# A general exponential kernel to handle mixed-categorical variables for Gaussian process

**Paul SAVES**

***ISAE-SUPAERO & ONERA***

***Toulouse, France***

**Youssef Diouane (Polytechnique Montréal)**
**Nathalie Bartoli (ONERA)**
**Thierry Lefebvre (ONERA)**
**Joseph Morlier (ISAE-SUPAERO/ICA)**

P. Saves, Y. Diouane, N. Bartoli, T. Lefebvre, and J. Morlier. A mixed-categorical correlation kernel for gaussian process, 2022. Submitted to Neurocomputing. Available at SSRN: https://dx.doi.org/10.2139/ssrn.4296794

# Context at ONERA

# Context at ONERA

New aircraft configurations = Unknown behaviors

# Context at ONERA

New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

# Context at ONERA

New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

**Aircraft design developments required:**

# Context at ONERA

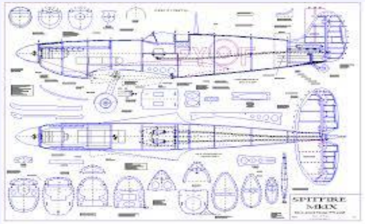
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

**Aircraft design developments required:**

- a large number of design variables

**High-dimension**



# Context at ONERA

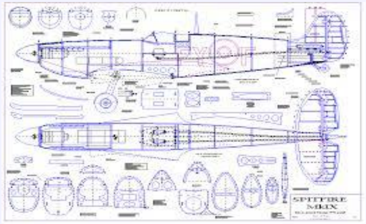
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

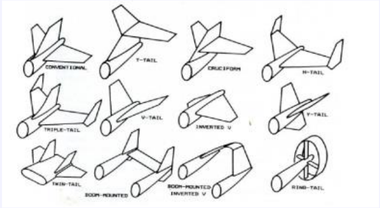
**Aircraft design developments required:**

- a large number of design variables
- mixed integer variables

**High-dimension**



**Discrete variables**



# Context at ONERA

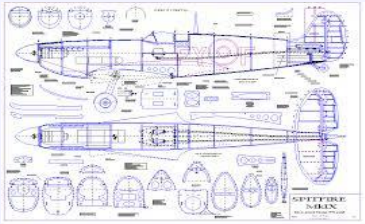
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

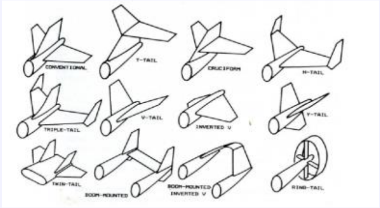
**Aircraft design developments required:**

- a large number of design variables
- mixed integer variables
- multimodal constraints (convex or non-convex, equality or inequality,…)

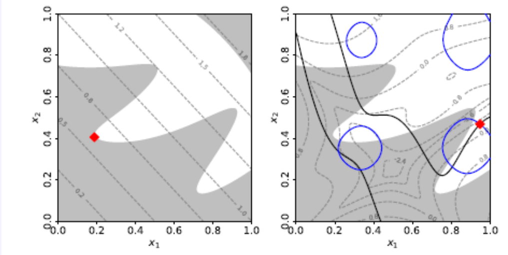
**High-dimension**



**Discrete variables**



**Multimodal constraints**



# Context at ONERA

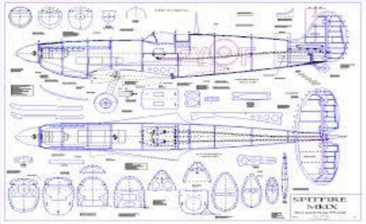
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

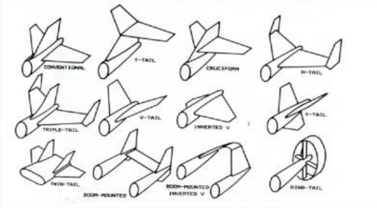
**Aircraft design developments required:**

- a large number of design variables
- mixed integer variables
- multimodal constraints (convex or non-convex, equality or inequality,…)
- multiobjective optimization (OWE vs Fuel mass,…)

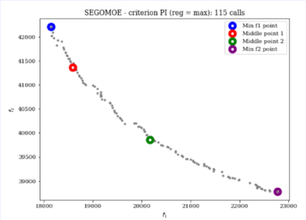
## High-dimension

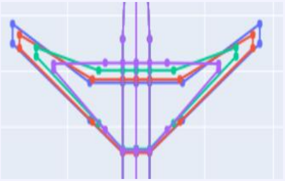


## Discrete variables

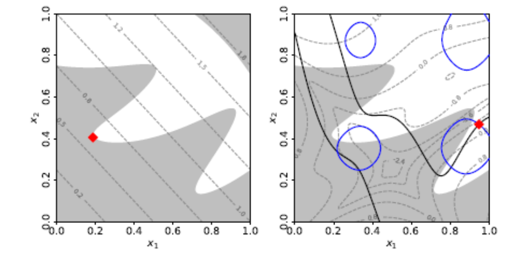


## Multiobjective optimization



## Multimodal constraints



# Context at ONERA

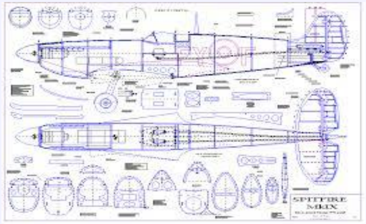
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

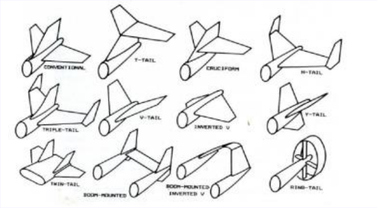
**Aircraft design developments required:**

- a large number of design variables
- mixed integer variables
- multimodal constraints (convex or non-convex, equality or inequality,…)
- multiobjective optimization (OWE vs Fuel mass,…)
- multifidelity data from simulations and computer experiments

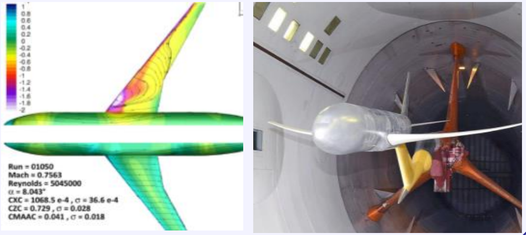
**High-dimension**



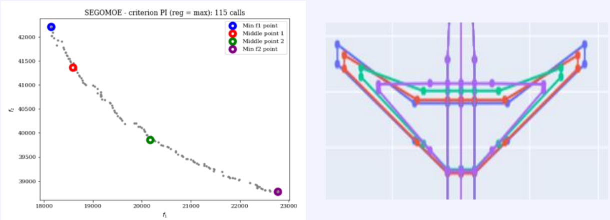
**Discrete variables**



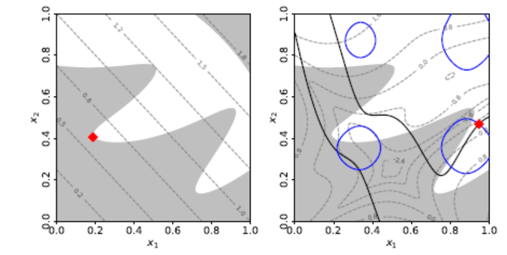
**Multifidelity database**




**Multiobjective optimization**




**Multimodal constraints**



# Context at ONERA

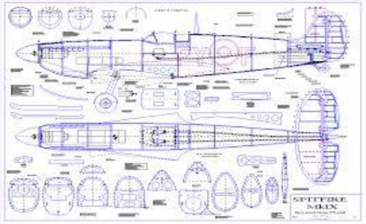
New aircraft configurations = Unknown behaviors

➔ built as a **black box** (derivative-free & expensive-to-evaluate)

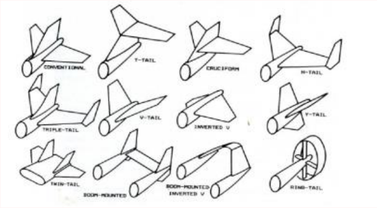
**Aircraft design developments required:**

- a large number of design variables
- mixed integer variables
- multimodal constraints (convex or non-convex, equality or inequality,…)
- multiobjective optimization (OWE vs Fuel mass,…)
- multifidelity data from simulations and computer experiments

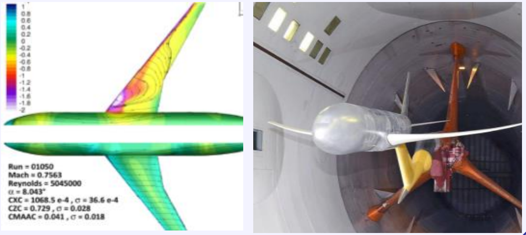
**High-dimension**



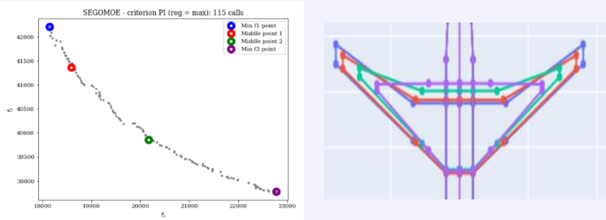
**Discrete variables**



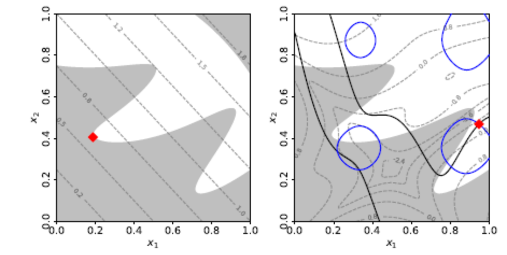
**Multifidelity database**




**Multiobjective optimization**

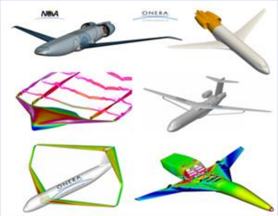



**Multimodal constraints**



# Bayesian optimization applied to aircraft design

**New concepts**

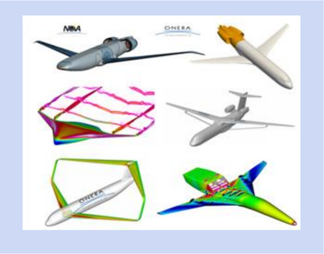


$\Omega$

Kim, S. H., and Boukouvala, F., "Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems," Computers & Chemical Engineering, Vol. 140, 2020, p. 106847.

# Bayesian optimization applied to aircraft design

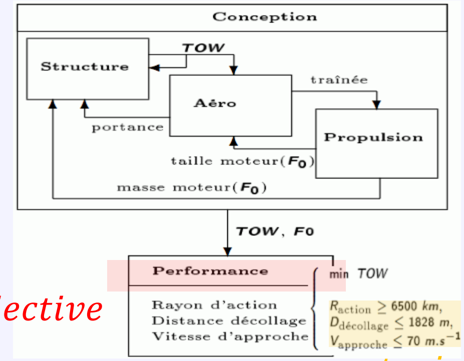
## New concepts



Ω

Analyze

## Multidisciplinary Design Analysis



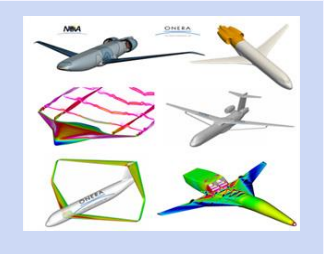

*objective*

*constraints*

Kim, S. H., and Boukouvala, F., "Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems," Computers & Chemical Engineering, Vol. 140, 2020, p. 106847.

# Bayesian optimization applied to aircraft design

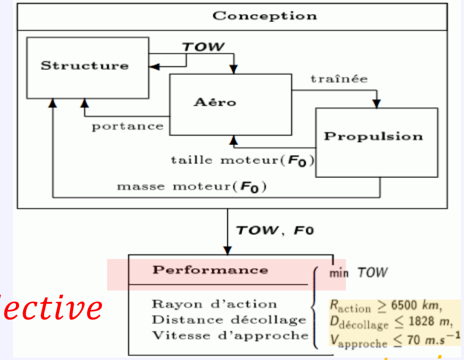
## New concepts




$\Omega$

**Analyze**

## Multidisciplinary Design Analysis




*objective*

*constraints*

**Improvements**

## Global optimization

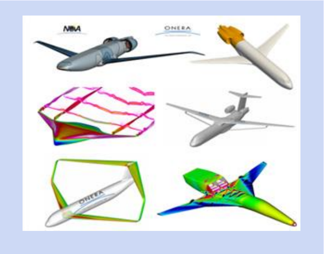
## MDF Formulation

$$\min_{\text{inputs} \in \Omega} \ objective\,(\text{inputs})$$
$$s.t \quad constraints\,(\text{inputs})$$

Kim, S. H., and Boukouvala, F., "Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems," Computers & Chemical Engineering, Vol. 140, 2020, p. 106847.

# Bayesian optimization applied to aircraft design

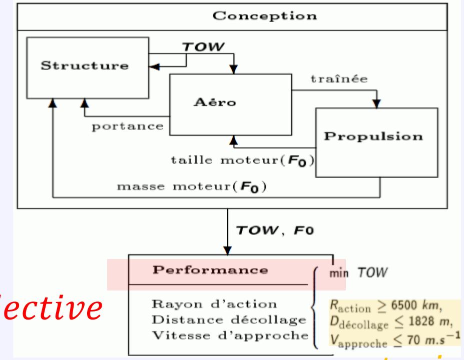
## New concepts




$\Omega$

**Analyze**

## Multidisciplinary Design Analysis




*objective*

*constraints*

**Improvements**

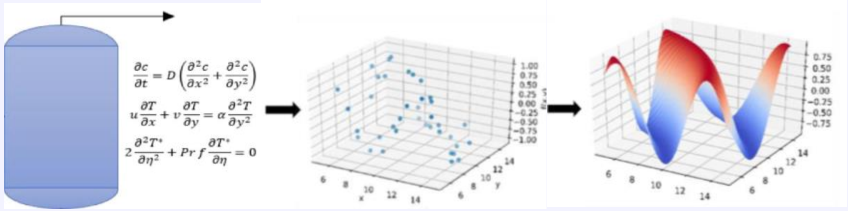
## Global optimization

## MDF Formulation

$$\min_{\text{inputs} \in \Omega} \ objective\ (\text{inputs})$$
$$s.t \quad constraints\ (\text{inputs})$$

**Expensive computations**

## Surrogate model

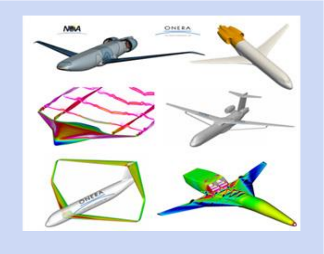



**Expensive black-box** **Design of Experiments** **Gaussian processes**

Kim, S. H., and Boukouvala, F., "Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems," Computers & Chemical Engineering, Vol. 140, 2020, p. 106847.

