

A Differentiable Distribution-Centric Gas Chromatographic Model for Gradient-Based Optimization of Temperature Programs

Y.S. Zhigalenok*, V.A. Dubrovskiy, T.V. Kan, A.K. Galeyeva, A.A. Starodubtseva

Farabi University, Al-Farabi Ave., 71, Almaty, Kazakhstan

ARTICLE INFO

Received
28.07.2026

Received in revised form
26.08.2026

Accepted
14.09.2026

Keywords:

gas chromatography;
temperature program;
distribution-centric
model; automatic
differentiation; gradient-
based optimization;
peak resolution;
polycyclic aromatic
hydrocarbons

ABSTRACT

Blumberg's distribution-centric model, the analyte migration integral and a band-broadening layer based on the Golay equation are reformulated in a differentiable form and validated against a published simulation for the 16 EPA priority PAHs. The resolution of every peak pair is thereby an explicit differentiable function of the temperature program, and the exact gradient of any criterion built from these resolutions is returned by backpropagation through the simulator, irrespective of the number of program parameters. Separation conditions can therefore be selected purposefully, for programs of any dimensionality. The method returns the attainable front of the separation-time trade-off rather than a single program; the three forms of objective function compared here reach different regions of that front, so the criterion stays a free part of the formulation, written to suit the analysis at hand. Pairs that no program brings to baseline separation are identified from the search and excluded from the criterion. At an equal number of calls to the model, backpropagation reaches an objective value that direct search, an evolution strategy and finite differences do not attain. The attainable separation improves as the profile is allowed more nodes, and the optimizer does not reproduce the conventional shape of a few linear ramps, it smooths the heating where the separation physics rewards it. The number of ramps that current instruments accept is a restriction of their control software, and it limits the separation quality the same column could otherwise deliver.

1. Introduction

Method development in gas chromatography (GC) is a resource-intensive process – selecting the temperature program, the flow and the column traditionally requires many experimental runs. Empirical rules for choosing these conditions exist [1], but for complex multicomponent mixtures they give only a coarse approximation, and the effort of experimental tuning grows rapidly with the number of components. The scale of the task is apparent even in routine environmental analysis – the determination of the 16 EPA priority polycyclic aromatic hydrocarbons (PAHs) in air, soil and water still rests on temperature programs several tens of minutes long that are re-tuned for every column and matrix [2]. Computer simulation of analyte retention allows a substantial part of this work to be moved into calculation and reduces the consumption of reagents

and instrument time [3]. Such simulations rest on thermodynamic retention models that relate the retention factor k to the temperature T . The most convenient of these is Blumberg's distribution-centric three-parameter (K-centric) model, in which the retention of each compound on a given phase is described by three parameters – the characteristic temperature T_{char} , the characteristic thermal constant θ_{char} and the isobaric heat capacity correction ΔC_p [4]. Such parameters are collected into open databases – the best known being the database [5], which currently holds 1873 entries on 24 stationary phases and is maintained as an open repository [6]. The K-centric parameterization remains under active development: the theory of band focusing in columns with exponential retention has been extended [7], and alternative measures of column performance have been proposed in place of the plate number [8]. A considerable effort is directed

*Corresponding author: Y.S. Zhigalenok; E-mail address: yaroslav.zsv@gmail.com

at obtaining the parameters themselves at a lower experimental cost – estimating them directly from temperature-programmed runs instead of a series of isothermal measurements [9], relating the characteristic temperature to the directly measurable elution temperature [10], [11] and predicting the K-centric parameters from linear solvation energy relationships (LSER) without chromatographic measurements at all [12].

The forward problem of computing a chromatogram from known parameters and a given program is solved today. The GasChromatographySimulator.jl package numerically integrates the migration and band broadening equations and reproduces experimental retention times with a deviation of less than 1% [13, 14]. An independent implementation of the same class of models, covering thermal gradient conditions, has been reported [15], and the simulation core has recently been extended to thermally modulated comprehensive two-dimensional gas chromatography through a graph-based description of the capillary network [16]. The modelling and data-processing options available in GC×GC have been reviewed separately [17]. Other tools provide similar functionality. Prediction of retention times from thermodynamic parameters back-calculated from temperature-programmed runs is implemented in the free web tool GC Retention Predictor [18]. The ProEZGC web application (Restek) models chromatograms and selects the column and separation conditions [19]. A common limitation of this class of tools is that they rely on extensive proprietary libraries of experimental retention data and enumerate a discrete set of typical conditions; optimization of the temperature program as a continuous object is not provided for.

A parallel line of work replaces the thermodynamic description with a data-driven prediction of retention. Deep neural networks trained on retention-index libraries predict Kovats indices from molecular structure with a mean absolute error of a few tens of index units [20, 21], and subsequent refinements have raised the accuracy attainable for isothermal indices [22]. A multimodal model combining a graph network with a recurrent encoder of the heating profile predicts the retention time directly for an arbitrary temperature program and on that basis ranks candidate conditions for the separation of isomers [23]. Models of this class extend the coverage to analytes whose thermodynamic parameters have not been measured, but they are trained rather than derived, they extrapolate poorly beyond the conditions represented in the training set, and

the search for conditions with them still amounts to an enumeration over a prescribed set of candidates.

Of practical interest, however, is the inverse problem – the search for a temperature program that is optimal with respect to separation and analysis time. Existing approaches rely on gradient-free and heuristic strategies – analytical constructions based on the separability ranges of adjacent pairs [24], retention prediction followed by simplex optimization of the program [25], and design of experiments. A similar problem in liquid chromatography motivated a move to closed-loop Bayesian optimization [26], but that approach too treats the simulator as a black box and does not exploit its internal structure. That line of work has since developed rapidly – evolutionary algorithms have been applied to one- and two-dimensional gradient separations [27], interpretive method development has been automated end to end on the basis of retention models fitted to a few scouting runs [28, 29], Bayesian optimization has been extended to LC×LC in a multi-task formulation [30] and taken as far as fully operator-free control of the instrument [31]. A systematic comparison of optimizers for gradient design showed that no gradient-free algorithm dominates the rest and that the number of experiments, or of model evaluations, remains the limiting resource [32]. The approaches listed share a high computational cost: all of them require a large number of forward runs of the model and scale poorly as the dimensionality of the program grows (several ramps and holds).

A question common to all these schemes is the form of the objective function. In chromatography the pairwise resolutions are customarily aggregated into a single chromatographic response function, and the manner of aggregation – the worst pair, a product, a sum of saturating terms, the weight given to the analysis time – determines which solution the optimizer will regard as the best; the weighting factor itself may have to be adapted in the course of the optimization [33].

In a number of adjacent fields, reverse-mode automatic differentiation through the simulator itself has become standard. If the forward model is composed of smooth operations, backpropagation makes it possible to compute the exact gradient of the objective function with respect to all parameters in a time of the order of a single forward pass, irrespective of their number [34], [35]. Differentiable physics and optimal control [36], as well as inverse design in nanophotonics [37], are built on this. Over the past five years the same principle has spread widely through the physical sciences – it underlies

hybrid physics-informed models [38], the differentiable solvers used to accelerate computational fluid dynamics [39], and differentiable quantum chemistry packages, in which molecular properties are obtained as derivatives of a differentiable self-consistent field procedure [40]. These works form the methodological basis of the present study. In separation science the approach has so far been used in the opposite direction: a differentiable model of solute migration was employed to recover the K-centric retention parameters from temperature-programmed runs, that is, to fit the model to an experiment [9]. To the optimization of the gas chromatographic temperature program itself – the design variable of the separation – reverse-mode differentiation has not, to our knowledge, been applied before.

In the present work the K-centric retention model, the migration integral and a layer that computes the width of the eluting band are implemented in the differentiable computational environment PyTorch. The resolution of each peak pair becomes a differentiable function of the temperature program, and the program itself becomes an object of gradient-based optimization with exact, automatically computed gradients. On open data (16 EPA priority PAHs, Rxi-17SilMS phase) it is shown that the forward model reproduces the published simulation; that the form of the objective function determines the shape of the attainable time-resolution set, the hinge function with an explicit target resolution providing the most reproducible optimum; that which pairs of the mixture can be separated at all is best decided from the result of the search rather than assumed in advance; that backpropagation requires substantially fewer calls to the model than gradient-free methods; and that the freedom given to the shape of the temperature profile has a measurable value, which the differentiable formulation makes available at practically no computational cost.

2. Experimental Part

2.1. Retention model and analyte migration along the column

Retention in GC is governed by the distribution of the analyte between the mobile (gas) and stationary (liquid film) phases, characterized by the distribution coefficient K . It is related to the retention factor k through the phase ratio of the column β :

$$K = \beta k \quad (1)$$

The temperature dependence of K through the Gibbs free energy $\Delta G = \Delta H - T\Delta S$ leads to the clas-

sical van't Hoff model; accounting for the temperature dependence of ΔH and ΔS adds a third parameter – the change in isobaric heat capacity ΔC_p – and gives the three-parameter Clarke–Glew model [41]:

$$\ln K = A + \frac{B}{T} + C \ln \left(\frac{T}{T_1} \right), \quad T_1 = 1K \quad (2)$$

An equivalent but chromatographically more meaningful parameterization is Blumberg's K-centric model [4]:

$$\ln K = \left(\frac{\Delta C_p}{R} + \frac{T_{\text{char}}}{\theta_{\text{char}}} \right) \left(\frac{T_{\text{char}}}{T} - 1 \right) + \frac{\Delta C_p}{R} \ln \left(\frac{T}{T_{\text{char}}} \right) \quad (3)$$

where: R – the universal gas constant. The parameters have a direct physical meaning. T_{char} – the temperature at which $\ln k = 0$ ($k = 1$), that is, the analyte is equally distributed between the phases; it sets the elution order. θ_{char} – the inverse slope of the dependence $\ln k(T)$ at the point $T = T_{\text{char}}$: raising the temperature by θ_{char} lowers k approximately by a factor of $e \approx 2.72$. ΔC_p describes the deviation from the two-parameter model far from T_{char} .

