MED for GP Model Selection

Kristyn Pantoja

1/30/2020

MMED for GPs

Sequential M-MED for GP

Space-filling input points

Close-together input points

Increasingly farther away input points

Finer Grid

MMED for GPs

Applying M-MED to Gaussian Process Model Selection

- Goal: Choose a design that will distinguish the two gaussian process models.
- Distinguishing functions vs. distributions over functions:
  - For regression models, we use $f_D(\mathbf{x}) = \text{Wasserstein}(\phi_{0,\mathbf{x}}, \phi_{1,\mathbf{x}})$. What is the distance function now? What are $\phi_{0,\mathbf{x}}, \phi_{0,\mathbf{x}}$?
  - Key Question: Do we need to consider the predictive distribution for each GP model?
    - Doing so would give us an option for $\phi_{0,\mathbf{x}}, \phi_{0,\mathbf{x}}$.
    - We would need to have at least some data in order to model each Gaussian Process (training set) and use M-MED to select points for comparing them.

Simulations Set-Up

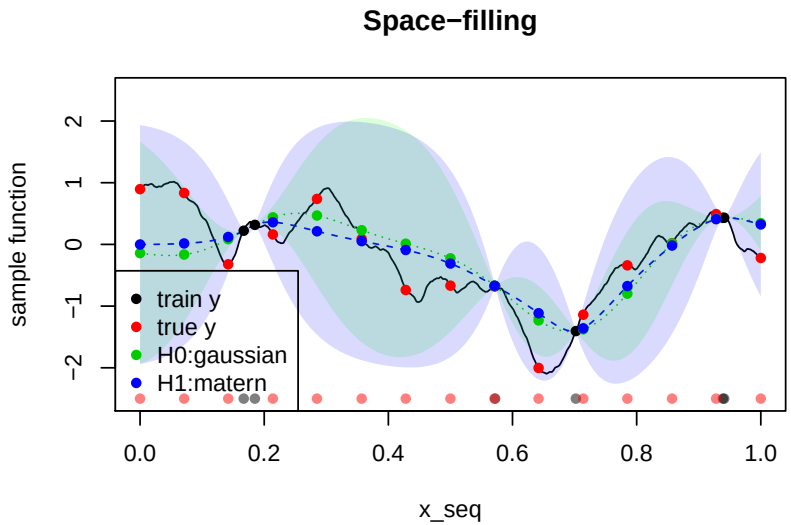
- I consider two cases:
  - Gaussian vs. Matern kernels, where the true function is generated from the Matern kernel
  - Matern vs. Periodic kernels, where the true function is generated from the Periodic kernel
- To evaluate MED for each case, I draw uniformly selected input points for my training set, and then apply MED to the data.
- I consider two measures for comparing MED to a space-filling design:
  - ratio of RSS for each hypothesized kernel:

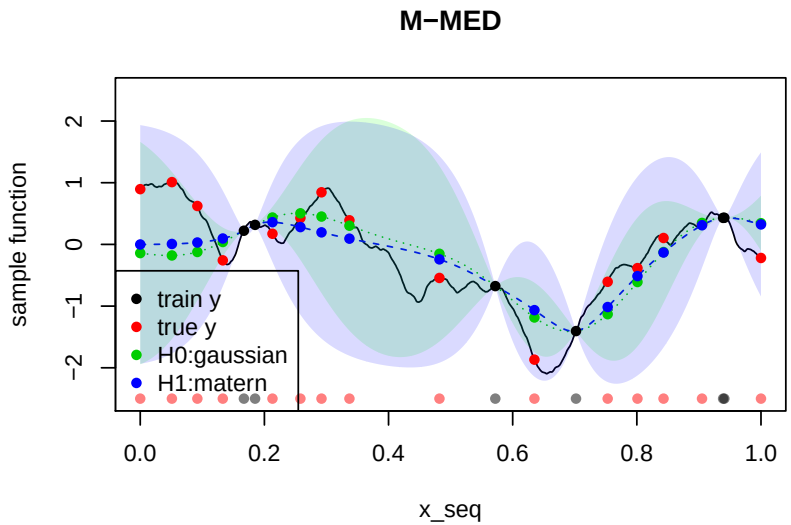
$$\frac{\sum_{i \in \mathbf{D}} (y_i^{\text{pred}_0} - y_i^{\text{new}})^2}{\sum_{i \in \mathbf{D}} (y_i^{\text{pred}_1} - y_i^{\text{new}})^2}$$

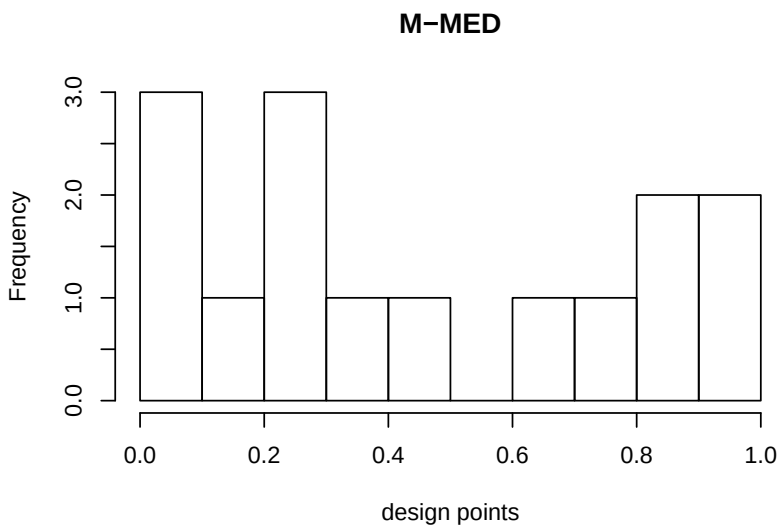
  - likelihood ratio:

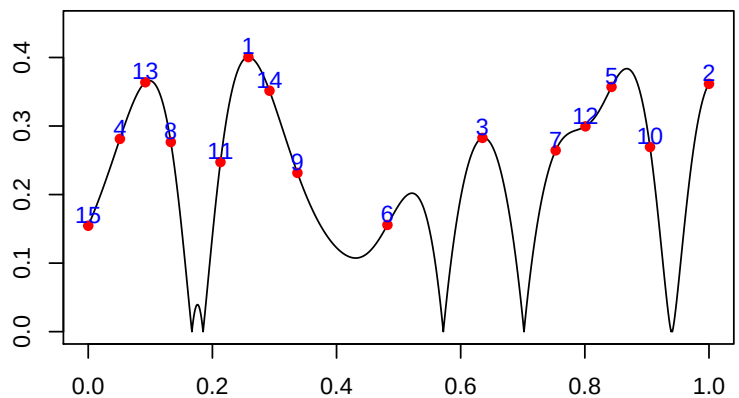
$$\frac{L(y^{\text{new}} | \xi, y^{\text{obs}}, \mathbf{X}^{\text{obs}}, \Theta = 0)}{L(y^{\text{new}} | \xi, y^{\text{obs}}, \mathbf{X}^{\text{obs}}, \Theta = 1)}$$

Gaussian vs. Matern (simulation)









Gaussian vs. Matern: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.62	0.053	0.25	0.54	0.62	5.5

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.42	0.042	0.14	0.42	0.39	4.8

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.68
```

Gaussian vs. Matern: log ratio of predictive densities

(after removing NAs caused from non-invertible matrix)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-Inf	-2.9e+08	-4.8e+07	-Inf	-9.9e+05	-1.3e+05

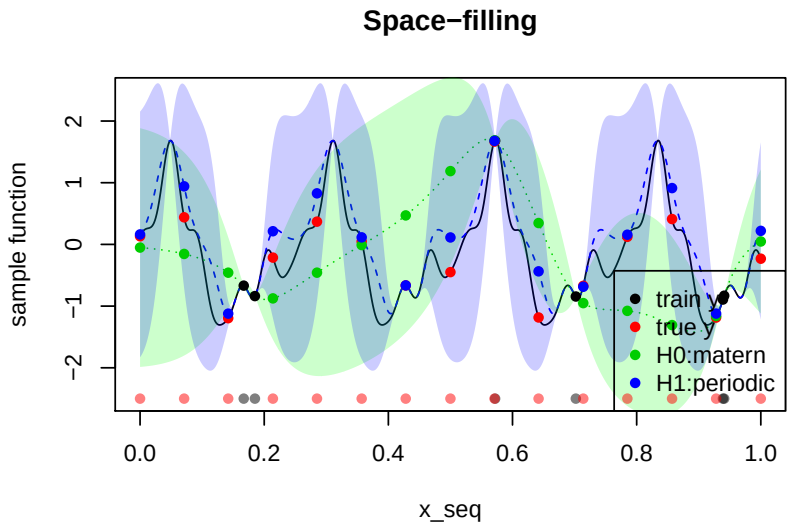
Space-filling

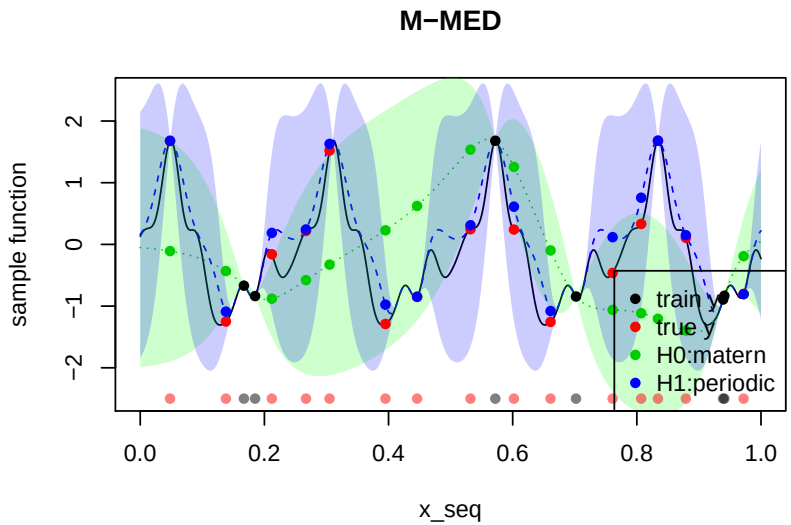
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.7e+10	-1.2e+06	-3.7e+05	-1.1e+09	-9.6e+04	-1.8e+04

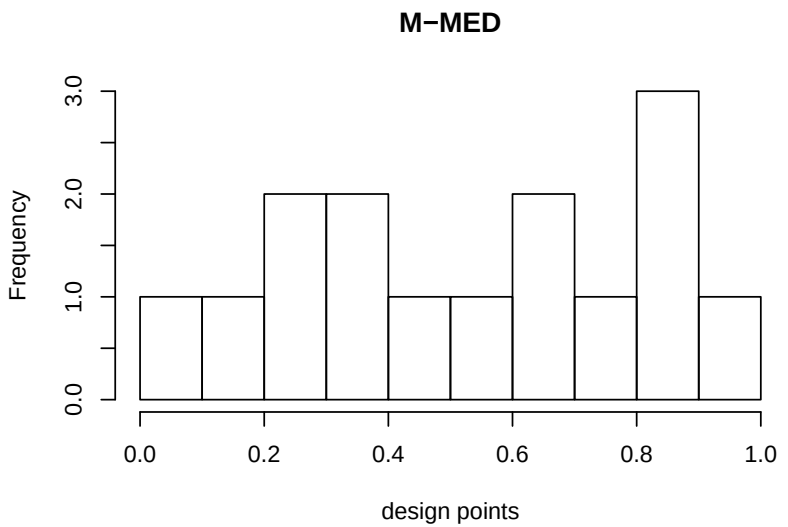
Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

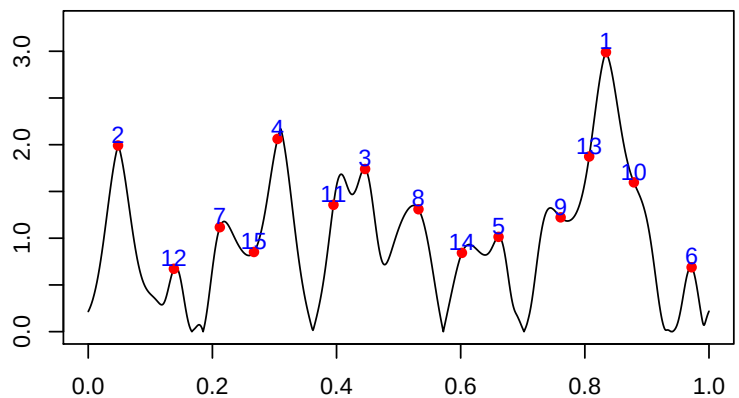
```
## [1] 0.875
```

Matern vs. Periodic (simulation)









Matern vs. Periodic: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.56	1.3	1.8	2	2.2	3.6

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.094	0.72	1	1.1	1.4	3

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 1
```

Matern vs. Periodic: log ratio of predictive densities

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-280	-160	-130	-140	-110	-65

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-Inf	-99	-85	-Inf	-60	-38

Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

```
## [1] 0.92
```

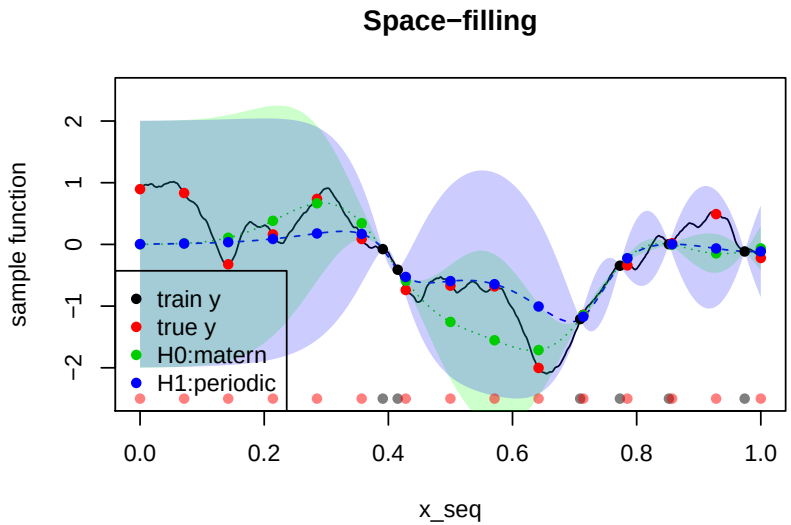
Sequential M-MED for GP

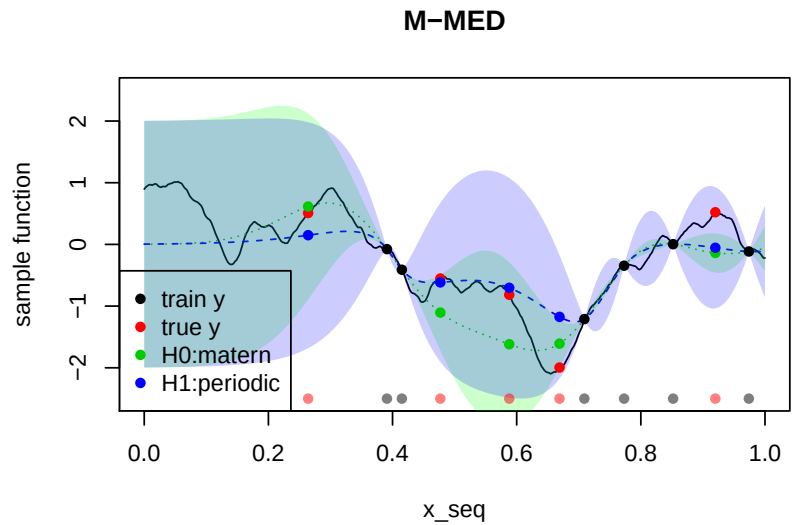
Will a sequential design improve results?

