Surrogate model

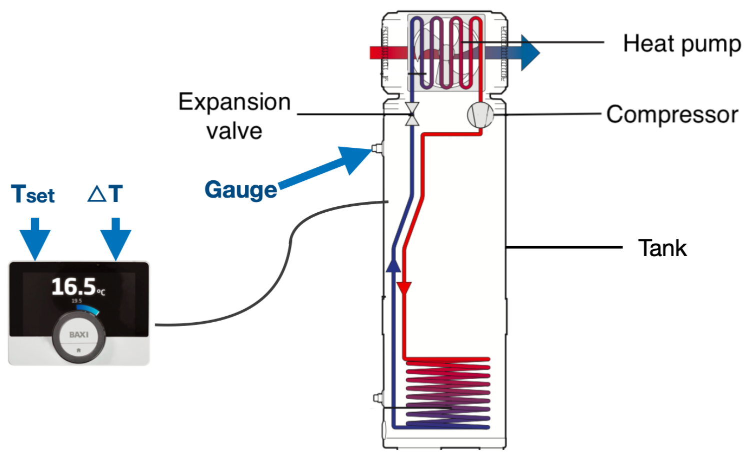
GAUTHIER Léonard
under the supervision of Aurélien Massein

August 21, 2022

0.1 Context

BDR Thermea are a company who develop innovative products and services that save energy and cut carbon emissions and supply them to building owners and users. Their connected and integrated technologies enable the global shift towards fully sustainable energy carriers.

The principle of a heat pump is to transfer heat from one medium to another by using the phase change properties of a refrigerant: condensation of a fluid requires a loss of heat, while evaporation of a fluid requires a supply of heat. Physically, the energy delivered is the sum of the energy pumped from the hot source and the energy consumed from the electric source; and of various energy losses. The coefficient of performance of a heat pump, or its efficiency, depends solely on the energy delivered over the energy consumed. It is necessary to establish a normative test setting out the temperature conditions and uses of the product.



During the development of their heat pump products, they have to improve several performances and to satisfy several constraints due to markets standards. These norm are: the European standard EN16147 and the LCIE 103-15 who establish a coefficient of performance COP_{DHW} and a rate $Star$ based on threshold.

The standard EN16147 test can last last from few days to several weeks depending upon its control parameters, blocking a climate chamber to certify its performances.

To develop and design their products, they resort to numerical simulations in order to adjust and to determine the optimal parameters according to the company aims. Their computation time and their design space are quite important such as they need to explore wisely possible design solutions(some minutes for each solution); therefore using surrogate models for performances of interest is a promising trail. The surrogate models will calcule solutions in the matter of few microseconds after his build.

They want to improve their product optimisation process, aiming to be fast, robust and versatile. Robustness is mandatory, as manufacturing and sensing processes are prone to uncertainties.

They need to propagate parameters uncertainties to their performances objective or constraint, in order to tackle and to determine performances confidence interval. In fine, optimal design solutions should be selected upon their worst performance cases.

This project is a continuation of several projects in M1[6] and M2[7] who has been done in Python, focusing on multi-objective optimisation and surrogate modelling, from a complete numerical models.