In programmed mode the column temperature $T(t)$ varies in time. The motion of the analyte is described by the fraction of the column travelled $z(t) \in [0, 1]$, a dimensionless coordinate that equals 0 at the inlet and 1 at the outlet, with the velocity:

$$\frac{dz}{dt} = \frac{1}{t_M(T(t))[1 + k(T(t))]} \quad (4)$$

and the moment of elution t_R is defined by the elution condition $z(t_R) = 1$:

$$\int_0^{t_R} \frac{dt}{t_M(T(t))[1 + k(T(t))]} = 1 \quad (5)$$

The hold-up time $t_M(T)$ depends on the column geometry (length L , internal diameter d) and on the carrier gas viscosity $\eta(T)$ [5]:

$$t_M = \frac{128}{3} \cdot \frac{L^2}{d^2} \cdot \eta(T) \cdot \frac{p_i^3 - p_o^3}{(p_i^2 - p_o^2)^2} \quad (6)$$

where: p_i and p_o are the inlet and outlet pressures. In constant mass flow mode (as in the source database) the controller maintains $(p_i^2 - p_o^2)/\eta(T) = \text{const}$, that is, the inlet pressure rises with temperature:

$$p_i(T) = \sqrt{p_o^2 + \frac{\eta(T)}{\eta(T_{\text{ref}})} (p_{i,\text{ref}}^2 - p_o^2)} \quad (7)$$

where: T_{ref} – the temperature at which the inlet pressure is set – the initial temperature of the program, 70 °C – and $p_{i,\text{ref}}$ is the inlet pressure at that temperature; the pair of values fixes the mass flow that the controller then maintains. The viscosity of helium is

approximated by the power law $\eta(T) = \eta_0 (T/273.15 \text{ K})^{0.685}$ over the range 300–700 K [3].

2.2. Automatic differentiation

Any numerical program, including a chromatogram simulator, is a composition of elementary operations (addition, multiplication, exponential, logarithm, and so on) whose derivatives are known exactly. The derivative of such a composition with respect to an input quantity is computed by the rule for differentiating a composite function (the chain rule). For a two-link composition $y = f(u)$, where $u = g(x)$, it gives:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \quad (8)$$

that is, the derivative of a composition equals the product of the derivatives of the steps that make it up. A chromatogram simulator is a chain of many such operations, so its derivative with respect to any input quantity is the product of the derivatives of all intermediate steps. Automatic differentiation (auto-diff) evaluates this product numerically and returns the derivatives of the result with respect to the input quantities without symbolic manipulation and without the approximation error inherent in finite differences [34].

The distinction between the two modes of auto-diff, corresponding to the two orders of multiplying the factors in the chain rule, is essential. In forward mode the derivatives are accumulated in the order of computation, from input to output: a single pass yields the derivatives of all outputs with respect to one input, so the gradient with respect to N parameters requires N passes. In reverse mode (backpropagation) the same factors are multiplied in the reverse order, from output to inputs: first an ordinary forward computation is carried out with the intermediate quantities stored, and then the derivatives are propagated from the result backwards along the chain of operations, so that a single backward pass yields the derivatives of one scalar output with respect to all inputs at once. For optimization this is the optimal configuration – there is one output, the value of the objective function, and many inputs, the temperature values at the profile nodes, and the full gradient is computed in a single backward pass, at a cost of the order of one forward chromatogram calculation and independently of the number of parameters N [34]. A finite-difference estimate of the same gradient perturbs each parameter separately; with central differences, which are the ones used in

the numerical comparison of section 3.4, this costs $2N$ calls to the model per gradient and introduces an approximation error in addition.

The applicability of reverse-mode differentiation requires that the forward model be built from operations that possess a derivative at every point, so that a small change of the input produces a proportional, well-defined change of the output. In the present implementation the calculation of $k(T)$ by Eq. (3) and of $t_m(T)$ by Eq. (6), the band-broadening layer of section 2.3, the numerical integration of the elution condition Eq. (5) by the trapezoidal rule on a fine time grid, and the determination of t_R by linear interpolation of the point where $z = 1$ holds, are all carried out with such operations.

The objective functions of section 2.6 and the upper temperature limit of section 2.5 are a separate matter. The function $\max(0, \cdot)$ has a kink at zero and clipping has a kink at the boundary, so the objective, as a function of the program parameters, is piecewise smooth rather than smooth. The set of points at which its derivative is undefined has measure zero; at such a point the framework returns one of the one-sided derivatives, that is, a subgradient – for $\max(0, x)$ at $x = 0$ it returns zero – and Adam uses it exactly as it uses a gradient. This is the standard situation in gradient-based training with rectifier nonlinearities [42].

2.3. Band broadening and the width of the eluting peak

Two narrow peaks standing close together can be separated better than two broad peaks standing further apart, so the retention time alone does not describe a separation. The width of a peak follows from the dispersion that the analyte band accumulates while travelling along the column.

The local plate height H – the length of column over which the band acquires one unit of spatial variance – is given by the Golay equation for an open tubular column [43]:

$$H = \frac{2D_g}{u} + \frac{1+6k+11k^2}{24(1+k)^2} \frac{d^2}{D_g} u + \frac{2}{3} \frac{k}{(1+k)^2} \frac{d_f^2}{D_s} u \quad (9)$$

The three terms are, in order, longitudinal diffusion in the gas, resistance to mass transfer in the gas and resistance to mass transfer in the stationary film; u is the local velocity of the carrier gas, D_g and D_s the diffusion coefficients of the analyte in the gas and in the film, d the internal diameter of the column and d_f the thickness of the stationary film. The retention

factor k enters through Eq. (3), so H depends on the temperature and hence on time.

The gas-phase diffusion coefficient is estimated from the correlation of Fuller, Ensley and Giddings [44]:

$$D_g = \frac{1.43 \cdot 10^{-3} T^{1.75}}{p M_{AB}^{\frac{1}{2}} \left(\Sigma v_A^{\frac{1}{3}} + \Sigma v_B^{\frac{1}{3}} \right)^2} \quad (10)$$

where: $M_{AB} = 2/(1/M_A + 1/M_B)$, M_A and M_B are the molar masses of the analyte and of the carrier gas, and Σv are the diffusion volumes, obtained for the analytes by summing the atomic increments of Fuller, Ensley and Giddings – 15.9 per carbon atom and 2.31 per hydrogen atom, with a correction of -18.3 per aromatic ring – and taken as 2.67 for helium [44]. Eq. (10) is dimensional: with T in K, p in bar and the molar masses in g/mol it returns D_g in cm^2/s . Because the carrier gas is compressible, the pressure falls along the column as $p(x) = [p_i^2 - (p_i^2 - p_o^2)x/L]^{1/2}$, so that u and D_g depend on the position x as well as on the temperature. The local velocity of the carrier gas follows from the same profile, and the integral of $1/u$ along the column reproduces the hold-up time of Eq. (6). Because H is evaluated locally, with the local velocity and the local diffusion coefficient, and the variance is integrated along the column, the decompression of the carrier gas is accounted for directly.

Eq. (11): the denominator of the second term is set as " ∂ " with x as a subscript; it must be " ∂_x " run in the spatial scale:

$$\frac{d\sigma_x^2}{dt} = H u_b + 2\sigma_x^2 \frac{\partial u_b}{\partial x}, \quad u_b = \frac{u}{1+k} \quad (11)$$

where: u_b is the migration velocity of the center of the band. The first term generates dispersion; the second rescales the band as it moves into a region where the migration velocity differs. In a conventional oven the temperature is uniform along the column, so at any instant k is the same everywhere and the derivative arises from the decompression of the carrier gas alone: the term is then positive throughout the run and the band is stretched as it accelerates towards the outlet. At the moment of elution the spatial width is converted into the time scale:

$$\sigma_t = \frac{\sigma_x}{u_b} \Big|_{x=L} \quad (12)$$

The spatial form is used because the equivalent equation written in the time scale is numerically unstable while the analyte is still standing at the column inlet with k of the order of 10^9 . The band is tak-

en to enter the column without initial width, $\sigma_x(0) = 0$, so the widths computed here are the contribution of the column alone and contain no injection or extra-column dispersion.

The film thickness follows from the phase ratio quoted with the retention parameters, $d_f = d/(4\beta) = 0.25 \mu\text{m}$ for $\beta = 250$ and $d = 0.25 \text{ mm}$. The diffusion coefficient in the stationary film, D_s , is not tabulated for these analytes and is taken as $D_g/10^4$, which is its usual order of magnitude. With a $0.25 \mu\text{m}$ film the stationary-phase term contributes 0.17% of the accumulated variance, against 78% for longitudinal diffusion and 16% for mass transfer in the gas; and changing D_g/D_s over two orders of magnitude, from 10^3 to 10^5 , moves the median σ_t from 1.057 to 1.065 s, that is, by less than 1%.

