Introduction ooo[o](#bookmark25" \o "Current Document)[o](#bookmark26" \o "Current Document)[oo](#bookmark27" \o "Current Document)oo

The Statistical Approach OOOOOOO O[O](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document) [O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix References

An introduction to Kriging metamodels

Didier Rulliere

Preliminary version, please indicate any typo to [didier.rulliere@emse.fr](mailto:didier.rulliere@emse.fr)

December 2020 - PART I



picture: mining headframe (chevalement) at Saint-Etienne

*Majeure Science des donnees, UP4*

Introduction

O[O](#bookmark15" \o "Current Document)OO[OO](#bookmark26" \o "Current Document)OOO

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO OOOOOOO OOOO

Appendix

References

Acknowledgements

This course is

an overview of Kriging metamodeling and Gaussian Process Regression

This material is partly recycled from previous classes by Nicolas Durrande [2], Roldolphe Le Riche [5], Xavier Bay and many others, thanks a lot!

All errors are mine, do not hesitate to tell me.

Introduction

O[O](#bookmark15" \o "Current Document)OO[OO](#bookmark26" \o "Current Document)OOO

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Outline

Introduction

。The Statistical Approach

。Simple Kriging

。Other Kriging techniques

[3，[The Gaussian Process Approach](#bookmark64" \o "Current Document) [Random and Gaussian Processes](#bookmark68" \o "Current Document) Gaussian process regression Kriging noisy data

[Introduction](#bookmark8" \o "Current Document)

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

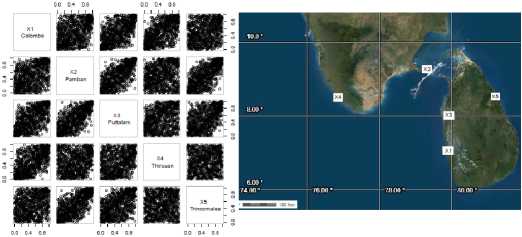
OOOO

Appendix

References

Introduction example : rainfall data

An example of rainfall data in Sri Lanka





How to predict rainfall somewhere, if it is only measured on few specific sites?

Which sites exhibits more correlation ?

Is this in link with spatial distance between sites ?

How would you do to predict between two sites ?

Introduction

[0](#bookmark15" \o "Current Document)\*00[0](#bookmark25" \o "Current Document)[0](#bookmark26" \o "Current Document)[0](#bookmark27" \o "Current Document)00

The Statistical Approach

OOOOOOO

[O](#bookmark51" \o "Current Document)OOO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

OOOOO OOOOOOO OOOO

Appendix

References

The Origins of Kriging

• ... ok about rain but...



A.

B.

C.

1.

2.

3.

• How to predict gold concentration specific sites ?

somewhere, if it is only measured on few

Who is this guy ?

Danie Spline

Danie Krige

Danie Kernel

Where is this mining engineer in the picture ? South Africa

Bermuda

Couriot Mine in Saint-Etienne

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

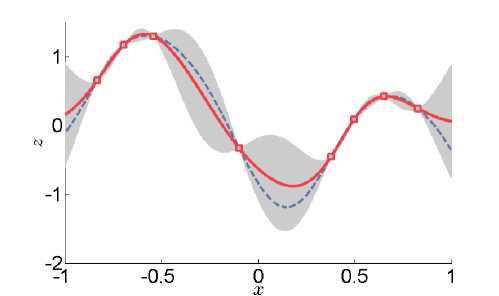
OOOO

Appendix

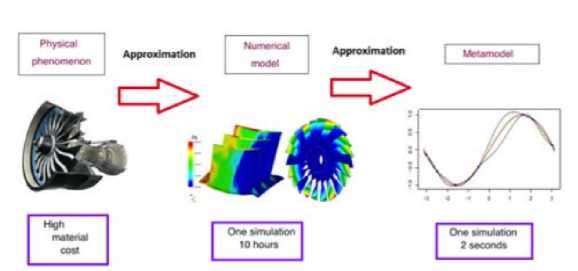
References

Kriging?

