A novel soot concentration field estimator applied to sooting ethylene/air laminar flames

L. Zimmer, S. Kostic and S. B. Dworkin
Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, Ontario, Canada

ABSTRACT
Soot formation modeling, when incorporated into computational fluid dynamics of industrial devices, can be numerically prohibitive. Nonetheless, there remains a significant push to predict soot formation, so as to aid in environmentally sustainable design. The present work features a redesign of an inexpensive soot estimator, that has been developed and applied to laminar flames with great success. It is much more accurate and efficient than previous versions. The soot estimator consists of a library generated from validated sooting flame models, in which the Lagrangian histories of soot-containing fluid parcels are stored. The library is used in post-processing to estimate soot concentrations. For the first time, the estimator framework is used to predict the entire soot field. Also, important parameters to the estimator technique are analyzed. This work is conducted for nine different sooting ethylene/air coflow diffusion flames. The framework successfully predicts the entire soot field. When the data from many flames were combined into one library based on mixture fraction, temperature, and H₂ histories, it could predict all flames with high accuracy. Finally, two scenarios were considered to assess the framework with an independent set of data, and the predictor presented very good accuracy in capturing soot formation.

1. Introduction
The knowledge of soot formation and oxidation has been advancing in the past decades. So too has the complexity of leading-edge models. Additionally, new emission regulations have also been pushing industry to account for particulate matter (soot) reduction in their applications (International Civil Aviation Organization, 2006; International Energy Agency, 2017; Williams & Minjares, 2016). Therefore, searching for and developing techniques to reduce soot formation and emissions has become an important concern for researchers and industry. The coupling of detailed soot modeling with the already complex fluid dynamic and combustion phenomena, which industrial applications require, can become numerically prohibitive or unattractive.

In order to reduce the computational burden of ascertaining combustion and soot characteristics, the complexity of the chemistry and therefore the time involved to predict the state of the system can be reduced. Such reduction may be obtained by employing steady state and partial equilibrium assumptions. Also, automatic reduction techniques are in common use, such as the Intrinsic Low Dimensional Manifold (ILDM) (Maas & Pope, 1992), the Steady Laminar Flamelet Model (SLFM) (Peters, 1984) and the Flamelet Generated Manifold (FGM) (van Oijen & de Goey, 2000). Even so, using these reduced chemistry techniques does not guarantee the exact prediction of Polycyclic Aromatic Hydrocarbons (PAHs), which are gas species that are essential to modeling nucleation and condensation processes of soot particles, making the utilization of expensive detailed soot models impractical. Another approach is to reduce the complexity of the soot model, using semi-empirical models (Fairweather, Jones, & Lindstedt, 1992; Leung, Lindstedt, & Jones, 1991), but then the predictive capability of the model for a range of conditions as well as the accuracy of soot predictions may be affected.

To address this problem, a novel solution has recently been proposed by Alexander, Bozorgzadeh, Khosousi, and Dworkin (2018), which leverages the knowledge from accurate soot modeling with detailed chemistry to estimate soot concentrations. This post-processing tool uses a library, built from Lagrangian hysteresis of key parameters that correlate soot formation from inception up to oxidation, to predict soot concentrations based on existing numerical Computational Fluid Dynamics (CFD) data. A series of numerical solutions of well validated sooting laminar diffusion flames were used as a...
training and validation dataset. This post-processing tool was demonstrated to be computationally inexpensive and was applied to laminar flames with some success.

It is important to state that this technique is not a model for soot formation and oxidation, but rather a framework which uses a library to estimate soot properties using correlations and interpolation (Alexander et al., 2018).

In Alexander et al. (2018), the assessment of the soot estimator was carried out analyzing the soot volume fraction $f_v$ along two representative pathlines in coflow laminar diffusion flames, one at the centerline and the other passing through the point of maximum soot volume fraction, leaving the question of whether or not this technique could have broader applicability or if it can be applied to estimate the entire field of $f_v$. Therefore, in the present study the estimator framework is extended to predict the entire field of $f_v$ in nine atmospheric laminar diffusion flames, and the accuracy of the prediction is analyzed and assessed. A new procedure is added to reconstruct the soot formation field in Eulerian coordinates from the pathlines, which are in Lagrangian coordinates. Also, important parameters to the estimator technique are analyzed, such as the number of pathlines from each validated flame that is needed to build the library, the resolution of the library, and the choice of variable histories to which $f_v$ should be correlated in order to achieve a reasonable prediction for the $f_v$ field. A rigorous validation is done at the end, in which the estimator technique is tested against with data that did not go into the library.

2. Numerical methods

In this section the soot estimator is explained briefly, followed by the test cases considered in this work. Then the method of post-processing the history of fluid parcels from a flame is described. An explanation follows on how this information is used to estimate soot volume fraction. Lastly, the numerical procedures are detailed.

2.1. Soot estimator

Background:
The soot estimator is developed to aid combustion device designers in reducing soot emissions (Alexander et al., 2018). From experimental and validated numerical data available in the literature, a library is created to be used as a post-processing tool to predict soot concentrations based on existing numerical Computational Fluid Dynamics (CFD) data. This library is built from Lagrangian hystereses of key parameters that correlate soot formation from inception up to oxidation.

Theoretically, the soot formation and destruction rates can be written as a function of local characteristics at any given time $t$ as

$$\frac{df_v(t)}{dt} = f(T(t), Y_i(t), P(t), f_v(t), A_s(t), \ldots),$$

where $T$ is the temperature, $Y_i$ is the mass fraction of any given species $i$, $P$ is the pressure, $f_v$ is the soot volume fraction, and $A_s$ is the average soot surface area. Nevertheless, due to the intrinsic non-linearity of the problem and the fact that the soot time scale is longer than the flow and gas-phase chemistry scales, the evolution of soot cannot be tabulated into a library as a function of just local field characteristics.