0.2 Surrogate Model

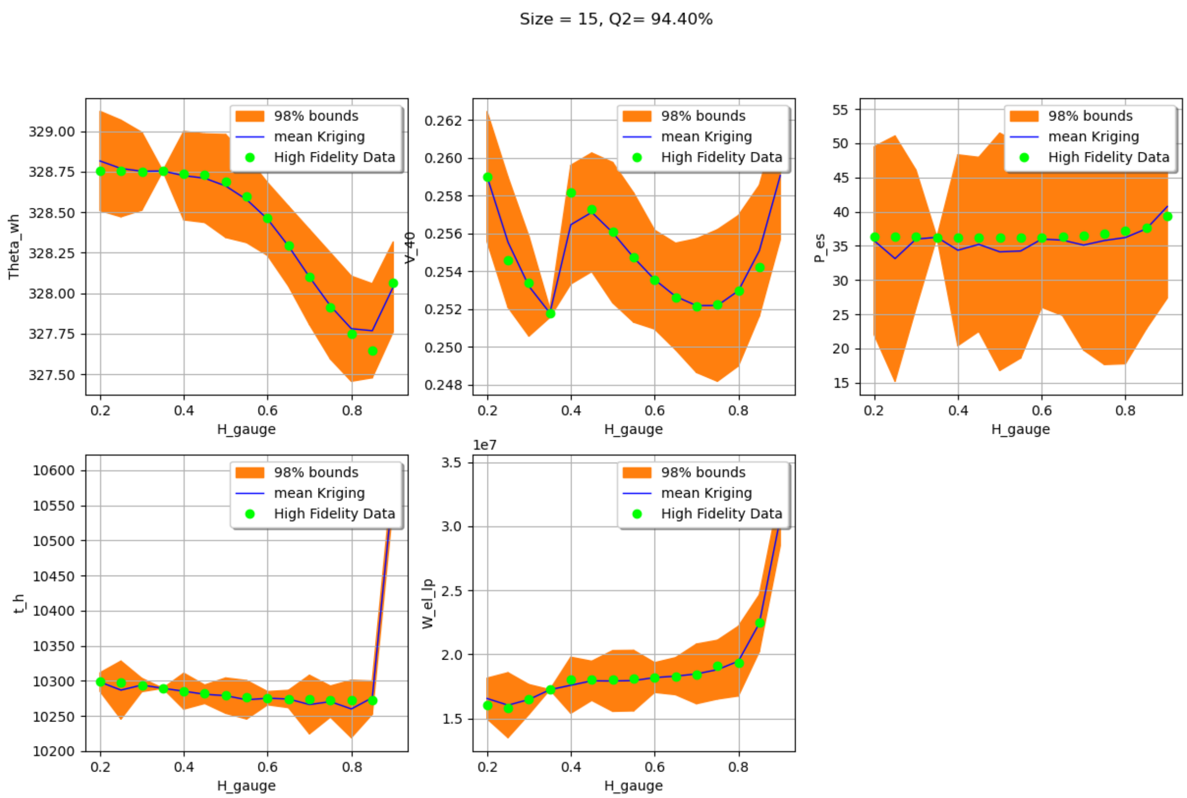
We will have 3 parameters we can manipulate (the input of our surrogate model):

- H_{gauge}: the height in m of the temperature sensor placed in the water heater which measures and regulates the temperature of the water in the tank
- T_{set}: the set-point temperature in $Kelvin$, i.e. the maximum temperature to be reached to stop the heating process
- $\Delta T_{hysteresis}$: the temperature in $Kelvin$ such that: $T_{min} = T_{set} - \Delta T_{hysteresis}$ are the minimum temperature to be reached to start the heating process

We choose to seperate the COP_{DHW} and $Star$ in their respective variables. It's justified by the unregularity of $Star$ as it's based on threshold.

The surrogate model that we want to create will be 5 functions: $R^3 \to R^1$
With these 5 values and other values which are constants defined according to the characteristics of the heat pump or according to the tests of the standard EN16147, we can calculate COP_{DHW} and $Star$

To create our surrogate model, we will use the Gaussian process regression (also known as Kriging), a technique of regression used in geostatistics, developed by G.MATHERON and based on D.G.KRIGE works.



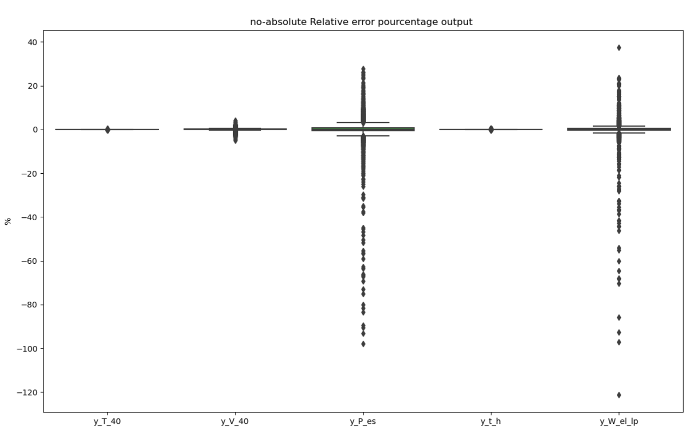
But it's not on the output of our models that we want with precision but on COP_{DHW} and $Star$. We will therefore also base our calculations on the percentage of absolute and relative error of our $Star$ because it is the hardest objective to have good results.

nb of training values	nb of error	absolute error	relativ error
100	245	3.96%	4.02%
200	134	2.16%	2.24%
300	137	2.21%	2.33%
400	114	1.84%	1.97%
500	110	1.78%	1.93%
750	78	1.26%	1.43%
1000	74	1.20%	1.43%

