Quantification of out-of-range peaks

Yuri Kalambet, Ampersand International, Inc., Beachwood, OH
Ksenie Mikhailova, Ampersand Ltd., Moscow, Russia;
Yuri P. Kozmin, Shemyakin Institute of Bioorganic Chemistry, Moscow, Russia;
Igor Nagaev, Institute of Molecular Genetics RAS, Moscow, Russia;
Pavel Tikhonov, The Russian Research Institute of Metrological Service, Moscow, Russia

Mailto: **kalambet@ampersand.ru**

Abstract

The Exponentially Modified Gaussian (EMG) peak shape [1] is widely used for peak approximation in chromatography. We constructed the EMG peak deconvolution routine for chromatography, using a combination of two EMG formulas [1,2] and linear optimization methods. This routine accounts for the maximum linear range of the detector and can work with out-of range peaks.

The optimization routine is applied to the reconstruction of out-of-range peaks using the shape of correctly measured part, so that an analyst can get an idea of the height, area and concentration of such peaks. We have found that in many cases such reconstruction provides a reasonable prediction error. Peak reconstuction helps in the reducing number of chromatographic runs during method development and routine work. The possibility of reconstructing out-of-range peaks using the pre-defined peak shape obtained while calibrating is also discussed.

Introduction

Exponentially modified Gaussian [] is probably the best formula, describing chromatographic peak shape. One of the basic advantages of this formula is the physical model behind it: mixing chamber after ideal chromatographic column. Formula, describing EMG peak shape in most cases is written as

$$F(t)=\frac{h\cdot\sigma}{\tau}\sqrt{\frac{\pi}{2}}\cdot e^{\left(\frac{\sigma^2}{2\tau^2}-\frac{t-\mu}{\tau}\right)}\cdot\left(1-\mathbf{erf}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma}+\frac{\sigma}{\tau}\right)\right)\right) \qquad 1)$$

where *t* is time, *h* – Gaussian height, *σ* – Gaussian sigma, *μ* – position of unmodified Gaussian, *τ* – relaxation time, parameter of exponent used to modify Gaussian and $\operatorname{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$

Mathematically more precisely:

$$F(t)=\frac{h\cdot\sigma}{\tau}\sqrt{\frac{\pi}{2}}\cdot e^{\left(\frac{\mu-t}{\tau}+\frac{\sigma^2}{2\tau^2}\right)}\cdot\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma}+\frac{\sigma}{\tau}\right)\right) \qquad 2)$$

where $\operatorname{erfc}(z)=1-\operatorname{erf}(z)$.

Theory of Deconvolution

Well-forgotten alternative formula of EMG was derived by Deley [2] and in modern notations can be written as

$$F(t)=h\cdot e^{\frac{-(\mu-t)^2}{2\sigma^2}}\cdot\frac{\sigma}{\tau}\sqrt{\frac{\pi}{2}}\cdot\operatorname{erfcx}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma}+\frac{\sigma}{\tau}\right)\right) \qquad 3)$$

where $\mathbf{erfcx(}z\mathbf{)}=e^{z^2}\,\mathbf{erfc(}z\mathbf{)}$. Notably, Formula (3) has the initial Gaussian as a multiplicand.

It can be shown, that in the extreme case of very small *τ/σ*

$$F(t)\approx\frac{h\cdot e^{\frac{-(\mu-t)^2}{2\sigma^2}}}{\left(1+\frac{(\mu-t)\cdot\tau}{\sigma^2}\right)} \qquad 4)$$

The example of EMG is one of very rare cases, when one curve can be described by different formulas (2, 3) with completely different computational properties. Really, formula 2) has computational instability: in the case of small τ exponent can become quite large, and can easily become larger than largest possible double precision floating-point number. Formula 3) has no problems in the case of small τ, although it has its own region of computational

instability. Formula 4) is an extreme particular case of formula 3). These formulas together allowed an easy calculation of EMG for any set of its parameters, whereas any one of them does not allow such calculation.

The Fourier transform procedure of EMG peaks deconvolution [] cannot be used in the case of the overloaded peak, as the Fourier spectrum of such peak is very broad and is not suitable for deconvolution. Traditional linear optimization methods were used instead.

Materials and methods

Chromatograph (Bischoff Chromatography):

Detector – mod. Lambda 1010;

Pump – mod.2250, 0.01-4.99 ml/min pumphead;

Variotherm column thermostat, 35°C.

