Sequential Designs for Computer Experiments

Stat 890-4 – SFU
Stat 547L – UBC
Math 4013 A1 / 5843 A1 – Acadia

(Fall 2014)

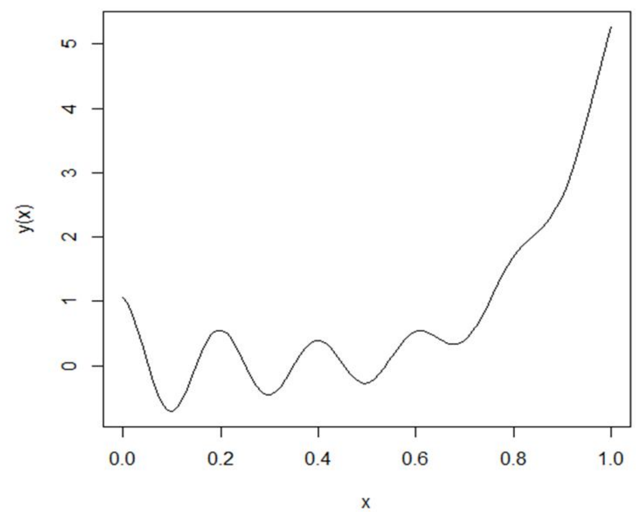
Pritam Ranjan – Acadia University

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Why do we need it?

- Suppose we wish to minimize the outputs of a deterministic computer simulator.

$$f(x) = \sin(2\pi x) + (x-1)^4$$
for $x \in (0.5, 2.5)$



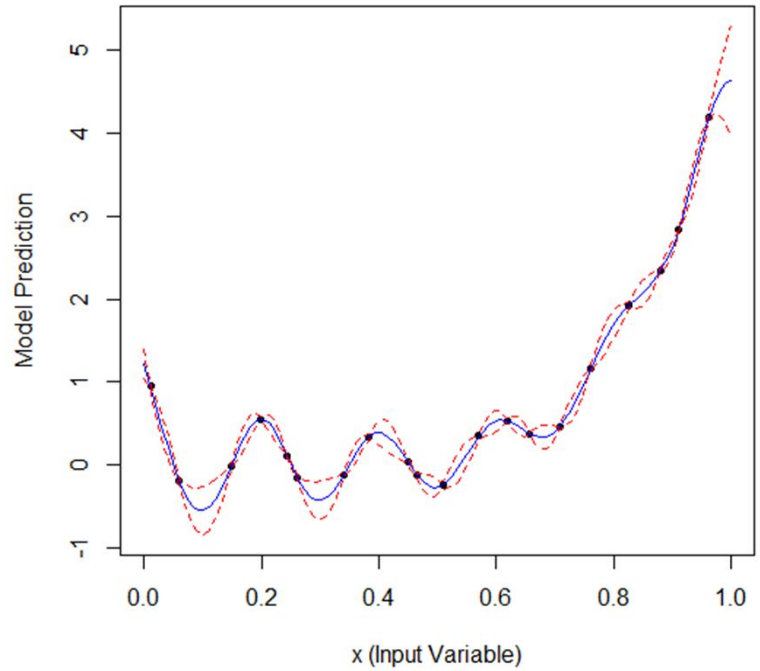
Case I: Evaluation of $f(x)$ is inexpensive

Method 1: use gradient based approach

Method 2: use stochastic algorithm (genetic algorithm, simulated annealing, particle swarm optimization, etc)

Why do we need it?

- **Case II**: Evaluation of $f(x)$ is <u>expensive</u>
  - budget is fixed (say) $N = 20$



Naïve approach:

- Use a 20-point maximinLHD
- Fit a GP model $\hat{f}(x)$
- Estimate the minimum using $\hat{f}(x)$

Is this a good method?

- No – why? We are wasting resources in uninteresting region

Possible alternative?

Sequential Designs

or

Adaptive Designs

Sequential Designs

- Particularly useful when the objective is to estimate a **pre-specified** process feature
  - Global minimum, maximum, local optima
  - Change points
  - Contours, percentiles, confidence intervals
  - Probability of failure in reliability
  - Overall surface

What is a sequential design?

Design scheme

1) Choose $n_0 (< N)$ points. Set $n = n_0$.

2) Fit a statistical surrogate model using

$$\{(x_i, y(x_i)), i = 1, \dots, n\}.$$

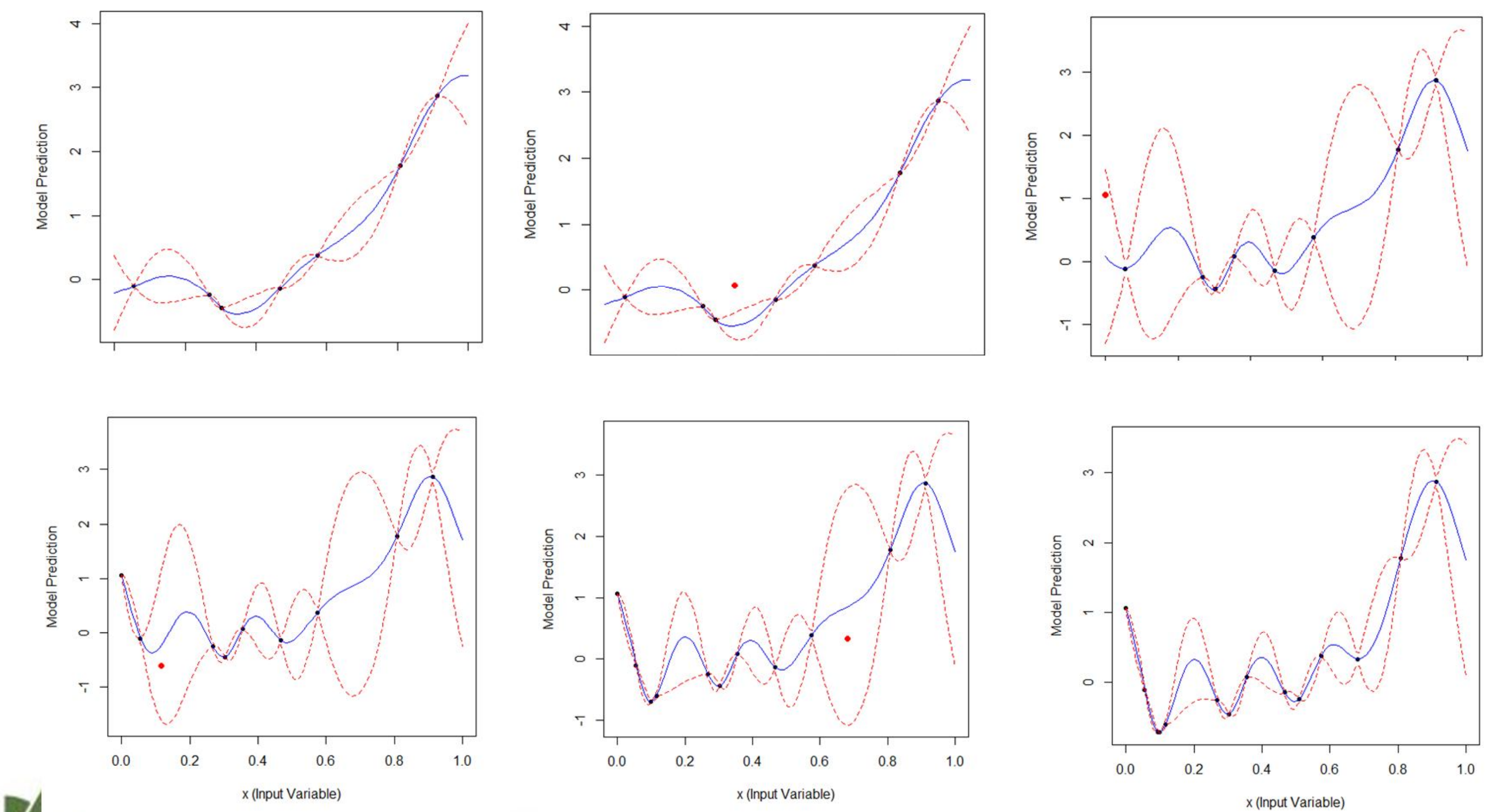
1) Choose a new trial x_{new}.

2) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.

3) Go to Step 2 if $n < N$.

Illustration

- Started with $n_0 = 7$ points & added 13 new points



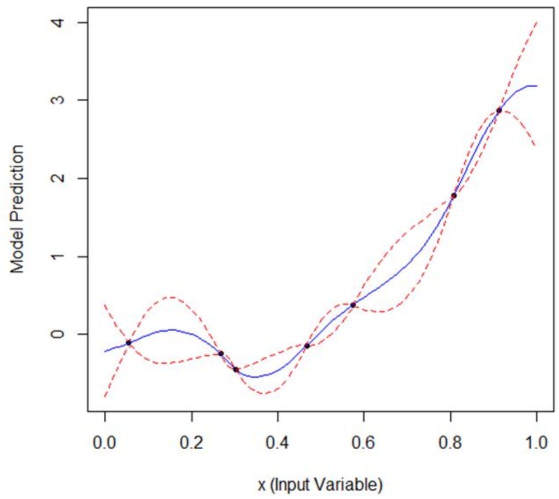
Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

CANSSI INCASS

ACADIA UNIVERSITY

Details of Design Scheme – 1

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.

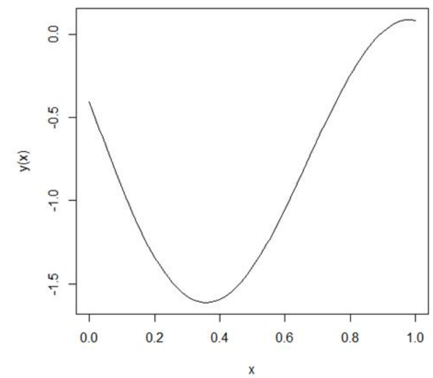


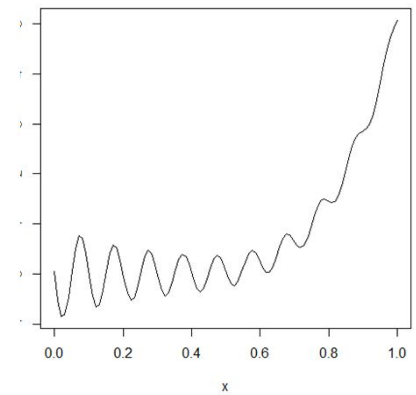
<u>Important issues:</u>

- How do we choose n_0 points?
  - Objective: understanding of overall surface
  - Popular choices: Space-filling designs
    - Distance based (maximin, uniform, etc.)
    - Space-filling LHDs
    - I-optimal, D-optimal designs

Details of Design Scheme – 1

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.