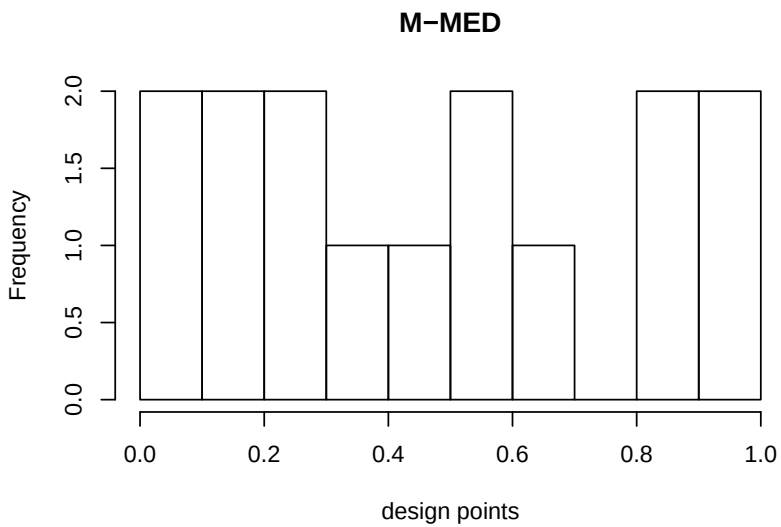
For the sequential designs, I:

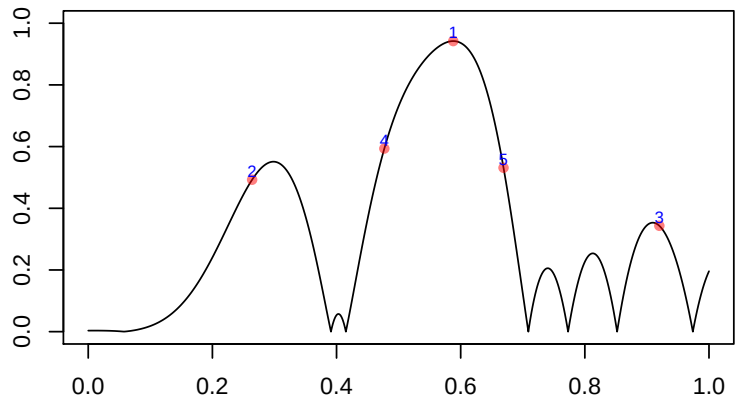
1. Start with 6 input data
2. Use SMMED to sequentially gather 15 new data points in 3 steps (5 new points at each step)
3. To compare SMMED to a space-filling design, I use the previous evaluations on the 15 new points (pretending that data was not gathered for them yet)