However, if the evolution is then seen from a history-point-of-view, theoretically, soot concentration can be determined entirely by the history of a fluid parcel in which it is contained, from before inception to oxidation or smoke emission. This idea is supported by evidences shown in Veshkini, Dworkin, and Thomson (2014) and Kholghy, Veshkini, and Thomson (2016), in which it has been shown that soot evolution has a stronger dependency on the temperature history of soot particles than just particle size and local flame temperature. In Veshkini et al. (2014) the concept of temperature aging was used to develop a new rate for soot growth processes; and in Kholghy et al. (2016) important processes for soot maturity, such as the particle dehydrogenation, PAH molecular weight growth and shell formation, were regulated by temperature-time history of particles. So, within the Soot Estimator technique, it is assumed that the local instantaneous soot volume fraction $f_v$ can be correlated to the history of thermo-chemical properties of a combustion system, or key parameters, to which a fluid parcel has been exposed as

$$f_v(t) = f(T_h(t), Y_{ih}(t), P_h(t), f_v,h(t), A_{s,h}(t), \ldots),$$

where $T_h$ is the temperature history, $Y_{ih}$ is the history of species $i$, $P_h$ is the time-integrated history of pressure, $f_{v,h}$ is integrated time history of soot volume fraction by time, and $A_{s,h}$ is the average soot surface area history. Therefore, if one tracks the history of all key parameters it would be possible to estimate the local $f_v$ without solving soot model equations. Equation (2) presents the fundamental idea from which the Soot Estimator technique is created.

Overview:
The soot estimator consists of a library of a series of numerical solutions of well validated sooting flames. Sectional detailed soot modeling are employed as well as a detailed kinetic mechanism to predict PAHs, which are the gas species responsible for the nucleation of the
soot particles (Eaves et al., 2012, 2016; Saffaripour, Kholghy, Dworkin, & Thomson, 2013; Veshkini et al., 2014; Zhang, Thomson, Guo, Liu, & Smallwood, 2009). From these numerical solutions the Lagrangian histories of soot-containing fluid parcels are algorithmically gathered and stored in a library. For populating the library, different validated numerical sooting flames are used. The library is built upon the integration in time of variables that are important to soot formation, such as temperature, mixture fraction $MF$, $C_2H_2$, $C_6H_6$, $OH$ and/or $O_2$. Once the library is created it can be used as a post-processor in conjunction with existing numerical data so no further modeling is required if only resulting soot quantities are desired. This technique can not only be applied to estimate $f_v$ as Equation (2) suggests, but it can also applied to correlate to soot characteristics parameters, such as particle number density, primary particle diameter, etc. For the purpose of the current study only $f_v$ predictions are analyzed. More information on the soot estimator can be found in Alexander et al. (2018).

Creation of library and usage:
The creation and usage of the Soot Estimator technique can be broken down into various steps, and for completeness the main steps are presented here in a summarized manner.

For creating the library:

1. Gathering the flames (see Section 2.2 for more details);
2. Extracting pathlines (see Section 2.3 for more details);
3. Computing histories from pathlines (see Section 2.4 for more details);
4. Averaging and storing $f_v$ associated with histories into a database (see Sections 2.4 and 2.7 for more details);

For using the library, similar steps to library creation are used:

1. Gathering the target flame(s);
2. Extracting pathlines;
3. Computing histories from pathlines;
4. Interpolating $f_v$ values from histories using the library database (see Section 2.4 for more details);
5. Using $f_v$ field estimation framework to compute $f_v$ field (see Section 2.5 for more details).

Most of these steps are described in Alexander et al. (2018). The most important steps for the present work are repeated and expanded here.

Advantages, disadvantages and expected errors of the technique:
As more flames are being simulated and validated, more data can be added so that the library is expanded, making it the biggest advantage of this framework.

A drawback is that the accuracy of the predictions is related to how close the target flame characteristics are to the flames stored in the library.

Species conservation is explicitly accounted for in the simulations that generate the flame CFD data, which get used to generate the library. Therefore, the library contains entries for soot concentration that are borne out of data from simulations that included conservation. When the library is then used to estimate soot concentrations in a given system, the quality and accuracy of the estimation will reflect the quality of the data used to estimate soot for the system in question. For example, if experimental data is used, then conservation will implicitly be accounted for. If numerical data is used and soot concentrations are estimated to be low ($\sim O(1 \text{ppm})$ or lower), then having neglected conservation will have minimal impact (Zimmer et al., 2017). If the soot concentrations are estimated to be high, then the error associated with neglecting conservation will be greater. If the numerical results are tuned to experimental temperatures, which is common practice in the gas turbine industry, the error associated with neglecting conservation will be minimal (Zimmer et al., 2017).

