Automatic Pyrolysis Mass Loss Modeling from Thermo-Gravimetric Analysis Data Using Genetic Programming

Kenneth Holladay
Southwest Research Institute
San Antonio, Texas
kholladay@swri.org

J. Marshall Sharp
Southwest Research Institute
San Antonio, Texas
marshall.sharp@swri.org

Marc Janssens
Southwest Research Institute
San Antonio, Texas

ABSTRACT
Modeling to predict flame spread and fire growth is an active area of research in Fire Safety Engineering. A significant limitation to current approaches has been the lack of thermophysical material properties necessary for the simplified pyrolysis models embedded within the models. Researchers have worked to derive physical properties such as density, specific heat capacity, and thermal conductivity from data obtained using bench-scale fire tests such as Thermo-Gravimetric Analysis (TGA). While Genetic Algorithms (GA) have been successfully used to solve for constants in empirical models, it has been shown that the resulting parameters are not valid individually as material properties, especially for complex materials such as wood. This paper describes an alternate approach using Genetic Programming (GP) to automatically derive a mass loss model directly from TGA data.

Categories and Subject Descriptors
I.2.2 [Artificial Intelligence]: Automatic Programming – Genetic Programming.

General Terms
Algorithms, Measurement, Experimentation, Languages

Keywords
Genetic Programming, Thermo-Gravimetric Analysis

1. INTRODUCTION
Passive fire protection for a building often involves subdividing the structure into compartments that are separated by fire resistant wall, floor, and ceiling assemblies. The required fire ratings for these assemblies are specified by building codes, but standardized testing is necessary to determine whether a specific assembly is in compliance. Using a large test furnace, the assembly is mounted in a frame and exposed to a standard fire that is controlled to produce a specific temperature versus time curve. This full scale fire-resistance testing is expensive, and any significant change to the design of an assembly usually requires additional testing. This approach, often referred to as prescriptive based design, has been in use for decades.

The cost-effectiveness of furnace tests could be greatly improved by supplementing the furnace data with computer simulations to predict the structural and thermal effects of design changes. Modeling flame spread and fire growth is an active area of research in Computational Fluid Dynamics (CFD). Popular software packages such as the Fire Dynamics Simulator (FDS) solve forms of the Navier-Stokes equations using finite difference methods to approximate solutions to partial derivatives of conservation equations for mass, momentum, and energy[3]. One limitation of this approach has been the lack of accurate thermophysical properties necessary for the simplified pyrolysis models embedded within the CFD software. This is especially a problem for materials with complex chemical compositions such as wood or composites where the physical properties are not constants but vary with temperature and degree of decomposition.

Recently, researchers have attempted to derive physical property functions for density, specific heat capacity, and thermal conductivity from experimental data obtained using common bench-scale fire tests such as Thermo-Gravimetric Analysis (TGA), Differential Scanning Calorimetry (DSC), and Cone Calorimetry. One of the more successful techniques has been the use of Genetic Algorithms (GA) to solve for constants in empirical models derived from the Arrhenius function[2, 4]. This is often an iterative process requiring the researcher to try multiple model forms (e.g., the number of active components) before applying GA to solve for the resulting parameter set. Even when reasonable solutions have been obtained, however, Webster[7] notes that “while the parameter estimation techniques are successful at providing calibrated models that reproduce experimental data, the values of the model parameters cannot be considered as representative of physical or chemical properties.” In addition, Stoliarov [6] and Sharp [5] have found that GA did not produce acceptable predictions for high heating rate TGA data.

This paper explores an alternate approach for developing mass loss models for direct integration into FDS. Instead of fitting parameters to a predefined pyrolysis model, both the model form and the model parameters are automatically derived directly from bench scale TGA data using Genetic Programming (GP). Section 2 describes the TGA data collection and preparation, Section 3 describes the GP results, and the final section discusses the practical implications of the work.
2. TGA DATA PREPARATION

TGA is a common laboratory technique used to characterize the thermal decomposition of a material. The procedure consists of increasing the temperature of a small sample of the material at a constant heating rate while simultaneously measuring the mass and temperature of the specimen. The data from a single heating rate is usually sufficient to determine temperature ranges for degradation reactions. When data is collected at a variety of heating rates, a surprisingly detailed analysis of the kinetics of a material can be assessed.