Gaussian vs. Matern (sequentially): Step 1

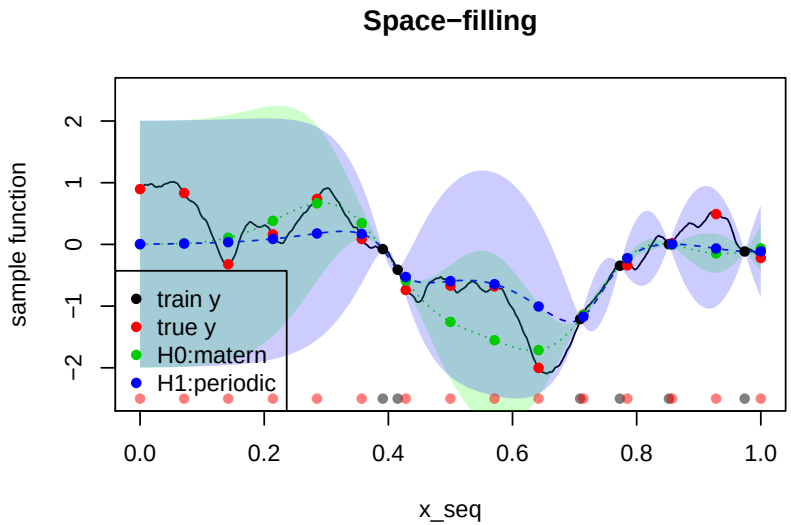


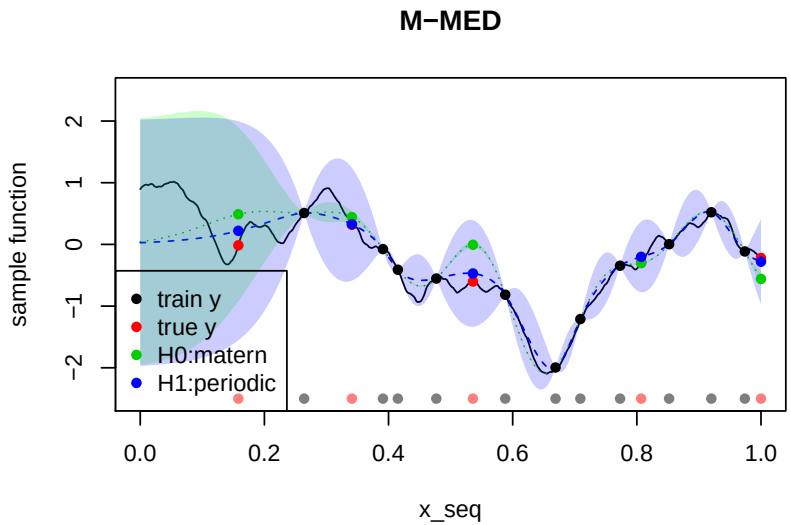


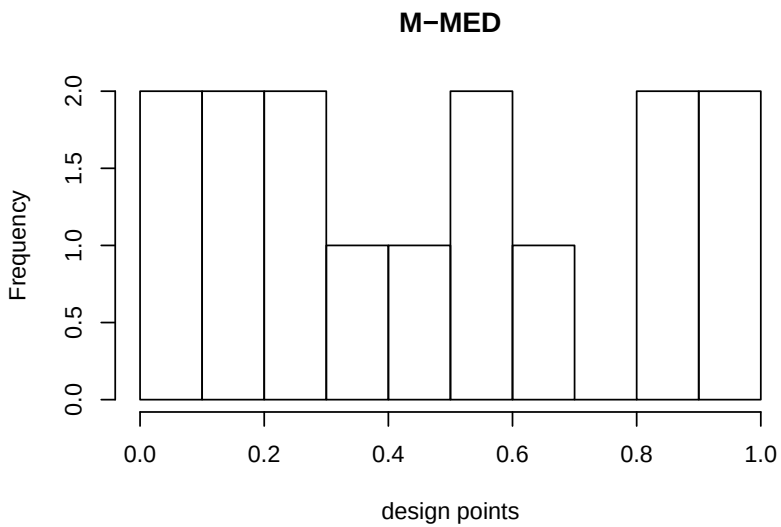


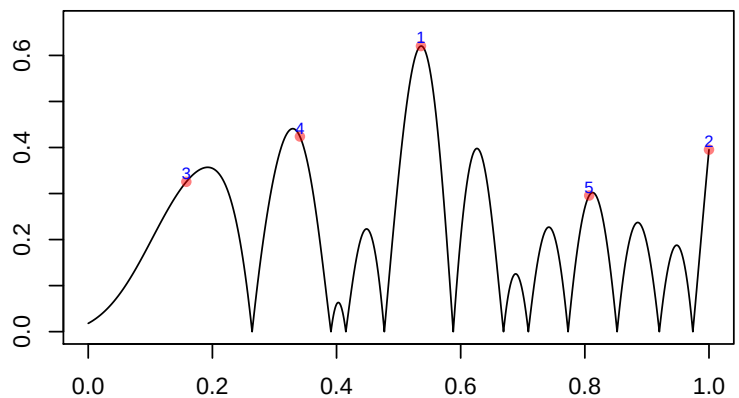


Step 2

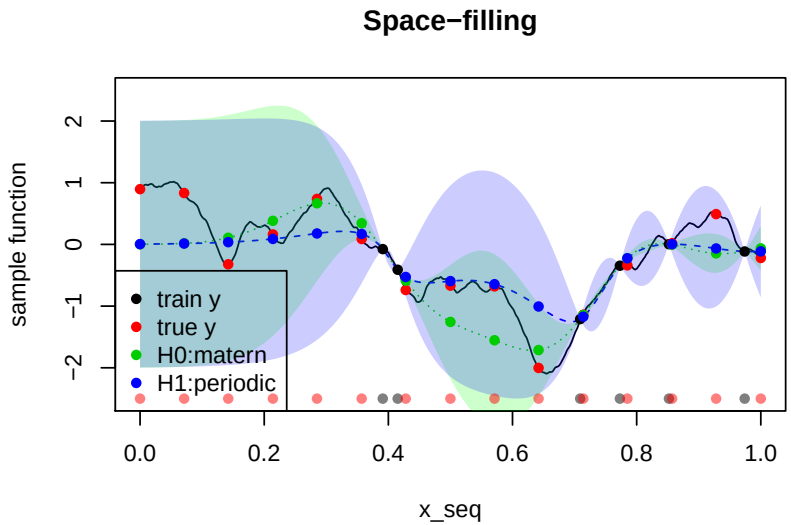

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

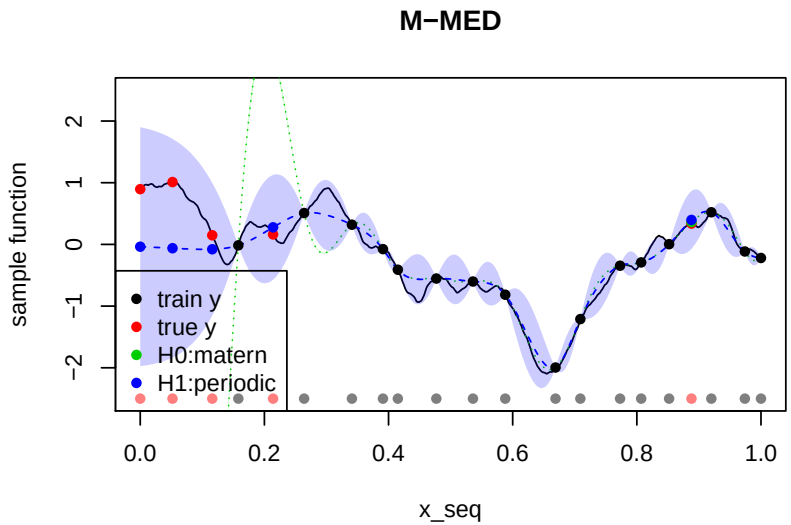

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

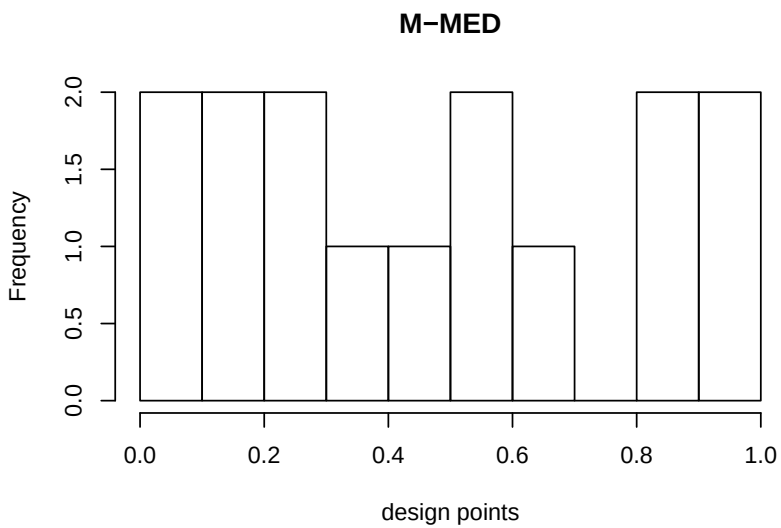

M−MED
Frequency
design points

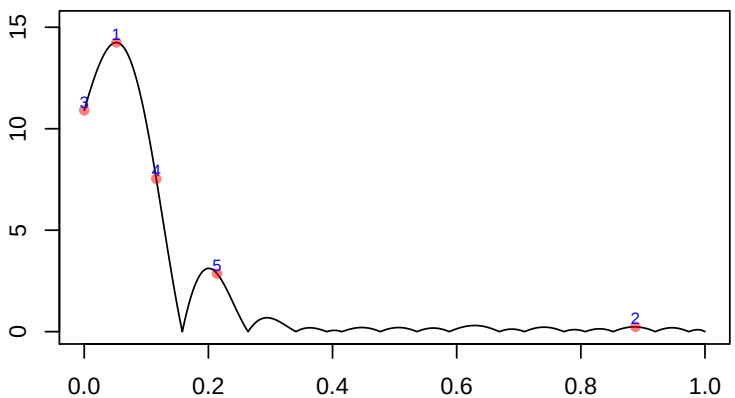

Wasserstein(x)
1
2
3
4
5

Step 3

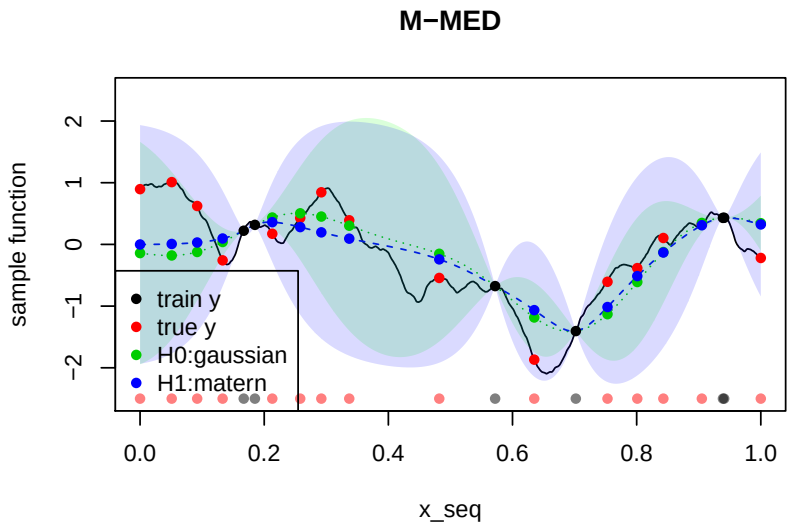

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

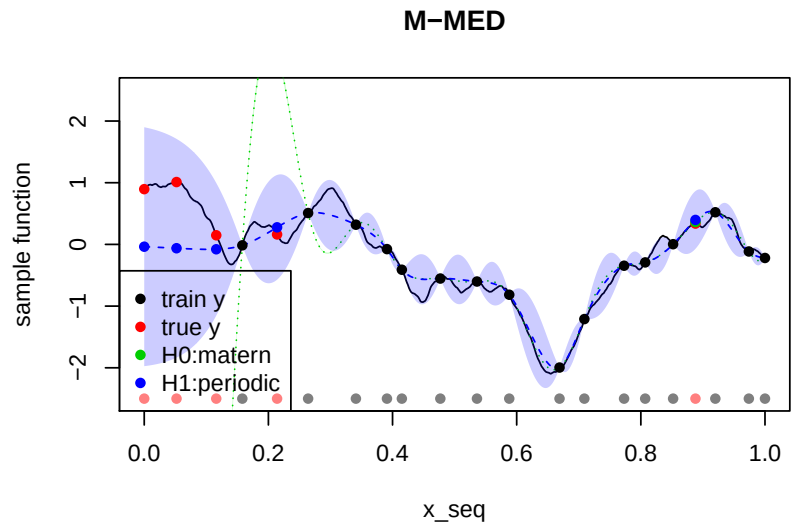

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

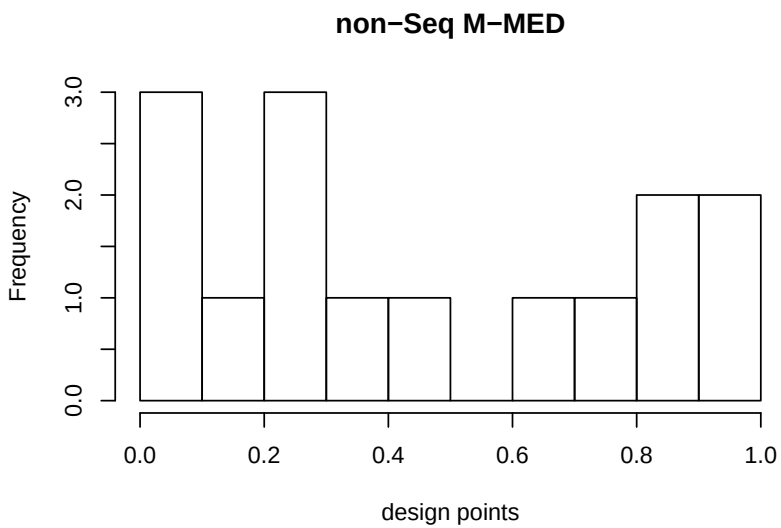

M−MED
Frequency
design points

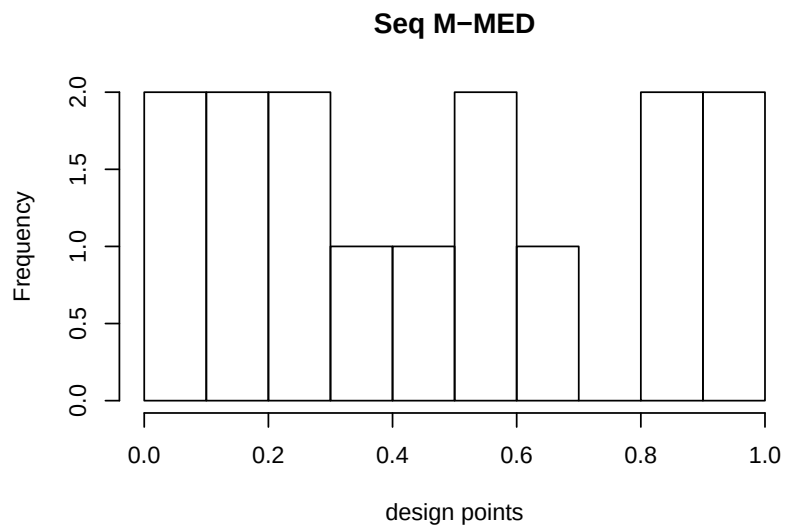

Wasserstein(x)

Compare to non-Sequential M-MED









Gaussian vs. Matern: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.49	-0.082	0.27	0.48	0.64	2.8

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.53	-0.21	0.24	0.45	0.57	2.5

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.64
```

Gaussian vs. Matern: log ratio of predictive densities

(after removing NAs caused from non-invertible matrix)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-8.7e+05	-1.2e+05	-5.5e+04	-1.3e+05	-3.1e+04	-5.8e+03

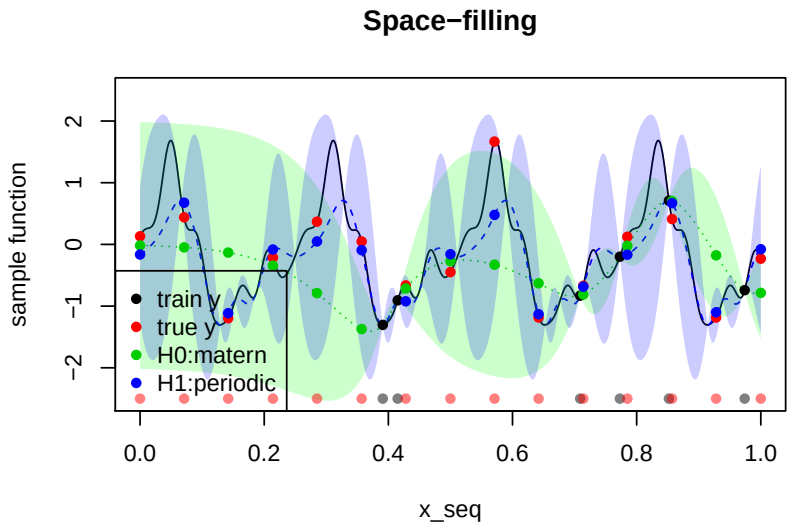
Space-filling

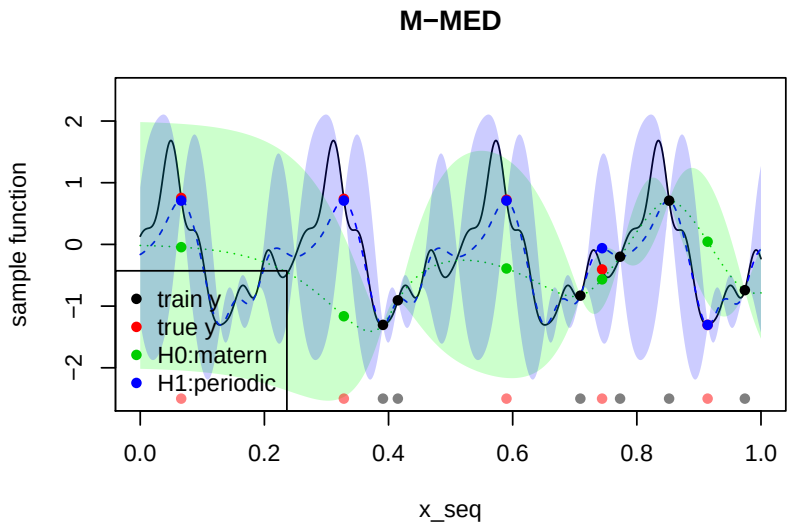
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-9.6e+08	-1.0e+07	-2.8e+06	-6.6e+07	-3.6e+05	-2.9e+04

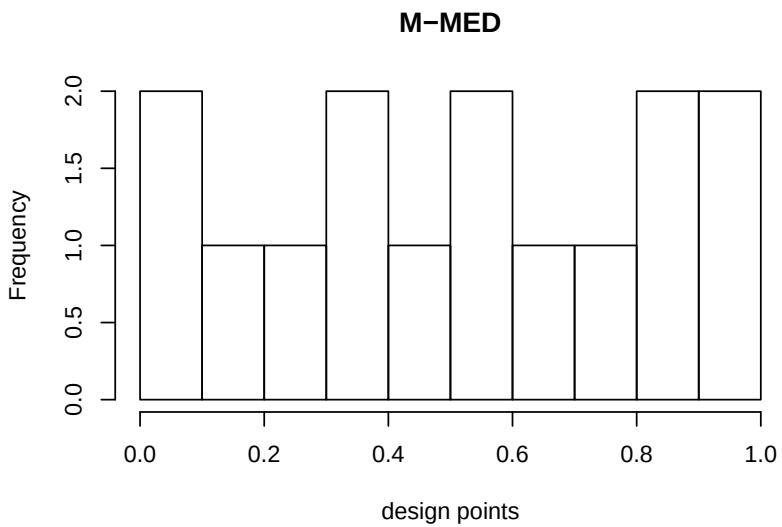
Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

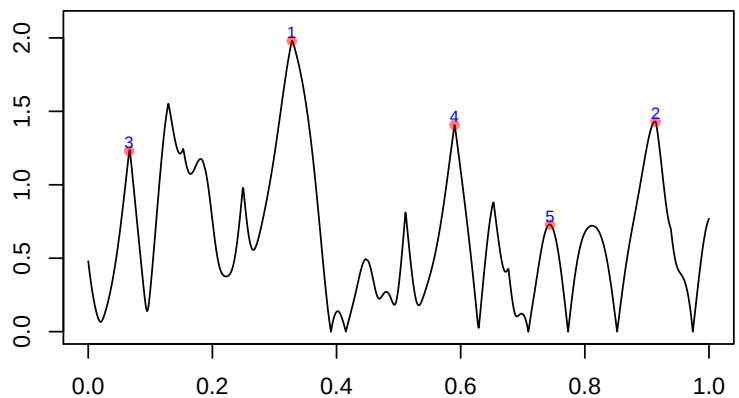
```
## [1] 0
```

Matern vs. Periodic (sequentially): Step 1

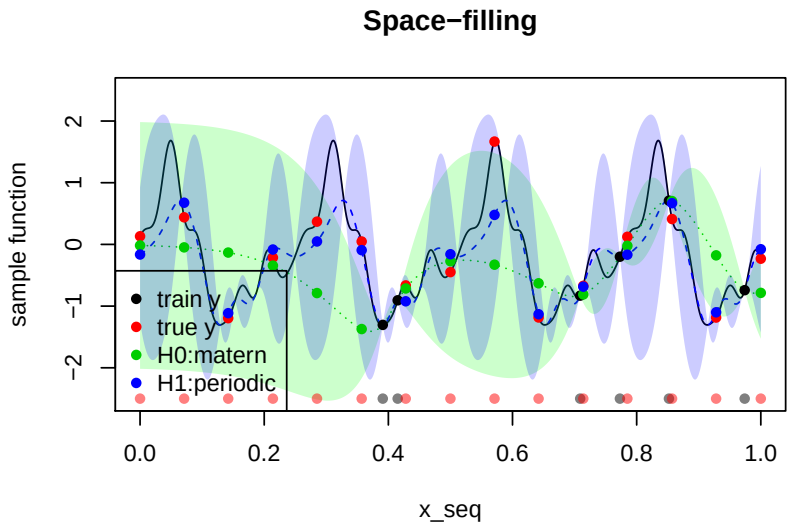


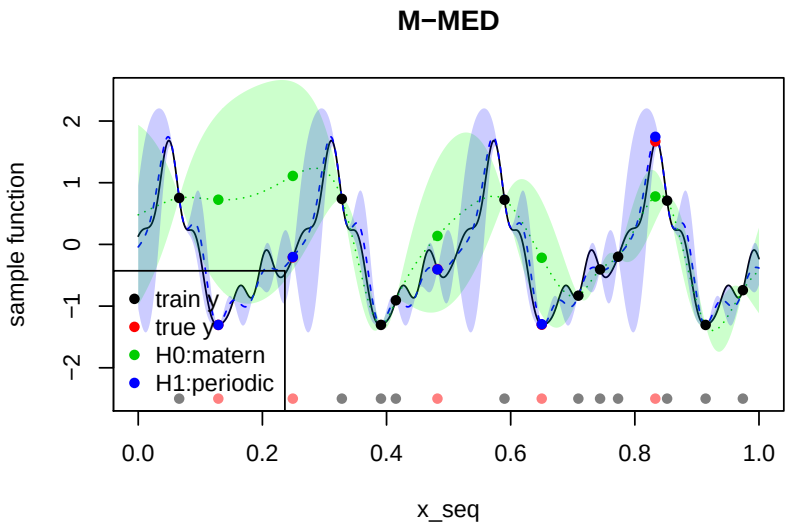


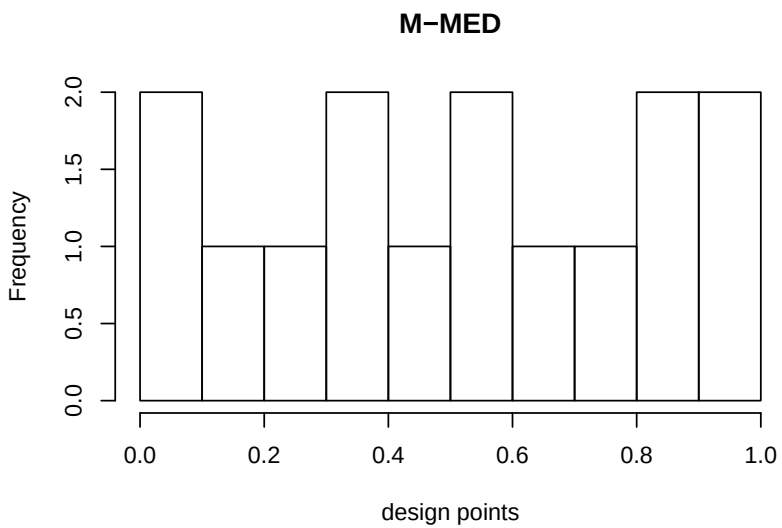


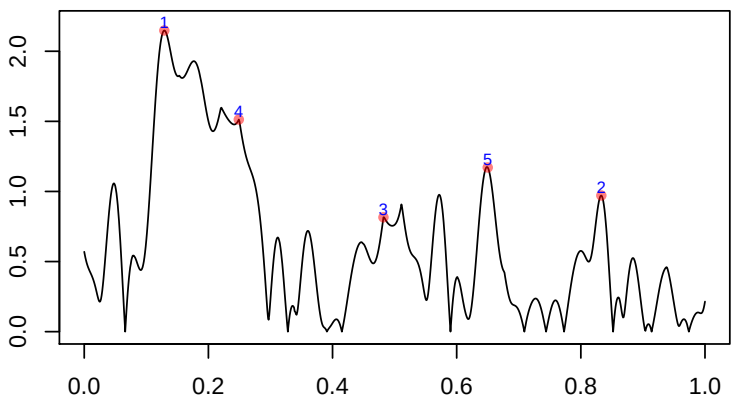


Step 2

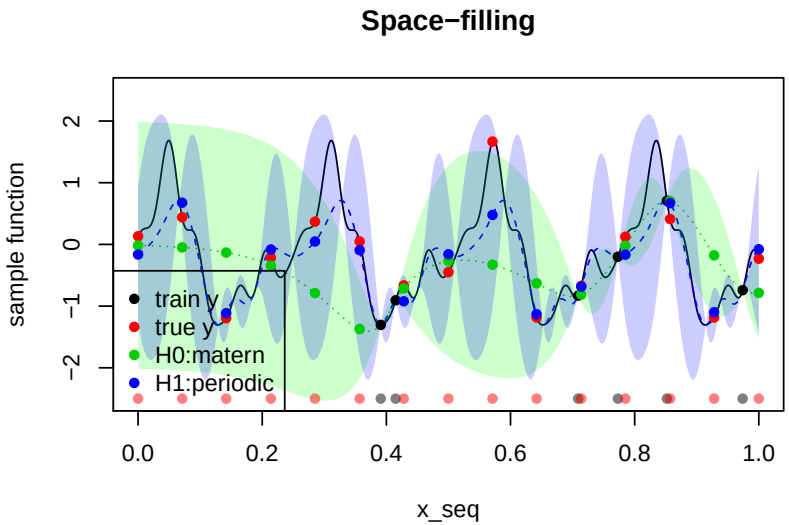

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

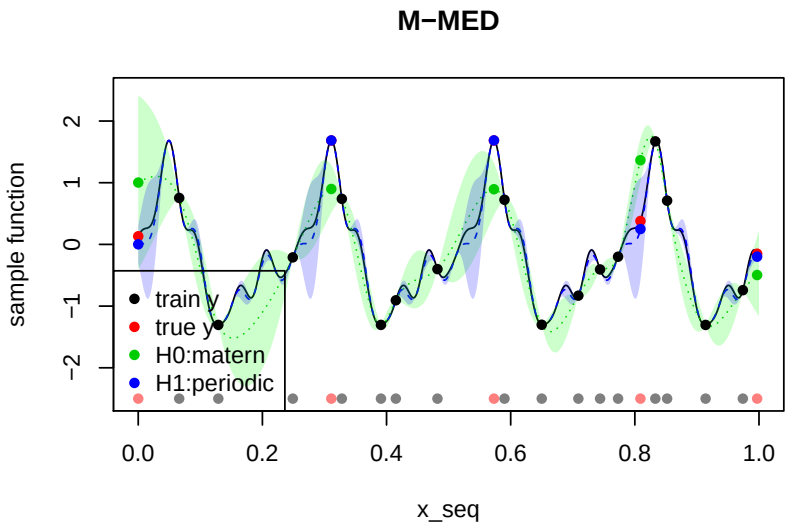

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

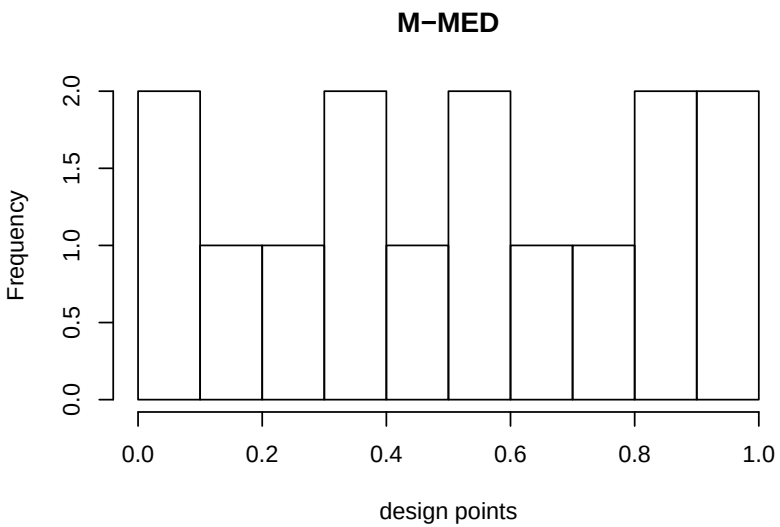

M−MED
Frequency
design points

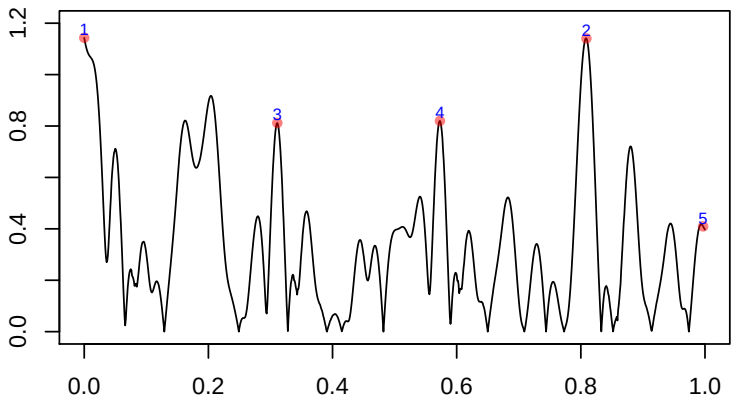

Wasserstein(x)
1
4
3
5
2

Step 3

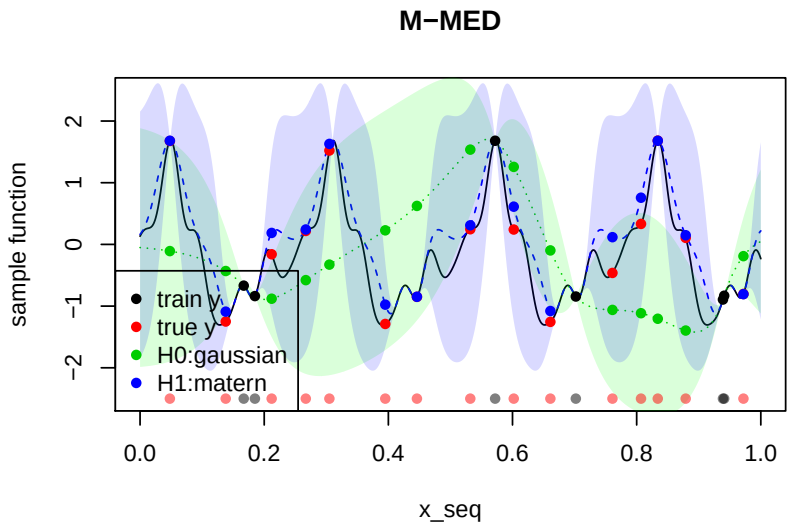

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

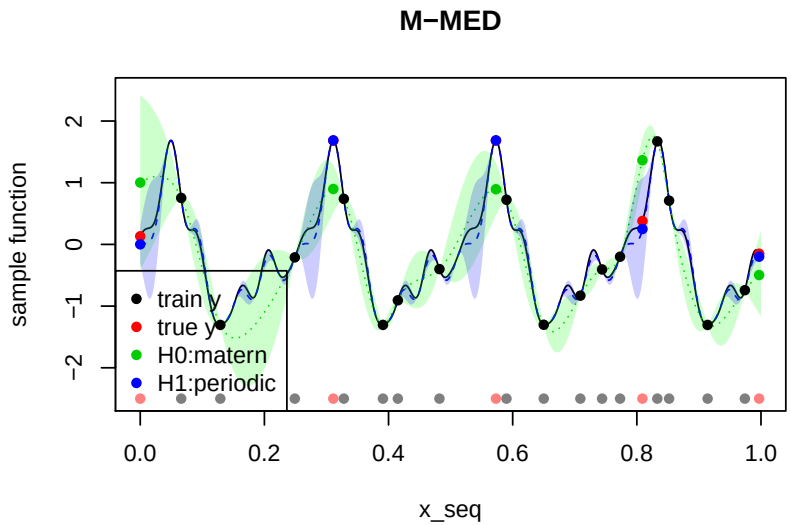

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

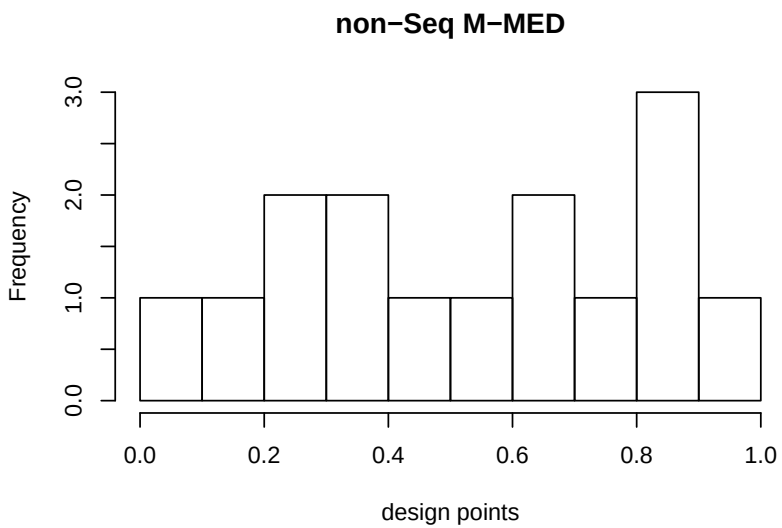

M−MED
Frequency
design points

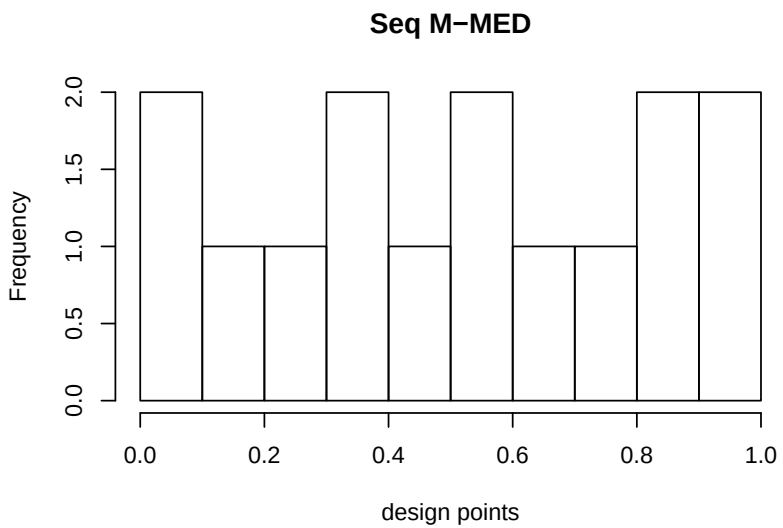

Wasserstein(x)

Compare to non-Sequential M-MED









Matern vs. Periodic: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.26	1.2	1.5	1.6	2	4.7

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.35	0.9	1.3	1.3	1.7	2.5

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.56
```