*"In statistics, originally in geostatistics, kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariances (...)".* Wikipedia (citation and curve)



Mathematical formalization by Georges Matheron (Ecole des Mines de Paris, student of Paul Levy) in *Memoires du BRGM.*



The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Kriging is most often used in the context of expensive experiments (simulators)

*Illustration from your previous lecture Design of Experiment.*

Many possible applications

Use with computer experiments

Many possible domains

Geostatistic (climate, mining)

Industry (crash tests, computer experiments)

Insurance (mortality tables, Economic Scenario Generator, nested simulations).

Saint-Etienne, December 2020, D [An introduction to Kriging](#bookmark0" \o "Current Document) 7/43

Introduction

O[O](#bookmark15" \o "Current Document)OO[»O](#bookmark26" \o "Current Document)OOO

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Context

Observations

Each experiment can be seen as a function of the input parameters.

input parameters £ *x* ——> (computer/physical/...) experiment ——> output £ R

so that *y = f (x*) where *f* is a **costly to evaluate function**.

In the following, we will assume that

-*x £ x* : There are *d* input variables. Usually (but not necessarily) *x* is R*d*.

-*y £* R : The output is a scalar. But extensions to GP regression with multiple outputs exist.

The interpolation problem

How to predict the output value for some new input parameters ?

Introduction

O[O](#bookmark15" \o "Current Document)OO[O»](#bookmark26" \o "Current Document)OOO

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO OOOOOOO OOOO

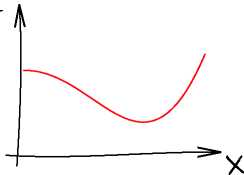
Appendix

References

*f* costly

The fact that *f* is **costly to evaluate** changes a lot of things...

1. Representing the function is not possible...



Introduction

O[O](#bookmark15" \o "Current Document)OO[OO](#bookmark26" \o "Current Document)»OO

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

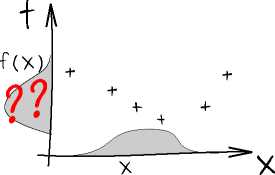
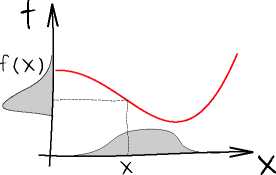
Appendix

References

*f* costly

The fact that *f* is **costly to evaluate** changes a lot of things...

1. Uncertainty propagation is not possible...



The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO OOOOOOO OOOO

Appendix

References

*f* costly

The fact that *f* is **costly to evaluate** changes a lot of things...

1. Optimisation is also tricky...

十 十

十

+

H x

1. Computing integrals is not possible...
2. Sensitivity analysis is not possible...

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Need of a metamodel

Metamodel

Need to replace the costly *f* by a metamodel

* that can give a mean interpolation
* that can also measure the uncertainty associated with this interpolation

The presented one

Here we present Kriging metamodels, also known as Gaussian Process Regression (GPR) under some Gaussian assumptions.

Many other metamodels exist : splines, Inverse Distance Weighting, decomposition in basis functions, etc.

They are sometimes related to Kriging.

The Statistical Approach

Notations

Observations

The Statistical Approach •OOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

set of possible sites : *x* = R*”*

*n* observations sites : X = (*x*i*, xn*) £ *xn*

*n* observed responses : **Y**x = ( *YX*i *,..., Yxn* )T £ R*n* indices : */* = {1*,..., n*}

Quantity of interest

(e.g. rainfall *%* = R2)

(e.g. *n* city locations)

(e.g. annual rainfall quantity)

One new prediction site : *x* £ *x*

Unknown response at this site : *Yx £* R

(e.g. one new city location)

(e.g. rainfall to be predicted)

Assumptions

• all *Yx；* are random variables with finite mean and finite variance

Covariances matrix **K** = *(K；j)；日* and vector **k***x* = *(k；(x))j^/* are known.

where *K；j* and *kj(x)*

*=*Cov *[Yx；, Yxj]*

covariance between responses,

*=* Cov *[Yx；, Yx*] covariance with target *Yx*.

The Statistical Approach

O»OOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Simple Kriging : the model

The idea

Most natural idea : your prediction is a linear combination of observed responses.