# Bayesian optimization applied to aircraft design

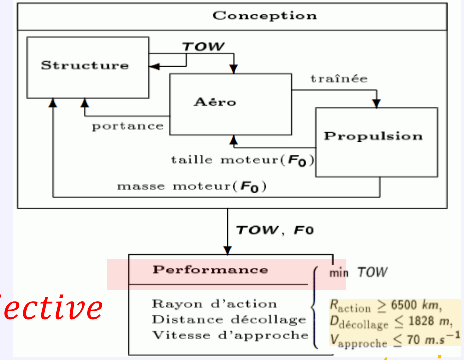
## New concepts



$\Omega$

**Analyze**

## Multidisciplinary Design Analysis




*objective*

*constraints*

**Improvements**

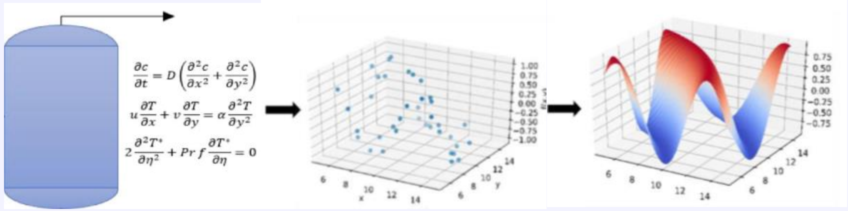
## Global optimization
## MDF Formulation

$$\min_{\text{inputs} \in \Omega} \ \textit{objective}\,(\text{inputs})$$
$$s.t \quad \textit{constraints}\,(\text{inputs})$$

**Expensive black-box optimization**

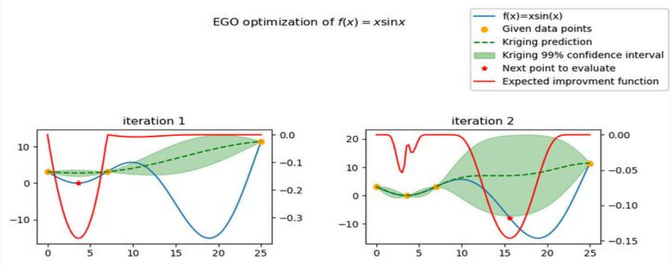
**Expensive computations**

## Surrogate model




**Expensive black-box** **Design of Experiments** **Gaussian processes**

## Efficient Global Optimization




Kim, S. H., and Boukouvala, F., “Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems,” Computers & Chemical Engineering, Vol. 140, 2020, p. 106847.

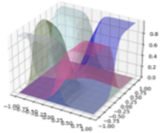
# Two frameworks

# Two frameworks

SMT: Surrogate Modeling Toolbox

github.com/SMTorg/smt

- Open source python toolbox: surrogate modeling methods, sampling techniques, and benchmarking functions
- Focus on derivatives (training derivatives used for gradient-enhanced modeling, prediction derivatives)
- New Kriging based surrogate models for higher dimension (KPLS and KPLS-K)
- Noisy Kriging to handle uncertainties on data
- Multifidelity Kriging with or without noise (MFK, MFKPLS)
- Mixture of experts technique for heterogeneous functions



- Mixed integer Kriging to handle **discrete and categorical variables**

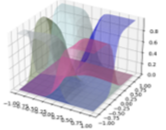
# Two frameworks

## SMT

github.com/SMTorg/smt

SMT: Surrogate Modeling Toolbox

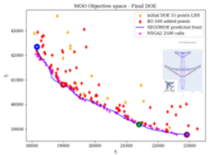
- Open source python toolbox: surrogate modeling methods, sampling techniques, and benchmarking functions
- Focus on derivatives (training derivatives used for gradient-enhanced modeling, prediction derivatives)
- New Kriging based surrogate models for higher dimension (KPLS and KPLS-K)
- Noisy Kriging to handle uncertainties on data
- Multifidelity Kriging with or without noise (MFK, MFKPLS)
- Mixture of experts technique for heterogeneous functions



- Mixed integer Kriging to handle **discrete and categorical variables**

## SEGOMOE

- Mono & multi objective Bayesian optimizer
- Mono & Multi fidelity sources
- Handling non linear objectives & constraints (black box, no derivative available)
- Equality & inequality constraints (1 ~ 100 constraints)
- Intermediate dimension problem (1 ~ 100 variables)




- Heterogenous variables (continuous, discrete, categorical)
- Costly evaluation (CFD, FEM, objective and/or constraints)
- Based on SMT toolbox for surrogate models
- Remote access via a **web interface**

ONERA WhatsOpt

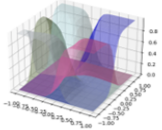
# Two frameworks

SMT

github.com/SMTorg/smt

SMT: Surrogate Modeling Toolbox

- Open source python toolbox: surrogate modeling methods, sampling techniques, and benchmarking functions
- Focus on derivatives (training derivatives used for gradient-enhanced modeling, prediction derivatives)
- New Kriging based surrogate models for higher dimension (KPLS and KPLS-K)
- Noisy Kriging to handle uncertainties on data
- Multifidelity Kriging with or without noise (MFK, MFKPLS)
- Mixture of experts technique for heterogeneous functions



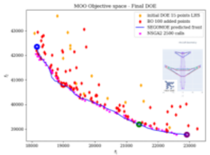
- Mixed integer Kriging to handle **discrete and categorical variables**

## SEGOMOE

- Mono & multi objective Bayesian optimizer
- Mono & Multi fidelity sources
- Handling non linear objectives & constraints (black box, no derivative available)
- Equality & inequality constraints

  (1 ~ 100 constraints)
- Intermediate dimension problem

  (1 ~ 100 variables)



- Heterogenous variables (continuous, discrete, categorical)
- Costly evaluation (CFD, FEM, objective and/or constraints)
- Based on SMT toolbox for surrogate models
- Remote access via a **web interface**

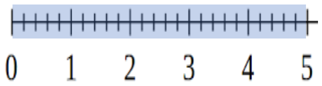
ONERA WhatsOpt

# Models to handle mixed variables (continuous, discrete, categorical)

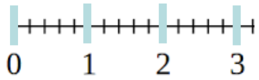
**Hybrid variables**

**Variables types:**

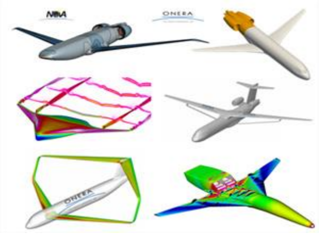
- **Continuous (x) Ex:** wing length




- **Integer (z) Ex:** winglet number




- **Categorical (u) Ex:** Plane shape / material properties

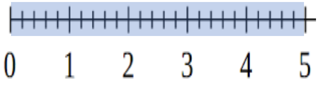


# Models to handle mixed variables (continuous, discrete, categorical)

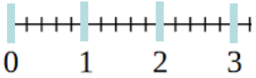
**Hybrid variables**

**Variables types:**

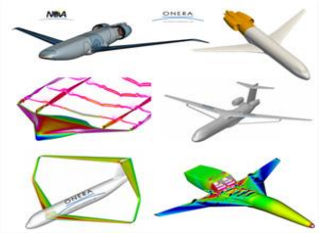
- **Continuous (x) Ex:** wing length




- **Integer (z) Ex:** winglet number




- **Categorical (u) Ex:** Plane shape / material properties



**Categorical variables: n variables, n=2**
u1= shape
u2= color

**Levels:** $L_i$ levels for I in 1,..n, $L_1$=3, $L_2$=2
Levels(u1)= square, circle, rhombus
Levels(u2)= blue, red

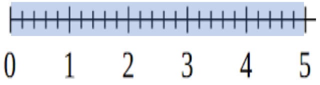
**Categories:** $\prod_{i=1}^{n} L_i$, 2*3=6
- Blue square
- Blue circle
- Blue rhombus
- Red square
- Red circle
- Red rhombus

# Models to handle mixed variables (continuous, discrete, categorical)

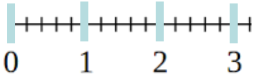
**Hybrid variables**

**Variables types:**

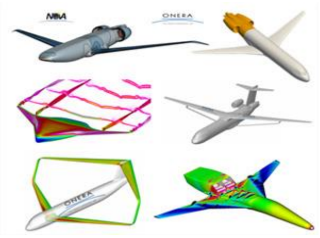
- **Continuous (x) Ex:** wing length




- **Integer (z) Ex:** winglet number




- **Categorical (u) Ex:** Plane shape / material properties



**Categorical variables: n variables, n=2**
u1= shape
u2= color

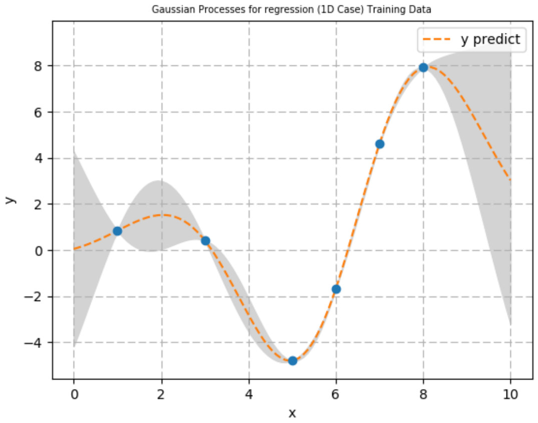
**Levels:** $L_i$ levels for I in 1,..n, $L_1$=3, $L_2$=2
Levels(u1)= square, circle, rhombus
Levels(u2)= blue, red

**Categories:** $\prod_{i=1}^{n} L_i$, 2*3=6
- Blue square
- Blue circle
- Blue rhombus
- Red square
- Red circle
- Red rhombus

6 possibilities

# From continuous to mixed-integer Gaussian process

# From continuous to mixed-integer Gaussian process



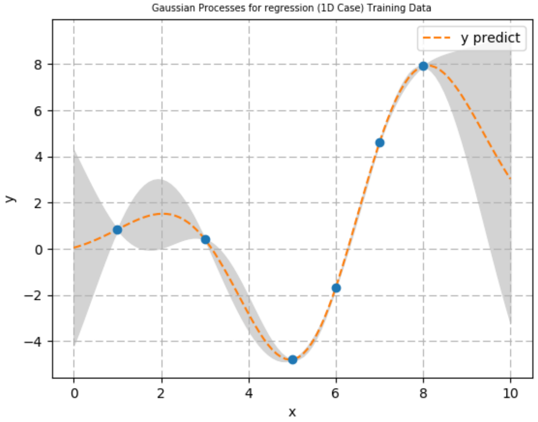

$x \in [0,10]$

## Exponential kernels

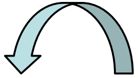
$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|]$$

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|^2]$$

# From continuous to mixed-integer Gaussian process




$x \in [0,10]$



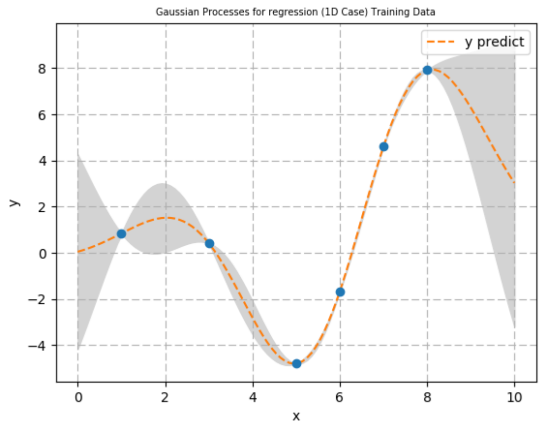
$c$: Blue color / Red color / Green color

## Exponential kernels

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|]$$

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|^2]$$

# From continuous to mixed-integer Gaussian process




$x \in [0,10]$

## Exponential kernels

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|]$$

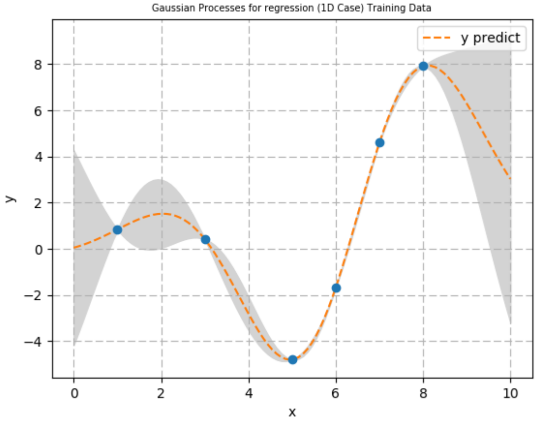
$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|^2]$$

