Alignment Using Variable Penalty Dynamic Time Warping

8 **AUTHORS**, INCLUDING:

Glenn Stone
Western Sydney University

43 PUBLICATIONS **771** CITATIONS

Ivan Montoliu
Nestlé S.A.

36 PUBLICATIONS **608** CITATIONS

François-Pierre J. Martin
Nestlé S.A.

77 PUBLICATIONS **2,072** CITATIONS

Sunil Kochhar
Nestlé S.A.

119 PUBLICATIONS **4,710** CITATIONS

Alignment Using Variable Penalty Dynamic Time Warping

David Clifford,*,† Glenn Stone,† Ivan Montoliu,‡ Serge Rezzi,‡ François-Pierre Martin,‡ Philippe Guy,‡ Stephen Bruce,‡ and Sunil Kochhar‡

CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW 1670, Australia, and Metabonomics and Biomarkers Group, BioAnalytical Science Department, Nestec Ltd., Vers-Chez-Les-Blanc, 1000 Lausanne 26, Switzerland

In this article we highlight a novel variation on dynamic time warping (DTW) for aligning chromatogram signals. We are interested in sets of signals that can be aligned well locally, but not globally, by shifting individual signals in time. This kind of alignment is often sufficient for aligning gas chromatography data. Regular DTW often "over-warps" signals and introduces artificial features into the aligned data. To overcome this we introduce a variable penalty into the DTW process. The penalty is added to the distance metric whenever a nondiagonal step is taken. We select our penalty based on a morphological dilation of the two signals. We showcase our method by aligning GC/MS datafiles from 712 blood plasma samples processed in 23 batches over the course of 6 months. The use of variable penalty DTW significantly reduces the number of nondiagonal moves. In the examples presented here, this reduction is by a factor of 30, with no cost to visual quality of the alignment.

Dynamic time warping (DTW) is a fast and efficient method for alignment of signals. DTW was developed for speech recognition[1,2] and has been successfully applied to chromatograms.[3] DTW is easy to implement and memory efficient variations such as Sakoe−Chiba DTW exist for long signals.[4] DTW alters the original pair of signals by selectively contracting and expanding the time axes to minimize the distance between them. This process produces well aligned signals but often changes the shape and magnitude of peaks in such a way that the resulting signals can appear artificial. These artificial features occur when excessive numbers of nondiagonal moves are taken during the warping. We introduce the concept of a variable penalty to DTW as a way of reducing the number of nondiagonal moves.

We have used this novel variation on DTW to generate a good alignment of over 700 GC/MS datafiles. The background information for this data is given, and the data are used in two examples at the end of this paper. We summarize alignment methods currently used and some of the shortcomings of these methods. Throughout this paper we use GC/MS data from the NuGenOb project as our example. The NuGenOb project is an EU initiated project (Project Number QLK1-CT-2000-00618) designed to elucidate the role, in human obesity, of interactions between macronutrient compositions of the diet with particular emphasis on fat intake. In our examples we use the Euclidean distance metric and Sakoe−Chiba DTW. Our novel method can be incorporated into other forms of DTW such as Itakura Parallelogram dynamic time warping.[5] Our application uses an asymmetric alignment and can be extended to the symmetric alignment case.

Example Data: NuGenOb GC/MS Dataset. A total of 143 individuals provided samples for analysis by GC/MS as part of the NuGenOb Study on Obesity. The NuGenOb Study is an EU initiated project (Project Number QLK1-CT-2000-00618) designed to elucidate the role, in human obesity, of interactions between macronutrient compositions of the diet with particular emphasis on fat intake. The goal of the analysis of the NuGenOb GC/MS data set is to find differences in the quantities of metabolites present in the samples that are predictive of different subgroups in the data.

Blood plasma samples were collected from each individual over a time period of several hours. In total, the NuGenOb data contains 712 files of GC/MS data. These samples were processed in 23 batches over the course of 6 months.

Figure 1 shows the chromatograms for a single batch over a small range: clearly there is no need to align these signals. The quality of the alignment in this figure is representative of the quality of alignment over other parts of the elution profile for this batch and within other batches.

Figure 2 shows the average chromatograms from three different batches: a delay between signals is obvious suggesting that these chromatograms need to be aligned (we have only included three of the batches to avoid cluttering the image). Figure 2 also indicates that a simple shift of the time axes could achieve a good alignment over this portion of the elution profile.

Figure 3 shows the master chromatogram to which each of the batch averages will be aligned. While simple shifts within subsets of the elution profile would achieve a good alignment to the master, a shift of the whole elution profile will not work globally. The magnitude of shift needed at early elution times is much smaller than the magnitude of shift needed at later elution times, see Figures

---

* Corresponding author. E-mail: David.Clifford@csiro.au.
† CSIRO Mathematical and Information Sciences.
‡ Nestec Ltd.