2.2. Data source: flames
The numerical solution for nine atmospheric pressure laminar diffusion flames simulated in Veshkini et al. (2014) were used in the present work to generate the libraries and as the target flames. In Veshkini et al. (2014) detailed models were used: For the gas-phase kinetic modeling, the fully coupled elliptical conservation equations for mass, momentum, energy, and species mass fraction were solved. The equations were solved in the two-dimensional ($z$ and $r$) cylindrical coordinate system. The DLR detailed kinetic mechanism (Dworkin, Zhang, Thomson, Slavinskaya, & Riedel, 2011; Slavinskaya & Frank, 2009; Slavinskaya, Riedel, Dworkin, & Thomson, 2012), consisting of 93 species and 719 reactions, was used to describe the oxidation of the fuel and the formation of PAHs. Soot was modeled using a detailed fully coupled sectional aerosol dynamics model (Park & Rogak, 2004; Park, Rogak, Bushe, Wen, & Thomson, 2005), and the soot particle mass range was divided into 35 sections. For each soot section two equations were solved, one for soot aggregate number densities, and another for primary particle number densities. The nucleation process is based on the collision of two
Table 1. Flame main characteristics.

| Flame code | Inner dia. (mm) | Fuel veloc. (cm/s) | Air veloc. (cm/s) | Fuel composition (volume) | Computeda peak f_v (ppm) |
|------------|----------------|-------------------|------------------|-------------------------|-------------------------|
| SA^a       | 11             | 5.06              | 13.3             | 100% C_H_4              | 10.74                   |
| SM32^b     | 4              | 35                | 35               | 32% C_H_4/68% N_2       | 0.12                    |
| SM40^b     | 4              | 35                | 35               | 40% C_H_4/60% N_2       | 0.36                    |
| SM60^b     | 4              | 35                | 35               | 60% C_H_4/40% N_2       | 1.68                    |
| SM80,2^b   | 4              | 17.5              | 17.5             | 80% C_H_4/20% N_2       | 1.80                    |
| SM80^b     | 4              | 35                | 35               | 80% C_H_4/20% N_2       | 3.21                    |
| SY41^c     | 11             | 4.1               | 8.7              | 100% C_H_4              | 12.65                   |
| SY46^c     | 11             | 4.6               | 8.7              | 100% C_H_4              | 14.76                   |
| SY48^c     | 11             | 4.8               | 8.7              | 100% C_H_4              | 16.39                   |

Notes: Experimental of ^aSantoro et al. (1983), ^bSmooke et al. (2005), and ^cShaddix and Smyth (1996); Computed of ^dVeshkini et al. (2014)

Pyrene molecules (C_{16}H_{10}) (Appel, Bockhorn, & Frenklach, 2000; Frenklach & Wang, 1994). Also, the processes of surface growth, based on the HACA mechanism (Appel et al., 2000), PAH condensation, coagulation, soot particle coalescence, and oxidation were taken into consideration. All the numerical simulations were validated against experimental data in Veshkini et al. (2014) to varying degrees.

It is important to note that the flame data used to populate the library need not be frozen. As flame simulation techniques continue to advance, newly computed data can be used to either regenerate or expand the library.

These nine flames represent the classical works from Santoro, Semerjian, and Dobbins (1983), Smooke, Long, Connelly, Colket, and Hall (2005), and Shaddix and Smyth (1996). They range from pure to diluted ethylene–air laminar coflow flames and from low soot concentration (f_v,max = 0.1 ppm) to moderate soot concentration (f_v,max = 16.3 ppm). The computed values for peak soot volume fraction are within the experimental range. The main characteristics of the set of numerical simulations are shown in Table 1.

These computed flames were used to compute the pathlines and their histories, which then were used to build the libraries and estimate the local f_v using the post-processor.

### 2.3 Pathlines calculation

The history of each variable can be expressed by the integral of each local variable with respect to time along a pathline traversed by a fluid parcel that may contain soot. Each pathline p is calculated from following a soot-containing fluid parcel position X_p which follows the fluid velocity û with respect to time t solving the ordinary differential equation:

\[
\frac{dX_p}{dt} = û(X_p(t), t), \quad X_{p,0} = X_{p,0}, \quad p = 1, \ldots, N_p, \quad (3)
\]

where each pathline p has an initial position X_{p,0}, and N_p is the total number of pathlines.

### 2.4 Histories calculation, library generation, and library usage

In theory, the number of key parameters used in Equation (2) could be as high as the total number of variables traceable in the detailed solution. Nevertheless, storing such an amount of information would be unfeasible. Currently, the choice of the key parameters is based on the background knowledge of soot formation, which means that histories that can track conditions for soot nucleation, growth and oxidation are preferred, and it is also partially based on a trial-and-error process (Alexander et al., 2018).

So, from the previous work (Alexander et al., 2018), at least three key parameters are necessary to estimate f_v, and the following are the starting point of the current study: the histories of temperature, mixture fraction MF (calculated with Bilger’s expression (Bilger, 1989)) and O_2 molar fraction X_{O_2}. They are calculated for each pathline as:

\[
T_h = \int_{\text{pathline}} T(\bar{X}, t) \, dt, \quad (4)
\]

\[
MF_h = \int_{\text{pathline}} MF(\bar{X}, t) \, dt, \quad (5)
\]

\[
X_{O_2,h} = \int_{\text{pathline}} X_{O_2}(\bar{X}, t) \, dt, \quad (6)
\]

where T_h is the temperature history, MF_h is mixture fraction history, and X_{O_2,h} is the O_2 molar fraction history.

Once the histories are calculated for all N_p pathlines, the histories and the associated f_v values are stored in the soot estimator library, so that f_v ≈ f(T_h, MF_h, X_{O_2,h}). The range of each history is divided into a specific number of bins and the final f_v value associated to each bin is averaged from the number of entry points for that bin.
In the last step of the soot estimator, the dependent variables are post-processed from an existing numerical solution, using the same steps described above, and then the local \( f_v \) is estimated for each position \( \vec{X}_p \), for each pathline, using the library and some interpolation technique.