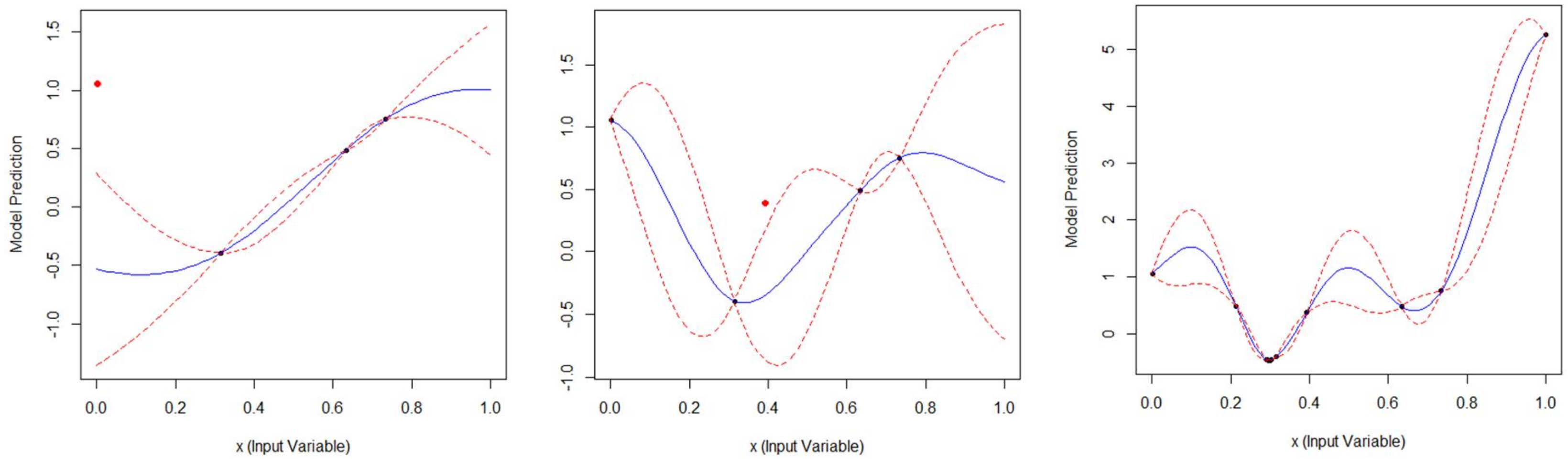


Important issues:

- What is the right choice of n_0?
  - My experience – depends on the complexity of f.
  - Even for $d = 1$, sometimes $n_0 = 5$ is enough, whereas, in some cases 15 points are not sufficient for n_0.
  - A few suggestion: $n_0 = 10d$ or $n_0 \approx N/3$ or $n_0 \approx N/4$.
  - n_0 should NOT be too small or too big

Details of Design Scheme – 1

- What is the right choice of n_0?
- Case 1: $n_0 = 3, N = 20$

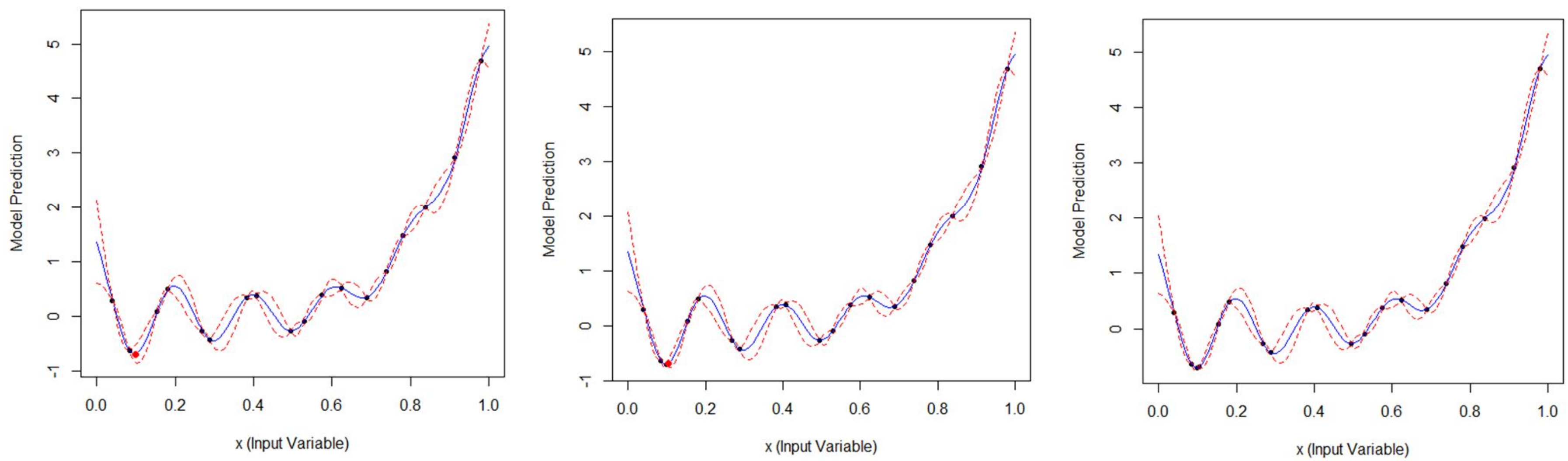


k=1 k=2 after k=17

➢ You get stuck in local optima. So, n_0 too small is not a good idea.

Details of Design Scheme – 1

- What is the right choice of n_0?
- Case 2: $n_0 = 18, N = 20$



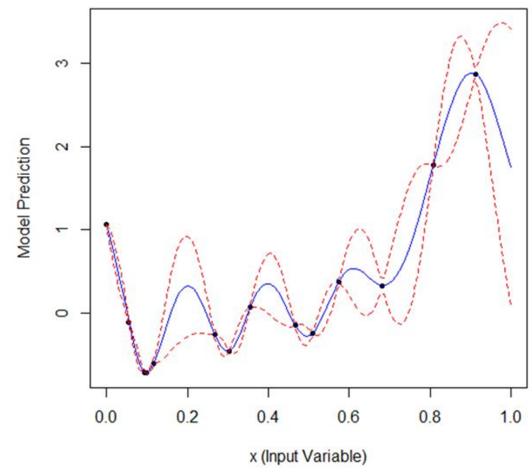
k=1 k=2 after k=2

➢ You still need to improve. So, n_0 too large is also a waste of resources.

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Details of Design Scheme – 2

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.

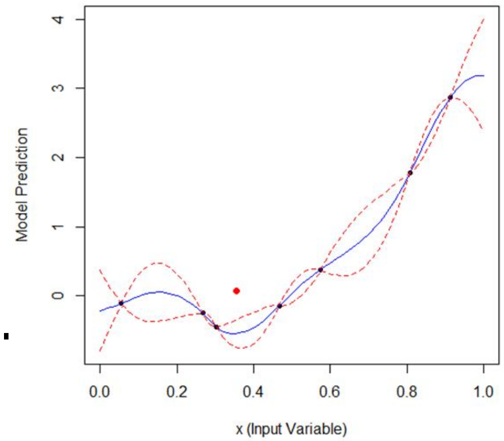


- Important issues:
  - Choice of surrogate model
    - Deterministic stationary process: $y(x) = \mu + Z(x)$
    - Noisy stationary process: $y(x) = \mu + Z(x) + \varepsilon$
    - Non-stationary process : TGP / BART / etc.

- The sequential design scheme is not restricted to only GP model

Details of Design Scheme – 3

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.



- <u>Important questions:</u>
  - How do we choose the new trial locations?
  - Do we have to choose only one trial at-a-time?
    - Complete sequential vs batch sequential

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

How do we choose a new trial?

1) Randomly – easy, but perhaps not very efficient

2) Based on a specific criterion
   - Popular choice – Expected Improvement (EI)
     - Easy to develop
     - Depends on the overall objective (overall surface fit, process optimization, estimating contours, percentiles, probability of failure, etc.)
     - See Bingham, Ranjan and Welch (2014) for a review.

- Is this the only criterion?
  - There are plenty more that can be used, but, the EI-class is huge.

Expected Improvement

- $EI(x)$ is defined over the entire input space $x \in [0,1]^d$
- The choice of $(n+1)$-th follow-up trial location is

$$x_{n+1} = \underset{x \in [0,1]^d}{\operatorname{argmax}} EI(x)$$

- Ideally, $EI(x)$ is the expectation of $I(x)$ over the predictive distribution

$$E\{I(x)\} = \int I(x) f(y|x) dy$$

  – i.e., $EI(x) = E\{I(x)\}$
  – In GP model, $y(x) \sim N\big(\hat{y}(x), s^2(x)\big)$.

- Improvement = negative loss (as in risk = expected loss)

$$I(x) = h\big(x, ; \hat{y}_{(n)}; \psi_n(y)\big)$$

$\psi_n(y)$ represents the feature of interest (e.g., min, max, contour, etc.)

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Expected Improvement

- In most cases, an EI – criterion is
  - Easy to construct (Is it a good news?)
  - It is a function of both
    - $\psi_n(y)$: the feature of interest
    - the prediction uncertainty introduced via $\int * f(y|x)dy$

Expected Improvement

- In most cases, an EI – criterion is
  - Easy to construct (Is it a good news?)
  - It is a function of both
    - $\psi_n(y)$: the feature of interest
    - the prediction uncertainty introduced via $\int * f(y|x)dy$

- Example: interested in global minimum (Jones, Schonlau and Welch 1998)
  - Deterministic stationary process
  - GP model

$$I(x) = \max\left\{y_{min}^{(n)} - y(x), 0\right\}$$

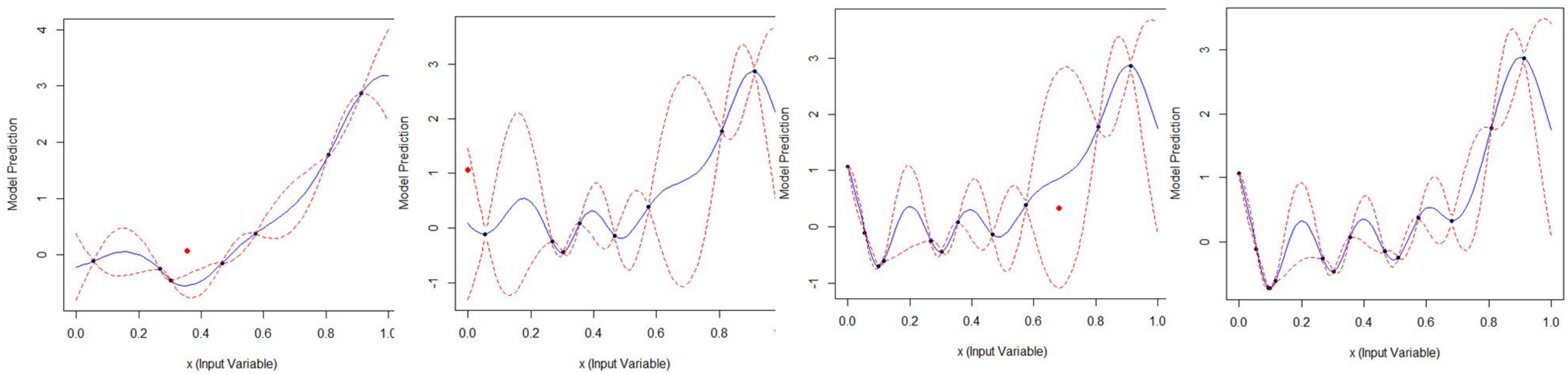
$$E\{I(x)\} = s(x)\phi(u) + \left\{y_{min}^{(n)} - \hat{y}(x)\right\}\Phi(u), \quad \text{where } u = \left\{y_{min}^{(n)} - \hat{y}(x)\right\}/s(x)$$

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

EI – Illustration (Jones et al.)

- Started with $n_0 = 7$ points & added 13 new points



- $E\{I(x)\} = s(x)\phi(u)$ (supports global search – exploration)

  $+ \left\{y_{min}^{(n)} - \hat{y}(x)\right\}\Phi(u)$ (encourages local search – exploitation)

  – Facilitates a balance between global and local search

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

EI - construction

- Easy to construct – a few examples for **process minimization**:
- Schonlau, Welch and Jones (1998) – for deterministic stationary process

$$I(x) = \max\left\{(y_{min}^{(n)} - y(x))^g, 0\right\} \text{ for } g = 1, 2, \dots$$

- Sobester, Leary and Keane (2005) – for deterministic stationary process

$$E\{I(x)\} = w * s(x)\phi(u) + (1 - w) * \left\{y_{min}^{(n)} - \hat{y}(x)\right\}\Phi(u)$$

- Ranjan (2013) – for noisy stationary process (GP-based model)

$$I(x) = \max\left\{ (q_{min}^{(n)} - Q(x))^g, 0\right\} \text{ for } g = 1, 2, \dots$$

Where $Q(x) = y(x) - 1.96 * s(x)$, and $q_{min}^{(n)} = \min\{\hat{Q}(x_i), i = 1, \dots, n\}$

- Chipman, Ranjan and Wang (2012) – for deterministic non-stationary process (BART)