(1) Vintsyuk, T. K. *Kibernetika* **1968**, *4*, 81–88.
(2) Sakoe, H.; Chiba, S. *Proceedings of the International Congress on Acoustics*, Budapest, Hungary, 1971; paper 20 c 13.
(3) Wang, C. P.; Isenhour, T. L. *Anal. Chem.* **1987**, *59*, 649–654.
(4) Sakoe, H.; Chiba, S. *IEEE Trans. Acoust., Speech, Signal Process.* **1978**, *26* (1), 43–49.
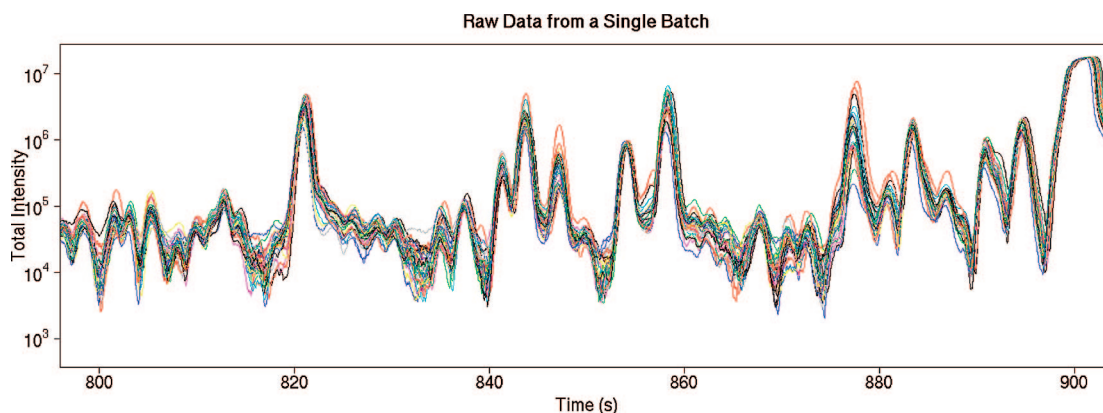(5) Itakura, F. *IEEE Trans. Acoust., Speech, Signal Process.* **1975**, *23* (1), 67–72.

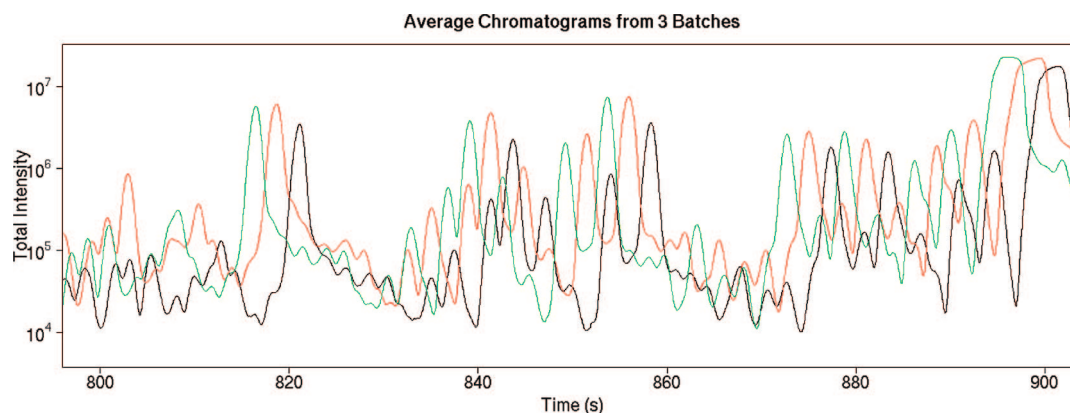**Figure 1.** Chromatograms from all data files for a single batch over a subset of the elution period.

Figure 2. Average chromatograms from three different batches plotted over a subset of the elution period. No background correction is needed between the different batches.
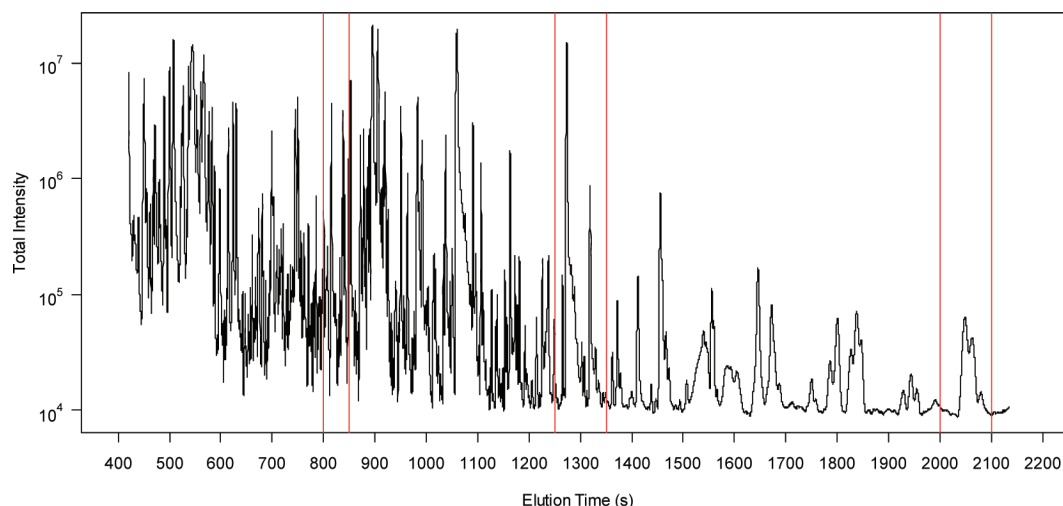

Figure 3. This chromatogram is the master, to which we will align the other signals. The length of the chromatograms are about 34 000 intensity values. Vertical red lines indicate three subsets of the elution period that will be examined more closely in subsequent images.

5−7. Also, rescaling the time axis will not align all of the peaks (details not shown). Therefore, a nonlinear alignment is needed to achieve a satisfactory global fit. The aim of this paper is to describe the alignment of this data and our novel approach to achieving this alignment using variable penalty DTW.

Alignment Algorithms. Penalized DTW has been used in various applications, including detection of low-rate TCP attacks in Internet traffic[6] and alignment of somatosensory evoked potentials in neurology.[7] In both of these applications a constant

penalty is added to the optimization function when a nondiagonal move is taken by DTW. More advanced optimization functions using signal derivatives are considered in other applications as a way of improving the alignment results.[8] Adaptively weighted

(6) Sun, H.; Lui, J. C. S.; Yau, D. K. Y. *Comput. Networks* **2006**, *50*, 2312–2330.