Separation:

Eluent: methanol-water-acetic acid (45:55:0.1), isocratic mode

Flow rate: 1 ml/min.

Column 1: Reprosil pur C18aq, 5um, 4*150 mm

Column 2: Kromasil 100C18, 6 мкм, 4*150 мм

Detection wavelength: 250 nm

Software: Chrom&Spec by Ampersand International, Inc.

Sample: nipagin (methyl n-hydroxybenzoat), solution 1mg/ml in methanol

Results

For evaluation of our reconstruction procedure we selected 44 standalone peaks with good signal to noise ratio (not less than 10% of the linear range of the detector) from different LC chromatograms. Peak shape was found to be close to EMG: residual relative RSD was not higher than 3%, and average relative RSD was found to be 1.4%.

Every peak was approximated by EMG using points with response value not exceeding 10, 25, 50 and 100 percent of the true peak height. Example screenshot of the overlaid initial and reconstructed peaks are shown on Fig.1.

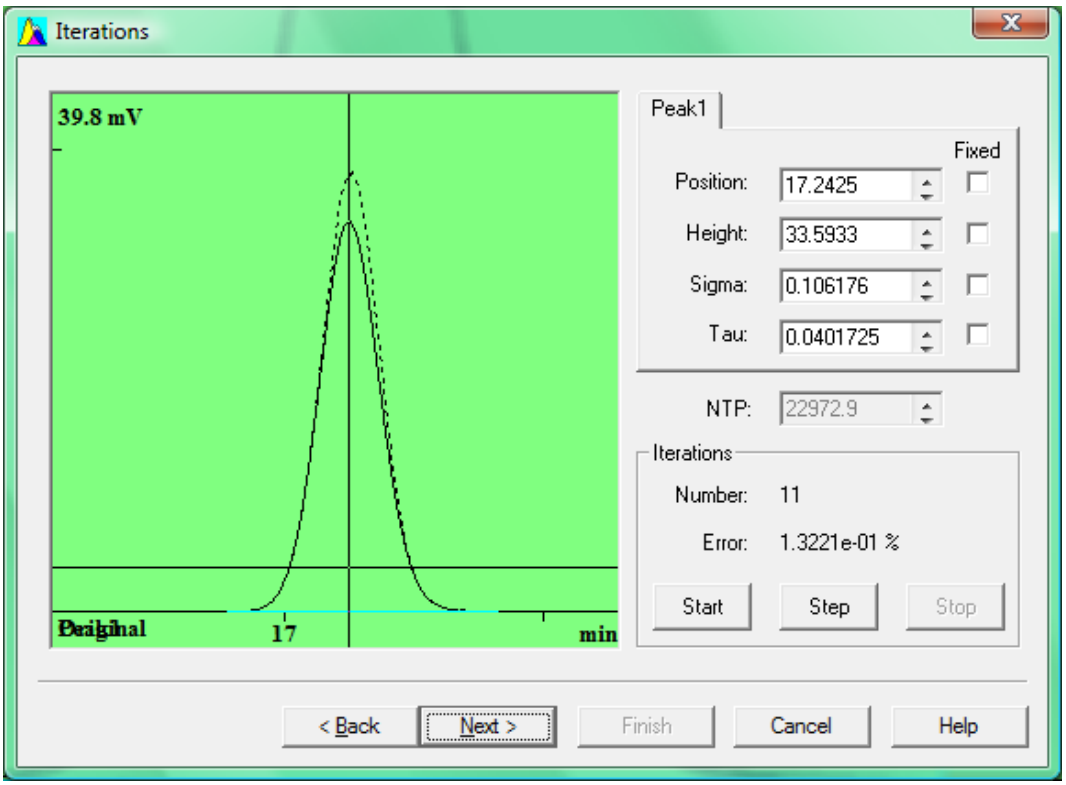

Figure 1. Window of the software deconvolution module. Reconstruction of the peak using 10% of the peak height. Original data are drawn by the dotted line, reconstructed – by the solid line. Only points between baseline and the horizontal line above the baseline are used for the reconstruction of the peak.

For every approximation we measured relative prediction error for height and area:

$$dV = 100*abs(Vr-V)/V \qquad 5)$$

where V stays for the parameter measured (area or height) and Vr – the same parameter of the reconstructed peak. Average values of prediction errors as a function of the used percent of height are presented in Fig.1.