For this series of experiments, all TGA was performed using a TA Instruments Q600 TGA/DSC. This device has a balance sensitivity of 0.1 μg and a differential thermal analysis (DTA) sensitivity of 0.001 °C. The tests were conducted using nitrogen purge gas at a flow rate of 100mL/min. The samples were contained in alumina (Al₂O₃) crucibles with modified lids to incorporate a 50-micron hole to allow for release of free moisture. Samples of Douglas fir and type “X” gypsum wallboard were heated from room temperature to 1,000 °C using heating rates of 5, 10, 20, and 60 °C per minute. An additional data set was collected for Douglas fir at a rate of 40 °C per minute.

The purpose of the data preparation phase was to transform the raw TGA data into a form suitable for use with GP. The raw TGA data was recorded in Excel format and consisted of three columns of measurements: time (minutes), temperature (°C), and mass (milligrams). Table 1 summarizes a few important characteristics of the raw TGA data for Douglas fir. Since the initial mass differed for every experiment, all calculations were performed on the mass fraction, which is calculated by dividing each mass value in a data set by the initial mass. Any material that remains on completion of a TGA test is referred to as the residual mass.

Table 1 – Douglas fir raw TGA data summary.

<table>
<thead>
<tr>
<th>Rate (°C/min)</th>
<th>Initial (mg)</th>
<th>Final (mg)</th>
<th>Duration (minutes)</th>
<th># Data Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>12.278</td>
<td>2.374</td>
<td>194.728</td>
<td>18999</td>
</tr>
<tr>
<td>10</td>
<td>8.754</td>
<td>1.428</td>
<td>97.578</td>
<td>11711</td>
</tr>
<tr>
<td>20</td>
<td>16.400</td>
<td>2.970</td>
<td>48.742</td>
<td>5851</td>
</tr>
<tr>
<td>40</td>
<td>12.784</td>
<td>2.491</td>
<td>24.428</td>
<td>2933</td>
</tr>
<tr>
<td>60</td>
<td>11.869</td>
<td>1.953</td>
<td>16.254</td>
<td>1952</td>
</tr>
</tbody>
</table>

Examining the raw data revealed that, as with all laboratory experiments, there were areas of concern that could adversely affect subsequent analysis. Concerns included the sample rate, the heating rate, measurement noise, data discontinuities, and the disparity in the number of points collected for each heating rate.

TGA data is normally collected at a constant sample rate. Calculating the differential time values in the raw data showed that the sample intervals ranged from 0.008323 to 0.008340 minutes, or between 119 to 120 samples per minute. Since this amounts to only 0.2% error, the sample interval was considered to be constant for filtering purposes.

While the intent of TGA is to expose the subject material to a constant heating rate over the entire duration of the test, some initial ramp up time is required. Figure 1 shows a plot of the recorded temperature versus time for the first few minutes of each TGA test for Douglas fir. Between 0.3 and 1.1 minutes were required to achieve the desired heating rate. During that period, the heating rate (dT/dt) was not constant. Since the instantaneous heating rate was expected to be an important GP terminal value, the actual heating rate during each time interval had to be calculated for inclusion in the cleaned data set.

![Figure 1 – Startup temperature profile for the Douglas fir TGA data.](image1)

The raw data also exhibited measurement noise for both the mass and temperature measurements. Since the TGA process should result in a continuously increasing temperature and a continuously decreasing mass, any point where the mass increases and/or the temperature decreases with respect to the previous measurement essentially represents a noise artifact. As illustrated in Figure 2, mass measurement noise was especially evident during the startup of each TGA test and for the entire duration of the test at the lowest heating rate. This resulted in a very poor estimation of the instantaneous mass loss rate (derivative of the mass fraction with respect to time), which is one of the values that evolved models need to predict.

![Figure 2 – Mass measurement noise during startup for the gypsum TGA data.](image2)
The final concern with the raw TGA data was the disparity in the number of data points collected at the low (5 °C/min) versus the high (60 °C/min) heating rates, amounting to a 10-fold difference. Since common GP fitness functions calculate an average error over the total number of points presented, the raw data distribution would create an unacceptable bias towards the low heating rate.