### 2.5. \( f_v \) field estimation framework

The governing equations for the simulation done in Veshkini et al. (2014) were solved in Eulerian coordinates, therefore any variable \( \phi(r,z) \) field is presented in \( r \) and \( z \). From this field the pathlines are calculated through Equation (3), so that each pathline is in Lagrangian coordinates and presented in \( \vec{X}_p(t) \). In this calculation process each pathline \( p \), at each time \( t \), is associated to a specific location in \( r \) and \( z \) as \( \vec{X}_p(t) = f(r,z) \). These pathline points are scattered in the \( r \) and \( z \) coordinate, which means that they do not follow the uniform grid. Then using interpolation, it is possible to interpolate this scattered data over the original uniform grid (Eulerian coordinates). In the present study the nearest neighbor interpolation method with Delaunay triangulation of the data is used to retrieve the estimated \( f_v \) field after the estimator library has been used for every pathline. This procedure is schematically shown in Figure 1.

**Figure 1.** Schematic of \( f_v \) field estimation framework.

In Step 1, the pathlines are calculated from the Eulerian coordinates. Three pathlines are shown in color and each time step for each pathline \( (\vec{X}_p(t) = f(r,z)) \) is represented as a black dot. In Step 2, the soot estimator is applied to all points of these pathlines to get the estimated \( f_v(X_p(t)) \), represented as red dots. In Step 3, a scattered data interpolation technique is used to get the estimated \( f_v \) field in \( r \) and \( z \), represented as a red uniform grid.

### 2.6. Important parameters

In this subsection important parameters to the estimator technique are discussed and analyzed, such as the number of pathlines from each flame to build the library, the variable histories to which \( f_v \) should correlate in order to achieve a reasonable \( f_v \) solution, and the resolution of the library.

**Pathlines: quantity and location**

The number of calculated pathlines \( N_p \) is determined by the level of detail required to capture the desired field (Lebedeva & Osiptsov, 2016) (the \( f_v \) field in the present case). In order to do so, every pathline starts at the inlet boundary and ends up at the outlet boundary. At the inlet, every computational node is used to start every pathline and extra pathlines are needed to cover the sooting area. This sooting area is defined from the flame centerline to the extent of the wings along the radial coordinate and the pathlines. In order to ensure that the peak \( f_v \) is mapped out an additional pathline is added to track the pathline that goes through the maximum \( f_v \). To study the effect of the number of pathlines needed to create the library, three different cases are tested. These cases leverage the already refined mesh of the computed domain to cover the sooting area. In order for the radial distance between each node to be kept at the same \( dr \), it is divided by two \( dr/2 \), and then it is divided by three \( dr/3 \), and these cases are presented in Table 2.

For each flame the \( N_p \) is different, since each flame requires a unique grid to capture the soot field. The number of pathlines for the Smooke flames are generally half of the other flames, as those flames are smaller and thinner than the others, and therefore need fewer grid points to be solved. Another point to note is that the extent of
the major variables found in any combustion system. The particles, and the can be a strong indicator for relative heat transfer into and/or oxidation. For example: the temperature history directly or indirectly to the reactions of soot formation affect soot surface reactions. Each variable can contribute major species which lead to PAH formation and thus the dimensionality (number of variable histories) of the solution and destruction of a range of flames. The effect of the dimensionality (number of variable histories) of the database was analyzed in the previous work of (Alexander et al., 2018). The authors found out that the more dimensions, the better soot predictions were, but with a higher computational cost. In the current study, only three-dimension libraries are used. So, the effect of the choice of these three histories is presented in Section 3.3.

### 2.7. Numerical method

The pathlines, Equation (3), are computed using a second-order time integration scheme, with an adaptive time-stepping based on the Courant Number \( CN = 0.5 \). This method ensured similar accuracy as constant time-stepping of \( \Delta t = 10^{-6} \), but with fewer iterations.

A new algorithm for library creation was developed in the present work in order to achieve higher computational speed. The library creation process is composed mainly of two parts: first, sorting all entry points of the data between the bins boundaries; and second, averaging the \( f_v \) values in those bins which received multiple entries. The improvement made was on the routine to sort and populate the database, since this step is the most computationally demanding. Every pathline evolves differently in the database space and only a small portion of the bins are actually populated. As each pathline evolves in a highly non-linear way, we found that we are unable to use functions to classify the entries into the bins. This information was taken into account to improve searching and then populating steps of the current populating procedure. So, in the sorting process every entry is sorted one at a time following the list of entries (order as pathline evolution) only once. Also, the bin location associated to the previous entry is used as the initial searching point for the next entry. In this way, the sorting process looks only for the bins which are closer to that entry and not through all bins of all dimensions as in (Alexander et al., 2018). With this new algorithm higher resolution is possible with less computational cost. For example, let’s compare the computational time for a 3D library with 100 bins for each independent variable for the SA flame, with \( Np = 134 \), which contains 133,866 entry points. The old algorithm (Alexander et al., 2018) took 26 minutes on a standard desktop PC and the new one took only 0.32 seconds, a speed up of 4,875. The library is built from the histories of pathlines of the sooting area only, so that it only contains relevant information. All numerical implementation of this work was done in MATLAB R2018a framework.