2.4. Measures of separation

The separation of a pair of peaks i and j is characterized by the resolution

$$R_{s,i,j} = \frac{|t_{R,i} - t_{R,j}|}{2(\sigma_i + \sigma_j)} \quad (13)$$

which, with the width at the base taken as $w = 4\sigma$, is the usual $R_s = 2|\Delta t_R|/(w_i + w_j)$. A value of $R_s = 1.5$ corresponds to baseline separation of two Gaussian peaks of comparable height and is used below as the threshold of a successfully separated pair. The resolution is evaluated for all $C(16, 2) = 120$ pairs, because the elution order is not fixed in advance. The widths σ_i and σ_j entering Eq. (13) are the time-scale standard deviations of Eq. (12), taken for each compound at its own elution time.

Resolution is not the only measure of separation, and the alternatives differ in what they describe. The selectivity, or separation factor, $\alpha_{ij} = k_j/k_i$, is the ratio of the retention factors of two compounds at a given temperature; it carries no information about the peak widths and characterizes the pair on a given stationary phase, which is why it is the quantity reported when phases or columns are compared [3]. The separation number and the peak capacity count how many peaks a run can accommodate, and so describe the run as a whole rather than a particular pair. Aggregate chromatographic response functions combine the pairwise resolutions of a chromatogram into a single number and are what automated method-development schemes minimize. The manner of the aggregation is itself a design decision, and the weight given to the analysis time may have to be adapted in the course of the optimization [33]. Here the phase and the column are fixed and the

design variable is the temperature program, so a measure that characterizes the phase does not bear on the object being optimized, and a measure that describes the run as a whole cannot express a requirement placed on one pair.

2.5. Parameterization of the temperature profile

The temperature program is specified by the values of T at K nodes equally spaced over the time horizon $t_{\max} = 52$ min, with linear interpolation between them. The first node is the initial temperature of the program, $T_{\text{init}} = 70$ °C, and the $N = K - 1$ increments between successive nodes are the free parameters. The main calculations use $K = 300$ nodes, that is, $N = 299$ parameters; the effect of K is examined separately in section 3.6.

Two physical constraints are built into the parameterization. The oven in programmed GC is not cooled, so the temperature must not decrease; and the heating rate cannot exceed a limit $R_{\max}$, taken here as 40 °C/min. Both hold by construction if the increment between the nodes m and $m + 1$ is written as

$$\Delta T_m = \text{sigmoid}(\theta_m) R_{\max} \Delta \tau, \quad \text{sigmoid}(\theta) = \frac{1}{1 + e^{-\theta}}, \quad \Delta \tau = \frac{t_{\max}}{K-1} \quad (14)$$

where: θ_m are unconstrained real parameters, which are what the optimizer actually varies, and $\Delta \tau$ is the spacing of the nodes in time. The sigmoid maps the whole real line into the interval (0, 1): a large negative θ_m gives a zero increment, that is, an isothermal hold, a large positive one gives heating at the limiting rate, and no value of θ_m can produce a negative increment or a rate above $R_{\max}$. The upper temperature limit $T \leq T_{\max} = 360$ °C, set by the column, is imposed by clipping.

2.6. Objective functions

All three objective functions act on the vector of resolutions $R_{s,i,j}$ of the pairs belonging to a set P and are supplemented by a penalty on the total analysis time, λt_{last} , where t_{last} is the elution time of the last peak and λ is a weighting coefficient. A penalty is a term that grows for an undesirable outcome, here for a long analysis, and therefore steers the optimizer away from such solutions.

Eq. (15): the minus sign has been lost from softmin. It must read $\text{smin}_a(x) = -(1/a) \ln \sum e^{-ax_i}$ of a set of numbers x_i are

$$s \max_b(x) = \frac{1}{b} \ln \sum_i e^{bx_i}, \quad s \min_a(x) = \frac{1}{a} \ln \sum_i e^{-ax_i} \quad (15)$$

The first expression is the log-sum-exp function. It is almost equal to the largest of the numbers, but, unlike the ordinary maximum, it has a derivative with respect to every one of them at once, so that the numbers which are not currently the largest still receive a gradient. The second is the same construction for the minimum, and is what is meant below by softmin. Both are used with a moderate sharpness ($a = 6$, $b = 3$), and t_{last} is computed as the smooth maximum of the retention times.

The first objective function maximizes the resolution of the worst pair:

$$L_{\min} = -\text{smin}_{(i,j) \in P} R_{s,i,j} + \lambda t_{\text{last}} \quad (16)$$

Its value is determined by a single critical pair, and improving any other pair does not change it.

The second sums saturating rewards for the resolution of each pair:

$$L_{\text{sat}} = - \sum_{(i,j) \in P} s \tanh\left(\frac{R_{s,i,j}}{s}\right) + \lambda t_{\text{last}} \quad (17)$$

The contribution of a pair grows almost linearly while $R_{s,i,j} \ll s$ and saturates for $R_{s,i,j} \geq s$ (here $s = 1$), so this criterion raises the resolution of all pairs on average and is not tied to any particular target.

The third is a piecewise-linear penalty, or hinge function; hinge is the name adopted in machine learning for the function $\max(0, \cdot)$:

$$L_{\text{hinge}} = \sum_{(i,j) \in P} \max(0, R^* - R_{s,i,j}) + \lambda t_{\text{last}} \quad (18)$$

The penalty is switched on only for a pair that has fallen short of the target resolution R^* , and it equals the shortfall. A pair that has reached the target drops out of the sum, and once all pairs have reached it the separation part of the loss vanishes – function Eq. (18) does not demand excess separation and does not spend analysis time on it. Functions Eq. (16) and Eq. (17) have no target resolution, so the set of solutions they can attain is traced by varying the weight λ ; for function Eq. (18) the set is traced by varying R^* at a fixed $\lambda = 0.02$.

2.7. Optimization algorithms

Minimization is performed by the Adam method [45] – gradient descent with an adaptive step – using the exact gradient obtained by backpropagation through the model; this combination is denoted below as autodiff-Adam. The learning rate is 0.1 in all the simulations reported below; its choice is examined in section 3.4.

Three reference methods are used for comparison, all of them applied to exactly the same problem, from the same initial profile and within the same

budget of calls to the forward model. FD-Adam is the same Adam algorithm with the gradient estimated by central finite differences with a step of 10^{-3} , that is, at a cost of $2N = 598$ calls per gradient at $N = 299$. Nelder–Mead [46] is a direct search method that maintains a simplex of $N + 1$ trial points in parameter space and deforms it – reflecting, expanding and contracting it away from the worst vertex – without using derivatives at all. CMA-ES [47], the covariance matrix adaptation evolution strategy, draws a population of trial profiles from a multivariate normal distribution; after each generation it shifts the mean of that distribution towards the better-performing samples and updates its covariance matrix, thereby building an anisotropic model of the local landscape and, in effect, learning which directions of the 299-dimensional space are worth exploring. It uses no derivatives. Methods of these two classes, direct search and evolution strategies, are precisely the ones in current use for the optimization of chromatographic separations [27, 32].

The cost of all methods is counted in calls to the forward model. The cost of one autodiff-Adam step is charged at the standard upper bound of the cheap-gradient result: reverse-mode differentiation returns the gradient with respect to all parameters at a cost of at most about 4 evaluations of the function, independently of their number [48].

Because the objective functions turn out to be

non-convex, the optimization is launched independently from many initial profiles, a procedure known as multistart – a linear ramp with small random perturbations, and fully random profiles.

2.8. Data, implementation and validation of the forward model

The demonstration mixture is the 16 EPA priority PAHs on the Rxi-17SiIMS phase, with a phase ratio $\beta = 250$ and hence a film thickness $d_f = d/(4\beta) = 0.25 \mu\text{m}$, taken from the validated part of the database [5]; the K-centric parameters of the compounds are given in Table 1. The column is 30.222 m long with an internal diameter of 0.25 mm, the carrier gas is helium and the outlet is at atmospheric pressure. The reference program is the set of conditions from the same work [5]: 70 °C; ramp 20 °C/min to 150 °C, hold 5 min; 12 °C/min to 250 °C, hold 2 min; 15 °C/min to 360 °C, hold 5 min.

Since the forward model was rewritten in a new software environment, its correctness was verified independently. The published simulation of the GC separation of the 16 PAHs [5] was digitized and used as a control set of retention times. The inlet pressure is not reported in the source work, so the free parameter p_i was determined by one-dimensional numerical minimization of the root mean square deviation of the calculated times from the control ones.

Table 1. K-centric retention parameters of the 16 EPA priority PAHs on the Rxi-17SiIMS phase ($\beta = 250$) according to [5]; the compounds are ordered by increasing T_{char} , which coincides with the elution order

No.	Compound	CAS	T_{char} , °C	θ_{char} , °C	ΔC_p , J·mol ⁻¹ ·K ⁻¹
1	Naphthalene	91-20-3	157.31	35.06	57.2
2	Acenaphthylene	208-96-8	204.51	38.03	48.3
3	Acenaphthene	83-32-9	207.59	37.89	51.1
4	Fluorene	86-73-7	220.75	38.33	55.3
5	Phenanthrene	85-01-8	251.47	40.76	64.6
6	Anthracene	120-12-7	252.13	40.36	55.0
7	Fluoranthene	206-44-0	286.09	42.41	52.1
8	Pyrene	129-00-0	295.10	43.43	48.8
9	Benz[a]anthracene	56-55-3	326.45	43.82	53.9
10	Chrysene	218-01-9	329.02	44.10	54.3
11	Benzo[b]fluoranthene	205-99-2	356.39	46.28	78.2
12	Benzo[k]fluoranthene	207-08-9	356.85	46.19	78.2
13	Benzo[a]pyrene	50-32-8	368.27	47.91	81.5
14	Dibenzo[a,h]anthracene	53-70-3	391.69	48.82	100.8
15	Indeno[1,2,3-cd]pyrene	193-39-5	393.38	49.57	89.2
16	Benzo[g,h,i]perylene	191-24-2	399.69	48.82	69.7

3. Results and Discussion

3.1. Inlet pressure calibration and validation of the forward model

The forward model contains a quantity that is not specified in the data source – the inlet pressure p_i . Meanwhile p_i sets the scale of the hold-up time $t_M(T)$ (Eq. (6)) and therefore shifts all calculated retention times proportionally. For this reason p_i was determined by minimizing the root mean square deviation of the calculated retention times from the digitized ones. The calibrated quantity enters the model as a common scale factor of t_M and cannot change either the elution order or the relative positions of the peaks, which are set by the retention factor $k(T)$. Consequently, the agreement in the mutual arrangement of the whole set of t_R remains an independent test of the retention model rather than a consequence of fitting.

