A FULLY-CUSTOMIZED BASELINE REMOVAL FRAMEWORK FOR SPECTROSCOPIC APPLICATIONS. Stephen Giguere¹, M. Darby Dyar², Thomas Boucher¹, CJ Carey¹, and Sridhar Mahadevan¹, ¹School of Computer Science, University of Massachusetts Amherst, Amherst MA 01003, USA, sgiguere@cs.umass.edu, ²Dept. of Astronomy, Mount Holyoke College, South Hadley, MA 01075.

Introduction: In applications for which the shape, size, and distribution of spectral peaks present are important, spectroscopists often confront the problem of applying baseline removal (BLR) to their data. In particular, various physical phenomenon can introduce a slowly varying signal into each spectrum based on the type of spectroscopy or the instruments and techniques used during acquisition. This signal, called the baseline or continuum, may vary dramatically even among individual spectra from the same data set, and must be removed prior to quantitative analysis.

In the past two decades, many automated solutions to the BLR problem have been proposed for different applications. Commonly, these methods have multiple adjustable parameters that must be tuned to the user's specific data and task. Identifying the optimal parameter values is often a time-consuming process. Manual approaches, in which baseline parameters are adjusted until the resultant spectrum “looks good” or yields the best are still used in many applications. This approach performs sufficiently well for simple datasets and tasks, but can be very difficult to execute when applied to complex problems and large data sets. Our recent work [2] has demonstrated for LIBS and FTIR spectroscopies that no single BLR method outperforms the others in every case. In fact, applying the wrong method can yield worse results than not applying baseline removal at all, even after tuning parameters. Clearly it is necessary to test multiple BLR methods to be confident in the quality of results obtained.

The goal of this work to make full optimization of BLR accessible to practitioners by providing a system that can automatically synthesize baseline removal methods that are customized and tuned to a user's task. Given a specific set of data, our system constructs a multitude of BLR methods and tunes their parameters, often discovering novel and effective methods in the process. To make this possible, we introduce two major components: a conceptual framework for analyzing existing BLR methods that forms the basis for constructing new ones, and an optimization system that builds and tunes the methods. Results of this new technique applied to LIBS data are given in [1], but our new method is applicable to any type of spectroscopy sharing the BLR problem.

Framework Overview: To date, we docu-

![Figure 1](https://example.com/figure1.png)  
**Figure 1.** Illustration of effects of treating a generic spectrum (top panel) with peak identification, baseline extraction, and smoothing/filtering.
output of baseline extraction depends heavily on the intensity of channels representing the baseline, while channels containing peaks have minimal effect on the output. Finally, the smoothing and filtering subtask represents reduction of noise in the input signal, often applied as preprocessing for other subtasks. Although both this and the baseline extraction subtask involve computing a new signal from an input, the smoothing and filtering subtask is distinguished by the fact that it produces a relatively unbiased approximation to the input signal.

This framework for analyzing BLR methods provides the necessary foundation for constructing new methods using optimization. The fact that the same few subtasks appear repeatedly in existing methods provides evidence that those operations are generally useful for baseline removal. However, unlike the operations themselves, there is little agreement in the literature over how the subtasks are implemented or sequenced. We find that novel yet effective baseline removal methods can be constructed by using the same subtasks utilized previously, but varying their implementations (parameters) and their arrangement.

System Overview: The first step of our system involves building a large number of candidate BLR methods that will later be tuned. Our system first constructs an arrangement of subtasks that defines the ordering of the subtasks, as well as any subtasks that will be applied repeatedly in a loop. To limit the complexity of the method, the user specifies the maximum number of subtasks the arrangement may contain. Next, implementations are chosen for each subtask to define the precise computational steps that will be applied to perform the subtask. These are drawn from existing BLR methods. Multiple distinct methods are synthesized by repeating this process with different arrangements and implementations. Special care is taken to ensure that the methods produce valid baselines and do not waste computation.

Given the synthesized methods and a specification of the user's task, the second step of our system determines the method and parameters that perform best at the task. The strategy used to determine the best configuration can be as simple as repeatedly testing random methods and parameters using trial-and-error, although more sophisticated optimization methods are able to find good configurations more quickly. Once the optimization is finished, the BLR method and parameters found are saved, allowing them to be used in the future without repeating the optimization.

Summary: Our system for customized baseline removal makes full-scale optimization of BLR accessible to spectroscopists. The large number of available BLR methods and the cost associated with tuning them makes it difficult for users to select the correct method for their task. Nonetheless, the choice of baseline removal method and parameter values often has a large effect on the quality of results obtained in subsequent analyses. Our system addresses this problem by automatically constructing novel algorithms that are specifically tailored to the chosen task. These methods are composed of operations that have been effective in existing techniques, as identified using a new conceptual framework for analyzing BLR methods. By providing a fully customized method, our system allows spectroscopists to be confident that the baseline correction being applied is right for their application while requiring minimal effort on their part.

In future work, we intend to use this framework to analyze the relationship between the structure of the BLR method and the quality of results obtained using the corrected spectra. For example, one question of particular interest is whether there are systematic differences in the structure of methods that are optimal for different types of spectra, such as Raman, FTIR, or LIBS spectra. Answers to such questions may identify previously unknown characteristics of different types of spectra, inform future experimental design in laboratory, and provide valuable insights to guide development of algorithms for particular applications.

Acknowledgments: Work supported by NSF grant CHE-1306133 and NASA grants NNX12AK84G and NNX15AC82G.