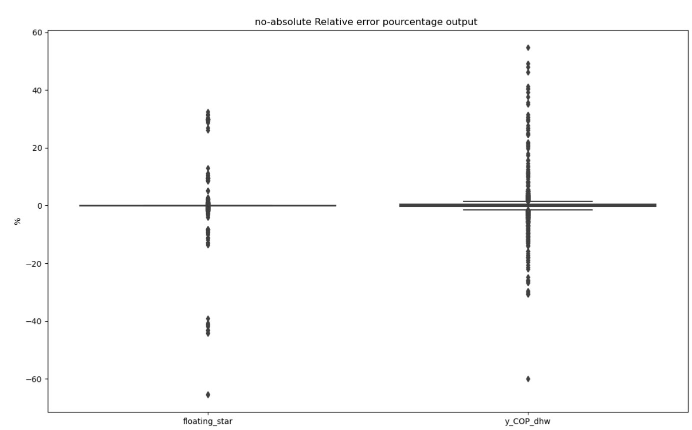
For the rest of this report, we will choose 1000 training values for our surrogate model
Mean Error relative (for 1000):

```
y_T_40       0.010192
y_V_40       0.268372
y_P_es       2.230229
y_t_h        0.017981
y_W_el_lp    1.248267

floating_star    0.345387
y_COP_dhw        1.180497
```



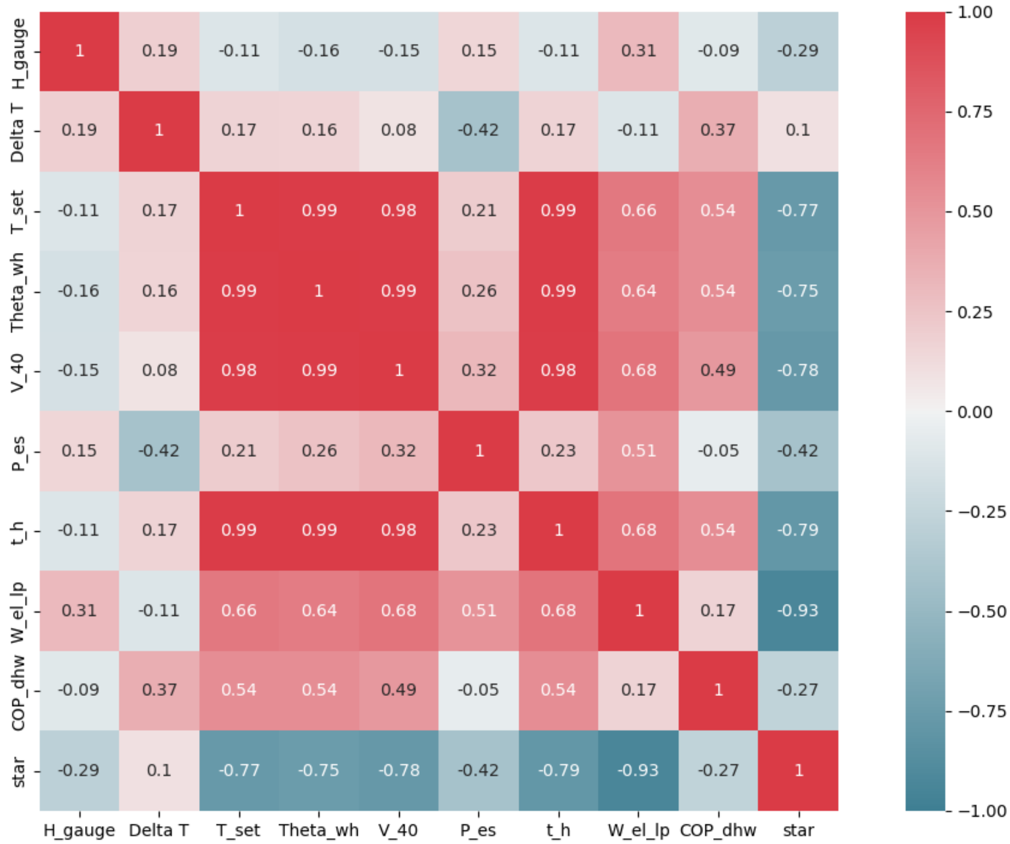
(a) error for $Star$



(b) error for COP_{DHW}

0.3 Analyse Surrogate Model

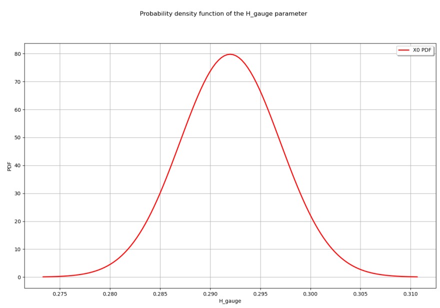
We can have a first idea of the correlation between the input and output parameters thanks to a comparison between the values of the CSV file and the output of the surrogate model. The closer they are, the more the 2 parameters are correlated.



