Computation and illustration of the generalized order restricted criterion (GORIC) for shape constrained spline functions

Kris Villez

September 6, 2020

OAK RIDGE NATIONAL LABORATORY

Electrical and Electronics Systems Research (EESR) Division

Computation and illustration of
the generalized order restricted criterion (GORIC)
for shape constrained spline functions

Kris Villez

Date Published: May 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

DOF	Degrees of freedom
GORIC	Generalized order restricted information criterion
LP	Linear program
QP	Quadratic program
RMSR	Root mean squared residual
SSR	Sum of squared residuals

ABSTRACT

The selection of a model among alternative models that include inequality constraints is a complex task. The generalized order restricted criterion (GORIC) is a suitable approach to do so. However, the GORIC proves challenging to compute in the general case. This report discusses all known methods to compute the GORIC and illustrates their use with a set of didactic examples.

1. Introduction

1.1 Inequality constrained regression

Inequality constrained regression models are regression models for which one or more parameters are subject to inequality constraints. Such inequality constraints often expresses basic ground truths about the system under study. For example, a regression model may be constrained to predict positive values only. This is sensible for variables like chemical concentrations or probability densities. Inequality constraints are also useful to express the shape of a functional relationship between input and output variables. For example, Meyer and Habtzghi [2011] discuss models to estimate hazard rate functions subject to shape constraints. A wide variety of statistical methods have been developed to estimate regression models subject to shape constraints [Delecroix et al., 1996, Mammen and Thomas-Agnan, 1999, Beliakov, 2000, Delecroix and Thomas-Agnan, 2000, Turlach, 2005, Meyer, 2008, Hazelton and Turlach, 2011, Meyer et al., 2011, Meyer and Habtzghi, 2011, Habtzghi and Datta, 2012, Papp and Alizadeh, 2014]. A fairly comprehensive treatment is given in Silvapulle and Sen [2011]. These statistical methods generally assume that the applicable constraints are specified completely for a given model. This means *(a)* that the shape of the fitted function is defined by means of signs of the first and/or second derivative and *(b)* that these signs are known everywhere in the function's domain.

1.2 Model selection in the presence of inequality constraints

In some areas of research, there may exist mulitple candidate models for a given system, each with their own set of inequality constraints. These candidate models are also described as informative hypotheses Hoijtink et al. [2008]. The selection of a hypothesis supported by the available evidence constitutes a model selection problem, for which multiple methods have been proposed [Mulder et al., 2010, Gu et al., 2018]. The generalized order restricted information criterion (GORIC), as proposed and studied in Kuiper et al. [2011, 2012], Kuiper and Hoijtink [2013] is especially appealing due to its similarity with the Akaike information criterion [AIC, Akaike], which is a applicable for model selection in absence of inequality constraints [e.g. Hipel et al., 1977, Aho et al., 2014].

The GORIC could be also useful for process diagnosis and automation on the basis of shape-constrained spline function fitting. Indeed, promising results based on shape constrained function fitting methods have been shown already in Villez et al. [2013] and Thürlimann et al. [2019]. However, these methods require the use of decision criteria that are tuned in a subjective manner. Potentially, the GORIC, or another criterion derived form it, could allow for a more objective trend-based decision-making in process monitoring and control.

1.3 Computational challenges

The GORIC, while appealing at a conceptual level, poses a computational challenge. In all but the most simple of cases, computing the GORIC requires a computationally intensive Monte Carlo procedure. To evaluate the need for further research in this direction, all known methods to compute the GORIC are listed in this report. In addition, several of these methods have been implemented and evaluated in terms of their computational efficiency.

In this report, the GORIC is first illustrated as a selection criterion for data series segmentation, particularly into segments that are defined by the shape of the function (e.g., increasing and decreasing segments, convex and concave segments). After illustration, special attention is given to the computation of the expected degrees of freedom, which is the term in the GORIC that is particulary challenging to compute.

2. Illustration of the generalized order restricted criterion (GORIC)

2.1 Data set for demonstration

We apply the GORIC for segmentation of a data series. To this end, a range of shape-constrained spline models, each corresponding to a different candidate segmentation, are fit to the data and the GORIC is used as the selection to find the best model. We use the Titanium data set from de Boor [1978] to demonstrate the methods. As in Villez et al. [2013], this data series is chosen here because of its distinct shape profile. The data series is described well as consisting of an increasing and decreasing segment. Each of these two segments can be decomposed further into a convex and concave segment.
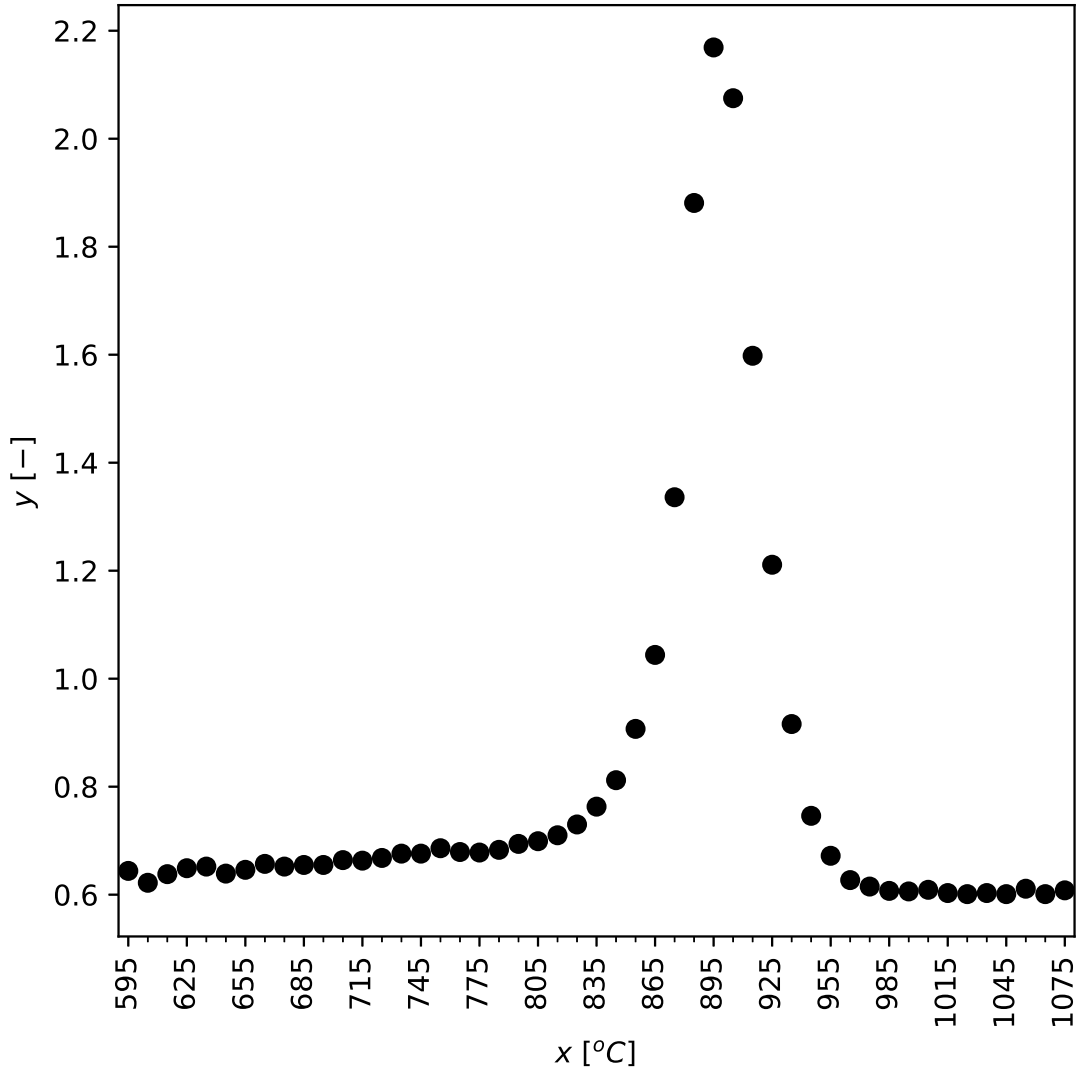
Figure 1. Data set used for demonstration purposes. Measurements (y) as a function of the independent variable (temperature, x).

This univariate data series is shown in Fig. 1 and consists of $N = 49$ pairs $(x_i, \tilde{y}_i)$ with x_i the known value of an independent variable and $\tilde{y}_i$ the measurement of the corresponding dependent variable ($i = 1, \ldots, N$). These data are

organized into two N-dimensional vectors $\boldsymbol{x}$ and $\tilde{\boldsymbol{y}}$ with:

$$x_i = \boldsymbol{x}(i) \tag{1}$$

$$\tilde{y}_i = \tilde{\boldsymbol{y}}(i) \tag{2}$$

2.2 Constraining the first derivative

We first fit piece-wise linear functions to the data while constraining the shape of the function to consist have a distinct shape. This shape is always defined by two consecutive segments, which are - in this order - *(1)* increasing (isotonic) and *(2)* decreasing (antitonic). Practically, this is achieved by solving the following quadratic program (QP), which minimizes the sum of squared residuals (SSR) while ensuring that the imposed shape constraints are satisfied:

$$\hat{\boldsymbol{\beta}} := \arg\min_{\beta} SSR = \sum_i^N (y_i - \tilde{y}_i)^2 = (\boldsymbol{y} - \tilde{\boldsymbol{y}})^T \cdot \boldsymbol{W} \cdot (\boldsymbol{y} - \tilde{\boldsymbol{y}}) \tag{3}$$

$$\boldsymbol{B}_0 \cdot \boldsymbol{\beta} = \boldsymbol{y} \tag{4}$$

$$\boldsymbol{S}_1 \cdot \boldsymbol{B}_1 \cdot \boldsymbol{\beta} \geq \boldsymbol{0} \tag{5}$$

with $\boldsymbol{W} := \boldsymbol{I}_{N \times N}$ and $\boldsymbol{B}_0 := \boldsymbol{I}_{N \times N}$. The $N - 1 \times N$-dimensional matrix $\boldsymbol{B}_1$ is used to compute the first-order differences along $\boldsymbol{y}$:

$$\boldsymbol{B}_1 := \begin{bmatrix} -1 & +1 & 0 & 0 & \dots & 0 \\ 0 & -1 & +1 & 0 & \dots & 0 \\ 0 & 0 & -1 & +1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & +1 \end{bmatrix} \tag{6}$$

The diagonal $N - 1 \times N - 1$-dimensional matrix $\boldsymbol{S}_1$ is used to specify the signs of the first-order difference along $\boldsymbol{y}$:

$$\boldsymbol{S}_1 := \mathrm{diag}\,(\boldsymbol{s}_1) \tag{7}$$

with $\boldsymbol{s}_1$ the $N - 1$-dimensional vector of signs (+1 or -1) for the first-order difference. These signs are chosen such that the fitted function satisfies the isotonic-antitonic shape described above. There are $N - 2 = 47$ vectors that satisfy this requirement for $\boldsymbol{s}_1$. Fig. 2 shows exemplary solutions of the above QP for three candidates for $\boldsymbol{s}_1$. The top panel in Fig. 3 shows the objective function of the QP as a function of the location of the change in sign, which is the location of the sign change in $\boldsymbol{s}_1$. Note that placing the change in sign at $T = 595$ means that the fitted function is isotonic over the whole domain. Similarly, placing the sign change at sign at $T = 1075$ means that the fitted function is antitonic over the whole domain. By including these two monotone functions (isotonic and antitonic) also, one obtains a total of $M = 49$ models.

We now aim to select a parsimonious model among the M models. To this end, the generalized order restricted information criterion (GORIC) is used. The GORIC is a model selection criterion similar to the AIC but is especially adapted for inequality constrained regression models, like the shape constrained spline functions fitted above [Kuiper et al., 2012]. It is written as:

$$GORIC := -2\,LL + 2\,\bar{D} \tag{8}$$

$$= -2\,\log\left(\left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{N/2} \exp\left(-\frac{SSR}{2\hat{\sigma}^2}\right)\right) + 2\,\bar{D} \tag{9}$$

where the first term is equal to twice the negative log-likelihood, thus reflecting the fit of the model, and where the second term is a penalty term equal to twice the expected degrees of freedom $\bar{D}$.

Figure 2. Exemplary fits of three shape constrained piece-wise linear functions. Cyan sectors indicate the selected location of the change of sign in the first derivative.

The estimate of the measurement error standard deviation ($\hat{\sigma}$) is set equal to its maximum likelihood estimate, which in this case is equal to the root mean squared residual (RMSR):

$$\hat{\sigma} := RMSR = \sqrt{\frac{SSR}{N}} \tag{10}$$

so that the GORIC simplifies to:

$$GORIC := N\left(1 + \log(2\pi) + \log\left(\hat{\sigma}^2\right)\right) + 2\bar{D} \tag{11}$$

As in Kuiper et al. [2012], the expected degrees of freedom is given by the number of parameters (β, σ) minus the expected number of active constraints, $\bar{C}$, among the inequality constraints in the QP:

$$\bar{D} := N + 1 - \bar{C} \tag{12}$$

4
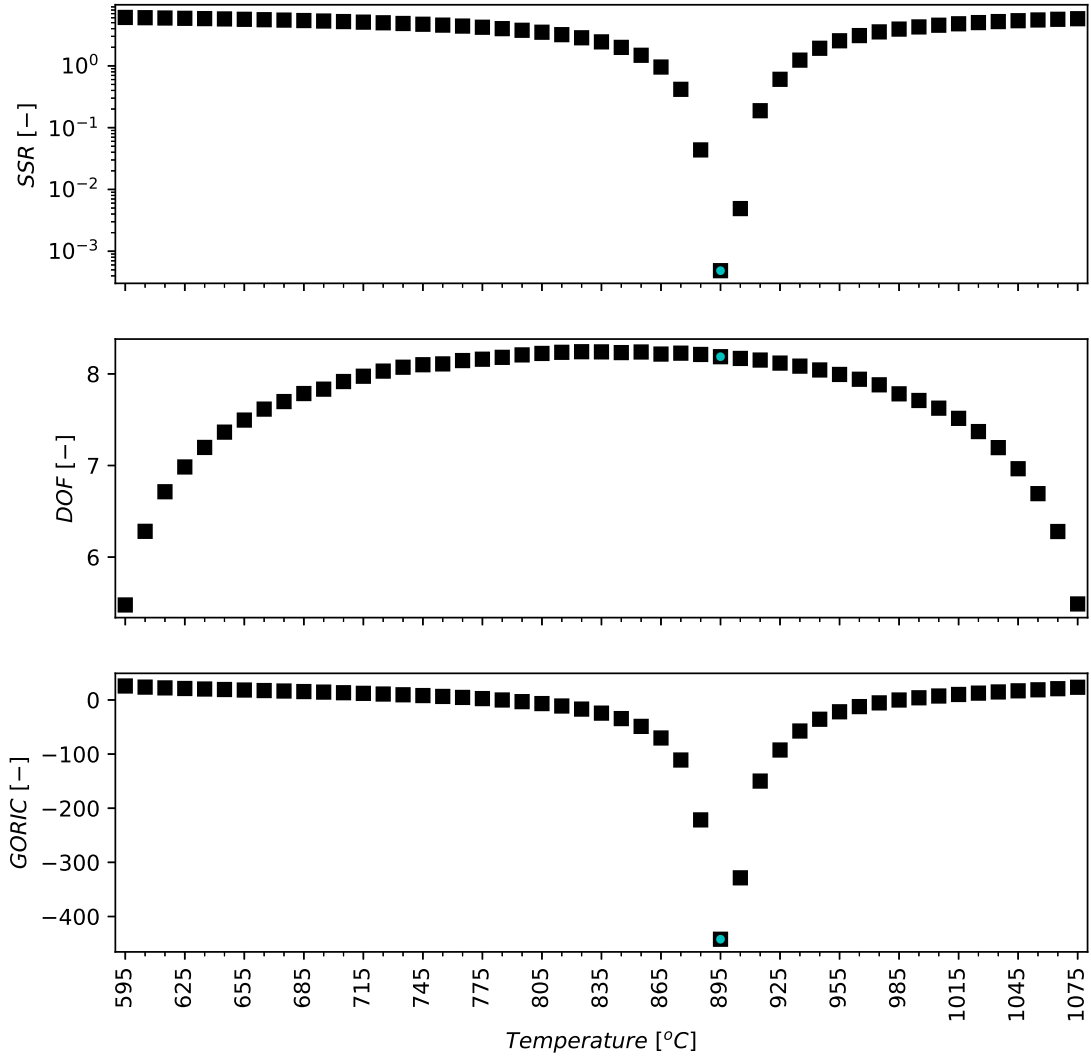
Figure 3. Model selection. Top: sum of squared residuals (SSR); Middle: degrees of freedom (DOF); Bottom: generalized order restricted information criterion (GORIC). All measures are shown as a function of the location of the sign change in the first-order differences.