Matern vs. Periodic: log ratio of predictive densities

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-490	-250	-160	-210	-130	-89

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-150	-96	-75	-77	-55	-32

Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

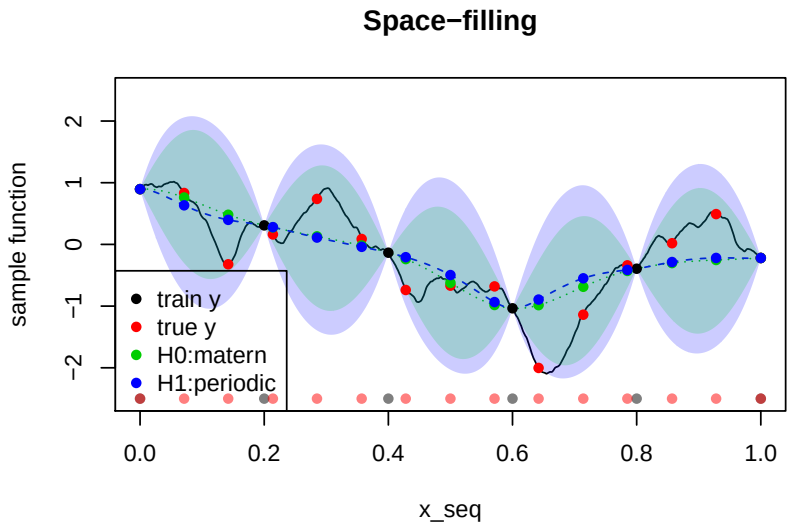
```
## [1] 1
```

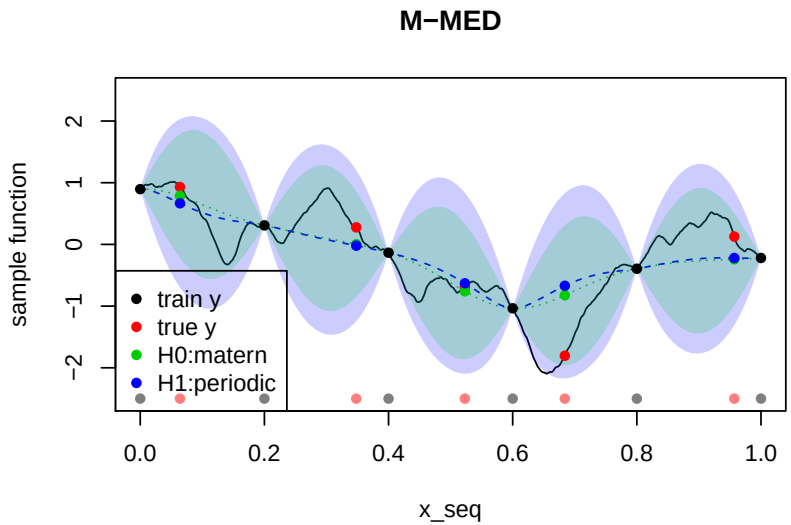
Space-filling input points

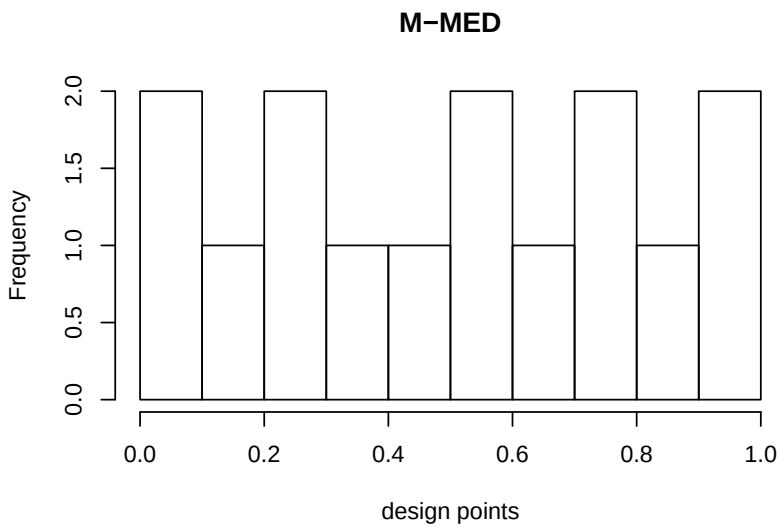
Motivation for space-filling input points

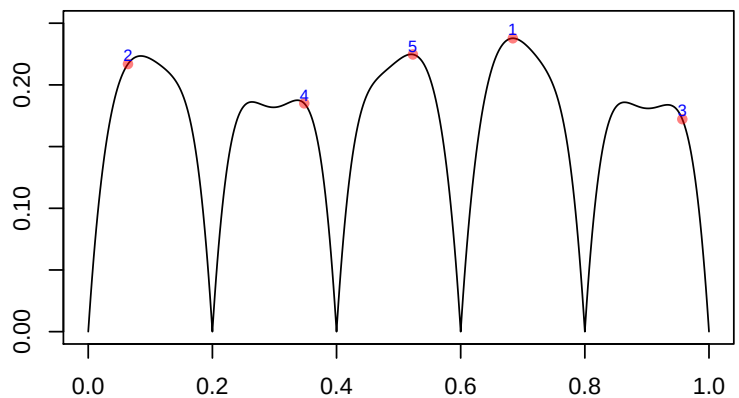
Since the Gaussian process first wants to fill in areas where the uncertainty is highest generally, a space-filling design would evenly distribute those uncertainty regions so that perhaps the Gaussian Process could start to focus on different things.