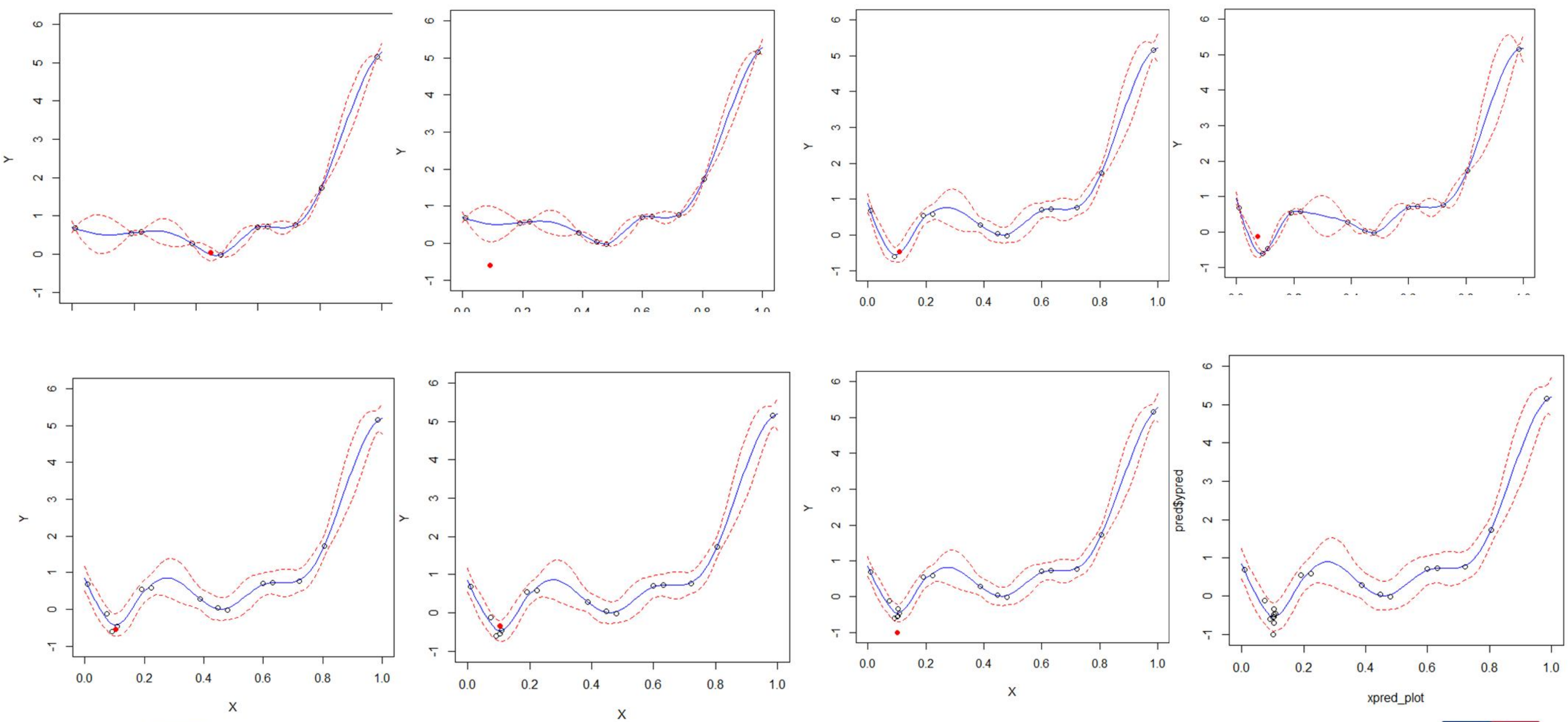
$$I(x) = \max\left\{(y_{min}^{(n)} - y(x))^g, 0\right\} \text{ for } g = 1, 2, \dots$$

(the expectation was taken over posterior realizations)

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

EI – Illustration (noisy)

- Ranjan (2013) – for noisy stationary process (GP-based model, $g = 1$)

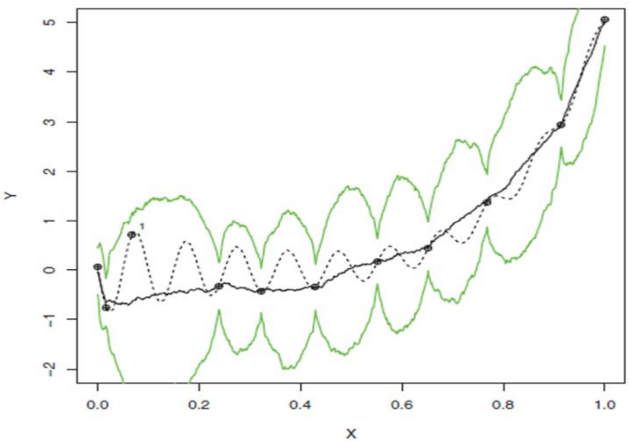
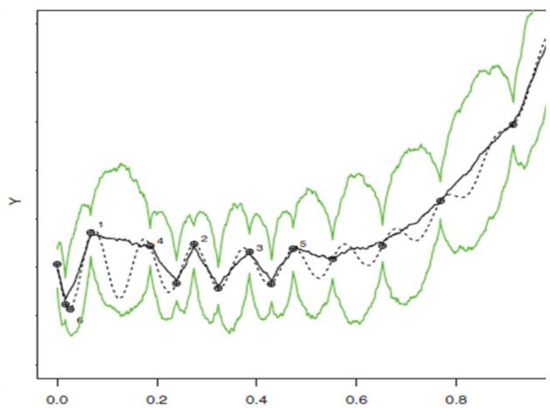
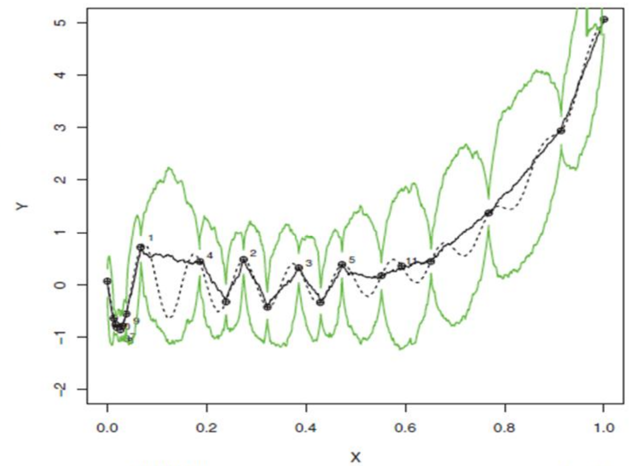


Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

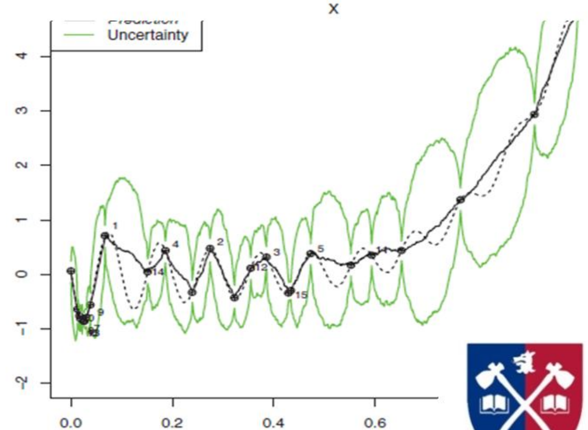
ACADIA
UNIVERSITY

EI – Illustration (non-stationary)

- Chipman, Ranjan and Wang (2012) – for deterministic non-stationary process using BART ($n_0 = 10, N = 25$)

We used $I(x) = \max\left\{ (y_{min}^{(n)} - y(x))^g, 0\right\}$ for $g = 1$



Perhaps, $g \geq 2$ would be better

- .

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

EI - construction

- Easy to construct – a few more examples for pre-specified features
- Ranjan, Bingham and Michailidis (2008) – for **contour** estimation
  $I(x) = \epsilon^2 - \min\{|y(x) - a|^2, \epsilon^2\}$, where $\epsilon(x) = 1.96 * s(x)$
- Roy and Notz (2013) – for **percentile** estimation
  $I(x) = \epsilon^g - \min\left\{|y(x) - \hat{v}_p|^g, \epsilon^g\right\}$, where $g = 1,2, \ldots,$ and $a = \hat{v}_p$.
- Bichon et al. (2008) – for estimating **probability of failure**
  $I(x) = \epsilon - \min\{|y(x) - a|, \epsilon\}$
- Bingham, Ranjan and welch (2013) – for **multiple contours** estimation
  $I(x) = \epsilon^2 - \min\{|y(x) - a_1|^2, |y(x) - a_2|^2, \ldots, |y(x) - a_m|^2, \epsilon^2\}$

.

EI – contour

- Ranjan, Bingham and Michailidis (2008) – for **contour** estimation

  $I(x) = \epsilon^2 - \min\{|y(x) - a|^2, \epsilon^2\}$, where $\epsilon(x) = 1.96 * s(x)$

  Expected improvement

  $$E\{I(x)\} = \int_{a-\epsilon}^{a+\epsilon} [\epsilon^2 - |y - a|^2] f(y|x) dy$$

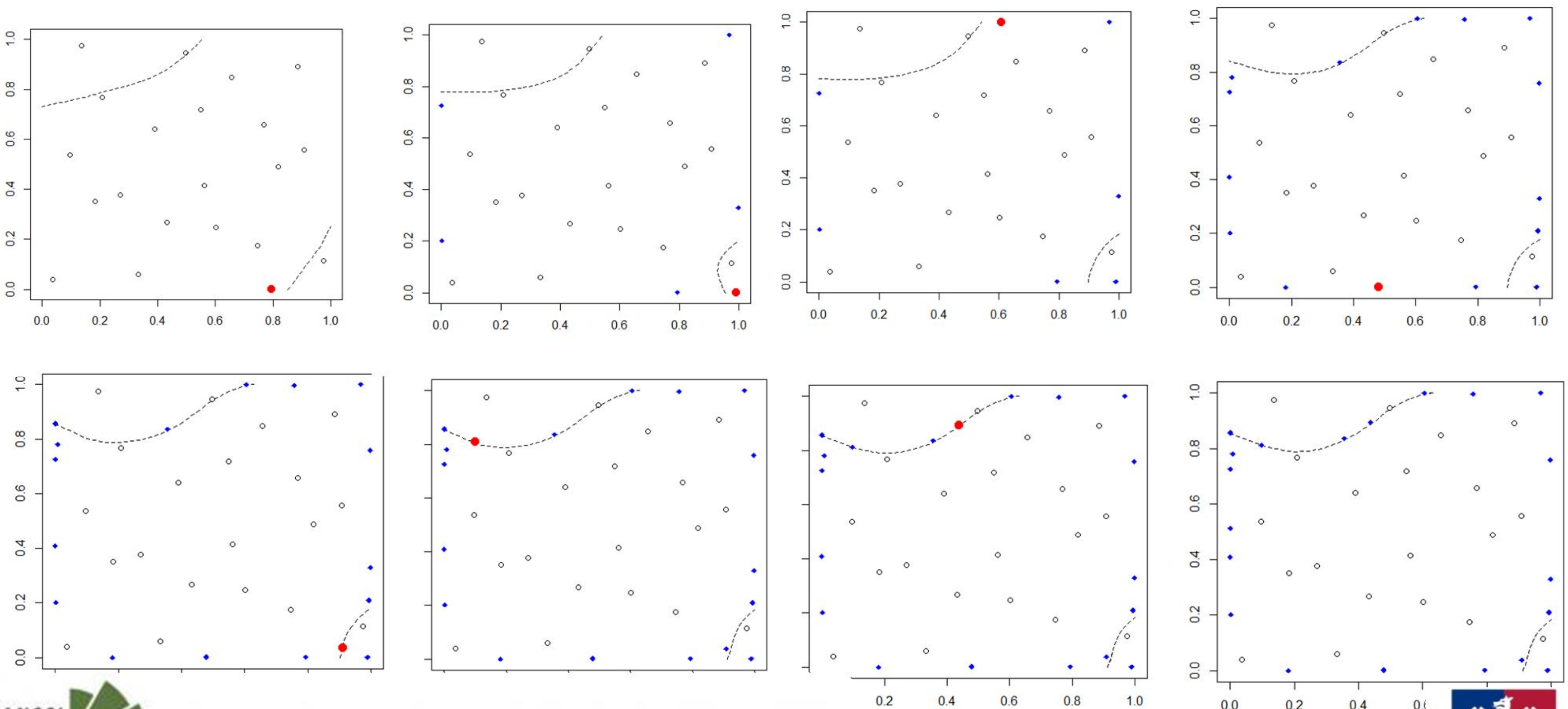
  Fortunately, we have closed form expression

  $$E\{I(x)\} = [\epsilon^2 - (\hat{y}(x) - a)^2 - s^2(x)]\big(\Phi(u_2) - \Phi(u_1)\big) + s^2(x)(u_2\phi(u_2) - u_1\phi(u_1)) + 2(\hat{y}(x) - a)s(x)(\phi(u_2) - \phi(u_1))$$

As before, the expectation over the prediction distribution facilitate a balance between global vs. local search.

EI – contour – illustration

- Ranjan, Bingham and Michailidis (2008) – for **contour** estimation
  (n_0 = 20 and N = 40)



EI - construction

- There are numerous variations/extensions of Jones – EI
  - Ginsbourger, Helbert and Carraro (2008) – Weighted EI for optimization
  - Benassi, Bect and Vazquez (2011) – Student EI
  - Kleijnen, van Beers and Nieuwenhuyse (2012) – Bootstrap EI
  - HenkenJohann and Kunert (2007) – optimization for multivariate response
  - Huang et al. (2006) – optimization for multi-fidelity process

- IMSE, maximum MSE, average MSE criteria can also be viewed as EI for appropriately defined Improvement function.