In this case, $\bar{C}$ is estimated by means of Monte Carlo integration using $K = 10000$ random samples. This is explained in further detail below.

The estimates $\bar{D}$ are shown in the middle panel of Fig. 3 as a function of the location of the sign changes in s_1. One can see that spline functions with sign changes at the extremes (close to the beginning or end of the data series) lead to a smaller degrees of freedom, indicating that these models are less flexible. The values for $\bar{D}$ lie between 5 and 9, which is substantially lower than the number of parameters (50).

Figure 4. Selected model. Top: data (dots) and fitted function (line); Bottom: first-order differences of the data (dots) and slopes of fitted piece-wise linear function (line). Cyan sectors indicate the selected location of the change of sign in the first derivative.

6

2.3 Constraining the second derivative

For a second demonstration of the GORIC, we fit piece-wise linear functions to the data while constraining the curvature of the function. The imposed shape is always defined by three consecutive segments, which are - in this order - *(1)* convex, *(2)* concave, and *(3)* convex. Practically, this is achieved by solving the following QP:

$$\hat{\boldsymbol{\beta}} := \arg\min_{\beta} \sum_{i}^{N} (y_i - \tilde{y}_i)^2 = (\boldsymbol{y} - \tilde{\boldsymbol{y}})^T \cdot \boldsymbol{W} \cdot (\boldsymbol{y} - \tilde{\boldsymbol{y}}) \tag{13}$$

$$\boldsymbol{B}_0 \cdot \boldsymbol{\beta} = \boldsymbol{y} \tag{14}$$

$$\boldsymbol{S}_2 \cdot \boldsymbol{B}_2 \cdot \boldsymbol{\beta} \geq \boldsymbol{0} \tag{15}$$

with $\boldsymbol{W} := \boldsymbol{I}_{N\times N}$ and $\boldsymbol{B}_0 := \boldsymbol{I}_{N\times N}$. The $N - 2 \times N$-dimensional matrix $\boldsymbol{B}_2$ is used to compute the second-order differences along $\boldsymbol{y}$:

$$\boldsymbol{B}_1 := \begin{bmatrix} +1 & -2 & +1 & 0 & 0 & \dots & 0 \\ 0 & +1 & -2 & +1 & 0 & \dots & 0 \\ 0 & 0 & +1 & -2 & +1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & +1 \end{bmatrix} \tag{16}$$

The diagonal $N - 2 \times N - 2$-dimensional matrix $\boldsymbol{S}_2$ is used to specify the signs of the second-order difference along $\boldsymbol{y}$:

$$\boldsymbol{S}_2 := \mathrm{diag}\,(\boldsymbol{s}_2) \tag{17}$$

with $\boldsymbol{s}_2$ the $N - 2$-dimensional vector of signs (+1 or -1) for the second-order difference. These signs are chosen such that the fitted function satisfies the convex-concave-convex shape described above. There are $M = (N - 4)(N - 5) = 1980$ vectors that satisfy this requirement for $\boldsymbol{s}_2$. Fig. 5 shows 3 examples of such a vector and the associated solution of the above QP. Fig. 6 shows the objective function (SSR) as a function of the location of the sign changes in $\boldsymbol{s}_2$. Note that the cells at left border and top border correspond to simpler shapes, namely concave-convex (left border) and convex-concave (top). The very top-left cell corresponds to a concave shape over the whole domain.

We now aim to select a parsimonious model among the $M = (N - 4)(N - 5)$ fitted functions. To this end, the GORIC is used again. Once more, this requires Monte Carlo integration to compute the degrees of freedom $\bar{D}$. In this case, $K = 10000$ samples are used. The resulting values for $\bar{D}$ are shown in Fig. 7. The values range between 4 and 8, again showing that application of shape constraints is an effective way to control the complexity of a spline-based model. Note that the DOF for functions with a simpler shape (left-border and top-border, as above) are considerably lower. This means that the computed DOF matches the intuition that increasing the number of segments in the imposed shape increases the complexity of the fitted model. Conversely, this means the DOF provides a quantifiable (and possibly objective) measure of this intuited complexity.

The DOF and SSR values are combined to compute the GORIC (as above). The GORIC values are shown in Fig. 8. The selected model, i.e. the model with minimum GORIC value, is illustrated in Fig. 9.
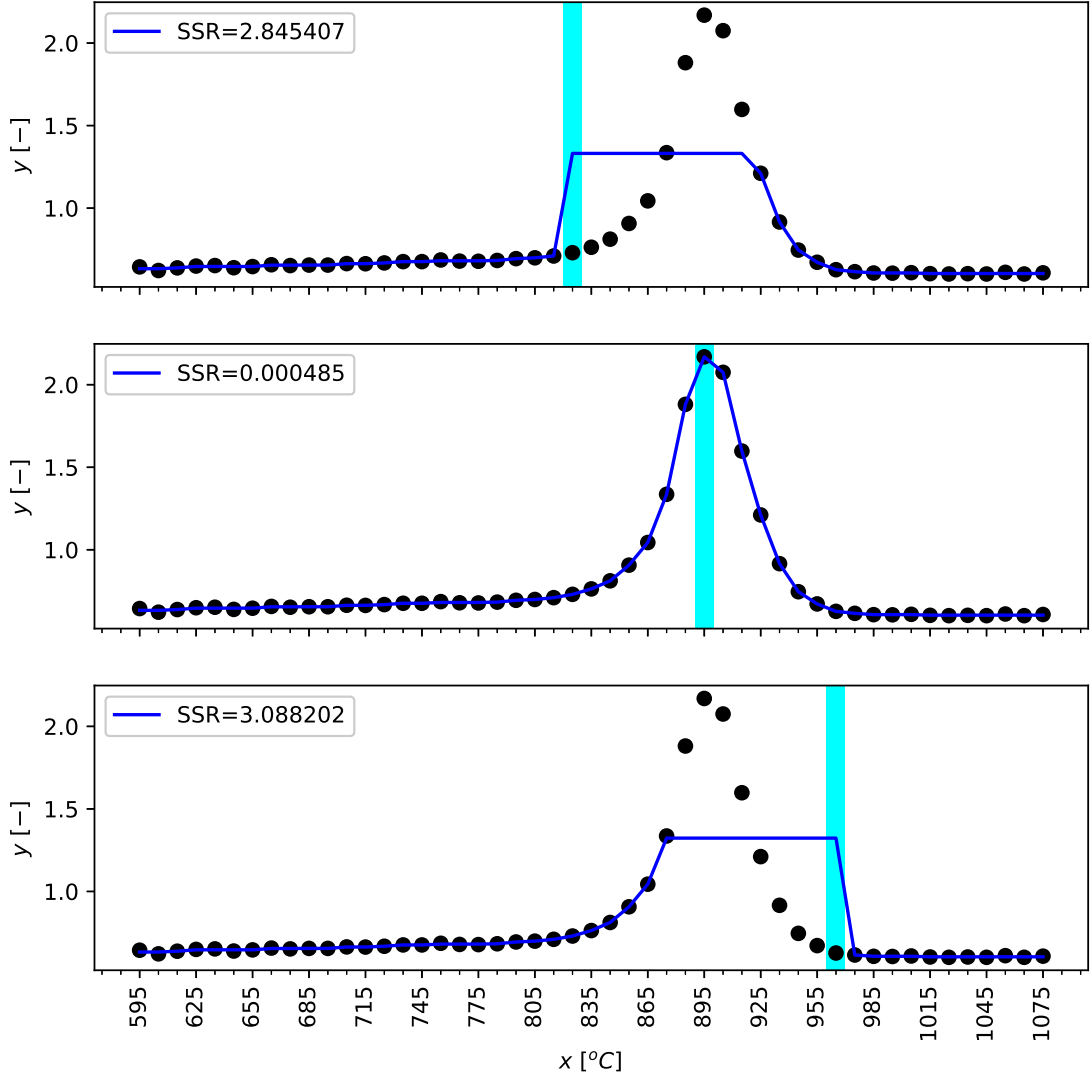
Figure 5. Exemplary fits of three shape constrained piece-wise linear functions. Colored bars indicate the selected location of the change of sign in the second derivative.

Figure 6. Heatmap of the sum of squared residuals (SSR) as a function of the locations of sign changes in the second derivative.
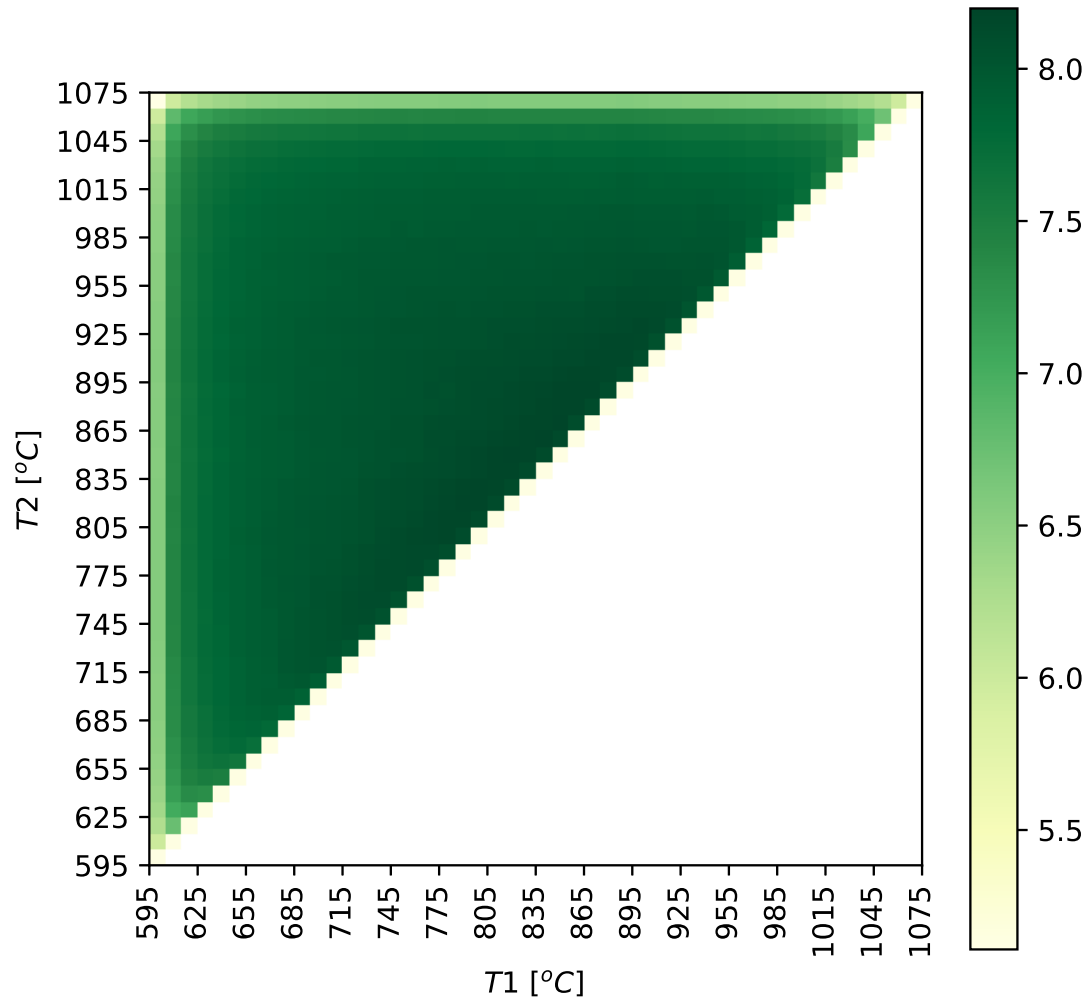
Figure 7. Heatmap of the degrees of freedom (DOF) as a function of the locations of sign changes in the second derivative.
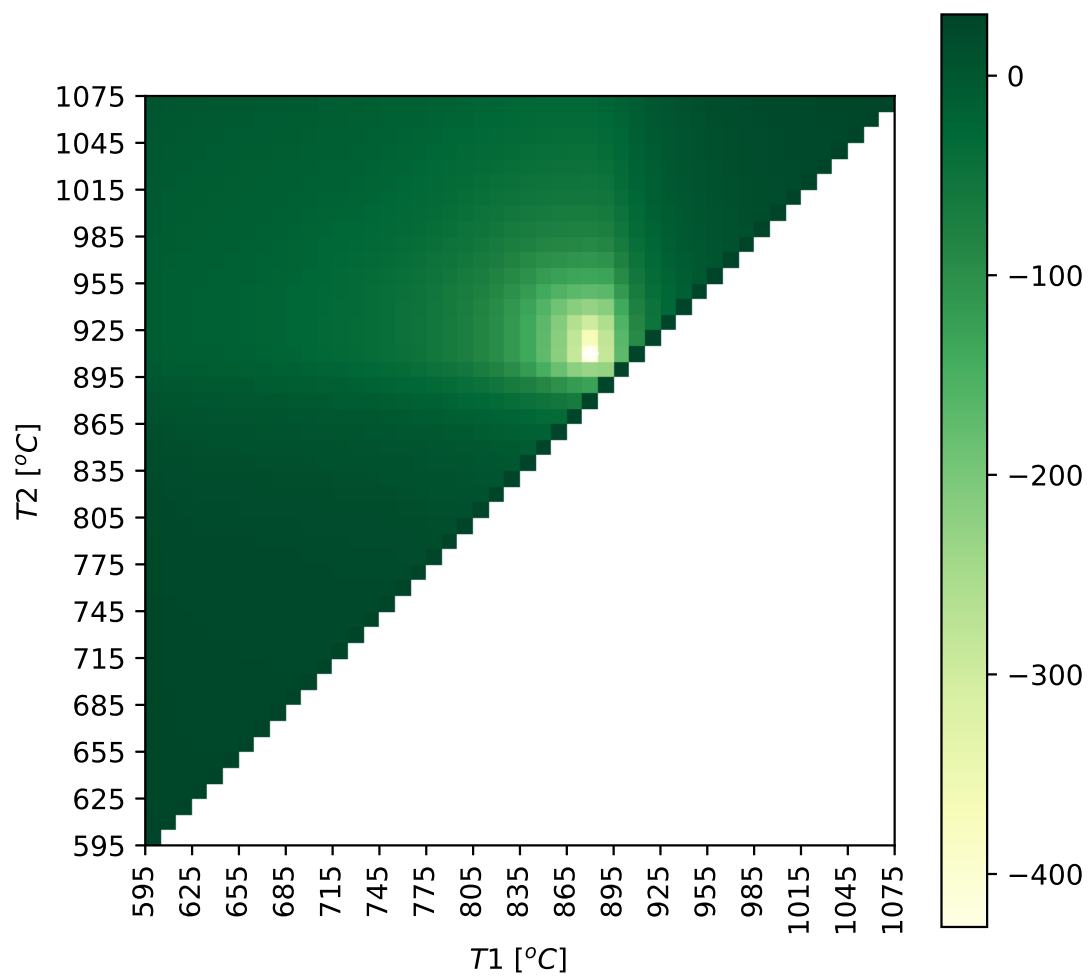
Figure 8. Heatmap of the generalized order restricted information criterion (GORIC) as a function of the locations of sign changes in the second derivative.
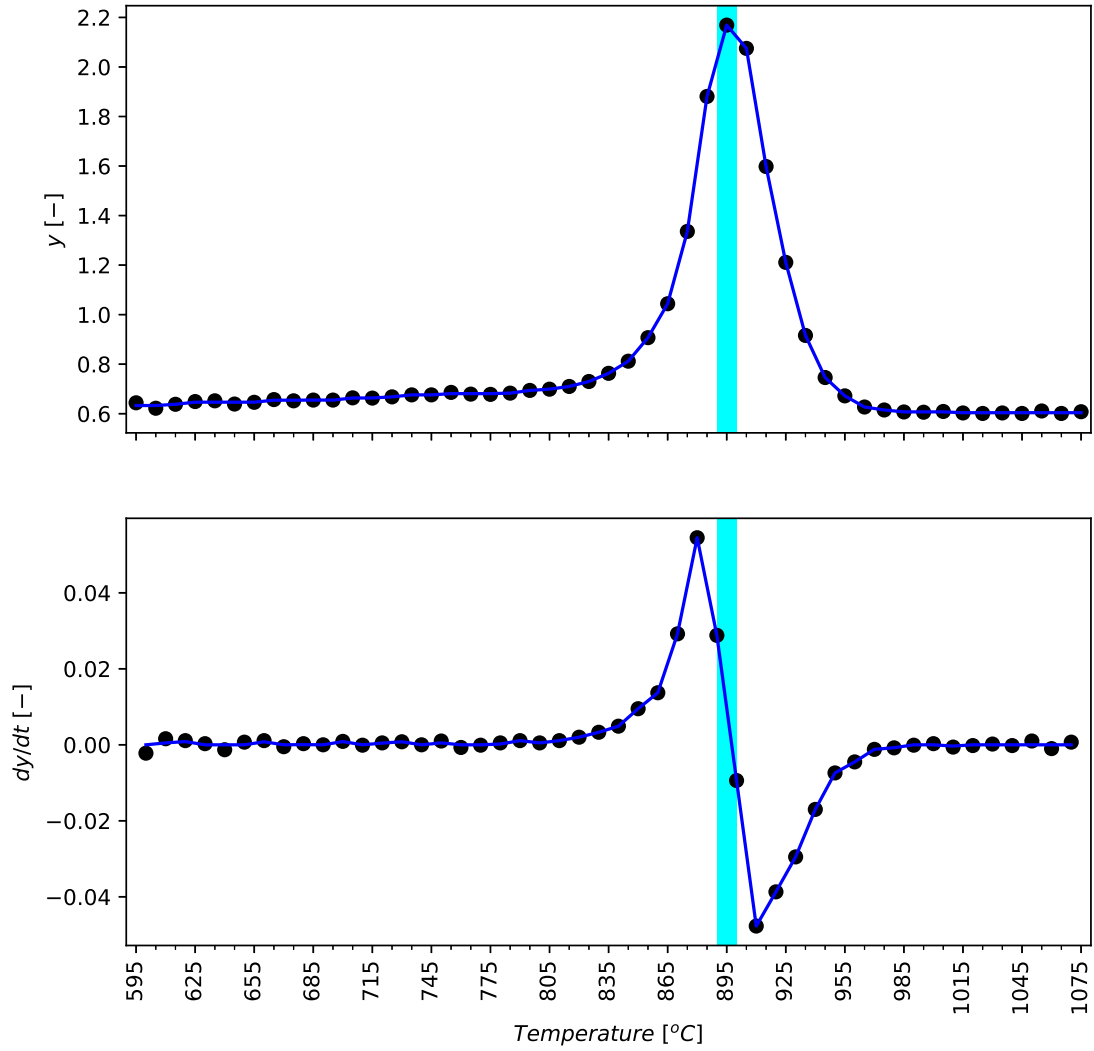
Figure 9. Selected model. Top: data (dots) and fitted function (line); Bottom: first-order differences of the data (dots) and slopes of fitted piece-wise linear function (line). Colored bars indicate the selected locations of the change of sign in the second derivative.

