_t|\theta, t)$$

$$= \prod_{t=1}^{n} N(CO2_t | c_0 + c_1 t + c_2 \cos(2\pi(t + c_3)), c_4^2),$$

which is to say that the likelihood of observing a given datum of CO2 ($CO2_t$) at time index $t$ is Normally distributed, centered at a time-indexed linear-cosine function but varying with a small noise term. This is a bread-and-butter likelihood, easily encoding a practitioner's intuitions about the overall behavior of the CO2 data, while also allowing for uncertainty. As a notation shorthand, $\theta = \{c_0, c_1, c_2, c_3, c_4\}$. Finally, while we can claim a fixed $n = 3242$, the total number of observations in the training set, the next portion of this project will focus on forecasting, which will extend the time series by 20 or so years to $n* = 3242 + 52 * 20 = 4,282$ observations.

As such, the full Bayesian model has 5 parameters, each of which has a prior distribution informed by the initial exploratory analysis.




(a) Parameter $c_0$: Historical $CO_2$ Levels.

$$p(c_0) \sim N(c_0 | 300, 20)$$

This is naively the intercept term (seen earlier to be about 310 ppm), but it is more accurately described as the CO2 levels in the absence of long-term variability, ironically entailing the need for some variability. As such, we are constrained to a positive parameter, positioned around 300 ppm with some uncertainty in a Normal prior. The uncertainty ($\sigma^2$) term is set to 20 as we know that 95

(b) Parameter $c_1$: Long-Term Linear Trend.

$$p(c_1) \sim N(c_1 | 0, 5)$$

As a coefficient of the time-index, $c_1$ can be thought of as the slope: the change in CO2 each year * in the absence of seasonal and noise components*. This is also constrained to be positive, so 95

During exploratory analysis, we observed indication that the trend, while positive, was not linear. The detrended data had a pronounced bow shape, suggesting an exponential or quadratic fit. In the interest of a benchmark, simpler model, we will use a linear trend.

(c) Parameter $c_2$: Seasonal Amplitude.

$$p(c_2) \sim N(c_2|0, 5)$$

The amplitude defines the lower and upper extent of seasonal variation, which traces a sinusoidal pattern. The sinusoidal trace has two features. First, a negative amplitude is simply a reflection about the x-axis. Second, the phase space of its period (1 year) is explored within a $[-\pi\ ,\ \pi]$ phase shift (parameter $c_3$). In other words, it is self-similar under constrained parameters. This is a boon when using MCMC methods for posterior sampling (as we soon will), as it significantly reduces the number of first-order moments the sampler typically fixates on (and samples poorly as a result). Following a coincidentally identical rationale above, we use a positive-constrained Normal centered at 0, with a standard deviation of 5.

(d) Parameter $c_3$: Phase Shift.

$$p(c_3) \sim \text{Uniform}(c_3| - \pi, \pi)$$

In this and the subsequent model, we assume a fixed time period $T$ of 1 year. As a result, $\omega = \frac{2\pi}{T} = 2\pi$, leading to an overall phase space of $[-\pi\ ,\ \pi]$. We don't have any intuitions on the phase shift, so we use a flat distribution constrained to be in the range of $[-\pi, \pi]$. The author likes to call this a uniform *pi*er. (To detail why the entire phase space is explored within this constrained range, see Phase Plot in Appendix A)

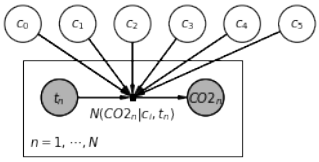
(e) Parameter $c_4$: Random Fluctuations away from the layered trend.

$$p(c_4) \sim \text{InvGamma}(c_4|2.5, 2)$$

Based on the distribution of noise marginals in the data analysis, we know that the noise is very fine-grained and less than one. An Inverse-Gamma distribution is appropriate for this parameter, with a shape parameter of 0 and a scale parameter of 0.1, as it distributions the majority of its mass within the range [0, 0.1] ppm. This is positively constrained and very sharp, as we know that noise plays a very modest role in the trend.