To mitigate the problems outlined above, each data set was first filtered to reduce the noise and to smooth the first derivative. After trying several filters, including moving average and Kaiser, qualitative observation of the resulting time plots indicated that the Kaiser filter was more effective. After filtering, the data was resampled to have the same number of points for each heating rate. This resulted in a different sample rate for each data set, but the sample rate (dt) was included as a terminal in the GP runs. Also note that since most reaction kinetic equations use an absolute temperature reference, the temperature values used for GP were converted from Celsius to Kelvin. The final cleaning process was implemented in MATLAB and performed the following steps on the raw data collected for each TGA test (a single material at a single heating rate):

- Calculate the raw mass fraction (m)
- Calculate the raw mass loss rate (dm/dt)
- Convert the raw temperature (T) from °C to °K
- Calculate the temperature derivative (dT/dt)
- Filter the raw m, dm/dt, T, and dT/dt
- Align the filtered data to a common starting temperature
- Create a 1000 point equal interval time vector covering the test duration (the new dt)
- Resample the filtered data using the new time vector
- Realign the time to start at zero

The results were stored in tab delimited files (one file per material per heating rate) with the column labels (or field names) as shown in Table 2. Four fields (TRate, dt, T, and m) were used as GP terminals, while three fields (m nxt, dm, and dmdt) were candidate model output values for fitness calculation.

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Unique (per material) identifier for each data point</td>
</tr>
<tr>
<td>t</td>
<td>Elapsed time (minutes)</td>
</tr>
<tr>
<td>TRate</td>
<td>Heating rate (°C per minute)</td>
</tr>
<tr>
<td>dt</td>
<td>Time interval between data points (minutes)</td>
</tr>
<tr>
<td>T</td>
<td>Temperature (°K)</td>
</tr>
<tr>
<td>m</td>
<td>Mass fraction at time t</td>
</tr>
<tr>
<td>mnxt</td>
<td>Mass fraction at time t+dt</td>
</tr>
<tr>
<td>dm</td>
<td>Change in mass fraction (mnxt – m)</td>
</tr>
<tr>
<td>dmdt</td>
<td>Rate of change of the mass fraction (mnxt-m)/dt</td>
</tr>
</tbody>
</table>

Figure 3 shows an example plot of the raw data (dotted line) and the cleaned data (solid line) for Douglas fir at the 40 °C/min heating rate. The mass fraction is a very close fit, with the solid line obscuring the underlying dotted line, and the clean first derivative is significantly smoother than the raw. Figure 4 shows the same information for Type X Gypsum heated at 60 °C/min. A comparison of the upper graph in each figure (mass fraction vs. temperature) highlights the significant difference between the total mass lost by each material. The Gypsum sample retained almost 80% of its initial mass, while the Douglas fir sample was reduced to less than 20% of its initial mass.
These figures also reveal the significance of the first derivative. For each material, there are at least three major pyrolysis reactions occurring over the test period, but they are difficult to isolate from the upper mass fraction plot. On the mass loss rate (dm/dt) plot, however, local peaks are easily distinguished, as indicated by the three arrows. For Douglas fir, the first peak represents moisture loss, while the other two peaks probably represent lignin and cellulose reactions. For Gypsum, the early peak is also related to moisture loss, while the others constitute a less aggressive chemical degradation. Note that the shapes and locations of the peaks are significantly different for each material.

3. GP MODELING

The goal of the Genetic Programming portion of this project was to derive a mass loss prediction model for each material from the cleaned TGA data. Since a future goal is to incorporate this model into the FDS software, the model must be able to accurately predict both mass loss and mass loss rate. For most numerical analysis techniques that use some form of time step (Δt), these two features are effectively related as shown in equation 1.

\[
\left(\frac{dm}{dt}\right)_i \approx \frac{m_{i+1} - m_i}{t_{i+1} - t_i} = \frac{\Delta m_i}{\Delta t_i}
\]

This leads to three equivalent forms for the model based on whether the prediction is the next mass value, the mass loss, or the mass loss rate. We selected the first and third forms for the GP experiments. The first form (equation 2), models the curve in the top graph of Figure 3, while the third form (equation 3) models the bottom graph in Figure 3.

\[
m_{i+1} = f(T_i, m_i, \Delta t_i, \left(\frac{dT}{dt}\right)_i)
\]

\[
\left(\frac{dm}{dt}\right)_i = f(T_i, m_i, \Delta t_i, \left(\frac{dT}{dt}\right)_i)
\]

Following the usual GP preparatory steps, Table 2 outlines the terminal set, function set, fitness function, and various run parameters used for the GP experiments. The table reflects an aggregation of numerous runs that employed variations within the ranges and elements specified. For each material, one of the data sets was omitted from the training data used during evolution and then reintroduced as part of the test and evaluation of any resulting models. For Douglas fir, the 20 °C/min set was omitted, while for gypsum, the 10 °C/min set was omitted.