2.4 Constraining the first and second derivative

For a third demonstration of the GORIC, we fit piece-wise linear functions to the data while constraining both the slope and curvature of the function. The imposed shape is always defined by three consecutive segments, which are - in this order - *(1)* convex-increasing, *(2)* concave, and *(3)* convex-decreasing. Practically, this is achieved by solving the following QP:

$$\hat{\boldsymbol{\beta}} := \arg\min_{\beta} \sum_{i}^{N} (y_i - \tilde{y}_i)^2 = (\boldsymbol{y} - \tilde{\boldsymbol{y}})^T \cdot \boldsymbol{W} \cdot (\boldsymbol{y} - \tilde{\boldsymbol{y}}) \tag{18}$$

$$\boldsymbol{B}_0 \cdot \boldsymbol{\beta} = \boldsymbol{y} \tag{19}$$

$$\boldsymbol{S}_1 \cdot \boldsymbol{B}_1 \cdot \boldsymbol{\beta} \geq \boldsymbol{0} \tag{20}$$

$$\boldsymbol{S}_2 \cdot \boldsymbol{B}_2 \cdot \boldsymbol{\beta} \geq \boldsymbol{0} \tag{21}$$

with $\boldsymbol{W}$, $\boldsymbol{B}_0$, $\boldsymbol{B}_1$, and $\boldsymbol{B}_2$ as defined above.

The diagonal matrix $\boldsymbol{S}_1$ ($\boldsymbol{S}_2$) is used to specify the signs of the first-order (second-order) differences along $\boldsymbol{y}$:

$$\boldsymbol{S}_1 := \operatorname{diag}(\boldsymbol{s}_1) \tag{22}$$

$$\boldsymbol{S}_2 := \operatorname{diag}(\boldsymbol{s}_2) \tag{23}$$

with $\boldsymbol{s}_1$ ($\boldsymbol{s}_2$) the vector of signs (+1 or -1) for the first-order (second-order) difference. The signs in $\boldsymbol{s}_1$ ($\boldsymbol{s}_2$) are chosen such that the fitted function satisfies the shape described above. The elements of $\boldsymbol{s}_1$ corresponding to the concave segment are set to zero, so that the corresponding inequality constraints are satisfied regardless of the values of $\boldsymbol{\beta}$. Note that the constraints for the first derivative can be simplified further by only including the constraints for the left-most and right-most data point (lowest and highest value for x). All other constraints on the first derivative will be inactive as long as all other constraints are satisfied. As a result, the number of non-redundant constraints equals $N - 1 + 2 = 50$, one more than the number of dimensions (N).

There are $M = (N - 4)(N - 5) = 1980$ pairs of vectors that satisfy the above requirement. Fig. 10 illustrates this with 3 exemplary solutions of the above QP. Fig. 11 shows the objective function (SSR) as a function of the location of the shape changes. As before, the cells at left border and top border correspond to simpler shapes, consisting of two segments only. The very top-left cell corresponds to a concave shape over the whole domain, while constraining the function's slope to be positive (negative) for lowest (highest) value of x.