(7) Roberts, K.; Lawrence, P.; Eisen, A.; Hoirch, M. *IEEE Trans. Biomed. Eng.* **1987**, *34* (6), 397–405.

(8) Keogh, E.; Pazzani, M. *First SIAM International Conference on Data Mining*, Chicago, IL, 2001.
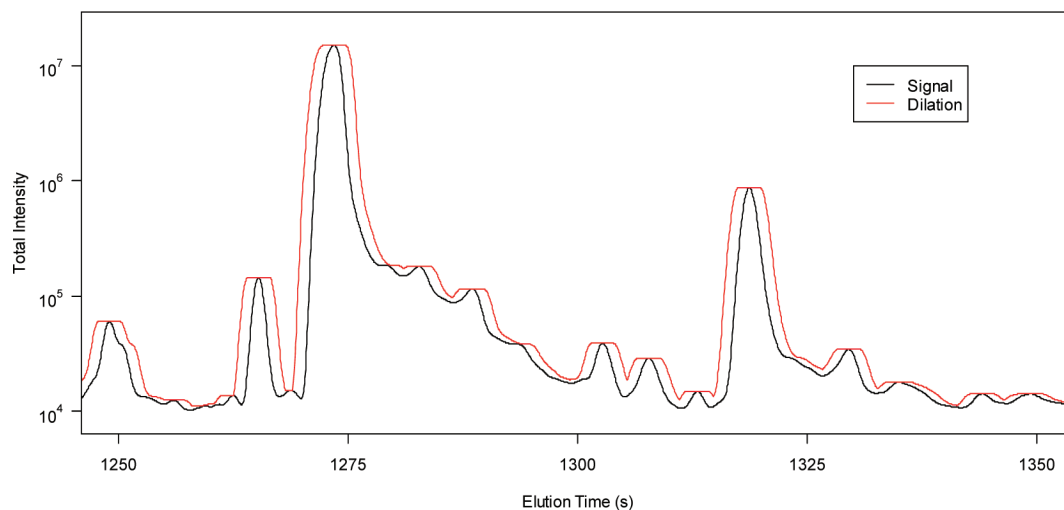
Figure 4. This image shows the morphological dilation of the master signal over a subset of the elution profile.
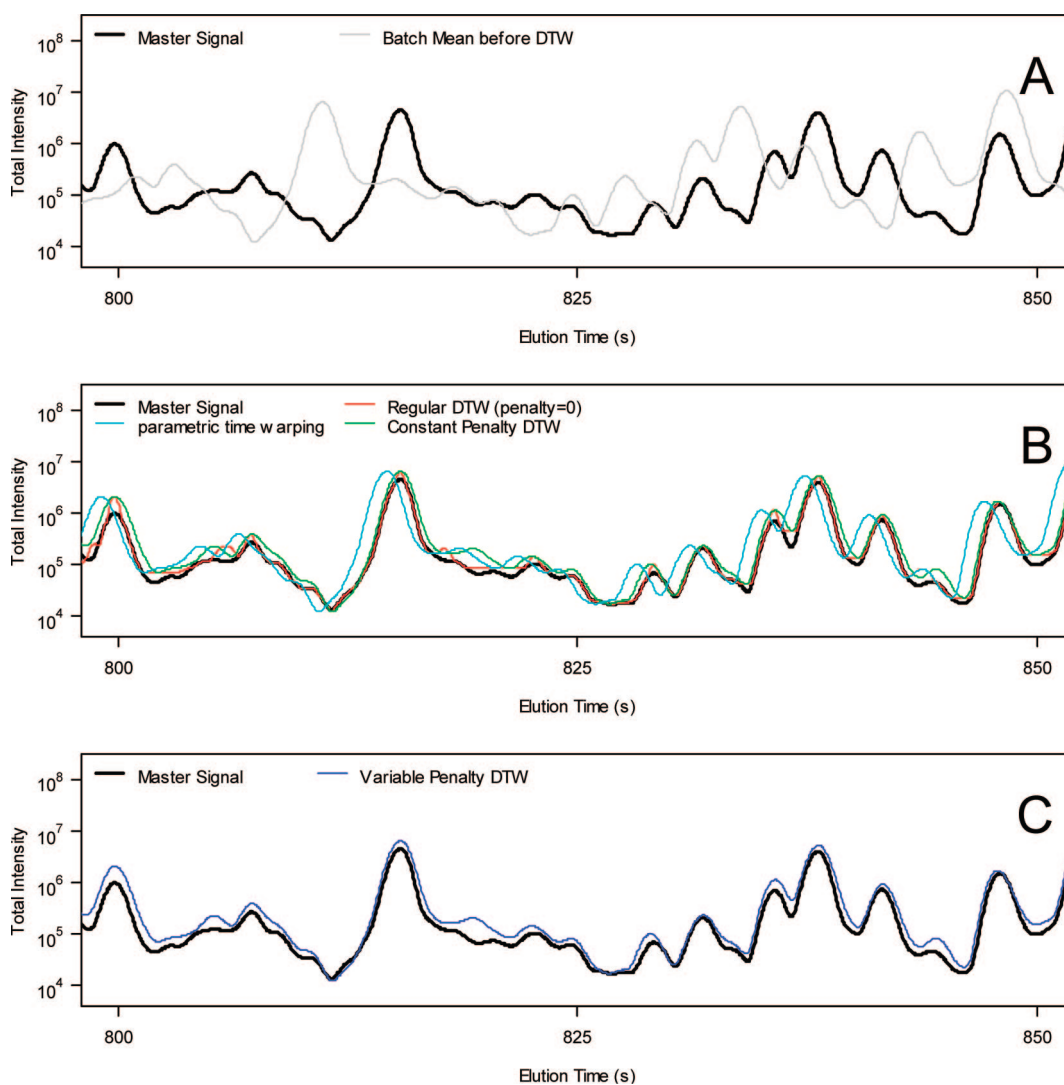

Figure 5. The results of alignment over the elution period between 800 and 850 s. Panel A shows the signal and the master. Panel B shows the master together with the results of parametric time warping, regular DTW, and constant penalty DTW. Panel C shows the master with the result of variable penalty DTW.