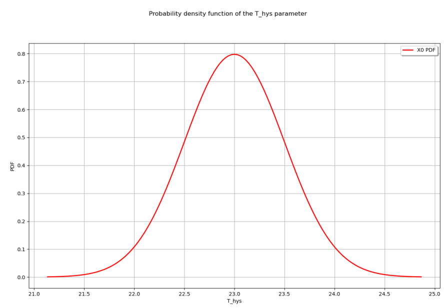
With the correlation matrix, we can see that T_{set} is strongly correlated to our outputs as well as for COP_{DHW} (with also T_{hys}) and that $Star$ is inversely correlated which is in line with reality: the less we set the temperature high, the better our energy consumption will be, the better our $Star$.

To analyze the sensitivity of our surrogate model, we will submit a point with an uncertainty distribution. For that we will take normal laws scaled to correspond to variations corresponding to reality: the probe can be misplaced by ± 2cm, and the temperatures set in an interval of ± 2°K. the point of reference will be $[0.292, 23, 327.15]$ with the "Normal" distribution:

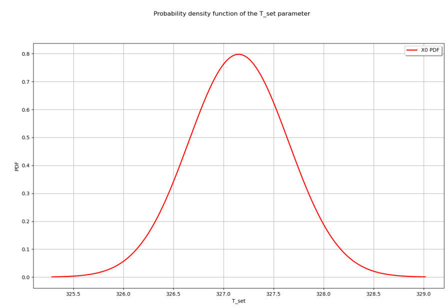
```
distribution = [Normal()/200,Normal()/2,Normal()/2]
```



(c) *Height* distribution)