The Simple Kriging Model

One assumes **Y**x and *Yx* centered : V*x,* E *[Yx*] = 0. Define a predictor *M(x)* as

*M(x*) = ^2 *ai(x')Yxi*

where weights 必=*(aj(x))i=1..n* are minimizing

△(*x*) = E *[(M(x*) - *Yx*)2]

Check that unbiasedness holds : E *[M(x*)] = E *[Yx*]

The Statistical Approach

OO»OOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Simple Kriging : calculations (1)

Step 1 : Develop △(*x*), express it as a function of covariances **K** and **k***x*

Recall that **k***x* is the covariance vector between *Yx* and the vector **Y**x, and **K** is the covariance matrix of *Y\** Using *M(x*) =***。****[****丫****艾*,let us develop

*△(x) =* E *[(M(x*) 一 *Yx*)2]*.*

*do your calculations here :*

The Statistical Approach

OOO»OOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Simple Kriging : calculations (2)

Step 2 : find the weights ***a****x* that minimize *△(x)*

Now let us minimize on ***a****x*

△(*x*) = ***a***j **K*a****x* — 2***a***J **k***x* + constant

*do your calculations here :*

The Statistical Approach

OOOO»OO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Simple Kriging : Result (1)

Optimal weights

This leads to the vector of weights

***a****x* = **K ' k***x*

where **k***x* is the covariance vector between *Yx* and the vector **Y**x, and **K** is the covariance matrix of **Y**x.

Predictor and variance

From that follows the expression of *M(x*) and A(*x*):

*/ M(x*) = **k**J **K**-1**Y**x

[A(*x*) = *ax -* **k**j**K**-1**k***x*

Notice that A(*x*) does not depend on observed responses **Y**x.

The Statistical Approach

OOOOO»O

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Simple Kriging : Result (1)

Results remain valid for *q* prediction points. Given a specific instance **Y**x = **y**, we get :

Simple Kriging

One assumes that **Y**x and *Yx* Linear Unbiased Predictor of square error A(*x*):

are centered. Kriging mean corresponds to the Best *Yx* given **Y**x = **y**, and Kriging variance to the mean

**k**j**KTy**

*—* **k** J **K** -1 **k***x*

where **K** = Cov [**Y**x*,* **Y**x] is *n* covariance matrix.

*X n* covariance matrix, and **k***x* = Cov [**Y**x*, Y***X**] is a *n x q*

At home, for *x, x*' G *X、*determine A(*x, x*z) = E *[(M(x)* — *Yx)(M(x*z) — \*/)], compare with *c(x, x*') in the next section.

The Statistical Approach

OOOOOO\*

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

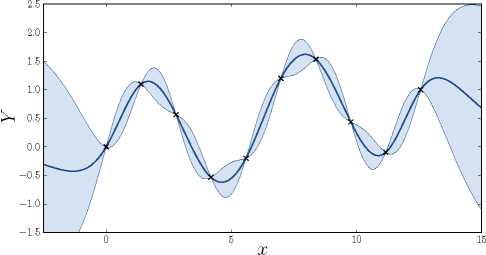
[O](#bookmark68" \o "Current Document)OOOO OOOOOOO OOOO

Appendix

References

Simple Kriging : illustration

It can summarized by a mean function *m(x*) and 95% confidence intervals corresponding to the variance *v*(*x*) (under a distribution assumption).





The kriging predictor is interpolating *m(xj) = Yx；* for all *i*, why?

The Statistical Approach

OOOOOOO

[•O](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Ordinary Kriging (1)

One assumes *Yx；*, *i* £ */* and *Yx* have the same unknown mean *冉* The predictor *M(x*) writes as previously :

*M* (*x*)=

*y^ai(x ^Yx；*

(4)

but unbiasedness condition E *[M(x*)] = E *[Yx；*] implies E*^/ ai(x) =* 1.

