Çankaya University

INTERNATIONAL WORKSHOP ON MATHEMATICAL METHODS IN ENGINEERING

MME 2017

27 - 29 APRIL 2017

A Review of R(C)MARS and (C)MARS with a Comparison Study

Ayşe Özmen

*Gerhard-Wilhelm Weber**

Institute of Applied Mathematics, METU, Ankara, Turkey

* Faculty of Economics, Management and Law, University of Siegen, Germany
School of Science, Information Technology and Engineering, University of Ballarat, Australia
Center for Research on Optimization and Control, University of Aveiro, Portugal

Content

- Introduction
- Regression
- Robustification
- MARS
- CMARS
- Idea of Robust (C)MARS
- Robust Optimization
- RCMARS under Polyhedral Uncertainty
- A Real-World Application by RCMARS
- Conclusion and Outlook on Future Studies

Introduction

learning from data has become very important in every field of science and technology, e.g., in

- ***financial sector***,
- ***quality improvent in manufacturing***,
- **computational biology**,
- **medicine** and
- **engineering**.

Learning enables for doing **estimation** and **prediction**.

Regression

- **Regression is mainly based on the methods of**
  - Least squares estimation,
  - Maximum likelihood estimation.
- **There are many regression models**
  - Linear regression models,
  - Nonlinear regression models,
  - Generalized linear models,
  - Nonparametric regression models,
  - Additive models,
  - Generalized additive models.

Robustification

- The known statistical methods suppose that the input data are exactly known (fixed) in developing models. This introduces a weakness to the methods since, in real-life, both output and input data include uncertainty in the form of noise.

- The presence of noise and **data uncertainty** raises *critical problems* to be dealt with on theoretical and computational side.

- Consequently, after the recent financial crisis, it has been realized that the known statistical models may not give *trustworthy* results.

- ***Robustification*** has started to draw more attention in many fields and the corresponding regression problems usually depend on complex data bases that are affected by noise and uncertainty.

- Robust Optimization has gained lots of concentration from both a theoretical and a practical point of view as a modeling framework for "immunizing" against parametric uncertainties in mathematical optimization.

- It is a modeling methodology to process optimization problems in which the data are uncertain, and are only known to belong to some uncertainty set.

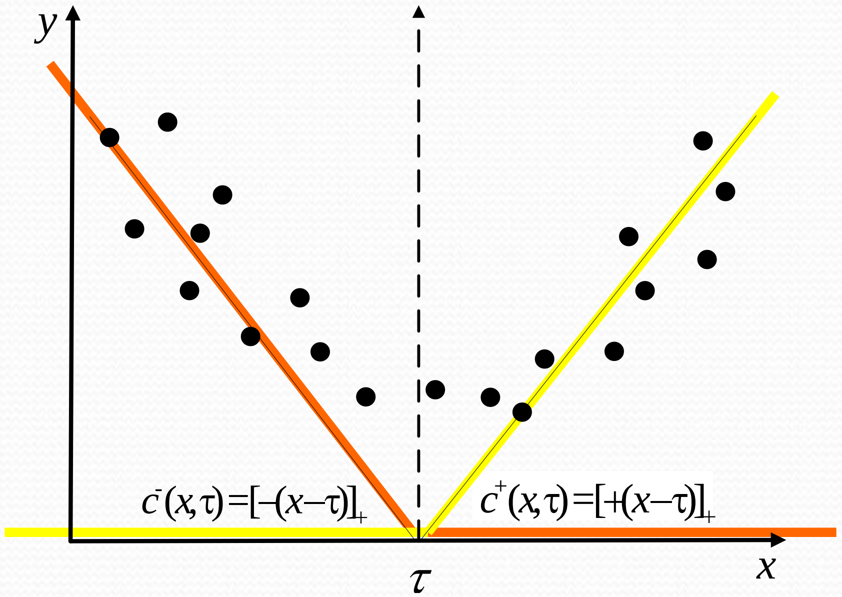
MARS: Multivariate Adaptive Regression Spline

- To estimate general functions of high-dimensional arguments.
- An adaptive procedure.
- A nonparametric regression procedure.
- **No** specific assumption about the underlying functional relationship between the dependent and independent variables.
- Ability to estimate the contributions of the basis functions so that ***additive*** and ***interactive*** effects of the predictors are allowed to determine the response variable.
- Uses expansions in piecewise linear basis functions of the form

$$c^{+}(x,\tau)=[+(x-\tau)]_{+}\,,\qquad c^{-}(x,\tau)=[-(x-\tau)]_{+}\,.$$

$$[q]_{+}:=\max\{0,q\}$$

MARS



- Let us consider $Y = f(\boldsymbol{X}) + \varepsilon,$ $\boldsymbol{X} = (X_1, X_2, \dots, X_p)^T.$
- The goal is to construct reflected pairs for each input $X_j \ (j = 1, 2, \dots, p).$

MARS

- Set of basis functions:

$$\wp := \left\{ (X_j - \tau)_+, (\tau - X_j)_+ \mid \tau \in \left\{ \tilde{x}_{1,j}, \tilde{x}_{2,j}, ..., \tilde{x}_{N,j} \right\},\ j \in \{1, 2, ..., p\} \right\}.$$

- Thus, $f(\boldsymbol{X})$ can be represented by

$$Y = \theta_0 + \sum_{m=1}^{M} \theta_m \psi_m(\boldsymbol{X}^m) + \varepsilon.$$

- $\psi_m \ (m = 1, 2, ..., M)$ are basis functions from $\wp$ or products of two or more such functions; interaction basis functions are created by multiplying an existing basis function with a truncated linear function involving a new variable.
- Provided the observations represented by the data $(\tilde{\boldsymbol{x}}_i, \tilde{y}_i) \ (i = 1, 2, ..., N)$:

$$\psi_m(\boldsymbol{x}^m) := \prod_{j=1}^{K_m} [s_{\kappa_j^m} \cdot (x_{\kappa_j^m} - \tau_{\kappa_j^m})]_+,$$

where $\boldsymbol{x}^m$: subvectors of $\boldsymbol{x}$.

MARS

Two subalgorithms:

(i) ***Forward stepwise algorithm***:

- Search for the basis functions.
- Minimization of some *"lack of fit" criterion*.
- The process stops when a user-specified value M_{max} is reached.
- Overfitting.

This model typically overfits the data; so a *backward* deletion procedure is applied.

(ii) ***Backward stepwise algorithm***:

- Prevents from over-fitting by decreasing the complexity of the model ***without degrading the fit to the data***.

MARS

- *Remove* from the model basis functions that contribute to the ***smallest increase*** in the residual squared error *at each stage*, producing an optimally estimated model $\hat{f}_\mu$ with respect to each number of terms, called μ.

- μ is related with some *complexity* of the estimation.
- To estimate the optimal value of μ:

$$LOF(\hat{f}_\mu) = GCV(\mu) := \frac{\sum_{i=1}^{N} (y_i - \hat{f}_\mu(\boldsymbol{x}_i))^2}{(1 - M(\mu)/N)^2}$$

$M(\mu) := u + d\,K$

N := number of samples,

u := number of independent basis functions,

K := number of knots selected by forward stepwise algorithm,

d := cost of optimal basis.