The first set of GP experiments used a typical GP symbolic regression technique to model the mass loss. The data points in the training set were systematically run through the population, and the fitness was calculated based on the squared difference between the model output and the expected output. Results with good fitness were often produced within a few generations. Figure 5 shows a typical plot where the model predictions are plotted with solid lines that fit closely enough to obscure the underlying data values plotted with dotted lines. Note that the model also fits the 20 °C/min rate that was not included in the training set.

Table 3 – GP experiment parameters

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminals</td>
<td>m, dt, T, TRate</td>
</tr>
<tr>
<td>Constants</td>
<td>-1, 1, 8.314, Ephemeral Random (ERC)</td>
</tr>
<tr>
<td>Functions</td>
<td>+ - * / ^ EXP SIN COS</td>
</tr>
<tr>
<td>Fitness Function</td>
<td>Error Squared</td>
</tr>
<tr>
<td>Fitness Target</td>
<td>0.00005 – 0.000005</td>
</tr>
<tr>
<td>Max Length</td>
<td>500 to 900 tokens</td>
</tr>
<tr>
<td>Population</td>
<td>50 to 200 per CPU</td>
</tr>
<tr>
<td>Ranking</td>
<td>Exponential (0.8 bias)</td>
</tr>
<tr>
<td>Evolution</td>
<td>Mutation (0.15 – 0.58), Point Mutation (0.00 – 0.15), 2 Point Crossover (0.20 - 0.35), 1 Point Crossover (0.20 - 0.30), Reduction (0.02 – 0.05)</td>
</tr>
<tr>
<td># CPUs</td>
<td>64 to 256</td>
</tr>
</tbody>
</table>

Figure 5 – Example 1, Douglas fir model predicting m.

Figure 6 – Example 1, Douglas fir model calculating dm/dt.
However, further analysis showed two major problems. The mass loss rate calculation shown in Figure 6 contained significant jitter, which would adversely affect performance if used within the FDS software. Of greater importance, the numerical analysis techniques that comprise most fire prediction software packages are iterative over time, using the mass loss rate \( \frac{dm}{dt} \) output of one time interval to calculate the mass input to the next time interval as follows:

\[
m_{i+1} \approx m_i + \left( \frac{dm}{dt} \right)_i \Delta t
\]  

(4)

Figure 7 shows the same example model as Figure 5 evaluated with iterative time steps. The model output (solid lines) obviously does not follow the TGA data (dotted lines), indicating that the performance of this model would be unacceptable.

Postulating that a smoother first derivative would ameliorate both of these problems, the next set of GP experiments were formulated to directly predict the mass loss rate (equation 3). This proved to be more challenging for GP. Using a maximum of 300 generations, only 1 in 5 runs converged to the desired fitness, with most runs effectively stalling after about 200 generations.

Figure 8 shows one of the better models produced for Douglas fir. As expected, the model exhibits a smoother first derivative than in Figure 6. The dotted lines represent \( \frac{dm}{dt} \) calculated from the TGA data, and the solid lines of the same color represent the model output. Although the lines are not superimposed, they do capture the reaction peaks reasonably well, including the 20 °C/min curve that was not included in the training data set.

Figure 9 shows the subsequent back calculation of the mass loss curves, which once again fit the experimental data very well. Figure 10 shows the results of iteratively calculating the mass. Compared to Figure 7, the general shapes of the curves are significantly better, but the residual mass for each heating rate is higher than expected.

The gypsum data proved to be more difficult to fit than the Douglas fir. The models produced by predicting the mass loss directly had very rough first derivatives, and the models produced by predicting \( \frac{dm}{dt} \) were generally less fit that those for Douglas fir.
4. DISCUSSION

While the results achieved to date are encouraging, the models produced so far are not significantly better than those developed using GA to fit parameters to simplified pyrolysis models. The GP models fit the high heating rate data better, but they tend to skew low temperature reactions at slow heating rates. For the next group of GP experiments, we will be employing two new features designed to improve the resulting models.

