EVALUATING THE QUALITY OF STEADY-STATE MULTIVARIATE EXPERIMENTAL DATA IN VARIOUS ORC EXPERIMENTAL SETUPS

S. Quoilin, O. Dumont, R. Dickes, V. Lemort

Thermodynamics Laboratory, University of Liège

September 15th 2017

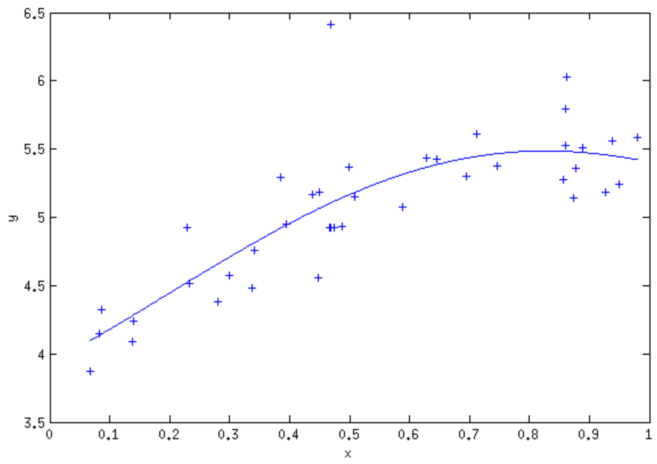
ORC 2017 Conference

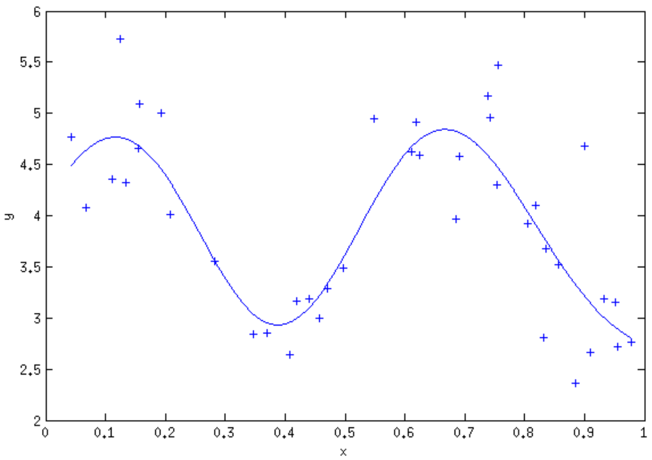
Introduction

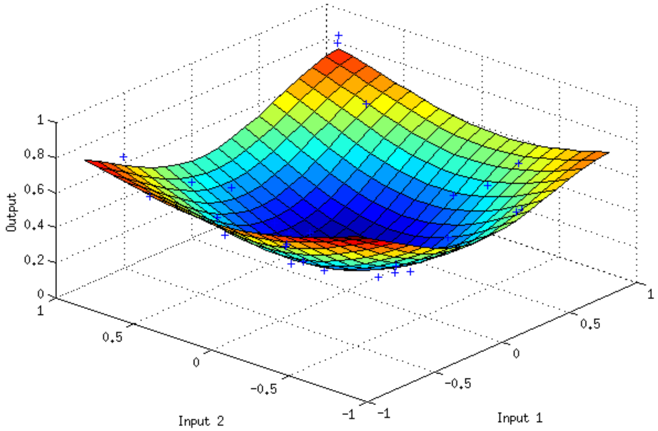
Quality of experimental data

- ✓ High quality data is required for the development, the calibration and the validation of models
- ✓ Different types of models:
  - ✓ Deterministic models
  - ✓ Semi-empirical models
  - ✓ Empirical models
- ✓ Experimental data is subject to many measurements errors, test bench malfunctions, operator misuse or misinterpretation, etc.
- ✓ Special focus on **multidimensional inputs**
- ✓ Goal of this work: provide an **open-source tool** to assess the quality of experimental data and its « explainability »
- ✓ Several key questions to be answered

Question 1:
"Most likely" shape of the function explaining the data?

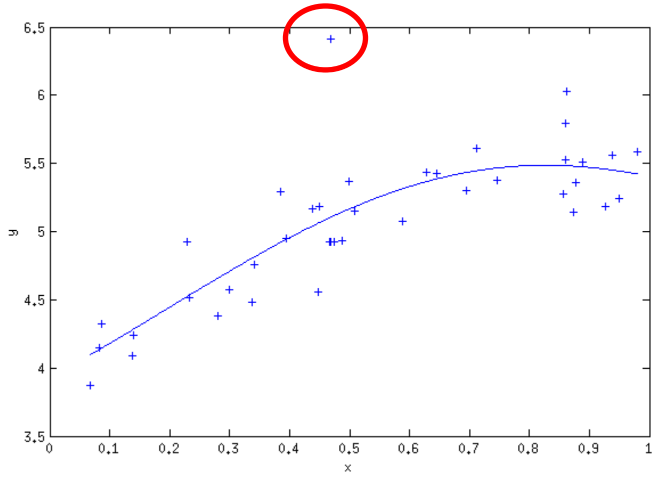


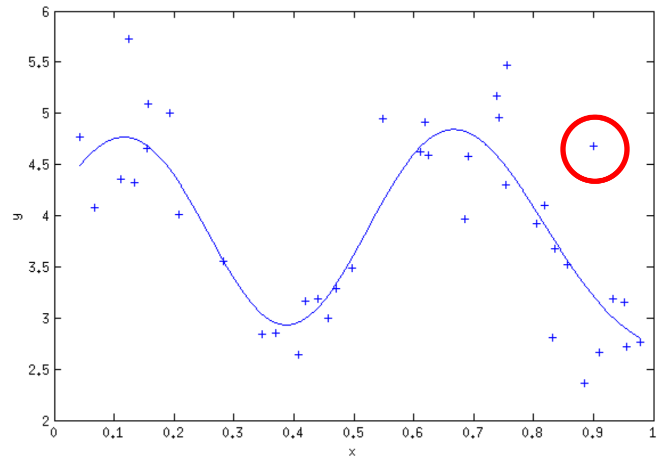


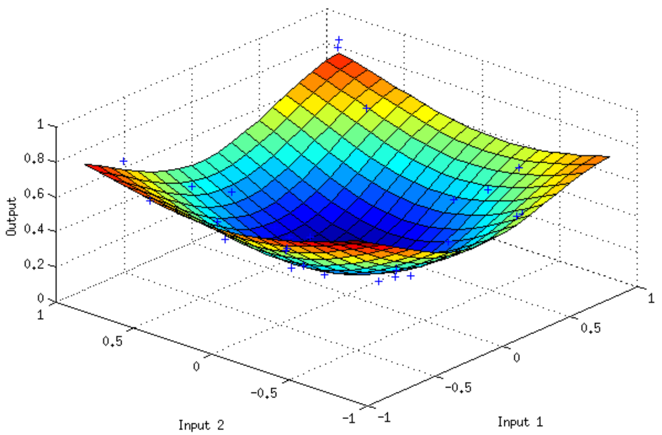


Question 2:

Repeatability and detection of outliers







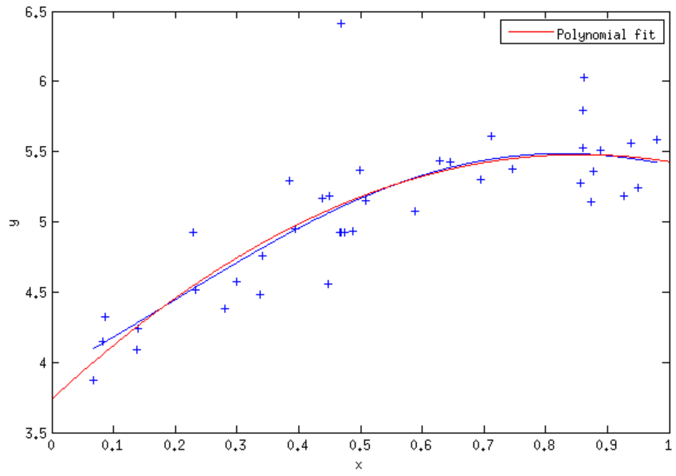
Additional questions:

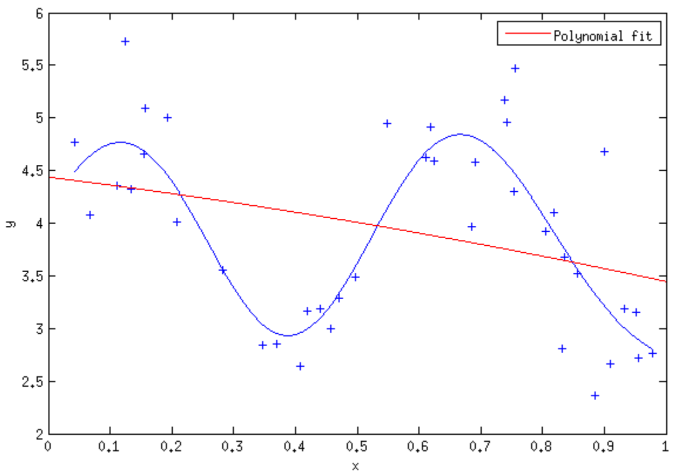
- *Best explanatory variables?*
  - Goal: explain the data with the smallest possible set of input variables
- *Data accuracy / noise level?*
  - Goal: determine experimentally what it the data accuracy => what would be the best accuracy a model could reach with this data
  - Necessity to de-noise the data

Data analysis:

Use of Gaussian Process regressions / Kriging interpolation

Gaussian process regression





- Traditional regression (e.g. based on least-squares):
  - Use of parametric functions
  - Function is defined *a-priori*
- Gaussian process regression:
  - Probabilistic distribution of the function with respect to the data points
  - Instead of the definition of a parametric function, definition of a **Covariance Matrix**
- Bayesian formulation:

$$p(f|y) = \frac{p(y|f)p(f)}{p(y)}$$

- Covariance functions (kernel):

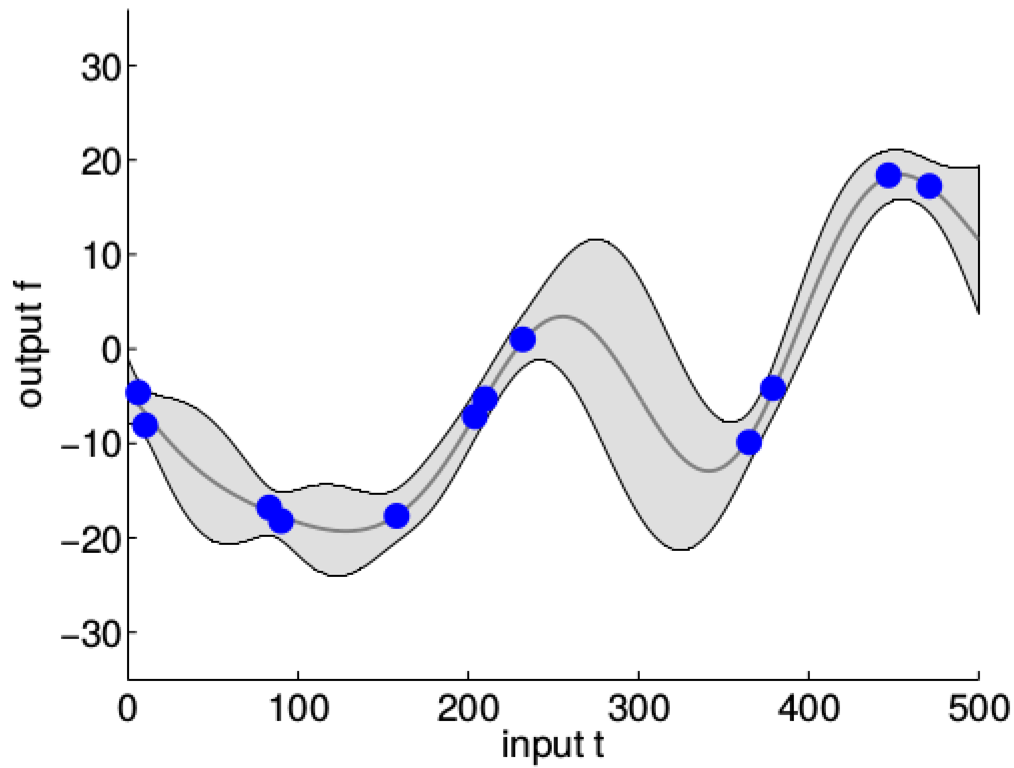
$$K(x, x') = \sigma_0{}^2 \exp\left[-\frac{1}{2}\left(\frac{x - x'}{\lambda}\right)^2\right]$$

LABORATOIRE DE THERMODYNAMIQUE UNIVERSITE DE LIEGE

Gaussian process

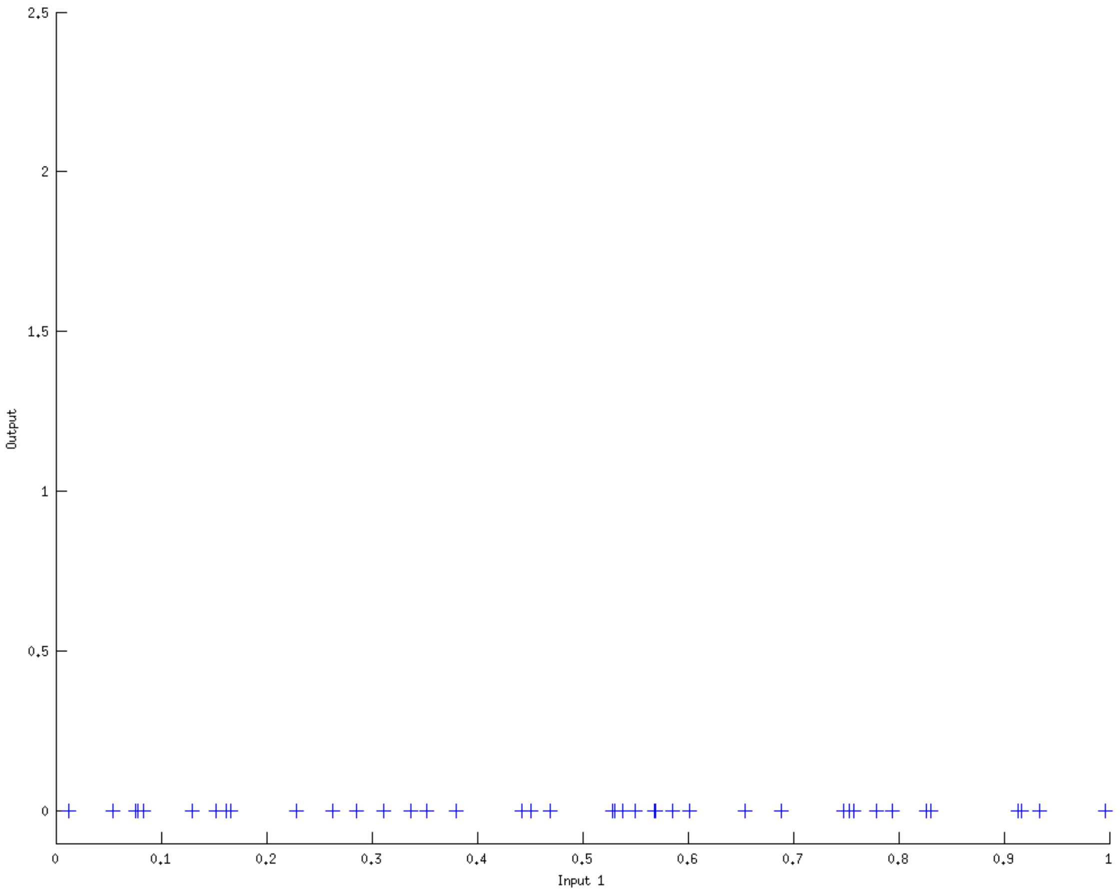
1. Definition of a GP prior
2. Use Baysian inference to update the probability distribution