The dependence of the root mean square deviation on the inlet pressure has a single minimum over the interval studied, 1.4–4.0 bar (Fig. 1A), attained at $p_i = 2.184$ bar, RMSE = 0.255 min. The residuals at the calibrated value (Fig. 1B) do not exceed 30 s in absolute value, yet they are distributed in a regular way: eleven peaks elute earlier in the calculation than in the publication (by 1–26 s) and five later (by 5–30 s), the largest lag, 21–30 s, falling on the three last heavy PAHs. This means that the discrepancy is not fully removed by the calibrated parameter; the likely sources of the residual systematic deviation are the error of digitizing the source figure, the difference between the viscosity models – here $\eta(T)$ is a power law with a fixed exponent, whereas [13] uses a correlation in which the exponent itself depends

on temperature, and unknown details of the flow mode and of the conditions at the column outlet.

At the calibrated pressure the calculated retention times reproduce the published simulation with a correlation coefficient $r = 0.9998$ (Fig. 1c); the mean relative deviation is 1.37%. The maximum relative deviation (6.96%) falls on the first, least retained peak (naphthalene, $t_R \approx 6$ min): at a small retention time an absolute deviation of the order of 0.4 min corresponds to a large relative value.

3.2. Effect of the form of the objective function on the attainable time–resolution set

The three objective functions of section 2.6 were compared on the full set of 120 pairs. For each of them a sweep was performed over its own control parameter – λ for functions Eq. (16) and Eq. (17), R^* for function Eq. (18) – with 24 independent random starts at every point of the sweep. Every simulation is an independent point in Fig. 2; the line drawn through each cloud is the set of its non-dominated points, that is, the front of what the given form of the objective function was able to attain – a shorter analysis at the same worst-case resolution, or a higher resolution at the same analysis time. The optimization itself is carried out on all 120 pairs; the vertical axis reports the worst resolution among the pairs classified as attainable in section 3.3, a classification applied after the fact and identically to all three forms, so that it affects only the way the result is displayed.

Function Eq. (16) is determined by a single pair, the worst one, so the remaining 119 pairs generate no gradient at all and the optimizer receives infor-

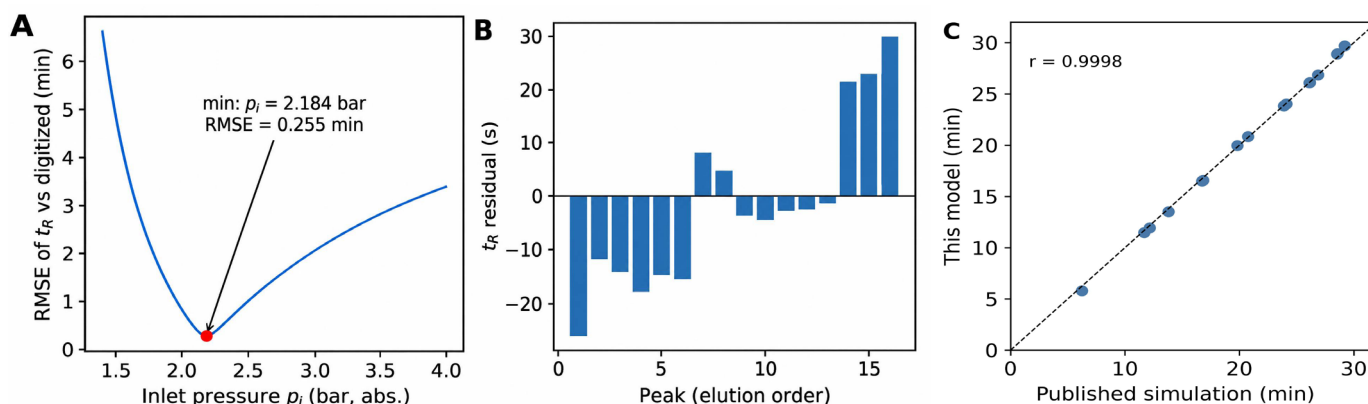


Fig. 1. Calibration of the inlet pressure. (A) Dependence of the root mean square deviation of the calculated retention times from the digitized simulation [5] on the inlet pressure; the minimum corresponds to $p_i = 2.184$ bar. (B) Residuals of t_R by peak at the calibrated pressure; the numbering follows the elution order (Table 1). (C) Comparison of the calculated and published [5] retention times after calibration of the inlet pressure ($r = 0.9998$); the dashed line denotes perfect agreement.

mation about one pair at a time. Function Eq. (17) rewards total separation, so it continues to separate pairs that are already well separated and pays analysis time for it. Function Eq. (18) penalizes only the shortfall to the target and stops as soon as the target is reached, which is why its front extends furthest and why the operating point along the front is set directly, by the value of R^* , rather than indirectly through a weight.

The spread of a cloud at a fixed value of the control parameter measures the reproducibility of the result. If the loss surface were convex, gradient descent would arrive at the same point from any start and the spread would be zero. It is not: at $\lambda = 0.02$ the worst-pair resolution attained by function Eq. (16) is 2.237 with a coefficient of variation of 3.0% over 24 simulations, function Eq. (17) gives 2.133 with 3.9%, and function Eq. (18) at $R^* = 2.5$ gives 2.404 with 2.5%. All three are thus reproducible to within a few per cent at a fixed setting, and the choice between them is decided by the extent of the front rather than by the spread of the cloud: soft-min occupies the shortest interval of analysis times, saturating lies below the other two over most of the range, and hinge spans the whole range and reaches the highest resolutions on it. All the results that follow were obtained with the hinge function.

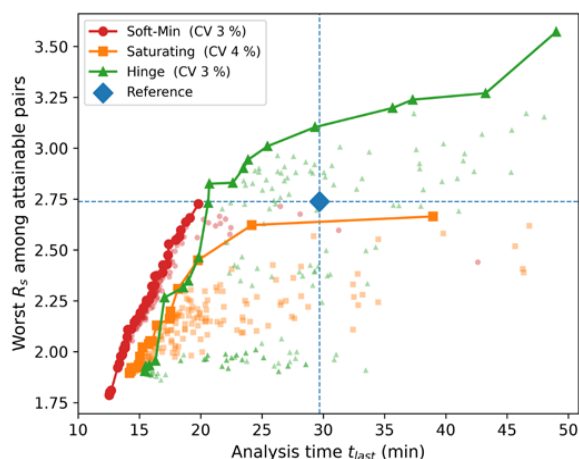


Fig. 2. Attainable time-resolution sets of the three objective functions on the full set of 120 pairs. Each faint marker is one independent simulation (24 random starts at every point of the sweep, 504 simulations in total); the solid lines connect the non-dominated points of each cloud. The sweep is over λ for soft-min Eq. (16) and saturating Eq. (17) and over R^* for hinge Eq. (18) at $\lambda = 0.02$. The vertical axis is the resolution of the worst pair among those classified as attainable in section 3.3. The blue diamond and the dashed lines mark the reference program.

3.3. Determination of the separable pairs

Not every pair of the sixteen compounds can be brought to baseline separation on this column. The resolution attainable for a pair is bounded from above by the selectivity of the phase towards the two compounds: if the phase barely distinguishes them, no temperature program will separate them within a reasonable time. Such a pair behaves destructively inside the objective function – it takes on the main part of the penalty and does not respond to it, so that its contribution to the gradient acts as a constant bias rather than as a signal, while the optimizer pays analysis time in an attempt to move it.

Which pairs these are is decided here from the result of the search. A pair is called attainable if at least one of the 504 programs found in the sweep of section 3.2 brought it to $R_s \geq 1.5$.

Of the 120 pairs, 117 are attainable and three are not: phenanthrene/anthracene, benzo[b]fluoranthene/benzo[k]fluoranthene and indeno[1,2,3-cd]pyrene/dibenzo[a,h]anthracene. Figure 3A shows the five most difficult pairs, every simulation contributing a point.

How much of this is a property of the phase and how much a limitation of the search was checked separately, by an isothermal scan that involves no optimizer at all: for each of the three pairs the resolution was evaluated on a grid of constant temperatures. The scan gives an upper estimate of 1.47 at 170 °C for phenanthrene/anthracene, which agrees with the 1.49 found by the search; 1.42 at 192 °C for benzo[b]fluoranthene/benzo[k]fluoranthene, but at a retention time of 335 min; and 3.50 at 222 °C for indeno[1,2,3-cd]pyrene/dibenzo[a,h]anthracene, at a retention time of 318 min. Both maxima lie beyond five hours of elution, so the separation of these two pairs is not forbidden by the phase but is unreachable at any practical analysis time.