We now aim to select a parsimonious model among the $M = (N - 4)(N - 5)$ fitted functions. As before, the GORIC is used for this purpose. In this case, $K = 10000$ samples are used to compute the degrees of freedom $\bar{D}$. The resulting values for $\bar{D}$ are shown in Fig. 12. The values range between 3.3 and 5.2, again showing that application of shape constraints is an effective way to control the complexity of a spline-based model. Not surprisingly, the range of the DOF for these functions, for which both the 1st and 2nd derivative are constrained, is lower than the range obtained for functions with constraints for the 1st or the 2nd derivative only (see Fig. 3 and Fig. 7). More surprising is that the DOF for segmentations with a short concave segment (diagonal line) or a long concave segment (top-left corner) is higher than the DOF obtained for models with in-between lengths for the concave segment, which are shown in the center of coloured section of the heatmap in Fig. 12. The DOF and SSR values are combined to compute the GORIC (as above). The GORIC values are shown in Fig. 8. The selected model, i.e. the model with minimum GORIC value, is illustrated in Fig. 9.
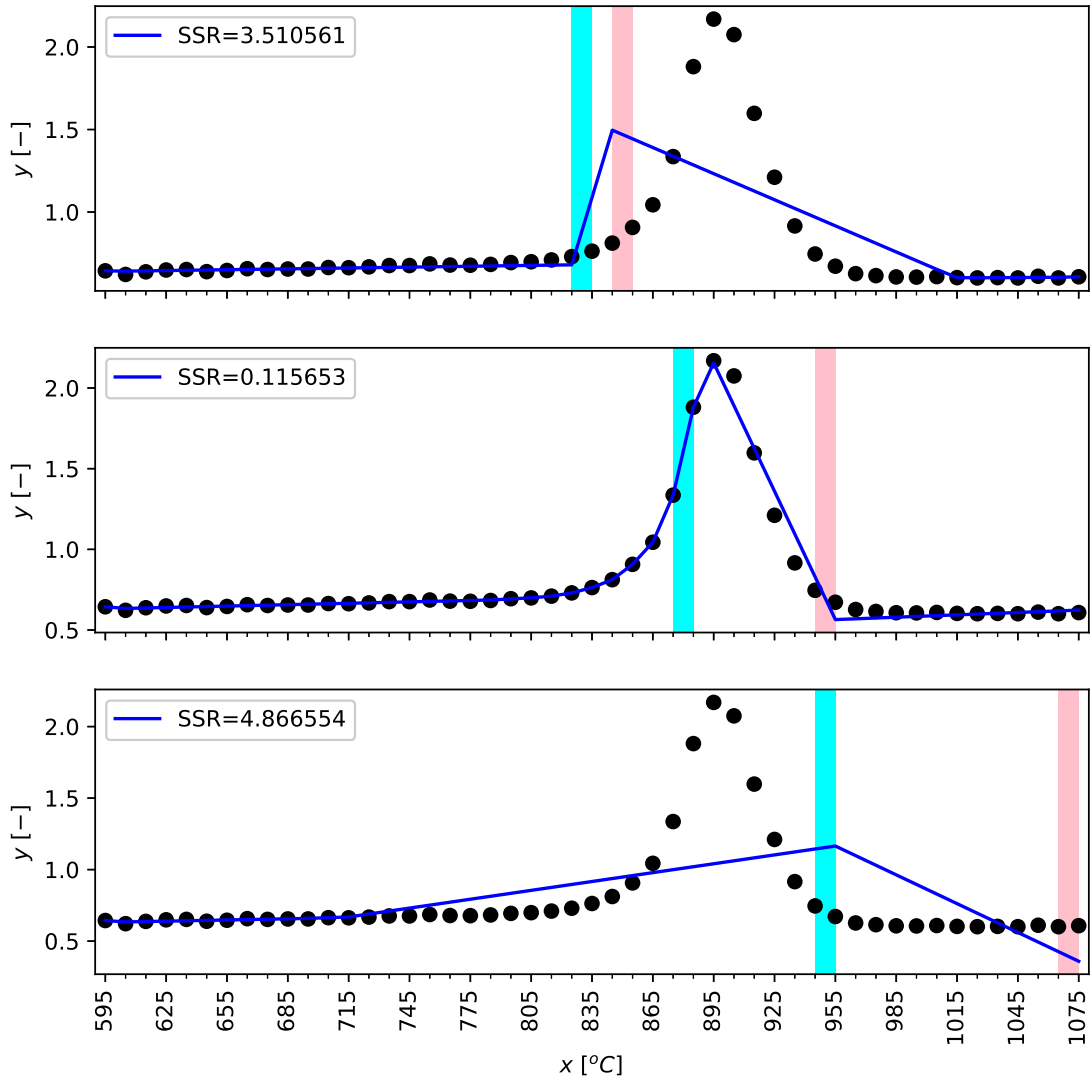
Figure 10. Exemplary fits of three shape constrained piece-wise linear functions. Colored bars indicate the selected location of the change of shape.
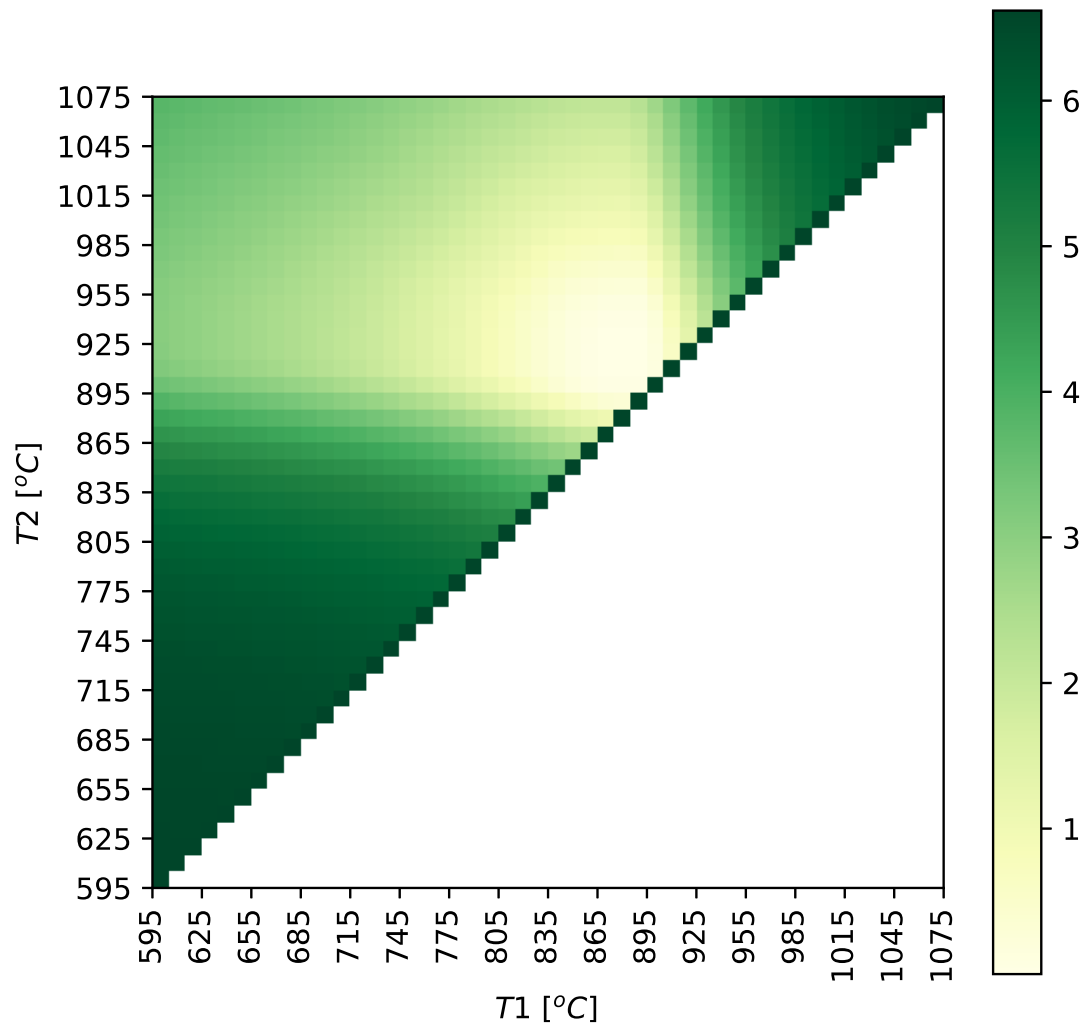
**Figure 11. Heatmap of the sum of squared residuals (SSR) as a function of the locations of shape changes.**
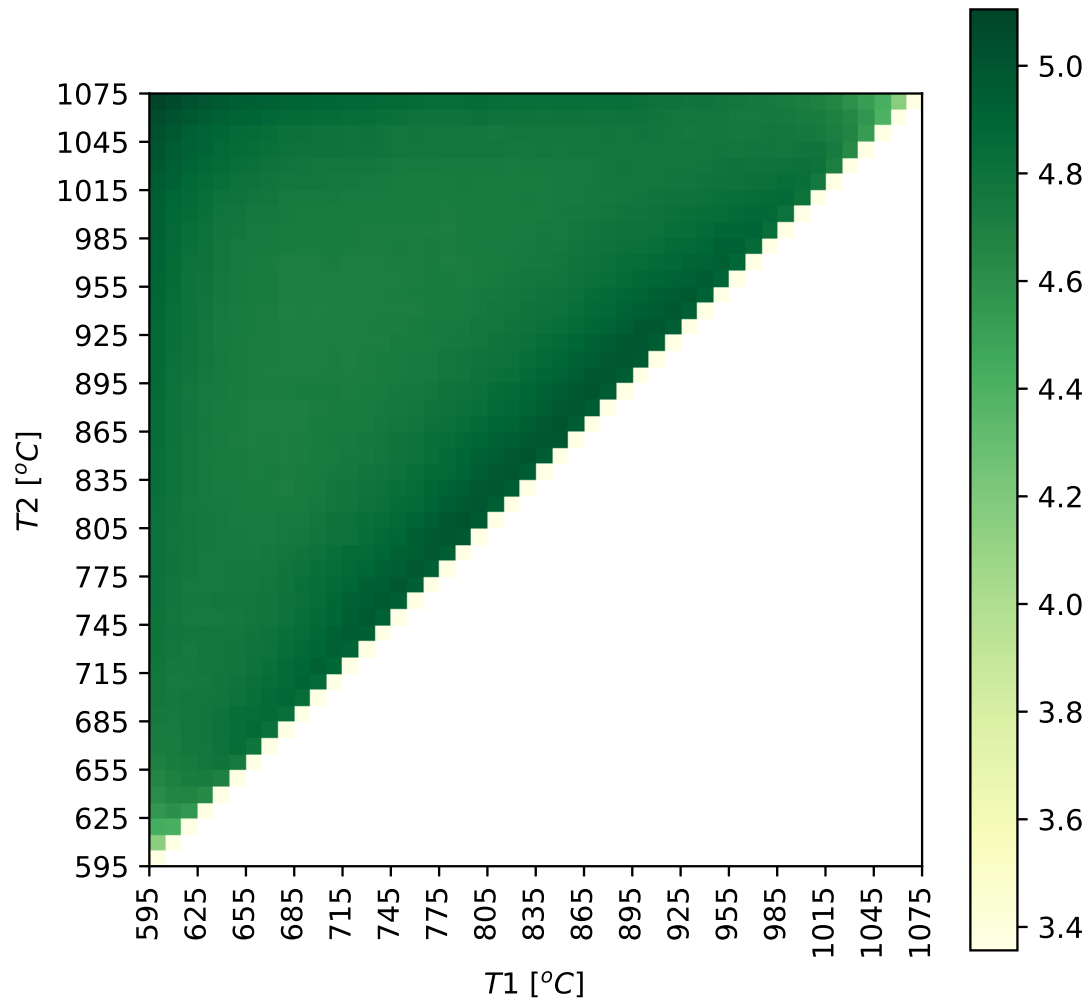
Figure 12. Heatmap of the degrees of freedom (DOF) as a function of the locations of shape changes.
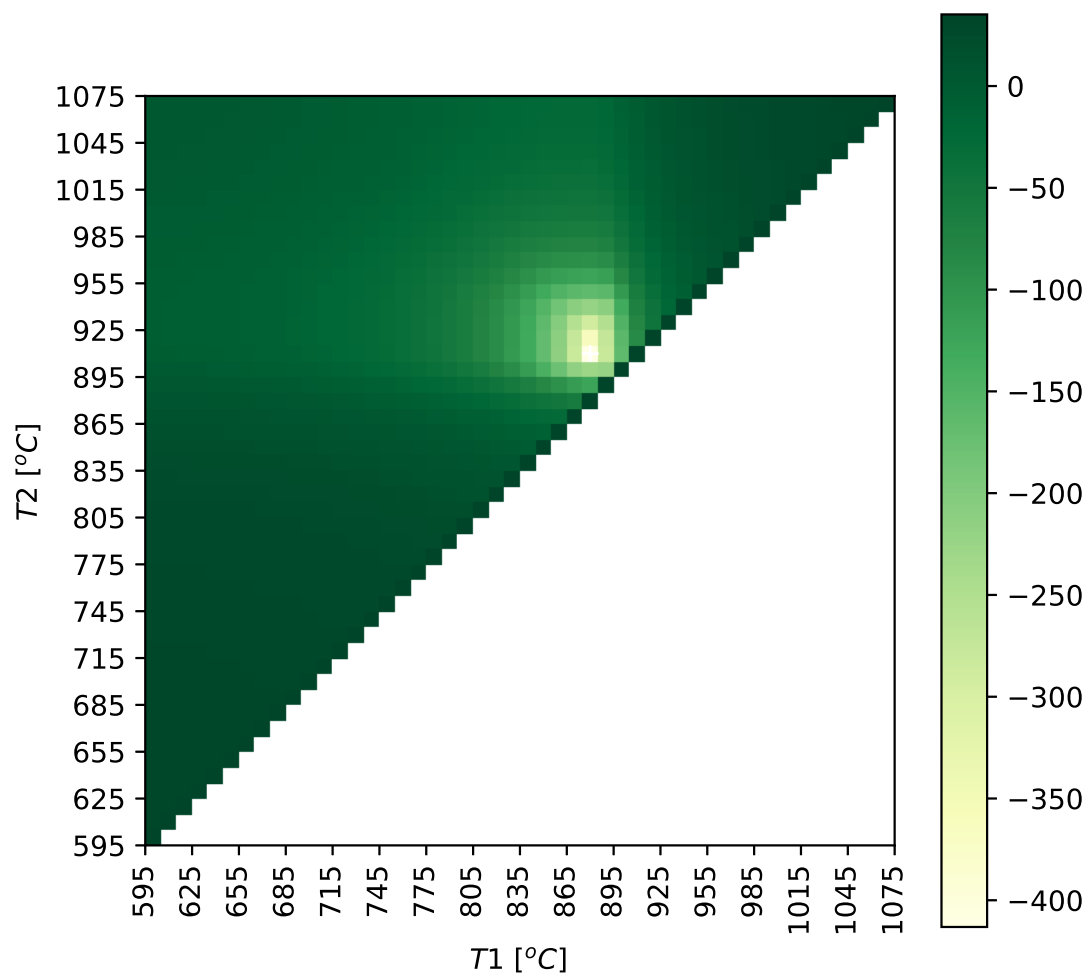
Figure 13. Heatmap of the generalized order restricted information criterion (GORIC) as a function of the locations of sign changes in the second derivative.
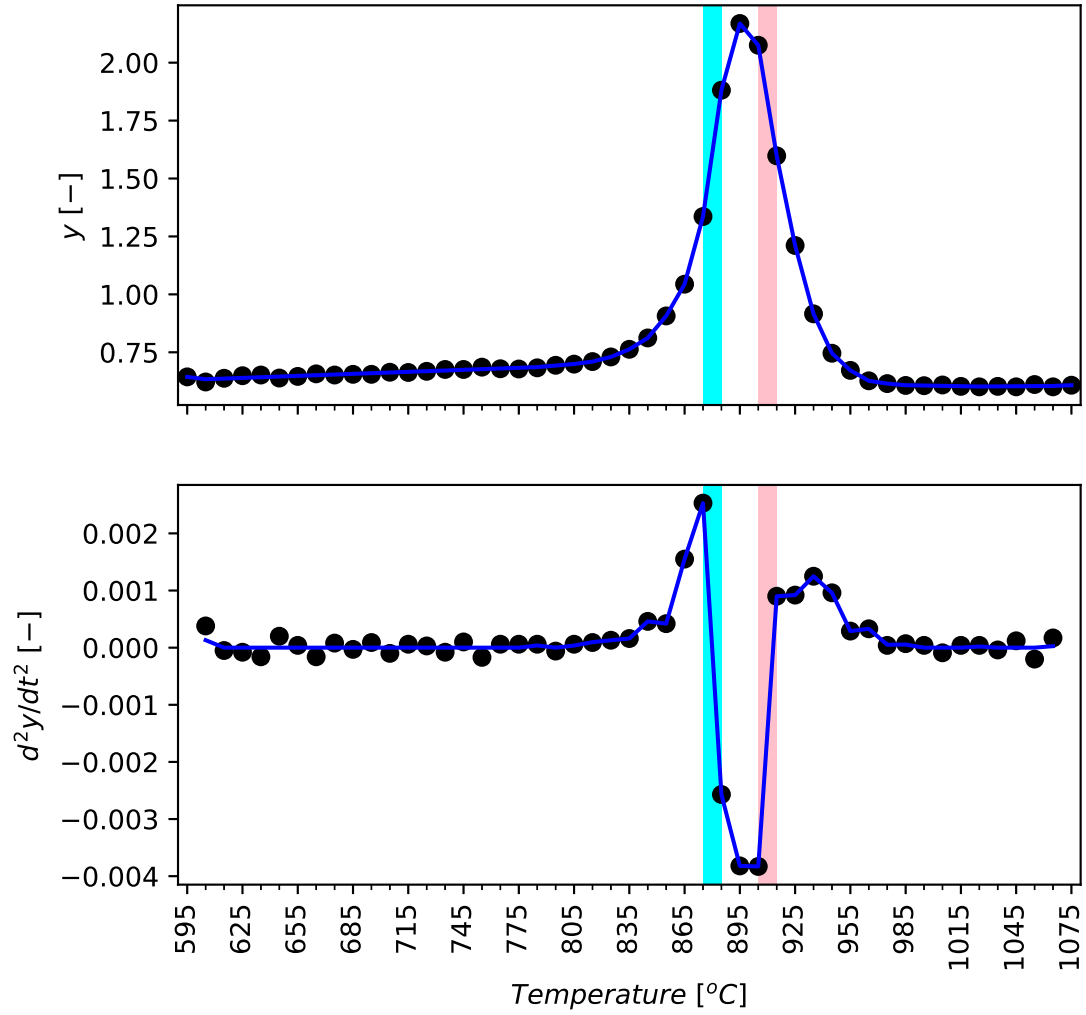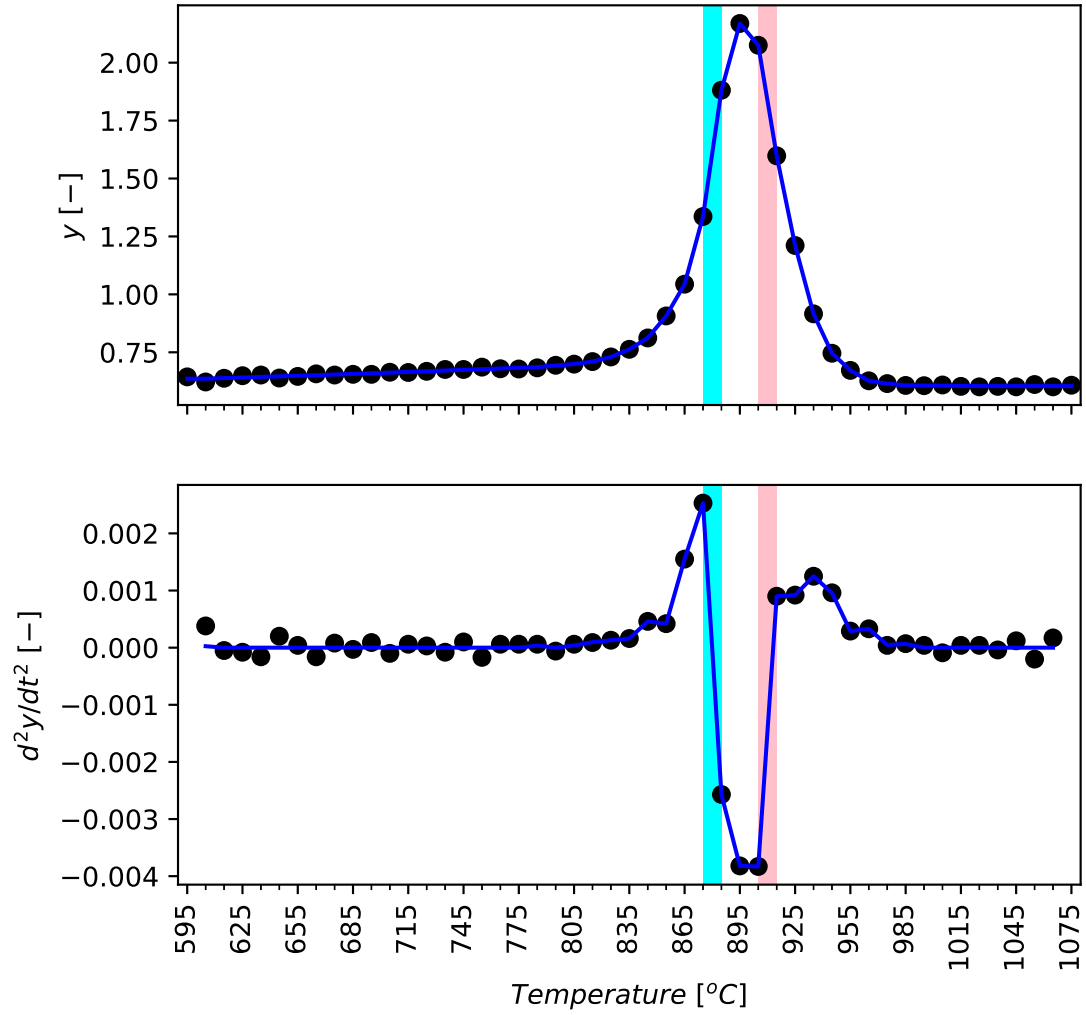
Figure 14. Selected model. Top: data (dots) and fitted function (line); Bottom: first-order differences of the data (dots) and slopes of fitted piece-wise linear function (line). Colored bars indicate the selected locations of the change of sign in the second derivative.