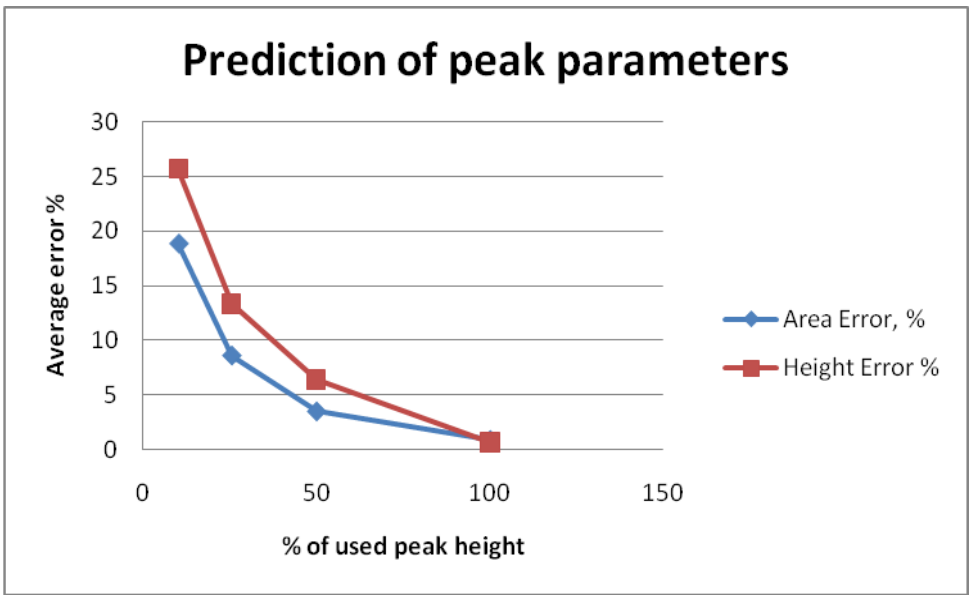

Figure 2. Prediction of the peak height and area for 44 stand-alone peaks from different chromatograms (all LC)

Fig.1 shows, that in the optimistic cases, to which our examples belong, one could hope for quite good prediction of peak parameters: average error of prediction for just 10% of the height used for reconstruction is about 25%.

To test the reconstruction scheme in real experiment we conducted two series of analyses of the substance Nipagin using two different columns. Series 1 included 4-point calibration and 3 overloaded peaks. Overlaid chromatograms of this series are presented on Fig.3.