- ***Alternative:***

PRSS for MARS

$$PRSS := \sum_{i=1}^{N}(y_i - f(\tilde{\boldsymbol{x}}_i))^2 + \sum_{m=1}^{M_{\max}} \phi_m \sum_{\substack{|\boldsymbol{\theta}|=1 \\ \boldsymbol{\theta}^T=(\theta_1,\theta_2)}}^{2} \sum_{\substack{r<s \\ r,s\in V(m)}} \int \alpha_m^2 [D_{r,s}^{\boldsymbol{\theta}} \psi_m(\boldsymbol{t}^m)]^2 d\boldsymbol{t}^m$$

$$V(m) := \left\{\kappa_j^m \mid j = 1, 2, ..., K_m\right\}$$

$$\boldsymbol{t}^m := (t_{m_1}, t_{m_2}, ..., t_{m_{K_m}})^T$$

$$\boldsymbol{\theta}^T = (\theta_1, \theta_2)$$

$|\theta| := \theta_1 + \theta_2$, where $\theta_1, \theta_2 \in \{0,1\}$

$$D_{r,s}^{\boldsymbol{\theta}} \psi_m(\boldsymbol{t}^m) := \left(\partial^{\theta} \psi_m / \partial^{\theta_1} t_r^m \partial^{\theta_2} t_s^m\right)(\boldsymbol{t}^m)$$

- ***Tradeoff*** between both *accuracy* and *complexity*.
- Penalty parameters ϕ_m .

CQP and Tikhonov Regularization for MARS

$$\boldsymbol{\psi}(\tilde{\boldsymbol{b}}_i) := \left(1, \psi_1(\tilde{x}_i^1), \ldots, \psi_M(\tilde{x}_i^M), \psi_{M+1}(\tilde{\boldsymbol{x}}_i^{M+1}), \ldots, \psi_{M_{\max}}(\tilde{\boldsymbol{x}}_i^{M_{\max}})\right)^T$$

$$\tilde{\boldsymbol{b}}_i := (\tilde{x}_i^1, \tilde{x}_i^2, \ldots, \tilde{x}_i^M, \tilde{x}_i^{M+1}, \tilde{x}_i^{M+2}, \ldots, \tilde{x}_i^{M_{\max}})^T$$

$$\boldsymbol{\alpha} := \left(\alpha_0, \alpha_1, \ldots, \alpha_{M_{\max}}\right)^T,$$

$$(\sigma^{\kappa_j})_{j \in \{1,2,\ldots,K_m\}} \in \{0,1,2,\ldots,N+1\}^{K_m}$$

$$\hat{\boldsymbol{x}}_i^m = \left(\tilde{x}_{l^{\kappa_1^m}_{\sigma^{\kappa_1}}, \kappa_1^m}, \tilde{x}_{l^{\kappa_2^m}_{\sigma^{\kappa_2}}, \kappa_2^m}, \ldots, \tilde{x}_{l^{\kappa_{K_m}^m}_{\sigma^{\kappa_{K_m}}}, \kappa_{K_m}^m} \right),$$

$$\Delta\hat{\boldsymbol{x}}_i^m := \prod_{j=1}^{K_m} \left(\tilde{x}_{l^{\kappa_j^m}_{\sigma^{\kappa_j}+1}, \kappa_j^m} - \tilde{x}_{l^{\kappa_j^m}_{\sigma^{\kappa_j}}, \kappa_j^m} \right)$$

$$\boldsymbol{\psi}(\tilde{\boldsymbol{b}}) := \left(\boldsymbol{\psi}(\tilde{\boldsymbol{b}}_1), \ldots, \boldsymbol{\psi}(\tilde{\boldsymbol{b}}_N)\right)^T$$

$$L_{im} := \left[\left(\sum_{\substack{|\boldsymbol{\theta}|=1 \\ \boldsymbol{\theta}^T=(\theta_1,\theta_2)}}^{2} \sum_{\substack{r<s \\ r,s \in V(m)}} \left[D_{r,s}^{\theta} \psi_m(\hat{\boldsymbol{x}}_i^m) \right]^2 \right) \Delta\hat{\boldsymbol{x}}_i^m \right]^{1/2}.$$

$\boldsymbol{L}$ is an $(M_{\max}+1) \times (M_{\max}+1)$ matrix.

CQP and Tikhonov Regularization for MARS

- For a short representation, we can rewrite the approximate relation as

$$PRSS = \left\| \boldsymbol{y} - \boldsymbol{\psi}(\tilde{\boldsymbol{b}})\boldsymbol{\alpha} \right\|_2^2 + \sum_{m=1}^{M_{\max}} \phi_m \sum_{i=1}^{(N+1)^{K_m}} L_{im}^2 \, \alpha_m^2 .$$

- In case of the *same* penalty parameter $\phi = \phi_m (=: \lambda^2)$, then:

$$PRSS = \left\| \boldsymbol{y} - \boldsymbol{\psi}(\tilde{\boldsymbol{b}})\boldsymbol{\alpha} \right\|_2^2 + \phi \left\| \boldsymbol{L}\boldsymbol{\alpha} \right\|_2^2 .$$

Tikhonov regularization

CQP for MARS

- ***Conic quadratic*** programming:

$$\min_{t,\theta} \quad t,$$

$$\text{subject to} \quad \left\|\boldsymbol{\psi}(\tilde{\boldsymbol{b}})\boldsymbol{\alpha} - \boldsymbol{y}\right\|_2 \le t,$$

$$\left\|\boldsymbol{L}\boldsymbol{\alpha}\right\|_2 \le \sqrt{M}.$$

In general : $\min_{\boldsymbol{x}} \ \boldsymbol{c}^T\boldsymbol{x},$ subject to $\left\|\boldsymbol{D}_i\boldsymbol{x} - \boldsymbol{d}_i\right\|_2 \le \boldsymbol{p}_i^T\boldsymbol{x} - q_i \quad (i = 1, 2, ..., k).$

CQP for MARS

- Moreover, $(t, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\eta}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ is a ***primal dual optimal solution*** if and only if

$$\boldsymbol{\chi} := \begin{pmatrix} \mathbf{0}_N & \boldsymbol{\psi}(\tilde{\boldsymbol{b}}) \\ 1 & \mathbf{0}^T_{M_{\max}+1} \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{y} \\ 0 \end{pmatrix},$$

$$\boldsymbol{\eta} := \begin{pmatrix} \mathbf{0}_{M_{\max}+1} & \boldsymbol{L} \\ 0 & \mathbf{0}^T_{M_{\max}+1} \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{M_{\max}+1} \\ \sqrt{M} \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{0}^T_N & 1 \\ \boldsymbol{\psi}(\tilde{\boldsymbol{b}})^T & \mathbf{0}_{M_{\max}+1} \end{pmatrix} \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}^T_{M_{\max}+1} & 0 \\ \boldsymbol{L}^T & \mathbf{0}_{M_{\max}+1} \end{pmatrix} \boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{M_{\max}+1} \end{pmatrix},$$

$$\boldsymbol{\omega}_1^T \boldsymbol{\chi} = 0, \quad \boldsymbol{\omega}_2^T \boldsymbol{\eta} = 0,$$

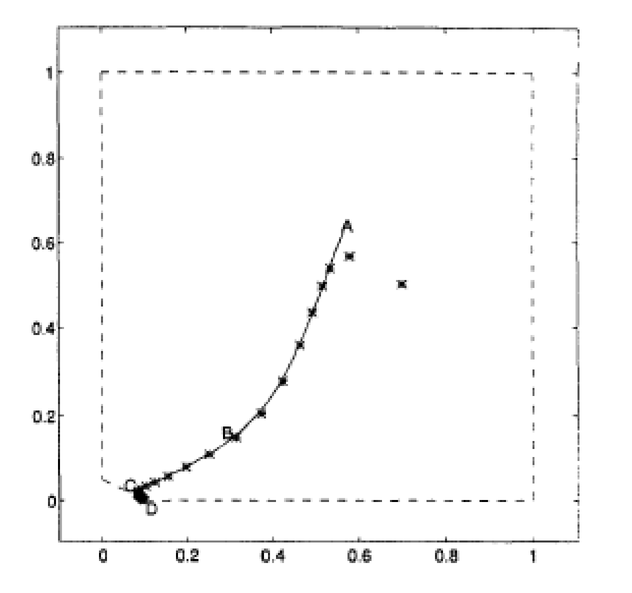
$$\boldsymbol{\omega}_1 \in L^{N+1}, \quad \boldsymbol{\omega}_2 \in L^{M_{\max}+2},$$

$$\boldsymbol{\chi} \in L^{N+1}, \quad \boldsymbol{\eta} \in L^{M_{\max}+2}.$$

CQP for MARS

- CQPs belong to the ***well-****structured* convex problems.
- *Interior Point Methods.*
- Better *complexity* bounds.
- *Better practical* performance.