Find the weights minimizing A(*x*)=

E*i* 曰 *a*(*x*) =1

E [(*Yx* — *M(x*))2|, subject to

| Introduction | The Statistical Approach | [The Gaussian Process Approach](#bookmark64" \o "Current Document) | Appendix | References |
| --- | --- | --- | --- | --- |
| o[o](#bookmark15" \o "Current Document)oo[oo](#bookmark26" \o "Current Document)ooo | ooooooo | [o](#bookmark68" \o "Current Document)oooo |  |  |
|  | [o«](#bookmark51" \o "Current Document)oo | ooooooo |  |  |
|  |  | oooo |  |  |

Ordinary Kriging (2)

Using a Lagrange multiplier, we minimize in ***a****x*

△ (*x*) - 2*A*(**1**t***q****x* - 1) = ***a***j**K*a****x* - 2***a***J**k***x* + *死-*2*A*(**1**T***a****x* - 1) (5)

after few calculation this gives

Ordinary Kriging

Under the assumption E *[Yxi*]= variance are

E*[Yx*] = *",* for all *i* £ */*, Ordinary Kriging mean and

with ***a****x* = **K** —1

J *m*(*x*) *[v* (*x*)

**k***x* +

***a***j**y**

***a***j **K*a****x* — 2***a***j **k***x* +

**1**t**K**-1 **1**

=*A*

Ordinary Kriging can be seen as a Simple Kriging on residuals, with :

*(" =*(**1**t**K**—1**1**)—1 **1**t*K*—1**Y**x

*\ m(x*) = *" +* **k**J**K**—1(**Y**x - *"***1**)

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)»O

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO OOOOOOO OOOO

Appendix

References

Universal Kriging

Consider given matrices of factors, e.g. *F*(X) = (**1***,*X) and *F*(*x*) = (1*, x*).

The universal Kriging predictor writes

*M(x*) = *F(x*)丁***月*** + *ai(x')Yxi* ⑥

The vector ***月*** does not depend on *x*. One can show *(Sacks et al.,* 1989):

Universal Kriging

The optimal coefficients ***月*** and ***a***(*x*) are the same as those obtained by :

1. doing a linear regression **Y**x = *F*(X)丁***月*** + ***e*** to estimate the *月*广s

***B*** *=(F* (X)t**K**-1*F* (X))—1 *F* (X)t**K**-1**Y**x

1. then doing a Simple Kriging on residuals

... so that no other results are needed O.

What happens when *F*(X) = **1** ?

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)O«

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Advantages of the Statistical Approach

Some advantages of the “statistical approach” (compared to other approaches)

Pro

* General : only requires random variables with two moments, no Gaussian assumption, manipulate only finite vectors
* Can be extended with other regression techniques : penalizations (LASSO, ridge), cross effects, quadratic terms, link functions...

*M(x*) =*a(x ^Yxj - A*

*M(x) = f (Yxi，*…*,Yxn,* ***a***)

• Can be nested using other estimators

*M(x)* =〉: *a(x) Mi (x)*

Cons

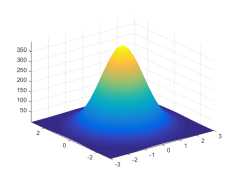
* Interpretation : No direct interpretation as a conditional process
* Theroretical : Conditional quantities sometimes hard to derive

The Gaussian Process Approach

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Introduction  o[o](#bookmark15" \o "Current Document)oo[o](#bookmark25" \o "Current Document)[o](#bookmark26" \o "Current Document)[o](#bookmark27" \o "Current Document)oo | The Statistical Approach ooooooo o[o](#bookmark51" \o "Current Document)oo | [The Gaussian Process Approach](#bookmark64" \o "Current Document) [o](#bookmark68" \o "Current Document)oooo  ooooooo  oooo | Appendix | References |
| Gaussian | Vectors |  |  |  |

A Gaussian Vector **Y** with mean ***卩*** and covariance matrix **£** is a random vector with density

小*，*…*，y*)= exp (-1 (**x** - ***m***)T (**x** - ***“***)) ⑼



* Non-degenerate if **£** definite positive : V**a** non zero, **a**T**£a***〉*0.
* Linear combinations of components of **Y** are Gaussian,
* thus components *Yi* are Gaussian, *i* = 1*,..., d* (reverse not true).