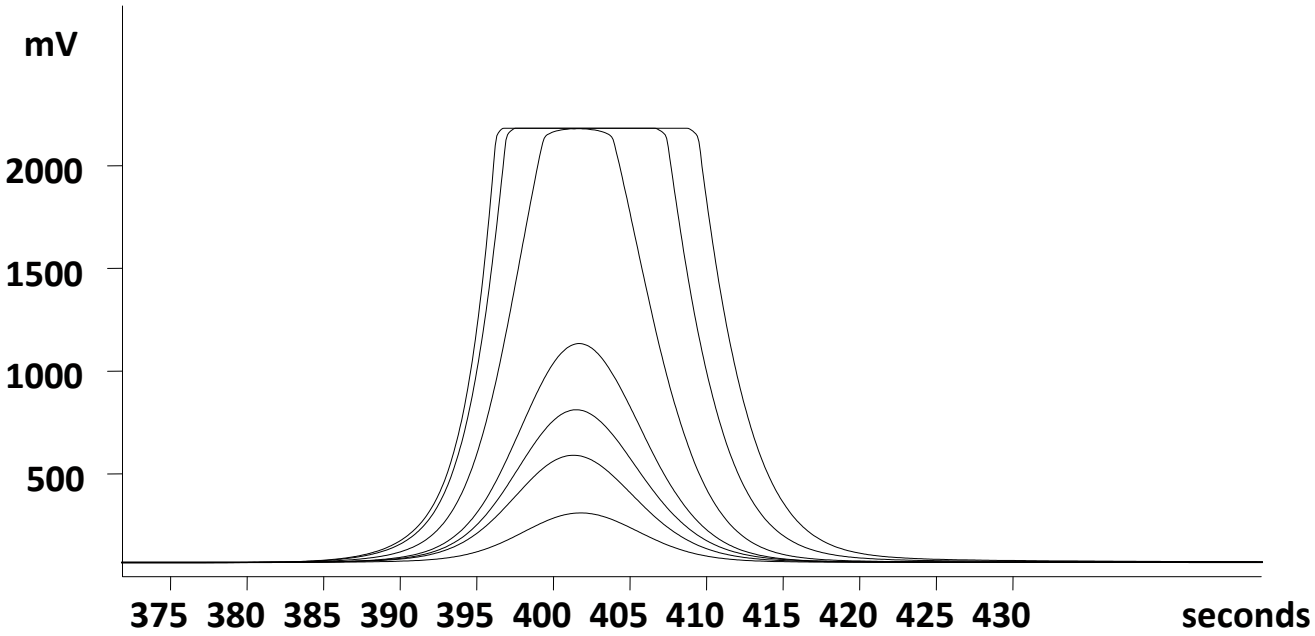

Figure 3. Series 1 of Nipagin chromatograms; concentrations 1: 2: 3: 4: 10: 20: 30. Time scale of some of chromatograms was shifted within 3 seconds limit for better visual comparison.

For the purpose of peak reconstruction, peak from the calibration point 4 (the highest peak in the calibration set) was selected as a Model, that is we accepted it as an example of the “true” peak shape. The height of this peak was 53% of the detector linear range, relative RSD 1.5% and $\tau/\sigma=0.54$.

Every out-of-range peak was reconstructed by several different procedures:

- EMG approximation without restrictions (free EMG shape);
- EMG approximation with fixed τ and σ, equal to those of the Model peak (fixed EMG shape);
- Gaussian approximation without restrictions.

An example of reconstructed peak is shown on Fig.4.

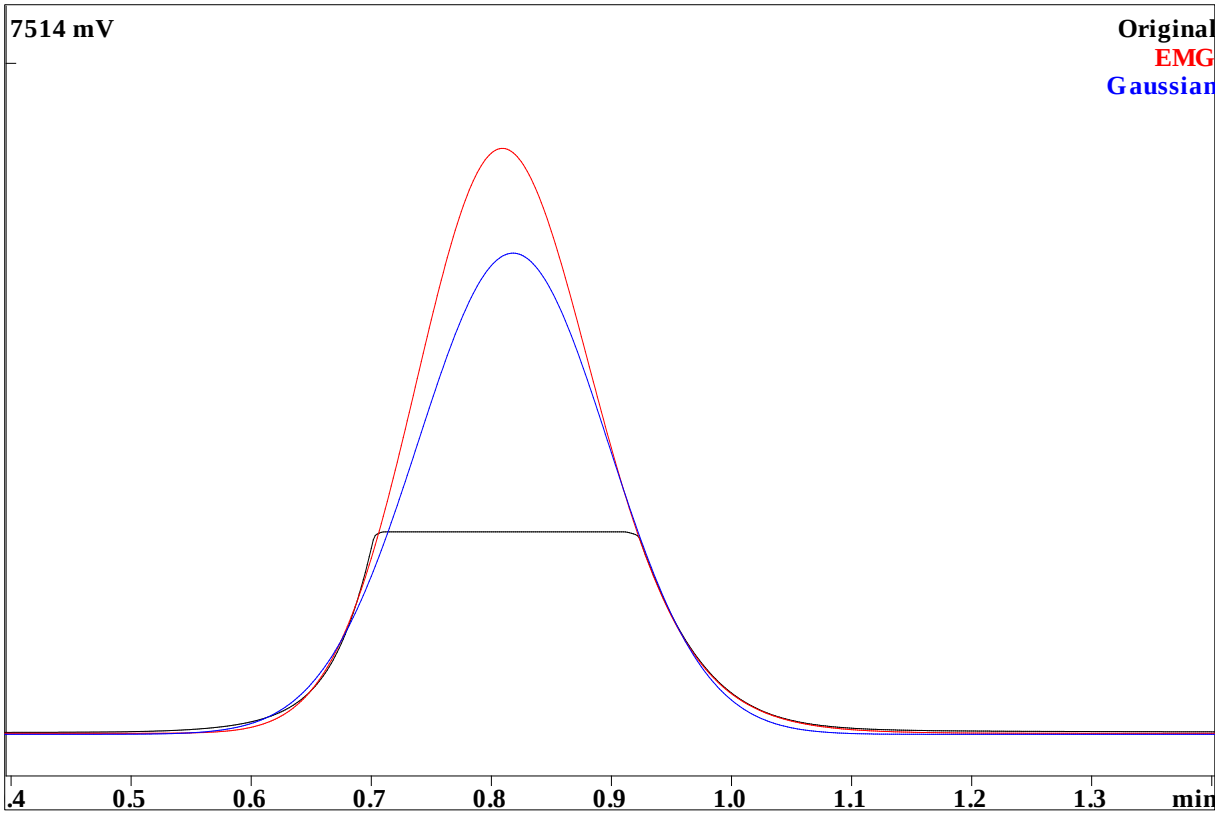

Figure 4. Reconstruction of the nipagin peak (black line) by EMG (red line) and Gaussian (blue line)

Results of out-of-range peak reconstruction were compared to the reconstruction of the Model peak using different signal levels. The results are shown on fig.5.

