OpenTURNS Users’ Day #11

Friday, the 15\textsuperscript{th}, June 2018

- The Galton Board (1822 - 1911) : british explorer, geograph, inventor, meteorologist, proto-geneticist, psychometrician, statistician
OpenTURNS User’s Day # 11

9h - 10h
➢ Reception at EDF Lab Saclay

10h – 11h
➢ Welcome speech: E. Duceau (Scientific Advisor Airbus)

➢ Open TURNS
   ✓ General features
   ✓ New features of the 1.10 & 1.11 releases
   ✓ On going & future developments

11h – 12h
➢ «Graph approach for high dimensional probabilistic model – otagram module»: PH. Wuillemin (Sorbonne University, LIP6) – R. Lebrun (Airbus)

Lunch (12h – 13h30)

13h30 – 15h30
➢ Studies using Open TURNS (1/2)
   ✓ « Anomaly detection in aircraft sensor data »: N. Huet (IMACS), J. Sen Gupta (Airbus)
   ✓ « ANATOLE: 3D tolerancing and sensitivity analysis »: B. Chaigne (IMACS)
   ✓ « New features of the GUI of OpenTURNS: system modelisations, dependence, validation»: M. Baudin (EDF), J. Schueller (Phimeca), A. Ladier (Phimeca)
   ✓ « Marges for the conception of complex systems »: A. Touboul (IRT)

Break (15h30 – 16h)

16h – 17h
➢ Studies using Open TURNS (2/2)
   ✓ « Estimation of penstock pipes reliability using system reliability methods»: E. Ardillon (EDF), A. Dumas (Phimeca)
   ✓ « Modelisation of the maximum annual magnitude of earthquakes with extreme value distributions»: A. Dutfoy (EDF)
Welcome speech

Eric Duceau
(Scientific Advisor Airbus)
What is OpenTURNS?

OpenTURNS
An Open source initiative for the Treatment of Uncertainties, Risks'N Statistics

- Multivariate probabilistic modeling including dependence
- Numerical tools dedicated to the treatment of uncertainties
- Generic coupling to any type of physical model
- Open source, LGPL licensed, C++/Python library

www.openturns.org

- Linux, Windows
- First version launched in 2007
- 4 full time software developers
- Project size: 720 classes

- Easy to use in Python world (Numpy, Scipy, Matplotlib)
- Algorithm parallelisation (TBB)
Data analysis

Visual analysis: QQ-Plot, Cobweb
Fitting tests: Kolmogorov, Chi2
Multivariate distribution: kernel smoothing (KDE), maximum likelihood
Process: covariance models, Welch and Whittle estimators
Bayesian calibration: Metropolis-Hastings, conditional distribution

Probabilistic modeling

Dependence modelling: elliptical, archimedian copulas.
Univariate distribution: Normal, Weibull
Multivariate distribution: Student, Dirichlet, Multinomial, User-defined
Process: Gaussian, ARMA, Random walk.
Covariance models: Matern, Exponential, User-defined

Meta modeling

Functional basis methods: orthogonal basis (polynomials, Fourier, Haar, Soize Ghanem)
Gaussian process regression: General linear model (GLM), Kriging
Spectral methods: functional chaos (PCE), Karhunen-Loeve, low-rank tensors

Reliability, sensitivity

Sampling methods: Monte Carlo, LHS, low discrepancy sequences
Variance reduction methods: importance sampling, subset sampling
Approximation methods: FORM, SORM
Indices: Spearman, Sobol, ANCOVA
Importance factors: perturbation method, FORM, Monte Carlo

Functional modeling

Numerical functions: symbolic, Python-defined, user-defined
Function operators: addition, product, composition, gradients
Function transformation: linear combination, aggregation, parametrization
Polynomials: orthogonal polynomial, algebra

Numerical methods

Integration: Gauss-Kronrod
Optimization: NLopt, Cobyla, TNC
Root finding: Brent, Bisection
Linear algebra: Matrix, HMat
Interpolation: piecewise linear, piecewise Hermite
Least squares: SVD, QR, Cholesky
Copulas

In this part, we will define the concept of copula.

To define the joint probability density function of the random input vector $X$ by composition, one needs:

- the specification of the copula of interest $C$ with its parameters,
- the specification of the $n$ marginal laws of interest $F_{X_i}$ of the $n$ input variables $X_i$.