The motivation for the first new feature comes from domain experience with the simplified pyrolysis models, which typically employ a conversion parameter that is active over a fixed temperature range for each significant reaction component. This allows separate Arrhenius equations to be summed together as shown in equation (5).

$$\frac{\hat{c}m}{\partial t} = \sum_{i=1}^{k} Z_i \left(1 - \alpha_i(T)\right)^{n_i} \cdot e^{-\frac{E_{a,i}}{R T}}$$  

$$\alpha_i(T) = \frac{m_{i,0} - y_i(T)}{m_{i,0} - m_{i,f}}$$  

(5)

Where \(k\) is number of reactions, and for each component \(i\): \(Z_i\) is the pre-exponential factor, \(n_i\) is the reaction order, \(E_{a,i}\) is the activation energy, \(y_i(T)\) is the mass fraction remaining at temperature \(T\), and \(\alpha_i(T)\) is the conversion at temperature \(T\).

Note that the conversion has a value of 0 before the reaction begins and continuously increases until the reaction is complete, at which point the value is 1. Typically, conversion parameters are manually identified by visual analysis of the TGA mass loss rate, while the pre-exponential factor, the reaction order, and the activation energy are automatically solved using GA. In practice, however, the TGA data does not provide the necessary distinction in the mass fractions of the reaction components, requiring simplifying assumptions that do not necessarily hold for complex materials.

Based on the observation that reactions are active over a limited subset of the test duration, we added two switches as GP functions. Both require three inputs: a value and two transition points. Figure 11 illustrates the pulse switch behavior, and Figure 12 illustrates the ramped switch behavior.

Preliminary results indicate that adding these to the function set may enable the evolution of range specific model behavior. One of the first runs with the Gypsum data that included the new switches produced the best model to date. Figure 13 shows the predicted mass loss rate where the fit to the low temperature reactions is adequate, but the high temperature reaction is still
missing. Figure 14 and Figure 15 show the calculated $m$ and the iteratively calculated $m$.

![Graph showing calculated $m$ and iteratively calculated $m$.](image)

Figure 15 – Example 3, Gypsum model iteratively predicting $\frac{dm}{dt}$, and then calculating $m$.

The second new feature consists of a new configuration option that has been incorporated into the GP software to mimic iterative application of the model. Understanding this option requires a brief explanation of the GPE5 XML configuration file. Normally, the file contains a tag that indicates which column in the data set (Table 2) is to be used in conjunction with the model output to calculate an error. Thus to predict the next mass loss rate we use:

\[
\text{<ResultField>dm/dt</ResultField>}
\]

Each row of the data set is presented to the model, and the resulting output is used by the fitness function to calculate the error between the model output and the expected output specified in `<ResultField>`.

Invoking the new option requires two additional tags. The first indicates which column in the data set is to be replaced with the model output. For the TGA problem, the output from the model will be used to calculate the next mass fraction input:

\[
\text{<DiscreteStepInputField>m</DiscreteStepInputField>}
\]

For each new data set, the initial value of all terminals is set from the first row of the data set. For subsequent rows in a data set, the `<PostCode>` tag contains the FIFTH code that will be executed to transform the model output into the specified next input. In this example, the calculation requires multiplying the mass loss rate by the time interval and then adding the current mass fraction:

\[
\text{<PostCode>DUP dt * m + SWAP</PostCode>}
\]

When this code is executed, the stack initially contains the output of the model for row $i$. After execution of the code, the stack must contain two items: the original model output on top and the replacement terminal value for row $i+1$ beneath.

5. SUMMARY

Although the development of black box pyrolysis models directly from TGA data seems well suited to GP symbolic regression, extensive searching through the archives of relevant journals and conference publications has found no previously published work. We believe that the characteristics of the problem (sufficient laboratory data; complex relationships between time, temperature, heating rate, and mass fraction; as well as poor performance of traditional modeling tools) indicate suitability for GP analysis. While results obtained to date have not produced models that are accurate enough to incorporate directly into fire prediction software, they have progressed sufficiently to warrant additional research.

6. ACKNOWLEDGEMENTS

This work was sponsored by the Southwest Research Institute Advisory Committee for Research (ACR).

The Genetic Programming Environment for FIFTH (GPE5) [1] was used to perform all GP experiments on this project. High Performance Computing resources were provided by the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (http://www.tacc.utexas.edu).

The complete source code for GPE5 and the raw data sets for Douglas fir and gypsum are available at fifth.swri.org.

7. REFERENCES