2. Exponential Cosine Model

To rectify a key problem with the previous model, we swap a linear-long term trend with an exponential trend:

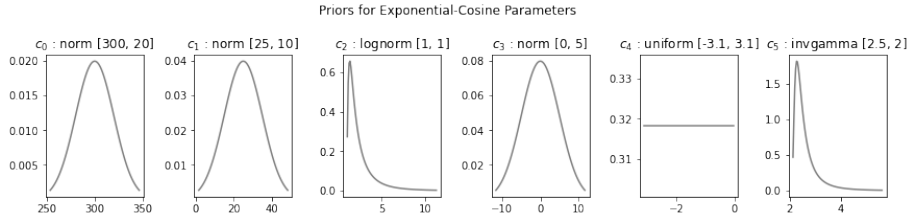



The factorization is nearly identical to the previous model, except with an additional 6th parameter and transformed time representation.

$$p(\mathbf{CO2}|\theta, \mathbf{t}) = \prod_{t=1}^{n} p(CO2_t|\theta, t)$$

$$= \prod_{t=1}^{n} N(CO2_t|c_0 + c_1 e^{c_2 t*} + c_3 \cos\left(2\pi(t + c_4)\right), c_5^2)$$

While the rationale for shared parameters are the same, the new parameters deserve introduction:




(a) Parameter $c_1$: Exponential Scale.

$$p(c_1) \sim N(c_1|25, 10)$$

This simply re-scales the exponential component. It is positively constrained, with a [5, 45] ppm probable range for the exponential scale, which is a good support.

(b) Parameter $c_2$: Exponential Shape.

$$p(c_2) \sim logN(c_2|1, 1)$$

This parameter controls the tautness of the arc of the exponential component. Larger values of $c_2$ result in a more pronounced, taut arc, while smaller values result in a smoother arc. A log-normal with a mean of 1.5 and a standard deviation of 1 is a reasonable choice, as we know that a shallow arc is obtained

from very small ($\leq 1$) values of $c_2$: the specified log-normal concentrates most of its mass in the range [0.5, 2.5], making this a good support for our prior knowledge.

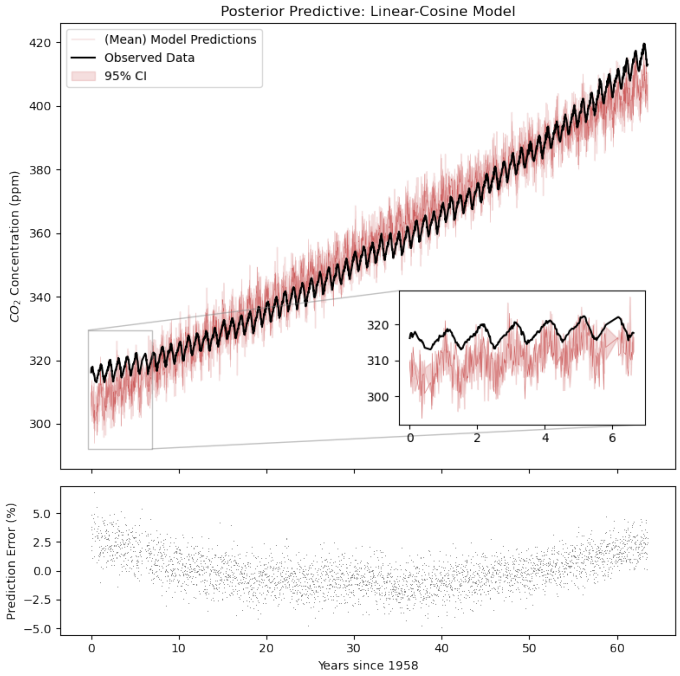
Since the growth in $t$ is exponential, we need to apply an affine transformation to it to condition the outputs: we do this by applying 't/t.max()' to yield a fractional $t*$, such that it falls between [0, 1]. In other words, $e^60 = 1e26$ (!) is a computationally intractable number, while $e^1 = 2.71$ is much more feasible to compute.