C-MARS



The Idea of Robust (C)MARS

- (C)MARS models depend on the parameters.
  Small perturbations in data may give different model parameters.
  This may cause unstable solutions.

- In R(C)MARS, the aim is to reduce the estimation error
  while keeping efficiency as high as possible.

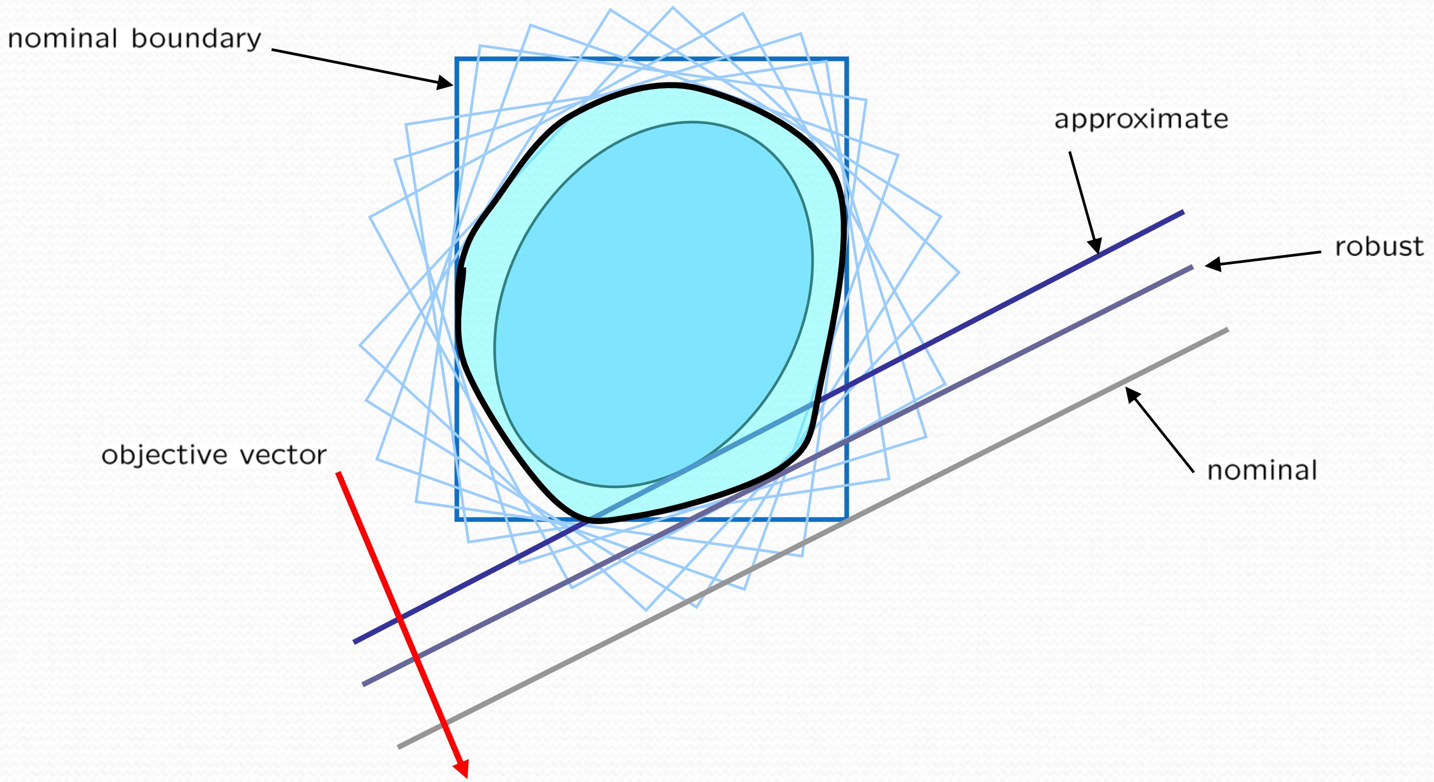
- In order to achieve this aim, we can use some approaches:
  - scenario optimization
  - robust counterpart
  - usage of more robust esrimators

- By using robustification in (C)MARS, the estimation variance will decrease.

Robust Optimization

- Robust counterpart may be much harder than original problem.

- Hence we may need to approximate



Robust Optimization

robust conic programming

- Conic program in dual form:

$$\max\ b^T y\ :\ c - A^T y \in K.$$

- Assume $A \in \mathcal{U}$, otherwise unknown.
- Robust conic program:

$$\max\ b^T y\ :\ \forall\, A \in \mathcal{U},\ \ c - A^T y \in K.$$

- Still convex, but maybe much harder than original conic program.

Robust Optimization

polytopic uncertainty

- Robust conic program:

$$\max\ b^T y\ :\ \forall\, A \in \mathcal{U},\ \ c - A^T y \in K.$$

- If $\mathcal{U}$ is a polytope described by its vertices:

$$\mathcal{U} = \mathbf{Conv}\{A_1, \ldots, A_L\},$$

  then robust conic program is an ordinary conic program:

$$\max\ b^T y\ :\ c - A_i^T y \in K,\ \ i = 1, \ldots, L.$$

- Drawback: in practice, vertices may not be known or too many to handle.

Robust Optimization

robust CQP

- Robust SOCP has the form

$$\min\ e^T x\ :\ \forall\, u \in \mathcal{U},\ \ \|A_i(u)x + b_i(u)\|_2 \le c_i(u)^T x + d_i(u)$$

  may be hard in general.

- Case of ellipsoidal uncertainty: if
  - $\mathcal{U}$ is an ellipsoid
  - data (A_i, b_i, c_i, d_i) are affine in u
  - uncertainties affecting (A_i, b_i) are independent of those affecting (c_i, d_i),

  then robust CQP can be solved as an SDP,

- see Ben Tal, Nemirovski, El Ghaoui (2000).

RCMARS

General model on the relation between input and response :

$$Y = f(\boldsymbol{X}) + \varepsilon, \qquad \boldsymbol{X} = (X_1, X_2, ..., X_p)^T$$

noisy input data value

X_j is random variable, and we assume that it is normally distributed.

$$X_j = \bar{X} + \zeta_j$$

mean

error term

To employ robust optimization on CMARS, a *"perturbation" (uncertainty)* is incorporated into the input data (for each dimension) and into the output data:

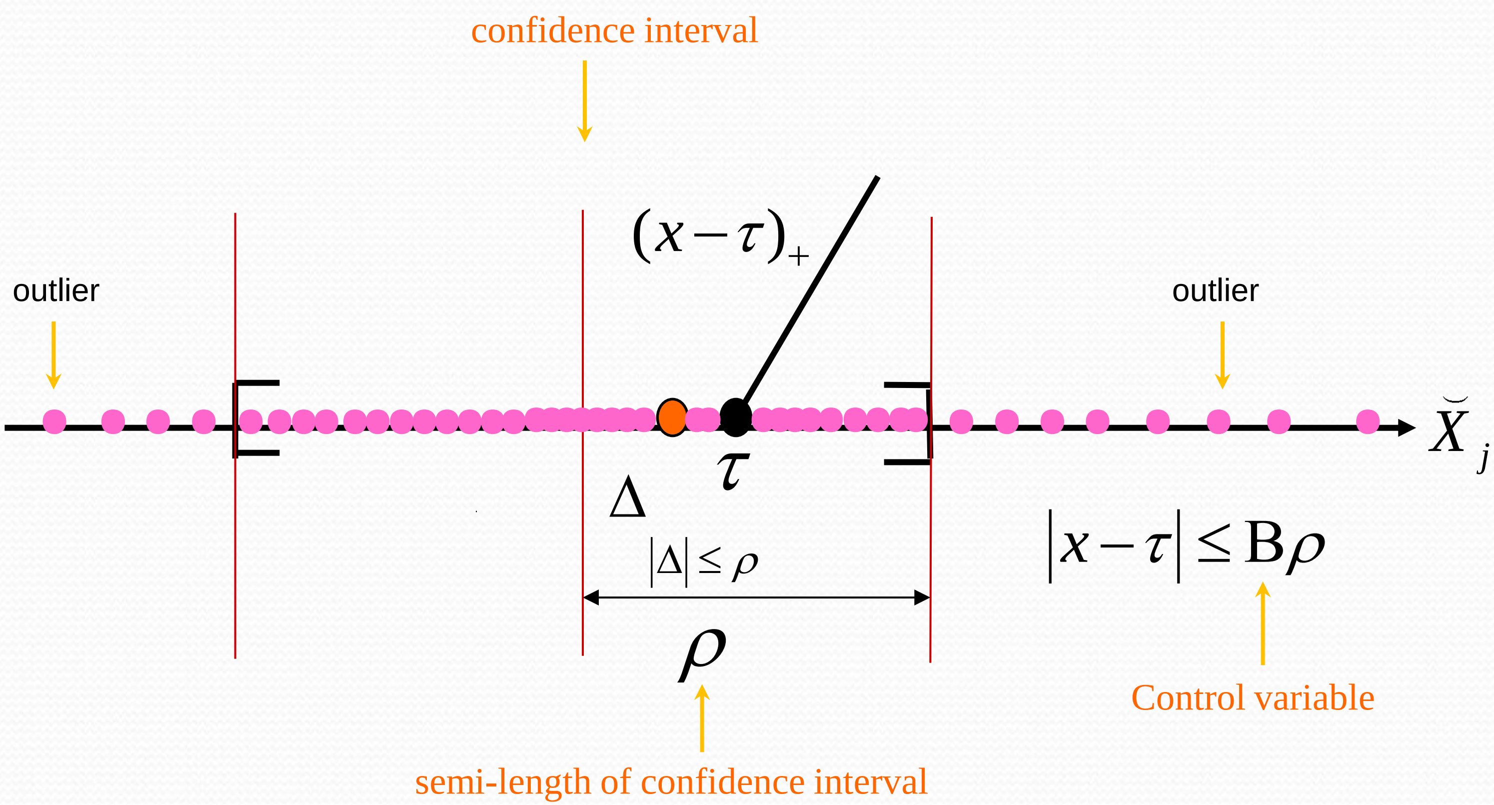
$\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, ..., x_{i,p})^T$ will be represented as $\breve{\boldsymbol{x}}_i = (\breve{x}_{i,1}, \breve{x}_{i,2}, ..., \breve{x}_{i,p})^T$,

after perturbation $\boldsymbol{\Delta}_i = (\Delta_{i,1}, \Delta_{i,2}, ..., \Delta_{i,p})^T$ $(i=1,2,...N)$.

$x_{ij} \rightarrow \breve{x}_{ij}$; $\breve{x}_{ij} = \bar{x}_j + \Delta_{ij}$, $\left|\Delta_{ij}\right| \leq \rho_{ij}$ $(j = 1, 2, ..., p;\ i=1, 2, ..., N)$.

Here, $\bar{x}_j$ $(j = 1, 2, ..., p)$ is the mean of the data $x_{i,j}$; the amount of perturbation in each dimension is restricted by ρ_{ij} , which is the semi-axis length of the confidence interval .

RCMARS



$$\mathrm{B} := \max\{\mathrm{B}_a \,|\, a = 1, \ldots, K_m - 1\}.$$

- B is normally equal to 2, but for outlier it will be bigger than 2.

RCMARS

The estimation of the basis functions

$$(\breve{x}_{i\kappa_m^j} - \tau_{i\kappa_m^j})_+ = \max\{0, \breve{x}_{i\kappa_m^j} - \tau_{\kappa_m^j}\} = \max\{0, (\overline{x}_{\kappa_m^j} + \Delta_{i\kappa_m^j}) - \tau_{\kappa_m^j}\}$$

$$\leq \max\{0, \overline{x}_{\kappa_m^j} - \tau_{\kappa_m^j}\} + \max\{0, \Delta_{i\kappa_m^j}\}$$

$$\leq \frac{1}{N}\sum_{l=1}^{N} \max\{0, (x_{l\kappa_m^j} - \tau_{\kappa_m^j})\} + \max\{0, \Delta_{i\kappa_m^j}\}$$

$$\leq \mathrm{A}^+_{i\kappa_m^j} + \max\{0, x_{i\kappa_m^j} - \tau_{\kappa_m^j}\} + \max\{0, \Delta_{\kappa_m^j}\}$$

$$= (x_{i\kappa_m^j} - \tau_{\kappa_m^j})_+ + \mathrm{A}^+_{i\kappa_m^j} + (\Delta_{i\kappa_m^j})_+ ,$$

$$(\breve{x}_{iK_m^j} - \tau_{K_m^j})_- = \max\{0, \tau_{\kappa_m^j} - \breve{x}_{i\kappa_m^j}\} = \max\{0, \tau_{\kappa_m^j} - (\overline{x}_{\kappa_m^j} + \Delta_{i\kappa_m^j})\}$$

$$\leq \max\{0, \tau_{\kappa_m^j} - \overline{x}_{\kappa_m^j}\} + \max\{0, -\Delta_{\kappa_m^j}\}$$

$$\leq \mathrm{A}^-_{i\kappa_m^j} + \max\{0, \tau_{\kappa_m^j} - x_{i\kappa_m^j}\} + \max\{0, -\Delta_{i\kappa_m^j}\}$$

$$= (x_{i\kappa_m^j} - \tau_{\kappa_m^j})_- + \mathrm{A}^-_{i\kappa_m^j} + (\Delta_{i\kappa_m^j})_- ,$$

RCMARS

By the combination of

$$(\breve{x}_{i\kappa_m^j} - \tau_{\kappa_m^j})_- \le (x_{i\kappa_m^j} - \tau_{i\kappa_m^j})_- + (\Delta_{i\kappa_m^j})_- + \mathrm{A}^-_{i\kappa_m^j} \quad \text{and} \quad (\breve{x}_{i\kappa_m^j} - \tau_{i\kappa_m^j})_+ \le (x_{i\kappa_m^j} - \tau_{\kappa_m^j})_+ + (\Delta_{i\kappa_m^j})_+ + \mathrm{A}^+_{i\kappa_m^j}$$

$$Y = f(\boldsymbol{X}) + \varepsilon,$$

$\mathrm{A}_{i\kappa_m^j} = \max\{\mathrm{A}^+_{i\kappa_m^j}, \mathrm{A}^-_{i\kappa_m^j}\}$: control variable.