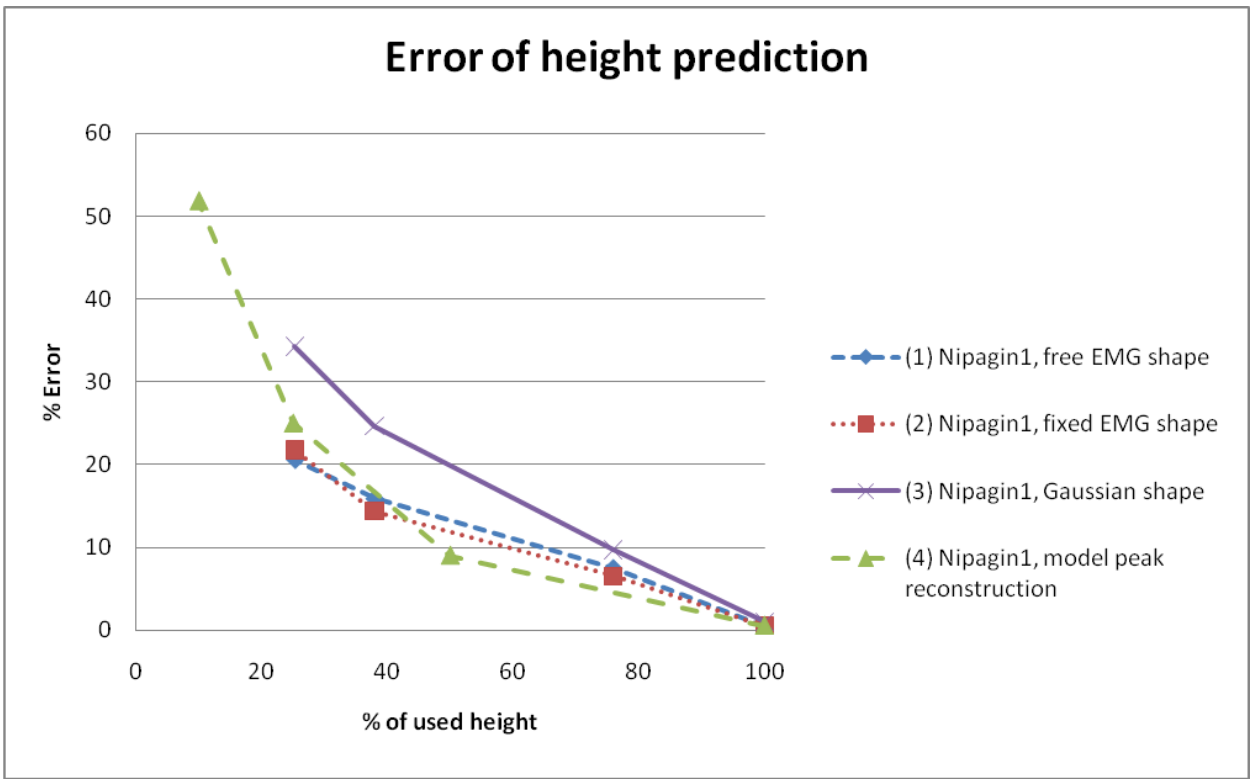

Figure 5. Peak reconstruction for Nipagin Series 1. For the fixed EMG shape σ and τ parameters of the EMG were fixed to the values, obtained for the Model peak; free EMG shape allow σ and τ to be optimized. The Model peak is the highest "normal" peak (peak with concentration of 4 in Series 1). "True" height and area were determined by the extrapolation of the calibration dependence.

Reconstruction results for this series (Line 4) are not too far from the expected values both for approximation without restrictions (Line 1), and for approximation with fixed EMG shape (Line 2). Approximation using Gaussian shape (Line 3) is significantly worse.