## Inference

pyStan was used to perform variational inference (see Appendix for code), with summary posterior predictive fits reproduced below.

Using 4000 posterior draws, we can perform a posterior predictive check on the observed data. In essence, we are using the mean, 2.5th and 97.5th percentiles of parameter samples to replicate the data, and in turn produce expected values and a 95 % CI uncertainty band.

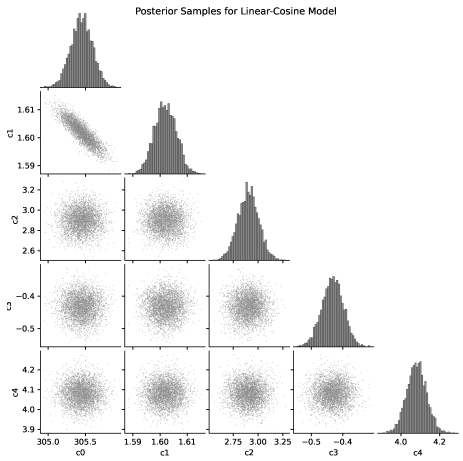
1. Linear-Cosine Model




The posterior predictive of the linear-cosine model evinces a moderately good fit: a significant portion of the data is enveloped by the confidence interval. Judged by

relative prediction error, however, the model is significantly impaired by the linear long-term trend component. A linear fit leads to an over-prediction of 5% at the start and end of the the time serious, while leads to an under-prediction of 5% around year 30.

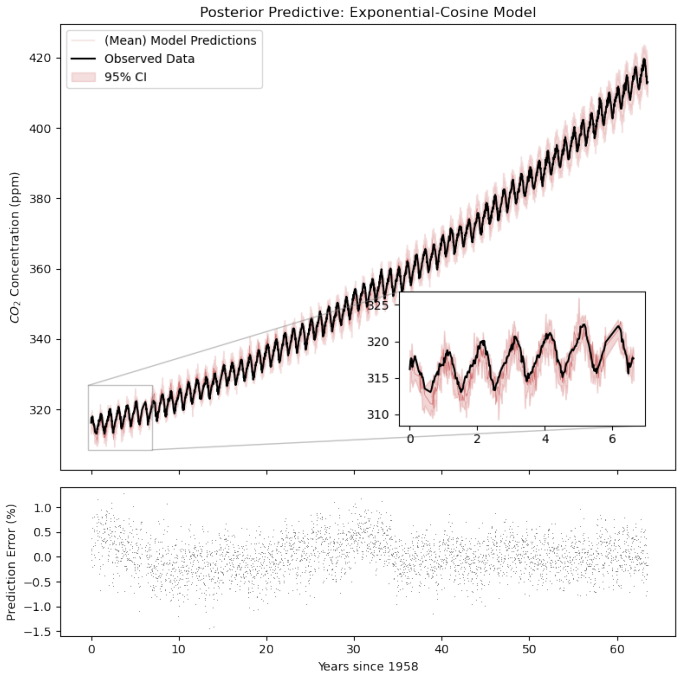
Using the magnified inset plot, we can see that the model attempts to compensate for the poor fit by increasing the scale of the noise component, leading to a highly erratic, quasi-random walk effect. The noise scale, $c_4$, has a mean posterior of about 4.1,



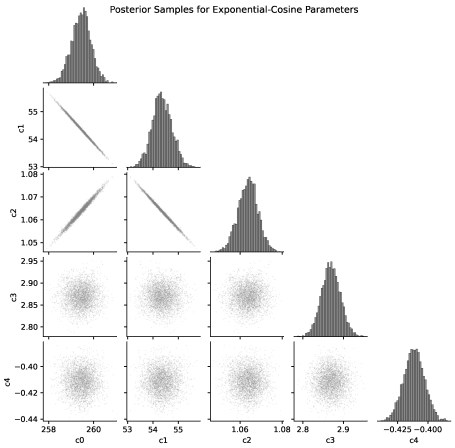

What is striking about this posterior mean is that is extremely sharp (the entire marginal posterior falls within [3.9,4.2]), and at the tail end of the specified prior (an inverse-gamma(2.5,2), seen earlier). Prediction error aside, this is a vital indication of poor fit.