EI - construction

- Lam and Notz (2008) proposed EI for overall good fit

$$I(x) = \{y(x) - y_{(n)}(x)\}^2$$

where $y_{(n)}(x) = y_{i^*}$ such that, $i^* = argmin\,\{||x - x_i||, i = 1, \ldots, n\}$

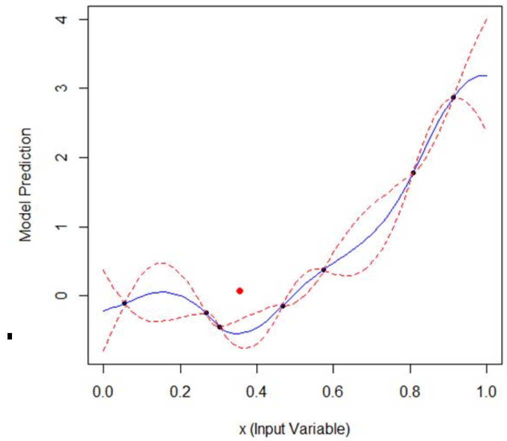
$$E\{I(x)\} = \{\hat{y}(x) - y_{(n)}(x)\}^2 + var(\hat{y}(x))$$

  - Compared the performance with IMSE, max MSE, etc.

- **Summary**:
  - Construction of EI is not difficult
  - all you need is a loss function.

Details of Design Scheme – 3

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.



- Important questions:
  - How do we choose the new trial locations?
  - Do we have to choose only one trial at-a-time?
    - Complete sequential vs batch sequential

Complete vs. Batch sequential

Batch Sequential - m follow-up trials at-a-time

- Why would someone want that?

- How is it possible?
  - Do we need to develop new EI criteria? Or modify the old ones?
  - Does the methodology depend on the feature of interest?

Batch sequential – EI

- Schonlau, Welch and Jones (1998) proposed Generalized Expected improvement

$$I_{MS}^{g}(x_{n+1}, \dots, x_{n+m}) = \left[\max\left\{y_{min}^{(n)} - y_{n+1}, \dots, y_{min}^{(n)} - y_{n+m}, 0\right\}\right]^{g}$$

- All EI criteria can be modified to choose a batch of m trials in $\chi = [0,1]^d$
  - (*Integrated Expected Improvement*)

$$X^{new} = \underset{X_c \in \chi^m}{\text{argmin}} \int_{x^* \in \chi} E\{I_{(n)}(x) | X_{(n)}, Y_{(n)}, X_c, \hat{Y}_c\} f(x^*) dx^*$$

Where X_c is the set of m candidate trials in χ^d and $\hat{Y}_c$ is the prediction based on n –point fit.

  - **Q: Why minimize it? Why not maximize it like EI?**

  - **Q: Can we avoid $m * d$ – dimensional optimization?**

Details of Design Scheme – 5

1) Choose $n_0 (< N)$ points. Set $n = n_0$.
2) Fit a statistical surrogate model using $\{(x_i, y(x_i)), i = 1, \ldots, n\}$.
3) Choose a new trial x_{new}.
4) Update the data: $x_{n+1} = x_{new}, y_{n+1} = f(x_{n+1})$.
5) Go to Step 2 if $n < N$.

- Important questions:
  - Do we proceed all the way up to N or stop before N?
    - How should we build stopping criteria?

.

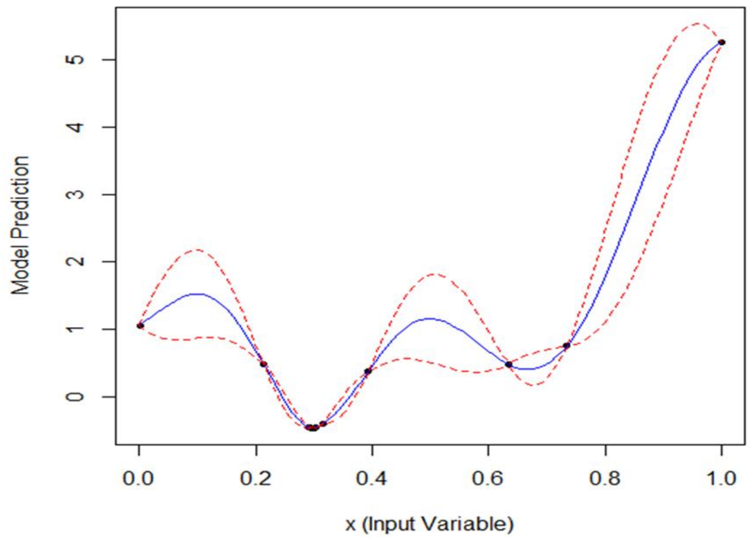
CANSSI INCASS

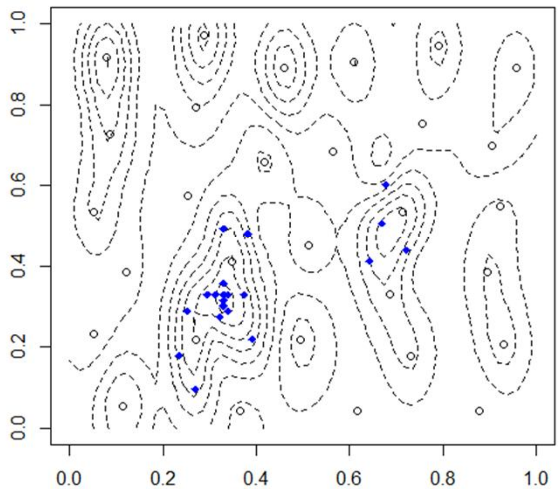
Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Potential project – 1

- Computational advantage in refitting (already have a good guess of θ) ??
- Ill-conditioning may arise if follow-up points start to pile-up (particularly in GP model without error term)



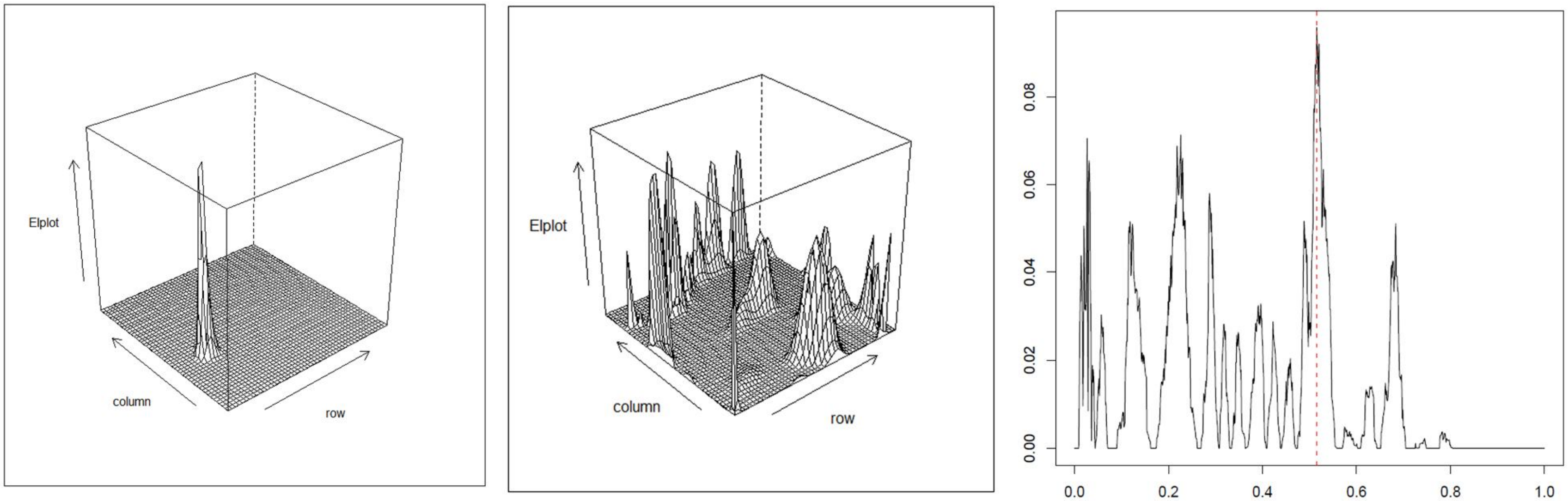


- **What can we do?**

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Potential project – 2

- EI optimization is often tricky (spiky, zeros)



- Any efficient way to optimize this ? Good news: EI- evaluation is cheap.
- Is it really important to find the global optimum of EI?

CANSSI INCASS
Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Potential project – 3, 4, ...

- Needs attention: EI criteria for
  - multiple contours
  - change points
  - local optima

- Can we develop a concept of optimal formulation for EI ?

- Integrated EI for batch sequential designs.

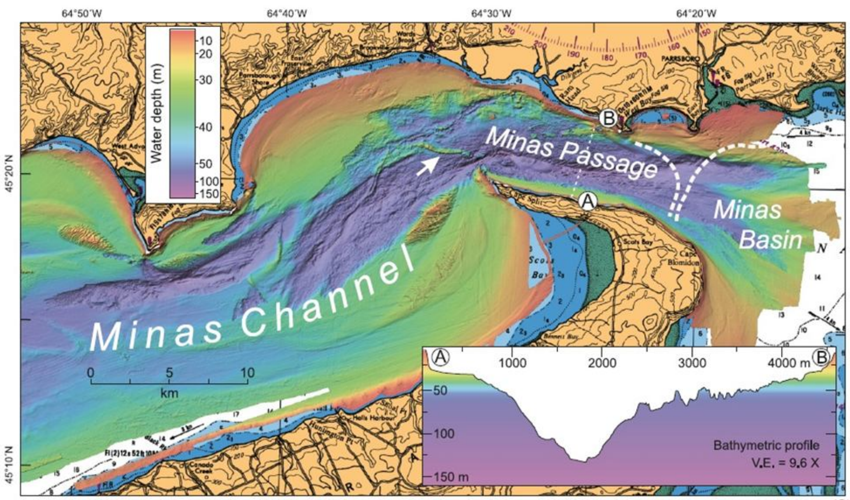
- EI criteria under noisy processes and/or non-GP processes