Gaussian vs. Matern (sequentially): Step 1

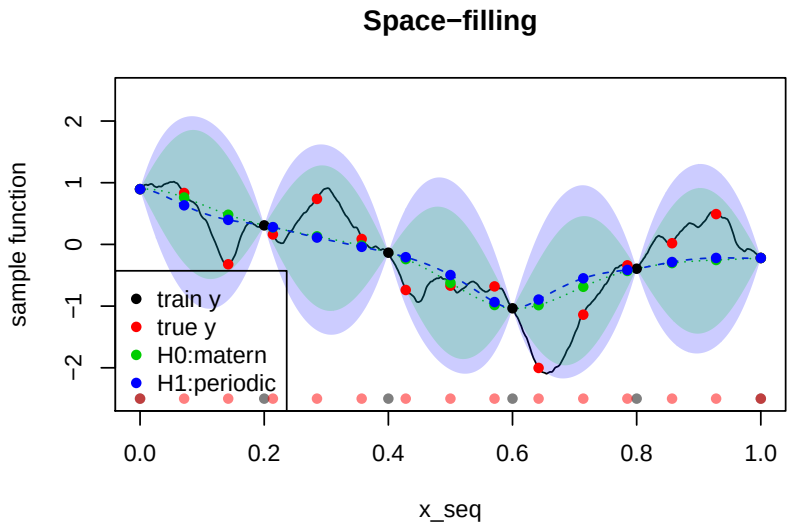


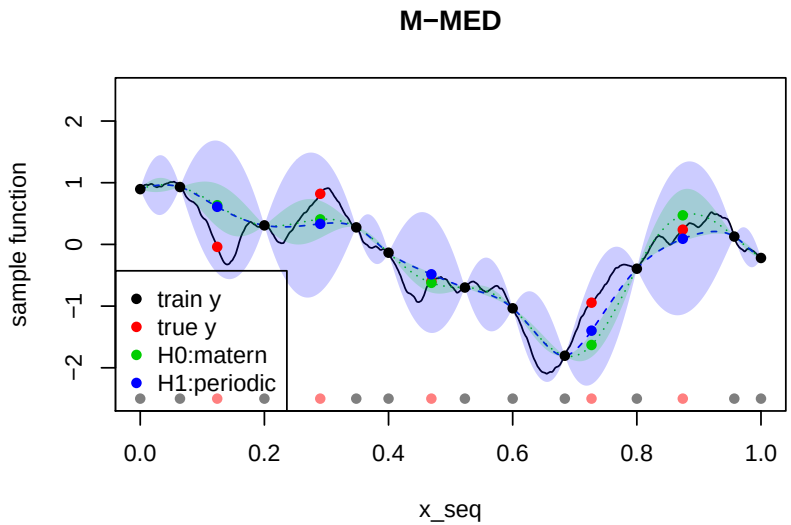


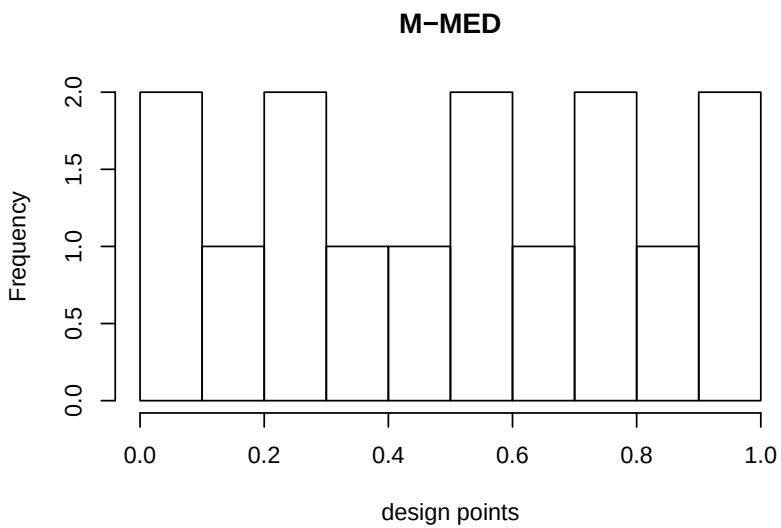


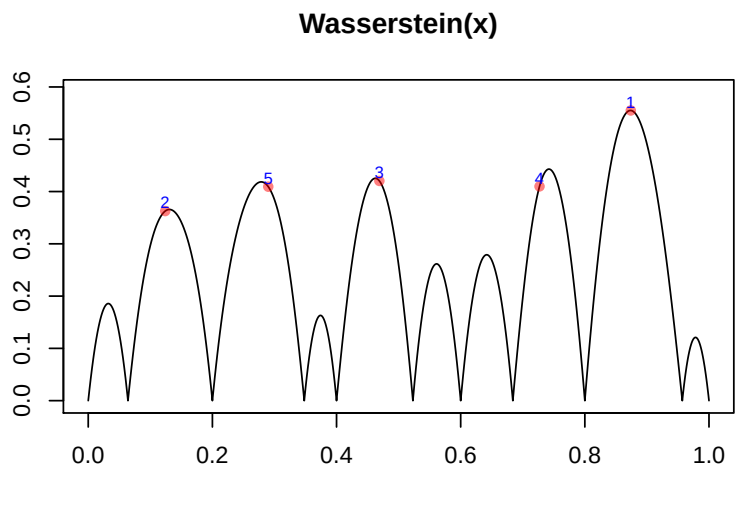


Step 2

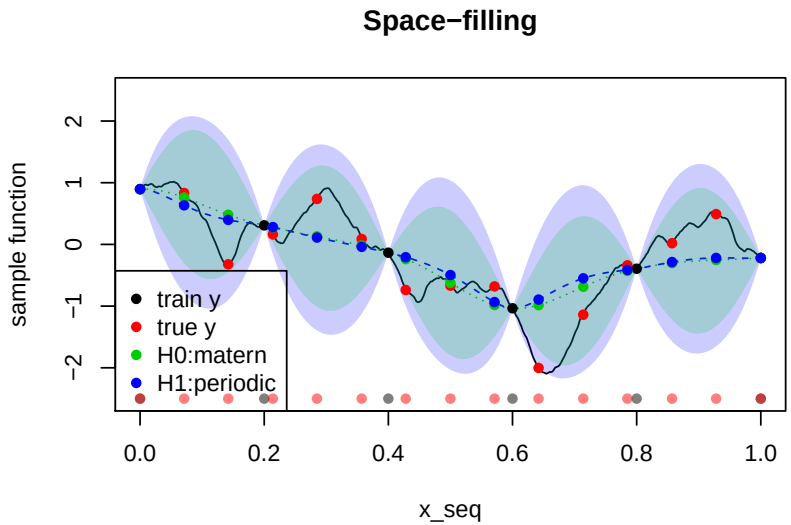

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

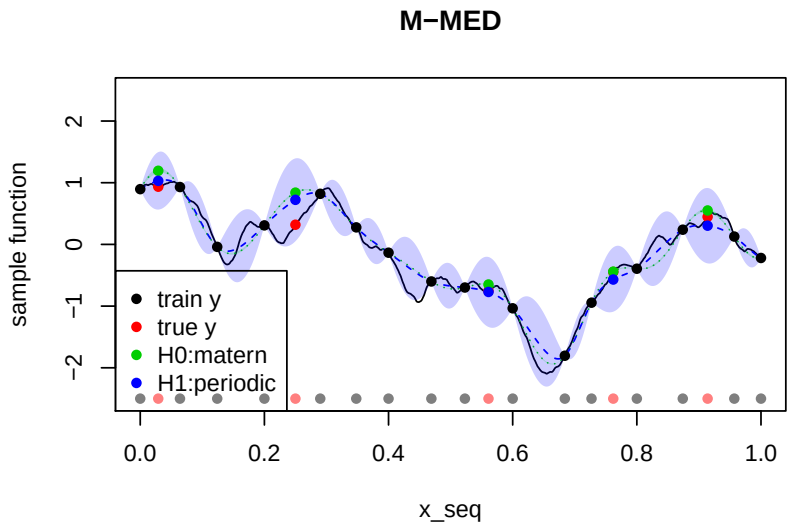

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

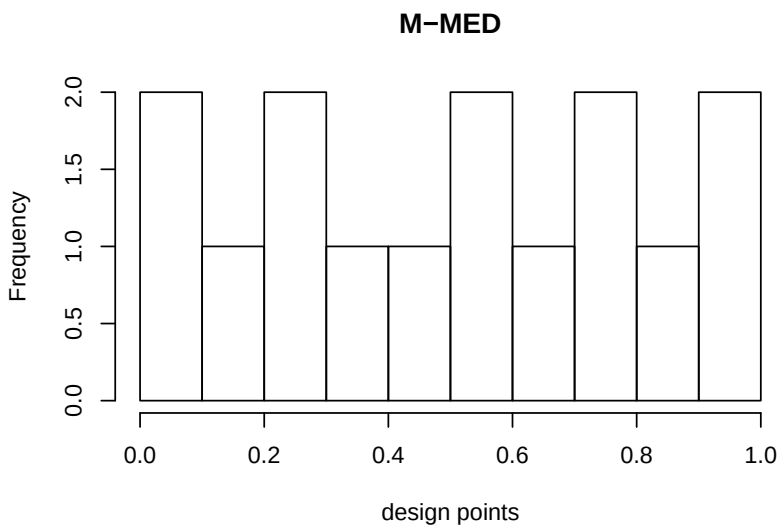

M−MED
Frequency
design points

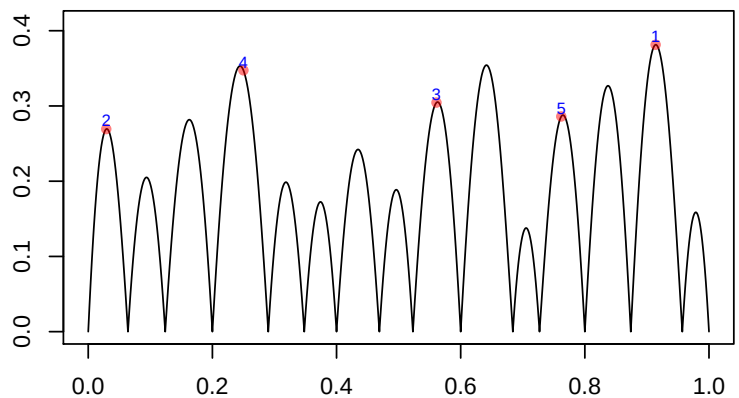

Wasserstein(x)
1
2
3
4
5

Step 3

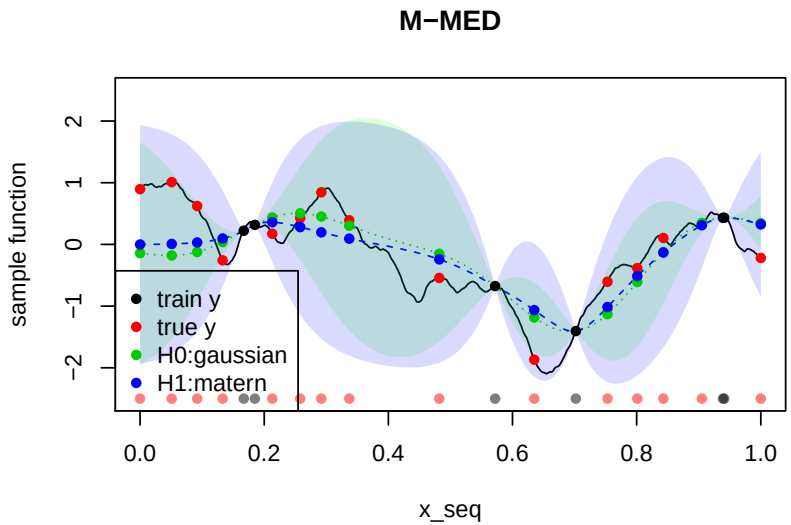

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

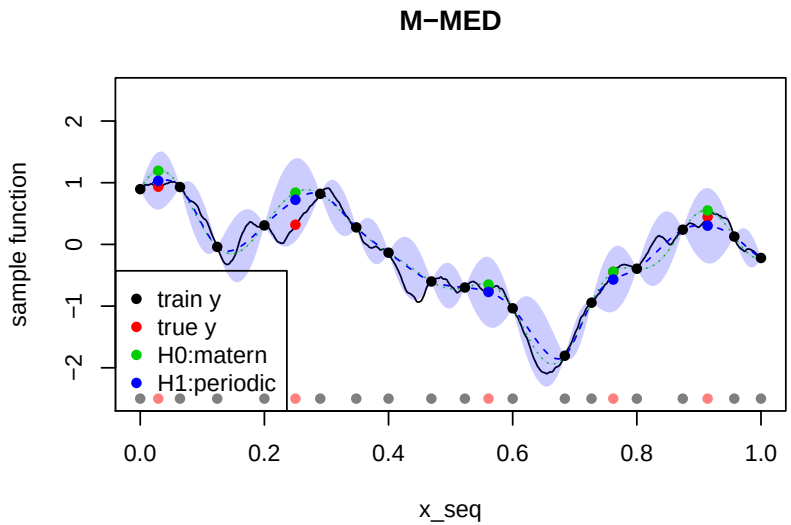

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

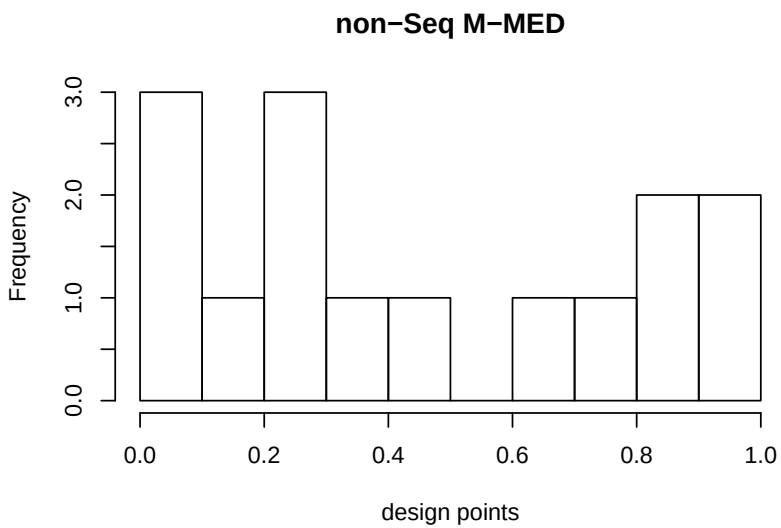

M−MED
Frequency
design points

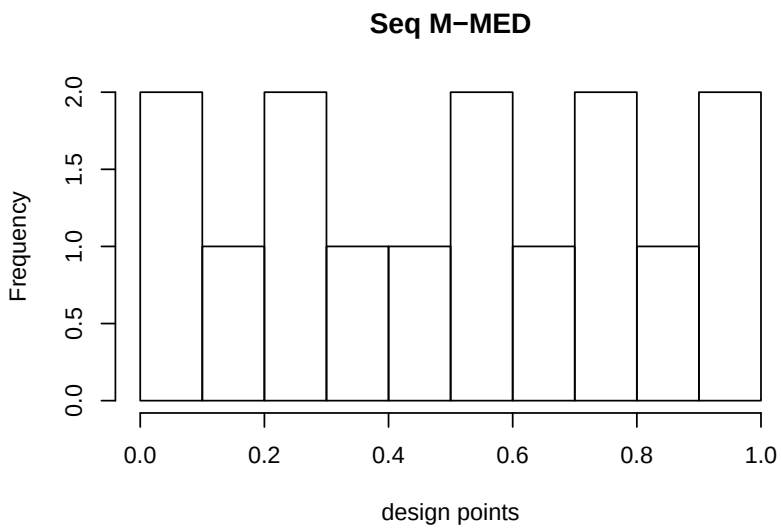

Wasserstein(x)