$c$: Blue color / Red color / Green color

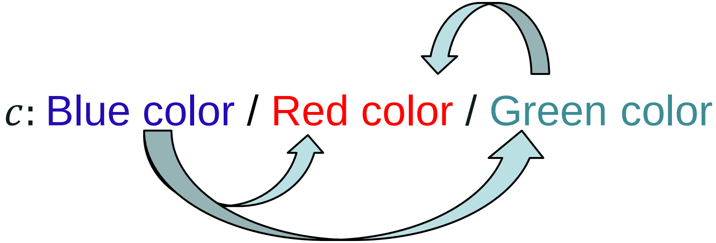
## ? kernel

$k(blue, red)$
$k(blue, green)$
$k(red, green)$

# From continuous to mixed-integer Gaussian process




$x \in [0{,}10]$




## Exponential kernels

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|]$$

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|^2]$$

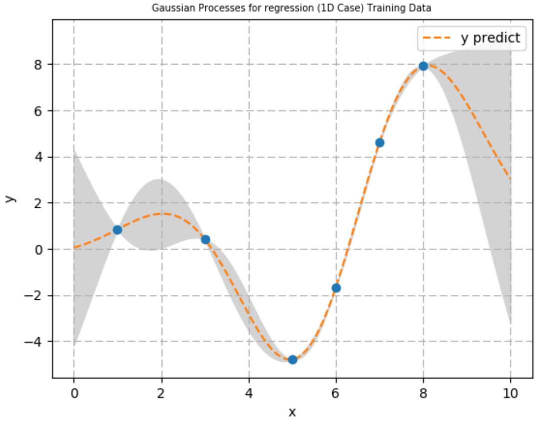
## ? kernel

$k(blue, red)$
$k(blue, green)$
$k(red, green)$

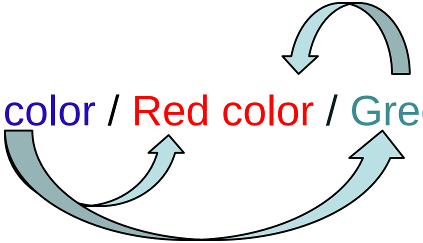
## Correlation matrix

$$\Theta_{color} = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix}$$

# From continuous to mixed-integer Gaussian process




$x \in [0,10]$



$c$: Blue color / Red color / Green color

## Exponential kernels

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|]$$

$$k(x^r, x^s) = \prod_{i=0}^{n} \exp[-\theta_i |x_i^r - x_i^s|^2]$$

To optimize

? kernel

$k(blue, red)$
$k(blue, green)$
$k(red, green)$

## Correlation matrix

$$\Theta_{color} = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix}$$

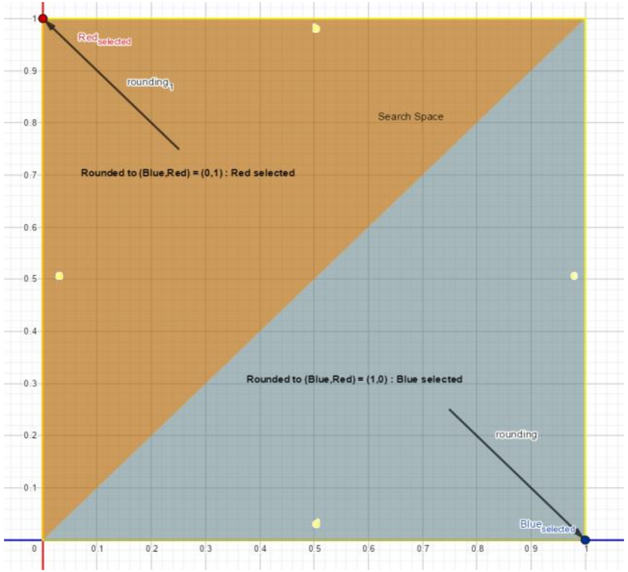
# State-of-the-art approaches

- *Continuous relaxation*

Example with 1 categorical variable and two levels
- Red color
- Blue color

➔ **One-hot encoding:** Categorical variable replaced by two continuous variables denoted by $X_1$ and $X_2$

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1., 0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0., 1.) \Rightarrow$ Red color




E. C. Garrido-Merchán, and D. Hernández-Lobato. "Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes". Neurocomputing, vol. 380 (2020), pages 20-35

# State-of-the-art approaches

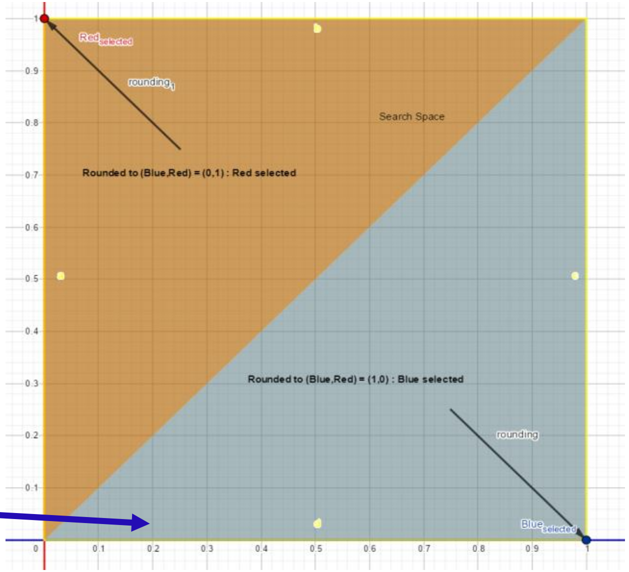
- *Continuous relaxation*

Example with 1 categorical variable and two levels

- Red color
- Blue color

➔ **One-hot encoding:** Categorical variable replaced by two continuous variables denoted by $X_1$ and $X_2$

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1., 0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0., 1.) \Rightarrow$ Red color




E. C. Garrido-Merchán, and D. Hernández-Lobato. “Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes”. Neurocomputing, vol. 380 (2020), pages 20-35

# State-of-the-art approaches

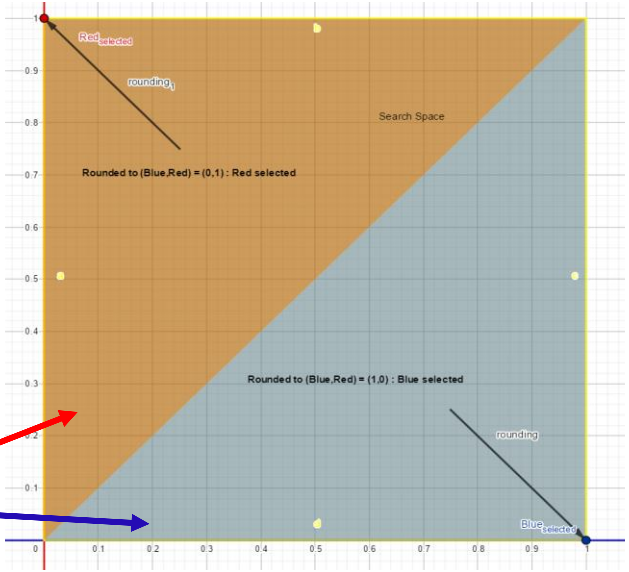
- *Continuous relaxation*

Example with 1 categorical variable and two levels

- Red color
- Blue color

➔ **One-hot encoding:** Categorical variable replaced by two continuous variables denoted by $X_1$ and $X_2$

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1.,0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0.,1.) \Rightarrow$ Red color




E. C. Garrido-Merchán, and D. Hernández-Lobato. "Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes". Neurocomputing, vol. 380 (2020), pages 20-35

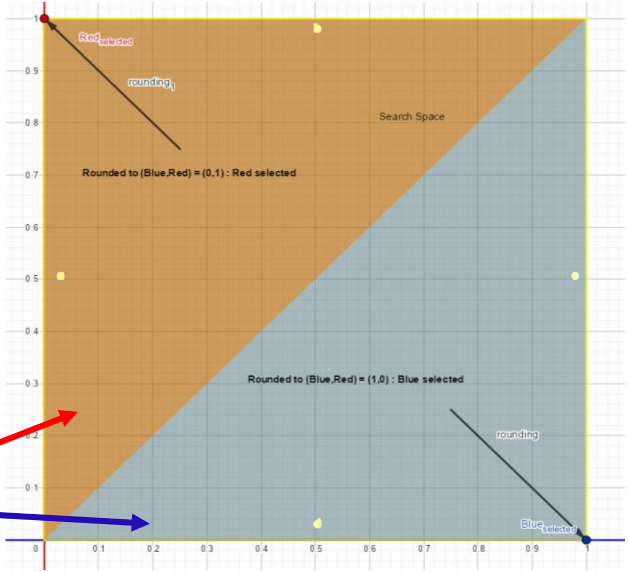
# State-of-the-art approaches

- *Continuous relaxation*

Example with 1 categorical variable and two levels
- Red color
- Blue color

➔ **One-hot encoding:** Categorical variable replaced by two continuous variables denoted by $X_1$ and $X_2$

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1.,0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0.,1.) \Rightarrow$ Red color




**$n$ relaxed dimension**

$$x^r, x^s \in \mathbb{R}^n$$

E. C. Garrido-Merchán, and D. Hernández-Lobato. “Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes”. Neurocomputing, vol. 380 (2020), pages 20-35

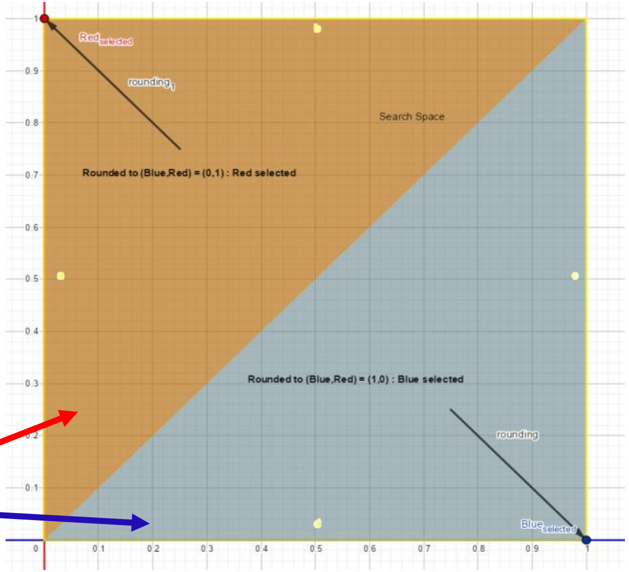
# State-of-the-art approaches

- *Continuous relaxation*

Example with 1 categorical variable and two levels
- Red color
- Blue color

➔ **One-hot encoding:** Categorical variable replaced by two continuous variables denoted by $X_1$ and $X_2$

- If $X_1 > X_2 \Rightarrow e_{c_1^b} = (1.,0.) \Rightarrow$ Blue color
- If $X_1 < X_2 \Rightarrow e_{c_1^r} = (0.,1.) \Rightarrow$ Red color




**$n$ relaxed dimension**

$$x^r, x^s \in \mathbb{R}^n$$

**A continuous kernel**

$$k(x^r, x^s, \theta^{cont}) = \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

E. C. Garrido-Merchán, and D. Hernández-Lobato. “Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes”. Neurocomputing, vol. 380 (2020), pages 20-35

# State-of-the-art approaches

- *Gower distance*

Example with 2 variables:

- $c_1$: Red color / Blue color,
- $x_1 \in [0,1]$

$$\Delta_{gow}\big((red, 0.2), (blue, 0.5)\big) = \boldsymbol{s}(\mathrm{red}, \mathrm{blue}) + |0.2 - 0.5| = 1.3$$
$$\Delta_{gow}\big((red, 0.2), (red, 0.5)\big) = \boldsymbol{s}(\mathrm{red}, \mathrm{red}) + |0.2 - 0.5| = 0.3$$

M. Halstrup. “Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems”. PhD thesis. TU Dortmund, 2016.

# State-of-the-art approaches

- *Gower distance*

Example with 2 variables:

- $c_1$: Red color / Blue color,
- $x_1 \in [0,1]$

$$\Delta_{gow}\big((red, 0.2), (blue, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{blue}) + |0.2 - 0.5| = 1.3$$
$$\Delta_{gow}\big((red, 0.2), (red, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{red}) + |0.2 - 0.5| = 0.3$$

$\boldsymbol{n}$ **continuous dimension**

$$\boldsymbol{x}^r, \boldsymbol{x}^s \in \mathbb{R}^n$$

$\boldsymbol{l}$ **categorical variables**

$$\boldsymbol{c}^r, \boldsymbol{c}^s \in \prod_{j=1}^{j=l} \{ \boldsymbol{L}_1, \dots, \boldsymbol{L}_j\}$$

$$\boldsymbol{w}^r = (\boldsymbol{x}^r, \boldsymbol{c}^r)$$

$$\boldsymbol{w}^s = (\boldsymbol{x}^s, \boldsymbol{c}^s)$$

$$\boldsymbol{s}\left(\boldsymbol{c}_j^r, \boldsymbol{c}_j^s\right) = \begin{cases} \boldsymbol{0}, \text{if } c_j^r = c_j^s \\ \boldsymbol{1}, \text{if } c_j^r \neq c_j^s \end{cases}$$

M. Halstrup. “Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems”. PhD thesis. TU Dortmund, 2016.

# State-of-the-art approaches

- *Gower distance*

Example with 2 variables:

- $c_1$: Red color / Blue color,
- $x_1 \in [0,1]$

$$\Delta_{gow}\big((red, 0.2), (blue, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{blue}) + |0.2 - 0.5| = 1.3$$

$$\Delta_{gow}\big((red, 0.2), (red, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{red}) + |0.2 - 0.5| = 0.3$$

$\boldsymbol{n}$ **continuous dimension**

$$\boldsymbol{x^r}, \boldsymbol{x^s} \in \mathbb{R}^{\boldsymbol{n}}$$

$\boldsymbol{l}$ **categorical variables**

$$\boldsymbol{c^r}, \boldsymbol{c^s} \in \prod_{j=1}^{j=l} \{\boldsymbol{L_1}, \dots, \boldsymbol{L_j}\}$$

$$\boldsymbol{w^r} = (\boldsymbol{x^r}, \boldsymbol{c^r})$$

$$\boldsymbol{w^s} = (\boldsymbol{x^s}, \boldsymbol{c^s})$$

$$\boldsymbol{s}(\boldsymbol{c_j^r}, \boldsymbol{c_j^s}) = \begin{cases} \boldsymbol{0}, \text{if } c_j^r = c_j^s \\ \boldsymbol{1}, \text{if } c_j^r \neq c_j^s \end{cases}$$

**A “mixed” kernel**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \exp\left(-s(c_j^r, c_j^s)\theta_j^{cat} s(c_j^r, c_j^s)\right) \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

M. Halstrup. “Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems”. PhD thesis. TU Dortmund, 2016.

# State-of-the-art approaches

- *Gower distance*

Example with 2 variables:

- $c_1$: Red color / Blue color,
- $x_1 \in [0,1]$

$$\Delta_{gow}\big((red, 0.2), (blue, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{blue}) + |0.2 - 0.5| = 1.3$$
$$\Delta_{gow}\big((red, 0.2), (red, 0.5)\big) = \boldsymbol{s}(\text{red}, \text{red}) + |0.2 - 0.5| = 0.3$$

**$n$ continuous dimension**

$$\boldsymbol{x^r}, \boldsymbol{x^s} \in \mathbb{R}^n$$

**$l$ categorical variables**

$$\boldsymbol{c^r}, \boldsymbol{c^s} \in \prod_{j=1}^{j=l} \{ \boldsymbol{L_1}, \dots, \boldsymbol{L_j} \}$$

$$\boldsymbol{w^r} = (\boldsymbol{x^r}, \boldsymbol{c^r})$$

$$\boldsymbol{w^s} = (\boldsymbol{x^s}, \boldsymbol{c^s})$$

$$\boldsymbol{s}(\boldsymbol{c_j^r}, \boldsymbol{c_j^s}) = \begin{cases} \boldsymbol{0}, \text{ if } c_j^r = c_j^s \\ \boldsymbol{1}, \text{ if } c_j^r \neq c_j^s \end{cases}$$

**A “mixed” kernel**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \exp\left(-s(c_j^r, c_j^s)\theta_j^{cat} s(c_j^r, c_j^s)\right) \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

**A $\theta_j$ for every categorical variable**

**A $\theta_j$ for every cont variable**

M. Halstrup. “Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems”. PhD thesis. TU Dortmund, 2016.

# State-of-the-art approaches

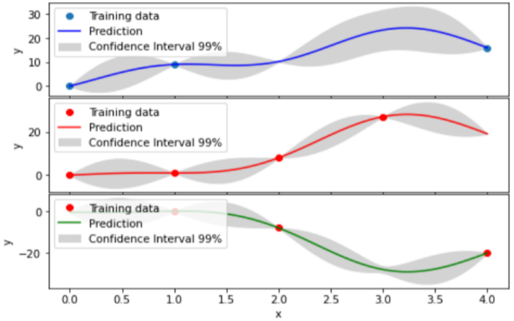
- *Homoscedastic hypersphere mixed kernel*

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$

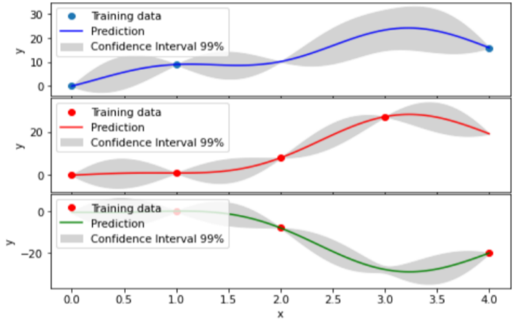



Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$



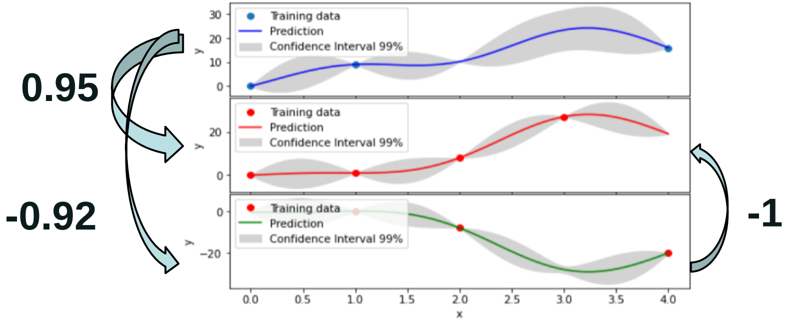

$$\Theta_1 = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix}$$

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$



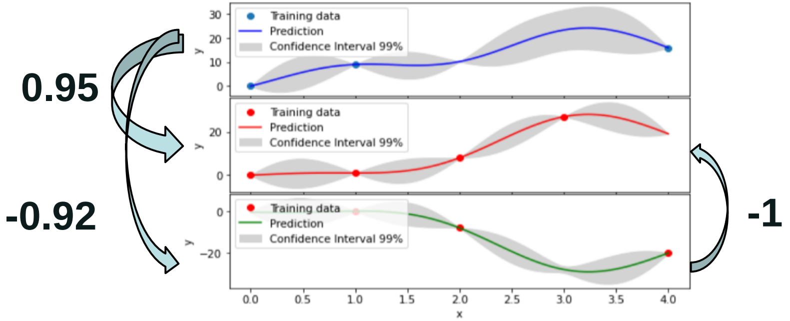

$$\Theta_1 = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.95 & -0.92 \\ 0.95 & 1 & -1 \\ -0.92 & -1 & 1 \end{pmatrix}$$


Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$




$$\Theta_1 = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.95 & -0.92 \\ 0.95 & 1 & -1 \\ -0.92 & -1 & 1 \end{pmatrix}$$

**Mixed kernel**

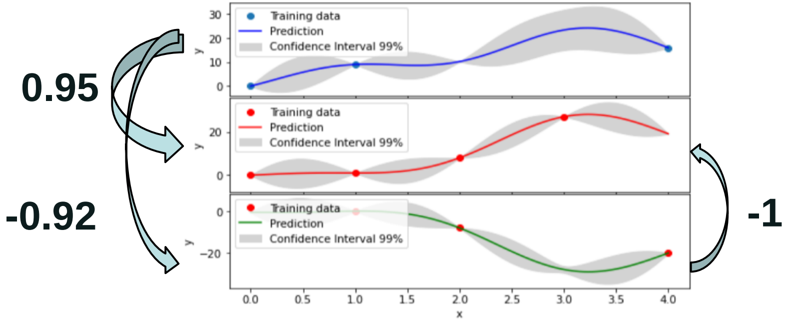
$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c^r_j, c^s_j} \prod_{j=1}^{n} \exp\left(-(x^r_j - x^s_j)\theta^{cont}_j (x^r_j - x^s_j)\right)$$

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$




$$\Theta_1 = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.95 & -0.92 \\ 0.95 & 1 & -1 \\ -0.92 & -1 & 1 \end{pmatrix}$$

**Mixed kernel**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c_j^r, c_j^s} \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

**A new categorical kernel for every categorical $c_j$ variable arises!**

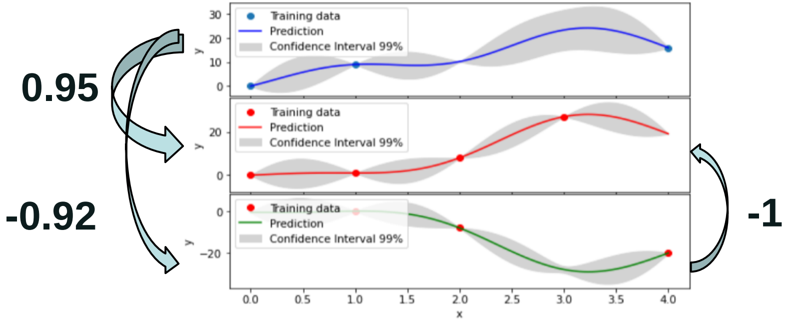
**The matrix $\Theta_j$ should be SPD!**

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# State-of-the-art approaches

- *Homoscedastic hypersphere mixed kernel*

- $c_1$: Blue color / Red color / Green color,
- $x_1 \in [0,4]$




$$\Theta_1 = \begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.95 & -0.92 \\ 0.95 & 1 & -1 \\ -0.92 & -1 & 1 \end{pmatrix}$$

**Mixed kernel**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c_j^r, c_j^s} \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

**A new categorical kernel for every categorical $c_j$ variable arises!**

**The matrix $\Theta_j$ should be SPD!** ⟹ Hypersphere Decomposition

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# Unification of the mixed models

- **Every model can be written as**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c^r_j, c^s_j} \prod_{j=1}^{n} \exp\left(-(x^r_j - x^s_j)\theta^{cont}_j (x^r_j - x^s_j)\right)$$

# Unification of the mixed models

- **Every model can be written as**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c_j^r, c_j^s} \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

$$= \prod_{j=1}^{l} [R_j(\Theta_j)]_{\ell_r^j, \ell_s^j}$$

**The input $c_j^r$ takes the level of index $l_r^j$**

- $c_1$: Blue color / Red color / Green color,

  1 2 3

**if** $c_1^r = \text{"}red\text{"},\ l_r^1 = 2\ \ (index)$

$$\begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \theta_{1,2} & \theta_{1,3} \\ & 1 & \theta_{2,3} \\ Sym & & 1 \end{pmatrix}$$

# Unification of the mixed models

- **Every model can be written as**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c_j^r, c_j^s} \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

$$= \prod_{j=1}^{l} [R_j(\Theta_j)]_{\ell_r^j, \ell_s^j}$$

$$= \prod_{j=1}^{l} \kappa([\Phi(\Theta_j)]_{\ell_j^r, \ell_j^s})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s, \ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^r, \ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s, \ell_j^s})$$

**The input $c_j^r$ takes the level of index $l_r^j$**

- $c_1$: Blue color / Red color / Green color,

  1 2 3

**if $c_1^r$ = "red", $l_r^1 = 2$ (index)**

$$\begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \theta_{1,2} & \theta_{1,3} \\ & 1 & \theta_{2,3} \\ Sym & & 1 \end{pmatrix}$$

# Unification of the mixed models

- **Every model can be written as**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c^r_j, c^s_j} \prod_{j=1}^{n} \exp\left(-(x^r_j - x^s_j)\theta^{cont}_j (x^r_j - x^s_j)\right)$$

$$= \prod_{j=1}^{l} [R_j(\Theta_j)]_{\ell^j_r, \ell^j_s}$$

$$= \prod_{j=1}^{l} \kappa([\Phi(\Theta_j)]_{\ell^r_j,\ell^s_j})\ \kappa([\Phi(\Theta_j)]_{\ell^s_j,\ell^r_j})\ \kappa([\Phi(\Theta_j)]_{\ell^r_j,\ell^r_j})\ \kappa([\Phi(\Theta_j)]_{\ell^s_j,\ell^s_j})$$

SPD SPD

**The input $c^r_j$ takes the level of index $l^j_r$**

- $c_1$: Blue color / Red color / Green color,

 1 2 3

if $c^r_1 = \text{"red"}$, $l^1_r = 2\ (index)$

$$\begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \theta_{1,2} & \theta_{1,3} \\ & 1 & \theta_{2,3} \\ Sym & & 1 \end{pmatrix}$$

# Unification of the mixed models

- **Every model can be written as**

$$k(w^r, w^s, \theta) = \prod_{j=1}^{l} \theta^{cat}_{c_j^r, c_j^s} \prod_{j=1}^{n} \exp\left(-(x_j^r - x_j^s)\theta_j^{cont}(x_j^r - x_j^s)\right)$$

$$= \prod_{j=1}^{l} [R_j(\Theta_j)]_{\ell_r^j, \ell_s^j}$$

$$= \prod_{j=1}^{l} \kappa([\Phi(\Theta_j)]_{\ell_j^r, \ell_j^s})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s, \ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^r, \ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s, \ell_j^s})$$

**SPD** **SPD**

**The input $c_j^r$ takes the level of index $l_r^j$**

- $c_1$: Blue color / Red color / Green color,

**1** **2** **3**

**if $c_1^r = \text{"red"}, l_r^1 = 2\ (index)$**

$$\begin{pmatrix} 1 & \theta_{blue/red} & \theta_{blue/green} \\ & 1 & \theta_{red/green} \\ Sym & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \theta_{1,2} & \theta_{1,3} \\ & 1 & \theta_{2,3} \\ Sym & & 1 \end{pmatrix}$$

*Example: Homoscedastic hypersphere mixed kernel*

- $[\Phi(\Theta_j)]_{r,s} = [C(\Theta_j)C(\Theta_j)^T]_{r,s}$ → Hypersphere Decomposition
- $\kappa = \mathbb{1}_{L_i}$

# The naïve Exponential kernel (Full model)

- *Extension of the Exponential kernel*

$$[R_i(\Phi(\Theta_i))]_{\ell^i_r,\ell^i_s} = \exp\left( - \sum_{j=1}^{L_i} \sum_{j'=1}^{L_i} \left| [e_{c^r_i} - e_{c^s_i}]_j \right|^{p/2} [\Phi(\Theta_i)]_{j,j'} \left| [e_{c^r_i} - e_{c^s_i}]_{j'} \right|^{p/2} \right)$$

$$[\boldsymbol{R_i}]_{\boldsymbol{r,s}}$$

# The naïve Exponential kernel (Full model)

- *Extension of the Exponential kernel*

$$[R_i(\Phi(\Theta_i))]_{\ell^i_r,\ell^i_s} = \exp\left(-\sum_{i=1}^{L_i}\sum_{i'=1}^{L_i} \left|[e_{c^r_i} - e_{c^s_i}]_j\right|^{p/2} [\Phi(\Theta_i)]_{j,j'} \left|[e_{c^r_i} - e_{c^s_i}]_{j'}\right|^{p/2}\right)$$

$$[\boldsymbol{R_i}]_{\boldsymbol{r,s}} = \exp\left(-\left([\Phi(\Theta_i)]_{\ell^i_r,\ell^i_r} + [\Phi(\Theta_i)]_{\ell^i_s,\ell^i_s} + [\Phi(\Theta_i)]_{\ell^i_r,\ell^i_s} + [\Phi(\Theta_i)]_{\ell^i_s,\ell^i_r}\right)\right)$$

# The naïve Exponential kernel (Full model)

- *Extension of the Exponential kernel*

$$[R_i(\Phi(\Theta_i))]_{\ell_r^i,\ell_s^i} = \exp\left(-\sum_{i=1}^{L_i}\sum_{i'=1}^{L_i} \left|[e_{c_i^r} - e_{c_i^s}]_j\right|^{p/2} [\Phi(\Theta_i)]_{j,j'} \left|[e_{c_i^r} - e_{c_i^s}]_{j'}\right|^{p/2}\right)$$

$$[\boldsymbol{R_i}]_{r,s} = \exp\left(-\left([\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_r^i}\right)\right)$$

$$= \exp\left(-[\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} - [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i}\right) \exp\left(-2[\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i}\right)$$