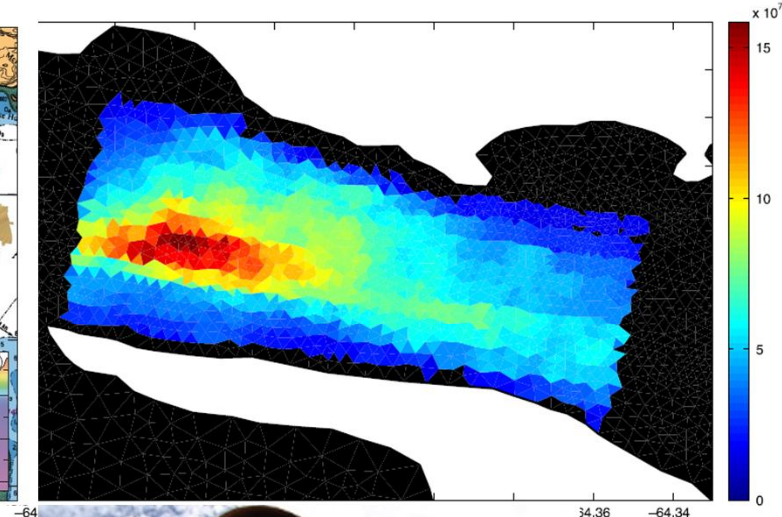
Real Application – 1

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Tidal power simulator – 1

- Objective: maximize the power function for installing turbine





CANSSI INCASS

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

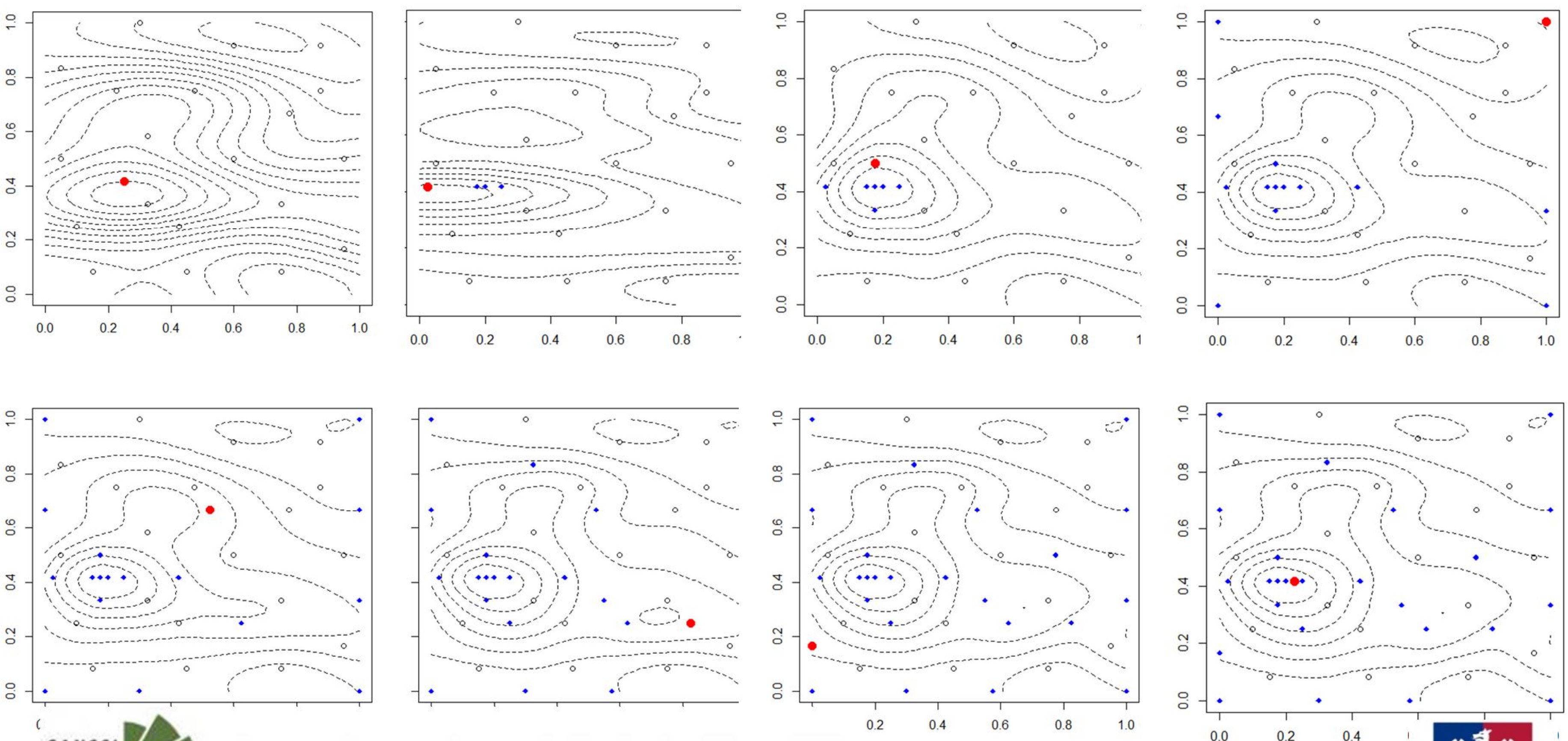
Tidal power simulator – 1

- Objective: maximize the power function for installing turbine
  - Simulator with $200m$ resolution
  - runs available only on 13×41 grid points
  - Q: How do we choose n_0 points?
    - MaximinLHS?

- Objective: maximize the power function for installing turbine
  - Simulator with $200m$ resolution
  - runs available only on 13×41 grid points
  - Q: How do we choose n_0 points?
    - MaximinLHS?

Tidal power simulator – 1

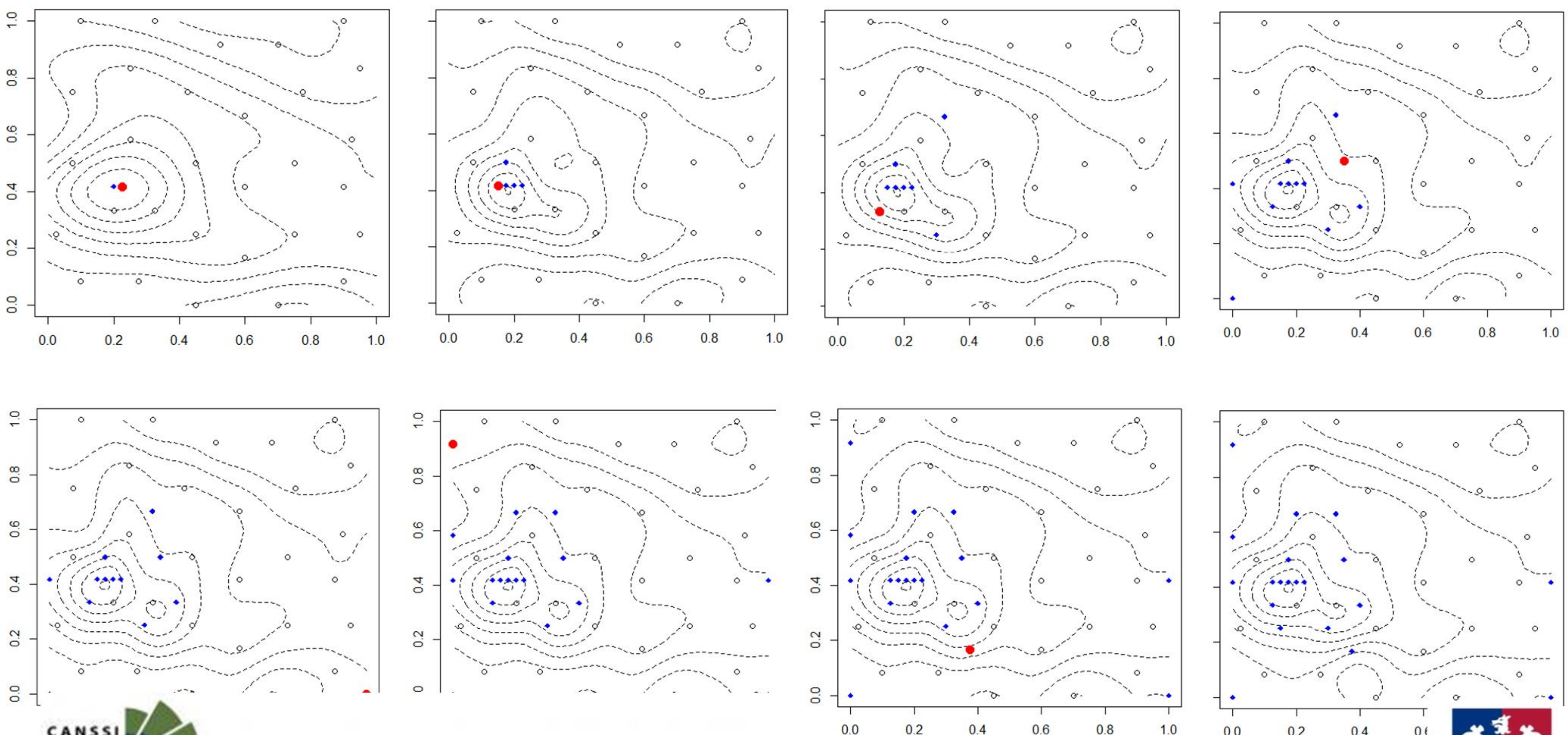
- Sequential design approach ($n_0 = 20, N = 50$)



Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Tidal power simulator – 1

- Sequential design approach ($n_0 = 30, N = 50$)