The figure is also a diagnostic for inference quality. Marginal posteriors are all sharp, uni-modal and uncorrelated, except the negative correlation between $c_0$ (intercept) and $c_1$ (gradient) parameters. This is expected, because to fit the same line, we could either raise the intercept and lower the gradient, or lower the intercept and raise the gradient. A large number of effective samples [1735, 3165] and ideal Rhat ( 1) values were also observed in Stan output (see Appendix) for each parameter. In summary, the inference produced quality samples for an undesirable model.

2. Exponential-Cosine Model




This model fares far better, with the range of prediction error reducing significantly to [-1%,1%]: the long-term, seasonal and noise components are each captured accurately. Unlike before, the errors do not change systematically to evince a poor choice of component. The confidence bands, as a result, cling tighter to mean predictions and enclose the observed data for the entirety of the time series.

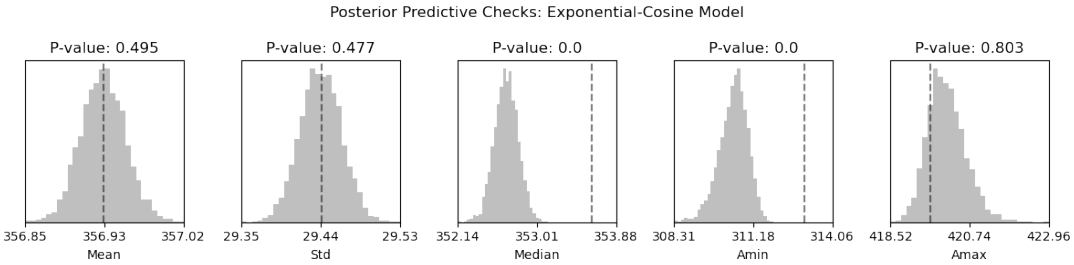



As before, we have evidence that the inference was conducted well. Posterior marginals are narrow, with the noise component being a much smaller. There is a negative corre-

lation between $c_1$ (Exponential Scale) and $c_2$ (Exponential Shape), which is expected. As we lower $c_2$, the arc bends outwards, which we can compensate for by scaling $c_1$ up, such that the same domain contains a shallower portion of the exponential.

Additionally, there is a negative correlation between $c_0$ (intercept) and $c_1$ (scale). There is a similar rationale: as we raise the intercept, we need a larger scale to claim a shallower portion of the exponential to fit the data over the same domain. Finally, a positive correlation between $c_0$(intercept) and $c_2$(bend) indicates that to compensate for a higher intercept, we need a shallower curve to fit the data. Stan reports high effective samples ($\geq$1000) and ideal Rhat ( 1) here too. In summary, we have quality sampled for a desirable model.