The general form of the *m*th basis function is represented as follows:

$$\underbrace{\prod_{j=1}^{K_m} (\breve{x}_{i\kappa_m^j} - \tau_{\kappa_m^j})_\pm}_{=:\psi_m(\breve{x}_i^m)} \le \underbrace{\prod_{j=1}^{K_m} (x_{i\kappa_m^j} - \tau_{\kappa_m^j})_\pm}_{\psi_m(\boldsymbol{x}_i^m)} + \underbrace{\sum_{\substack{A \subseteq \{1,\dots K_m\} \\ \neq}} \prod_{a \in A} (x_{ia} - \tau_a)_\pm \prod_{b \in \{1,\dots K_m\}/A} (\pm \mathrm{A}_{ib} + \Delta_{ib})_\pm}_{u_{im}}$$

$(i = 1, 2, \dots, N).$

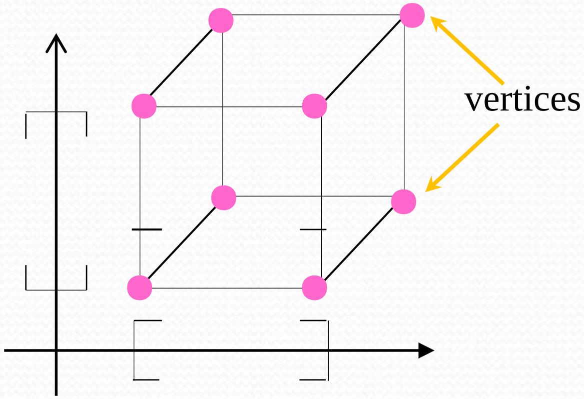
$\left|\psi_m(\breve{\boldsymbol{x}}_i^m) - \psi_m(\boldsymbol{x}_i^m)\right| = |u_{im}|$ can be estimated in this way:

$$|u_{im}| \le \sum_{\substack{A \subseteq \{1,\dots,K_m\} \\ \neq}} \prod_{a \in A} \underbrace{|x_{ia} - \tau_a|}_{\le \mathrm{B}_{ia}\rho_{ia}} \prod_{b \in \{1,\dots,K_m\}/A} \underbrace{|\pm \mathrm{A}_{ib} + \Delta_{ib}|}_{\le \gamma_{ib} + \rho_{ib}} \le \sum_{\substack{A \subseteq \{1,\dots,K_m\} \\ \neq}} \mathrm{B}_i^{|A|-1} \prod_{a \in A} \rho_{ia} \prod_{b \in \{1,\dots,K_m\}/A} (\gamma_{ib} + \rho_{ib}).$$

where $\quad \mathrm{B}_i := \max\{\mathrm{B}_{ia} \mid a = 1, 2, \dots, K_m - 1\} \quad (i = 1, 2, \dots, N).$

Polyhedral Uncertainty and Robust Counterpart

$$\boldsymbol{U}=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1M_{\max}} \\ u_{21} & u_{22} & \cdots & u_{2M_{\max}} \\ \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & \cdots & u_{NM_{\max}} \end{bmatrix}, \quad \boldsymbol{v}=\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}.$$



vertex

$$[\cdot,\cdot]\times[\cdot,\cdot]\times[\cdot,\cdot]\times\cdots\times[\cdot,\cdot]$$

Cartesian product

↓

polyhedral uncertainty sets $U_1,\ U_2$

Based on polyhedral uncertainty sets $U_1,\ U_2$, the robust counterpart

$$\min_{\alpha}\ \max_{\substack{\boldsymbol{W}\in U_1 \\ \boldsymbol{z}\in U_2}}\ \left\|\boldsymbol{z}-\boldsymbol{W}\boldsymbol{\alpha}\right\|_2^2+\phi\left\|\boldsymbol{L}\boldsymbol{\alpha}\right\|_2^2,$$

Polyhedral Uncertainty and Robust Counterpart

U_1 is a **polytope** with $2^{N\times M_{\max}}$ vertices $\boldsymbol{W}^1, \boldsymbol{W}^2 \ldots, \boldsymbol{W}^{2^{N\times M_{\max}}}$:

$$\boldsymbol{W} \in U_1 = \left\{ \sum_{j=1}^{2^{N\times M_{\max}}} \delta_j \boldsymbol{W}^j \;\middle|\; \delta_j \geq 0 \; (j = 1, \ldots, 2^{N\times M_{\max}}),\; \sum_{j=1}^{2^{N\times M_{\max}}} \delta_j = 1 \right\},$$

where $U_1 = \mathrm{conv}\{\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^{2^{N\times M_{\max}}}\}$ is the *convex hull*.

U_2 is the **polytope** with 2^N vertices $\boldsymbol{z}^1, \boldsymbol{z}^2, \ldots, \boldsymbol{z}^{2^N}$:

$$\boldsymbol{z} \in U_2 = \left\{ \sum_{i=1}^{2^N} \varphi_i \boldsymbol{z}^i \;\middle|\; \varphi_i \geq 0 \; (i = 1, 2, \ldots, 2^N),\; \sum_{i=1}^{2^N} \varphi_i = 1 \right\},$$

where $U_2 = \mathrm{conv}\{\boldsymbol{z}^1, \boldsymbol{z}^2, \ldots, \boldsymbol{z}^{2^N}\}$ is the *convex hull*.

Robust CQP with the Polytopic Uncertainity

Robust conic quadratic programming of **RCMARS**:

$$\min_{t,\boldsymbol{\alpha}} \quad t$$

$$\text{subject to} \quad \boldsymbol{z} - \boldsymbol{W}\boldsymbol{\alpha} \in \boldsymbol{L}$$

$$\forall\ \boldsymbol{W} \in U_1,\ \ \boldsymbol{z} \in U_2,$$

where L ice-cream (or second-order, or Lorentz) cones.

equivalently

$$\min_{t,\boldsymbol{\alpha}} \quad t$$

$$\text{subject to} \quad \boldsymbol{z}^i - \boldsymbol{W}^j\boldsymbol{\alpha} \in L \ \ (i = 1, \ldots, 2^N;\ j = 1, \ldots, 2^{N \times M_{\max}})$$

(Standard) Conic Quadratic Programming

Simulation Study

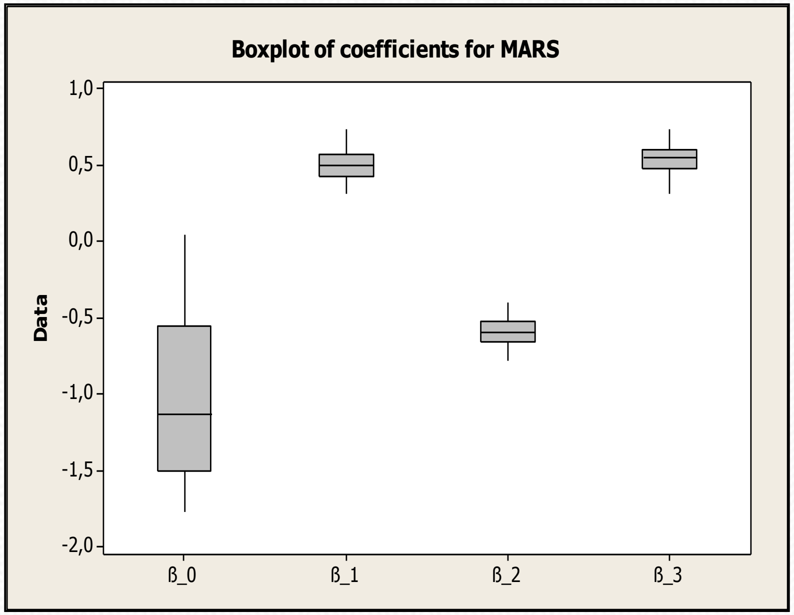
- Compare MARS and RMARS methods based on *variation of the parameter estimates*.
- Obtain 30 different random data sets to apply simulation for MARS algorithms.
- Construct 30 different data sets to apply simulation for RMARS algorithm with 30 different uncertainty scenarios.
- In order to see variation of model performance with parameter estimates, *estimation errors of simulation models* are also evaluated based on AAE and RMSE.
- Since it is not easy to illustrate the reduction of estimation variance in RMARS statistically, this reduction is represented graphically.