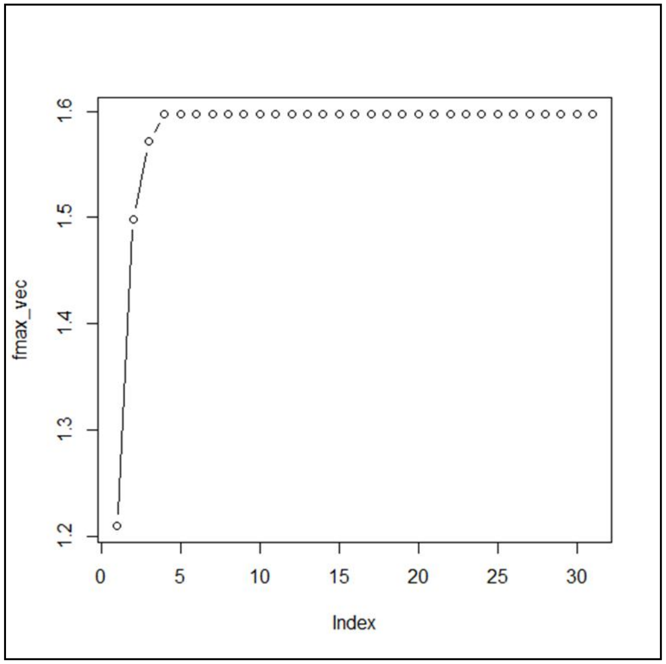


Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

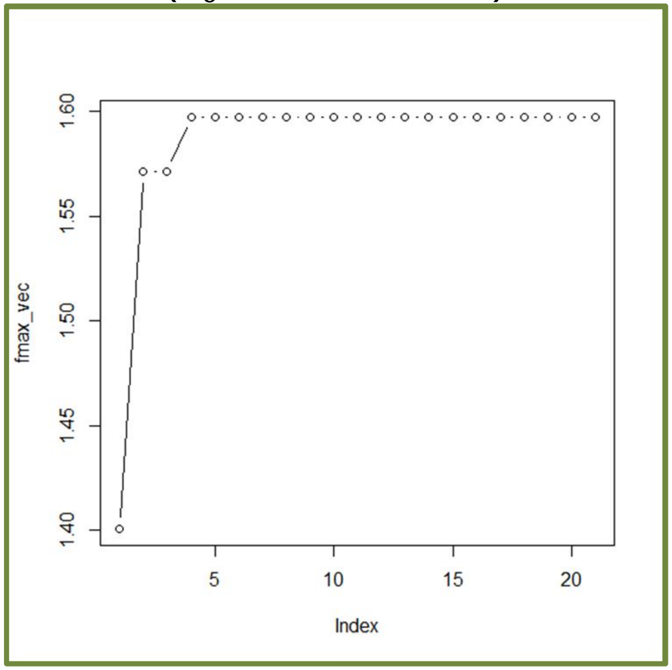
Tidal power simulator – 1

- Sequential design approach

$(n_0 = 20, N = 50)$



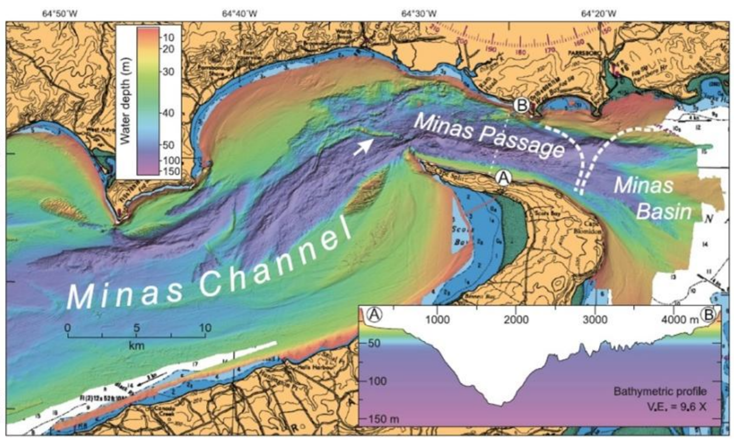
$(n_0 = 30, N = 50)$

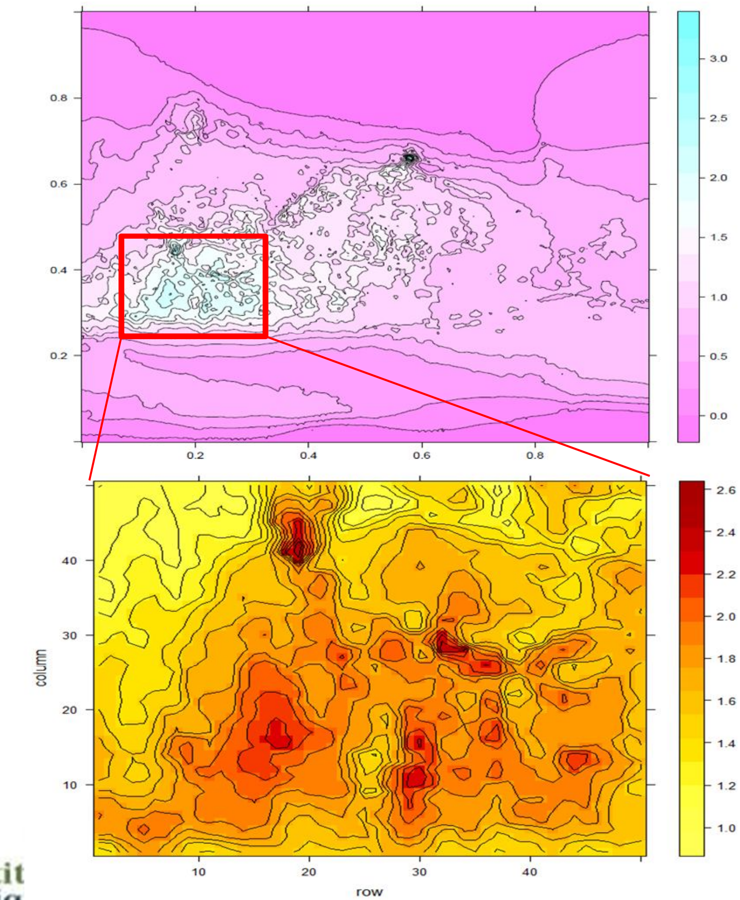


Real Application – 2

Tidal power simulator – 2

- Objective: maximize the power surface for installing **several turbines**





- $10/20\ m$ resolution simulator

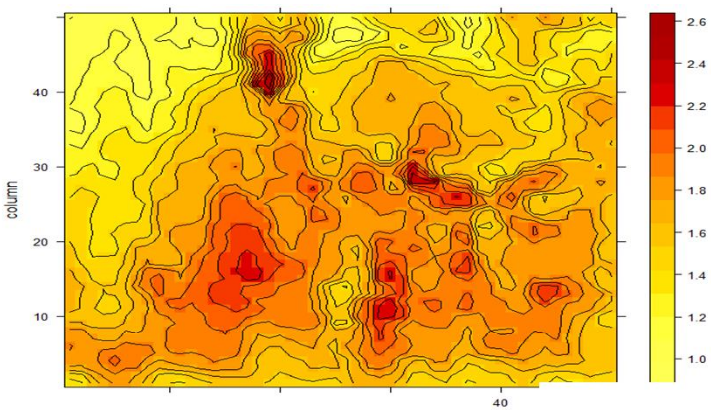
CANSSI INCASS

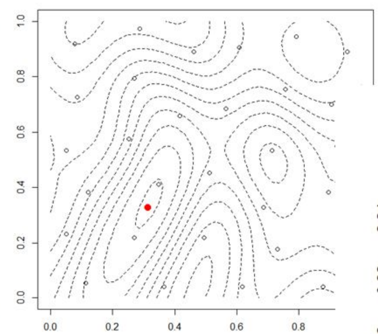
Canadian Statistical Sciences Instit
Institut canadien des sciences statistiq

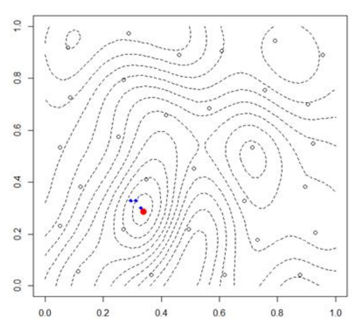
ACADIA UNIVERSITY

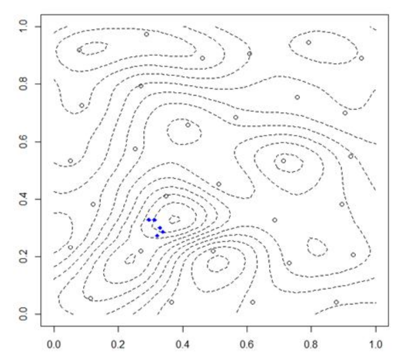
Tidal power simulator – 2

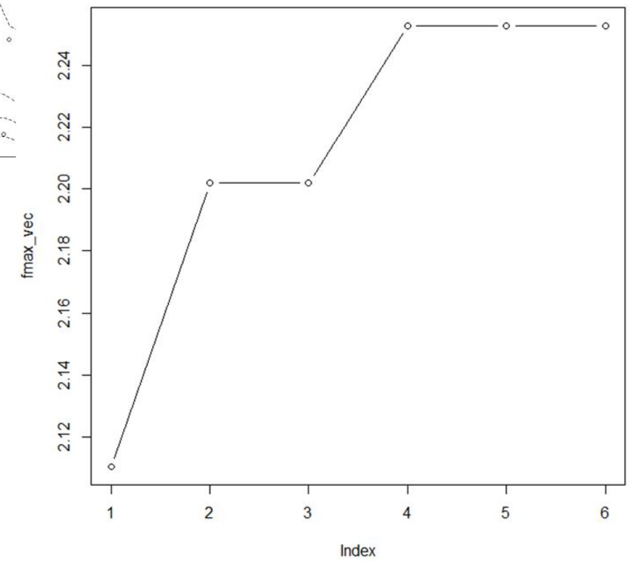
- Sequential design approach ($n_0\ =\ 30, N\ = 35$)







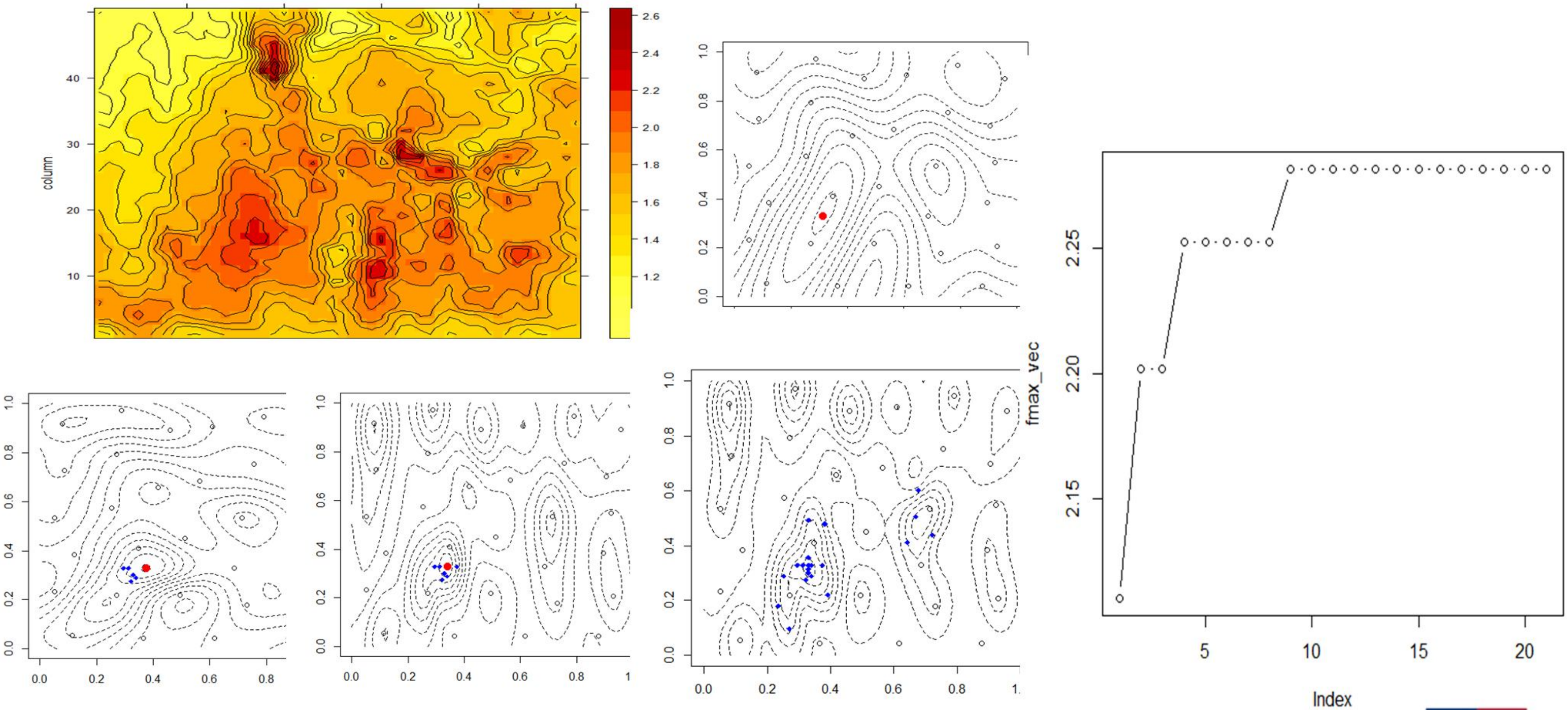




Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Tidal power simulator – 2

- Sequential design approach ($n_0 = 30, N = 50$)

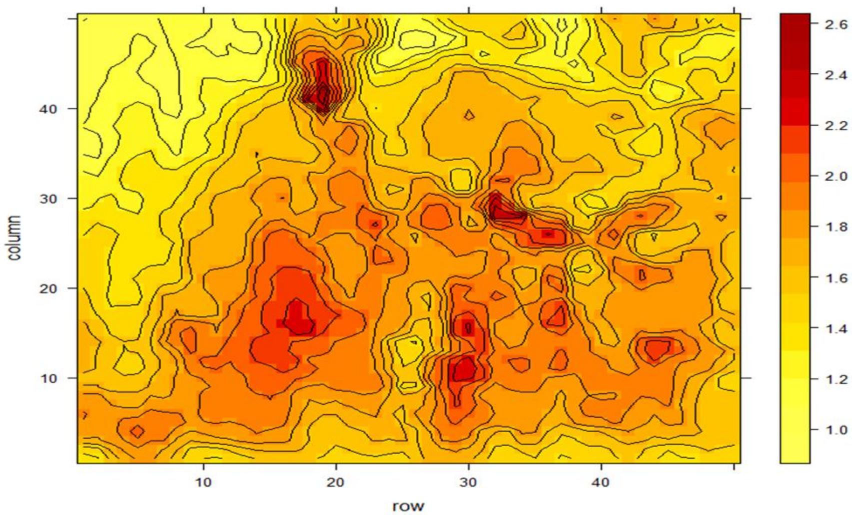


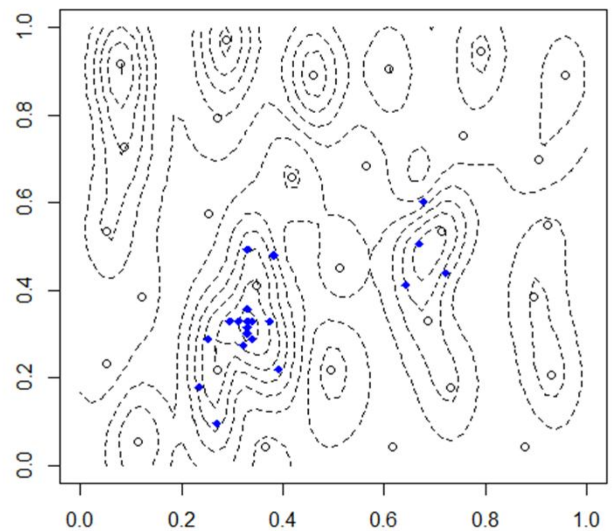
Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Tidal power simulator – 2

- Sequential design approach ($n_0 = 30, N = 50$)





- **Any ideas for getting better results?**

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Real Application – 3

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Tidal power modeling - issues

- One 1MW **OpenHydro** turbine was installed by *Fundy Ocean Research Center for Energy* (**FORCE**) in the Minas Passage during Nov 2009 – Dec 2010
  - Unfortunately, no access to the data

- FORCE and OpenHydro intend to deploy a 4MW tidal array by 2015

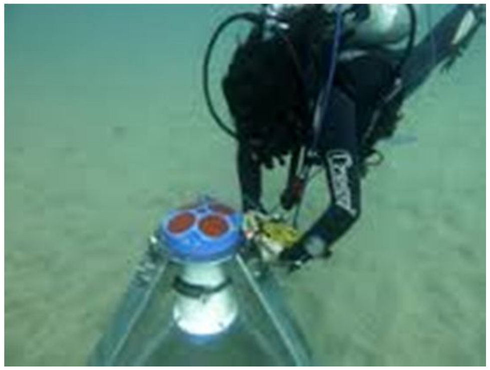
- $10-million **turbine was destroyed due to strong current**

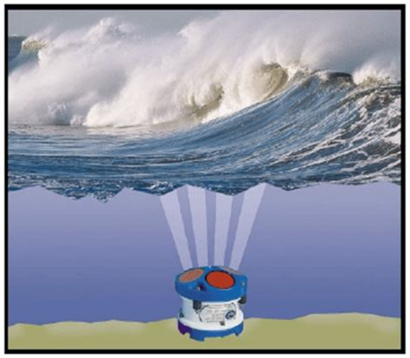
Turbine construction

- Successful development of turbines to generate electricity from tidal currents **requires more knowledge of the inflow conditions**.