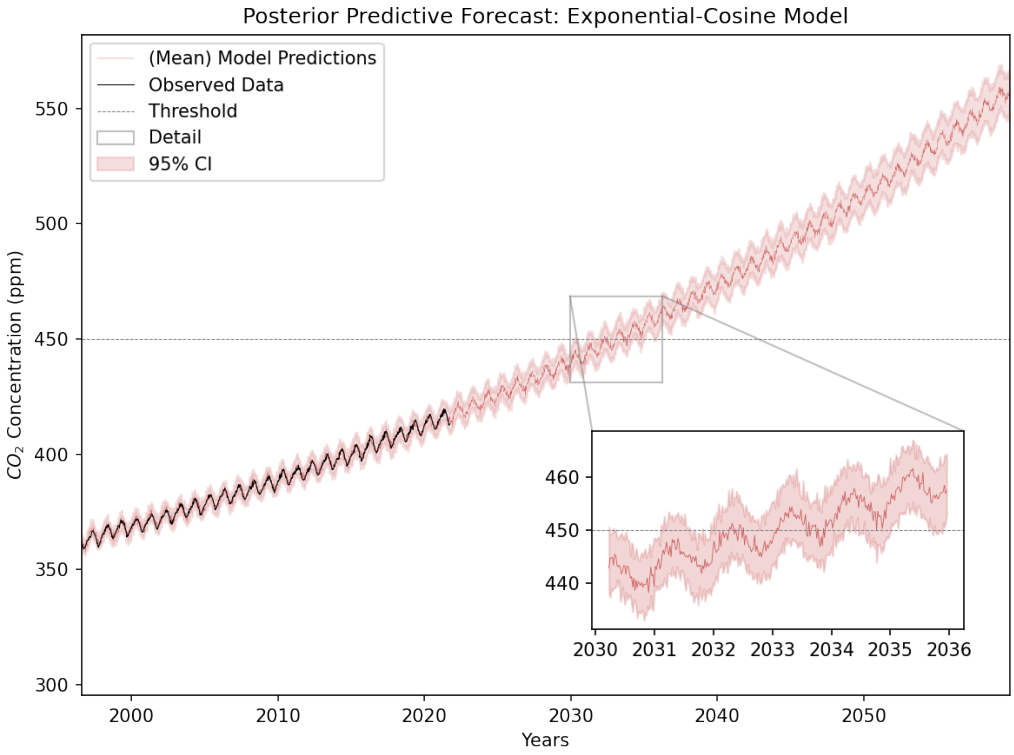
Equipped with some confidence in the updated model, we can perform more systematic checks using test statistics,




In a pattern endemic to Normal likelihoods, the improved model is good at predicting the mean and standard deviation (with near ideal p-values of 0.495 and 0.477 respectively). However, it performs poorly on predicting the median, minimum, and it performs moderately at predicting the maximum. This might be because there might be an underlying process to noise we are unaware of, and/or lack the knowledge to accurately model. For example, there might be local CO2 emission patterns that happen during specific intervals (a nearby volcanic vent emitting CO2 every month, for example). Nonetheless, these distributions enclose very narrow intervals, so the mean predictions are off by relatively insignificant amounts. A word of caution: to show that a model predicts well is not to say that it is "correct", instead, only that is not wrong.

## Predictions

We can now extend the prediction period to 2060 to include a 40-year forecast,




While the mean predictions tends upward predictably, the width of the confidence intervals increase over time. These intervals are measures of uncertainty over our predictions, and we expect to be increasingly uncertain as we move further away from the last recorded datum.