2.5 Constraining the function value

A last demonstration is provided by constraining the function value. We consider a segmentation of the data into three segments by fitting a piece-wise linear function whose sign of the fitted function value at the values in x is constrained to be below, above, and below 1. This leads to an the simplest QP program so far:

$$\hat{\beta} := \arg\min_{\beta} \sum_{i}^{N} (y_i - \tilde{y}_i)^2 = (y - \tilde{y})^T \cdot W \cdot (y - \tilde{y}) \tag{24}$$

$$B_0 \cdot \beta = y \tag{25}$$

$$S_0 \cdot B_0 \cdot \beta \geq 1 \tag{26}$$

The diagonal $N \times N$-dimensional matrix S_0 is used to specify the signs of the second-order difference along y:

$$S_0 := \text{diag}(s_0) \tag{27}$$

with s_0 the N-dimensional vector of signs (+1 or -1) for the second-order difference. These signs are chosen such that the fitted function satisfies the below-above-below shape described above.

There are $M = (N-2)(N-3) = 2162$ vectors that satisfy this requirement for s_0. Fig. 15 shows 3 examples of such a vector and the associated solution of the above QP. Fig. 16 shows the objective function (SSR) as a function of the location of the sign changes in s_0. Note that the cells at left border and top border correspond to simpler shapes, namely above-below (left border) and below-above (top). The very top-left cell corresponds to an above shape over the whole domain.

In this case, an analytic solution is available to computed the expected DOF, $\bar{D}$. Because the matrix W, B_0, and $S_0 \cdot B_0$ are all diagonal, it follows that $\bar{D} = 1 + N/2 = 25.5$. This is true regardless of the chosen s_0, so that the model complexity is the same for every possible segmentation. Consequentially, the best possible segmentation corresponds to the function with the lowest SSR. This model is illustrated in Fig. 17.
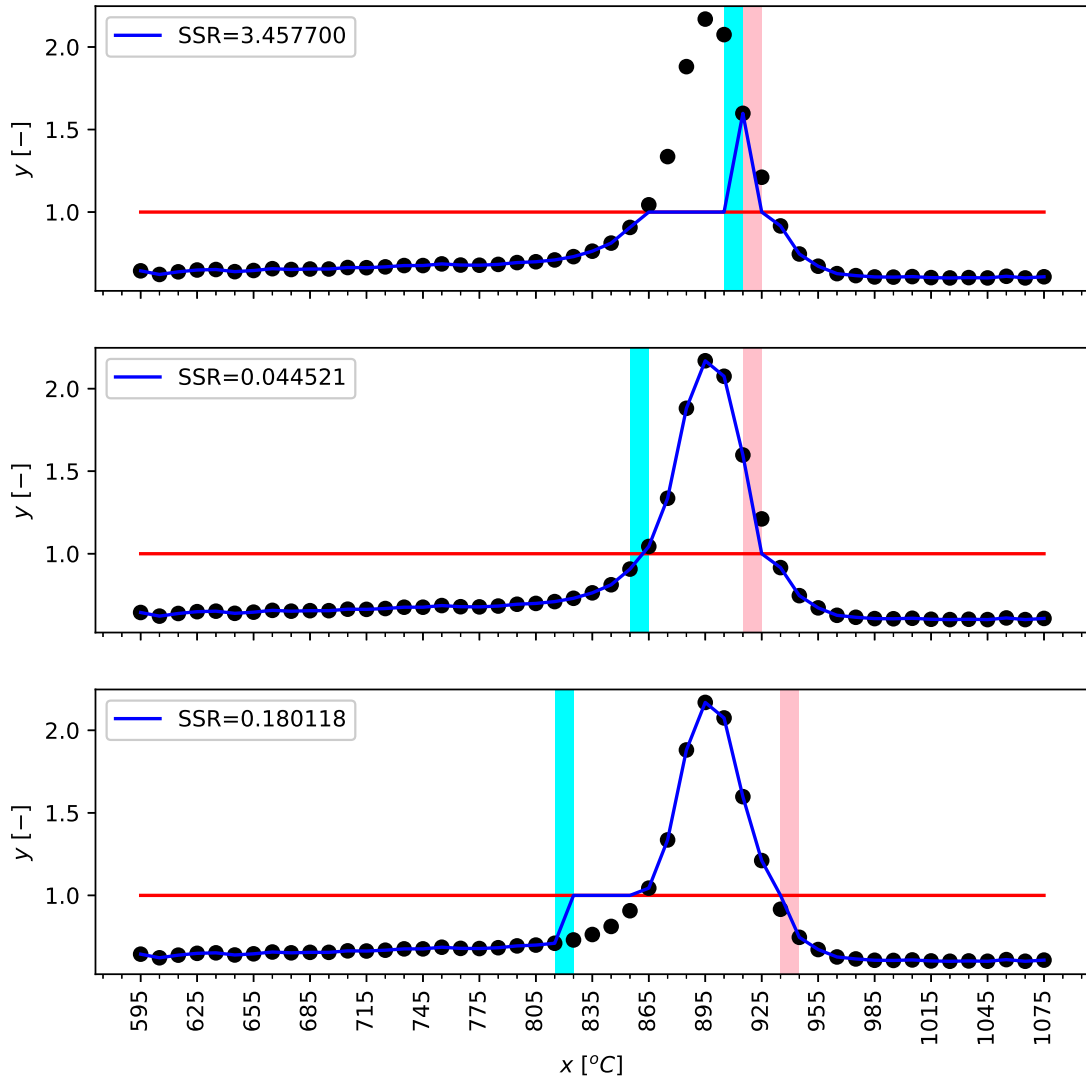
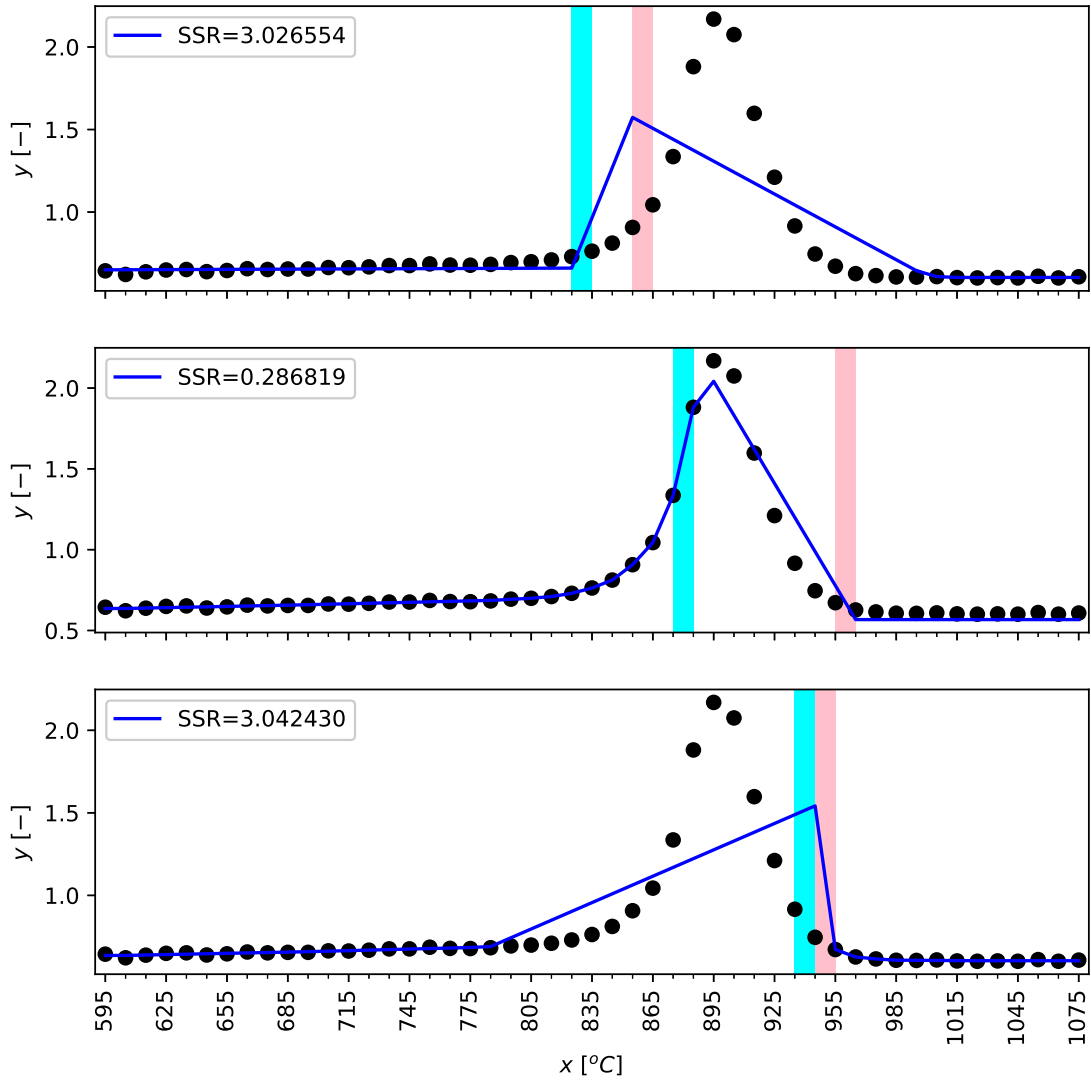**Figure 15.** Exemplary fits of three shape constrained piece-wise linear functions. Colored bars indicate the selected segmentation.

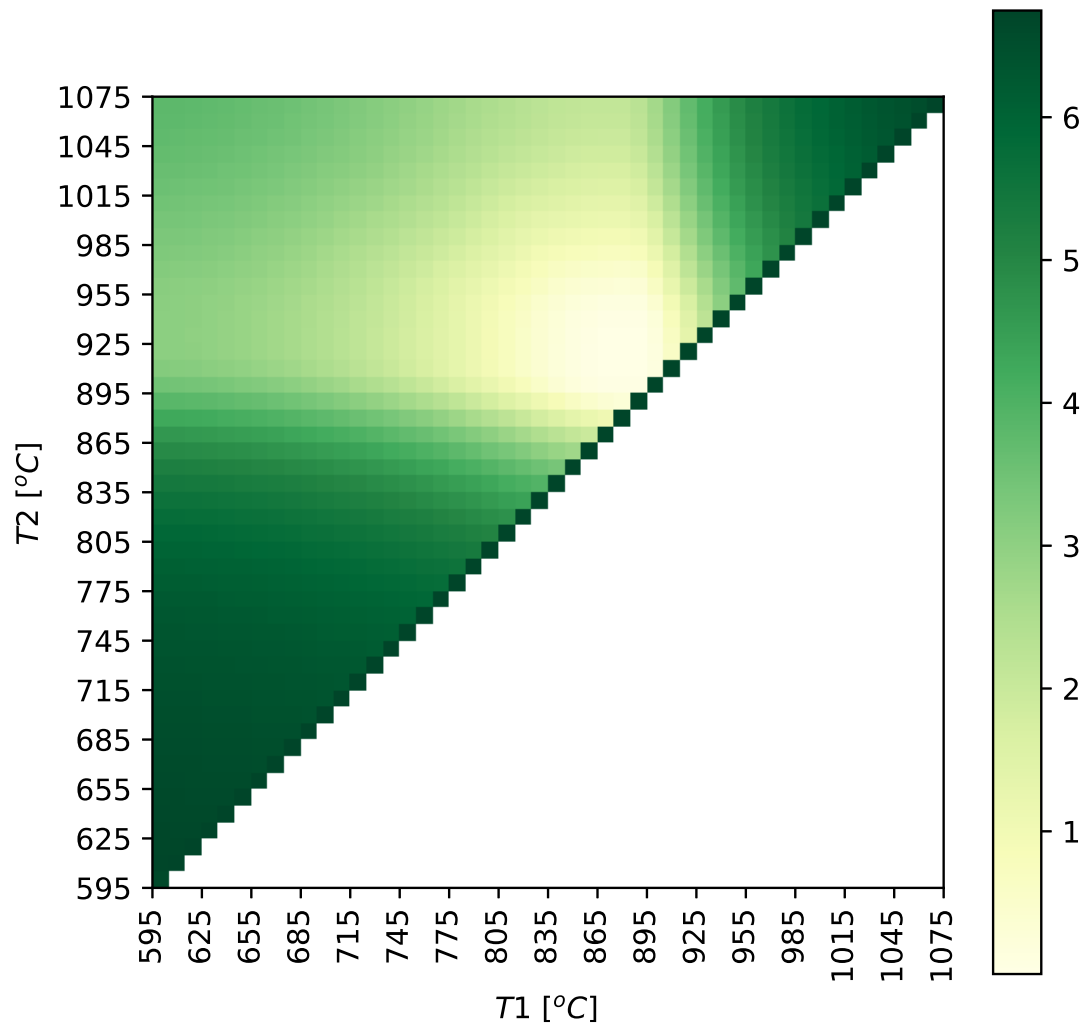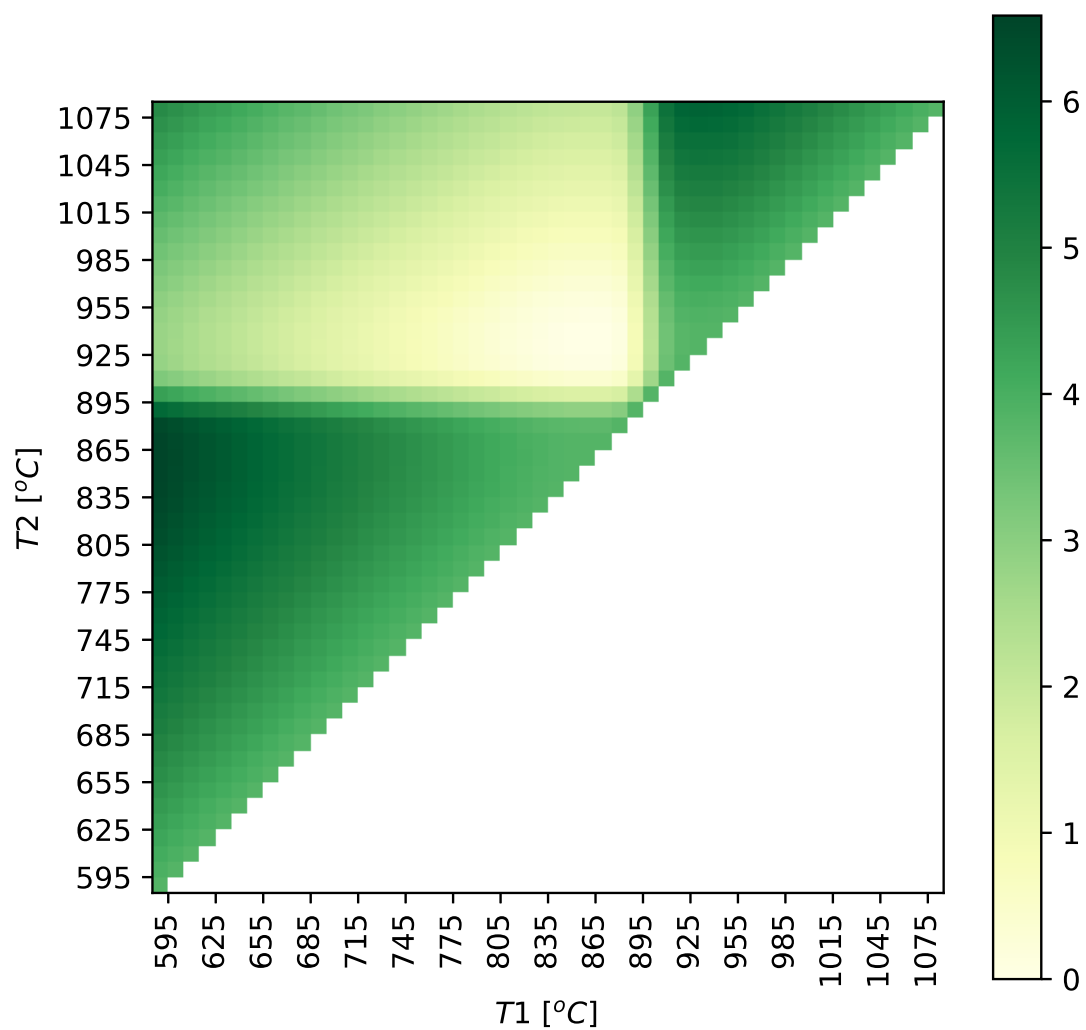**Figure 16.** **Heatmap of the sum of squared residuals (SSR) as a function of the locations of sign changes in the second derivative.**
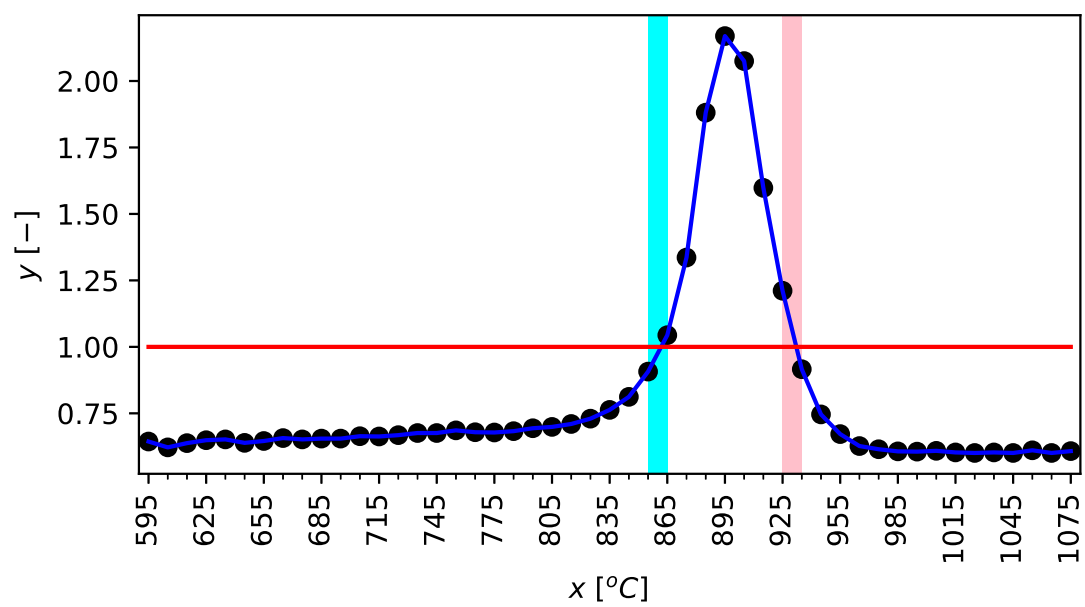
Figure 17. Selected model. Data (dots) and fitted function (line).

