LA-UR-16-21293
Approved for public release; distribution is unlimited.

Title: Statistical Modeling Efforts for Headspace Gas

Author(s): Weaver, Brian Phillip

Intended for: Report

Issued: 2016-03-17 (rev.1)

Statistical Modeling Efforts For Headspace Gas

Brian Weaver, CCS-6

March 15, 2016

The purpose of this document is to describe the statistical modeling effort for gas concentrations in WIPP storage containers. The work was performed primarily by Brian Weaver of CCS-6 (Statistical Sciences) and included input from Joanne Wendelberger (CCS-6), Bruce Robinson (ADEP), David Funk (ADEP), and Eric Heatwole (M-6).

Headspace Gas Data

Figure 1 shows the concentration (in ppm) of CO_2 in the headspace volume of standard waste box (SWB) 68685. The different colors represent the temperature that the measurement was taken where red denotes higher temperatures (in Celsius) and blue denotes lower temperatures. The data spans from May 19, 2014, to February 3, 2015. The goal of this analysis is to utilize the information within this data, along with current physics knowledge, to predict what future concentrations levels will be.

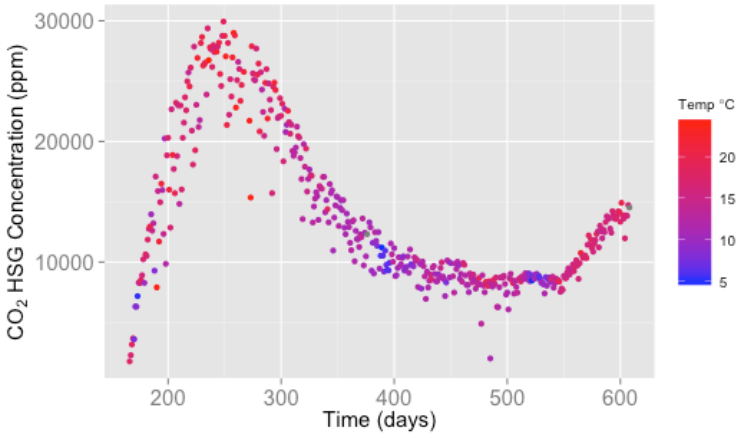

Figure 1: CO_2 gas concentration as a function of time and temperature (represented by color)

Modeling Efforts

Physical Model

Let $C(t,T)$ denote the concentration of a particular gas at time t (in days) in a headspace container at temperature T. Then the concentration changes according to the following model:

$$V_{HSG}\frac{dC}{dt} = -Q_{out}C(t,T) + Q_{in}C_{in} + M(t,T), \tag{1}$$

where

$$\begin{aligned} Q_{out} &= Q_{in} + Q_{gen}, \\ Q_{gen} &= \frac{M(t,T)R_1T}{P_{HSG}X_g}, \\ M(t,T) &= \chi(T)e^{-\beta t}, \\ \chi(T) &= Ae^{-E_a/R_2T}. \end{aligned}$$

The first term $-Q_{out}C(t,T)$ describes how the gas flows out of the SWB into the atmosphere, $Q_{in}C_{in}$ describes the flow of gas from the outside atmosphere into the SWB, and $M(t,T)$ describes how gas is generated by the substances of interest within the SWB for temperature T. $C(t,T)$ is given as the solution to the differential equation in Equation (1) and must be solved using numerical methods.

In this model, the unknown parameters, denoted by the vector θ, are $\theta = (Q_{in}, A, E_a, \beta)$ and are to be estimated using the data collected from the headspace volume. The remaining parameters are known and their values are given in Table 1.

Quantity	Value
P_{HSG}	1
R_1	0.08206
R_2	1.987×10^{-3}
X_g	0.429
C_{in}	400 (for CO_2)

Table 1: Known quantities and their values in Equation (1)

Data Model

Let $Y(t,T)$ represent the random variable associated with the measured concentration at time t for temperature T and let y be an observation of Y. Then our statistical model is

$$Y(t,T) = C(t,T) + \varepsilon \tag{2}$$

where ε represents random deviations from the physical model. Initially we assume that $\varepsilon \sim N(0, \sigma^2)$ independently. Here $\sigma > 0$ is the standard deviation of the random deviations. σ is also an unknown quantity and so it is estimated and added to our vector θ.