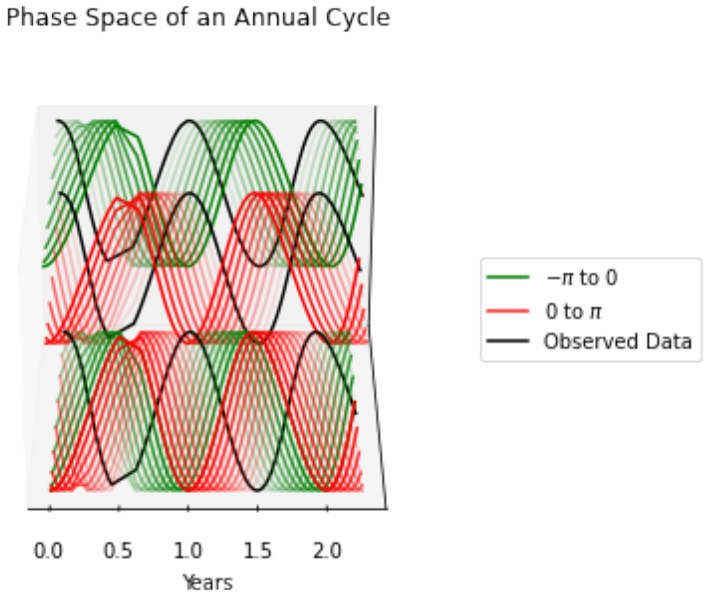
The 450 ppm threshold (dotted gray line) is recognized as a dangerous threshold for atmospheric CO2 concentration. Taking a horizontal slice of our confidence band and mean prediction, we can estimate confidence for when we will reach the 450 ppm threshold. According to the mean prediction, we will first reach this threshold on 11-04-2032. According to the lower bound of the confidence interval, we will reach this threshold on 26-03-2034, about two years later than the mean prediction predicts. According to the upper bound of the confidence interval, we will reach this threshold on 24-03-2030, about 2 years prior to the mean prediction. Interpreted another way, we can say with high confidence that we will reach this threshold between March 2030 and March 2034, with an expectation to reach this threshold in April 2032. "We should not expect to be below this threshold after March 2034" is a troubling statement for a 21 year-old author, who can barely time their midlife crisis to coincide with this end-of-times scenario.

While these results have been useful, we have to acknowledge that the model is designed according to pedestrian opinions about climatology. For example, parameters are constant across the entire time-scale. We know that the $c_2$ (exponential scale) parameter might

change if we successfully decarbonize our global economy, with the trend becoming shallow before (optimistically) reversing. To describe the reversal, we will need another model, and another set of parameters: a foreshortened exponential curve is increasingly linear. As a result, an extension of the model would include better priors from experts. For example, we could use historical, long-term ice core data to inform the intercept term, and decouple the remaining trend from the first recorded Mauna Loa observation. As of now, the author doesn't know what he doesn't know.

# Appendix

1. Phase Space Plot




2. Stan Code We map a unit vector to [-$\pi$, $\pi$] using the arctan function in transformed parameters. This limits posterior modalities to improve sampling efficiency.

```
basic = """
data {
    int<lower = 1> N;

    real<lower = 0> time[N];
    real<lower = 0> ppms[N];
}

parameters {
    real<lower = 0> c0;
    real<lower = 0> c1;

    real<lower = 0> c2;
    real<lower = 0> c4;

    unit_vector[2] c3_deg;
}
```

```
transformed parameters {
  real c3 = atan2(c3_deg[1], c3_deg[2]);
}


model {
    c0 ~ normal(280, 20);
    c1 ~ normal(0, 5);

    c2 ~ normal(0,5);
    c4 ~ inv_gamma(2.5,2);

    for (i in 1:N) {
        ppms[i] ~ normal(c0 + c1*time[i] + \
                         c2*cos(2*pi()*time[i] + c3),
                         c4);
    }
}

"""


exp = """
data {
    int<lower = 1> N;

    real<lower = 0> time[N];
    real<lower = 0> ppms[N];
    real<lower = 0> year[N];
}

parameters {
    real<lower = 0> c0;
    real<lower = 0> c1;
    real<lower = 0> c2;
    real<lower = 0> c3;
    unit_vector[2] c4_deg;
    real<lower = 0> c5;
}
```

```
transformed parameters {
  real c4 = atan2(c4_deg[1], c4_deg[2]);
}



model {
    c0 ~ normal(300, 20);
    c1 ~ normal(25,10);
    c2 ~ lognormal(1.5,1);
    c3 ~ normal(0,3);

    c4 ~ uniform(-pi(),pi());
    c5 ~ inv_gamma(2.5,2);

    for (i in 1:N) {
        ppms[i] ~ normal(c0 + c1*exp(c2*time[i]) + \
                         c3*cos(2*pi()*year[i] + c4),
                         c5);
    }
}

"""
```