3. Computation of the Expected Degrees of Freedom

The computation of the expected degrees of freedom is not trivial in the general case. The next sections describe the methods known to the author.

3.1 Quadratic Programs

It is assumed that the model fitting problem can be described as a convex QP. It is further assumed that the problem is feasible (non-empty solution set). This means the parameter estimation problem can be written as the minimization of a quadratic objective function and a set of linear equality and inequality constraints (QP0):

$$\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta}} J = \left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\right)^T \cdot \boldsymbol{W} \cdot \left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\right) \tag{28}$$

$$\boldsymbol{A}_{eq} \cdot \boldsymbol{\beta} = b_{eq} \tag{29}$$

$$\boldsymbol{A}_{ineq} \cdot \boldsymbol{\beta} \le b_{ineq} \tag{30}$$

where $\tilde{\boldsymbol{\beta}}$ represents the optimal solution for $\boldsymbol{\beta}$ in absence of any equality or inequality constraints.

For the purpose of this report, the studied QPs are restricted to the set of QPs that can be transformed equivalently to the following form (QP1):

$$\hat{\boldsymbol{\delta}} := \arg\min_{\delta} J = \left(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}\right)^T \cdot \left(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}\right) \tag{31}$$

$$\boldsymbol{A} \cdot \boldsymbol{\delta} \le 0 \tag{32}$$

where the vector $\tilde{\boldsymbol{\delta}}$ represents the optimal solution for $\boldsymbol{\delta}$ in absence of any (inequality) constraint. In QP1, $\boldsymbol{A}$ is always a matrix with the number of rows (K) equal to or larger than the number of columns (N). We further assume that rows in $\boldsymbol{A}$ representing constraints that are always inactive or can only be weakly active have been removed. Geometrically speaking, this means that $\boldsymbol{A}$ represents a pointed polyhedral cone with its point at the origin.

The following transformation steps are useful to transform a general-form QP (QP0) into the desired form (QP1):

1. Centering. The coordinate system is shifted so that right-hand sides of the constraints become equal to null vectors.

2. Whitening 1. The coordinate system is rotated and scaled so that the curvature of the objective function is constant (in every direction and in every point), thus making $\boldsymbol{W}$ equal to the identity matrix.

3. Dimension reduction. The number of parameters are reduced by:

   (a) Accounting for equality constraints by projecting into the null space of the equality constraints. The number of removed parameters equals the row rank of A_{eq}.

   (b) Accounting for unconstrained directions (null space of A_{ineq}) by removing these directions from the problem (no effect on the number of constraints due to previous steps).

   (c) Remove inequality constraints that are inactive always or weakly active always.

4. Whitening 2. The coordinate system is rotated (not scaled) so that the constraint matrix A contains a maximum number of rows that have only one non-zero element. Following this, rows and columns of A are permuted so that A becomes block-diagonal with a diagonal matrix block of maximal size.

Note that the first step may not be feasible always, thus restricting the set of models covered in this report. The other steps can be executed for any general-form convex QP. In what follows, we assume that the QP has been transformed into the QP1 form.

3.2 Computational methods

Three methods can be considered for computation, depending on the form taken by A. These are described next from low to high computational complexity, corresponding to increasing generality. All methods are based on the same general approach. One assumes a null model, also called null hypothesis, which assumes a distribution for random instances of $\tilde{\delta}$. In this work, these vectors are drawn from a standardized multivariate normal distribution (zero mean, identity covariance matrix), as is typical [Silvapulle and Sen, 2011, Kuiper et al., 2012]. The expected active number of constraints for an alternative model (or alternative hypothesis), specified with A, is then computed with data simulated according the null model. A small yet meaningful extension obtained through this work is that the number of constraints (K) is allowed to exceed the number of dimensions (N), unlike models considered in earlier works [Silvapulle and Sen, 2011, Kuiper et al., 2012].

3.2.1 Implementation

All methods are implemented with Python (3.7), except for the software by Kuiper and Hoijtink [2013], which is available as a stand-alone executable file programmed in Fortran 90. The latest version of the Python software is released as open software (Apache License 2.0) and is currently available at https://code.ornl.gov/2kv/goric.

3.2.2 Case I: Diagonal case

A extremely efficient method is available if the matrix A is square and the number of non-zero elements and every row and column is equal to 1. In this case, A has been made diagonal in the last transformation step discussed above ($A = A \cdot I$). In case I, the probability of one constraint being active is completely independent of the probability of another constraint being active. As a result, the number of active constraints follow a binomial distribution, $B(n, p)$, with the number of trials equal to the number of dimensions ($n = N$) and equal probability for any constraint being active or inactive ($p = 1/2$). Consequently, the expected DOF is equal to mean of the binomial distribution: $\bar{C} = N/2$.

3.2.2.1 Illustration. This method applies to example discussed in Section 2.5 above where $N = 49$. Figure Fig. 18 shows the binomial distribution reflecting the expected distribution of the number of active constraints k with the associated mean $\bar{C}$ indicated with a vertical line. In this case, it follows that $\bar{C} = N/2 = 24.5$. This is true for every segmentation considered in Section 2.5, i.e. the results is independent of the signs of the values on the diagonal of A.

3.2.3 Case II: Low-dimensional setup ($N \leq 3$)

An analytic solution is available in non-diagonal cases for low dimensions ($N \leq 3$). These are discussed next.

3.2.3.1 Case II – 1D ($N = 1$). In this case, there is at most one inequality represented by A. Since A is 1×1-dimensional, it is covered under Case I above. For randomly generated instances of $\tilde{\delta}$, drawn from an standard univariate normal distribution (zero mean, unit variance), it is fairly trivial to see that the expected number of active constraints is again $\bar{C} = N/2 = 1/2$.

3.2.3.2 Case II – 2D ($N = 2$). In this case, A is a square 2×2-dimensional matrix. Note that cases with $N = 2$ and one constraint only can be reduced to the case $N = 1$ (see paragraph on dimension reduction above). The matrix represents a two-dimensional cone, given as column vectors of a square 2×2-dimensional matrix Q. When a randomly sampled instance of $\tilde{\delta}$ falls into this cone, there are zero active constraints ($k = 0$). When the sample is outside of the cone, the sample can land in one of three regions. One region is defined by the polar cone, given as column vectors of a square 2×2-dimensional matrix P. Note also that $P = A.T$. If the sample is in this region, the number of active constraints equals 2 ($k = 2$). There are two remaining regions, each defined by taking one column of Q and P. If the sample lands in one of these regions, only one of the constraints is active ($k = 1$).

The relative size of each region can be measured as their interior angle. This corresponds to the expected number of samples landing in each region, assuming the same distribution as before (multivariate normal, zero mean, unit variance). The two regions where $k = 1$ each have an angle of $\pi/4$ radians.
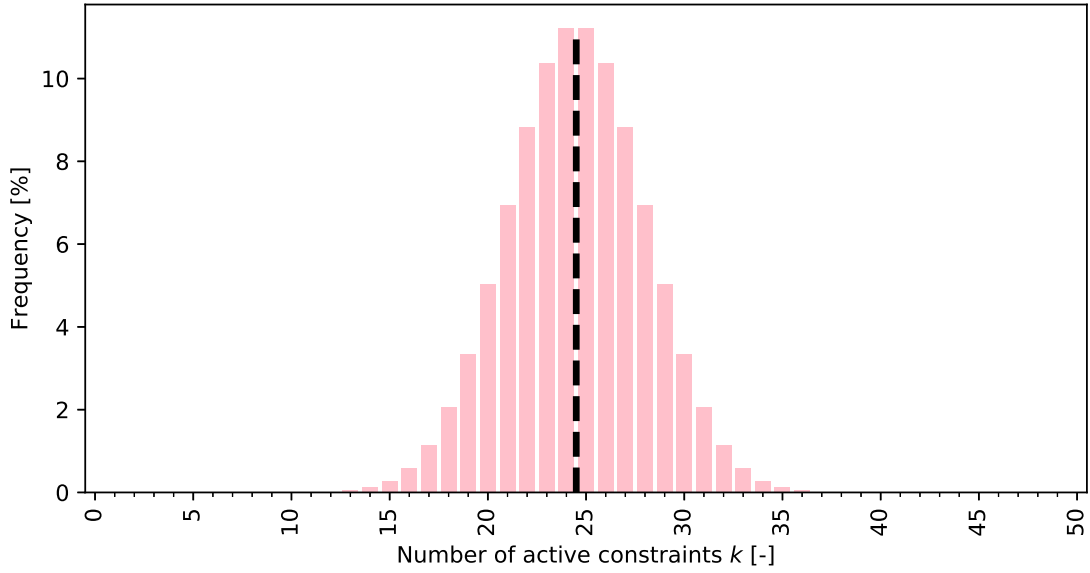
Figure 18. Case I: Binomial distribution of the active constraints given randomly generated vectors $\tilde{\delta}$ and a diagonal matrix A.

As the number of regions is small ($R = 2^N = 4$) and computing their interior angle (γ_r) is trivial, one can compute the expected number of active constraints analytically:

$$\bar{C} = \frac{1}{2\pi} \sum_{r=1}^{R} \gamma_r \cdot C_r = \frac{1}{2\pi} \sum_{k=0}^{N=2} \theta_k \cdot \binom{N}{k} \tag{33}$$

with C_r the number of active constraints for the rth region. The expression can also be reformulated with three angles θ_k, instead of four, each associated with a feasible number of active constraints k ($k = 0, ..., K$). These angles are multiplied with the corresponding binomial coefficient (i.e., N choose k), as in the equation above.