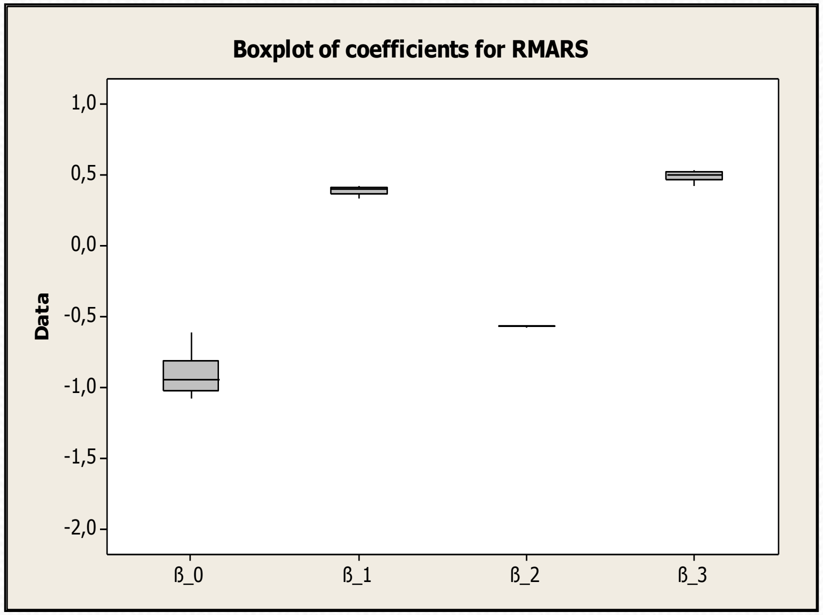
Simulation Study

The graphical representation for the variance of parameter estimates:

MARS

RMARS

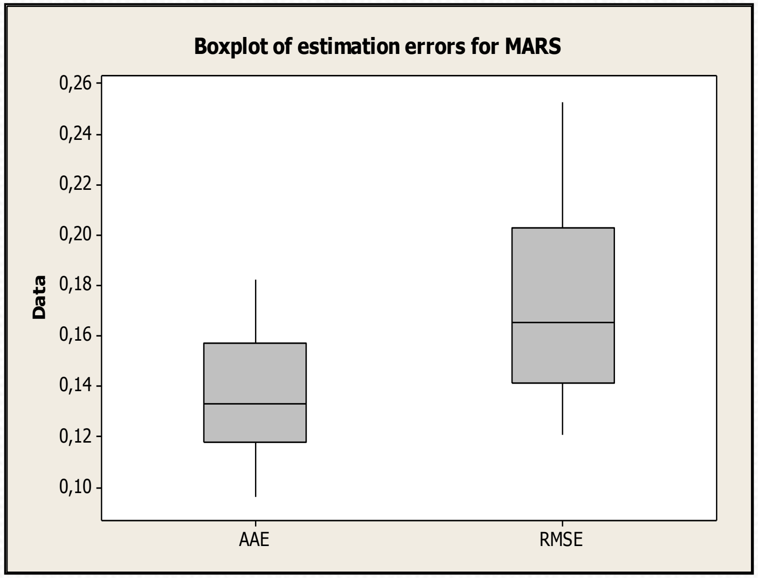




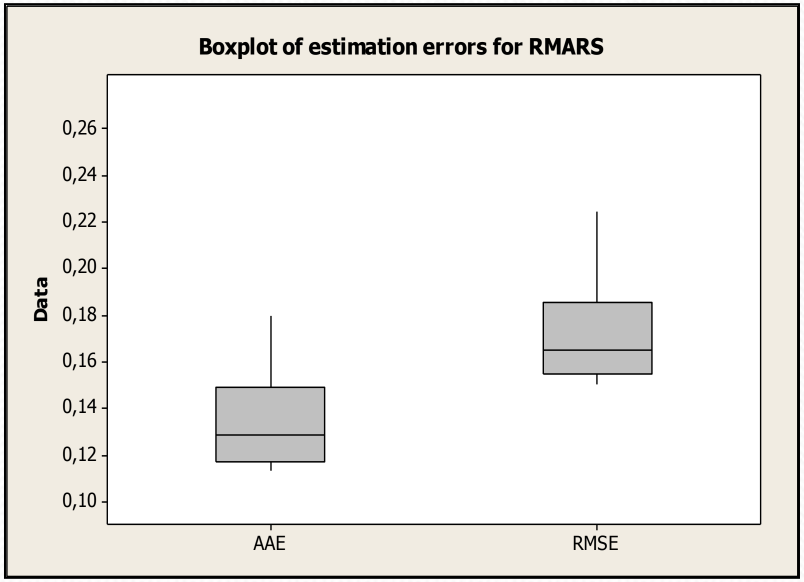
Simulation Study

The graphical representation for the variance of model performance:

MARS



RMARS



A Real-World Application by RCMARS

Precipitation models help in the design of *early warning systems*
for disasters such as severe storms, floods and droughts.

For this study, we aimed to build precipitation models for the *continental central Anatolia* (*CCA*) region of Turkey,
where drought has been a recurrent phenomenon for the last few decades.

For this purpose, the RCMARS, MARS and CMARS methods were used,
and then, the models developed were evaluated and compared with respect to several criteria including accuracy, precision and stability.

A Real-World Application by RCMARS

We used RCMARS for precipitation forecasting, and compared its performance to that of MARS and CMARS.

The dataset studied includes seven meteorological variables:

- monthly precipitation total (in millimeters),
- monthly mean temperature,
- monthly relative humidity (in percent),
- cloudiness,
- vapor pressure,
- surface air temperature,
- mean pressure,
- mixing ratio.

A Real-World Application by RCMARS

In RCMARS methodology, we had difficulties regarding computer capacity to solve the optimization problem using uncertainty matrices on a large amount of data,

consisting of seven variables with 420 rows (one for every month in 35 years), for each one of the 43 recording stations.

To overcome this problem, the size of data was reduced by taking yearly averages of each meteorological variable over all stations.

To compare the performances of prediction models constructed, we also applied the hold-out method as the validation technique, where the dataset is divided into two subsamples as training and test sets.

Since the dataset contains a time series of meteorological variables, it was not subdivided randomly.

Instead, the first 30 years (1976-2005) of each variable considered were assigned to be the training dataset, whereas the last 5 years of the series were assigned to be the test dataset.

A Real-World Application by RCMARS

Our basic performance measure to calculate the precision of the models was prediction variance (*PV*) which is the variance of the estimated response values.

Also, to compare the results concerning the accuracies of RCMARS, CMARS and MARS methods, the models developed were also evaluated based on some accuracy measures like

- multiple coefficients of determination (R^2),
- average absolute error (*AAE*),
- root mean squared error (*RMSE*),
- correlation coefficient (*r*).

A Real-World Application by RCMARS

We constructed model functions for these data using
MARS Software,
where we selected the maximum number of basis elements: $M_{\max} = 12$.

$$\psi_1(\boldsymbol{x}) = \max\{0,\ X_2 + 2.0927\}, \qquad \psi_2(\boldsymbol{x}) = \max\{0,\ X_3 + 1.4227\},$$
$$\psi_3(\boldsymbol{x}) = \max\{0,\ X_7 + 0.6001\}, \qquad \psi_4(\boldsymbol{x}) = \max\{0,\ -0.6001 - X_7\},$$
$$\psi_5(\boldsymbol{x}) = \max\{0,\ X_6 - 0.2563\}, \qquad \psi_6(\boldsymbol{x}) = \max\{0,\ 0.2563 - X_6\},$$
$$\psi_7(\boldsymbol{x}) = \max\{0,\ X_5 + 0.0875\}, \qquad \psi_8(\boldsymbol{x}) = \max\{0,\ -0.0875 - X_5\},$$
$$\psi_9(\boldsymbol{x}) = \max\{0,\ X_4 + 2.3288\}, \qquad \psi_{10}(\boldsymbol{x}) = \max\{0,\ X_1 + 2.4477\},$$
$$\psi_{11}(\boldsymbol{x}) = \max\{0,\ X_4 + 0.1409\}, \qquad \psi_{12}(\boldsymbol{x}) = \max\{0,\ -0.1409 - X_4\}.$$

Here, X_1, X_2 and X_3 are the normalized mean temperature, cloudiness, and vapor pressure;

X_4 and X_5 are the first-order lagged cloudiness and mean pressure;

X_6 and X_7 are the fifth-order lagged cloudiness and vapor pressure, respectively.