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Conditional Gaussian Vectors

Let ***?*** = [***r*** 1]

be Gaussian with mean ***卩****=*

and covariance **£** =

then

[£11

£i/|

**£**22,

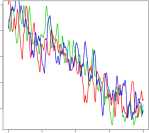
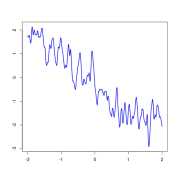
Conditional Gaussian Vector

The conditional distribution of **Y**1 given **Y**2 = **y**2 is Gaussian with mean and covariance

**J *M***2|1 = ***M***1 + **£**12**£**221 (**y**2 —***卩*** *2)*

(10)

**[£**2| 1 = **£**11 — **£**12**£**22X **£**21*-*



Repeat the random event (3x)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Introduction o[o](#bookmark15" \o "Current Document)oo[o](#bookmark25" \o "Current Document)[o](#bookmark26" \o "Current Document)[o](#bookmark27" \o "Current Document)oo | The Statistical Approach OOOOOOO o[o](#bookmark51" \o "Current Document)oo | [The Gaussian Process Approach](#bookmark64" \o "Current Document)  •OOOO  OOOOOOO  oooo | Appendix | References |
| Random | Process |  |  |  |

A random process is a set of RV's indexed by *x* 6 *x*

random event *3* 6 Q

(eg, weather)

This creates 3 trajectories *y*(*x*)'s. They are different, yet bear strong similarities.

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[o](#bookmark68" \o "Current Document)»ooo OOOOOOO OOOO

Appendix

References

Gaussian Process (1) : definition

Gaussian Process : one possible definition



A stochastic process is Gaussian

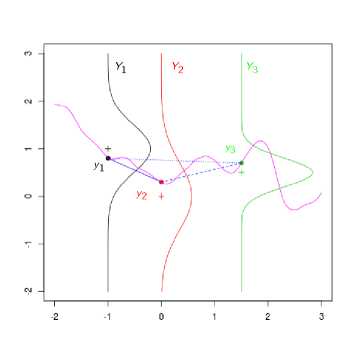
all finite subvectors are Gaussian

* implies that for any *x* £ *x*，*Yx* is a Gaussian RV (reverse not true).
* implies that any finite linear combination of some *Yx*'s is Gaussian.

[An introduction to Kriging](#bookmark0" \o "Current Document)

*y*(*x*)

27 / 43



The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)O»OO

OOOOOOO

OOOO

Appendix

References

Gaussian Process (2) : characterisation

For such a Gaussian Process (GP), we denote

*k(x, x*') = Cov *[Yx, Yxf* ]*.*

Gaussian Process (2) : characterisation

The distribution of a GP is fully characterised by :

* its mean function */ : x —* R :

以 *x* ) = E [\* ]

* its covariance function, or *kernel, k : x X x —* R :

*k(x, x*') = Cov *[Yx, Yx，*]

In particular, VX =

(3

e *xn*, **Y**x =

where **K** =*(庇妇,Kj =*

〜N(以X)*,* **K**),

Cov(&*, Yxj*) = *k(xi, xj*).

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OO»O OOOOOOO OOOO

Appendix

References

Gaussian Process (3) : covariance function

Conditions

... but conditions hold (detailed later) for the covariance function *k.)*!

Should *k(x, x') = k(x', x*) ? why?

For any ***a****,* variance of the random variable ***a***T **Y**x? consequence on *k(.,.)*

One example (for the moment)

The *Gaussian kernel,* or *Squared Exponential (SE) covariance function* :

|  |  |
| --- | --- |
| *k(x, x') = a*2 exp ( | -嘉 ll*x* -*x*』2) |

has two parameters, the variance *a*2 and the lengthscale *0.*

Matrix notations for kernels

for two vectors **u** £ *xq*, **v** £ *xn*, we often use the matrix notation : *k(****u****，***v**) = *(k(ui，uj*))*/*=1*,...,q*；*j*=1*,...,n* £ RX°

* **K** = *k*(X*,*X) *£* R*n*x*n*,
* **k***x* = *k*(X*, x*) = *k(x, X)T £* R*n*x1.