The joint cumulative density function is therefore defined by:

$P(X^1 \leq x^1, X^2 \leq x^2, \ldots, X^n \leq x^n) = C(F_{X_1}(x^1), F_{X_2}(x^2), \ldots, F_{X_n}(x^n))$

Copulas allow to represent the part of the joint cumulative density function which is not described by the marginal laws. It enables to represent the dependency structure of the input variables. A copula is a special cumulative density function defined on $[0, 1]^n$ whose marginal distributions are uniform on $[0, 1]$. The choice of the dependence structure is disconnected from the choice of the marginal distributions.

A copula, restricted to $[0, 1]^n$ is a $n$-dimensional cumulative density function with uniform marginals.

- $C(u) \geq 0, \forall u \in [0, 1]^n$
- $C(u) = u$, $\forall u = (1, \ldots, 1, u, 1, \ldots, 1)$
- For all $n$-box $B = [u_1, b_1] \times \cdots \times [u_n, b_n] \subset [0, 1]^n$, we have $V_c(B) \geq 0$, where:
  - $V_c(B) = \sum_{i=1}^{2^n} \text{sign}(u_i) \times C(u_i)$, the summation being made over the $2^n$ vertices $u_i$ of $B$.
  - $\text{sign}(u_i) = +1$ if $x_i^+ \in \{u_i \}$ for an even number of $|x_i^+ \in \{u_i \}$, $\text{sign}(u_i) = -1$ otherwise.

API:

- See the list of available copulas.

Examples:

- See Create a copula
- See Assemble copulas
- See Extract the copula from a distribution

References:

- Nelsen, Introduction to Copulas

Several documentations for several goals

- Mathematics: Reference Guide
- The way to use it: API
- Some script examples: Examples (basic & advanced)
- The library architecture: Developer’s Guide

The documentation wants to be ....:

- as exhaustive as possible
- illustrated : with access to the script used to produce the graph
- a link between different needs: crossed references between classes & ref doc
- an help providing validated examples (part of the tests catalog)
Open TURNS – Trainings

EDF

- « Uncertainty Management : Open TURNS »
  - 3 days to learn the use of Open TURNS : python TUI, wrapping aspects, application of the methodology from step A to step C & C’
  - next session : 10-12/09/2018, ITECH
  - EDF Lab Paris-Saclay

- « Uncertainty Management : Methodology »
  - 3 days to learn the Uncertainty Methodology
  - next session : 3-5/09/2018, ITECH
  - EDF Lab Chatou

Phimeca : 2 sessions each year for each modulus

- Probability & Statistics: 2 days
- Response Surfaces Models : 2 days
- Uncertainty propagation methods for sensitivity & dispersion analysis : 2 * 1 day
- Uncertainty propagation methods for reliability evaluation : 2 days

- Introduction to the use of OpenTURNS : 2 days
- Modelica & Python: 1 day
- Salome : 2 days

PRACE

- 1 annual training at Maison de la Simulation (Saclay)
- 3 days
- Méthodology + Practice on OpenTURNS or Uranie (CEA)
Where to load OpenTURNS?

**Sourceforge:**
- **Sources:** for those who want to compile
- **Windows Python Module** (architectures 32 et 64 bit + 3 python versions)
- **Native Windows Library**

**Sous Linux:**
- distributions Debian, Ubuntu, CentoS, Fedor, openSUZE, Archlinux, Parabola
- compiled packages installed by the **package manager**
  (care: Administrative privilege)

**Conda:**
- installation of python without privilege (USERS)
- **Python Module python on anaconda** (https://anaconda.org/openturns/openturns) since v1.7 (2016): windows / linux, several architectures, …

**SALOME::**
- OpenTURNS is now included in specific Salome version with a specific GUI
  - https://www.salome-platform.org/contributions/copy_of_combs

**Téléchargements:**
- 14 000 downloads from SourceForge since 2007 (1500 from 2017)
- 40 000 downloads Conda since the beginning of 2016
- + all the others!
Open TURNS publications

- **International conferences**
  - EuroScipy (08/ 2017): python community
    - 2-day conference in Europe
    - OpenTURNS presentation
    - Objective next year: OpenTURNS tutorial ➔ 1h30 live demonstration
  - Significative increase in OpenTURNS loads with conda
  - SIAMUQ SIAMUQ 2018 : uncertainty community
  - UNCECOMP 2017 : uncertainty community

- **Publication**
  - Springer Handbook of Uncertainty Quantification (2017): one chapter dedicated to OpenTURNS
Open TURNS in the academic world

Engineers trained to the Uncertainty Methodology and able to use OpenTURNS on the labor market very soon

Anthony Nouy: 3rd year curse on meta models (low rank tensors)

Uncertainty Management – Robust Optimization (Airbus)