The RCMARS model obtained is a "distributed lag" model due to the fact that it includes lagged independent variables

A Real-World Application by RCMARS

- To prevent non-differentiability in the optimization problem, the knot values selected are different from but very much close to the corresponding input data.

- As a result, the largest model can be described as follows:

$$\begin{aligned}
\hat{Y} &= \alpha_0 + \sum_{m=1}^{M_{\max}} \alpha_m \psi_m(\boldsymbol{x}) \\
&= \alpha_0 + \alpha_1\psi_1(\boldsymbol{x}) + \alpha_2\psi_2(\boldsymbol{x}) + \alpha_3\psi_3(\boldsymbol{x}) + \alpha_4\psi_4(\boldsymbol{x}) + \alpha_5\psi_5(\boldsymbol{x}) + \alpha_6\psi_6(\boldsymbol{x}) + \alpha_7\psi_7(\boldsymbol{x}) \\
&\quad + \alpha_8\psi_8(\boldsymbol{x}) + \alpha_9\psi_9(\boldsymbol{x}) + \alpha_{10}\psi_{10}(\boldsymbol{x}) + \alpha_{11}\psi_{11}(\boldsymbol{x}) + \alpha_{12}\psi_{12}(\boldsymbol{x}) \\
&= \alpha_0 + \alpha_1 \max\{0,\ X_2 + 2.09278\} + \alpha_2 \max\{0,\ X_3 + 1.4228\} \\
&\quad + \alpha_3 \max\{0,\ X_7 + 0.6002\} + \alpha_4 \max\{0,\ -0.6002 - X_7\} \\
&\quad + \alpha_5 \max\{0,\ X_6 - 0.2564\} + \alpha_6 \max\{0,\ 0.2564 - X_6\} \\
&\quad + \alpha_7 \max\{0,\ X_5 + 0.0876\} + \alpha_8 \max\{0,\ -0.0876 - X_5\} \\
&\quad + \alpha_9 \max\{0,\ X_4 + 2.3289\} + \alpha_{10} \max\{0,\ X_1 + 2.4478\} \\
&\quad + \alpha_{11} \max\{0,\ X_4 + 0.1410\} + \max\{0,\ -0.1410 - X_4\}.
\end{aligned}$$

A Real-World Application by RCMARS

Then, the *uncertainty* is evaluated for all input and output values, which are represented by *confidence intervals*(*CIs*).

Based on polyhedral uncertainty *sets*, the uncertainty matrices and vectors are obtained.

- The uncertainty matrix for input data has a huge size $(2^{30\times 12} = 2^{360})$.
- We do not have enough computer capacity.
- ***Tradeoff*** between *tractibility* and *robustification*.
- We formulate a submodel for each sample value (observation) using the *combinatorial approach*, which we call *weak robustification*.

To solve our problem:

- We obtain 30 different ***weak RCMARS*** (***WRCMARS***) submodels.
- We solve them separately by using MOSEK.
- After obtaining result for each of our observation, we select a model which has the maximum *t* value using the *worst-case approach*, and we continue with this model to calculate our parameter values.

A Real-World Application by RCMARS

- First, using the training dataset, several MARS models were developed using Salford System's MARS software.
- After picking the best one among them, the CMARS model was constructed and robustified under polyhedral uncertainty.
- While developing RCMARS models, a **sensitivity study** was conducted to determine the most suitable confidence limits on both input and output data.
- For this purpose, different *uncertainty matrices*, U, for the input data, $\boldsymbol{x}$, and different *uncertainty vectors*, v, for the output data, $\boldsymbol{y}$, were constructed by using four different intervals.
- These are represented by the pairs ±3/5, ±3/10, ±3/20 and 0 (i.e., zero-length interval).
- Here, the 0-length interval refers to a *special case* where the RCMARS model reduces to the CMARS model.
- We estimated our parameters with four different uncertainty scenarios using PRSS values under polyhedral uncertainty sets for our training data set.

A Real-World Application by RCMARS

Table 1 Performance measures of RCMARS models (Note:* indicates best performance of *U* for training, test and stability data with respect to the related performance measure)

U	±3/5											
v	±3/5			±3/10			±3/20			±0		
	training	test	stability	training	test	stability	training	test	stability	training	test	stability
R^2	0.782	0.833	0.939	0.813	0.849	0.958	0.823*	0.846	0.972*	0.767	0.850*	0.902
MAE	0.360	0.256	0.710*	0.327	0.206	0.631	0.312*	0.200*	0.641	0.374	0.256	0.683
RMSE	0.459	0.366	0.796	0.426	0.348	0.818	0.414*	0.351	0.848*	0.475	0.347*	0.730
r	0.907	0.942	0.962	0.914	0.935	0.977	0.916*	0.931	0.984*	0.898	0.944*	0.952
PV	0.504	0.523	0.963	0.602	0.669	0.899	0.649	0.741	0.876	0.495	0.585	0.846
U	±3/10											
v	±3/5			±3/10			±3/20			±0		
	training	test	stability	training	test	stability	training	test	stability	training	test	stability
R^2	0.934	0.649	0.695	0.941	0.609	0.647	0.943*	0.588	0.623	0.876	0.789*	0.901*
MAE	0.198	0.410	0.482	0.181	0.441	0.410	0.172*	0.462	0.372	0.273	0.311*	0.877*
RMSE	0.253	0.530	0.477	0.238	0.559	0.426	0.234*	0.574	0.408	0.346	0.411*	0.842*
r	0.973	0.823	0.846	0.975	0.811	0.831	0.975*	0.808	0.828	0.950	0.900*	0.947*
PV	0.752	0.800	0.940*	0.788	0.880	0.894	0.820	0.954	0.859	0.628*	0.687*	0.914
U	±3/20											
v	±3/5			±3/10			±3/20			±0		
	training	test	stability	training	test	stability	training	test	stability	training	test	stability
R^2	0.934	0.649	0.695	0.941	0.609	0.647	0.943*	0.588	0.623	0.876	0.789*	0.901*
MAE	0.198	0.410	0.482	0.181	0.441	0.410	0.172*	0.462	0.372	0.273	0.311*	0.877*
RMSE	0.253	0.530	0.477	0.238	0.559	0.426	0.234*	0.574	0.408	0.346	0.411*	0.842*
r	0.973	0.823	0.846	0.975*	0.811	0.831	0.975	0.808	0.828	0.950	0.900*	0.947*
PV	0.752	0.800	0.940*	0.788	0.880	0.894	0.820	0.954	0.859	0.628*	0.687*	0.914
U	±0											
v	±3/5			±3/10			±3/20			±0 (CMARS)		
	training	test	stability	training	test	stability	training	test	stability	training	test	stability
R^2	0.941	0.563	0.598	0.941	0.563*	0.598	0.941	0.563	0.598*	0.971*	0.2248	0.231
MAE	0.186	0.468	0.398	0.186	0.468	0.398	0.186	0.468*	0.398*	0.131*	0.6463	0.203
RMSE	0.239	0.591	0.404	0.239	0.591	0.404	0.239	0.591*	0.404*	0.166*	0.7875	0.211
r	0.977	0.769	0.787*	0.977	0.769	0.787*	0.977	0.769*	0.787*	0.986*	0.6719	0.682
PV	0.735*	0.788*	0.932*	0.735*	0.788	0.932*	0.735*	0.789	0.931	0.953	1.2409	0.768

A Real-World Application by RCMARS

- Based on the findings, the best RCMARS solution was determined for $U=\pm3/5$, $\pm3/10$ or $\pm3/20$ and $v=\pm0$.