# The naïve Exponential kernel (Full model)

- *Extension of the Exponential kernel*

$$[R_i(\Phi(\Theta_i))]_{\ell_r^i,\ell_s^i} = \exp\left(-\sum_{i=1}^{L_i}\sum_{i'=1}^{L_i} \left|[e_{c_i^r} - e_{c_i^s}]_j\right|^{p/2} [\Phi(\Theta_i)]_{j,j'} \left|[e_{c_i^r} - e_{c_i^s}]_{j'}\right|^{p/2}\right)$$

$$[\boldsymbol{R}_i]_{r,s} = \exp\left(-\left([\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_r^i}\right)\right)$$

$$= \exp\left(-[\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} - [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i}\right) \exp\left(-2[\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i}\right)$$

**Continuous Relaxation (CR)** $[\boldsymbol{S}_i]_{r,s}$

**Exponential Homoscedastic Hypersphere (EHH)** $[\boldsymbol{T}_i]_{r,s}$

# The naïve Exponential kernel  (Full model)

- *Extension of the Exponential kernel*

$$[R_i(\Phi(\Theta_i))]_{\ell_r^i,\ell_s^i} = \exp\left(-\sum_{i=1}^{L_i}\sum_{i'=1}^{L_i} \left|[e_{c_i^r} - e_{c_i^s}]_j\right|^{p/2} [\Phi(\Theta_i)]_{j,j'} \left|[e_{c_i^r} - e_{c_i^s}]_{j'}\right|^{p/2}\right)$$

$$[\boldsymbol{R_i}]_{\boldsymbol{r,s}} = \exp\left(-\left([\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i} + [\Phi(\Theta_i)]_{\ell_s^i,\ell_r^i}\right)\right)$$

$$= \underbrace{\exp\left(-[\Phi(\Theta_i)]_{\ell_r^i,\ell_r^i} - [\Phi(\Theta_i)]_{\ell_s^i,\ell_s^i}\right)}_{\text{Continuous Relaxation (CR)}} \underbrace{\exp\left(-2[\Phi(\Theta_i)]_{\ell_r^i,\ell_s^i}\right)}_{\text{Exponential Homoscedastic Hypersphere (EHH)}}$$

**Continuous Relaxation (CR)** $[\boldsymbol{S_i}]_{\boldsymbol{r,s}}$

**Exponential Homoscedastic Hypersphere (EHH)** $[\boldsymbol{T_i}]_{\boldsymbol{r,s}}$

**Warning:** $\exp(\boldsymbol{x}) > \boldsymbol{0}$

# The naïve Exponential kernel (Full model)

- *Construction of the correlation matrix* $[R_i]$

Rebonato, R. and Jäckel, P. (1999) The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. Journal of Risk, 2, 17-28.

# The naïve Exponential kernel (Full model)

- *Construction of the correlation matrix* $[R_i]$

$$[R_i]_{r,s} = [S_i]_{r,s}[T_i]_{r,s} \quad \text{if } r \neq s$$
$$[R_i]_{r,r} = 1,$$

Rebonato, R. and Jäckel, P. (1999) The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. Journal of Risk, 2, 17-28.

# The naïve Exponential kernel (Full model)

- *Construction of the correlation matrix* $[R_i]$

$$[R_i]_{r,s} = [S_i]_{r,s}[T_i]_{r,s} \quad \text{if } r \neq s$$
$$[R_i]_{r,r} = 1,$$

**Continuous Relaxation diagonal kernel**

$$[S_i]_{r,s} = \prod_{j=1}^{L_i} \exp\left(-[\Phi(\Theta_i)]_{j,j}\left((e_{c_i^r})_j - (e_{c_i^s})_j\right)^2\right)$$

**Exponential Homoscedastic Hypersphere**

$$[T_i]_{r,s} = \exp\left(-2[\Phi(\Theta_i)]_{r,s}\right)$$

$$[\Phi(\Theta_i)]_{j,j} := [\Theta_i]_{j,j} \geq 0$$
$$[\Phi(\Theta_i)]_{j,j'} := \frac{\log \epsilon}{2}([C(\Theta_i)C(\Theta_i)^\top]_{j,j'} - 1) \quad \text{if } j \neq j',$$

Rebonato, R. and Jäckel, P. (1999) The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. Journal of Risk, 2, 17-28.

# The naïve Exponential kernel (Full model)

- *Construction of the correlation matrix* $[R_i]$

$$[R_i]_{r,s} = [S_i]_{r,s}[T_i]_{r,s} \quad \text{if } r \neq s$$
$$[R_i]_{r,r} = 1,$$

**Continuous Relaxation diagonal kernel**

$$[S_i]_{r,s} = \prod_{j=1}^{L_i} \exp\left(-[\Phi(\Theta_i)]_{j,j}\left((e_{c_i^r})_j - (e_{c_i^s})_j\right)^2\right)$$

$$[S_i]_{r,s} \geq 0$$
$$[S_i]_{r,s} \leq 1$$

**Exponential Homoscedastic Hypersphere**

$$[T_i]_{r,s} = \exp\left(-2[\Phi(\Theta_i)]_{r,s}\right)$$

$$[T_i]_{r,s} = \epsilon \exp\left(-(\log \epsilon)[C_i C_i^T]_{r,s}\right)$$
$$[T_i]_{r,s} = \epsilon \exp\left(-(\log \epsilon)[C_i C_i^T]_{r,s}\right) \geq \epsilon \exp 0 = \epsilon$$
$$[T_i]_{r,s} = \epsilon \exp\left(-(\log \epsilon)[C_i C_i^T]_{r,s}\right) \leq \frac{\epsilon}{\epsilon} = 1$$

$$[\Phi(\Theta_i)]_{j,j} := [\Theta_i]_{j,j} \geq 0$$
$$[\Phi(\Theta_i)]_{j,j'} := \frac{\log \epsilon}{2}([C(\Theta_i)C(\Theta_i)^\top]_{j,j'} - 1) \quad \text{if } j \neq j',$$

$1 \gg \epsilon > 0$

$C_i$: **Homoscedastic hypersphere**

Rebonato, R. and Jäckel, P. (1999) The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. Journal of Risk, 2, 17-28.

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i (L_i - 1)$ |

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our full model (FE) | $\begin{pmatrix} [\Theta_i]_{11} & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & [\Theta_i]_{22} & [\Theta_i]_{23} \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))\ \exp(-2\,[\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i + 1)$ |

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our full model (FE) | $\begin{pmatrix} [\Theta_i]_{11} & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & [\Theta_i]_{22} & [\Theta_i]_{23} \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))\ \exp(-2[\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i + 1)$ |
| Our model as Exponential Homoscedastic Hypersphere (EHH) | $\begin{pmatrix} 0 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 0 & [\Theta_i]_{23} \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2[\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i - 1)$ |

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our full model (FE) | $\begin{pmatrix} [\Theta_i]_{11} & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & [\Theta_i]_{22} & [\Theta_i]_{23} \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s})) \exp(-2 [\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i + 1)$ |
| Our model as Exponential Homoscedastic Hypersphere (EHH) | $\begin{pmatrix} 0 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 0 & [\Theta_i]_{23} \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2 [\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our model as Continuous Relaxation (CR) | $\begin{pmatrix} [\Theta_i]_{11} & 0 & 0 \\ & [\Theta_i]_{22} & 0 \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))$ | $L_i$ |

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our full model (FE) | $\begin{pmatrix} [\Theta_i]_{11} & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & [\Theta_i]_{22} & [\Theta_i]_{23} \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))\ \exp(-2\,[\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i + 1)$ |
| Our model as Exponential Homoscedastic Hypersphere (EHH) | $\begin{pmatrix} 0 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 0 & [\Theta_i]_{23} \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2\,[\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i(L_i - 1)$ |
| Our model as Continuous Relaxation (CR) | $\begin{pmatrix} [\Theta_i]_{11} & 0 & 0 \\ & [\Theta_i]_{22} & 0 \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))$ | $L_i$ |
| Our model as Gower distance (GD) | $[\Theta_i]_{cov} \begin{pmatrix} 0 & 1 & 1 \\ & 0 & 1 \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2\,[\Theta_i]_{cov})$ | 1 |

# Modeling categorical kernels

| Model | $\Theta_i$ (example with 3 levels) | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of parameters |
|---|---|---|---|
| Homoscedastic Hypersphere (HH) | $\begin{pmatrix} 1 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 1 & [\Theta_i]_{23} \\ \text{Sym.} & & 1 \end{pmatrix}$ | $[\Theta_i]_{c_i^r, c_i^s}$ | $\frac{1}{2} L_i (L_i - 1)$ |
| Our full model (FE) | $\begin{pmatrix} [\Theta_i]_{11} & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & [\Theta_i]_{22} & [\Theta_i]_{23} \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s})) \exp(-2 [\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i (L_i + 1)$ |
| Our model as Exponential Homoscedastic Hypersphere (EHH) | $\begin{pmatrix} 0 & [\Theta_i]_{12} & [\Theta_i]_{13} \\ & 0 & [\Theta_i]_{23} \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2 [\Theta_i]_{c_i^r, c_i^s})$ | $\frac{1}{2} L_i (L_i - 1)$ |
| Our model as Continuous Relaxation (CR) | $\begin{pmatrix} [\Theta_i]_{11} & 0 & 0 \\ & [\Theta_i]_{22} & 0 \\ \text{Sym.} & & [\Theta_i]_{33} \end{pmatrix}$ | $\exp(-([\Theta_i]_{c_i^r, c_i^r} + [\Theta_i]_{c_i^s, c_i^s}))$ | $L_i$ |
| Our model as Gower distance (GD) | $[\Theta_i]_{cov} \begin{pmatrix} 0 & 1 & 1 \\ & 0 & 1 \\ \text{Sym.} & & 0 \end{pmatrix}$ | $\exp(-2 [\Theta_i]_{cov})$ | 1 |

# Summary for $\kappa$ and $\Phi$

## Every model can be written as

$$[R_j(\Theta_j)]_{\ell_r^j,\ell_s^j} = \prod_{j=1}^{l} \kappa([\Phi(\Theta_j)]_{\ell_j^r,\ell_j^s})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s,\ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^r,\ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s,\ell_j^s})$$

| Name | $\kappa$ | $\Phi(\Theta_i)$ |
|---|---|---|
| GD | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := \frac{1}{2}\theta_i \quad ; \quad [\Phi(\Theta_i)]_{j\neq j'} := 0$ |
| CR | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := [\Theta_i]_{j,j} \quad ; \quad [\Phi(\Theta_i)]_{j\neq j'} := 0$ |
| EHH | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := 0 \quad ; \quad [\Phi(\Theta_i)]_{j\neq j'} := \frac{\log \epsilon}{2}([C(\Theta_i)C(\Theta_i)^\top]_{j,j'} - 1)$ |
| HH | $\mathbb{I}_{L_i}$ | $[\Phi(\Theta_i)]_{j,j} := 1 \quad ; \quad [\Phi(\Theta_i)]_{j\neq j'} := ([C(\Theta_i)C(\Theta_i)^\top]_{j,j'}$ |

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

# Summary for $\kappa$ and $\Phi$

## Every model can be written as

$$[R_j(\Theta_j)]_{\ell_r^j,\ell_s^j} = \prod_{j=1}^{l} \kappa([\Phi(\Theta_j)]_{\ell_j^r,\ell_j^s})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s,\ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^r,\ell_j^r})\ \kappa([\Phi(\Theta_j)]_{\ell_j^s,\ell_j^s})$$

| Name | $\kappa$ | $\Phi(\Theta_i)$ |
|---|---|---|
| GD | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := \frac{1}{2}\theta_i$ ; $[\Phi(\Theta_i)]_{j\neq j'} := 0$ |
| CR | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := [\Theta_i]_{j,j}$ ; $[\Phi(\Theta_i)]_{j\neq j'} := 0$ |
| EHH | $\exp(.)$ | $[\Phi(\Theta_i)]_{j,j} := 0$ ; $[\Phi(\Theta_i)]_{j\neq j'} := \frac{\log \epsilon}{2}([C(\Theta_i)C(\Theta_i)^\top]_{j,j'} - 1)$ |
| HH | $\mathbb{I}_{L_i}$ | $[\Phi(\Theta_i)]_{j,j} := 1$ ; $[\Phi(\Theta_i)]_{j\neq j'} := ([C(\Theta_i)C(\Theta_i)^\top]_{j,j'}$ |

**Our new kernel EHH is Gaussian and categorical!**

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, Springer International Publishing, 2020, pp. 189–224

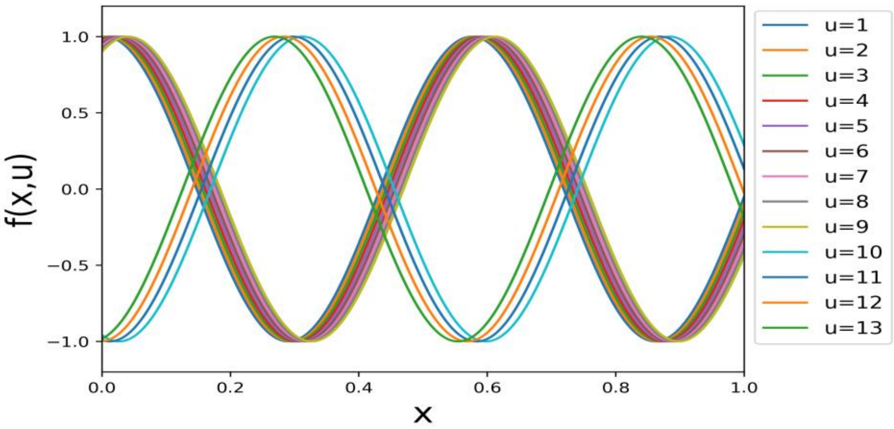
# Results

- *An illustration case*

2 design variables

- 1 continuous variable
- 1 categorical variable: 13 levels
- 14 dimensions

- DoE: LHS with 98 points (7 per dimension)




Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

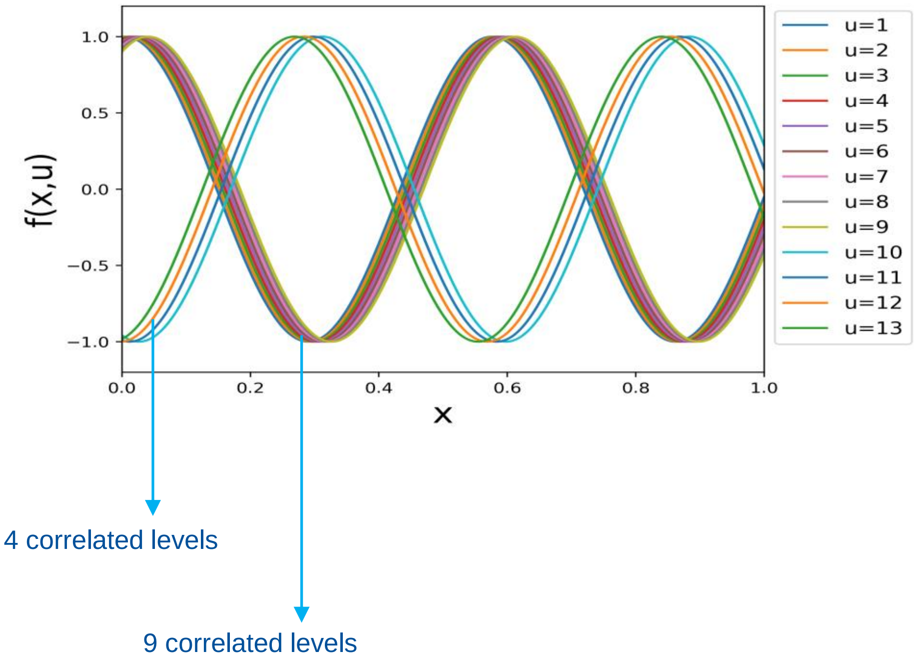
# Results

- *An illustration case*

2 design variables

- 1 continuous variable
- 1 categorical variable: 13 levels
- 14 dimensions

- DoE: LHS with 98 points (7 per dimension)



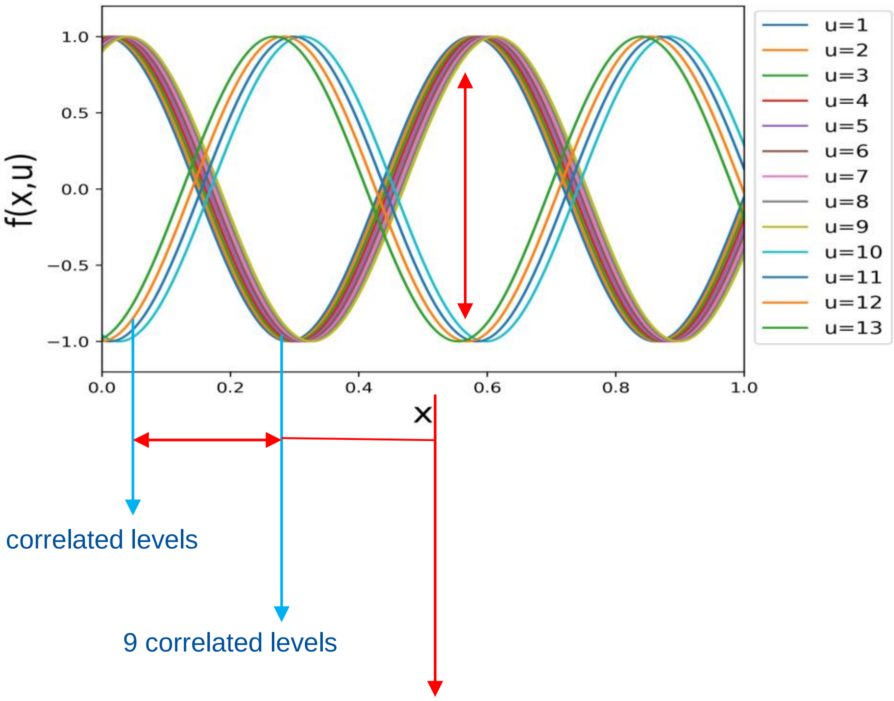

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

- *An illustration case*

2 design variables
- 1 continuous variable
- 1 categorical variable: 13 levels
- 14 dimensions

- DoE: LHS with 98 points (7 per dimension)




4 correlated levels

9 correlated levels

Anticorrelated groups

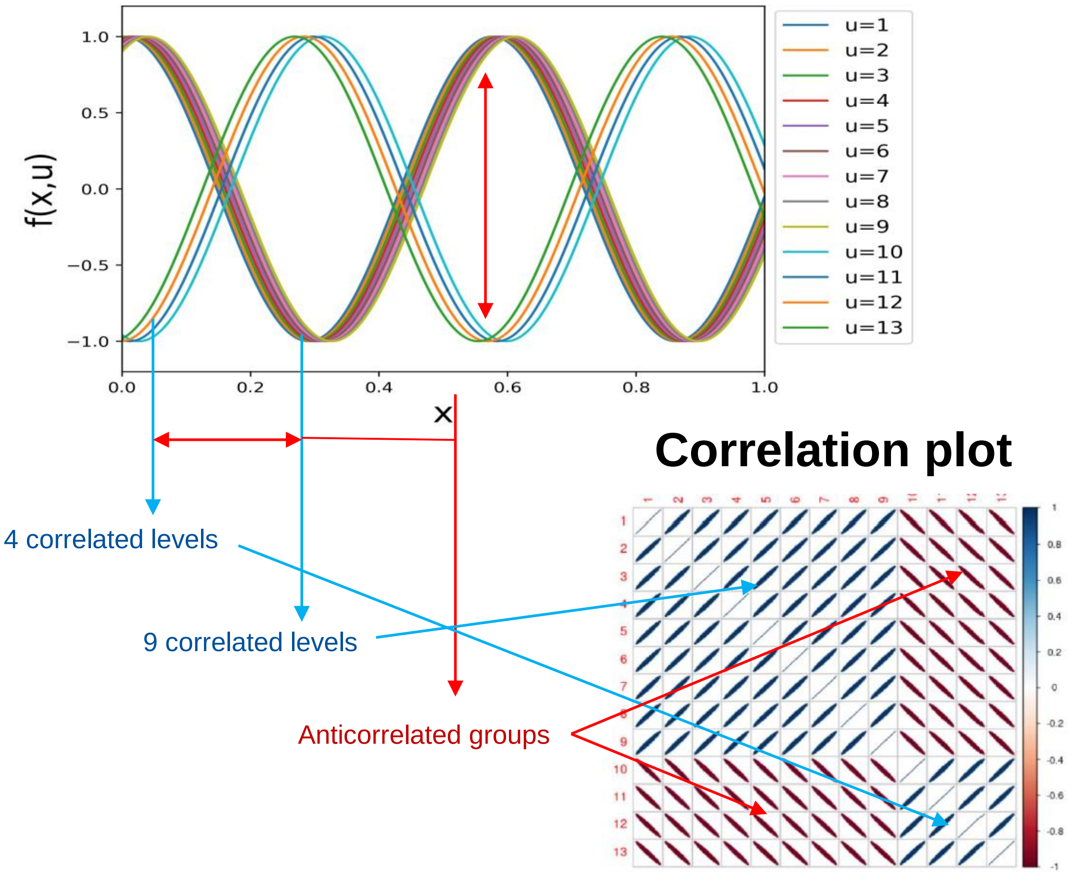
Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

- *An illustration case*

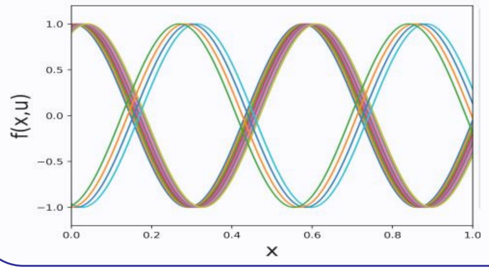
2 design variables
- 1 continuous variable
- 1 categorical variable: 13 levels
- 14 dimensions

- DoE: LHS with 98 points (7 per dimension)




Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results




**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

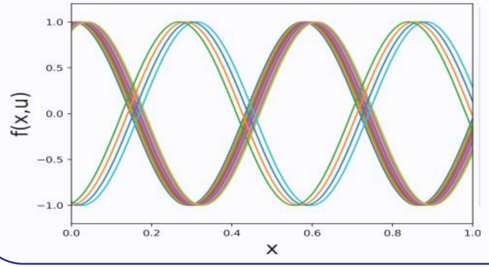
**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Full model**

- 92 hyperparameters
- Time = 642 seconds
- RMSE = 22.6



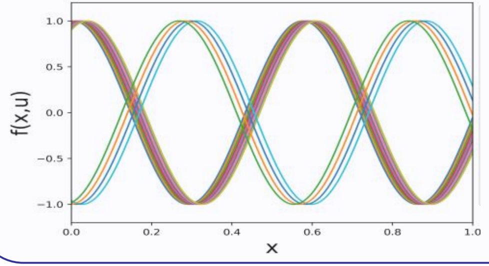

**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

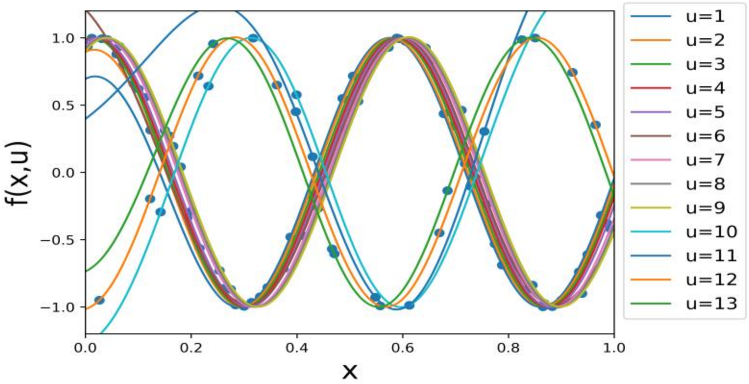


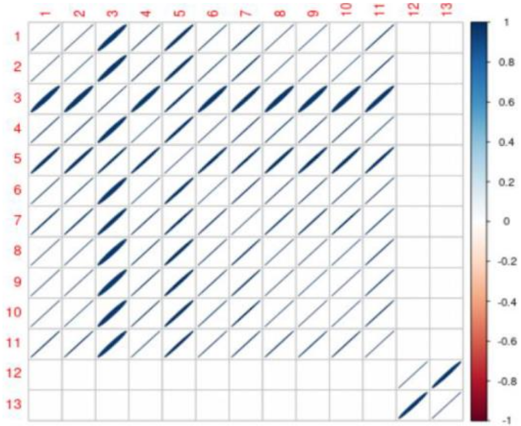

2 variables
- One continuous variable
- One categorical variable: 13 levels

⇒ 14 dimensions

DoE : LHS (n = 98)

**Full model**
- 92 hyperparameters
- Time = 642 seconds
- RMSE = 22.6

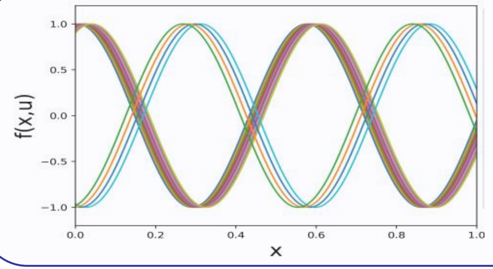


Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

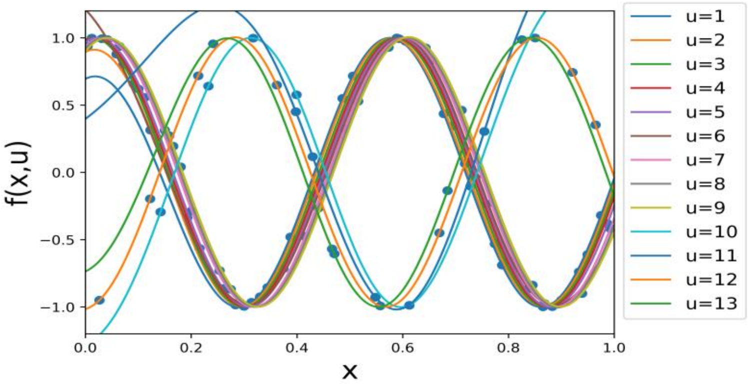
**Full model**
- 92 hyperparameters
- Time = 642 seconds
- RMSE = 22.6

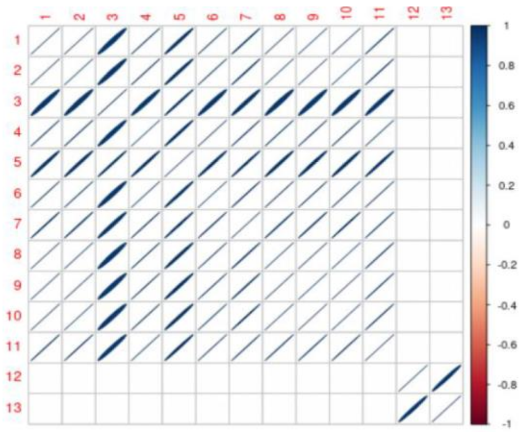



**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**

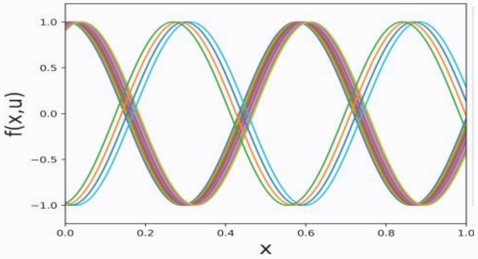


**Many hyperparameters are useless! Expensive with many numerical issues.**


Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results




**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

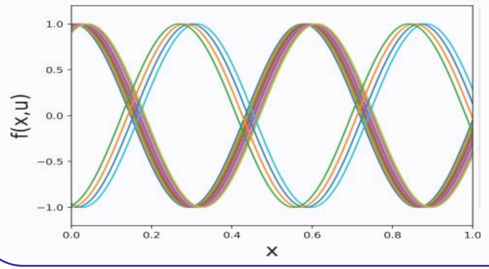
**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Gower case**