- The key parameters (turbulence intensity and turbulence spectra) are estimated by collecting real data using *acoustic Doppler current profiler* (ADCP) and *acoustic Doppler velocimeter* (ADV) devices.

- .



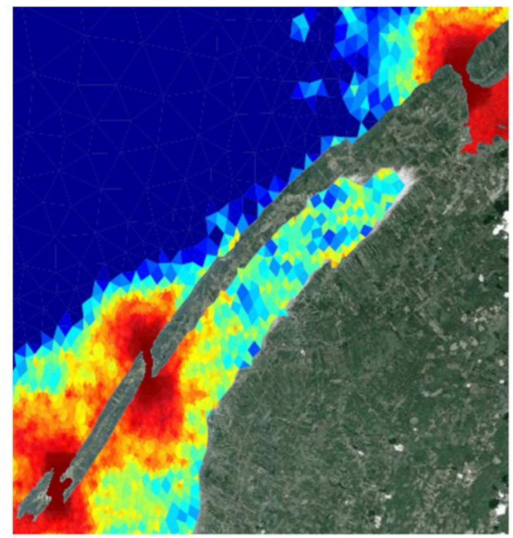


Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

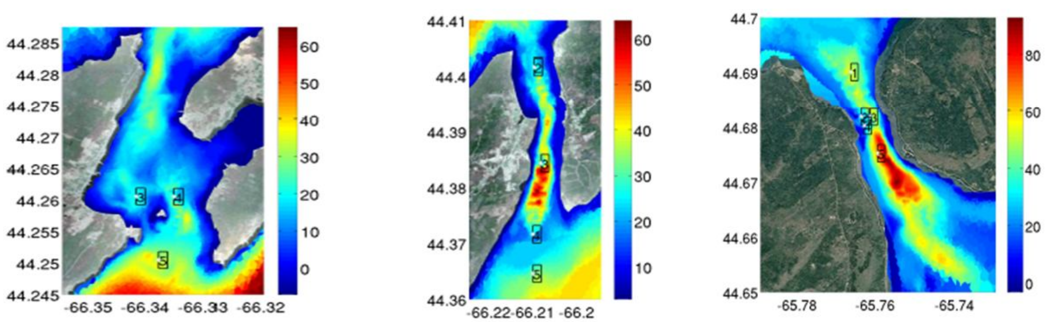
Calibration problem

- We have real ADCP data for 13 sites in Digby Neck region
- We also have simulator (DNgrid) data for these sites and more



- ➢ Time-series response (velocity)
- ➢ At each location the data was recorded for 1 month

  actual time lag 1sec - 2min (working with 10min avg lag)



(a) Grand Passage (b) Petit Passage (c) Digby Gut

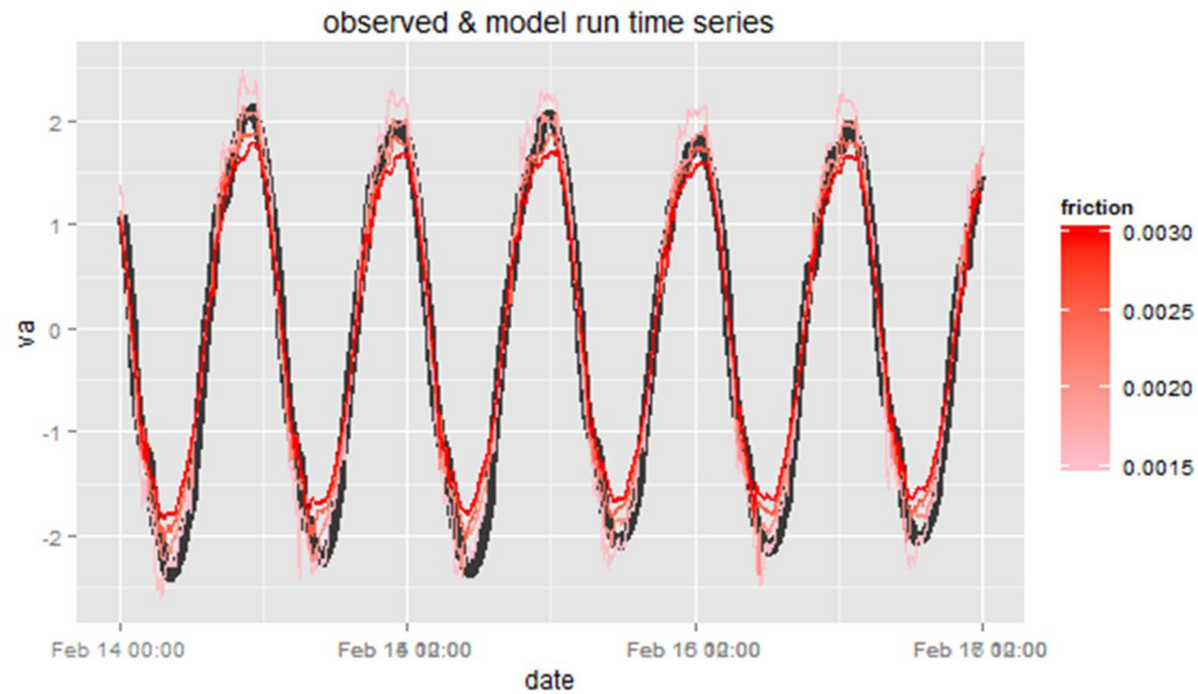
CANSSI INCASS

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Calibration problem

- Objective: find **bottom friction** (key parameter of DNGrid) that gives the best match



Calibration problem

- Statistical problem

- Field (ADCP) data: velocity time-series at 13 locations

$$W_{t,0}^{(l)}, l = 1, 2, \ldots, 13, t = 1, 2, \ldots, T$$

- Model (DNGrid) data: velocity time-series at 13 locations for a given bottom friction (b)

$$W_t^{(l)}(b), l = 1, 2, \ldots, 13, t = 1, 2, \ldots, T$$

  - Every model run gives the velocity time series for all 13 locations.

- Objective: To calibrate the computer model (find optimal b) to match reality

CANSSI INCASS

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Calibration problem

- Minimization problem

Find b that minimizes the following sum of squares

$$SS(b) = \sum_{l=1}^{13} \sum_{t=1}^{T} \left(W_t^{(l)}(b) - W_{t,0}^{(l)} \right)^2$$

- Used harmonic analysis to decompose the time series
- Used specific weights for choosing key constituents of harmonic analysis
- Used EI-based sequential design to optimize this SS

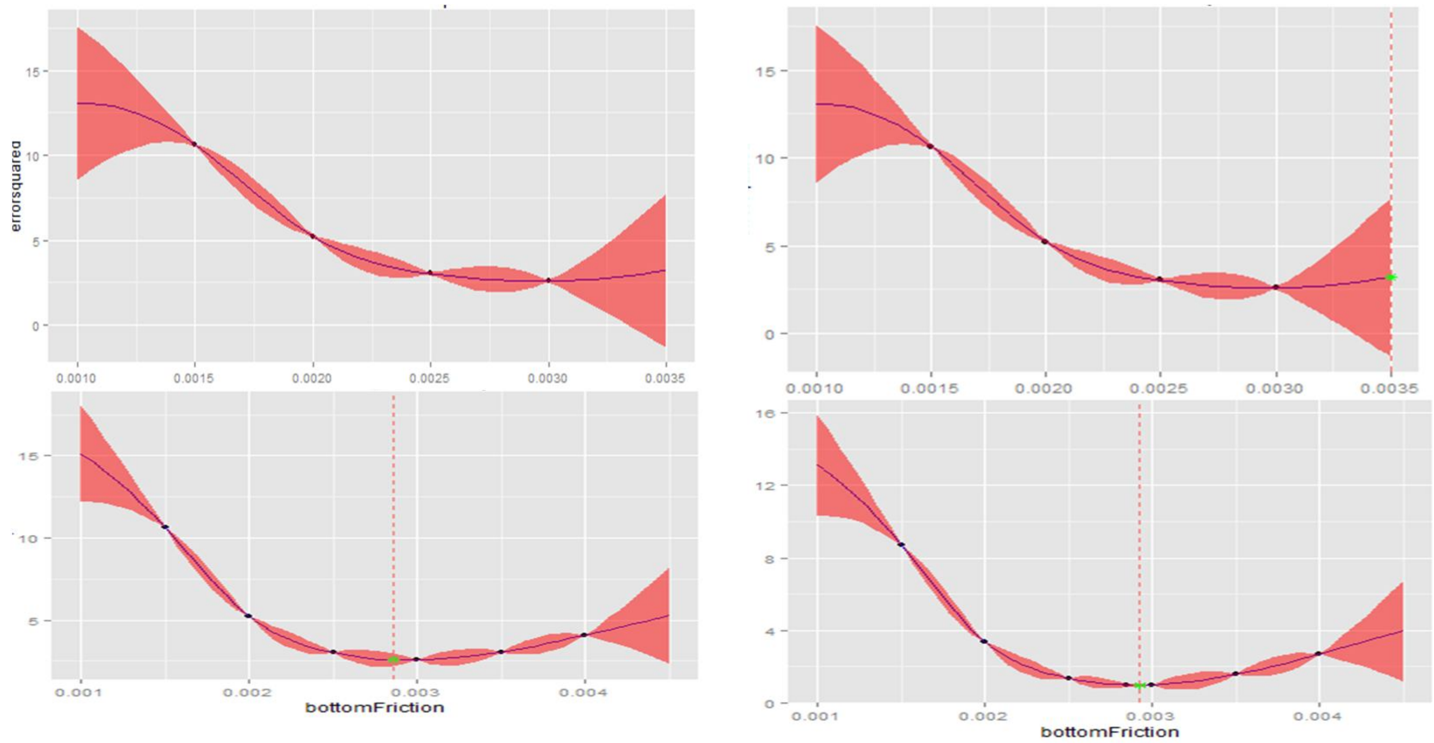
CANSSI INCASS

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Calibration problem

- <u>Minimization problem</u>



- Still working on validation, and sensitivity of harmonic constituents.

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

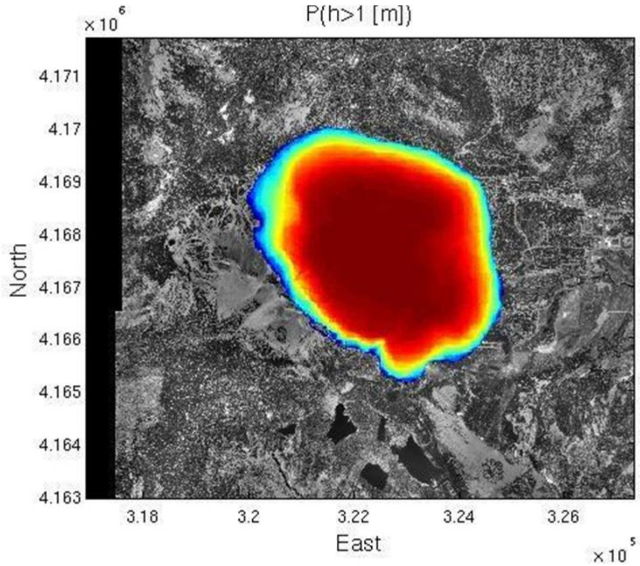
Real Application – 4

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Volcano simulator – TITAN2D

- Based on a study of Colima Volcano in Mexico (Elaine Spiller; Bayarri et al. 2009)
- Response: $y = \sqrt{z}$, where z is the maximum flow height at a particular critical location
- Predictors:
  - X_1 - pyroclastic flow volume
  - X_2 - basal fraction angle