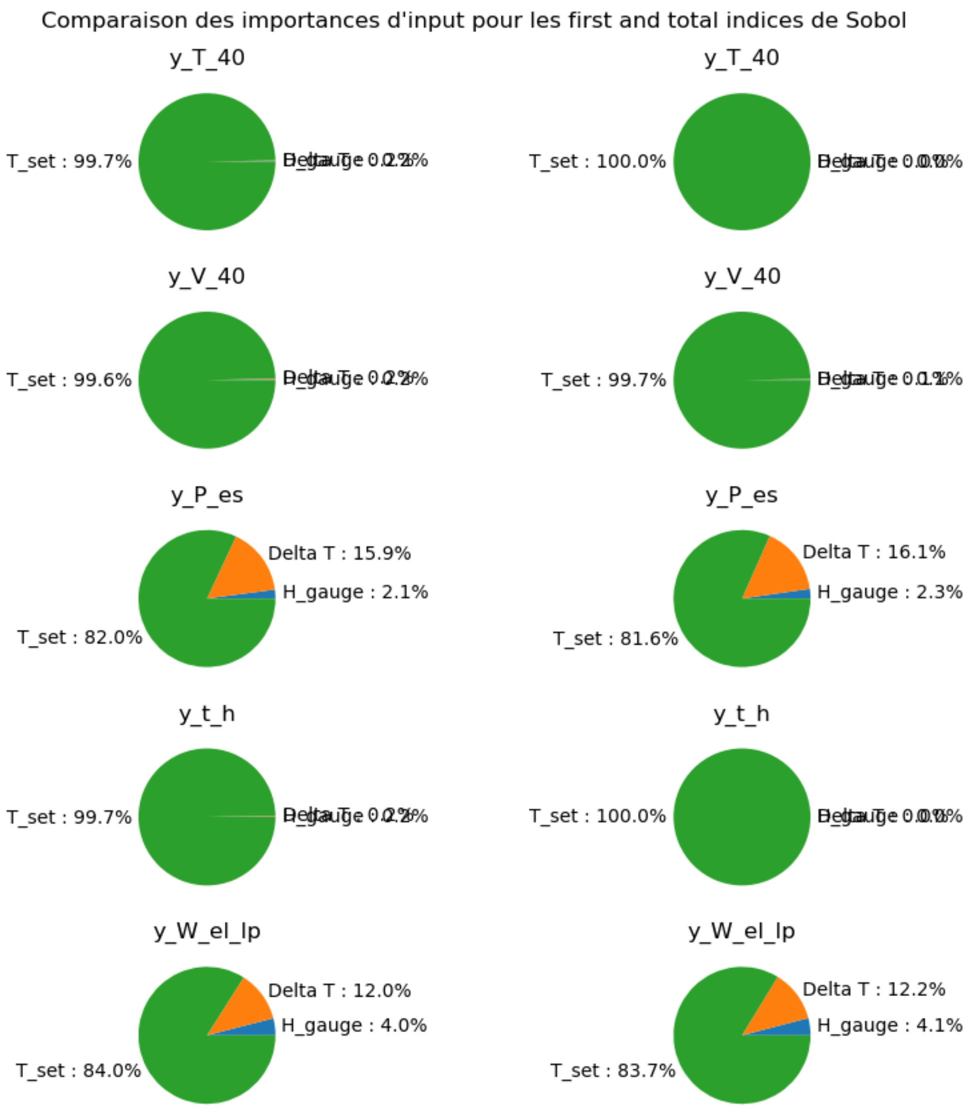


(d) T_{hys} distribution

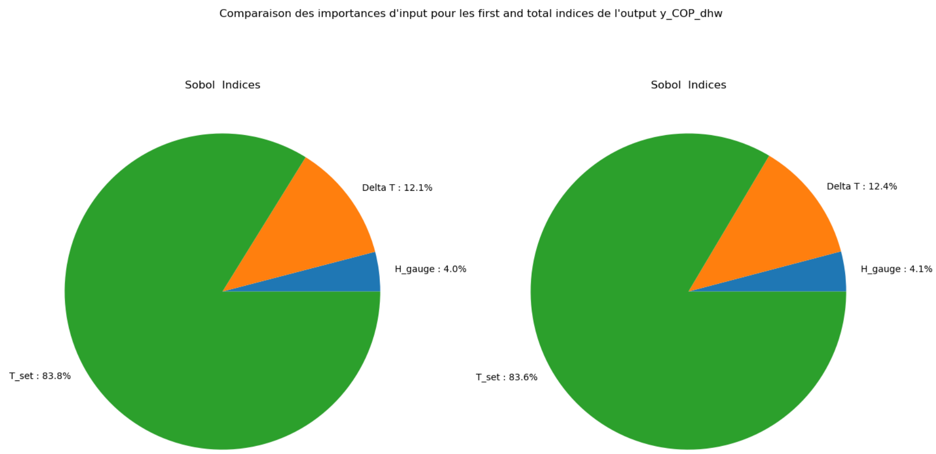


(e) T_{set} distribution

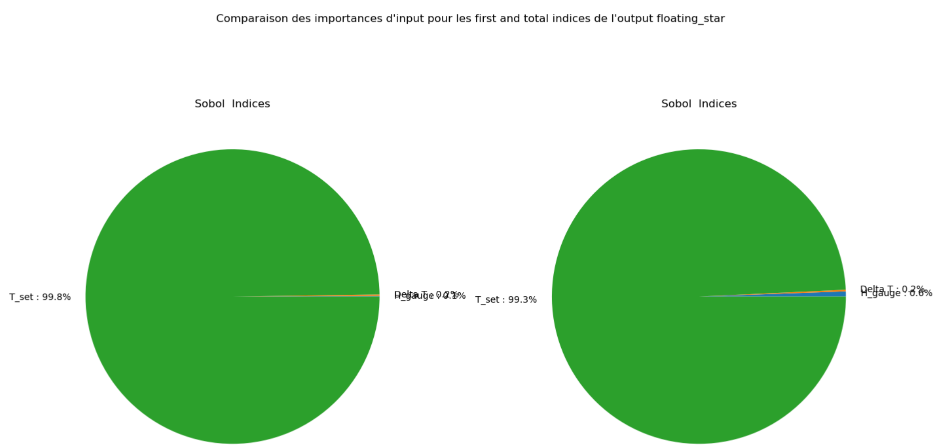
$$\text{distribution} = [\text{Normal}()/200, \text{Normal}()/2, \text{Normal}()/2]$$



We notice that for this distribution the surrogate depend mostly on T_{set}. So for "y_T_40," "y_V_40" and "y_t_h" we can based the surrogate model only on T_{set} without loss to the robustness. So