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOO»

OOOOOOO

OOOO

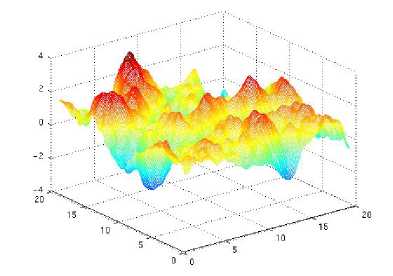
Appendix

References

Gaussian Process (4) : random fields

On previous illustrations *x* £ R, so that trajectories are functions R T R.

When *x* £ R*”*, with *d >* 1, trajectories are functions R*d* T R.



Nothing is changed, but we sometimes call the process a Gaussian Random Field.

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

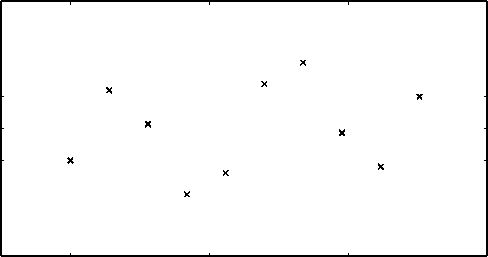
[O](#bookmark68" \o "Current Document)OOOO •OOOOOO OOOO

Appendix

References

Gaussian process regression

Assume we have observed a function *f*() over a set of points *X =* (*x*i*,...,Xn)*:



2.5

0.5

-1.5

2.0

1.5

1.0

0.0

-0.5

-1.0

0

5

10

15

The vector of observations is **y** = *f* (X), i.e. *y = f (xj)*.

The Statistical Approach

OOOOOOO [OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

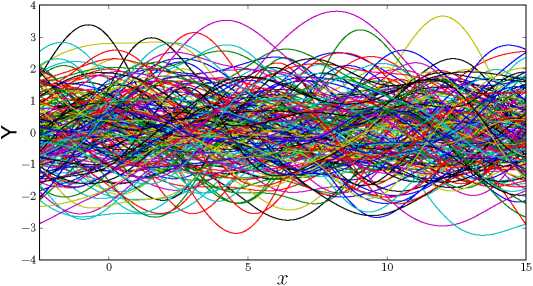
O»OOOOO

OOOO

Appendix

References

Since *f* () in unknown, we make the general assumption that it is the sample path of a Gaussian process *Y N(^(-), k*(*.,.*)):



(here *〃(x) =* 0)

The Statistical Approach

OOOOOOO [OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

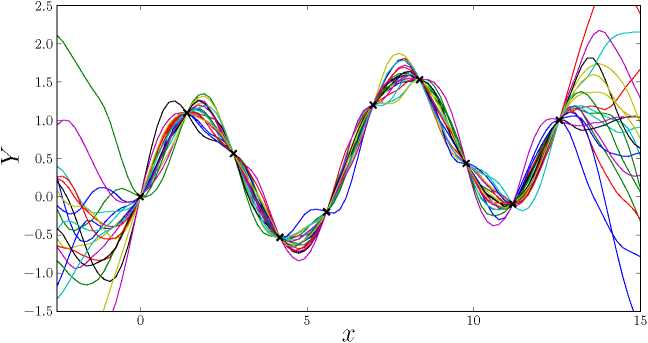
OO»OOOO

OOOO

Appendix

References

If we remove all the samples that do not interpolate the observations we obtain :



The Statistical Approach

OOOOOOO [OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

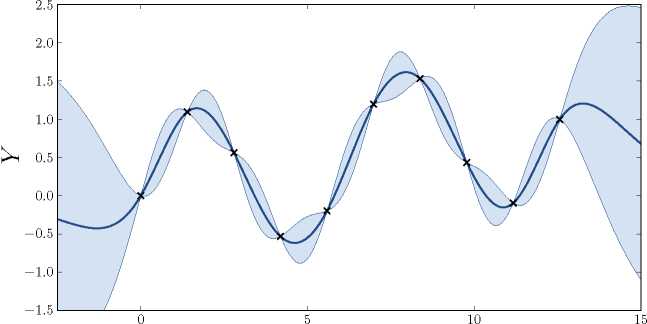
000\*000

OOOO

Appendix

References

It can summarized by a mean function and 95% confidence intervals.