Figure 3B compares the fronts obtained with all 120 pairs in the objective function and with the 117 attainable ones, at the same budget and from the same random starts. The front built on 117 pairs lies above the other over the short-time part of the range: the three immovable pairs otherwise take on a constant share of the penalty, and the analysis time extended in an attempt to resolve them yields no gain in resolution. In that pass none of the excluded pairs rose to the threshold, which confirms the classification. In all subsequent calculations these three pairs are excluded from the objective function.

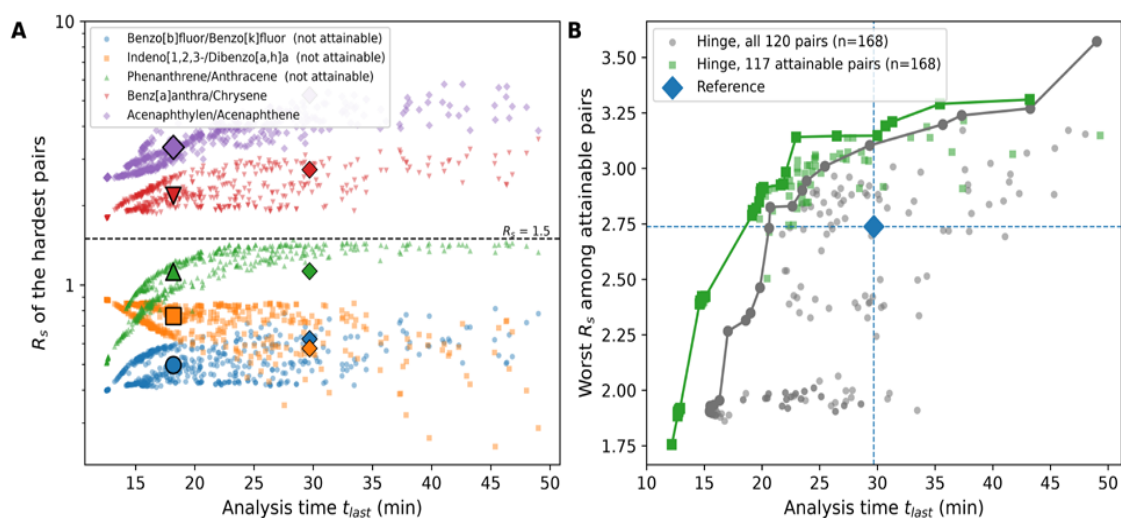


Fig. 3. Classification of the pairs by attainable resolution. (A) The five most difficult pairs of the mixture; each small marker is the resolution of that pair in one simulation of the sweep of Fig. 2, the large marker is the median over the simulations and the diamond is the value in the reference program. (B) The same hinge optimization performed twice with identical budgets and identical random starts, differing in the set of pairs entering the objective function. Markers are individual simulations, lines are the non-dominated fronts, the diamond and the dashed lines mark the reference program.

3.4. Computational cost of the optimization

Figure 4A compares the convergence of the four algorithms against the number of calls to the forward model at $N = 299$ free parameters. All of them solve the same problem – the hinge function at $R^* = 3.0$ over the 117 attainable pairs – from the same initial profile and within the same budget of 10 000 calls.

Within the budget of 10 000 calls, the final value of CMA-ES is reached by autodiff-Adam after 2 172 calls, 4.6 times fewer than CMA-ES itself needed; the final value of Nelder–Mead after 396 calls, 25 times fewer; and that of FD-Adam after 4 calls, three orders of magnitude fewer. The reason lies in the scaling. A central-difference gradient costs $2N = 598$ calls, so FD-Adam completes about seventeen steps within the budget, and the Nelder–Mead simplex of $N + 1 = 300$ vertices is deformed a few tens of times, whereas the cost of backpropagation does not grow with the number of parameters. Over the first few thousand calls CMA-ES descends faster than autodiff-Adam and is overtaken at about two thousand, so at small budgets it remains competitive.

The gradient sets the direction of the fastest decrease of the objective function, but not the length of the step along it; the latter is set by the learning rate – the coefficient that determines how far the optimizer moves in parameter space at each iteration. In the Adam algorithm the step is additionally normalized by the characteristic scale of the gradi-

ent, so the learning rate sets approximately the step size itself in parameter space rather than scaling together with the steepness of the landscape.

Figure 4B shows 20 simulations from random initial profiles for each of seven learning rates. The mean final value falls from 0.603 at a rate of 0.01 to 0.450 at 0.2 and rises again to 0.582 at 0.5, while the standard deviation grows over the whole range from 0.018 to 0.138. A larger step carries the trajectory out of the basin of attraction in which it started, which lowers the mean and at the same time widens the spread; at 0.5 the descent no longer settles within 1000 steps. All other simulations in this work used a rate of 0.1.

3.5. Optimized programs

Figure 5 shows three programs taken from the front of the hinge optimization, together with the chromatograms they produce; the bottom row is the reference program.

The optimizer knows nothing about the reference program: it does not start from it, is not steered towards it and does not have it in the objective function. It searches the space of profiles for the arrangement of the peaks that its own criterion – the resolution of the worst attainable pair, weighed against the analysis time – considers best. Consequently, the result is not an improved version of the reference program: a program that raises the worst pair may well

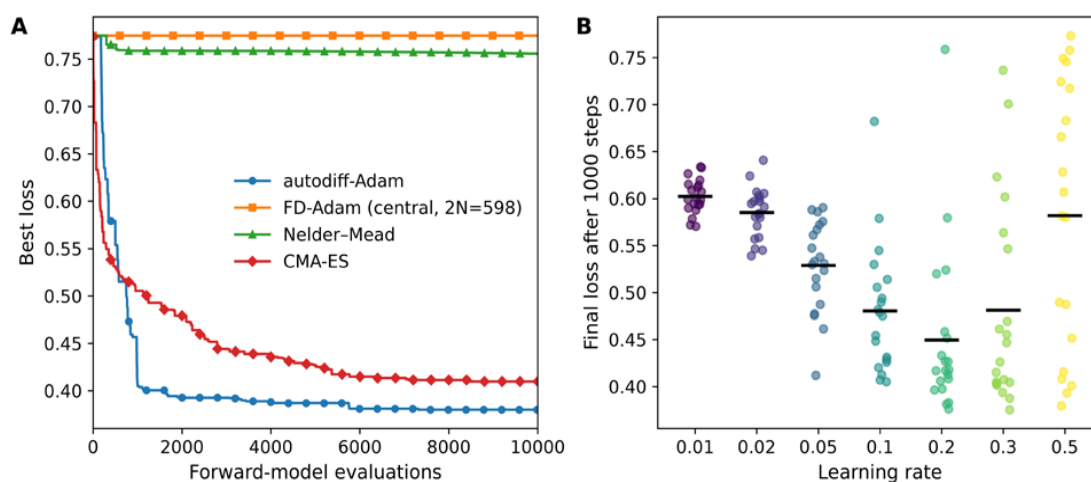


Fig. 4. Cost and tuning of the gradient-based optimization (hinge function, $R^* = 3.0$, $\lambda = 0.02$, 117 pairs, $N = 299$). (A) Best objective value reached against the number of calls to the forward model: autodiff-Adam, FD-Adam with a central-difference gradient costing $2N = 598$ calls per step, Nelder-Mead and CMA-ES. (B) Effect of the learning rate: 20 simulations from random initial profiles of 1000 steps each for every value; the points are the final objective value of an individual simulation and the horizontal segments are the means.

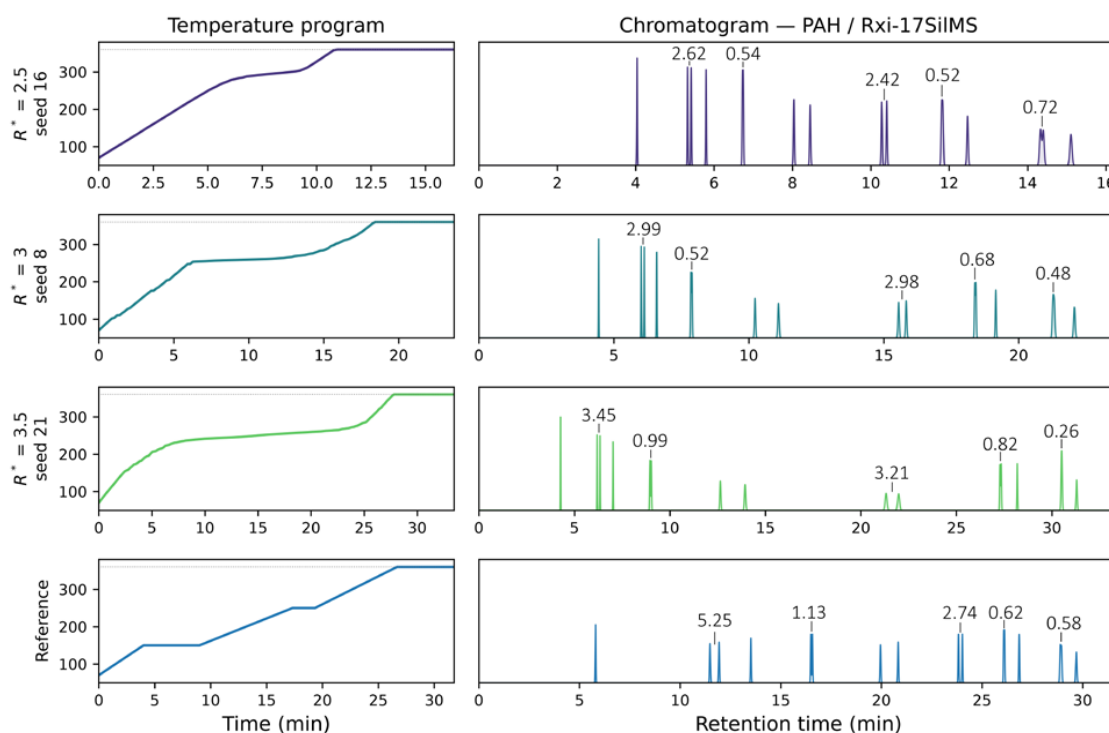


Fig. 5. Programs from the front of the hinge optimization at three values of the target resolution: temperature profiles (left) and simulated chromatograms (right); the bottom row is the reference program [5]. The peak numbering by elution order corresponds to Table 1. The peak widths are those calculated by the band-broadening layer of section 2.3; the figures above the chromatograms are the resolutions of the marked neighbouring pairs.

leave some comfortably resolved pair with less room than it had before. Only the stated criterion is optimized; but the criterion is stated explicitly and can be changed – if a particular pair is critical for a particular method, it is entered into the objective function with the required target, which is precisely what the hinge form makes possible.