$$p(f|y) = \frac{p(y|f)p(f)}{p(y)}$$

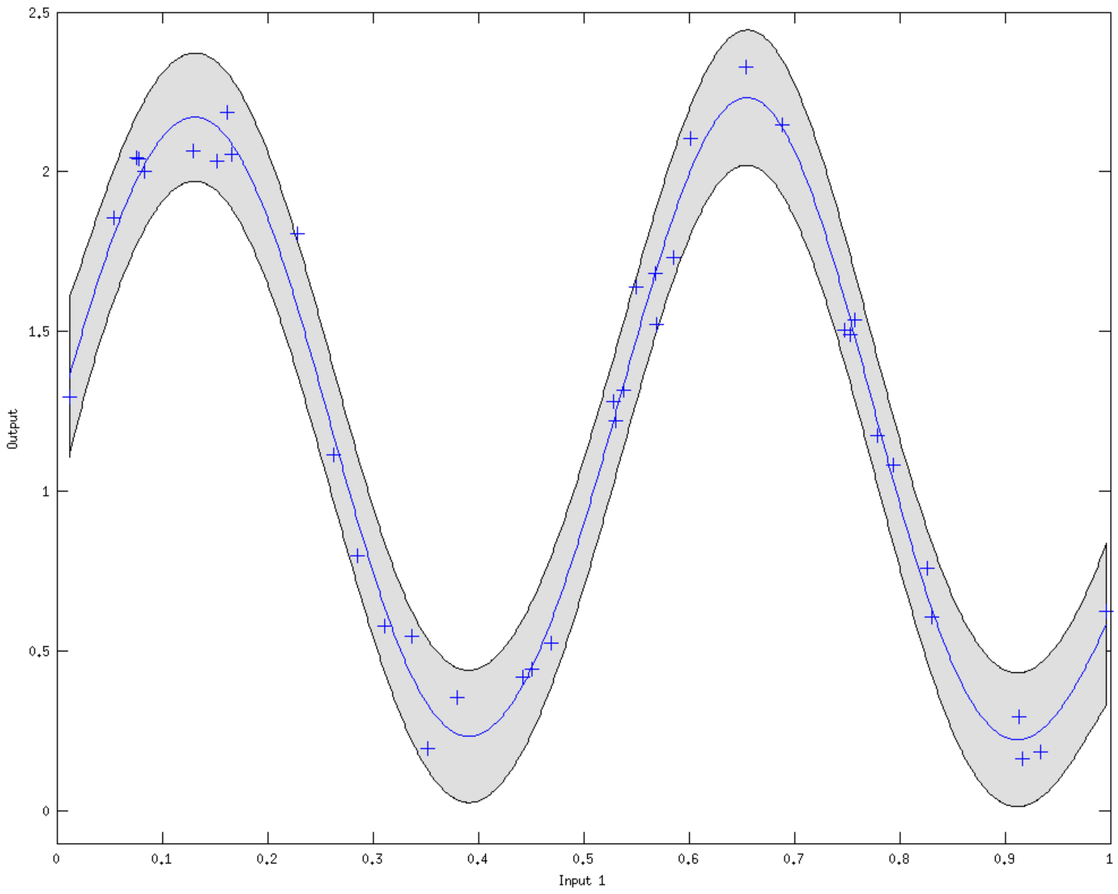


Source: John P. Cunningham, Gaussian Processes for Machine Learning, MLSS 2012

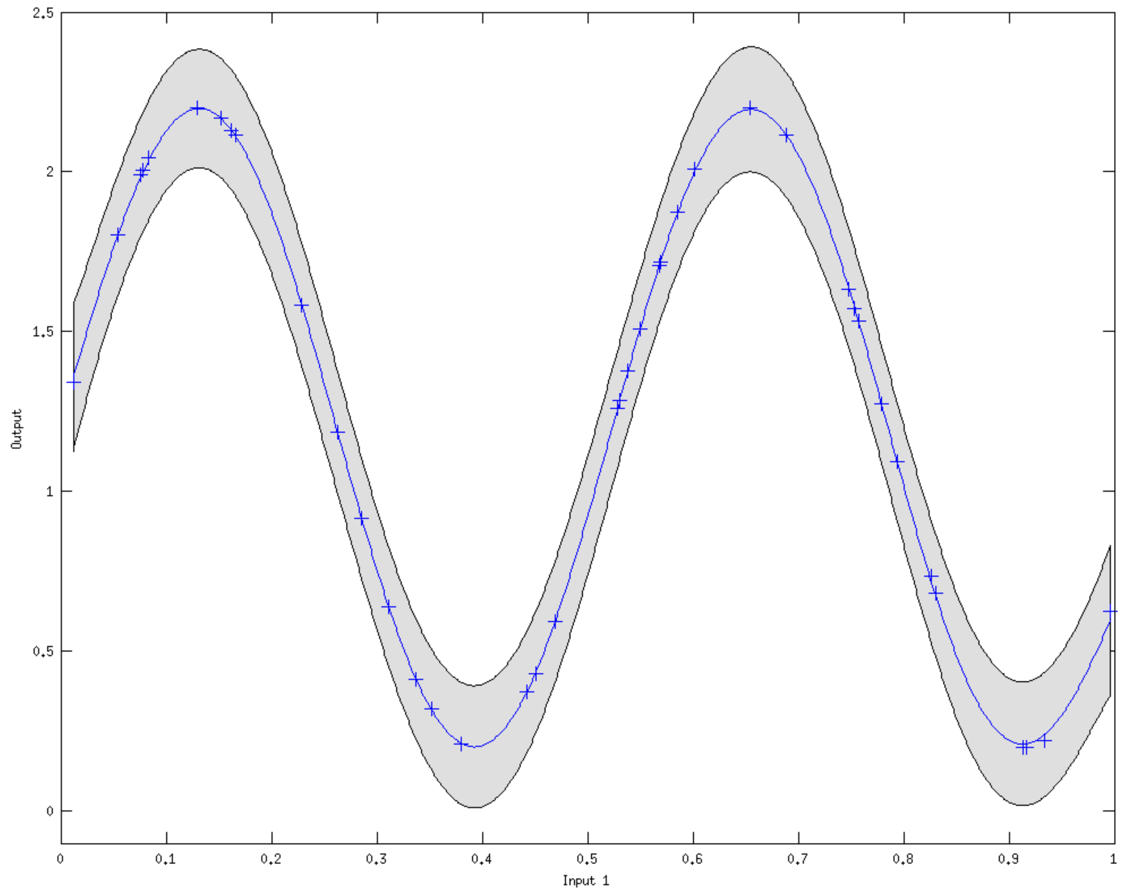
Gaussian process:
Regression



Gaussian process:
Effect of outliers



Gaussian process:
Effect of noise



Selection of the length scales

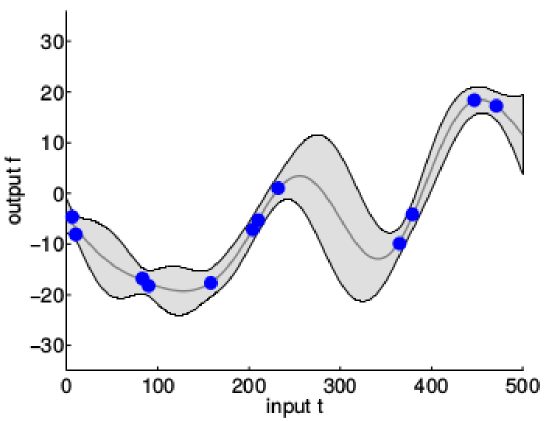
Preventing overfitting

Unidimensional Gaussian Process regression:

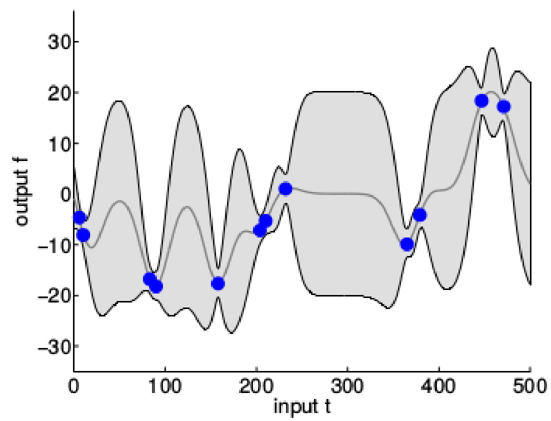
Use of a SE kernel:

$$f \sim \mathcal{GP}(0, k_{ff}), \text{ where } k_{ff}(t_i, t_j) = \sigma_f^2 \exp\left\{-\frac{1}{2\ell^2}(t_i - t_j)^2\right\}$$

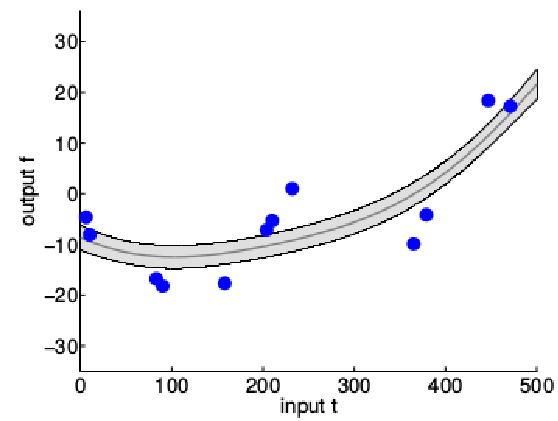
l=50: just right



l=15: Overfitting



l=150: Underfitting



Source: John P. Cunningham, Gaussian Processes for Machine Learning, MLSS 2012

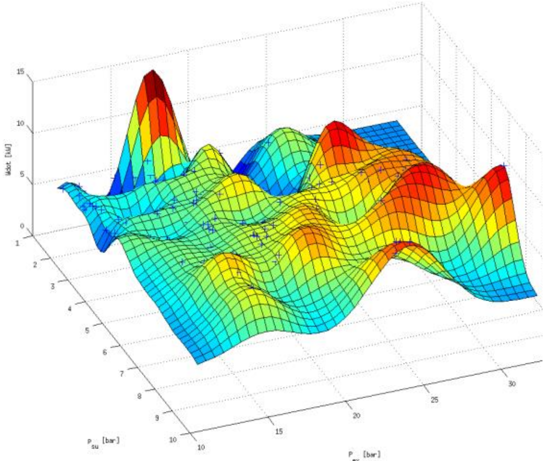
Selection of the length scales

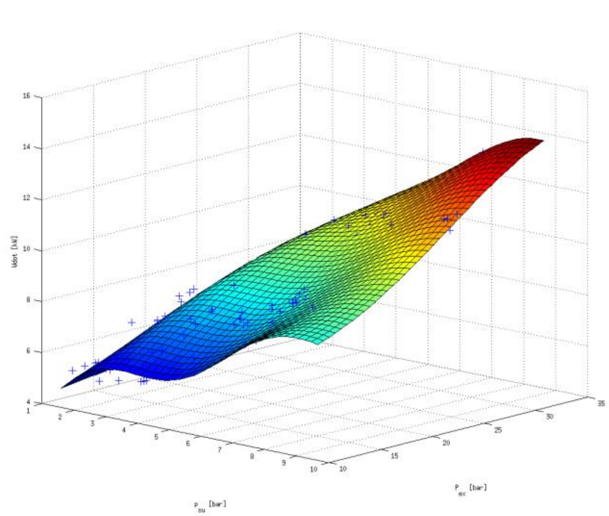
Preventing overfitting

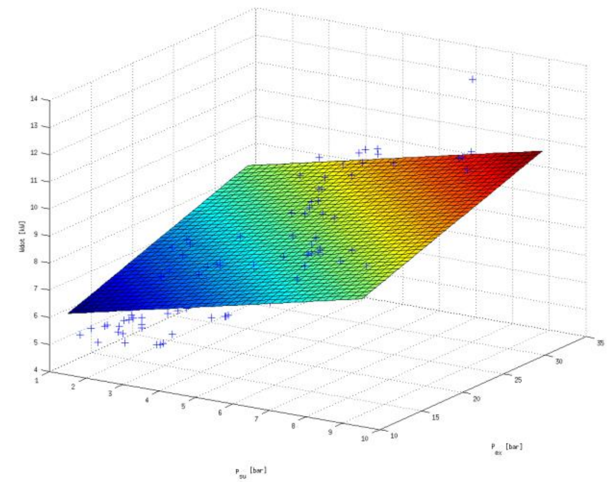
Multidimensional Gaussian Process regression:

Use of an SE ARD Kernel:

$$k_{ff}(x^{(i)}, x^{(j)}) = \sigma_f^2 \exp\left\{ -\sum_d \frac{1}{2\ell_d^2}(x_d^{(i)} - x_d^{(j)})^2 \right\}$$







➡ The hyperparameters l_d must be optimized to avoir under/over fitting

Selection of the length scales
Preventing overfitting

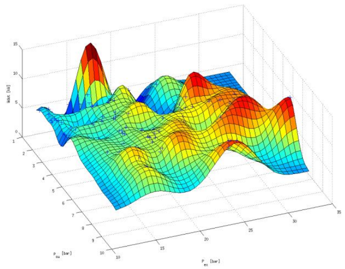
Three-step process:

1. Optimize the parameters (l1,l2,l3,…) to maximize the marginal likelihood

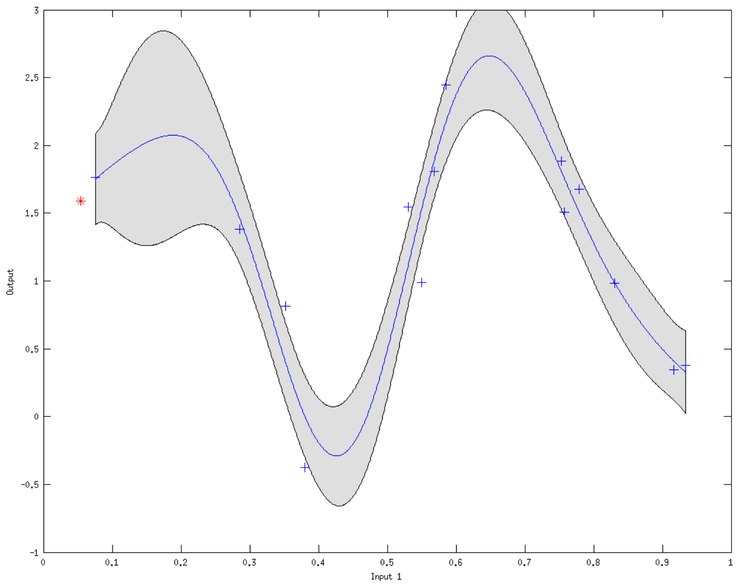
$$p(y) = \int p(y|f)p(f)df$$

2. Computation of the mean average error in Cross-Validation / Training

3. Visual verification:



Cross-validation, good fit:



Mean relative error:

MRE with all points:	2.7%
MRE in cross-validation:	7.6%

Selection of the length scales

Preventing overfitting

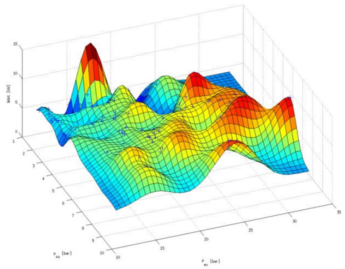
Three-step process:

1. Optimize the parameters (l1,l2,l3,…) to maximize the marginal likelihood

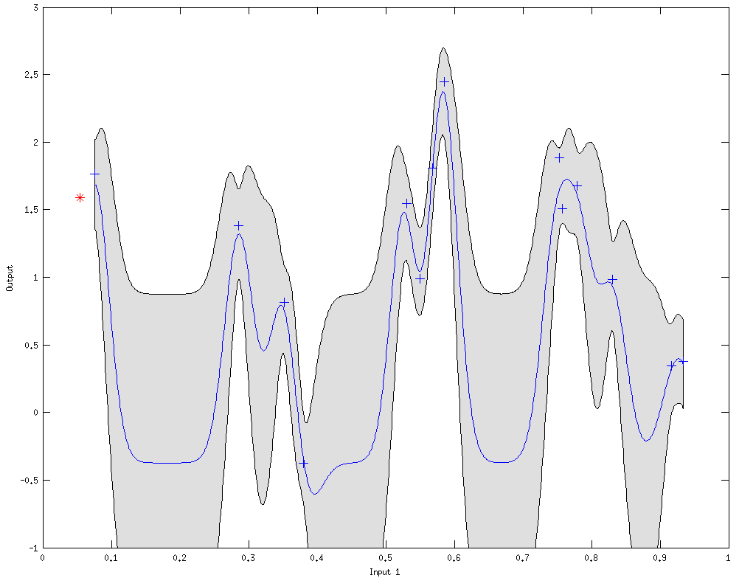
$$p(y) = \int p(y|f)p(f)df$$

2. Computation of the mean absolute error in Cross-Validation / Training

3. Visual verification:



Cross-validation, overfit:



Mean relative error:

MRE with all points:	1.1%
MRE in cross-validation:	18.5%

Examples of experimental test rig

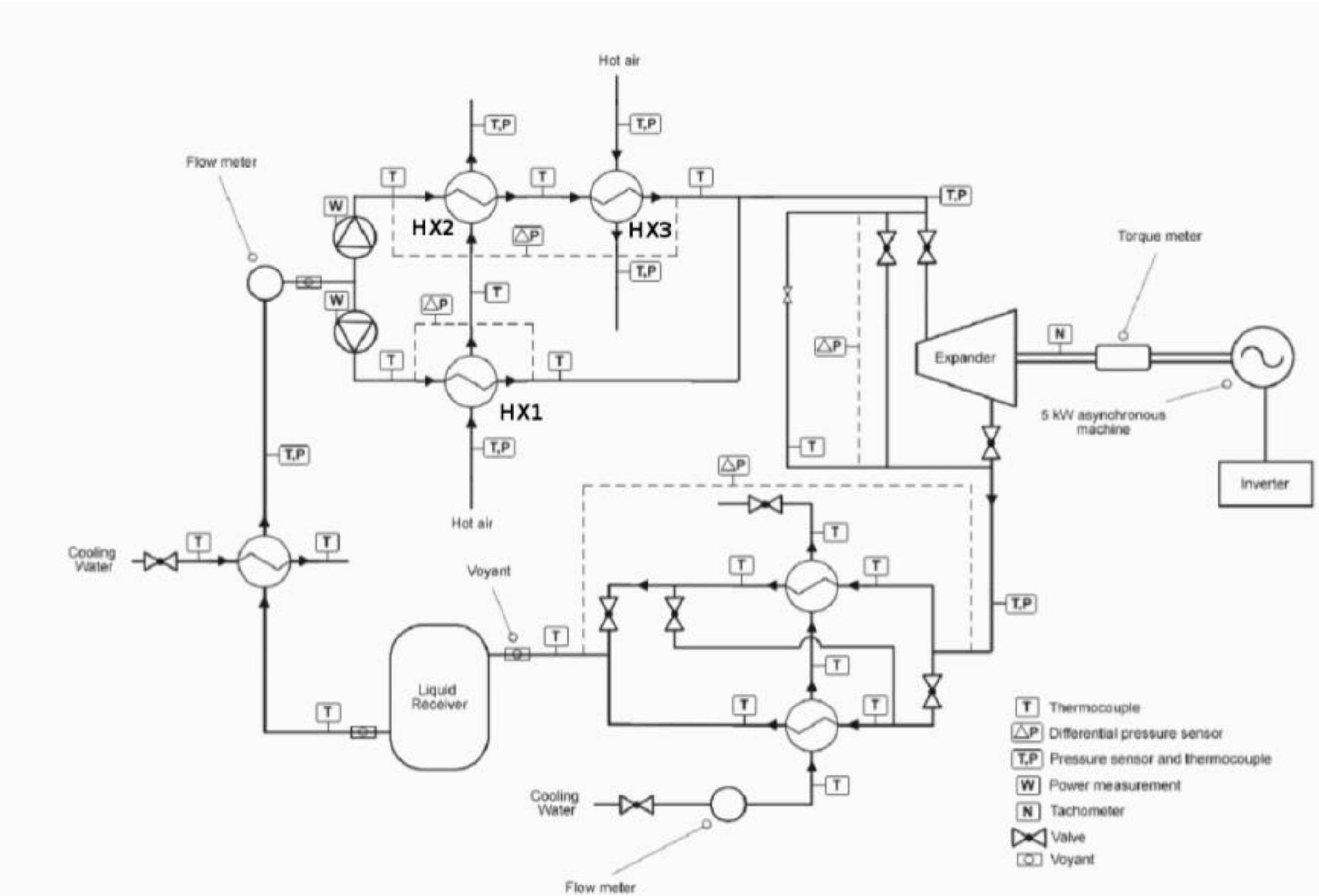
Experimental setups

Open-drive scroll compressor

- Converted into an expander
- Built in volumetric ratio : 3.94
- Absence of lubrication
- Not tight