DTW is used to align short melody roots from Hungarian folk songs.[9] We introduce a variable penalty into the DTW process, added to the distance metric whenever a nondiagonal move is made. Adding a variable penalty to the optimization function is similar to rescaling the optimization function by adaptive weights when the later is transformed using logs, but our

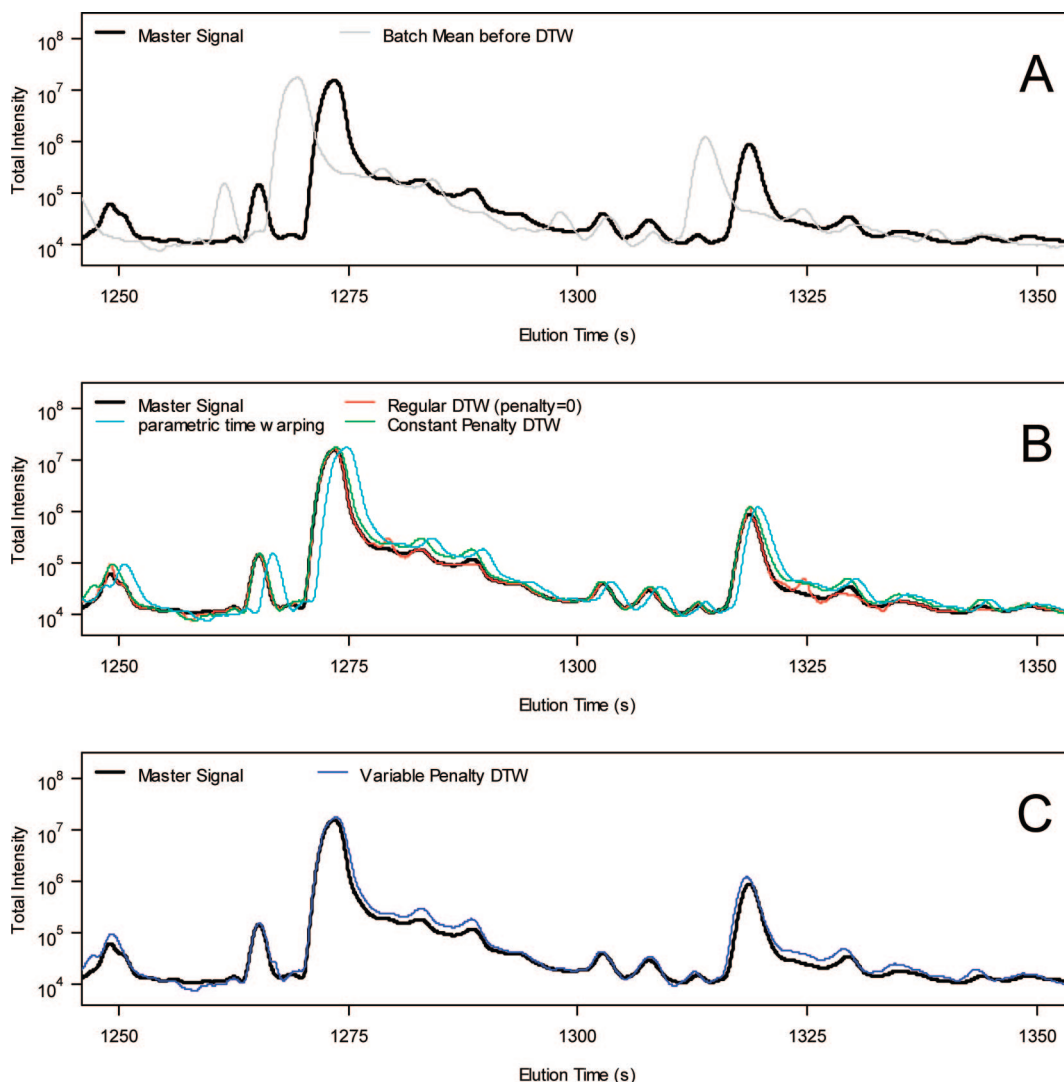(9) Juhász, Z. *Appl. Artif. Intell.* **2007**, *21*, 35–55.

Figure 6. The results of alignment over the elution period between 1250 and 1350 s. Panel A shows the signal and the master. Panel B shows the master together with the results of parametric time warping, regular DTW, and constant penalty DTW. Panel C shows the master with the result of variable penalty DTW.

approach adds the variable penalty only when nondiagonal moves are taken. Also, this variable penalty DTW is shown here with an application to chromatograms. These signals are much longer and have a larger dynamic range compared with the melody roots used by Juhász.[9]

Alignment algorithms for GC/MS and LC/MS data follow two different approaches. The first is based on finding peaks in the signals, then aligning these peaks.[10,11] The second approach performs alignment using the full signal. Our solution to the alignment problem is also based on the full signal.

Two recent algorithms that are based on the full signal include AMSRPM[12] and parametric time warping.[13] AMSRPM (robust point matching) uses smooth monotone spline regression[14] in its published software, but the algorithm is very computationally demanding. Other versions of AMSRPM using monotone kernel regression[15] have reduced the computational burden,[16] but these methods are not yet fully developed or publically available.

AMSRPM works for signals of length 2000 or so, but for longer signals the algorithm is extremely slow. We have found that AMSRPM exhibits poor symmetry (the result of aligning signal A to signal B can look very different from the result of aligning signal B to signal A). For signals of length 5000 units or more, the computation times were huge. The NuGenOb data discussed here involves chromatograms 7 times as long.