The method returns a front rather than a single solution. At $R^* = 2.5$ the program completes the analysis in 15.11 min with a worst-pair resolution of 2.42; at $R^* = 3.0$, in 22.06 min with 2.98; at $R^* = 3.5$, in 31.29 min with 3.21. The reference program, a single point, likely found experimentally, gives 29.69 min and 2.74. It lies below the front, so at its analysis

time there exist programs with a better worst-case resolution, and at its resolution there exist faster ones. Concretely, the middle program shortens the analysis by 26% and at the same time raises the worst attainable pair from 2.74 to 2.98; the fast program halves the analysis time at the price of 0.3 resolution units on the worst pair; the slow one shows a resolution of 3.21 for 5% more time than the reference. Which of the three to use is a decision for the method developer. In the general case a ready reference program does not exist at all and has to be tuned manually through a series of experiments, whereas the front is obtained from the retention parameters alone.

A manually assembled program consists of a few straight ramps and one or two isothermal holds, with sharp corners at the transitions. Having 299 free increments at its disposal, the optimizer does something else: where the criterion benefits from it, it rounds the corners, replacing an abrupt change of the heating rate by a gradual one, and it does so in particular places – as a rule before the region in which the critical pairs elute. The profile nevertheless remains piecewise linear, a sequence of linear segments on a uniform grid with a step of 0.174 min.

3.6. Effect of the number of profile nodes

The number of nodes K sets the dimensionality of the search space. To establish its effect on the attainable separation, the optimization was repeated at $K = 5, 10, 20, 100$ and 300 with all other conditions held equal (Fig. 6).

The best worst-pair resolution reached anywhere on the front rises from 2.77 at $K = 5$ to 3.26 at $K = 10$, 3.31 at $K = 20$, 3.33 at $K = 100$ and 3.53 at $K = 300$, while the fastest point of the front moves from 17.1 to 12.2 min. On that scale the loss at $K = 5$ – a program of four linear segments, that is, the conventional shape – is large and unambiguous, the loss at $K = 10$ is marginal, and from $K = 20$ upwards the fronts are indistinguishable from one another.

The highest worst-pair resolution is reached at the largest number of nodes examined, $K = 300$. For the sixteen-component mixture considered here a profile of 20 nodes recovers most of that value, but the optimum itself lies at the finest profile, and for mixtures of hundreds of components, in which far more pairs compete for the same temperature interval, a correspondingly finer profile is to be expected. Current instruments accept programs of 20 to 32 ramps [49, 50]. That number is a limit of the instrument control software rather than of the thermal hardware, and the results above show that it caps the separation quality obtainable from a given column and carrier gas.

3.7. Limitations

The model requires three parameters (T_{char} , θ_{char} , ΔC_p) per compound and stationary phase, obtained from a series of isothermal measurements, so the approach is applicable where such parameters are available. This restriction is progressively being relaxed: the parameters can be estimated from temperature-programmed runs [9] or predicted from

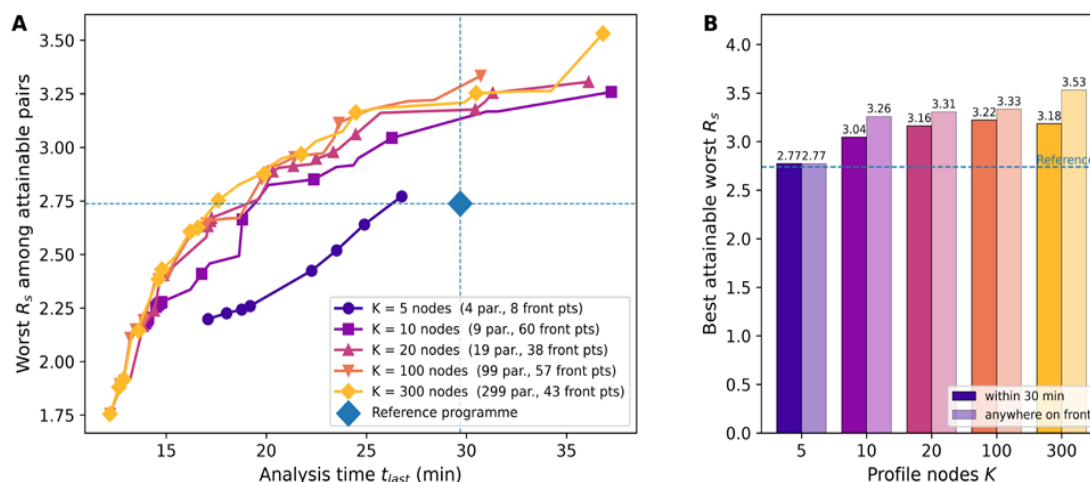


Fig. 6. Effect of the number of profile nodes on the attainable separation. (A) Non-dominated fronts obtained at 5, 10, 20, 100 and 300 profile nodes with an identical budget of simulations at every point of the sweep over R^* ; the diamond and the dashed lines mark the reference program. (B) Best attainable worst-pair resolution as a function of the number of nodes - anywhere on the front, and restricted to programs no longer than the reference program.

LSER descriptors [12], while data-driven models predict retention for compounds absent from the existing databases [22, 23]. The demonstration is limited to the composition and phases represented in the database used.

The band-broadening layer is a calculated estimate. It rests on the Golay equation and on the Fuller estimate of the gas-phase diffusion coefficient, and only the retention times were checked against published data. The contribution of the stationary phase is negligible at a film thickness of 0.25 μm , but the absolute values of R_s remain model values.

The calculation was performed for constant mass flow mode with atmospheric outlet. The inlet pressure is not reported in the source work and was calibrated against the published simulation rather than measured.

Finally, the optimum is sought in a continuous space of temperatures and under a single heating-rate limit, whereas a real instrument rounds its setpoints and its maximum heating rate depends on the temperature. The present study is accordingly a study of what the attainable set looks like when the temperature program is treated as a continuous design variable; the transfer of a particular found program to a particular instrument is a separate question.

4. Conclusion

The K-centric retention model, the equation of analyte motion along the column and a band-broadening layer based on the Golay equation have been reformulated in a differentiable computational environment and validated against a published simulation and by internal tests. The resolution of every peak pair is thereby an explicit differentiable function of the temperature program, and the exact gradient of any criterion built from these resolutions is returned by gradient backpropagation, irrespective of the number of parameters of the program. Separation conditions can therefore be selected purposefully, and for programs of practically any dimensionality, where manual or grid search becomes unmanageable.

The method returns the attainable front of the trade-off between separation and analysis time. Three forms compared here reach different regions of the front, and the piecewise-linear form with an explicit target resolution places the operating point directly, through the value of the target.

Backpropagation through the simulator outperforms the gradient-free algorithms compared with

it – direct search, an evolution strategy and finite differences – reaching, at an equal number of calls to the model, an objective value that none of them attains. The advantage grows with the number of parameters of the program, since the cost of a gradient does not depend on it, whereas for the other algorithms it does.

Finally, the attainable separation improves as the temperature profile is allowed more nodes, and the best result is obtained with the finest profile examined. Given that freedom the optimizer does not reproduce the conventional shape of a few linear ramps: it smooths the heating in the places where the physics of the separation rewards it. The number of ramps that current instruments accept is a restriction of their control software, and it limits the separation quality that the same column and the same carrier gas could otherwise deliver.

Author Contributions

Yaroslav Zhigalenok – Conceptualization; Methodology; Software, Formal Analysis, Visualization, Writing – Original Draft. **Vladislav Dubrovskiy** – Investigation, Resources, Data Curation, Writing – Review & Editing. **Tatyana Kan** – Validation, Data Curation, Writing – Review & Editing. **Alina Galeyeva** – Validation, Data Curation, Writing – Review & Editing. **Alena Starodubtseva** – Investigation, Data Curation, Writing – Review & Editing.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26197159). During the preparation of this work the authors used Claude Opus 4.8 (Anthropic) to assist with writing the Python code for the model implementation and numerical experiments. After using this tool, the authors reviewed and verified the generated code and all results it produced and take full responsibility for the content of the publication.