Bayesian Statistical Model

We use a Bayesian approach for estimating θ. The posterior distribution, $p(\theta|y_1, \ldots, y_n)$, is obtained using

$$p(\theta|y_1, \ldots, y_n) \propto L(\theta; y_1, \ldots, y_n)p(\theta)$$

where $L(\theta; y_1, \ldots, y_n)$ denotes the likelihood function and is derived using Equation (3) and $p(\theta)$ is the prior distribution for θ. The purpose of the likelihood is to describe which values of θ are most plausible (in some sense) given the observed data. $p(\theta)$ represents our current state of knowledge about θ (before observing any data) in the form of a probability distribution function. The posterior distribution is then a reweighting of $p(\theta)$ based on the information in the data through the likelihood. For this effort we assume uniform (flat) priors for our unknown parameters. Table 2 gives the upper and lower bound for these distributions for each parameter.

Quantity	Lower Bound	Upper Bound
Q_{in}	0	1
A	0	1,000,000
E_a	0	100
β	0	100
σ	0	100

Table 2: Upper and lower bounds for the uniform prior distributions assigned to the unknown parameters θ

Data Analysis

An adaptive Metropolis-Hastings algorithm was used to obtain draws from the posterior distribution for θ. Table 3 gives the posterior point estimates for θ along with the upper and lower values for their corresponding 95% credible intervals.

The posterior estimate of $C(t, T)$, along with its 95% credible interval is given in Figure 2. Notice that the physics model tends to capture the general trend of the data but is discrepant in some specific features. For example, the main peak for the data tends to occur earlier than described by our model.

Quantity	Estimate	Lower Bound	Upper Bound
Q_{in}	0.0014	0.00094	0.0095
A	378829.3	24395.5	516189.4
E_a	15.315	15.061	15.550
β	2.44×10^{-8}	2.26×10^{-8}	2.62×10^{-8}
σ	2254.2	2119.7	2426.1

Table 3: Posterior summaries for the unknown parameters θ

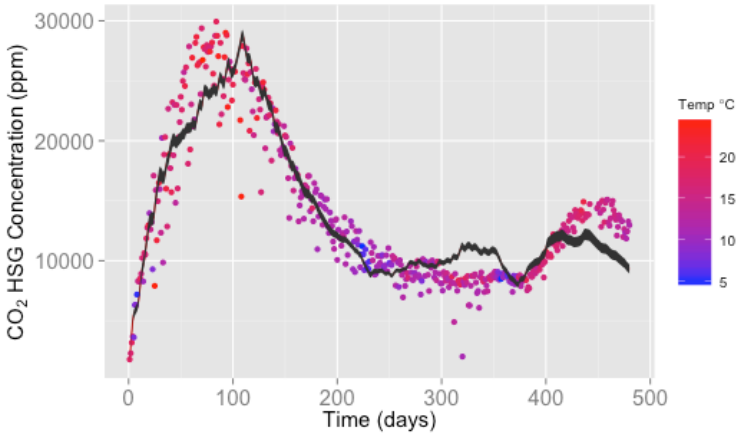

Figure 2: Posterior estimate of CO_2 gas concentration as a function of time and temperature along with its corresponding 95% credible interval (gray ribbon)

Figure 3 displays the residuals for the model fit, i.e., $Y(t,T) - \hat{C}(t,T)$ where $\hat{C}(t,T)$ is the estimate for the gas concentration as a function of both time (along the x-axis) and temperature (again indicated by color). The most striking feature is the large variability for earlier times. Additionally, it appears that the model is predicting higher gas concentrations for later times (say times larger than 350 days) than what is observed in the data.