### 3. Results and discussion

In this section the results are presented and qualitative analysis is made. The solution of the detailed soot equations reported in Veshkini et al. (2014) will be referred to

| Flames | \( dr \) | \( dr/2 \) | \( dr/3 \) |
|--------|--------|--------|--------|
| SA     | 70     | 134    | 198    |
| SM32   | 31     | 55     | 80     |
| SM40   | 32     | 58     | 84     |
| SM60   | 34     | 63     | 92     |
| SM80.2 | 39     | 72     | 104    |
| SM80   | 39     | 71     | 103    |
| SY41   | 71     | 135    | 200    |
| SY46   | 70     | 134    | 198    |
| SY48   | 70     | 134    | 198    |
as computed and the solution provided by the soot estimator will be referred to as estimated or predicted, interchangeably. For quantitative comparison, the relative errors \( \text{error} = \frac{\text{abs}(\phi_{\text{computed}} - \phi_{\text{estimated}})}{\phi_{\text{computed}}} \) of two important variables: 1) Peak \( f_v \), and 2) Integrated \( f_v \) are used to address the difference between the computed and estimated \( f_v \) fields. The integrated \( f_v \) represents the volumetric integration of the \( f_v \) in the entire domain and is calculated as:

\[
f_{v,\text{int}} = \iint_V 2\pi r f_v(r,z) \, dr \, dz. \quad (7)
\]

This variable is chosen since it brings a quantification of the overall soot formation in the entire domain.

The development of the soot predictor is composed by various steps, from testing a library generated by a single flame data on that flame, up to predicting a range of flames with a single library created by the same flames. All the steps of this development are addressed in this section, as well as the important parameters for building a framework which predicts the entire \( f_v \) field with quantitative accuracy. As the framework is designed for predictive work, a final step in this development is required - a rigorous test. This test is used to check its ability to predict soot formation in a system which itself did not contribute to the generation of the library.

So, in the first part of this section the framework is assessed for individual flames in order to test the overall concept of the soot prediction in the entire domain. Following that, a parametric study is conducted to see the influence of the number of pathlines and the number of bins on the soot prediction. Next, the chosen settings are used to assess the \( f_v \) field framework with a single combined library and a new study is done to find the best combination of variable histories. In the last part, a rigorous validation test, with independent data, is presented.

### 3.1. Assessing the \( f_v \) field framework: individual flame validation

Here the overall framework concept is tested for individual flames, in which the library is built from each specific flame's histories and then used to predict soot concentration within the same flame. The libraries are based on \( T_b, MF_{th}, \tau_{O_2,th} \), with 100 bins, and \( N_p = dr \).

The first flame assessed is the SA flame and the \( f_v \) field can be seen in Figure 2. The scales are different so that graph a) can emphasize how the soot volume fraction field is located in a very narrow region, and graphs b) and c) can show that the region close to the centerline (\( r = 0 \) cm) is not predicted (graph c) as well as the computed \( f_v \) field (graph b). All three graphs share the same peak \( f_v \), and therefore the same \( f_v \) contour level, presented in the right part of graph c) from 0 up to 10.7 ppm.

It is possible to see some important aspects of the prediction: First, the framework predicts both the magnitude and overall distribution of the \( f_v \) field in such a small area, and second, the soot values close to the centerline are not well captured. Similar results were found for the other flames, but are not shown here. The quantitative assessment of the results, for all nine flames, is shown in Table 3.

The table shows that the peak \( f_v \) is accurately predicted for all flames and that the \( f_v \) field, assessed by a volumetric integration, is captured within the same order of magnitude as the detailed solution. All flames presented integrated \( f_v \) errors around 10% except for the SM32 and SM40 flames, for which the errors were higher, even though the errors on the peak \( f_v \) for those flames

![Figure 2. \( f_v \) fields for SA flame; \( f_v \), level in color; (a) Computed \( f_v \) field, scale 1:1; (b) Computed \( f_v \) field, not in scale; (c) Estimated \( f_v \) field, not in scale.](image)

| Flames | Peak \( f_v \) estimator relative error (%) | Integrated \( f_v \) estimator relative error (%) |
|--------|-----------------------------------------|---------------------------------------|
| SA     | 0.65                                    | 11.69                                 |
| SM32   | 0.66                                    | 46.46                                 |
| SM40   | 0.68                                    | 33.33                                 |
| SM60   | 3.58                                    | 14.12                                 |
| SM80.2 | 1.38                                    | 14.79                                 |
| SM80   | 0.30                                    | 11.83                                 |
| SY41   | 0.80                                    | 10.62                                 |
| SY46   | 1.69                                    | 11.05                                 |
| SY48   | 2.58                                    | 10.35                                 |
were low. The reason for the errors is because any particular bin may contain training data from more than one streamline which gets averaged. Those bins then get used for predictions. In the combined limit of infinitely refined bins with perfectly accurate determination of dimensions and correlations, no errors should occur. However, practical constraints lead to the errors seen. Two conclusions can be drawn here; the first is that for predicting the entire domain, the integrated \( f_v \) is a better indicator than the peak \( f_v \), therefore it will be used herein as the best guide to develop this work; and the second conclusion is that for capturing the entire soot field of SM32 and SM40 flames properly, more investigation into the practical constrains of the technique must be completed.

### 3.2. Parametric study: effect of the number of pathlines \( N_p \) and the library resolution on the \( f_v \) prediction

In this section the effect of the number of pathlines \( N_p \) and the library resolution on predicting \( f_v \) field is presented. The starting point is the case presented in the previous section, where the libraries are based on \( T_h, M F_{h}, \) and \( X_{O_2,h} \), with 100 bins, and \( N_p = dr \). Three cases of library resolution of 100, 200, and 300 bins for each variable history, and three cases of varying numbers of pathlines, presented in Table 2, are tested. The libraries are built from each specific flame's histories and then used to predict the same flame. Table 4 presents the comparison of the relative error of \( f_{v,int} \) for each flame and case.