Parametric time warping is another algorithm based on the full signal. Matlab code to perform parametric time warping was provided by Paul Eilers.[13] Alignment using this method is extremely fast, but for our application the smaller features are not properly aligned.

Details of Variable Penalty DTW. Variable Penalty DTW is a novel variation on DTW with a user defined penalty added to nondiagonal moves made during alignment. The magnitude of the penalty is not necessarily constant, and we advocate larger penalties when making nondiagonal moves near peaks with high intensity values. Nondiagonal moves are penalized to reduce the number of artificial features in the aligned signals.

(10) America, A. H. P.; Cordewener, J. H. G.; van Geffen, M. H. A.; Lommen, A.; Vissers, J. P. C.; Bino, R. J.; Hall, R. D. *Proteomics* **2006**, *6* (2), 641–653.

(11) Robinson, M. D.; De Souza, D. P.; Wai Keen, W.; Saunders, E. C.; McConville, M. J.; Speed, T. P.; Likić, V. A. *BMC Bioinf.* **2007**, *8*, 419.

(12) Kirchner, M.; Saussen, B.; Steen, H.; Steen, J. A. J.; Hamprecht, F. A. *J. Stat. Software* **2007**, *18*, 4.
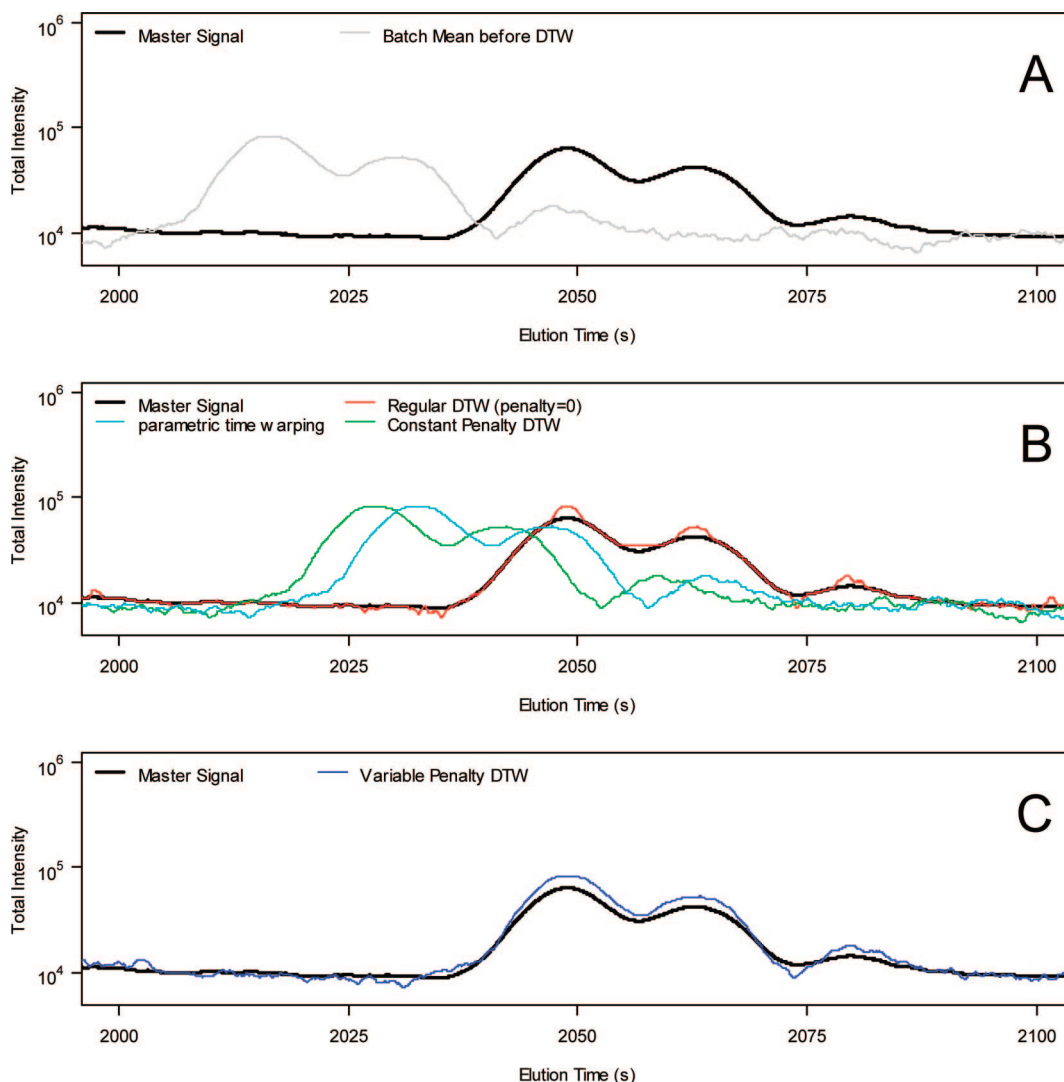
Figure 7. The results of alignment over the elution period between 2000 and 2100 s. Panel A shows the signal and the master. Panel B shows the master together with the results of parametric time warping, regular DTW, and constant penalty DTW. Panel C shows the master with the result of variable penalty DTW.

Using the same notation and set up as by Eilers,[13] suppose we are given two signals $x(t_i)$ and $y(t_i)$ for $i = 1 \ldots n$. Suppose also that y is a master signal and we wish to warp the time axis $t_i \rightarrow w(t_i)$ for signal x to minimize the distance between $y(t_i)$ and $x(w(t_i))$, thereby aligning signal x to signal y.