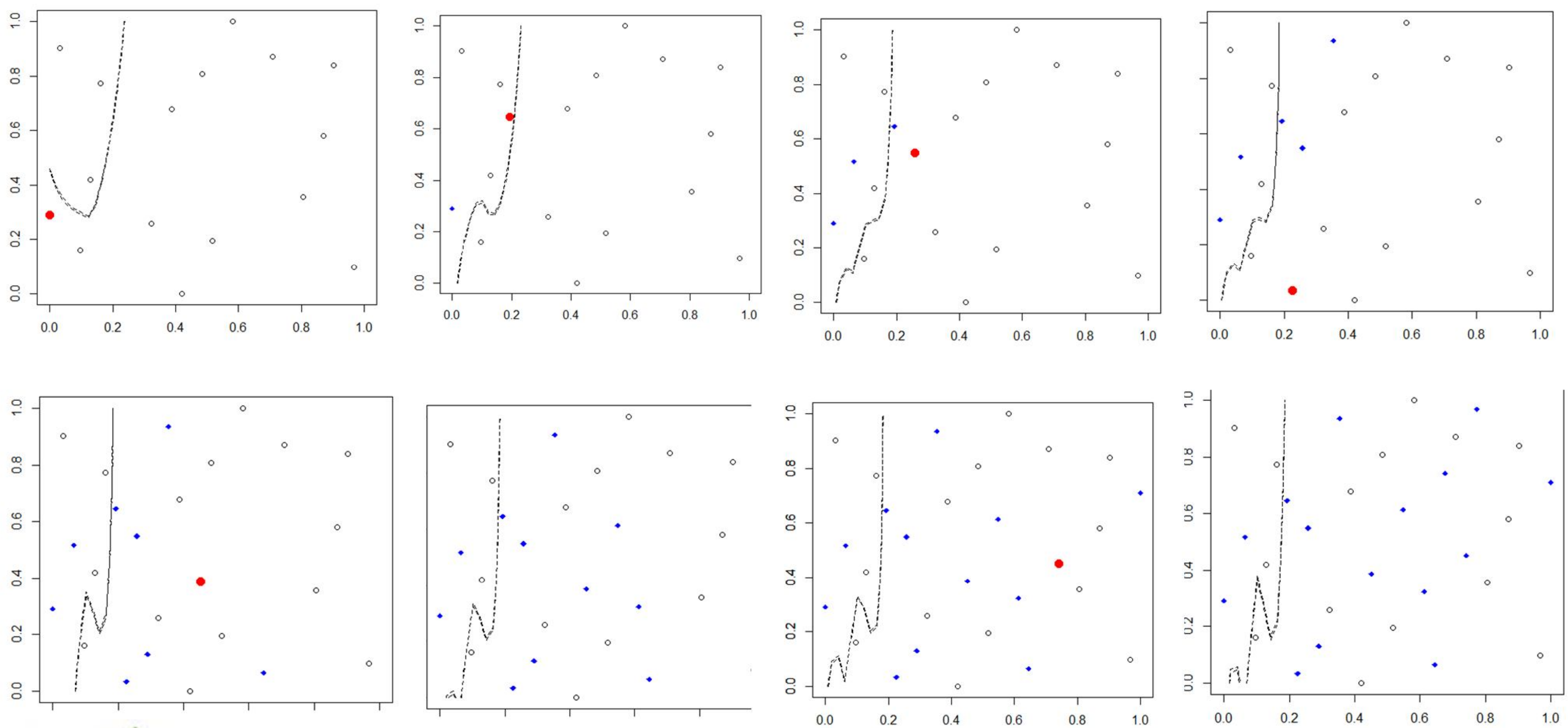
(random photo from internet) →



- Scientific objective: estimate the "catastrophic region", i.e., contour at $y(x) \geq 1$.

Volcano simulator

- Contour estimation with $n_0 = 15$, $N = 32$ (at $\boldsymbol{y(x) = 1}$)



Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Real Application – 5

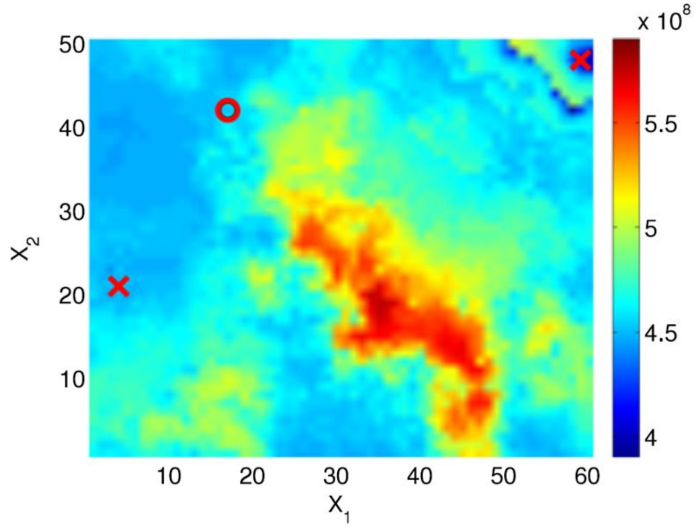
Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

Oil reservoir simulator

- Matlab Reservoir Simulator (MRST) (Lie et al., 2011; SINTEF Applied Mathematics, 2012).

- Response: the Net Present Value (NPV) of the produced oil
- Predictors: locations (x_1, x_2) of two injection and two production wells & several economical parameters

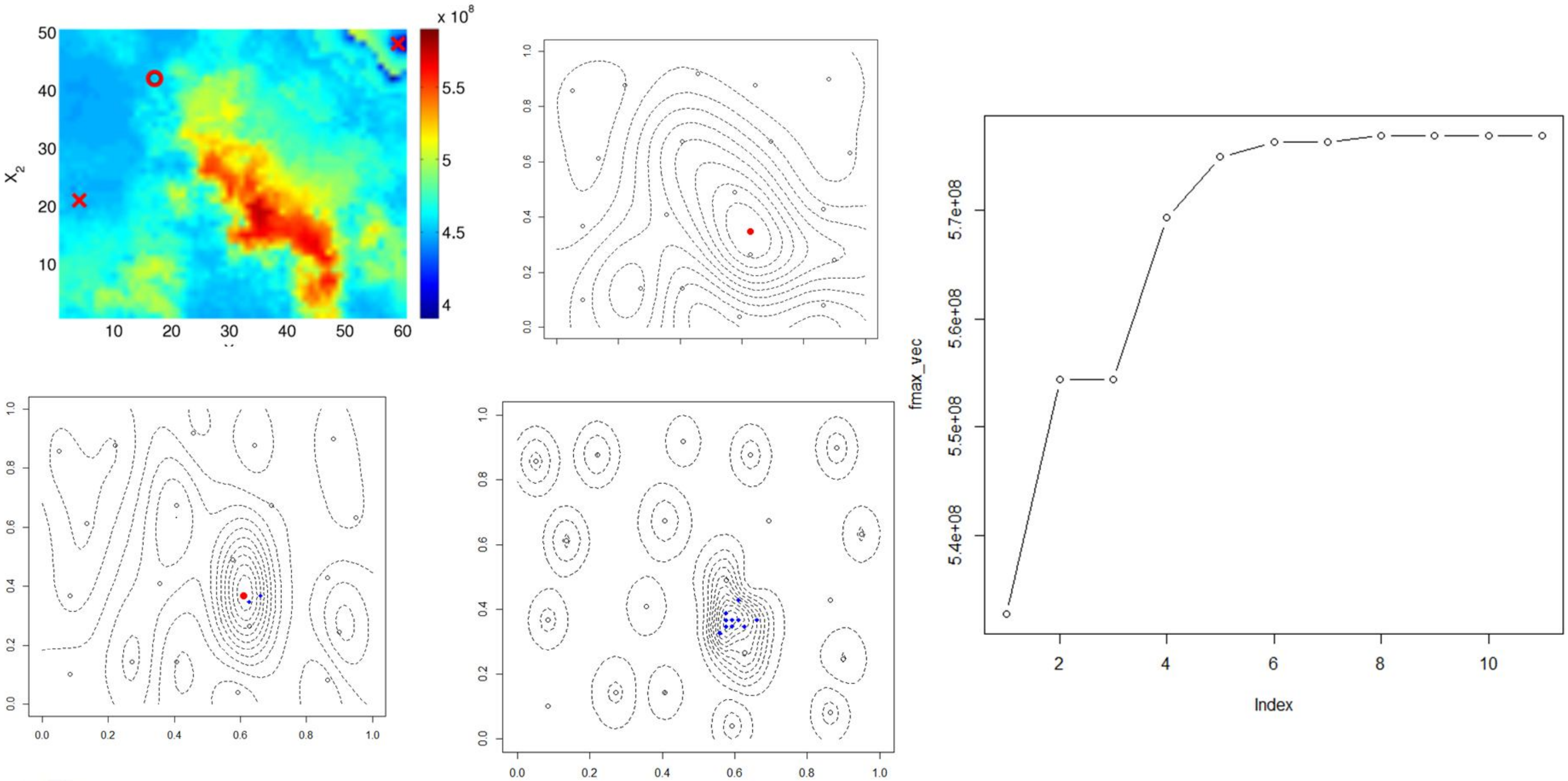
- Assume three well locations are already chosen
  - Two injection wells (x)
  - one production well (o)



- Objective: maximize NPV for finding an optimal location for drilling a production oil well

Oil reservoir simulator

- Global optimization with $n_0 = 20, N = 30$



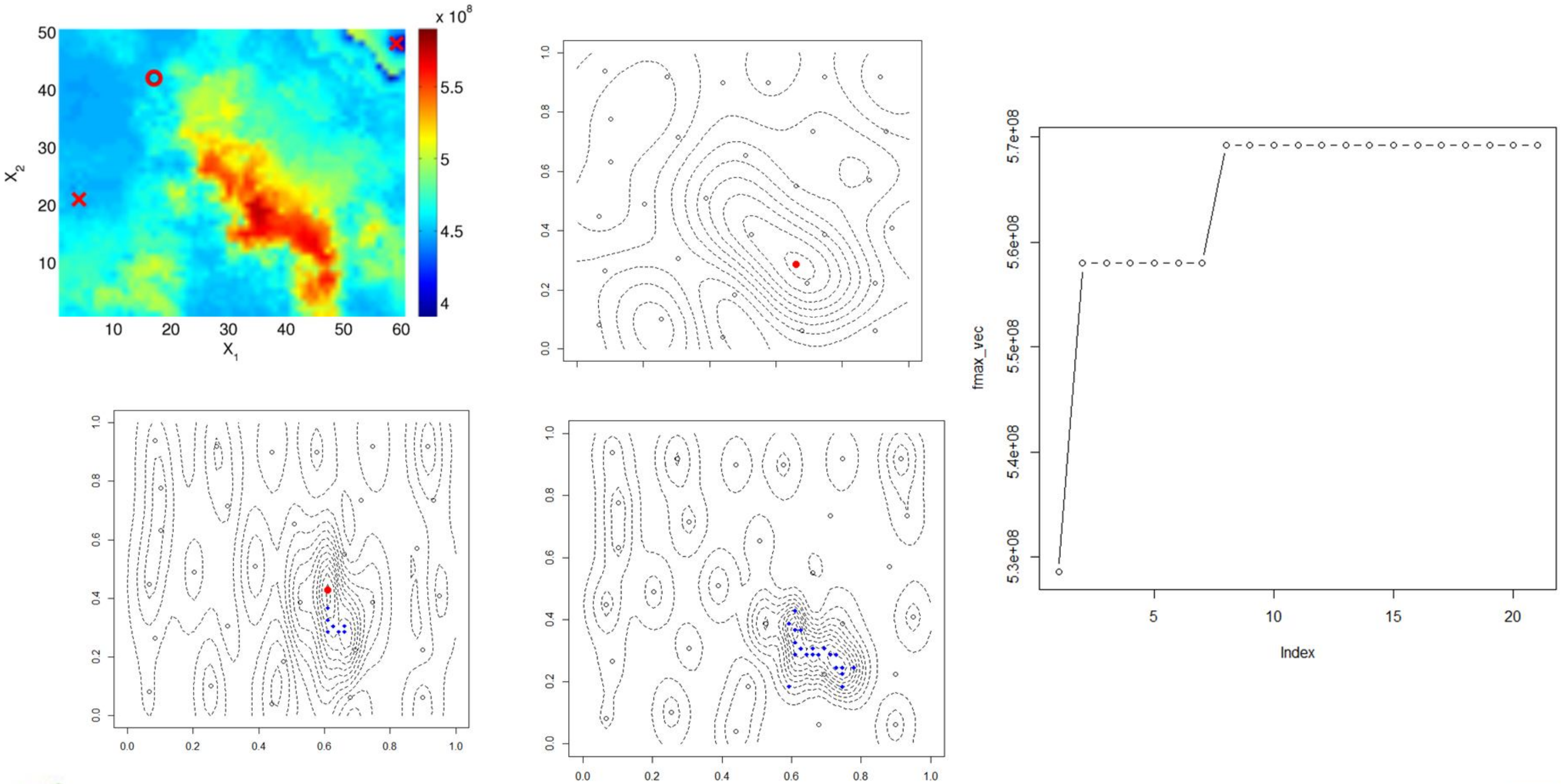
CANSSI INCASS

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

ACADIA UNIVERSITY

Oil reservoir simulator

- Global optimization with $n_0 = 30, N = 50$



Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

The end

Canadian Statistical Sciences Institute
Institut canadien des sciences statistiques

CANSSI INCASS

ACADIA UNIVERSITY