From the table it is possible to see that: First, looking at the average values, the number of bins have a larger effect on lowering the errors than the number of pathlines added to the library. Second, the case \( dr/2 \) has a higher influence (higher sensitivity) in reducing the errors from the previous case (\( dr \)), than the \( dr/3 \) case does in reducing the errors from \( dr/2 \). Thirdly, as seen previously, the SM flames have a higher error than the other flames for 100 bins, and similar results are found for 200 bins, but these errors vanish for the higher resolution libraries. This means that for the current library based on \( T_h, M F_{h}, X_{O_2,h} \), to capture the entire \( f_v \) field properly, a higher library resolution is needed. The setting of 300 bins and \( dr/2 \) is chosen for further study, since it gives the average \( f_{v, int} \) error value below 1%.

### 3.3. Effect of the choice of variable history with the combined libraries

In this section the ability of the soot estimator framework is assessed on reproducing all nine flames with a single combined library, built with all pathlines from these flames. In addition, the effect of variable history choice on predicting soot concentration for all flames is addressed here. In Section 2.6, 10 key variables were selected, and 120 possible combinations of three variables at a time were tested here.

Table 5 shows a detailed comparison between the baseline library combination of \( T_h, M F_{h}, X_{O_2,h} \) and the one with the lowest relative errors \( T_h, M F_{h}, X_{H_2,h} \).

Some important aspects can be seen in the table. Firstly, the single library with baseline combination \( (M F_{h}, T_h \) and \( X_{O_2,h}) \) could predict the entire \( f_v \) field of some flames better than others. The estimator was able to reproduce the flames with similar fuel composition and burner characteristics, SA and SY flames, while the SM
Figures were not accurately predicted. Secondly, the single library with the best combination \((MF_{th}, \, T_{th}, \, X_{H_2h})\) predicted all flames with high accuracy. This is the first time that one single library accurately predicts integrated \(f_v\) in such diverse flames with such accuracy. Lastly, the substitution of the \(X_{O_2h}\) to \(X_{H_2h}\) improved the prediction of all flames, especially the SM flames. To show the improvement, the comparison of the computed soot field and the predicted by the two libraries for the SM32 flame is shown in Figure 3. Note the differing scales in the colourmaps.

It is of interest to see the great improvement in the prediction of the \(f_v\) field from the baseline library combination to the one with \(H_2\). The distribution and the magnitude of the predictions with the baseline library were not close to the computed solution, but the library based on \(MF_{th}, \, T_{th}, \, X_{H_2h}\) was able to capture all flame features.

The overall improvements due to the substitution of \(X_{O_2h}\) with \(X_{H_2h}\) in the library can be explained by the importance of \(H_2\) in the surface growth mechanism (Appel et al., 2000) in three parts. The first, is that the history of \(H_2\) brings a direct correlation with the surface growth of soot particles since it plays an important role in the soot HACA mechanism. Second, the history of \(H_2\) is also correlated with other species due to the chemical mechanism, bringing additional correlation of the gas species to soot formation and oxidation reactions. Third, \(O_2\) species could not provide enough information in areas of the flame where it has low concentration and soot formation is of importance, such as at the centerline. Also, \(O_2\) is specially important only in the soot oxidation region, after soot has nucleated and grown. The last point is that when all three variables \(MF_{th}, \, T_{th}, \, X_{H_2h}\) were combined a three-dimensional space highly correlated with soot formation was created, as perceived by Equation (2), independently of the flame characteristics.

In order to provide an overall view of the effect of the other species and more options for the choice of variables, the five best, the baseline and the worst combinations are shown in Table 6.

From the table is interesting to see that all first five share similar characteristics, either in participating in the thermal exchange of energy \(T_{th}\), main combustion products \(CO_2\) and \(H_2O\), intermediate product \(CO\), or in the HACA mechanism as \(C_2H_2, \, H_2\) and \(CO\). Nevertheless, the most important aspect is that all five incorporate at least one species that participates in the soot growth processes, which can not be seen in the baseline (Rank 112) and in the worst composition (Rank 120).

Discussion about the Library

For the best library combination, the averaging process on those bins that have multiple entries produced a mean value with a variation of 9.6\%, with a confidence of 2 standard deviations. This small variation indicates that most of the points averaged are close to the mean, and that the
Table 7. Validation of the framework; Predicting new flames.

| Flames  | Integrated $f_v$ estimator relative error [%] | Peak $f_v$ estimator relative error [%] |
|---------|-----------------------------------------------|---------------------------------------|
| SA      | 15.39                                         | 4.17                                  |
| SM32    | 3.33                                          | 53.61                                 |
| SM40    | 0.59                                          | 86.38                                 |
| SM60    | 12.12                                         | 18.84                                 |
| SM60.2  | 18.37                                         | 45.03                                 |
| SM80    | 5.73                                          | 69.27                                 |
| SY41    | 8.52                                          | 4.03                                  |
| SY46    | 10.04                                         | 27.42                                 |
| SY48    | 9.32                                          | 19.49                                 |

outlier points do not create excessive noise on the final stored value.

During library formation, only a small part of it is populated, 0.67%, and during the process of predicting, an even smaller space is accessed to obtain the soot concentration to every single flame. This is expected, since the boundary limits are broad for considering all nine flames, each dimension is orthogonal to each other, and the bins are uniformly spaced along the domain.

3.4. Validating the combined library

In this section the chosen framework configuration is assessed in two scenarios in which the library is validated against an independent set of data. As the framework is designed for predictive work, it is necessary to rigorously test its ability to predict soot formation in a system which itself did not contribute to the generation of the library.

Predicting new flames:

In this scenario one flame from the dataset is selected as the target flame and the library is built from the remaining flames. In this case, the ability of the framework to predict the $f_v$ field on new or untrained data is assessed. This exercise stresses some important aspects of the technique, such as its ability to predict new flames and its dependency on the data that was used to create the library. Table 7 shows the relative error in predicting each of the nine flames with a library built from the remaining eight.