Variable penalty DTW requires a vector of non-negative penalties λ_i. The function we minimize by dynamic programming is

$$\sum_{i=1}^{n} (y(t_i) - x(w(t_i)))^2 + \sum_{i=2}^{n} \lambda_i \mathrm{I}(i\text{th move is nondiagonal})$$

$$= \sum_{i=1}^{n} (y(t_i) - x(w(t_i)))^2 + \sum_{i=2}^{n} \lambda_i \mathrm{I}(w(t_i) - w(t_{i-1}) = t_i - t_{i-1}) \quad (1)$$

The ith move is a diagonal move if the distance between t_i and t_{i-1} is the same as the distance between $w(t_i)$ and $w(t_{i-1})$. Otherwise the move is a nondiagonal move, and a penalty of magnitude λ_i is incurred. For signals defined at equally spaced time points, if we have aligned $y(t_{i-1})$ with $x(t_{i-1})$, then the ith move is defined to be a diagonal move if $y(t_i)$ is aligned with $x(t_i)$. Setting $\lambda_i = 0$ for each i gives us the usual unpenalized DTW algorithm. Variable penalty DTW is easy to implement as it involves a minor alteration of the optimization function used in DTW. We have chosen to use the Euclidean metric in our optimization function above. The L_1 distance can be used at a marginally higher computational cost, but this approach will be less sensitive to outliers in intensity.

We have examined several forms for the penalty vector. Our observation is that constant penalties $\lambda_i = \lambda$ for each value of i do not work well for chromatogram data in which the magnitudes of peaks change dramatically over the elution profile. In such a situation, it is difficult to balance our desire to align the features with our wish to reduce the number of artificial features in the aligned signal. We have found that a

(13) Eilers, P. H. C. *Anal. Chem.* **2004**, *76*, 404–411.
(14) Ramsay, J. O.; Silverman B. W. *Functional Data Analysis*; Springer-Verlag: New York, 2001.
(15) Pelckmans, K.; Espinoza, M.; De Moor Brabanter, J.; Suykens, J. A. K.; De, B. *Neur. Process. Lett.* **2005**, *22* (2), 171–182.
(16) Saussen, B. Retention Time Domain Registration of LCMS Data, Ph.D. Thesis, University of Heidelberg, Medical Informatics, 2007.
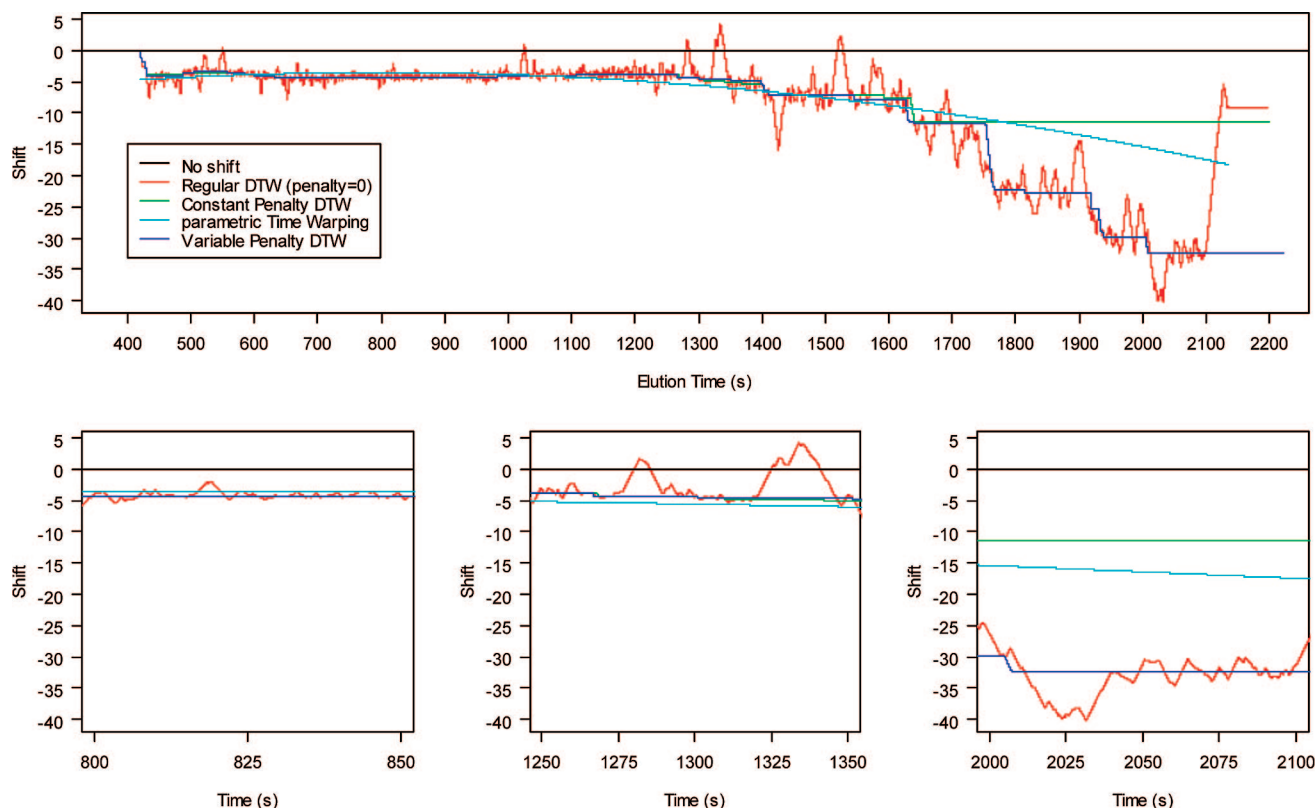
Figure 8. Shifts incurred by the four alignments over the full elution profile. Nondiagonal moves taken during alignment correspond to nonhorizontal moves in this image. A total of 60% of moves during regular DTW are nondiagonal compared to 2% of moves taken during variable penalty DTW.

penalty vector based on a morphological dilation of the signal works very well. Such a penalty changes as the magnitude of the peaks change. Thus we have called this approach a variable penalty DTW algorithm.