- For the purposes of comparison, we took $U=\pm3/10$ or $\pm3/20$ and $v=\pm0$.

Table 2 The performance measures of the precipitation models (Note:* indicates the best performance for training, test and stability data with respect to the corresponding performance measure)

Method	**MARS**			**CMARS**			**RCMARS**		
Measure	training	test	stability	training	test	stability	training	test	stability
R^2	0.957	0.139	0.145	0.971*	0.225	0.231	0.876	0.789*	0.901*
MAE	0.165	0.701	0.235	0.131*	0.6463	0.203	0.273	0.311*	0.877*
RMSE	0.204	0.830	0.246	0.166*	0.788	0.211	0.346	0.411*	0.842*
r	0.978	0.652	0.666	0.986*	0.672	0.680	0.950	0.900*	0.947*
PV	0.957	1.324	0.723	0.953	1.2409	0.768	0.628*	0.687*	0.914*

- For the training data, CMARS performed better than the other two methods with respect to all measures except *PV*; it was *the best* for RCMARS.

- For the test data and stabilities, RCMARS considerably outperformed the other two methods with respect to all measures.

Conclusion and Outlook on Future Studies

- By introducing a robustification in (C)MARS, we try to decrease the ***estimation variance*** and, we demonstrate their good performance with simulation studies, numerical experiences and real-world applications.

- According performance criteria, RCMARS can produce more accurate models with smaller variances when we compare with MARS and CMARS for testing.

- Although we received better performance by RCMARS, we have to solve an extra problem (by Software MARS, etc.), namely the *knot selection* (being not necessary for the linear part) in RCMARS.

- For this reason, as a future work, using the contribution of RCMARS, we study on *Robust Conic Generalized Partial Linear Model* (***RCGPLM***) for modeling, optimization and robustification to decrease the complexity of RCMARS; that complexity is measured by the number of variables applied in RCMARS.

- This new semiparametric model approach gains in importance to **reduce** complexity and variance of estimation.

References

- Aster, A., Borchers, B., and Thurber, C., *Parameter Estimation and Inverse Problems*, Academic Press, 2004.
- Ben-Tal, A., Nemirovski, A., Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, MPR-SIAM Series on Optimization, SIAM, Philadelphia, 2001.
- El Ghaoui, L., Robust Optimization and Applications, IMA Tutorial, 2003.
- Friedman, J.H., Multivariate adaptive regression splines, The Annals of Statistics 19, 1 (1991) 1-141.
- Hansen, P.C., *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998.
- Hastie, T., Tibshirani, R., and Friedman, J.H., *The Element of Statistical Learning*, Springer Verlag, NY, 2001.
- Krabs, W., and Pickl, S., A game-theoretic treatment of a time-discrete emission reduction model, International Game Theory Review 6(1) (2004) 21–34.
- Myers, R.H., and Montgomery, D.C., Response Surface Methodology: Process and Product Optimization Using Designed Experiments,New York: Wiley (2002).
- MARS® Salford Systems; software available at http://www.salfordsystems.com.
- MOSEK™; software available at http://www.mosek.com.
- Özmen, A., Weber, G.-W., and Batmaz, I., The New Robust CMARS (RCMARS) Method, 24th Mini-EURO Conference - On Continuous Optimization and Information-Based Technologies in the Financial Sector, MEC EurOPT 2010, Selected Papers, ISI Proceedings (Izmir, Turkey, June 23-26, 2010), 362-368.

- Özmen A., Weber G.-W., Batmaz I., and Kropat E., RCMARS: Robustification of CMARS with Different Scenarios under Polyhedral Uncertainty Set, Communications in Nonlinear Science and Numerical Simulation (CNSNS) 16 (12) (2011) 4780-4787.
- Özmen, A., and Weber, G.-W., Robust conic generalized partial linear models using RCMARS method, AIP Proceedings 1499 (2012) 337-343 (ISBN 978-0-7354-1113-5, ISSN 0094-243X) (AIP: American Institute of Physics).
- Özmen, A, G.-W. Weber and Kropat, E., Robustification of Conic Generalized Partial Linear Models under Polyhedral Uncertainty, International IFNA-ANS scientific Journal "Problems of Nonlinear Analysis in Engineering Systems, 18:2(38)(104-113), 2012.
- Özmen, A., Weber, G.-W., Çavuşoğlu, Z., and Deferli, Ö., The new robust conic GPLM method with an Application to Finance: prediction of credit default, Journal of Global Optimization (JOGO) 56 (2) (2013) 233-249.
- Özmen A., Weber G.-W., and Karimov, A., RCMARS: A new optimization supported tool applied on financial data, Pacific Journal of Optimization (PJO) 9(3) (2013) 535-552.
- Özmen, A., Batmaz, I., and Weber, G.-W., Precipitation Modeling by Polyhedral RCMARS and Comparison with MARS and CMARS, Environmental Modeling and Assessment 19, 82 (October 2014) 425–435.
- Özmen, A., Kropat, E., and Weber, G.-W., Spline regression models for complex multi-modal regulatory networks, Optimization Methods and Software (OMS) 29(3) (2014) 515-534.
- Özmen, A., and Weber, G.-W., RMARS: Robustification of Multivariate Adaptive Regression Spline under Polyhedral Uncertainty, Journal of Computational and Applied Mathematics (CAM) 259 (2014) 914–924.

- Roos, C., Terlaky T., and Vial, J., Interior Point Approach to Linear Optimization: Theory and Algorithms, John Wiley & Sons, New York, 1997.
- Roos, C., Terlaky T., and Vial, J., Interior Point Methods for Linear Optimization, Springer Science, Heidelberg/Boston, 2006.
- Taylan, P., Weber, G.-W., and Beck, A., New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology, Optim. 56, 5-6 (2007), pp. 675–698.
- Uğur, O., and Weber, G.-W., Optimization and dynamics of gene-environment networks with intervals. J. Ind. Manag. Optim. 3(2), 357-379, 2007.
- Weber, G.-W., Batmaz, I., Köksal, G., Taylan, P., and Yerlikaya, F., 2009. CMARS: A New Contribution to Nonparametric Regression with Multivariate Adaptive Regression Splines Supported by Continuous Optimisation, Inverse Problen in Science and Engineering.
- Werner, R., Consistency of robust portfolio estimates, Optimization in Finance, Coimbra, 2007.
- J.R. Harris, W. Nystad, P. Magnus, Using genes and environments to define asthma and related phenotypes: applications to multivariate data, Clinical and Experimental Allergy 28(1) (1998) 43–45.
- Weber, G.-W., Tezel, A., Taylan, P., Soyler, and A., Çetin, M., Mathematical contributions to dynamics and optimization of gene-environment networks. Optim. 57(2), 353-377, 2008.
- Weber, G.-W., Çavuşoğlu, Z., and Özmen, A., Predicting default probabilities in emerging markets by conic generalized partial linear models and their optimization, Optimization 61(4) (2012) 443-457.
- Yildirim, M.H., Özmen, A., Türker Bayrak, Ö., and Weber, G.-W., Electricity price modeling for Turkey, Operations Research Proceedings 2011, Selected Papers of the International Conference on Operations Research (OR 2011), August 30 - September 2, 2011, Zurich, Switzerland, (2012) 39-44.

Thank you