Master curse on Data Science

École des Ponts ParisTech

Uncertainty Management curse
(participation EDF and Airbus)

Henning Bruske: Master 2 curse

Master curse based on OpenTURNS(CEA, IMACS)
New features of the 1.10 & 1.11 releases

- Uncertainty modelling & quantification
  - SmoothedUniform($a$, $b$, $\sigma$) distribution of $X+Y$ where $X \sim U(a, b)$, $Y \sim N(0, \sigma)$, $X$ indep. $Y$
    - This distribution allows for further simplifications in RandomMixture (speed & accuracy gains)
  - FrechetFactory based on maximum likelihood estimates. Used eg in extreme values quantification.
New features of the 1.10 & 1.11 releases

- **Numerical methods**
  - Fehlberg to integrate systems of ordinary differential equations (explicit, adaptive)
  - Nested Runge Kutta schemes
  - Dynamic simulation grid by local error estimation
  - Results interpolated on the user grid.
  - Four schemes available (1/2 to 4/5)

- FilonQuadrature
  - Filon: 41 calls, 0.08% error
  - Gauss Legendre: 41 calls, >100000% error
  - Gauss Kronrod: 60 calls, >100000% error
New features of the 1.10 & 1.11 releases

- **New models with functional input or output**
  - **OT 1.9:**
    - Fonction to map $\mathbb{R}^n$ into $\mathbb{R}^q$
    - FieldFunction to map $F(\mathbb{R}^m, \mathbb{R}^n)$ into $F(\mathbb{R}^p, \mathbb{R}^q)$
  - **OT 1.10 adds**
    - PointToFieldFunction to map $\mathbb{R}^n$ into $F(\mathbb{R}^p, \mathbb{R}^q)$
    - FieldToPointFunction to map $F(\mathbb{R}^m, \mathbb{R}^n)$ into $\mathbb{R}^q$

+ all the classes to compose these models in any way: PointToPointConnection, PointToFieldConnection, FieldToPointConnection, FieldToFieldConnection

- You still have to present all the inputs as a unique field if at least one of them is a field, and all the inputs must share the same mesh. Same constraint for the output.

- The input and output meshes will be part of the model definition in the future version, only the sample of values will transit through the evaluation operator.
New features of the 1.10 & 1.11 releases

- **Sobol’ indices asymptotic variance estimators**
  - Previously only bootstrap estimator was available (asymptotic only for Martinez)
  - Asymptotic variance estimators using the “delta” method (Janon & al, 2014)
  - Variance estimators are declined for all Sobol’ estimators (Jansen, MauntzK, ...)
  - Access to the confidence interval as well as the distribution of the indices
  - Some corrections of the raw estimators of the indices (Saltelli, ...)

```python
import openturns as ot
X = ot.SobolIndicesExperiment(distribution, size).generate()
Y = f(X)
algo = ot.SaltelliSensitivityAlgorithm(X, Y)
algo.setUseAsymptoticDistribution(True)

fo = algo.getFirstOrderIndices()
to = algo.getTotalOrderIndices()
interval_fo = algo.getFirstOrderIndicesInterval()
interval_to = algo.getTotalOrderIndicesInterval()
dist_fo = algo.getFirstOrderIndicesDistribution()
dist_to = algo.getTotalOrderIndicesDistribution()
```
New features of the 1.10 & 1.11 releases

- New installation media
  - Python package index (PyPI)
  - otconda bundle: Anaconda-like offline installer of OpenTURNS and its modules

```
pip install openturns [--user]
```

**v1.11**

- jschueler released this 20 days ago

**Assets**

- [otconda2-1.11-Linux-x86_64.sh](#)
- [otconda2-1.11-MacOSX-x86_64.sh](#)
- [otconda2-1.11-Windows-x86_64.exe](#)
New features of the 1.10 & 1.11 releases

- **ExprTk**
  - Like muParser, allows to parse analytical formulas from C++:
    - + no external dependency, this library is a single C++ header file
    - + slightly faster evaluation than muParser
    - + supports multiple outputs
    - + full programming language (see its documentation)
    - - requires more memory at run time
  - Both muParser and ExprTk are currently supported
  - ResourceMap SymbolicParser-Backend entry set to ExprTk (default) or MuParser

```python
>>> import openturns as ot
>>> from openturns.viewer import View

>>> formula = "sin(2*pi_*x)*exp(-x^2/2)"
>>> f = ot.SymbolicFunction("x", formula)
>>> graph = f.draw(-1.2, 1.2)
>>> graph.setTitle("y0=\"+formula\")
>>> View(graph).show()

>>> import openturns as ot
>>> from math import sqrt, atan2

>>> formula = "r := sqrt(x * x + y * y); theta := atan2(y, x)"
>>> polar = ot.SymbolicFunction(["x", "y"], ["r", "theta"], formula)
>>> print(polar([1.0, 1.0]))
[1.41421, 0.785398]
```
New features of the 1.10 & 1.11 releases