A morphological dilation of a signal is a moving local maximum within a window of specified width.[17] Figure 4 illustrates what the dilation of the master chromatogram looks like for a window width of 50 observations. The dilation is always greater than or equal to the original signal. Our penalty vector is the dilation of the master chromatogram appropriately scaled. The scaling factor is chosen to balance the need to make some diagonal moves with our desire to penalize unnecessary nondiagonal moves. This kind of penalty is only appropriate if the signal is always positive. Note we do not mention a baseline correction in the description of the method here but one can always be used in conjunction with this alignment method. We have not mentioned it explicitly because in our application below a baseline correction was not needed; however, it may be required in other applications. If the baseline corrected signal is negative in places, then choosing the penalty using a dilation will not be appropriate (as it will lead to negative penalties).

Examples with Two GC/MS Chromatograms. We introduce our method by performing asymmetric alignment of data from the NuGenOb data set. We align the average chromatogram of one batch (the signal) to the average chromatogram of another batch (the master). We align the signal to the master using four

different approaches, and these are compared and contrasted here. Each signal has on the order of 34 000 observed intensities. The master is plotted in Figure 3, and three regions are marked by vertical lines. These subsections of the data are shown in greater detail in Figures 5−7. In each of these figures, panel A plots both the signal and the master. In each of these figures, panels B and C show the results of the different alignments. While we focus our attention on each of these figures in what follows, the alignments are performed on the full signals and not separately on each of these subsections of the data.

Our first alignment is with a constant zero penalty, i.e., $\lambda_i = 0$ for each i. This is the usual asymmetric DTW, and the results of this are plotted in red. A total of 60% of the moves made in this alignment are nondiagonal moves. Our second alignment uses a constant penalty of $\lambda_i = 3 \times 10^5$. The results of this alignment are plotted in green. Our third alignment is done using parametric time warping, and the results are plotted in light blue. Our final alignment uses a variable penalty found by rescaling the dilation of the signal, we used $\lambda = \text{dilation}(x, \text{width} = 1500)/10\,000$. Our dilation is based on a window of width 1500 observations; a width of 1500 observations was also used in the Sakoe−Chiba DTW code. The results of this alignment are plotted in blue. Only 2% of moves for this alignment are nondiagonal moves.

We examine Figures 5−7 together to get a sense for how well the alignments have performed. Regular DTW shows many artificial features in the data including flat regions of the signal

(17) Soille, P. *Morphological Image Analysis: Principles and Applications*; Springer: New York, 1999.
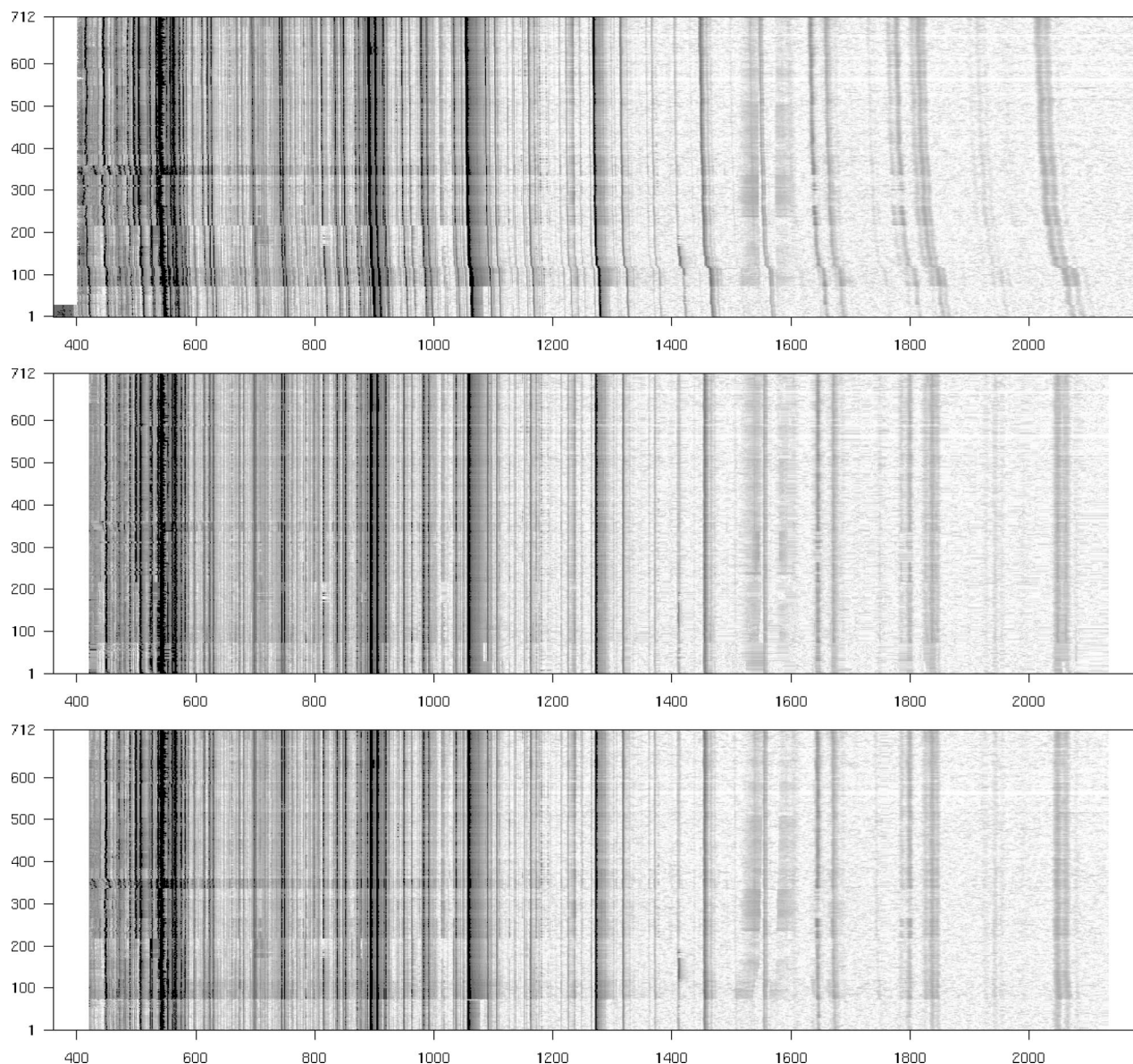
Figure 9. A total of 712 chromatograms: before alignment, after the usual DTW alignment, and after alignment using variable penalty DTW.