Experimental setups

Open-drive scroll compressor

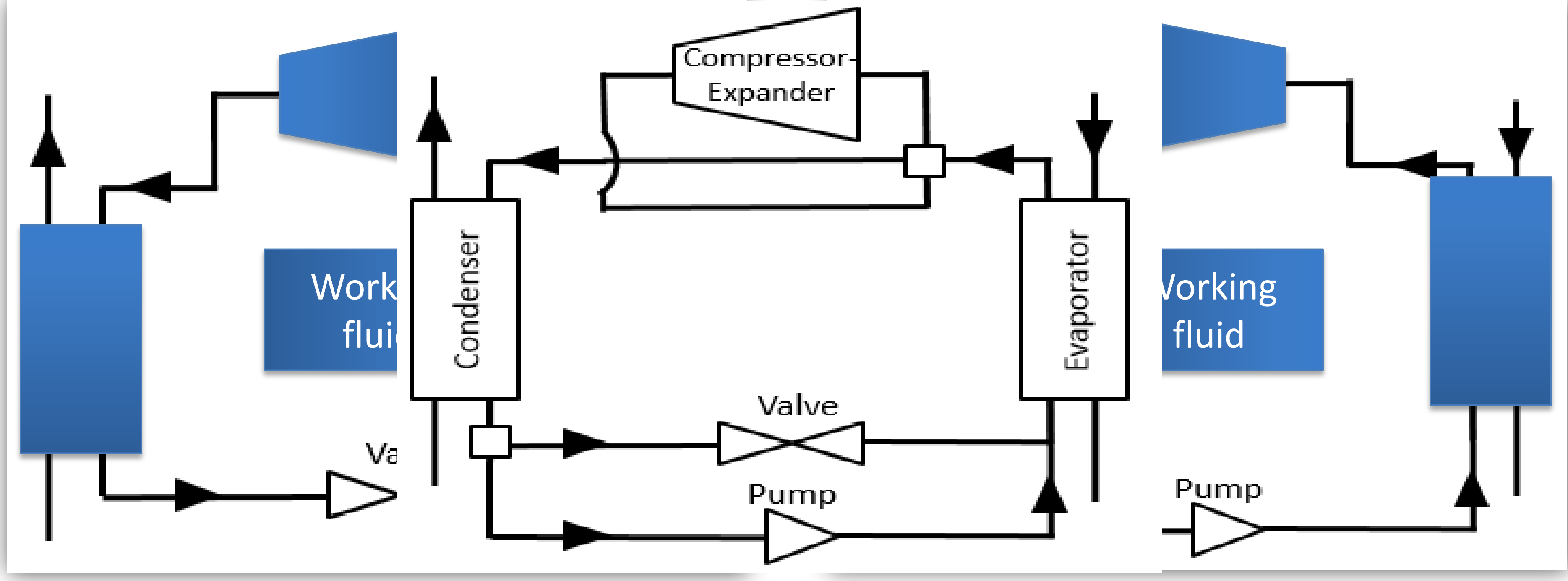


$$\varepsilon_s = \frac{\dot{W}_{sh}}{\dot{M}_r \cdot (h_{su,\exp} - h_{ex,\exp,s})}$$

Experimental setups

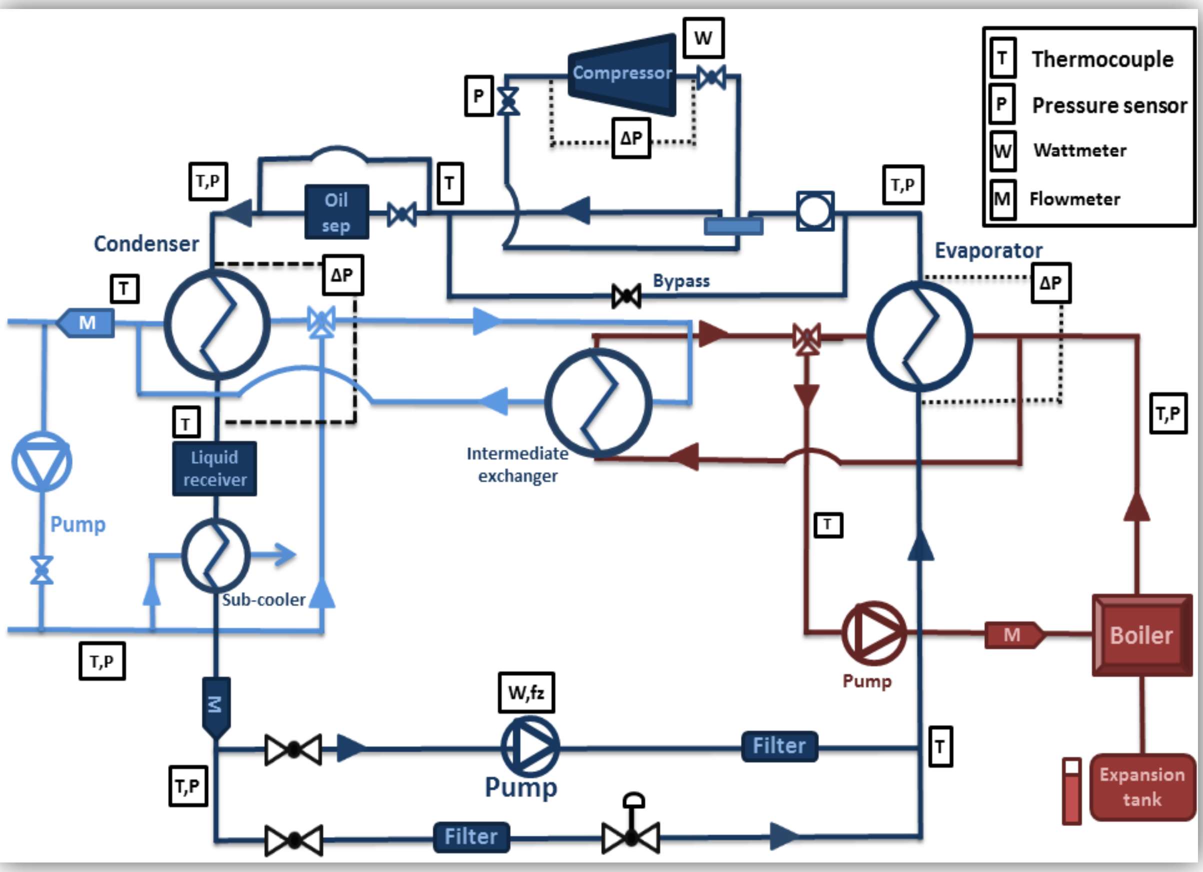
Reversible HP/ORC unit

Heat Pump → ORC → Reversible HP/ORC



Experimental investigations

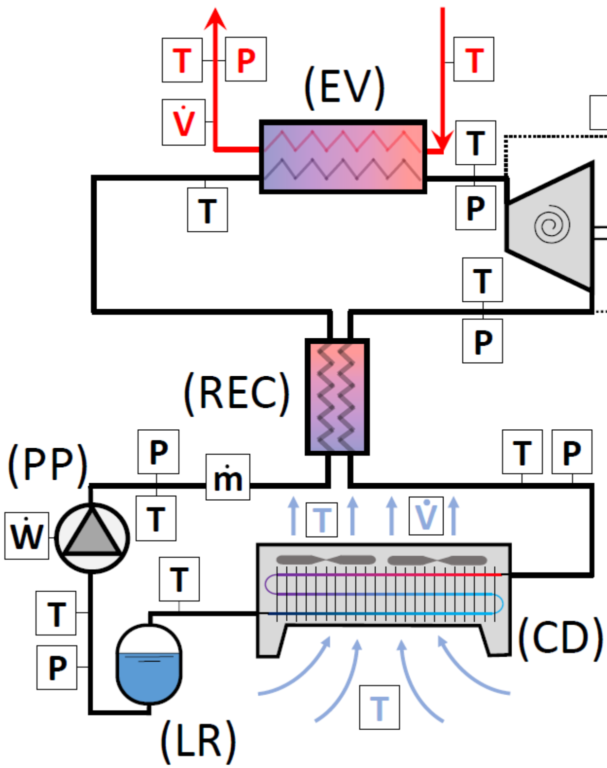
Reversible HP/ORC unit



$$\eta_{ORC} = \frac{W_{exp,el} - W_{pump,el}}{\dot{Q}_{ev}}$$

Experimental investigations

Sun2Power unit



P : pressure sensor - **T** : thermocouple - **$\dot{m}$** : mass flow meter
$\dot{W}$: power meter - **$\dot{V}$** : volumetric flow meter