The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOO»OO

OOOO

Appendix

References

Kriging equations (1/2)

The conditional distribution can be obtained analytically :

By definition, *(Yx,* **Y**x) is multivariate normal. Formulas on the conditioning of Gaussian vectors, in Equation [(10)](#bookmark1" \o "Current Document), give the distribution of *Yx*|**Y**x = **y**. It is N(*m*(*.*)*, c*(*.,,*)) with :

*m*(*x*) = E [幺|**Y**x=**y**]

=*片(x)* + *k(x, X)k*(X*,* X)-1(**y** — *〃(X))*

*c (x , x* ') = Cov *[Yx , Yx>* |**Y**x=**y**]

=*k(x, x!*) — *k(x, X)k(X,* X)-1*k*(X*, x*z)

Simple Kriging, Gaussian case

For a centered process, when *"(x) = N*(X) = 0, the simple Kriging predictor in Equation [(3)](#bookmark1" \o "Current Document) corresponds to

J *m*(*x*) = E [\*|**Y**x=**y**]

*[v(x*) = V *[Yx*|**Y**x=**y**]

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOO»O

OOOO

Appendix

References

Kriging equations (2/2)

Summary

The distribution of *Yx*|**Y**x = **y** is *Nc.*)) with mean and covariance

*(m(x) = 片(x*) + *k*(*x, X)k(X,* X)-1(**y** —*卩*(X))

*[c*(*x, X) = k(x, x*z) — *k(x, X)k*(X*,* X)-1*k*(X*, X)*

* *k*(X*,* X) = *[k(xj, Xj*)] : covariance matrix between observed outputs, sometimes named Gram matrix. It is of size *n x n.*
* *k(x, X) = [k(x, Xj),..., k(x, Xn*)] : *n x* 1 covariance vector between observed output and target *Yx*.

Remarks

* It is a Gaussian distribution : gives confidence intervals, can be sampled, this is actually how the previous slides were generated.
* It is Bayesian : *Yx*|**Y**x = **y** is the posterior distribution of *Yx* once **Y**x = **y** is observed.
* It is named *Gaussian Process Regression,* often identified with the term *Kriging.*

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOO\* \*

OOOO

Appendix

References

Properties

A few remarkable properties of Gaussian Process Regression (GPR) models

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

•OOO

Appendix

References

Kriging of noisy data

An important special case, noisy data **Z**x = **Y**x +

where 〜*N*(0*, n(.,.*)) independent of *Y*(*.*). Then,

J Cov (*Yx/* + *eX!, YXj* + *eXj*) = *k (xi, xj*) + *n(xi, xj*)

[Cov(*Yx, YXi* + *eXj) = k(x, xi*)

The expressions of GPR with noise become (just apply Gaussian vector conditioning with the above), when *x 宅* X and *ex*,財,*eX* mutually independent.

*m(x*) = E *[Zx* |**Z**x=**z**]

=*片(x)* + *k(x, X) (k(X, X)* + *n(X,* X))-1 (**z** -*卩(X)) c(x, x*z) = Cov [*Z*(*x*)*, Z(x*z)|*Z*(X)=z]

=*k(x, x*z) — *k(x, X) (k(X, X) + n(X,* X))-1 *k(X, x*z)

Remarks

* this is the same distribution as the one of *Yx*|**Z**x = **z**, *x 宅* X.
* usually *n(X, X)* diagonal, called nugget effect.
* can be used to help the inversion of *k*(*X, X*)

The Statistical Approach

OOOOOOO [OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

O»OO

Appendix

References

Examples of models with observation noise for *n(x, x*z) = *t*2*5X〃* :

(x)a+(x)z")z

8

*疽* I 0

8 7 1。  
HH(x)+(x)z 一z

8 7 I 。 I, HH(x)a+(x)z 一$z

I I I I I I -1 1 1 1 1 1- -1 1 1 1 1 1­

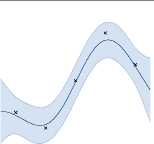
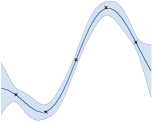
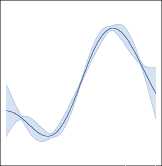
0.0 0.2 0.4 0.6 0.8 1.0 0.0 0.2 0.4 0.6 0.8 1.0 0.0 0.2 0.4 0.6 0.8 1.0

*x x x*

The values of *t*2 are respectively 0.001, 0.01 and 0.1.