Table 7 shows the relative error of integrated $f_v$, and, additionally, the error on the peak $f_v$ to complement the assessment. Some important aspects can be drawn to attention. First, the ability of predicting the overall soot formation for all flames; all presented errors are lower than 19%. Second, the predicted peak $f_v$ were all within the same order of magnitude, and for SA and SY41 flames the relative error was below 5%. Third, some flames shared similar characteristics and therefore were predicted with a higher accuracy, for example SA and SY41. For other flames for which flame characteristics and fuel composition were not found in the training set, the predicting capability of the soot estimator was more vigorously tested, such as flame SM60 for which reasonable errors of less than 20% were found for both integrated and peak $f_v$. It is important to state that the ideal case for using the soot estimator technique is to have a greater number of flame datasets in order to extrapolate the predictive capabilities to new flames. Nevertheless this assessment with only eight flames was adequate to demonstrate the potential of the technique.

Cross validation technique:

In the cross validation technique, the full dataset, with all nine flames, is partitioned into two random independent sets, one for building the library (or training) and the other for validating. In this scenario the predicting points will be found within the boundaries of the dataset, making this test a different assessment than the previous one. Here the $k$-fold cross validation (Geisser, 1975) is applied to assess the accuracy of the framework. In this technique the dataset, after being randomized, is divided into $k$ groups, or folds, of similar size where $(k-1)$ folds are used for training and the last $k$ fold is used for evaluation. This process is repeated $k$ times, and during the iteration each of the $k$ folds is assigned as validation data.

For a less biased prediction (Rodriguez, Perez, & Lozano, 2010) $k$ is set equal to 10 in this study, and all flames combined generated a full data set of 607,604 entries, so for each of the ten iterations, the training set is composed by 546,844 entries and the validation set by 60,760 entries.

The accuracy is measured with the relative error of $f_v$ ($\text{error} = \text{abs}((f_{v,\text{computed}} - f_{v,\text{estimated}}) / f_{v,\text{computed}})$) for every entry of the validation set. This is a more rigorous test than the ones of the previous sections, since every single entry is assessed instead of the overall $f_v$ field prediction. Nevertheless, this assessment also brings a challenge, since the exact prediction of $f_v$ close to numerical zero values results in errors and does not bring any valid information to the usage of this soot estimator. So, a threshold of $f_v = 0.0001$ ppm (or 0.1 ppb) is used to limit this behavior. The accuracy of the technique is shown in Table 8.

Table 8 presents a lot of information, but some results can be highlighted. First, the cross validation technique creates unique independent validation data for every iteration resulting in a less biased evaluation of the soot predictor. For every iteration the statistics of the error are different, especially the standard deviation which varies widely from 165 to 645. Second, looking at the averaged error values, it is observed that most of the errors are very small (median of 0.31), but there are some points for which the errors are larger resulting in an increase of the overall mean to 8.27% and the standard deviation to 314.98. This effect can be seen in the last three columns,
where most of the predicted $f_v$ points have low errors. It is important to remember that it is common to find measurement errors of about 30-40% on $f_v$ presented in the literature. If that is taken into account on using the present soot estimator, the predictor presented very good accuracy.

### Table 8. $k$-fold cross validation of the combined library.

| $k$-fold iter | Mean | Median | Std    | < 5% | < 10% | < 50% |
|---------------|------|--------|--------|------|-------|-------|
| 1             | 8.51 | 0.30   | 267.99 | 88.85| 93.86 | 99.19 |
| 2             | 8.16 | 0.30   | 224.04 | 88.70| 93.68 | 99.09 |
| 3             | 7.41 | 0.31   | 219.54 | 88.27| 93.46 | 99.05 |
| 4             | 10.00| 0.32   | 628.18 | 88.29| 93.57 | 99.15 |
| 5             | 10.31| 0.31   | 645.22 | 88.56| 93.62 | 99.08 |
| 6             | 8.71 | 0.31   | 382.55 | 88.67| 93.72 | 99.10 |
| 7             | 6.48 | 0.31   | 165.09 | 88.57| 93.79 | 99.09 |
| 8             | 8.86 | 0.30   | 306.88 | 88.53| 93.65 | 99.11 |
| 9             | 6.87 | 0.31   | 289.14 | 88.37| 93.52 | 99.12 |
| 10            | 7.45 | 0.30   | 291.15 | 88.60| 93.64 | 99.13 |
| Average       | 8.27 | 0.31   | 341.98 | 88.54| 93.65 | 99.11 |

The technique has demonstrated its potential to work with different flame characteristics and fuel compositions. For further usage, flame datasets similar to the target problem must be gathered so that the library can be improved.

### Notation

The following symbols are used in this paper:

- $A_s$: Average soot surface area, m$^2$
- $CN$: Courant Number, –
- $f_v$: Soot volume fraction, –
- $MF$: Mixture fraction, –
- $Np$: Total number of pathlines, –
- $P$: Pressure, atm
- $r$: Cylindrical radial coordinate, m
- $t$: Time, s
- $T$: Temperature, K
- $\mathbf{u}$: Fluid velocity vector, m/s
- $Y_i$: Mass fraction of any given species $i$, –
- $X_i$: Molar fraction of any given species $i$, –
- $\mathbf{X}_p$: Position vector, m
- $z$: Cylindrical axial coordinate, m

### Subscript:

- $h$: History or time-integrated
- $int$: Integrated
- max: Maximum value
- $s$: Soot

### Acknowledgments

This research was undertaken, thanks to funding from the Canada Research Chairs program, the Natural Sciences and Engineering Research Council of Canada (NSERC) with the grant number 401836-2012, the Ontario Ministry of Research,
Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

L. Zimmer http://orcid.org/0000-0002-4642-5122

References

Alexander R., Bozorgzadeh S., Khosousi A., & Dworkin S. B. (2018). Development and testing of a soot particle concentration estimator using Lagrangian post-processing. *Engineering Applications of Computational Fluid Mechanics, 12*(1), 236–249.