Compare to non-Sequential M-MED









Gaussian vs. Matern: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.36	-0.085	-0.0041	-0.00016	0.15	0.25

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.39	-0.11	-0.0053	-0.009	0.13	0.24

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.6
```

Gaussian vs. Matern: log ratio of predictive densities

(after removing NAs caused from non-invertible matrix)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-4.0e+05	-1.2e+05	-2.0e+04	-7.7e+04	-7.9e+03	-1.5e+03

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-5.3e+05	-1.2e+05	-6.4e+04	-1.0e+05	-1.5e+04	-9.3e+02

Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

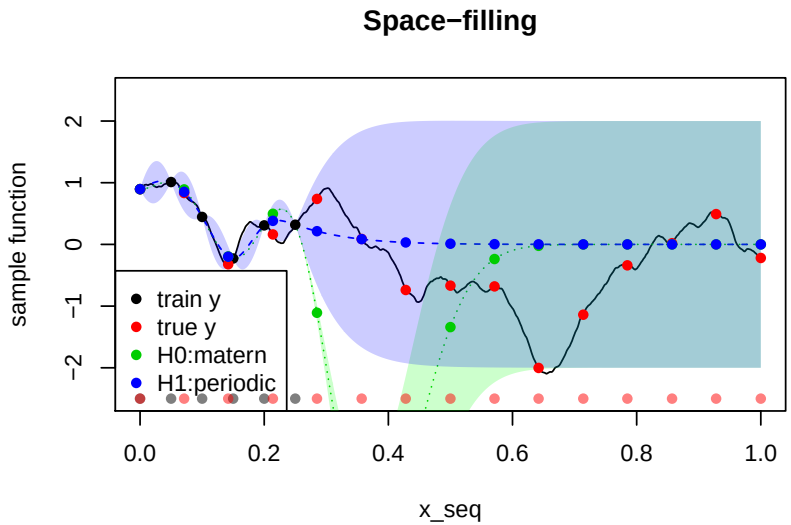
```
## [1] 0.32
```

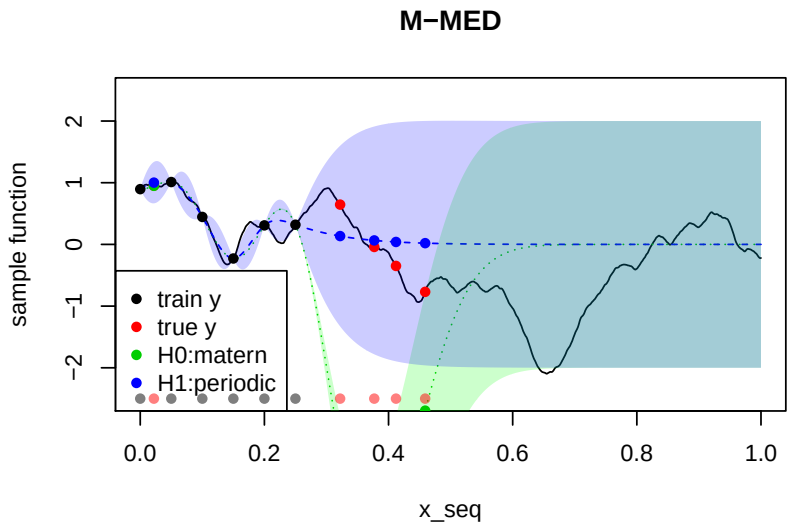
Close-together input points

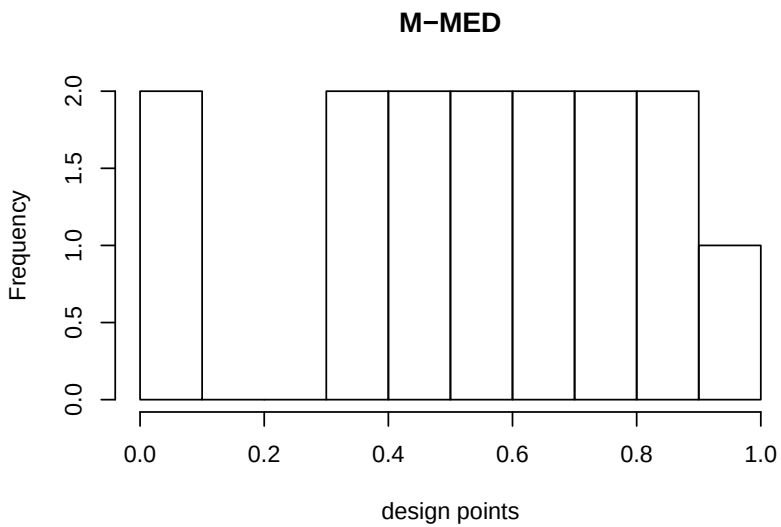
Motivation for close-together input points

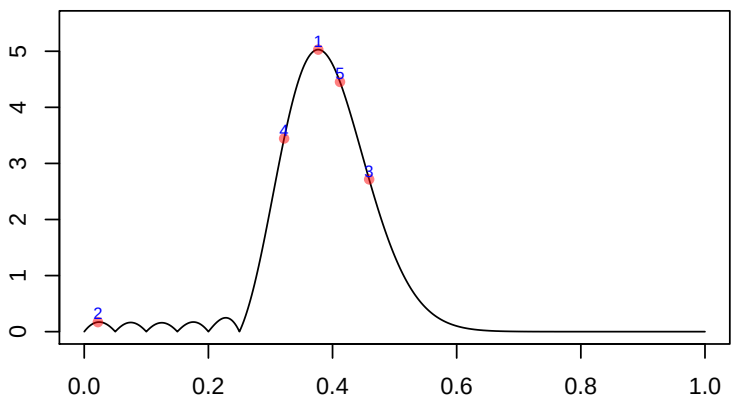
Seeing the space-filling input points seems to lend more evidence to the theory that for the Gaussian vs. Matern case, the main difference is the uncertainty quantification and hence a space-filling design is optimal and sought-after. Maybe the close-together input points will tell a similar story.

Gaussian vs. Matern (sequentially): Step 1

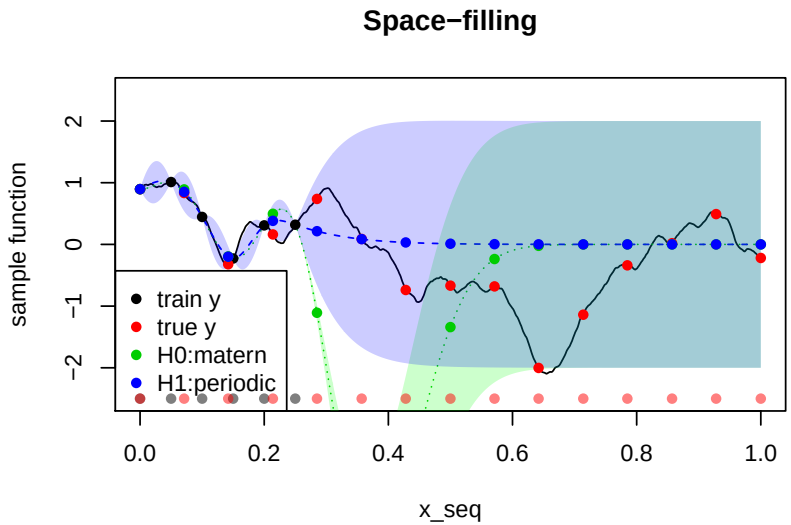


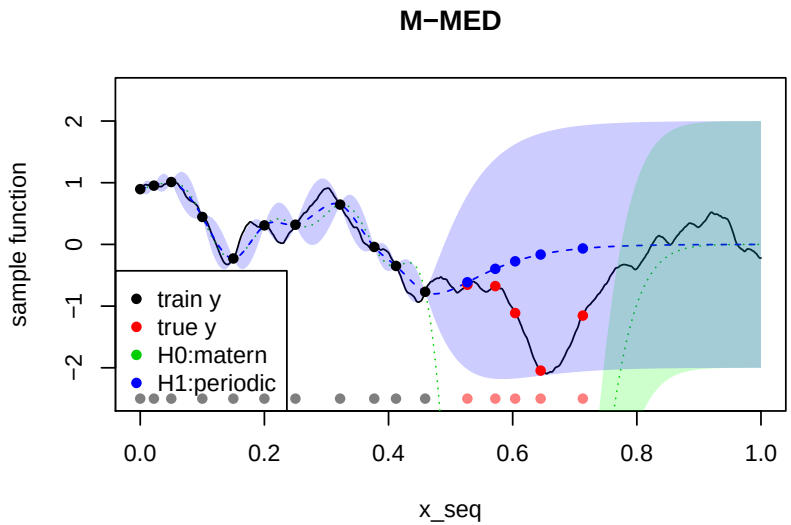


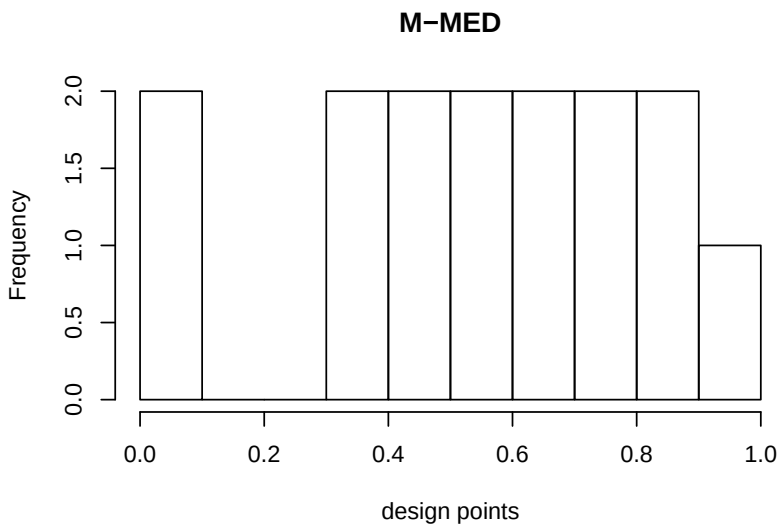


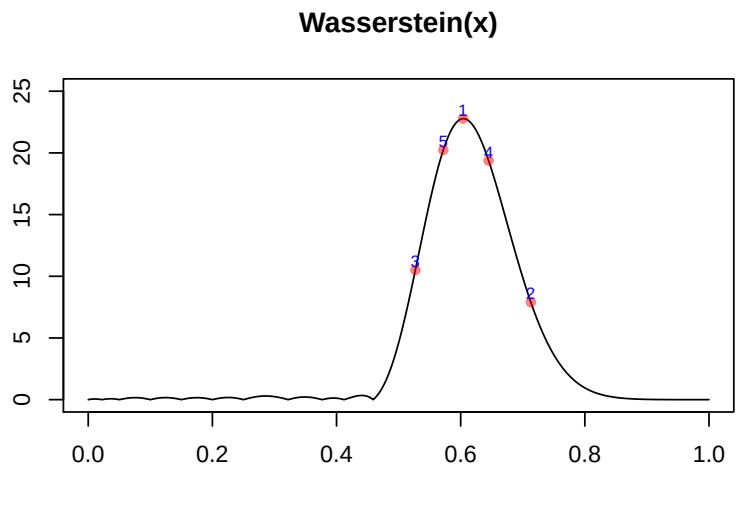


Step 2

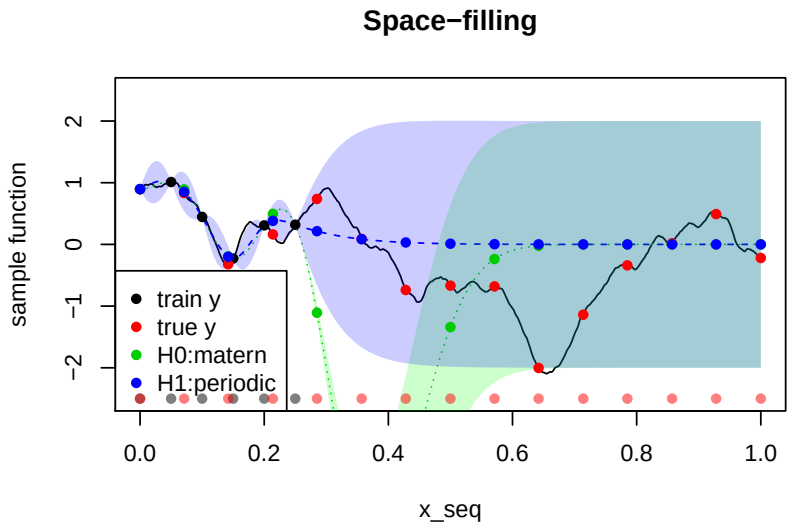

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

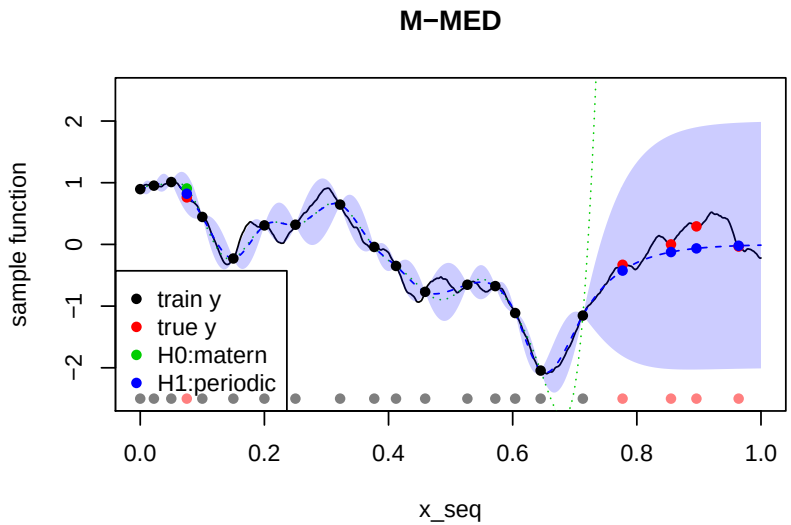

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

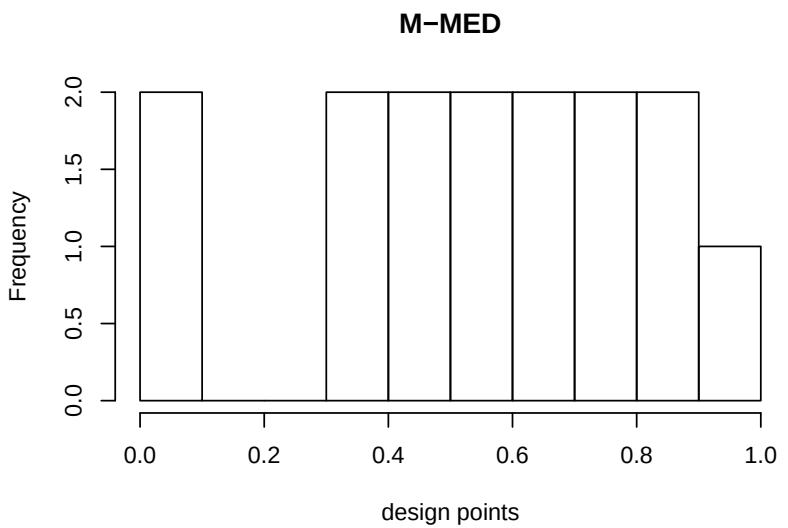

M−MED
Frequency
design points

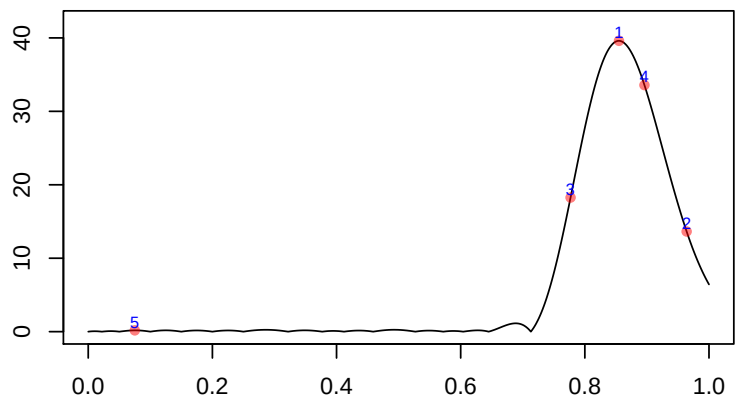

Wasserstein(x)

Step 3

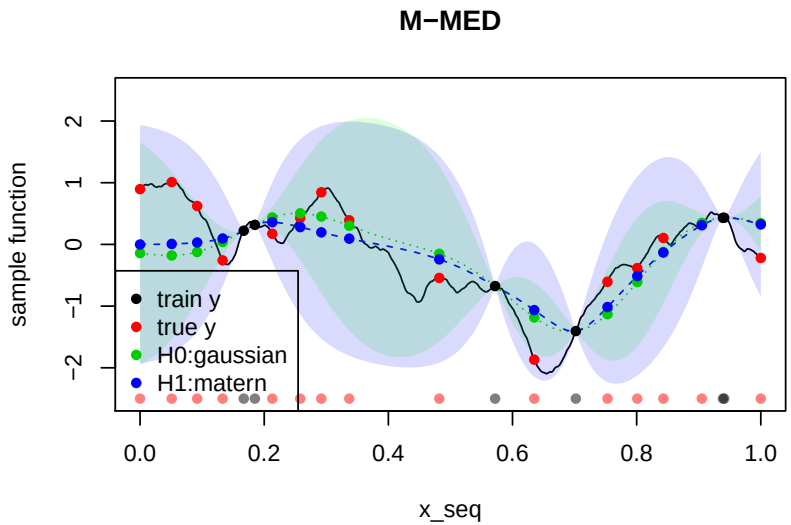

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

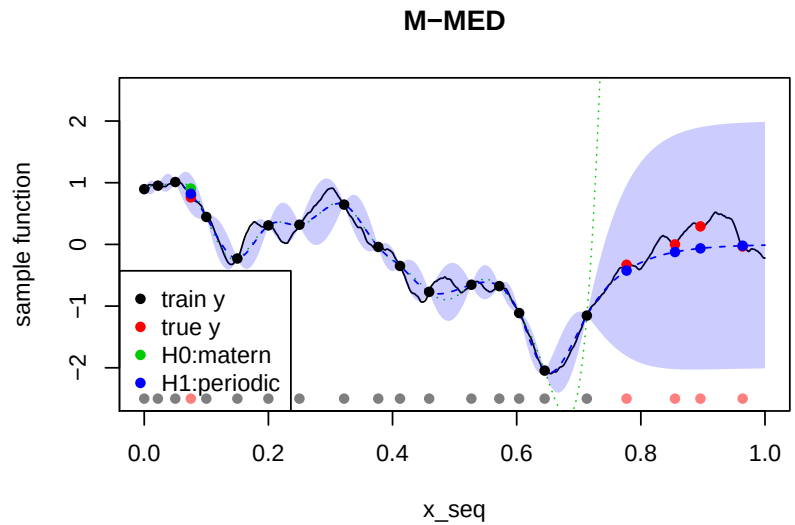

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

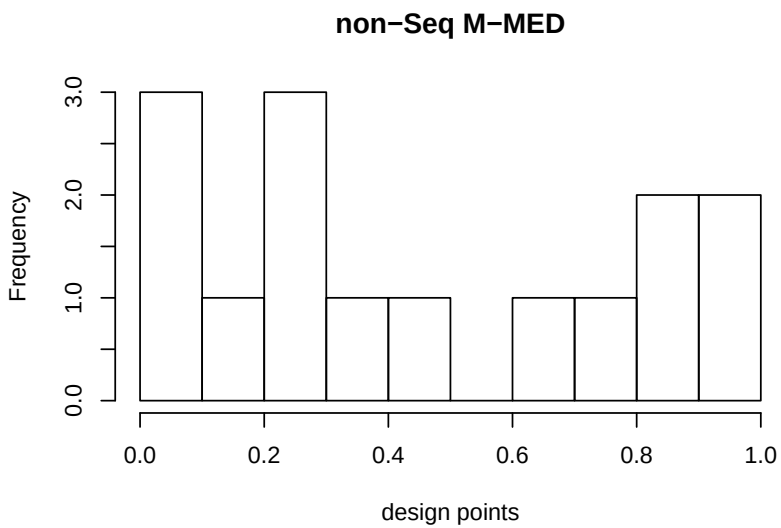

M−MED
Frequency
design points

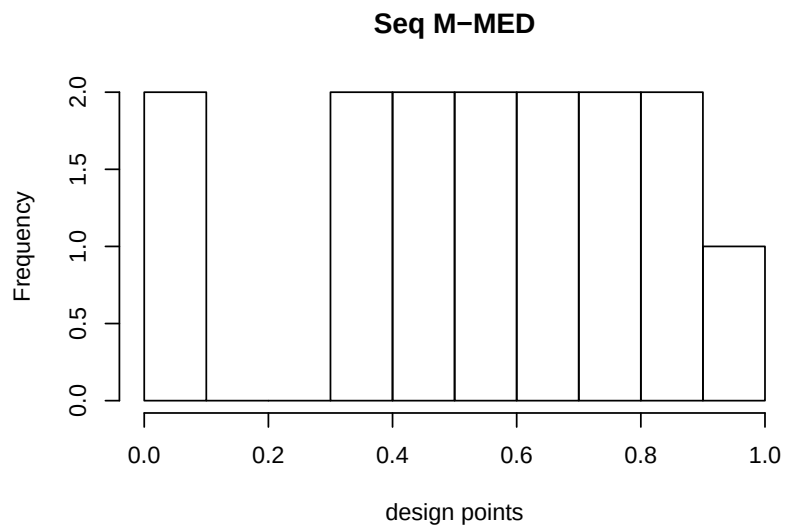

Wasserstein(x)

Compare to non-Sequential M-MED









Gaussian vs. Matern: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.58	0.47	1.6	1.4	2.1	3

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.48	0.41	1.5	1.3	2	2.7

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.76
```