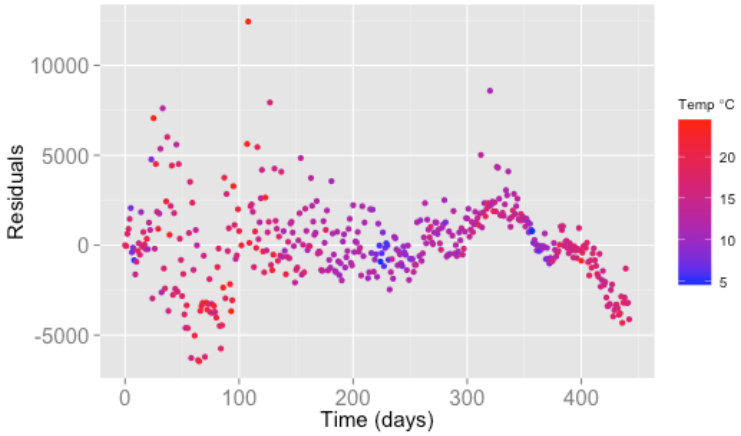

Figure 3: Model residuals as a function of time (x-axis) and temperature (color)

Lastly, Figure 4 displays concentration predictions for the last seven observations which were not used in the parameter estimation. The black points represent the posterior prediction and the vertical bars represent a 95% prediction interval. The actual observation is given as a red point. In all of these cases, the model has predicted the observation well because each of the red dots resides within the prediction interval.

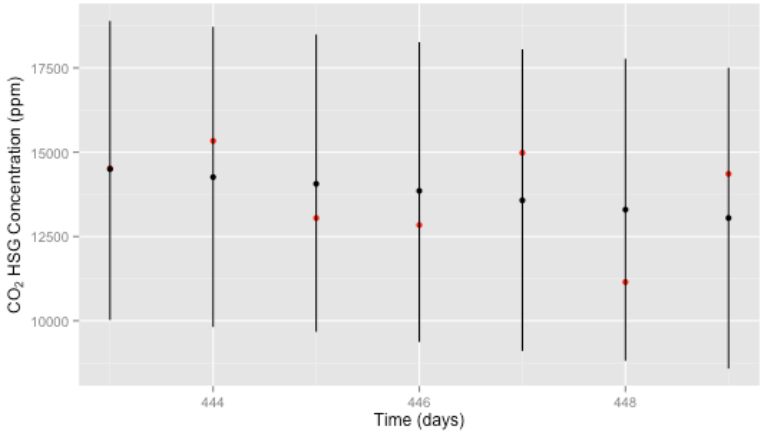

Figure 4: Posterior predicted gas concentrations (black dots) with corresponding 95% prediction intervals. The actual observations are given by red dots.

Potential Model Enhancements and Proposed Areas for Future Work

Figures 2 and 3 indicate various discrepancies associated with our full statistical model given in Equation (3). First, recall that in Figure 3 the variability in the residuals decreases as a function of time. This is a clear violation of our constant variance assumption in Equation (3). It is believed this change in variability is due to the researcher making the gas concentration measurements getting better at making the measurements with time. One potential improvement to our statistical model would be to incorporate this time dependence into the measurement error portion of the statistical model:

$$Y(t,T) = C(t,T) + f(t)\sigma\epsilon \tag{3}$$

for some appropriate function of time $f(t)$ and where $\epsilon \sim N(0,1)$.

One assumption to the physics model in Equation (1) is that gas flow is only occurring between the SWB and the surrounding atmosphere. It is observed in Figure 2 that physics model seems to be missing the peak concentration by about two weeks. This could in part be due to the additional flow of gas from the drum within the SWB and the atmosphere in the SWB. In total, gas can flow between the drum and the SWB and then between the SWB and the surrounding atmosphere. By accounting for the additional avenue of gas flow might help shift the peak concentration predicted by the model to what is observed in the data. A potential physics model could take the following form:

$$\frac{dC_2}{dt} = Ae^{-Ea/RT}e^{\beta t} - Q_{out,2}C_2 + Q_{in}C_1 \tag{4}$$

$$\frac{dC_1}{dt} = Q_{in}C_2 - Q_{out,1}C_1 + Q_{atm}C_{atm} \tag{5}$$

where C_2 and C_1 are the gas concentrations in the drum and SWB, respectively, $Ae^{-Ea/RT}e^{\beta t}$ describes the gas being added to the drum from chemical reactions, $Q_{out,2}C_2$ describes the gas leaving the drum and entering the SWB, $Q_{in}C_1$ describes the flow of gas from the SWB into the drum, $Q_{in}C_2$ describes the flow of gas from the drum to the SWB, $Q_{out,1}C_1$ describes the flow of gas from the SWB to the atmosphere, and $Q_{atm}C_{atm}$ describes the flow of gas from the atmosphere into the SWB.