References

- [1]. L.M. Blumberg, M.S. Klee. Optimal heating rate in gas chromatography. *Journal of Microcolumn*

- Separations 12 (2000) 508-514. [https://doi.org/10.1002/1520-667X\(2000\)12:9<508::AID-MCS5>3.0.CO;2-Y](https://doi.org/10.1002/1520-667X(2000)12:9<508::AID-MCS5>3.0.CO;2-Y).
- [2]. M.A. Aretaki, J. Desmet, M. Viana, B.L. van Drooge. Comprehensive methodology for semi-volatile organic compound determination in ambient air with emphasis on polycyclic aromatic hydrocarbons analysis by GC–MS/MS. *Journal of Chromatography A* 1730 (2024) 465086. <https://doi.org/10.1016/j.chroma.2024.465086>.
- [3]. L.M. Blumberg. *Temperature-Programmed Gas Chromatography*. Wiley-VCH, Weinheim, 2010, 350 p. <https://doi.org/10.1002/9783527630592>.
- [4]. L.M. Blumberg. Distribution-centric 3-parameter thermodynamic models of partition gas chromatography. *Journal of Chromatography A* 1491 (2017) 159-170. <https://doi.org/10.1016/j.chroma.2017.02.047>.
- [5]. T. Brehmer, B. Duong, M. Marquart, L. Friedemann, P.J. Faust, et al. Retention Database for Prediction, Simulation, and Optimization of GC Separations. *ACS Omega* 8 (2023) 19708-19718. <https://doi.org/10.1021/acsomega.3c01348>.
- [6]. GasChromatographyToolbox. RetentionData: A compilation of retention data in gas chromatography. Available online: <https://github.com/GasChromatographyToolbox/RetentionData>.
- [7]. L.M. Blumberg. Theory of linear focusing in chromatographic columns with exponential retention. Part 1: Basic solutions. *Journal of Chromatography A* 1696 (2023) 463967. <https://doi.org/10.1016/j.chroma.2023.463967>.
- [8]. L.M. Blumberg. Chromatographic parameters: Transport efficiency—A meaningful alternative to the plate number parameter. *Journal of Chromatography A* 1731 (2024) 465177. <https://doi.org/10.1016/j.chroma.2024.465177>.
- [9]. J. Leppert, T. Brehmer, M. Wüst, P. Boeker. Estimation of retention parameters from temperature programmed gas chromatography. *Journal of Chromatography A* 1699 (2023) 464008. <https://doi.org/10.1016/j.chroma.2023.464008>.
- [10]. T. Brehmer, P. Boeker, M. Wüst, J. Leppert. Relation between characteristic temperature and elution temperature in temperature programmed gas chromatography - part I: Influence of initial temperature and heating rate. *Journal of Chromatography A* 1707 (2023) 464301. <https://doi.org/10.1016/j.chroma.2023.464301>.
- [11]. T. Brehmer, P. Boeker, M. Wüst, J. Leppert. Relation between characteristic temperature and elution temperature in temperature programmed gas chromatography – Part II: Influence of column properties. *Journal of Chromatography A* 1728 (2024) 464997. <https://doi.org/10.1016/j.chroma.2024.464997>.
- [12]. T. Brehmer, B. Duong, P. Boeker, M. Wüst and J. Leppert. Simulation of gas chromatographic separations and estimation of distribution-centric retention parameters using linear solvation energy relationships. *Journal of Chromatography A* 1717 (2024) 464665. <https://doi.org/10.1016/j.chroma.2024.464665>.
- [13]. J. Leppert. GasChromatographySimulator.jl. *Journal of Open Source Software* 7 (2022) 4565. <https://doi.org/10.21105/joss.04565>.
- [14]. J. Leppert, P.J. Müller, M.D. Chopra, L.M. Blumberg, P. Boeker. Simulation of spatial thermal gradient gas chromatography. *Journal of Chromatography A* 1620 (2020) 460985. <https://doi.org/10.1016/j.chroma.2020.460985>.
- [15]. H.D. Tolley, S. Avila, B.D. Iverson, A.R. Foster, A.R. Hawkins, et al. Simulating Capillary Gas Chromatographic Separations including Thermal Gradient Conditions. *Analytical Chemistry* 93 (2021) 2291-2298. <https://doi.org/10.1021/acs.analchem.0c04160>.
- [16]. J. Leppert, T. Brehmer, M. Wüst. Prediction of Thermal Modulated Comprehensive Two-Dimensional Gas Chromatographic Separation Using a Modular, Graph-Based Simulation Platform. *Journal of Separation Science* 49 (2026) e70351. <https://doi.org/10.1002/jssc.70351>.
- [17]. M. Gaida, P.-H. Stefanuto, J.-F. Focant. Theoretical modeling and machine learning-based data processing workflows in comprehensive two-dimensional gas chromatography - A review. *Journal of Chromatography A* 1711 (2023) 464467. <https://doi.org/10.1016/j.chroma.2023.464467>.
- [18]. P.G. Boswell, P.W. Carr, J.D. Cohen, A.D. Hegeman. Easy and accurate calculation of programmed temperature gas chromatographic retention times by back-calculation of temperature and hold-up time profiles. *Journal of Chromatography A* 1263 (2012) 179-188. <https://doi.org/10.1016/j.chroma.2012.09.048>.
- [19]. Restek Corporation. ProEZGC Chromatogram Modeler. Available online: <https://ez.restek.com/proezgc>.
- [20]. C. Qu, B.I. Schneider, A.J. Kearsley, W. Keyrouz, T.C. Allison. Predicting Kováts Retention Indices Using Graph Neural Networks. *Journal of Chromatography A* 1646 (2021) 462100. <https://doi.org/10.1016/j.chroma.2021.462100>.
- [21]. T. Vrzal, M. Malečková, J. Olšovská. DeepRel: Deep learning-based gas chromatographic retention

- indexpredictor. *Analytica Chimica Acta* 1147 (2021) 64–71. <https://doi.org/10.1016/j.aca.2020.12.043>.
- [22]. A. Anjum, J. Liigand, R. Milford, V. Gautam, D.S. Wishart. Accurate prediction of isothermal gas chromatographic Kováts retention indices. *Journal of Chromatography A* 1705 (2023) 464176. <https://doi.org/10.1016/j.chroma.2023.464176>.
- [23]. J. Lin, L. Song, Y. Chen, C. Liu, S. Chen, et al. Multimodal learning in synthetic chemistry applications: gas chromatography retention time prediction and isomer separation optimization. *Digital Discovery* 4 (2025) 2465–2477. <https://doi.org/10.1039/d4dd000369a>.
- [24]. V. Bártů, S. Wičar, G.-J. Scherpenzeel, P.A. Leclercq. Optimization of temperature programming in gas chromatography with respect to separation time: I. Temperature programme optimization fundamentals. *Journal of Chromatography A* 370 (1986) 219–234. [https://doi.org/10.1016/S0021-9673\(00\)94694-6](https://doi.org/10.1016/S0021-9673(00)94694-6).
- [25]. H. Snijders, H.-G. Janssen, C. Cramers. Optimization of temperature-programmed gas chromatographic separations I. Prediction of retention times and peak widths from retention indices. *Journal of Chromatography A* 718 (1995) 339–355. [https://doi.org/10.1016/0021-9673\(95\)00692-3](https://doi.org/10.1016/0021-9673(95)00692-3).
- [26]. J. Boelrijk, B. Ensing, P. Forré, B.W.J. Pirok. Closed-loop automatic gradient design for liquid chromatography using Bayesian optimization. *Analytica Chimica Acta* 1242 (2023) 340789. <https://doi.org/10.1016/j.aca.2023.340789>.
- [27]. B. Huygens, K. Efthymiadis, A. Nowé, G. Desmet. Application of evolutionary algorithms to optimise one- and two-dimensional gradient chromatographic separations. *Journal of Chromatography A* 1628 (2020) 461435. <https://doi.org/10.1016/j.chroma.2020.461435>.
- [28]. T.S. Bos, J. Boelrijk, S.R.A. Molenaar, B. van 't Veer, L.E. Niezen, et al. Chemometric Strategies for Fully Automated Interpretive Method Development in Liquid Chromatography. *Analytical Chemistry* 94 (2022) 16060–16068. <https://doi.org/10.1021/acs.analchem.2c03160>.
- [29]. E. Bosten, M. Pardon, K. Chen, V. Koppen, G. Van Herck, et al. Assisted Active Learning for Model-Based Method Development in Liquid Chromatography. *Analytical Chemistry* 96 (2024) 13699–13709. <https://doi.org/10.1021/acs.analchem.4c02700>.
- [30]. J. Boelrijk, S.R.A. Molenaar, T.S. Bos, T.A. Dahlseid, B. Ensing, et al. Enhancing LC×LC separations through multi-task Bayesian optimization. *Journal of Chromatography A* 1726 (2024) 464941. <https://doi.org/10.1016/j.chroma.2024.464941>.
- [31]. T.M. Dixon, J. Williams, M. Besenhard, R.M. Howard, J. MacGregor, et al. Operator-free HPLC automated method development guided by Bayesian optimization. *Digital Discovery* 3 (2024) 1591–1601. <https://doi.org/10.1039/D4DD00062E>.
- [32]. G.B. van Henten, J. Boelrijk, C. Kattenberg, T.S. Bos, B. Ensing, et al. Comparison of optimization algorithms for automated method development of gradient profiles. *Journal of Chromatography A* 1742 (2025) 465626. <https://doi.org/10.1016/j.chroma.2024.465626>.
- [33]. L.E. Niezen, G. Desmet. A new chromatographic response function with automatically adapting weight factor for automated method development. *Journal of Chromatography A* 1727 (2024) 465008. <https://doi.org/10.1016/j.chroma.2024.465008>.
- [34]. A.G. Baydin, B.A. Pearlmutter, A.A. Radul, J.M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research* 18 (2018) 1–43. <https://www.jmlr.org/papers/v18/17-468.html>.
- [35]. R.T.Q. Chen, Y. Rubanova, J. Bettencourt, D.K. Duvenaud. Neural Ordinary Differential Equations. *Advances in Neural Information Processing Systems* 31 (2018) 6571–6583. <https://doi.org/10.48550/arXiv.1806.07366>.
- [36]. Y. Hu, L. Anderson, T.-M. Li, Q. Sun, N. Carr, et al. DiffTaichi: Differentiable Programming for Physical Simulation. 8th International Conference on Learning Representations (ICLR), Addis Ababa, 2020. <https://doi.org/10.48550/arXiv.1910.00935>.
- [37]. S. Molesky, Z. Lin, A.Y. Piggott, W. Jin, J. Vučković, et al. Inverse design in nanophotonics. *Nature Photonics* 12 (2018) 659–670. <https://doi.org/10.1038/s41566-018-0246-9>.
- [38]. G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, et al. Physics-informed machine learning. *Nature Reviews Physics* 3 (2021) 422–440. <https://doi.org/10.1038/s42254-021-00314-5>.
- [39]. D. Kochkov, J.A. Smith, A. Alieva, Q. Wang, M.P. Brenner, et al. Machine learning-accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 118 (2021) e2101784118. <https://doi.org/10.1073/pnas.2101784118>.
- [40]. M.F. Kasim, S. Lehtola, S.M. Vinko. DQC: A Python program package for differentiable quantum chemistry. *The Journal of Chemical Physics* 156 (2022) 084801. <https://doi.org/10.1063/5.0076202>.
- [41]. E.C.W. Clarke, D.N. Glew. Evaluation of