In Series 2 we made an attempt to move to more significant linear range overloads. This series consisted of 5 analyses, the first (Model) one corresponded to 91% of the detector linear range, relative RSD 0.93% and $\tau/\sigma=0.84$. In other analyses injected volume was increased 2, 3, 4 and 5 times and peaks were out of linear range of the detector. Resulting peaks were processed analogous to Series 1. Results of the reconstruction are presented on fig.6.

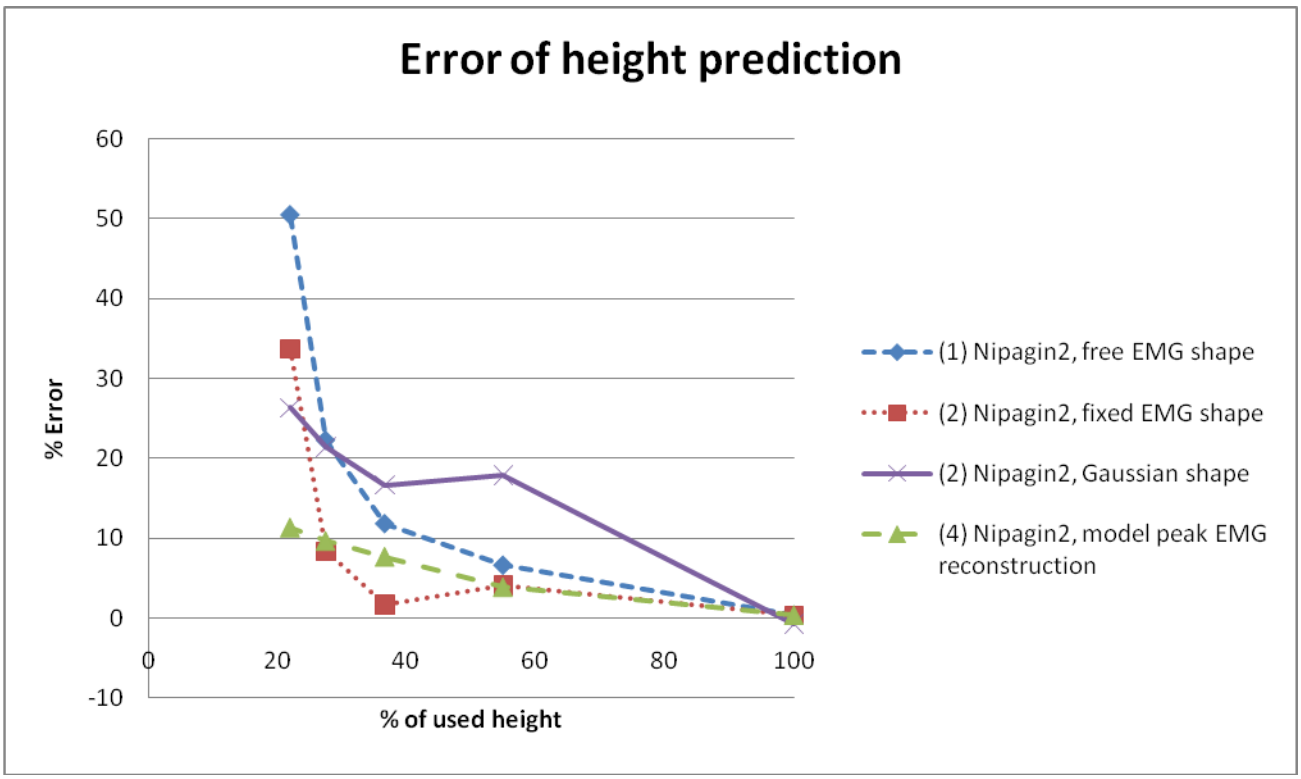

Figure 6. Peak reconstruction for Nipagin Series 2. Concentrations 1: 2: 3: 4: 5. Already, the second peak is out of the range. The "true" peak height was calculated by scaling of the Model (concentation = 1) peak height.

It's easy to see, that prediction results for bigger peaks are worse, than expected, allthough up to 4 times overload (25%) they are still acceptable.

There are many reasons of peak shape distortion in the case of out-of-range peaks:

- Detector nonlinearity;
- Peak broadening due to increased injected volume;
- Broadening or sharpening due to content difference in eluent and solute of injected sample;
- Column overload.

We do not pretend to list all possible reasons of peak shape changes and will not try to reveal, what is the main reason in our case.