Gaussian vs. Matern: log ratio of predictive densities

(after removing NAs caused from non-invertible matrix)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.9e+07	-2.0e+06	-9.7e+05	-3.5e+06	-1.1e+05	-1.4e+04

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-5.3e+05	-1.2e+05	-6.4e+04	-1.0e+05	-1.5e+04	-8.0e+02

Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

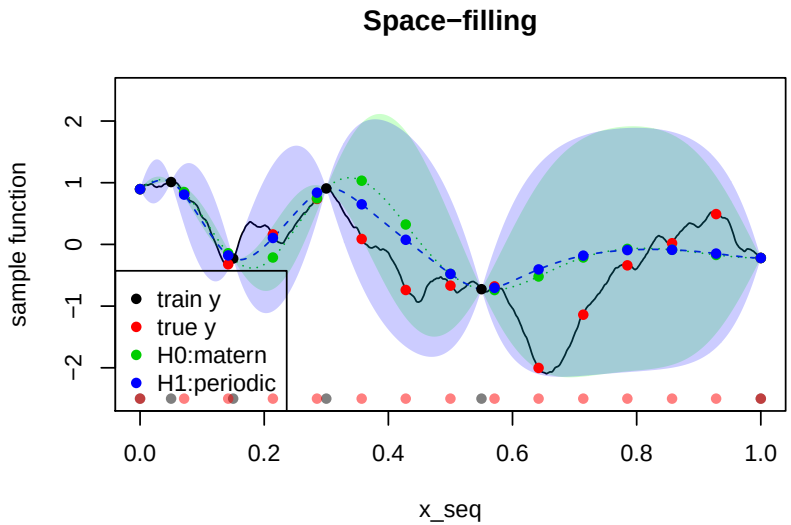
```
## [1] 0.92
```

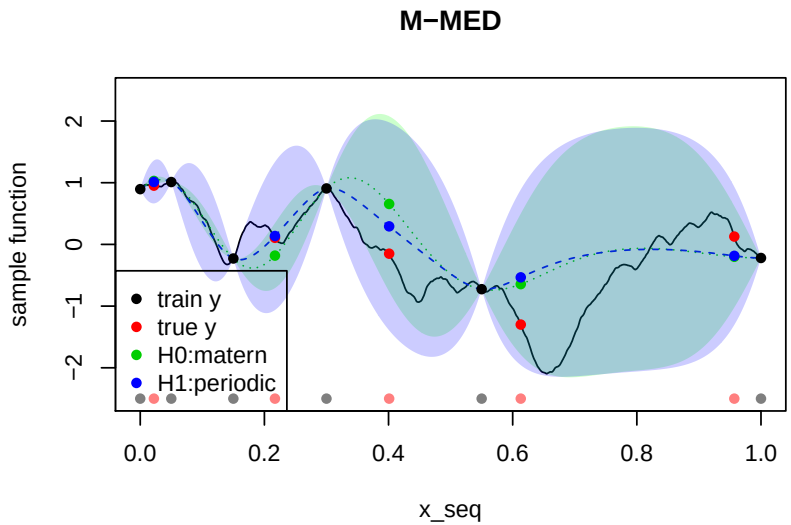
Increasingly farther away input points

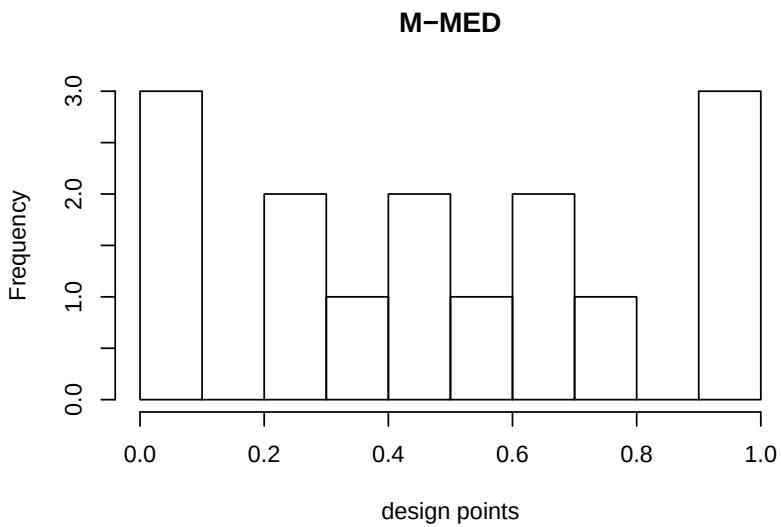
Motivation for increasingly farther away input points