Sun2Power unit:

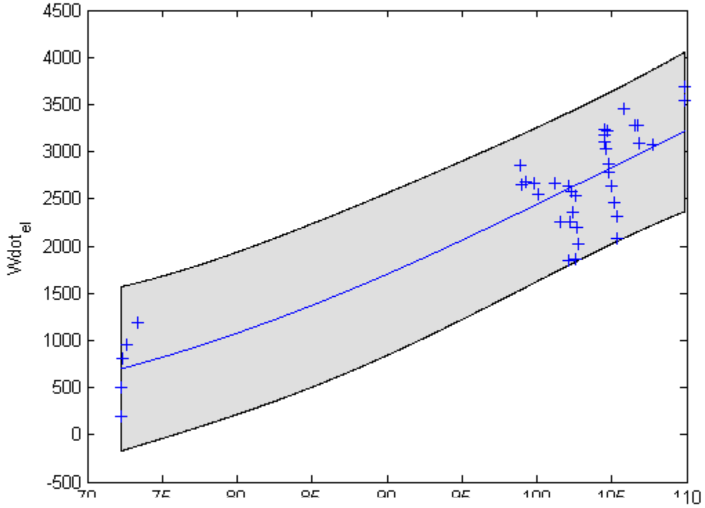
- 2kWe recuperative ORC
- R245fa as working fluid
- Scroll expander + diaphragm pump
- <u>Two BPHEXs (EV + REC)</u>
- <u>One fin coil air-cooled condenser</u>

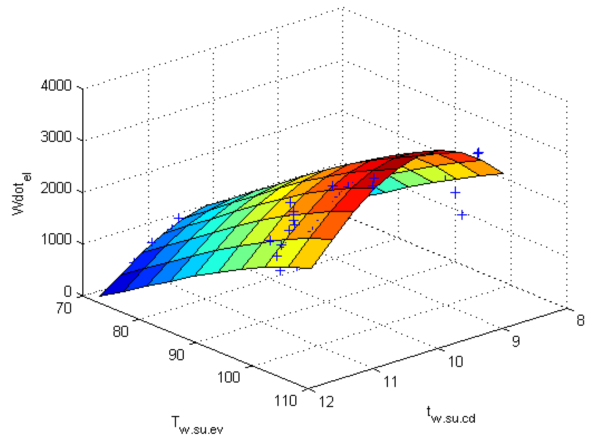
Reference database:

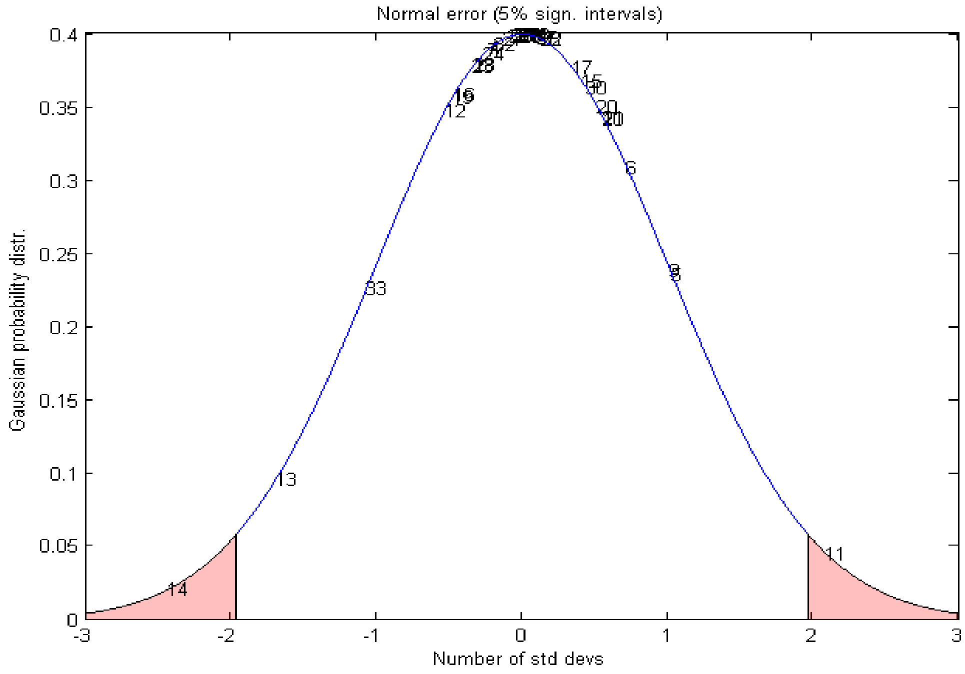
- Experimental measurements
- Complete range of conditions (40 pts)

Outlier detection

Example with the HP/ORC test rig

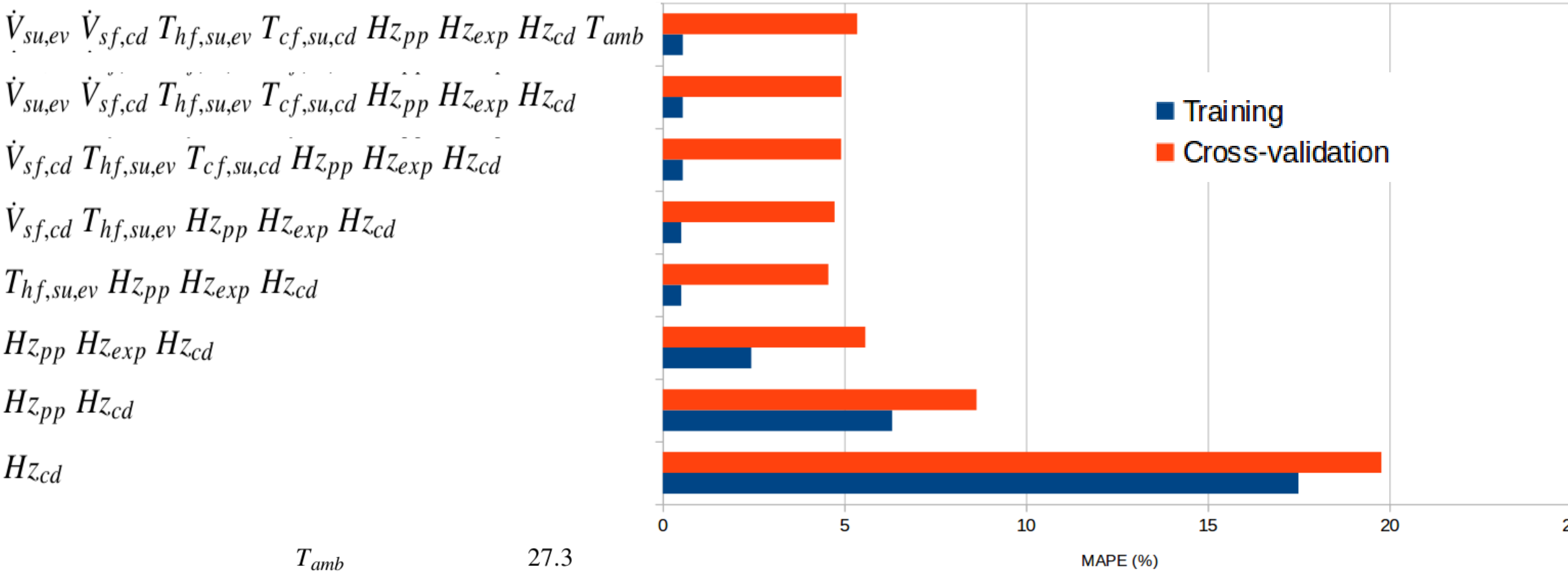






Feature selection

Example with the Sun2power test rig



Lengthscales:

T_{amb}	27.3
$\dot{V}_{su,ev}$	528
$T_{cf,su,cd}$	27.3
$\dot{V}_{sf,cd}$	11.5
$T_{hf,su,ev}$	1.95
Hz_{exp}	1.15
Hz_{pp}	1.31
Hz_{cd}	0.32

Relevant input variable

Summary

Table 2. Inputs variables of the three considered processes

HP/ORC		Sun2Power		Expander	
Heat source flow rate:	$\dot{M}_{su,ev}[kg/s]$	Heat source flow rate:	$\dot{V}_{su,ev}[kg/s]$	Inlet pressure:	$P_{su}[Pa]$
Heat sink flow rate:	$\dot{M}_{sf,cd}[kg/s]$	Heat sink flow rate:	$\dot{V}_{sf,cd}[kg/s]$	Outlet pressure:	$P_{ex}[Pa]$
Heat source temperature	$T_{hf,su,ev}[K]$	Heat source temperature	$T_{hf,su,ev}[K]$	Rotating speed:	$N_{rot}[rpm]$
Heat sink temperature	$T_{cf,su,cd}[K]$	Heat sink temperature	$T_{cf,su,cd}[K]$	Inlet temperature:	$T_{su}[K]$
Pump speed:	$N_{pp}[rpm]$	Expander Rotating speed:	$Hz_{pp}[s^{-1}]$	Ambient temperature:	$T_{amb}[K]$
		Expander Rotating speed:	$Hz_{exp}[s^{-1}]$		
		Condenser fan speed:	$Hz_{cd}[s^{-1}]$		
		Ambient temperature:	$T_{amb}[K]$		

Predicted variable: Power output
2 outliers
MAPE, GP: 1,92%
MAPE, physical model: 2.45 %

Predicted variable: Power output
No outlier
MAPE, GP: 4.56%
MAPE, physical model: 8.12 %

Predicted variable: Power output
2 outliers
MAPE, GP: 0.99%
MAPE, physical model: 1.94 %

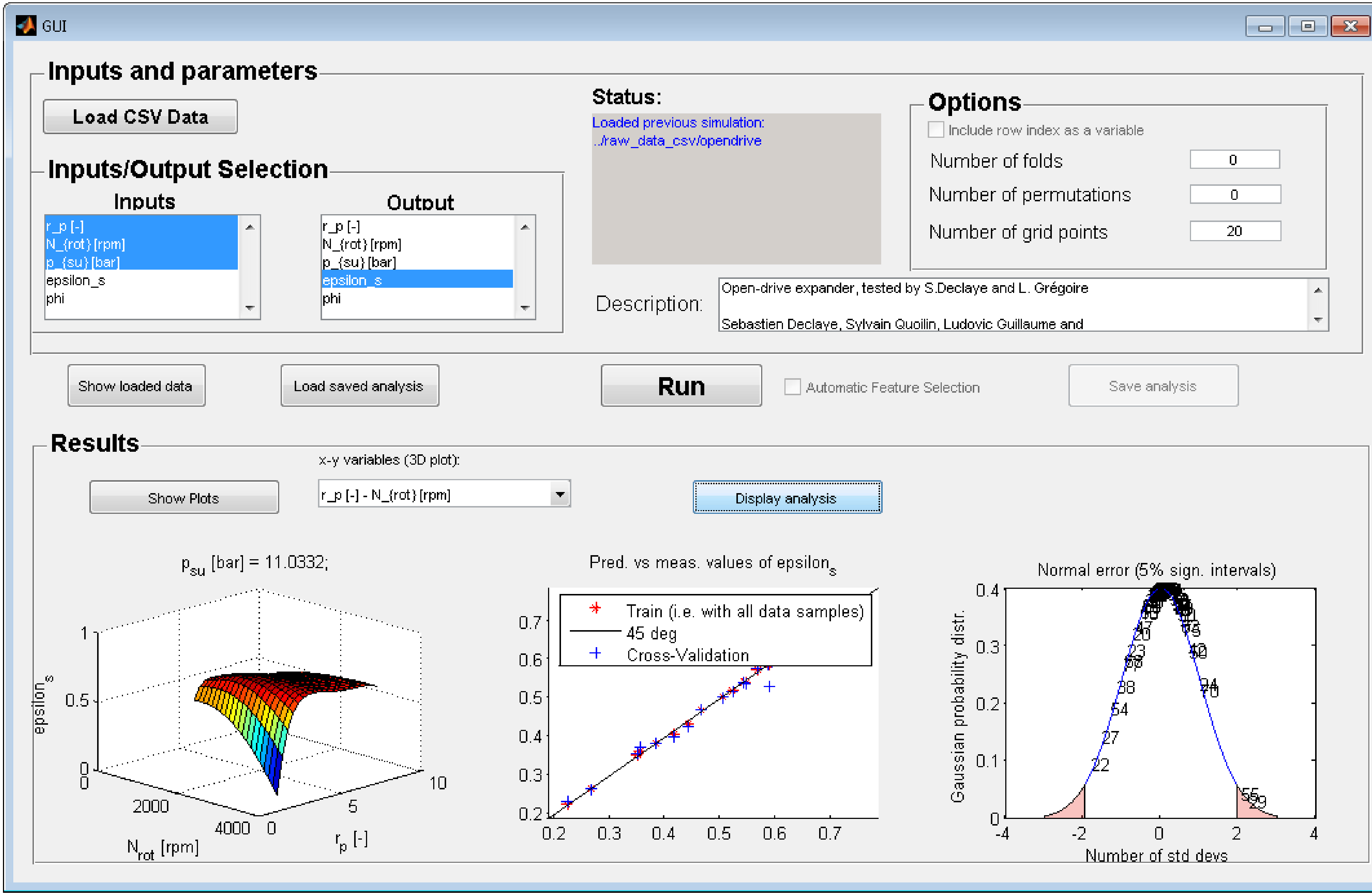
Conclusions

Starting from three ORC-related datasets, the proposed framework allowed us to:

- ✓ Perform non-linear, non-parametric regression
- ✓ Detect doubtful data points (outliers)
- ✓ Select the relevant input variables for the process to model (feature selection)
- ✓ Plot the effect of each relevant input variable by keeping the others constant (response surface)
- ✓ Evaluate the noise level in the data (i.e. the maximum model accuracy)
- ✓ Compare the "quality" of various datasets

The GPExp tool:

- ✓ Open-source
- ✓ Easy to download and run in Matlab
- ✓ Graphical user interface (GUI)
- ✓ External contributions and improvements are welcome



Download GPExp:

https://github.com/squoilin/GPExp

EVALUATING THE QUALITY OF STEADY-STATE MULTIVARIATE EXPERIMENTAL DATA IN VARIOUS ORC EXPERIMENTAL SETUPS

S. Quoilin, O. Dumont, R. Dickes, V. Lemort

Thermodynamics Laboratory, University of Liège

September 15th 2017

ORC 2017 Conference