3.2.3.3 Illustration. The 2D case is illustrated with an instance of QP1 where A takes the following form:

$$A = \begin{bmatrix} +1 & 0 \\ -1 & +1 \end{bmatrix} \tag{34}$$

In Figure Fig. 19, one can see the generating vectors of the four polyhedral regions (all blue for feasible region, all red for region where all constraints active). The feasible set, where $k = 0$, is a cone with an interior angle of $\pi/8$. Given that the regions where $k = 1$ each have an angle $\pi/2$, the remaining polar cone has angle $3\pi/8$. Consequently, the expected number of active constraints equals $\bar{C} = 5/8 \cdot N = 5/4 = 1.25$.
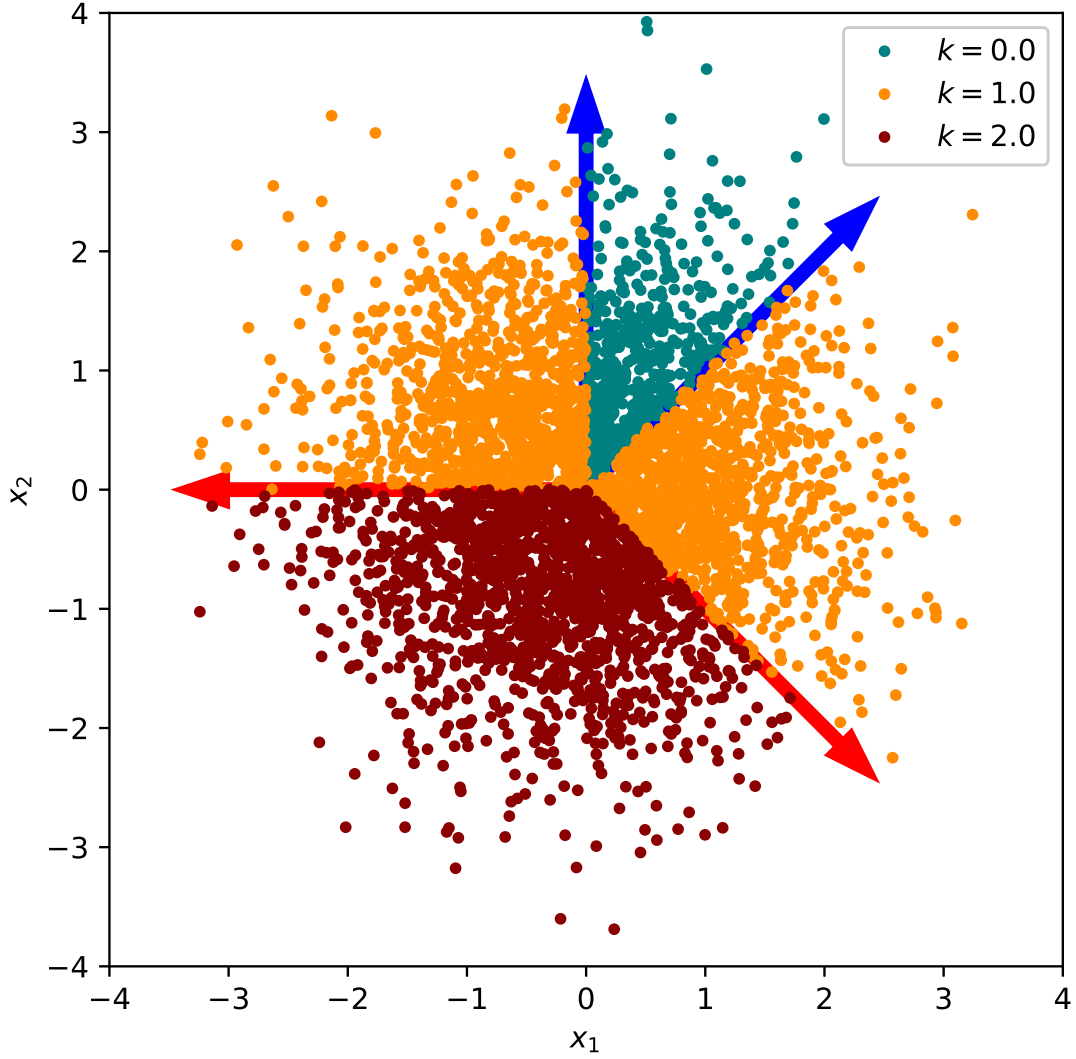
Figure 19. Case II – 2D example ($N = 2$, $K = 2$)**: Scatter plot showing the generating vectors of the polyhedral regions (Q: blue arrows, P: red arrows). A number of randomly sampled vectors $\tilde{\delta}$ are shown for illustration with colours indicating the number of active constraints (k).**

3.2.3.4 Case II – 3D ($N = 3$). Cases with $K < 3$ constraints can be transformed to one of the above cases ($N = 1$ or $N = 2$). In all remaining cases, A is an $K \times 3$-dimensional matrix with $K \geq N$ representing a three-dimensional cone. This cone has K generating vectors, given as column vectors of Q. As in the previous case ($N = 2$), this cone represents the region where zero constraints are active. One can compute the polar cone also, which is also defined with K generating vectors, the columns of P, and defines the region where $N = 3$ constraints are active. Note also that $P = A.T$. Additional regions are defined by taking $N = 3$ column vectors from P and Q, selecting one vector from Q and two from P. The columns in Q and P are ordered such that each of the K regions is produced by selecting the same $N = 3$ column positions in Q and P first, following by selecting either the column vector Q or P for a given column position, never selecting vectors in the same column position. A sample falling into one of these additional regions will result in one active constraint. Similarly, one can define K more regions by taking one generating vector from P and two from Q. A sample falling in one of these K regions will produce two active constraints. For each of the $2K + 2$ regions, the expected fraction of randomly generated samples (if drawn from a standardized multivariate normal distribution) is proportional to the relative size as measured by its solid angle. Fortunately, the solid angle can be computed analytically in this case, so that one can write:

$$\bar{C} = \sum_{r=1}^{R} \Omega_r \cdot C_r \tag{35}$$

with Ω_r and C_r the solid angle and number of active constraints in each region. Each polyhedral region can exhibit a unique value for the solid angle so that a sum over the $R = 2K + 2$ regions is necessary in general.

3.2.3.5 Illustration. The 3D case is illustrated with an instance of QP1 where A takes the following 5×3-dimensional form:

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ +1 & -1 & 0 \\ 0 & -1 & +1 \end{bmatrix} \tag{36}$$

Note that the second row in A corresponds to a redundant constraint, which is always inactive or weakly active. This constraint can be removed to produce a $K \times N = 4 \times 3$-dimensional matrix.

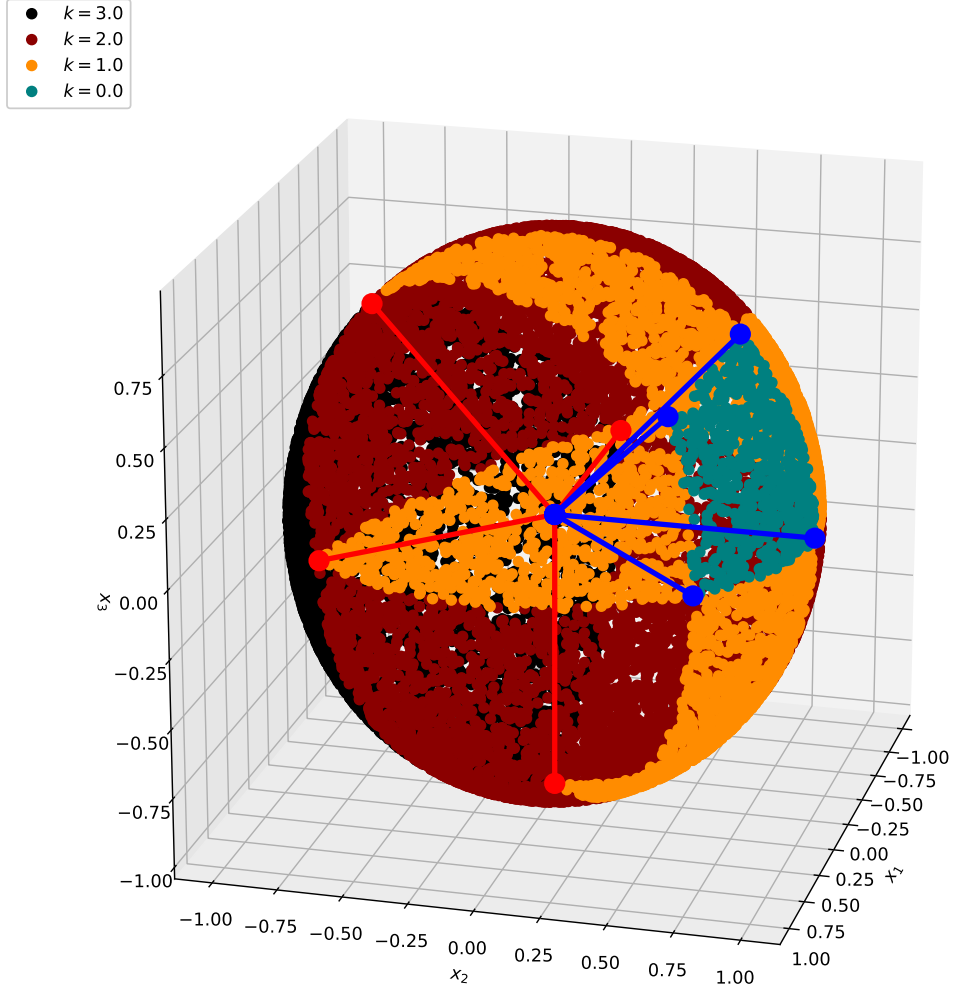**Figure 20. Case II – 3D example** ($N = 3$, $K = 4$)**: Scatter plot showing the generating vectors of the polyhedral regions** (Q**: blue lines,** P**: red lines). A number of randomly sampled vectors** $\tilde{\delta}$ **on the sphere are used to illustrate the locations of the polyhedral regions with colors indicating the number of active constraints** (k**).**

3.2.4 Case III: General Setup

When none of the above cases apply, one needs to resort to approximation methods. Two methods are known.

3.2.4.1 Taylor series. As in the cases above, it is feasible to define R non-overlapping polyhedral cones in the N-dimensional case, each with their own number of active constraints C_r. The challenge then is to compute the solid angles of each of these regions. A Taylor series expansion is available for this purpose [Aomoto, 1977, Ribando, 2006, Beck et al., 2009]. This approach has not been tested in the frame of this study. The number of polyhedral regions equals $R = 2^N$ when $N = K$ and thus increases exponentially in the worst-case scenario. For this reason, it is hypothesized that increasing N reduces the convergence rate of the series rather fast.

3.2.4.2 Monte Carlo. An alternative method that is easier at a conceptual level is to generate M random samples $\tilde{\delta}_j$ and solve QP1. For each sample, j ($j = 1, \ldots, M$), count the number of constraints that are active (C_j). Then, after solving QP1 for a sufficiently large number of samples, one can estimate the expected number of active constraints as:

$$\bar{C} = \frac{1}{M} \cdot \sum_{j=1}^{M} C_j \tag{37}$$

Evaluating the number of active constraints (C_j) constitutes a practical numerical challenge. When using a typical solver, the number of active constraints can be identified by counting the Lagrange multipliers that are (strictly) positive. Lagrange multipliers are negative or equal to zero represent inactive constraints. Because computers do not exhibit absolute precision, it is generally difficult to determine whether a Lagrange multiplier is exactly equal to zero or not. As a result, determining the active constraints requires the use of a non-zero tolerance (e.g., $\epsilon = 10^{-6}$) and leads to an imprecise evaluation of the number of active constraints. Alternative approaches are desirable as *(a)* it is not obvious how to set the value of this tolerance and *(b)* the optimal tolerance value likely depends on the specific instance of QP1 under study.

3.2.4.3 Monte Carlo – QP. One possible solution consits of using the Goldfarb/Idnani dual quadratic programming algorithm [Goldfarb and Idnani, 1983]. This algorithm identifies the active constraints during execution and thus permits counting the active constraints upon convergence to the final solution. This approach is used as the basis for a stand-alone program to compute the GORIC. It is based on an efficient implementation in Fortran90. Although the GORIC is said to be restricted to cases where $K \leq N$, the implementation of [e.g., Kuiper et al., 2012] appears to deliver accurate results even when $K > N$ (not shown). The stand-alone nature of this software makes integration in other software platform somewhat challenging. For this reason, the same approach is implemented here based on the *quadprog* package for Python [McGibbon and Turlach, 2019], which, in turn makes use of efficient Cython code [Behnel et al., 2011].

3.2.4.4 Monte Carlo – LP. To avoid potential numerical issues with conventional QP solvers, the QP1 problem can be converted into an equivalent linear complementarity problem (LCP). Concretely, the following LCP is solved for α and ζ for the given $\tilde{\delta}_j$:

$$\boldsymbol{Q} \cdot \boldsymbol{\alpha} + \boldsymbol{P} \cdot \boldsymbol{\zeta} - \tilde{\delta}_j = 0 \tag{38}$$

$$\boldsymbol{\alpha} \geq 0 \tag{39}$$

$$\boldsymbol{\zeta} \geq 0 \tag{40}$$

$$\boldsymbol{\alpha}^T \cdot \boldsymbol{\zeta} = 0 \tag{41}$$

where the column vectors of $\boldsymbol{Q}$ are the generating vectors for the polyhedral cone describing the feasible set of QP1 and $\boldsymbol{P}$ are the generating vectors for the polar of the feasible set. As before, the columns of $\boldsymbol{Q}$ and $\boldsymbol{P}$ are ordered such that each region is a polyhedral cone defined by N generating vectors, picking one column from either $\boldsymbol{Q}$ or $\boldsymbol{P}$ for each column position. The number of active constraints for problem QP1 and a given sample $\tilde{\delta}_j$ is then given as the

number of coefficients in ζ that are non-zero. As with the Lagrange multipliers, this leads to issues due to limited numerical precision. To avoid this, a relative comparison of the magnitude of the coefficients in α and ζ can be used instead:

$$C_j = \sum_i H\left(\zeta i, j - \alpha_{i,j}\right) \tag{42}$$

with $H(\cdot)$ the Heaviside function. The last constraint (bilinear complementarity) in the above LCP can be removed by solving the following linear program (LP) instead:

$$\hat{\alpha}_j, \hat{\zeta}_j := \arg \min_{\alpha, \zeta} J = \sum_i^N \zeta_i \tag{43}$$

$$Q \cdot \alpha + P \cdot \zeta - \tilde{\delta}_j = 0 \tag{44}$$

$$\alpha \geq 0 \tag{45}$$

$$\beta \geq 0 \tag{46}$$

This alternative formulation is expected to work due to the following properties of the problem. The polyhedral regions of the input space within which the number of active constraints is constant *(a)* are convex, *(b)* are pointed cones with their point at the origin, *(c)* cover the whole N-dimensional space, and *(d)* are non-overlapping.

3.2.4.5 Monte Carlo - mpLP. Another solution consists of *multi-parametric linear programming* [mpLP, Pistikopoulos et al., 2007] and is specifically aimed at reducing computational costs associated with solving many instances of the same problem. The idea is to provide a function that maps any N-dimensional vector $\tilde{\delta}_j$ (now called parameters) to the corresponding estimates $\hat{\delta}_j$ (solution). For quadratic problems, like instances of QP1, this function is piece-wise linear. The same holds for the LP form discussed above, leading to *multi-parametric linear programming*, which is used here. In the present case (polyhedral cone as feasible space), this function consists of R linear functions, each valid for a region of the input space defined by a set of inequality constraints for the input vector ($\tilde{\delta}$). In addition, the number of active constraints is constant in each region. Consequently, the number of active constraints can be obtained by means of a lookup table. First, find the region that contains the new sample and then report the number of active constraints for that region. This is computationally efficient. Effects of imperfect numerical precision can be minimized by selecting the region for which the distance from the point to the interior of the region is minimal. Unfortunately, the number of regions R scales exponentially with the number of dimensions (N). This means the memory requirements for the lookup table quickly exceed the capacity of typical hardware.

3.2.4.6 Illustration and comparison. The top panel of Fig. 21 shows the obtained values for $\bar{D}$ (= $N - \bar{C}$) and the computational requirements for isotonic regression as a function of the number of dimensions (N). The largest value for N ($N = 49$) corresponds to one of the models considered in Section 2.2. The number of samples equals $M = 4096$. All methods are in close agreement with each other and indicate that applying isotonicity constraints to reduce the degrees of freedom is a very effective method. Indeed, the marginal increase of the degrees of freedom ($\bar{D}$) becomes smaller as N increases. The bottom panel of Fig. 21 shows the time used for computation. For low values of N, the mpLP solution is on par with the analytic solution. However, its efficiency is dominated by all methods that solve each instance of QP1 separately for $N = 16$ and higher. The QP-based methods [Kuiper and Hoijtink, 2013, McGibbon and Turlach, 2019] are more efficient than the LP method. It is assumed that this is at least in part attributed to the efficient implementation of the dual QP algorithm.