Appel J., Bockhorn H., & Frenklach M. (2000). Kinetic modeling of soot formation with detailed chemistry and physics: Laminar premixed flames of c2 hydrocarbons. *Combustion and Flame, 121*(1–2), 122–136.

Bilger R. (1989). The structure of turbulent nonpremixed flames. *Symposium (International) on Combustion, 22*(1), 475–488.

Dworkin S. B., Zhang Q., Thomson M. J., Slavinskaya N. A., & Riedel U. (2011). Application of an enhanced pah growth model to soot formation in a laminar coflow ethylene/air diffusion flame. *Combustion and Flame, 158*(9), 1682–1695.

Eaves N. A., Veshkini A., Riese C., Zhang Q., Dworkin S. B., & Thomsons M. J. (2012). A numerical study of high pressure, laminar, sooting, ethane–air air coflow diffusion flames. *Combustion and Flame, 159*(10), 3179–3190.

Eaves N. A., Zhang Q., Liu F., Guo H., Dworkin S. B., & Thomson M. J. (2016). Collage: A refined and validated numerical algorithm for modeling sooting laminar coflow diffusion flames. *Computer Physics Communications, 207*, 464–477.

Fairweather M., Jones W., & Lindstedt R. (1992). Predictions of radiative transfer from a turbulent reacting jet in a crosswind. *Combustion and Flame, 89*(1), 45–63.

Frenklach M., & Wang H. (1994). Detailed mechanism and modeling of soot particle formation. In *Soot formation in combustion* (pp. 165–192). Springer.

Geisser S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70*(350), 320–328.

International Civil Aviation Organization (2006). Annex 16, environmental protection: Aircraft engine emissions, vol. 2. *International Standards and recommended Practices. International Energy Agency* (2017). 2017 key world energy statistics report. International Energy Agency.

Kholghy M. R., Veshkini A., & Thomson M. J. (2016). The core–shell internal nanostructure of soot – a criterion to model soot maturity. *Carbon, 100*, 508–536.

Lebedeva N. A., & Osipov A. N. (2016, September 1). A combined Lagrangian method for simulation of axisymmetric gas-particle vortex flows. *Fluid Dynamics, 51*(5), 647–659.

Leung K., Lindstedt R., & Jones W. (1991). A simplified reaction mechanism for soot formation in nonpremixed flames. *Combustion and Flame, 87*(34), 289–305.

Maas U., & Pope S. (1992). Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combustion and Flame, 88*(3–4), 239–264.

Park S., & Rogak S. (2004). A novel fixed-sectional model for the formation and growth of aerosol agglomerates. *Journal of Aerosol Science, 35*(11), 1385–1404.

Park S. H., Rogak S. N., Bushe W. K., Wen J. Z., & Thomson M. J. (2005). An aerosol model to predict size and structure of soot particles. *Combustion Theory and Modelling, 9*(3), 499–513.

Peters N. (1984). Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in Energy and Combustion Science, 10*(3), 319–339.

Rodriguez J. D., Perez A., & Lozano J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(3), 569–575.

Saffaripour M., Kholghy M., Dworkin S., & Thomson M. (2013). A numerical and experimental study of soot formation in a laminar coflow diffusion flame of a jet a-1 surrogate. *Proceedings of the Combustion Institute, 34*(1), 1057–1065.

Santoro R., Semerjian H., & Dobbins R. (1983). Soot particle measurements in diffusion flames. *Combustion and Flame, 51*, 203–218.

Shaddix C. R., & Smyth K. C. (1996). Laser-induced incandescence measurements of soot production in steady and flickering methane, propane, and ethylene diffusion flames. *Combustion and Flame, 107*(4), 418–452.

Slavinskaya N., & Frank P. (2009). A modeling study of aromatic soot precursors formation in laminar methane and ethene flames. *Combustion and Flame, 156*(9), 1705–1722.

Slavinskaya N. A., Riedel U., Dworkin S. B., & Thomson M. J. (2012). Detailed numerical modeling of pah formation and growth in non-premixed ethylene and ethane flames. *Combustion and Flame, 159*(3), 979–995.

Smooke M., Long M., Connelly B., Colket M., & Hall R. (2005). Soot formation in laminar diffusion flames. *Combustion and Flame, 143*(4), 613–628.

van Oijen J., & de Goey L. (2000). Modelling of premixed laminar flames using flamelet-generated manifolds. *Combustion Science and Technology, 161*, 113–137.

Veshkini A., Dworkin S. B., & Thomson M. J. (2014). A soot particle surface reactivity model applied to a wide range of laminar ethylene/air flames. *Combustion and Flame, 161*(12), 3191–3200.

Williams M., & Minjares R. (2016). A technical summary of euro 6/vi vehicle emission standards. *International Council for Clean Transportation (ICCT)* (2017).

Zhang Q., Thomson M., Guo H., Liu F., & Smallwood G. (2009). A numerical study of soot aggregate formation in a laminar coflow diffusion flame. *Combustion and Flame, 156*(3), 697–705.

Zimmer L., Pereira F. M., van Oijen J. A., & de Goey L. P. H. (2017). Investigation of mass and energy coupling between soot particles and gas species in modelling ethylene counterflow diffusion flames. *Combustion Theory and Modelling, 21*(2), 358–379.