Kriging with noise kernel (nugget) does not interpolate the data.

A small *t*2 (e.g., 10-10) often used to make the covariance matrix invertible (more on regularization of GPs in [6]).



The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OO»O

Appendix

References

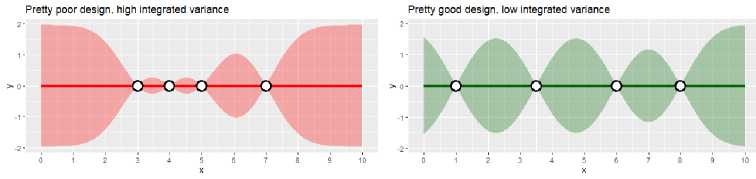
Kriging based Design

Write here *k*(X*,* X) = Cov [**Y**x*,* **Y**x] = **K** and *k*(X*, x*) = Cov [**Y**x*, Yx*] = *k*(*x,* X)丁 = **k***x*.

Lower prediction variance is better : optimize X in order to minimize the sum of prediction variances over *x* for a given measure *冉*

*IMSE*(X) *= I v*(*x)d^(x) = I* 侦2 — *k(x, X)k*(X*,* X)-1*k*(*x,* X)) *d卩(x)*

*x x*



Other criterions : maximizing entropy or Mutual Information *Krause et al.,* 2008.

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOO»

Appendix

References

to be continued...

in the rest of the lecture, we will detail more results on

* Kernels, covariance functions.
* (Hyper)-parameters estimation.

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Some references I

1. Abrahamsen, P. (1997). A review of gaussian random fields and correlation functions.
2. Durrande, N. and Le Riche, R. (2017). Introduction to Gaussian Process Surrogate Models. Lecture at 4th MDIS form@ter workshop, Clermont-Fd, France. HAL report cel-01618068.
3. Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data. *arXiv preprint arXiv :1309.6835.*
4. Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in gaussian processes : Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research,* 9(Feb) :235—284.
5. Le Riche, R. (2014). Introduction to Kriging. Lecture at mnmuq2014 summer school, Porquerolles, France. HAL report cel-01081304.
6. Le Riche, R., Mohammadi, H., Durrande, N., Touboul, E., and Bay, X. (2017). A Comparison of Regularization Methods for Gaussian Processes. slides of talk at siam conference on optimization op17 and accompanying technical report hal-01264192, Vancouver, BC, Canada.

<https://www.emse.fr/~leriche/op17_R_LeRiche_slides_v2.pdf>and [https://hal.archives-ouvertes.fr/hal-01264192.](https://hal.archives-ouvertes.fr/hal-01264192)

The Statistical Approach

OOOOOOO

[OO](#bookmark51" \o "Current Document)OO

[The Gaussian Process Approach](#bookmark64" \o "Current Document)

[O](#bookmark68" \o "Current Document)OOOO

OOOOOOO

OOOO

Appendix

References

Some references II

1. Lopez-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018). Finite-dimensional gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification,* 6(3) :1224—1255.
2. Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer School on Machine Learning,* pages 63—71. Springer.
3. Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim : Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1).
4. Rulliere, D., Durrande, N., Bachoc, F., and Chevalier, C. (2018). Nested kriging predictions for datasets with a large number of observations. *Statistics and Computing,* 28(4) :849—867.
5. Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science,* pages 409—423.

* They (can) interpolate the data-points
* The prediction variance does not depend on the observations **y**
* The mean predictor does not depend on the scale of variances parameters

• They (usually) come back to the a priori trend ) when we are far away from the observations.

proofs left as exercise