Sobol indice for COP_{DHW}



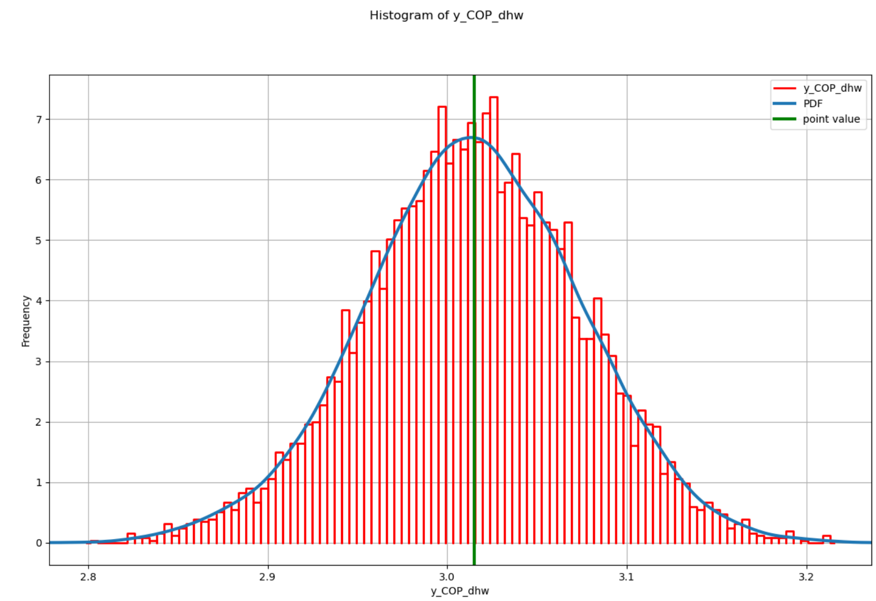
Sobol indice for $Star$

We conclude that our COP_{DHW} and $Star$ are very sensitive to T_{set}

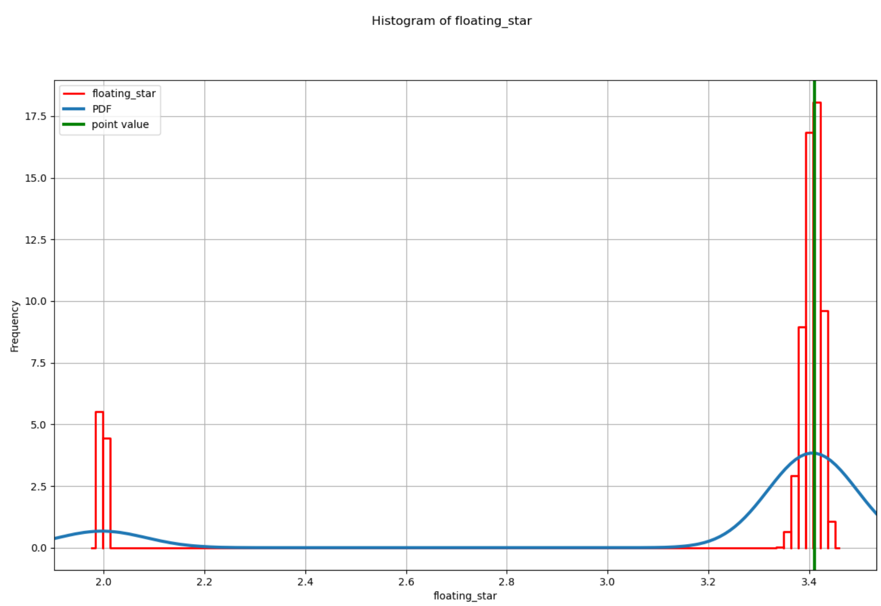
0.4 Uncertainty analysis

For the uncertainty analysis, we keep the same distribution as for the sensitivity:
For the 1000 surrogate model, we get these COP_{DHW} and $Star$ by varying:

If we vary all 3 input at the same time:



Histogram of COP_{DHW}



Histogram of $Star$

We can observe the possibility of getting a lower star rating than the original, we were hoping to find a 3 star score for all but with this distribution there is a small population of points that get the 1 star rating! (On the histogram just above, all $floating\_star$ values are below 2 which results in being truncated to 1 for the final $Star$ rating).
We can deduce that this comes from the variation on T_{set} since their histograms of $Star$ are similar.
In reality, we would avoid taking these parameters to create our product because it has ~15% of having 1 star rather than 3 during the verification of the standard. The company would probably prefer to take other more robust parameters to get 3 stars or even prefer to guarantee a 2 star score.

0.5 Reference

[6] DUMONT Quentin, KLOCKENBRING Dimitri. 2021. "*Optimisation multi-objectifs sous contraintes*". University of Strasbourg.

[7] DUMONT, Quentin. 2021. "*Project report - BDR thermea group*". University of Strasbourg.