While the search for optimal peak reconstruction it was noted, that reconstruction using fixed peak shape performed much better, than optimization without restrictions: it never came to wrong lokal minimum, whereas unrestricted optimization can stick to local minima quite easily, and requires careful selection of initial conditions. The best result was always close to the minimum of the optimization with fixed shape, so that typical procedure can consist of "rough" optimization with fixed shape (if this shape is known) and "fine" optimization using full set of variables, including σ and τ.

Good results of peak modeling using fixed peak shape raises a possibility of wide using of the known peak shapes, but we have to be sure, that peak shape is really constant. Checking conservation of peak shape can be done using calibration runs in the case they are available. A good indicator of EMG peak shape is a τ/σ ratio, and if it remains constand while calibration runs, we have a good reason to expect that peak shape will remain the same in the analysis runs.

We made a short investigation of the constancy of the peak shape in the calibrations available at our disposal. In most cases τ/σ ratio remained quite constant, showing behaviour similar to Nipagin (Figure 7, Line 3), sometimes with less asymmetry for low concentrations, but some cases where peak shape was far from being constant were also met (Figure 7, Lines 1 and 2)

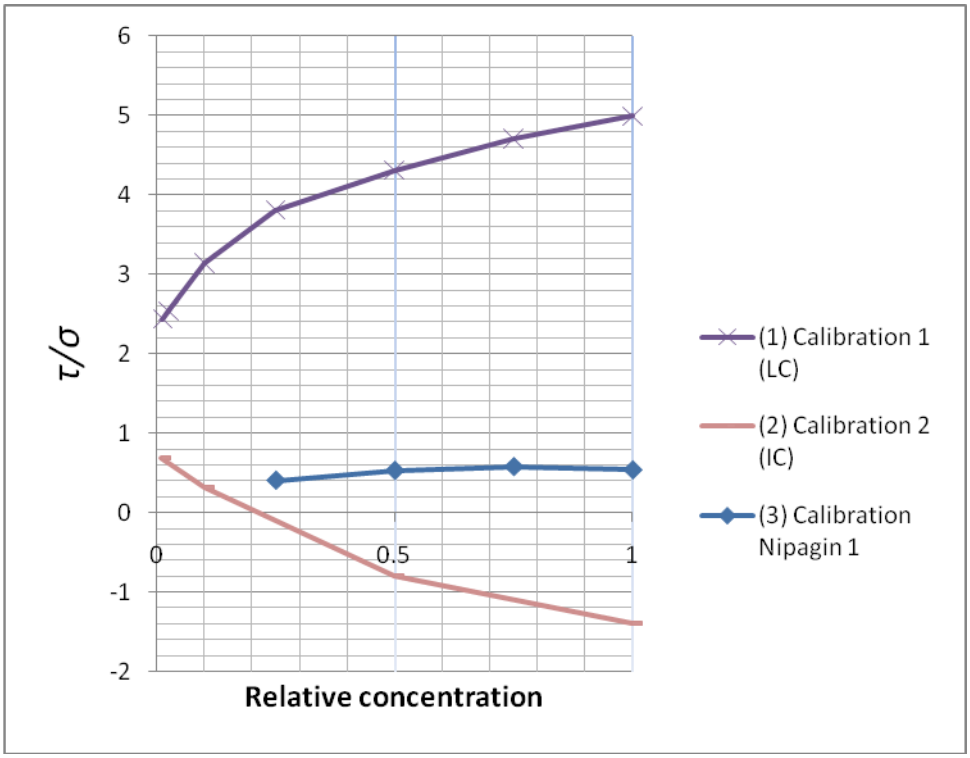

Figure 7. Dependence of peak shape on the concentration. Calibration 1 and 2 are the extreme cases found after investigation of more than a hundred calibrations of different types.

Conclusions:

1. Peaks may be reconstructed with reasonable accuracy for up to 4-times overload.
2. Area is reconstructed better than height.
3. Reconstruction using known peak shape may give good results, but monitoring of peak shape while calibration is required in this case.
4. Column overload cannot be accounted for by this technology due to peak shape distortion.
5. The technology can be used for getting an estimate of required dilution on early steps of method development or in the case of big error in concentration. Numerical results may be suitable for internal use of the investigation laboratory.

References

[1] McWilliam, I. G.; Bolton, H. C. Anal. Chem. 1969, 41, 1755-1762.

[2] Delley, R. *Anal. Chem*. 1985, 57, 388