- 2 hyperparameters
- Time = 1.4 seconds
- RMSE = 30.1




**2 variables**

- **One continuous variable**
- **One categorical variable: 13 levels**

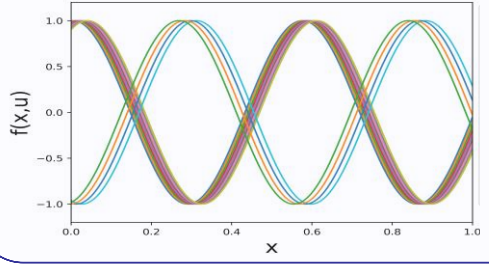
⇒ **14 dimensions**

**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

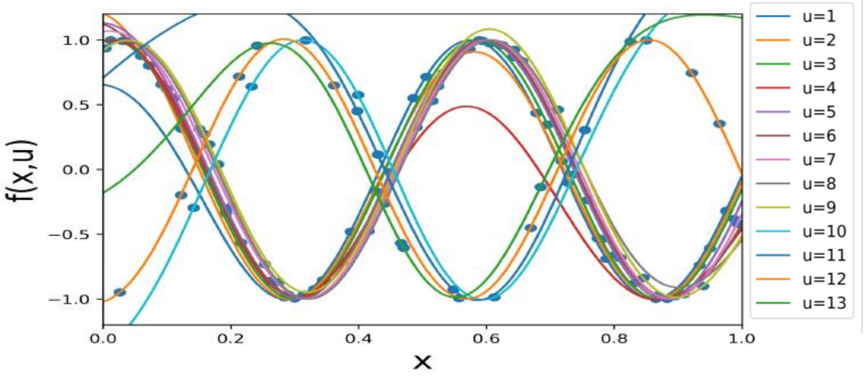
**Gower case**
- 2 hyperparameters
- Time = 1.4 seconds
- RMSE = 30.1

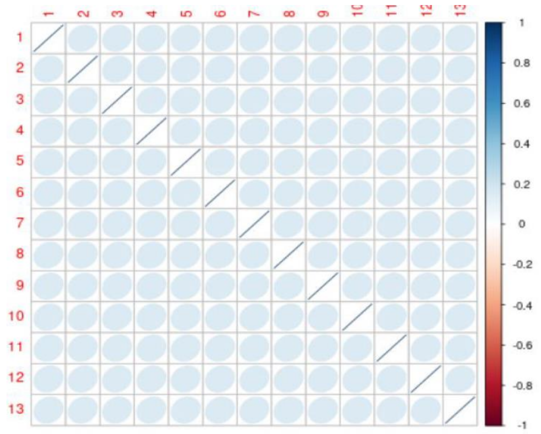



**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**



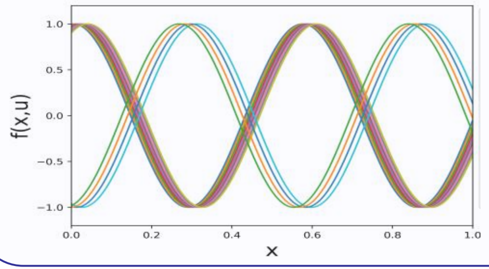



Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Gower case**

- 2 hyperparameters
- Time = 1.4 seconds
- RMSE = 30.1

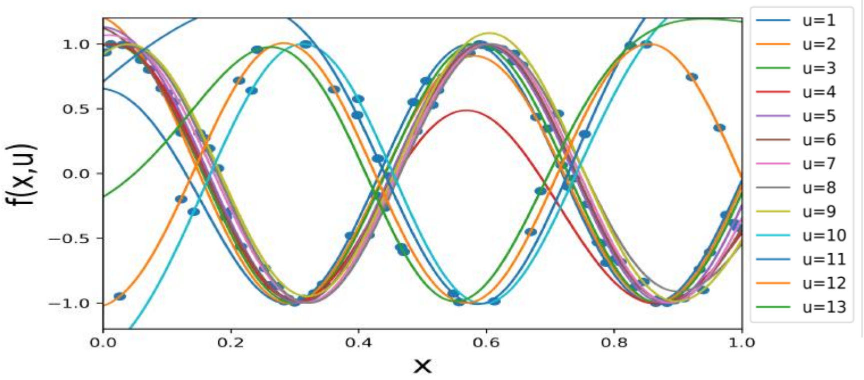


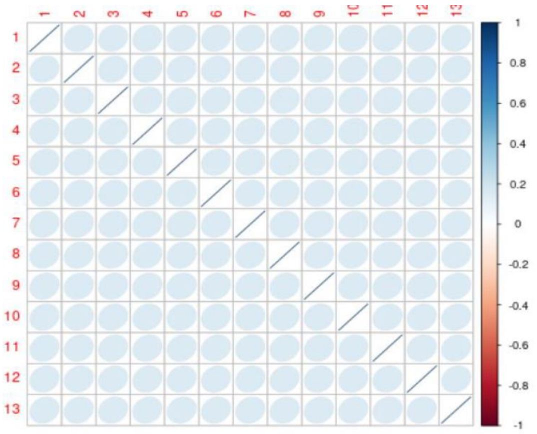

**2 variables**

- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**

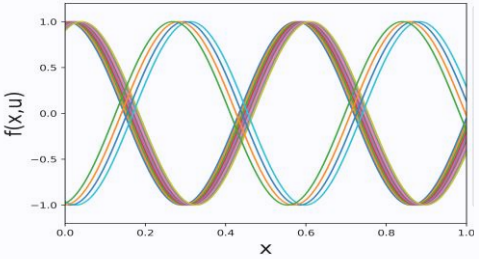



**Only few hyperpameters are used which may not be enough.**


Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results




**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

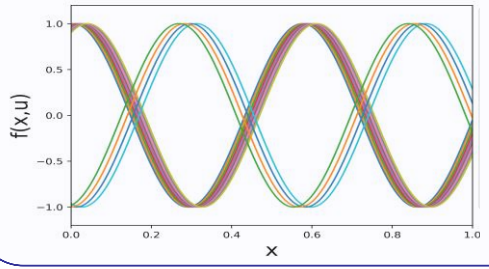
**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Continuous relaxation case**

- 14 hyperparameters
- Time = 24.5 seconds
- RMSE = 22.3




**2 variables**

- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

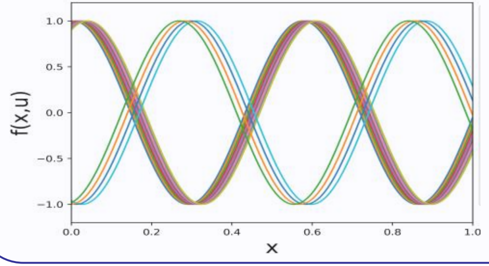
**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

## Continuous relaxation case

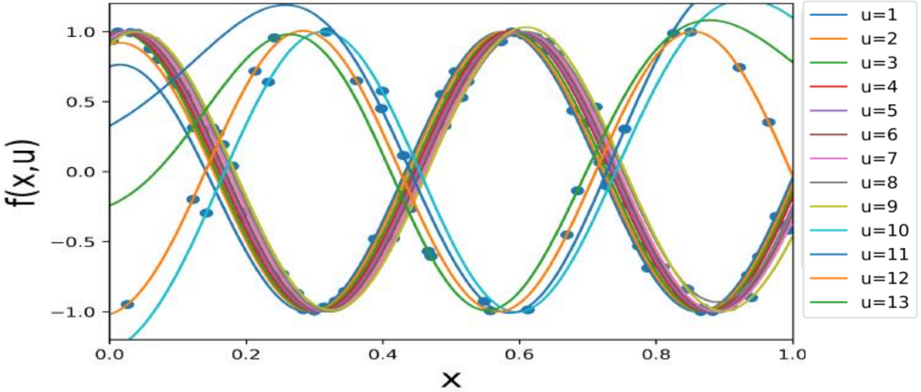
- 14 hyperparameters
- Time = 24.5 seconds
- RMSE = 22.3

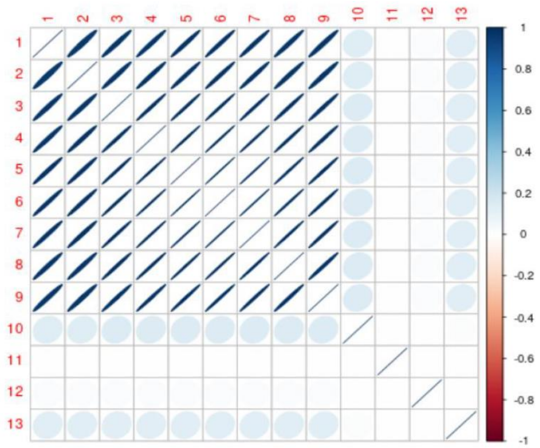



**2 variables**
- One continuous variable
- One categorical variable: 13 levels

⇒ 14 dimensions

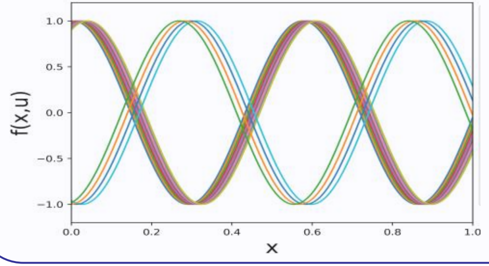
DoE : LHS (n = 98)




Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results



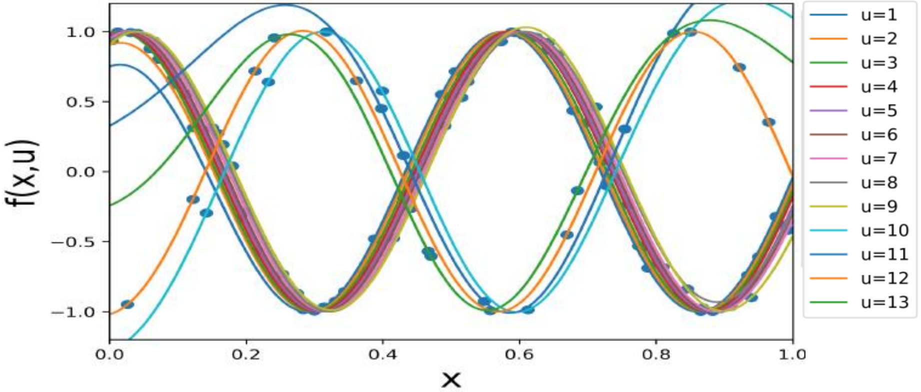

2 variables
- One continuous variable
- One categorical variable: 13 levels

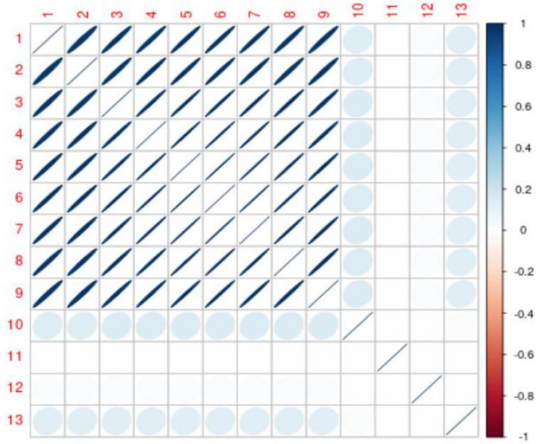
⇒ 14 dimensions

DoE : LHS (n = 98)

## Continuous relaxation case

- 14 hyperparameters
- Time = 24.5 seconds
- RMSE = 22.3

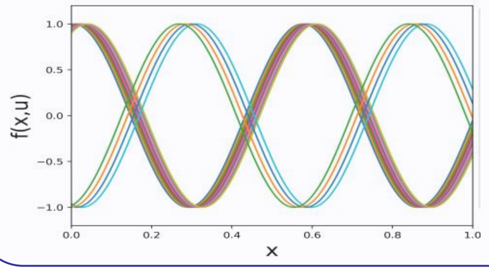


**Best trade-off: number of hyperparameters vs model accuracy!**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results




**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

$\Rightarrow$ **14 dimensions**

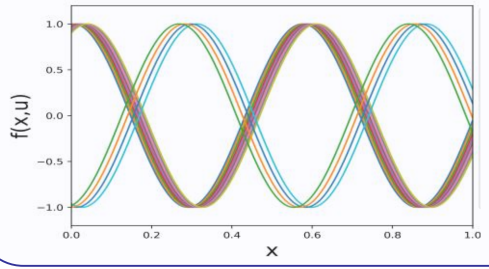
**DoE : LHS (n = 98)**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Exponential Homoscedastic hypersphere case**

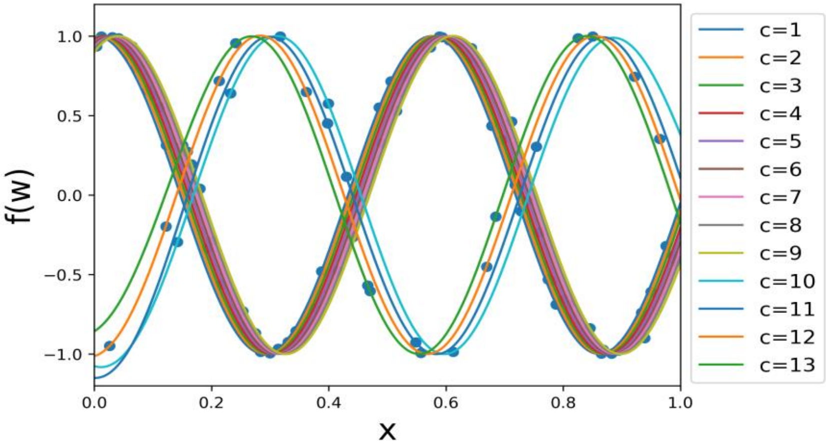
- 79 hyperparameters
- 514.5 seconds
- RMSE = 1.8