(Figure 5) and sudden un-natural looking features (see Figure 7 for an example where we have peaks on top of peaks instead of individual smooth peaks). Constant penalty DTW looks good in all but Figure 7 where the peaks fail to align with the peaks of the master signal. The blue signal achieves good alignment of all features in these plots.

The parametric time warping method is much faster than any of the dynamic programming methods mentioned here. However, the alignment is not as accurate as the alignment achieved using the dynamic time warping methods. The differences are minor in Figures 5 and 6. In Figure 7 we see that the parametric time warping is better than the constant penalty dynamic time warping but not as good as variable penalty dynamic time warping.

Our main criticism of the usual DTW alignment (plotted in red) is that in many regions of the plots there is no difference between the red and black signals, see Figure 6. This is worrying because we are looking for differences between subgroups of the data. If our alignment method removes many of these differences (where a shift of the signals does not) in the first stage of data

processing, then there is no way to recover these differences at a later stage and subsequently investigate what metabolites are responsible for such differences.

The performance of the constant penalty DTW may well be improved by optimizing the choice of the penalty. This alignment does a good job compared with the regular DTW in that there are less of the artificial features where the peak intensities are very high, but we see that the alignment fails for the peaks at 2050 s in Figure 7. The alignment fails in this region of the data because our penalty to move these features closer together is too high relative to the size of these peaks. Lowering the penalty will improve the performance of the alignment in this region of the data; however doing so could increase the number of artificial features at earlier elution times where peak intensities are high. The variable penalty DTW achieves our goal of reducing the number of artificial features, while also achieving good alignment across the full elution profile. This alignment is achieved with only a fraction of the nondiagonal moves needed for the usual DTW.

Figure 8 shows how far each point is shifted by each of our alignment methods. We see that our variable penalty DTW results in a warping function that involves local shifts of the signal. These local shifts are chosen to closely follow the warping function from regular DTW. Most points are shifted exactly the same as neighboring points indicating that a local shift in signal is used to align the signal to the master. This was our goal initially, and this results in a very strong alignment of signal to master.

Alignment of All NuGenOb GC/MS Chromatograms. We now return to our original problem, how to align all of the 712 chromatograms to the master chromatogram. We tackle this problem in exactly the manner outlined above. We compute the alignment on the log scale to account for the massive range in intensity values one finds with mass spectrometry data, so for us, $y(t)$ and $x(t)$ are the log intensity values at elution time t. For each batch of GC/MS datafiles, we compute the average log chromatogram across all of the datafiles and perform the usual DTW alignment as well as the variable penalty DTW alignment as described above. The warping incurred in this step is then applied to each of the original datafiles in that batch. This process is repeated across each of the batches. This results in 712 aligned GC/MS datafiles and chromatograms.

Figure 9 shows all the chromatograms stacked beside each other as an image, before alignment, after the usual DTW alignment, and after variable penalty DTW alignment. In this image, each row of pixels corresponds to a single chromatogram, with the total ion counts indicated by the shade of the pixels; lighter shades of gray correspond to lower intensity values.

The vertical bands (curved in the first panel and straight in the second and third panel) show where major peaks lie across all 712 samples. The fact that these bands are curved in the first panel indicates that a smooth drift of the GC/MS column occurs over time. The fact that these bands are straight in the second and third panel indicates that the alignments have been successful, across the full elution profile for all 712 datafiles.

DISCUSSION

This paper presents an extension of dynamic time warping, where nondiagonal steps in the optimal alignment path are penalized. This kind of warping encourages local shifts in signal. We use a penalty based on a dilation of the signals to allow more shifts to occur in regions where the peak intensities are lower. We show that variable penalty DTW achieves good alignment of peaks in GC/MS chromatograms, where the peaks heights can span several orders of magnitude.

There are extensions that can be made to this method of the alignment algorithm. Our alignment is based on the chromatogram values only. We could extend this in several ways to align signals based on the derivative of the signal or to include information from the mass spectrometry dimension to ensure two peaks are aligned if the metabolites that elute have similar m/z fingerprints.

We have illustrated our method using a large data set of GC/MS data on the log-intensity scale. The percentage of nondiagonal moves taken during the usual DTW range from 52% of moves to 70% of moves. The percentage of nondiagonal moves taken during variable penalty DTW range from 1% to 5%. Variable penalty DTW is carried out at no additional computational cost compared to the usual DTW alignment. The end result of both approaches is alignment of peaks across the full elution profile. The benefit of the variable penalty DTW lies with the fact that the reduction in the proportion of nondiagonal moves is highly significant and leads to a similar reduction in the number of artificial features in the aligned data.