Fig. 22 offers another look into the performance of the studied methods. The top panel shows the estimated value for $\bar{D}$ as a function of the number of Monte Carlo samples for one of the isotonic spline functions given above, specifically for $N = 11$. Note that the random samples are the same for all methods, except for the implementation by Kuiper and Hoijtink [2013]. One can see in the top panel that all methods converge fairly rapidly to the same estimate. A number of $N = 2^{14} = 16384$ samples appears sufficient for all computed values for $\bar{D}$ to be less than 0.01 apart. The bottom panel shows the time needed for computation as a function of the number of Monte Carlo samples. As one might expect, all methods exhibit increasing computational costs with increasing sample sizes. However, for

the mpLP method and for the QP-based method implemented by Kuiper and Hoijtink [2013], this is only apparent for $N > 1000$. The QP method by McGibbon and Turlach [2019] appears particularly efficient if the number of samples is small ($N \leq 1000$). For larger samples ($M \geq 10000$), the QP method as implemented by Kuiper and Hoijtink [2013] appears the most efficient. It is suspected that this is because McGibbon and Turlach [2019] only allows to execute a single instance of QP1, thus requiring M restarts of this routine initiated in Python, with random sampling and averaging executed in pure Python. In contrast, the code by Kuiper and Hoijtink [2013] executes the random sampling, all calls to the dual QP algorithm, and averaging within the same program implemented in Fortran90. Although a benefit is thus observed for the implementation by Kuiper and Hoijtink [2013] for large sample sizes, results discussed in Section 2.2 and Section 2.3 are obtained with the implementation of McGibbon and Turlach [2019], mainly to enable straightforward extension and reuse of the produced software.

Figure 21. Expected degrees of freedom for isotonic spline function fit as a function of N (with $M = 4096$) for 5 methods: (a) linear programming (LP), (b) multi-parametric linear programming (mpLP, $N \leq 20$), (c) quadratic programming with stand-alone executable (QP Kuiper), (d) quadratic programming based on Cython code (QP McGibbon), and (e) analytic solution ($N \leq 4$).

3.2.5 Case IV: Mixed methods

Special cases arise when A is a block-diagonal matrix. Each of these blocks can be treated independently in this case due to orthogonality of the constraints. Each of these blocks belongs to one of the above three cases and can thus be treated accordingly. This is expected to provide computational benefits as well as improved accuracy. After doing so, one simply needs to sum the expected number of active constraints for each block. This approach has not been tested yet because this kind of situation is considered rare.
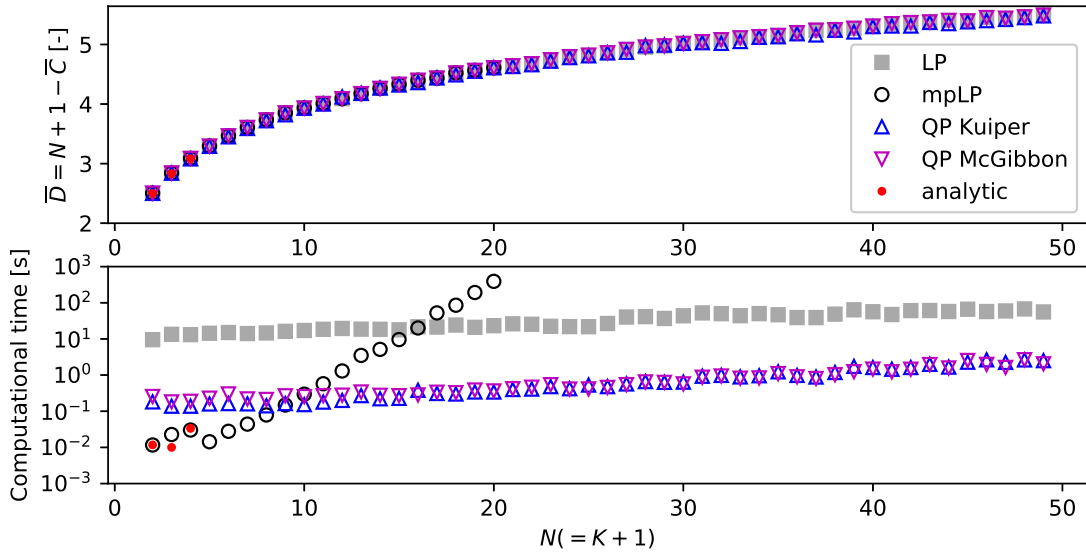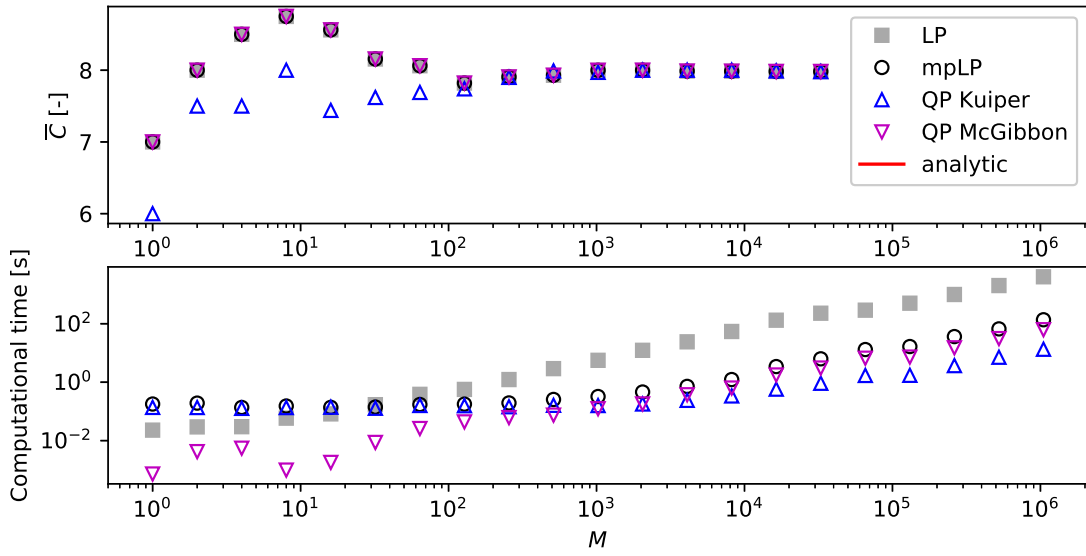
Figure 22. Expected degrees of freedom for isotonic spline function fit as a function of M (with $N = 11$) for 5 methods: (a) linear programming (LP), (b) multi-parametric linear programming (mpLP, $N \leq 20$), (c) quadratic programming with stand-alone executable (QP Kuiper), (d) quadratic programming based on Cython code (QP McGibbon), and (e) analytic solution ($N \leq 4$).

4. Conclusions and Perspectives

This report provides details on practical approaches to compute the generalized order restricted information criterion (GORIC) for inequality-constrained model selection, particularly to select the statistically optimal model among a set of shape-constrained spline functions. The GORIC trades model complexity, measured as the expected degrees of freedom, against the fit of the model, measured as the likelihood of the data conditioned to the fitted model parameters. In the context of the GORIC, the expected degrees of freedom is defined as the number of parameters in the model minus the expected number of active constraints for random instances of the data. Computing the expected number of active constraints constitutes a practical challenge to the application of GORIC. For this reason, the available methods to compute it have been listed and a selection of these has been studied in detail.

4.1   Conclusions

The results shown in this report support the following conclusions, which are not new:

- The expected degrees of freedom can be computed analytically in some restrictive cases: *(a)* low dimensional cases ($K \leq 3$) and *(b)* cases with an isotropic objective function and orthogonal constraints. Methods to do so have been enumerated exhaustively in this report.

- In the general case, computing the GORIC requires a numeric approximation. One method is based on a Taylor series and has not been tested. Another method is based on Monte Carlo integration. Several implementations of this method have been tested.

Some new conclusions are:

- Although the GORIC has been developed with a restriction on the number of constraints ($K \leq N$), results discussed in this report suggest that this restriction is not strictly necessary. It is speculated that this is true whenever the feasible set is a full-dimensional polyhedral set (i.e., a non-empty N-dimensional object). All cases studied in this report satisfy this property. When so, the number of active constraints among the K linear inequality constraints can never exceed the number of dimensions (N). All cases studied in this report satisfy this property.

- Multi-parametric programming is an effective way to reduce the computational effort. However, its benefits are only apparent if a sufficient number of samples is reached. With increasing dimension, the number of samples that need to be requested to see the benefit of the multi-parametric paradigm increases dramatically, probably exponentially. It is however suspected that the distribution of the number of active constraints becomes narrower with increasing dimension, in turn meaning that the number of samples required to obtain an accurate value of the GORIC decreases with increasing dimension.

- The dual quadratic program algorithm, when implemented such that the active constraints are reported, scales well with the size of the studied problems. The same holds for the linear programming approach.

- The implementations of the dual quadratic program algorithm, both making use of efficient code, prove very effective as a way to compute the GORIC in a rapid manner for all cases studied in this report.

4.2   Perspective

This report is focused on the use of GORIC for data series segmentation based on shape-constrained spline functions. Throughout this report, it is assumed that a shape is given, e.g. increasing-decreasing, and that only the change-points, e.g., the location of maxima and minima, need to be determined. The original motivation is however to determine the shape itself also. The idea is to select an optimal shape among many options (e.g., increasing, decreasing, increasing-decreasing, decreasing-increasing, increasing-decreasing-increasing, ...). This problem is more complex given that this requires to identify the number of segments, in addition to identifying their location.

Initial work in this direction made use of subjective model selection criteria [Villez et al., 2013, Thürlimann et al., 2019]. While successful, this requires elaborate tuning of subjective beliefs for each application. A more objective

model selection criterion could avoid the need for such tuning procedures. A provably optimal model selection criterion has not been established for this kind of problem. This report is expected to serve as inspiration to establish such a criterion as well as to implement it as an efficient piece of software.

5. Acknowledgments

References

K. Aho, D. Derryberry, and T. Peterson. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, 95 (3):631–636, 2014.

H. Akaike. Information theory and an extension of maximum likelihood principle. In *Proc. 2nd Int. Symp. on Information Theory (pp. 267-281)*.

K Aomoto. Analytic structure of Schläfli function. *Nagoya Mathematical Journal*, 68:1–16, 1977. doi: 10.1017/S0027763000017839.

M Beck, S Robins, and S V Sam. Positivity theorems for solid-angle polynomials, 2009. URL https://arxiv.org/abs/0906.4031.

S Behnel, R Bradshaw, C Citro, L Dalcin, D S Seljebotn, and K Smith. Cython: The best of both worlds. *Computing in Science and Engineering*, 13:31–39, 2011. doi: 10.1109/MCSE.2010.118.

G. Beliakov. Shape preserving approximation using least squares splines. *Analysis in Theory and Applications*, 16: 80–98, 2000. doi: 10.1007/BF02837633.

C de Boor. *A Practical Guide to Splines*. Springer, 1978.

M. Delecroix and C. Thomas-Agnan. *Smoothing and Regression: Approaches, Computation, and Application*, chapter Spline and Kernel Regression under Shape Restrictions, pages 109–134. Wiley, 2000.

M. Delecroix, M. Simioni, and C. Thomas-Agnan. Functional estimation under shape constraints. *Journal of Nonparametric Statistics*, 6:69–89, 1996. doi: 10.1080/10485259608832664.

S Diamond and S Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(1):2909–2913, 2016.

D Goldfarb and A Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming*, 27(1):1–33, 1983. doi: 10.1007/BF02591962.

X. Gu, J. Mulder, and H. Hoijtink. Approximated adjusted fractional bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, 71(2):229–261, 2018. doi: 10.1111/bmsp.12110.

D. Habtzghi and S. Datta. One sample goodness of fit tests in presence of shape restrictions on the hazard rate function. *Sankhya B - Applied and Interdisciplinary Statistics*, 74-B(2):171–194, 2012.

M. L. Hazelton and B. A. Turlach. Semiparametric regression with shape constrained penalized splines. *Computational Statistics & Data Analysis*, 55:2871–2879, 2011. doi: 10.1016/j.csda.2011.04.018.

K. W. Hipel, A. I. McLeod, and W. C. Lennox. Advances in Box-Jenkins modeling: 1. Model construction. *Water Resources Research*, 13(3):567–575, 1977. doi: 10.1029/WR013i003p00567.

H. Hoijtink, I. Klugkist, and P. Boelen. *Bayesian evaluation of informative hypotheses*. Springer Science & Business Media, 2008.

R. M. Kuiper and H. Hoijtink. A Fortran90 program for the generalization of the order-restricted information criterion. *Journal of Statistical Software*, 54:1–19, 2013. doi: 10.18637/jss.v054.i08.

R M Kuiper, H Hoijtink, and M J Silvapulle. An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, 98:495–501, 2011.

R M Kuiper, H Hoijtink, and M J Silvapulle. Generalization of the order-restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, 142:2454–2463, 2012. doi: 10.1016/j.jspi.2012.03.007.

E. Mammen and T. Thomas-Agnan. Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, 26:239–252, 1999. doi: 10.1111/1467-9469.00147.

R T McGibbon and B A Turlach. Quadprog 0.1.7. 2019. URL https://pypi.org/project/quadprog/.

M. C. Meyer. Inference using shape-restricted regression splines. *Annals of Applied Statistics*, 2:1013–1033, 2008. doi: 10.1214/08-AOAS167.

M. C. Meyer and D. Habtzghi. Nonparametric estimation of density and hazard rate functions with shape restrictions. *Journal of Nonparametric Statistics*, 23:455–470, 2011. doi: 10.1080/10485252.2010.531133.

M. C. Meyer, A. J. Hackstadt, and J. A. Hoeting. Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. *Journal of Nonparametric Statistics*, 23:867–884, 2011. doi: 10.1080/10485252.2011.597852.

J Mulder, H Hoijtink, and I Klugkist. Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140:887–906, 2010. doi: 10.1016/j.jspi.2009.09.022.

D Papp and F Alizadeh. Shape-constrained estimation using nonnegative splines. *Journal of Computational and Graphical Statistics*, 23:211–231, 2014. doi: 10.1080/10618600.2012.707343.

E Pistikopoulos, M Georgiadis, and V Dua. *Multi-Parametric Programming: Theory, Algorithms and Applications, Volume 1*. Wiley VCH, 2007.

J M Ribando. Measuring solid angles beyond dimension three. *Discrete & Computational Geometry*, 36(3):479–487, 2006. doi: 10.1007/s00454-006-1253-4.

M J Silvapulle and P K Sen. *Constrained statistical inference: Order, inequality, and shape constraints (Vol. 912)*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011.

C M Thürlimann, K M Udert, E Morgenroth, and K Villez. Stabilizing control of a urine nitrification process in the presence of sensor drift. *Water Research*, 165(114958), 2019. doi: 10.1016/j.watres.2019.114958.

B. A. Turlach. Shape constrained smoothing using smoothing splines. *Computational Statistics*, 20:81–103, 2005.

K Villez, V Venkatasubramanian, and R Rengaswamy. Generalized shape constrained spline fitting for qualitative analysis of trends. *Computers & Chemical Engineering*, 58:116–134, 2013. doi: 10.1016/j.compchemeng.2013.06.005.