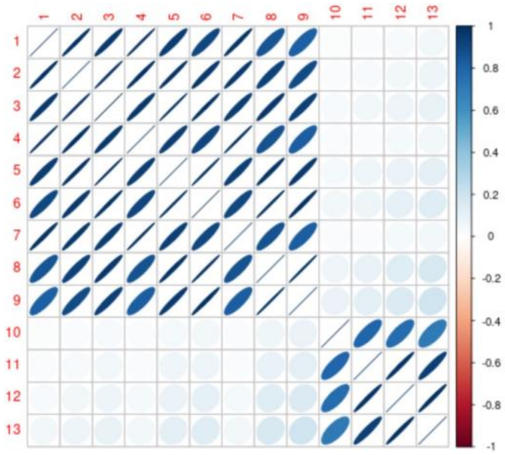



**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**

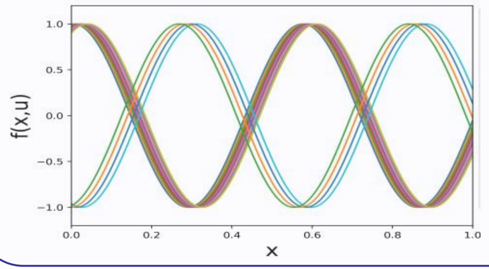


Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

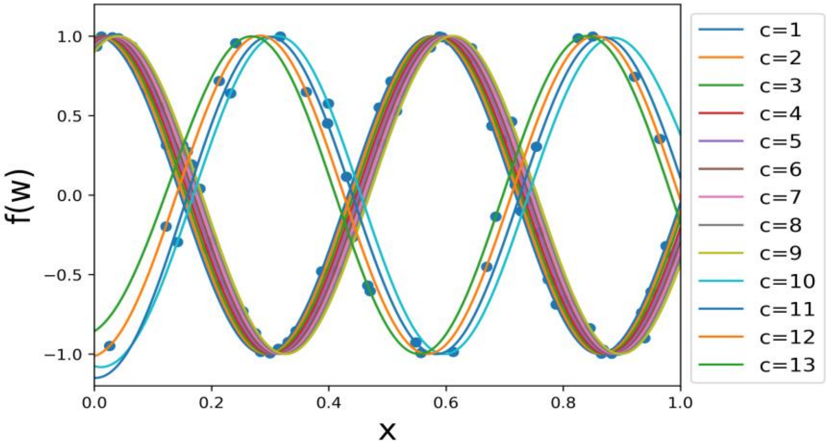
**Exponential Homoscedastic hypersphere case**
- 79 hyperparameters
- 514.5 seconds
- RMSE = 1.8



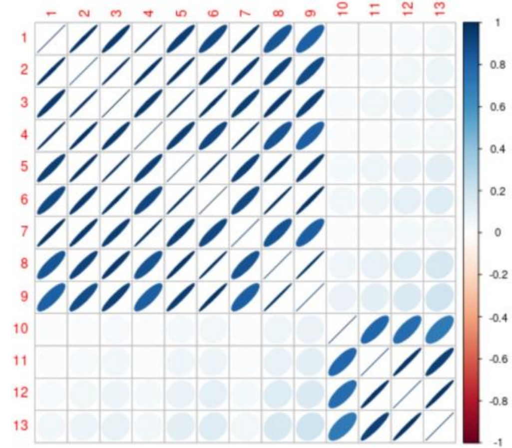

**2 variables**
- **One continuous variable**
- **One categorical variable: 13 levels**

⇒ **14 dimensions**

**DoE : LHS (n = 98)**




> 0

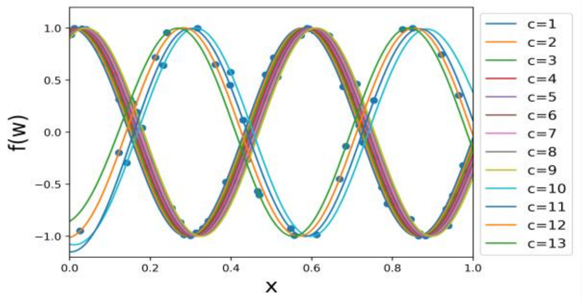


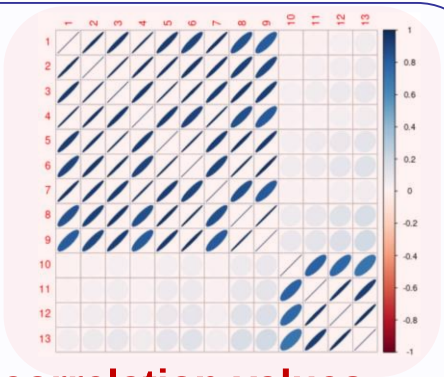
**Accurate model but expensive. Only positive correlations can be handled.**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Our Exponential homoscedastic hypersphere kernel**



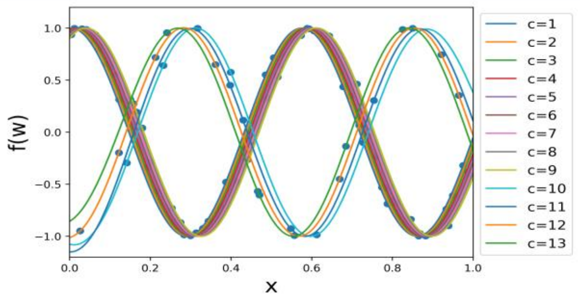
$\in [0, 1]$

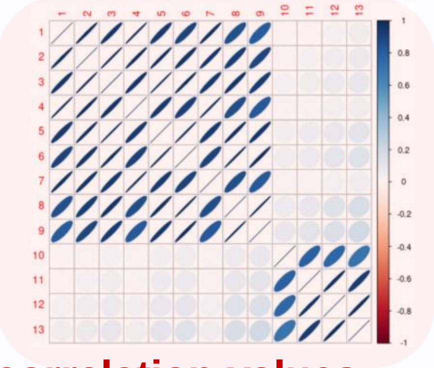
**Efficient method with positive correlation values**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

**Our Exponential homoscedastic hypersphere kernel**

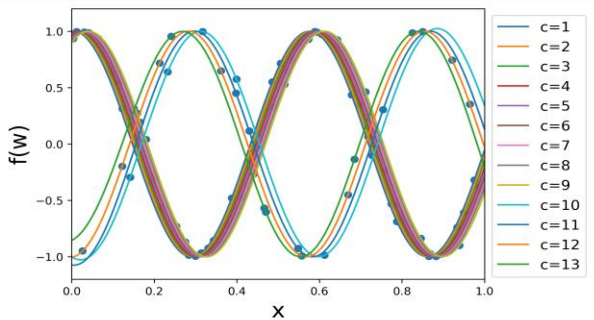


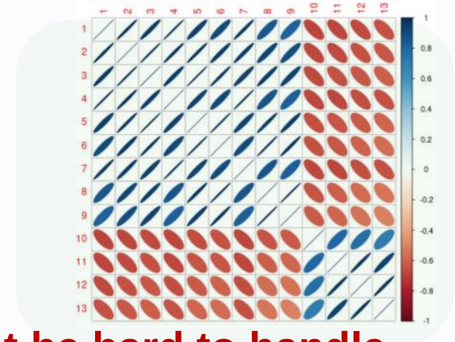



$\in [0,1]$

**Efficient method with positive correlation values**

**Homoscedastic hypersphere kernel** *(Pelamatti et al., 2020)*



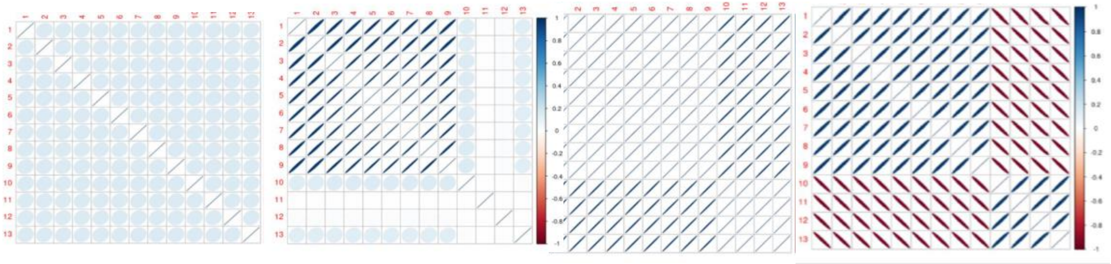
$\in [-1,1]$ **but hard to optimize**

**Include general correlation values but might be hard to handle**

Roustant, O., E. Padonou, Y. Deville, A. Clément, G. Perrin, J. Giorla, and H. Wynn (2020). Group kernels for Gaussian process metamodels with categorical inputs. SIAM J.Uncertain. 8, 775–806.

# Results

- *An illustration case*



2 design variables
- 1 continuous variable
- 1 categorical variable: 13 levels
- 14 dimensions

- DoE: LHS with 98 points (7 per dimension)

Validation set of 30603 points

$$RMSE = \left( \frac{1}{30603} \sum_{i=1}^{30603} \left( \hat{f}(w_i) - f(w_i) \right)^2 \right)^{1/2}$$

$$PVA = \log \left( \frac{1}{30603} \sum_{i=1}^{30603} \frac{\left( \hat{f}(w_i) - f(w_i) \right)^2}{\sigma^f(w_i)^2} \right)$$

Table 4: Kernel comparison for the cosine test case

| Kernel | # of Hyperparam. | RMSE | PVA | CPU time (s) |
|---|---|---|---|---|
| GD | 2 | 30.079 | 21.99 | 1.4 |
| CR | 14 | 22.347 | 23.04 | 24.5 |
| EHH | 79 | 1.882 | 23.74 | 514.5 |
| HH | 79 | 1.280 | 24.31 | 514.5 |

C.Demay, Iooss, B., Gratiet, L. L., and Marrel, A., "Model selection based on validation criteria for Gaussian process regression: An application with highlights on the predictive variance," Quality and Reliability Engineering International, Vol. 38, 2022, pp. 1482–1500.

# Exemple of application: Bayesian optimization

SMT

Table of Contents



Previous topic

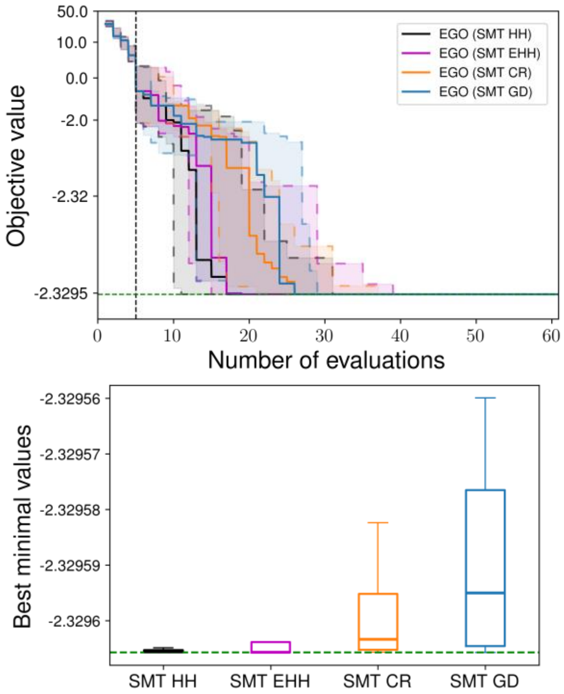

Mixed Integer and Hierarchical usage (Variables, Sampling and Context)


## Branin

```
#Import the Mixed Integer API
from smt.surrogate_models import (KRG,XSpecs,XType,
  XRole,MixIntKernelType)
from smt.applications.mixed_integer import(
MixedIntegerSamplingMethod as misamp,
MixedIntegerContext)
#Define the function
from smt.problems import Branin
fun = Branin(ndim=2)
#Define the mixed variables
xtypes = [XType.ORD,XType.FLOAT]
xlimits = fun.xlimits
xspecs = XSpecs(xtypes=xtypes, xlimits=xlimits)
#Perform a mixed integer sampling with LHS
from smt.sampling_methods import LHS
smp = misamp(LHS,xspecs,random_state=42)
xdoe = smp(10)
# Call the Bayesian optimizer
from smt.applications import EGO
criterion = "EI"  #'EI' or 'SBO' or 'LCB'
ego = EGO(xdoe=xdoe,
  n_iter=20,
  criterion="EI",
  random_state=42,
  surrogate=KRG(xspecs=xspecs,
   categorical_kernel=MixIntKernelType.GOWER))
x_opt, y_opt, _, _, _ = ego.optimize(fun=fun)
# Check if the result is correct
self.assertAlmostEqual(0.494, float(y_opt), delta=1)
```

Figure 7: Example of usage of mixed integer surrogates for Bayesian optimization

## Toy problem





M. M. Zuniga and D. Sinoquet. Global optimization for mixed categorical-continuous variables based on gaussian process models with a randomized categorical space exploration step. INFOR: Information Systems and Operational Research, 58:310–341, 2020

# Conclusion & perspectives

# Conclusion & perspectives

- *Contributions*

- Extended continuous Gaussian kernels to handle mixed-categorical variables.
- Proposed a unified framework for continuous relaxation and Gower distance based kernels.
- Validated the efficiency of the proposed method on a challenging mixed-categorical problem.

# Conclusion & perspectives

- *Contributions*

- Extended continuous Gaussian kernels to handle mixed-categorical variables.
- Proposed a unified framework for continuous relaxation and Gower distance based kernels.
- Validated the efficiency of the proposed method on a challenging mixed-categorical problem.

- *Current works*

Apply dimension reduction (*e.g.*, PLS) using the proposed Gaussian kernel to model large-scale mixed-categorical problems. “Mixed categorical Gaussian process for high-dimensional Bayesian optimization” in Structural and Multidisciplinary Optimization.

Couple mixed-discrete variables with hierarchical variables. SMT 2.0: A Surrogate Modeling Toolbox with a focus on Hierarchical and Mixed Variables Gaussian Processes” in Advances in Engineering Software.

# The End

## Thank you for your attention !

P. Saves, Y. Diouane, N. Bartoli, T. Lefebvre, and J. Morlier. A mixed-categorical correlation kernel for gaussian process, 2022. Submitted to Neurocomputing. Available at SSRN: https://dx.doi.org/10.2139/ssrn.4296794