- thermodynamic functions from equilibrium constants. *Transactions of the Faraday Society* 62 (1966) 539-547. <https://doi.org/10.1039/tf9666200539>.
- [42]. X. Glorot, A. Bordes and Y. Bengio. Deep sparse rectifier neural networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, 2011, pp. 315-323.
- [43]. M.J.E. Golay. Theory of chromatography in open and coated tubular columns with round and rectangular cross-sections. In: *Gas Chromatography 1958*, D.H. Desty (Ed.), Butterworths, London, 1958, pp. 36-55.
- [44]. E.N. Fuller, K. Ensley and J.C. Giddings. Diffusion of halogenated hydrocarbons in helium. The effect of structure on collision cross sections. *The Journal of Physical Chemistry* 73 (1969) 3679-3685. <https://doi.org/10.1021/j100845a020>.
- [45]. D.P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations (ICLR), San Diego, 2015. <https://doi.org/10.48550/arXiv.1412.6980>.
- [46]. J.A. Nelder, R. Mead. A Simplex Method for Function Minimization. *The Computer Journal* 7 (1965) 308-313. <https://doi.org/10.1093/comjnl/7.4.308>.
- [47]. N. Hansen, A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation* 9 (2001) 159-195. <https://doi.org/10.1162/106365601750190398>.
- [48]. C.C. Margossian. A review of automatic differentiation and its efficient implementation. *WIREs Data Mining and Knowledge Discovery* 9 (2019) e1305. <https://doi.org/10.1002/widm.1305>.
- [49]. Agilent Technologies. Agilent 8890 Gas Chromatograph. Data sheet 5994-0492EN. Available online: <https://www.agilent.com/cs/library/datasheets/public/datasheet-8890-gas-chromatograph-5994-0492en-agilent.pdf>.
- [50]. Thermo Fisher Scientific. TRACE 1300 Series Gas Chromatograph. Specifications. Available online: <https://documents.thermofisher.com/TFS-Assets/CMD/Specification-Sheets/PS-52260-TRACE-1300-GC-PS52260-EN.pdf>.

About the Authors

Y.S. Zhigalenok – PhD candidate, Junior Researcher, Farabi University, Almaty, Kazakhstan
E-mail: yaroslav.zsv@gmail.com
ORCID: <https://orcid.org/0000-0003-1452-1248>

V.A. Dubrovskiy – PhD candidate, Junior Researcher, Farabi University, Almaty, Kazakhstan
E-mail: vlad.dubrovskiy.2001@mail.ru
ORCID: <https://orcid.org/0000-0002-0925-335X>

T.V. Kan – Master of Science, Junior Researcher, Farabi University, Almaty, Kazakhstan
E-mail: tatyana.qan@gmail.com
ORCID: <https://orcid.org/0000-0002-1945-8471>

A.K. Galeyeva – Candidate of Chem. Sci., Vice-Rector for Academic Affairs, Farabi University, Almaty, Kazakhstan
E-mail: galeyeva8998@gmail.com
ORCID: <https://orcid.org/0000-0001-9303-5277>

A.A. Starodubtseva – Master of Science, Head of the Laboratory of Structural Design of Functional Materials, Farabi University, Almaty, Kazakhstan
E-mail: alena.a.rubanova@gmail.com
ORCID: <https://orcid.org/0000-0003-4344-2039>

Дифференцируемая модель газовой хроматографии, центрированная по коэффициенту распределения, для градиентной оптимизации температурных программ

Я.С. Жигаленок*, В.А. Дубровский, Т.В. Кан, А.К. Галеева, А.А. Стародубцева

Farabi University, пр. аль-Фараби, 71, Алматы, Казахстан

АННОТАЦИЯ

Центрированная по коэффициенту распределения (K-centric) модель удерживания Блумберга, интеграл миграции аналита и слой уширения полос на основе уравнения Голя переформулированы в дифференцируемом виде и валидированы относительно опубликованной симуляции для 16 приоритетных ПАУ ЕРА. Разрешение каждой пары пиков становится, тем самым, явной дифференцируемой функцией температурной программы, а точный градиент любого построенного на этих разрешениях критерия возвращается обратным распространением сквозь симулятор независимо от числа параметров программы. Условия разделения, таким образом, могут выбираться направленно и для программ любой размерности. Метод возвращает не единственную программу, а достижимый фронт компромисса «разделение – время анализа»; три сравниваемые здесь формы целевой функции выходят на разные участки этого фронта, так что критерий остается свободной частью постановки и формулируется под конкретную задачу. Пары, которые ни одна программа не выводит на базовое разделение, выявляются из самого поиска и исключаются из критерия. При равном числе обращений к модели обратное дифференцирование достигает значения целевой функции, недостижимого для прямого поиска, эволюционной стратегии и конечных разностей. Достижимое разделение улучшается по мере увеличения числа узлов профиля, причем оптимизатор не воспроизводит привычную форму из нескольких линейных рамп, он сглаживает нагрев там, где это выгодно с точки зрения физики разделения. Число рамп, допускаемое современными приборами, есть ограничение их управляющего программного обеспечения, и оно ограничивает качество разделения, которое та же колонка могла бы обеспечить.

Ключевые слова: газовая хроматография, температурная программа, центрированная по коэффициенту распределения модель, автоматическое дифференцирование, градиентная оптимизация, разрешение пиков, полициклические ароматические углеводороды

Температуралық бағдарламаларды градиенттік оңтайландыруға арналған таралу коэффициентіне негізделген дифференциалданатын газ хроматографиясының моделі

Я.С. Жигаленок*, В.А. Дубровский, Т.В. Кан, А.К. Галеева, А.А. Стародубцева

Farabi University, әл-Фараби даң., 71, Алматы, Қазақстан

АНДАТПА

Блумбергтің таралу коэффициентіне негізделген (K-centric) ұстау моделі, аналиттің миграция интегралы және Голей теңдеуіне негізделген жолақтың кеңею қабаты дифференциалданатын түрде қайта тұжырымдалды және ЕРА-ның 16 басым ПАК-і үшін жарияланған симуляцияға қатысты валидацияланды. Осының арқасында әрбір пик жұбының ажыратымдылығы температуралық бағдарламаның айқын дифференциалданатын функциясына айналады, ал осы ажыратымдылықтардан құралған кез келген критерийдің дәл градиенті бағдарлама параметрлерінің санына тәуелсіз түрде симулятор арқылы кері таралу жолымен алынады. Демек, бөлу шарттарын кез келген өлшемділіктегі бағдарламалар үшін мақсатты түрде таңдауға болады. Әдіс жалғыз бағдарламаны емес, «бөлу – талдау уақыты» ымырасының қолжетімді фронтын қайтарады; мұнда салыстырылған мақсат функциясының үш түрі осы фронттың әртүрлі бөліктеріне шығады, сондықтан критерий тұжырымның еркін бөлігі болып қалады және нақты талдауға қарай жазылады. Ешбір бағдарлама базалық бөлінуге жеткізе алмайтын жұптар іздеудің өзінен анықталып, критерийден шығарылады. Модельге жүгіну саны тең болғанда кері дифференциалдау тікелей іздеу, эволюциялық стратегия және ақырлы айырымдар жете алмайтын мақсат функциясының мәніне жетеді. Профиль түйіндерінің саны артқан сайын қолжетімді бөлу жақсарады, бұл ретте оптимизатор бірнеше сызықтық рампдан тұратын дәстүрлі пішінді қайталамайды, ол бөлу физикасы тұрғысынан тиімді болатын жерлерде қыздыруды тегістейді. Қазіргі аспаптар қабылдайтын рампалар саны – олардың басқару бағдарламалық қамтамасыз етуінің шектеуі, және ол сол колонка бере алатын бөлу сапасын шектейді.

Түйін сөздер: газ хроматографиясы, температуралық бағдарлама, таралу коэффициентіне негізделген модель, автоматты дифференциалдау, градиенттік оңтайландыру, пиктердің ажыратымдылығы, полициклді ароматты көмірсутектер