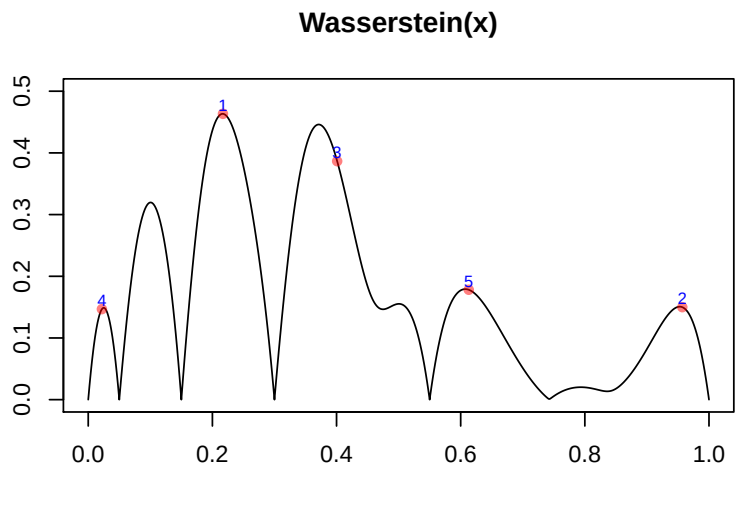
For completion

Gaussian vs. Matern (sequentially): Step 1

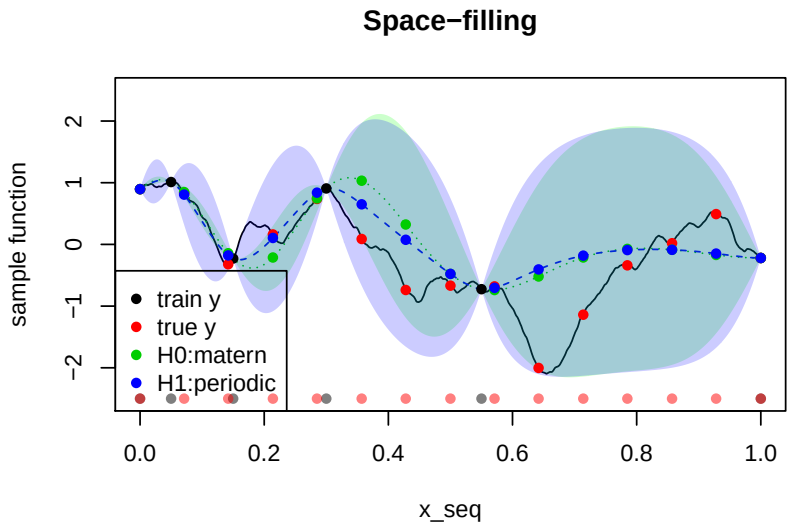


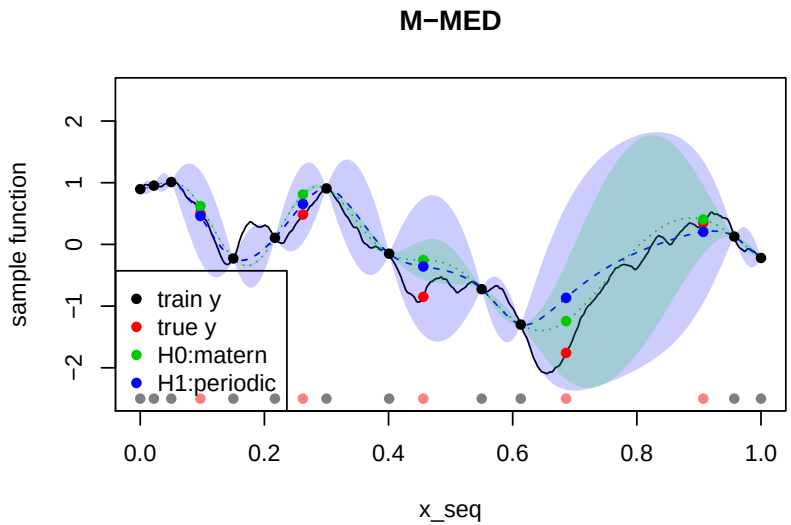


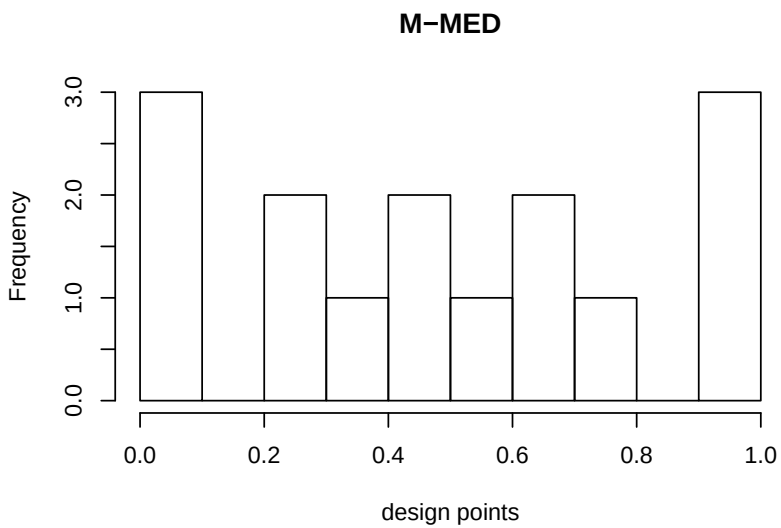


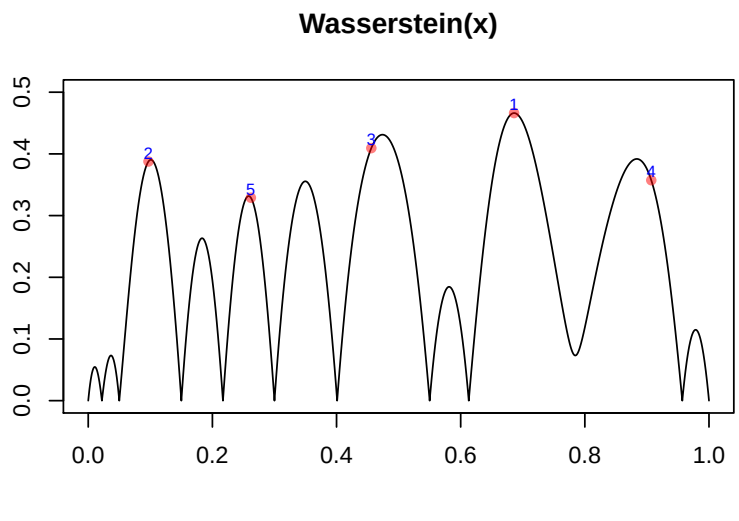


Step 2

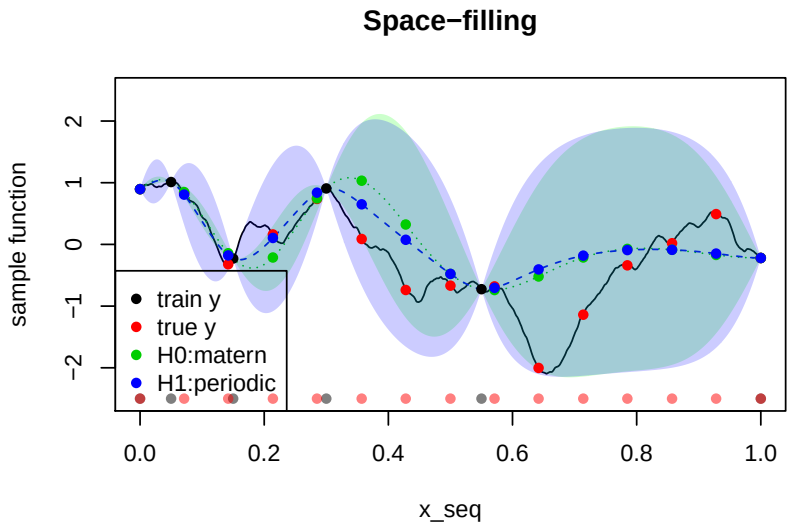

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

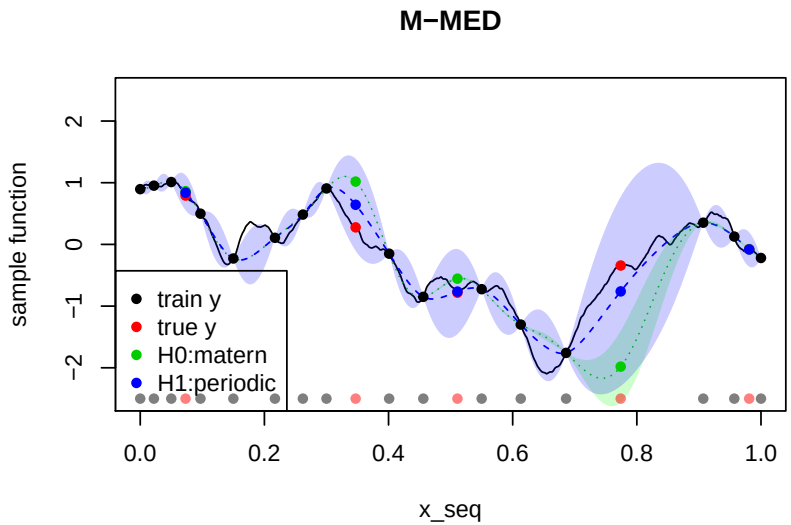

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

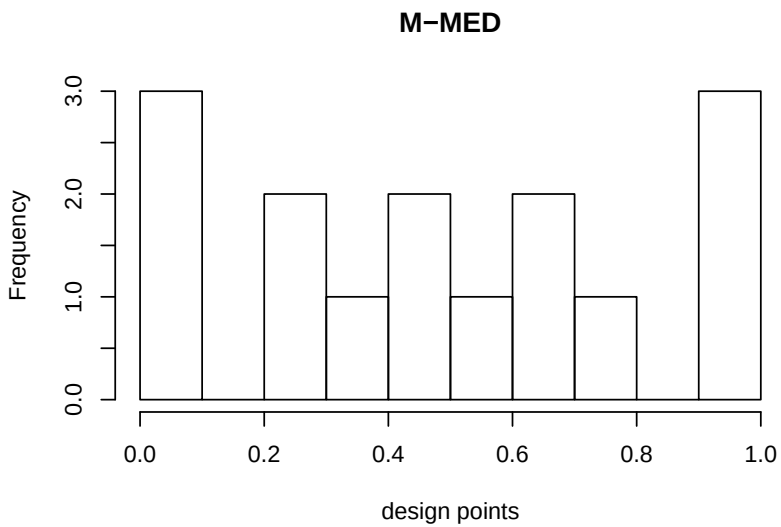

M−MED
Frequency
design points

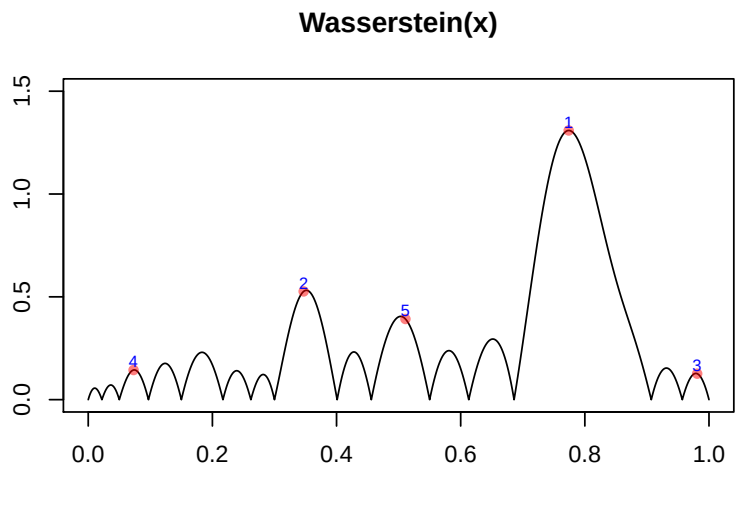

Wasserstein(x)
1
2
3
4
5

Step 3

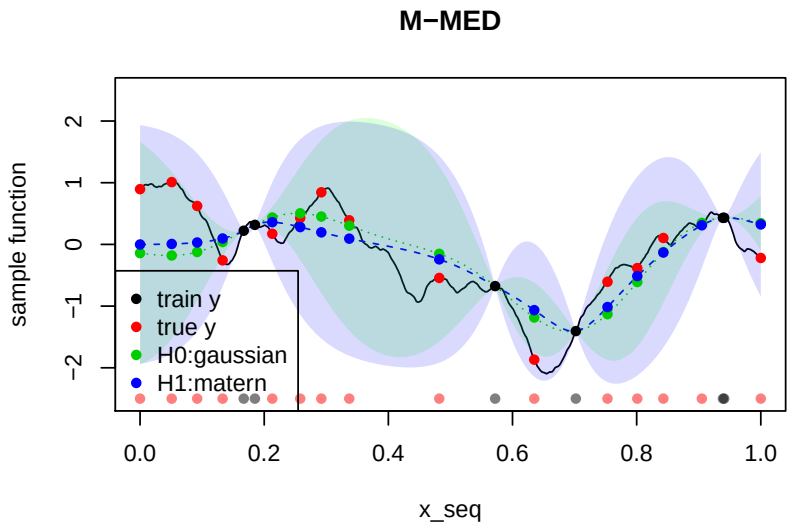

Space−filling
sample function
x_seq
train y
true y
H0:matern
H1:periodic

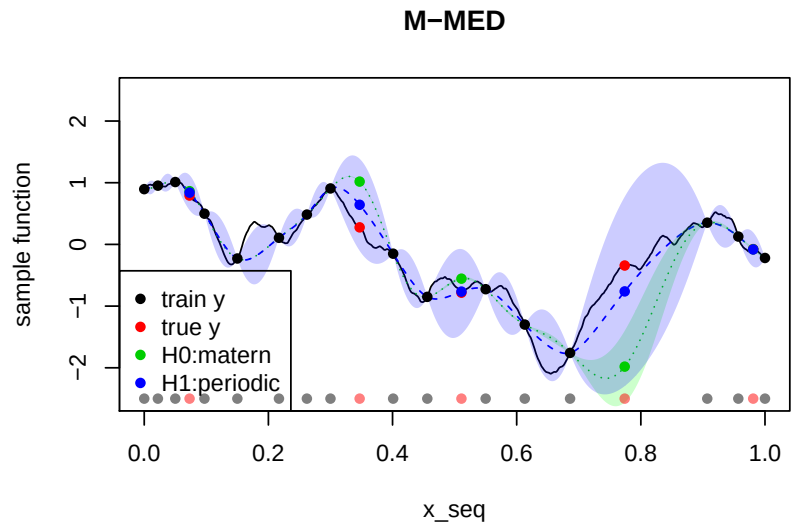

M−MED
sample function
x_seq
train y
true y
H0:matern
H1:periodic

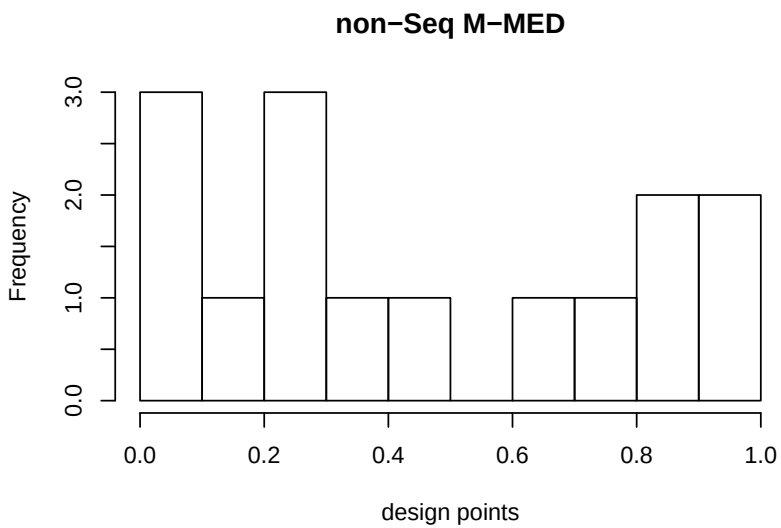

M−MED
Frequency
design points

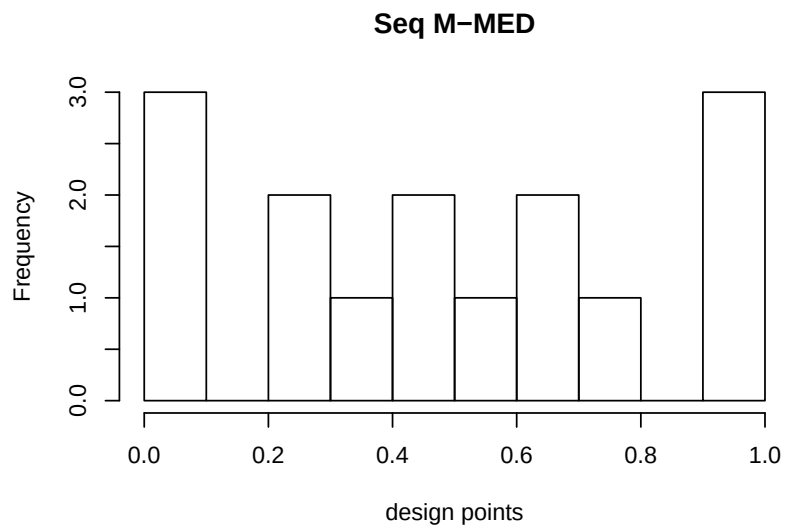

Wasserstein(x)

Compare to non-Sequential M-MED









Gaussian vs. Matern: log(RSS0/RSS1)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.37	-0.024	0.071	0.065	0.17	0.42

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.39	-0.029	0.07	0.066	0.2	0.44

Percentage of simulations that resulted in M-MED evaluations that were larger than Space-filling evaluations

```
## [1] 0.48
```

Gaussian vs. Matern: log ratio of predictive densities

(after removing `NA`s caused from non-invertible matrix)

M-MED

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.5e+06	-3.0e+05	-1.3e+05	-2.8e+05	-4.1e+04	-1.3e+03

Space-filling

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-5.3e+05	-1.2e+05	-6.4e+04	-1.0e+05	-1.5e+04	-9.3e+02

Percentage of simulations that resulted in M-MED evaluations that were smaller than Space-filling evaluations

```
## [1] 0.64
```

Finer Grid

If we choose a finer grid for x_seq