- Improve objects conversion Python ↔ OpenTURNS
  - Avoid copies when converting OT containers (Point, Sample, Matrix) into Numpy arrays
  - In general, there are still copies when converting Numpy arrays into OT containers
  - Functions defined in Python now receive an argument which is not a Sample but a view on a Sample, which only supports indexing; should work everywhere, call Sample(X) if you need to call Sample methods on X
  - Functions defined in Python can be as fast as a SymbolicFunction (defined in Python), but not trivial; implement evaluation on a Sample with Numpy arrays to avoid Python for-loops

```python
import openturns as ot
from math import cos, sin, tan, exp
import numpy as np

def f_point(X):
    x = X[0]
    return [exp(cos(sin(tan(x)))) \ 
            + cos(exp(sin(tan(x))))]

def f_sample(X):
    x = np.array(X)
    return np.exp(np.cos(np.sin(np.tan(x)))) \ 
           + np.cos(np.exp(np.sin(np.tan(x))))

f0 = ot.PythonFunction(1, 1, f_point)
f1 = ot.PythonFunction(1, 1, func_sample=f_sample)
f2 = ot.SymbolicFunction("x",
                        "exp(cos(sin(tan(x))))+cos(exp(sin(tan(x))))")
```

<table>
<thead>
<tr>
<th></th>
<th>OT 1.9 &amp; 1.10</th>
<th>OT 1.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>f0</td>
<td>5.85s</td>
<td>1.15s</td>
</tr>
<tr>
<td>f1</td>
<td>0.60s</td>
<td>0.17s</td>
</tr>
<tr>
<td>f2</td>
<td>0.35s</td>
<td>0.22s</td>
</tr>
</tbody>
</table>

Evaluation on a Sample of size 1E6
New features of the 1.10 & 1.11 releases

- Changes in handling of cache and history of functions
  - Cache and history was defined in lowest-level class, and derived classes had to take care of implement them if needed
  - Now these methods have been removed, and a single class MemoizeFunction is added

OT < 1.11

```python
import openturns as ot
...
heavyFunction.enableCache()
heavyFunction.enableHistory()
outSample = heavyFunction(inputSample)
nrCalls = heavyFunction.getCallsNumber()
```

OT ≥ 1.11

```python
import openturns as ot
...
myFunc = ot.MemoizeFunction(heavyFunction)
myFunc.enableCache()
myFunc.enableHistory()
outSample = myFunc(inputSample)
nrCalls = myFunc.getCallsNumber()
```
New features of the 1.10 & 1.11 releases

Geometry

- Domain algebra: eg allows to define complex failure events

\[
\begin{align*}
D_1 &= \text{DomainComplement}(\text{cube}) \\
D_2 &= \text{DomainDifference}(\text{cube}, \text{sphere}) \\
D_3 &= \text{DomainDisjunctiveUnion}(\text{cube}, \text{sphere}) \\
D_4 &= \text{DomainIntersection}(\text{cube}, \text{sphere}) \\
D_5 &= \text{DomainUnion}(\text{cube}, \text{sphere})
\end{align*}
\]

\[
\begin{align*}
X &= \text{RandomVector}(\text{Normal}(2)) \\
E_1 &= \text{EventDomain}(X, D_1) \ # \ or \ D_2, D_3...
\end{align*}
\]
New features of the 1.10 & 1.11 releases

- **Geometry (cont’ed)**
  - Up to 1.10, Mesh contained a kd-tree to speed-up nearest neighbour searches
  - Mesh is now only a container of vertices and simplices
  - Services for spatial lookup are moved into new classes
  - Mesh no more derives from Domain
  - MeshDomain(mesh) converts a Mesh into a Domain

- **Spatial lookup**
  - New classes added in 1.11:
    - NearestNeighbourAlgorithm (interface class)
    - NearestNeighbour1D, RegularGridNearestNeighbour (1-d meshes)
    - KDTree, NaiveNearestNeighbour
  - These classes provide a query method to find the index of the nearest neighbour of a Point

```python
>>> import openturns as ot
>>> sample = ot.Normal(2).getSample(10000)
>>> finder = ot.NearestNeighbourAlgorithm(sample)
>>> neighbour = sample[finder.query([[0.1, 0.2]])]
>>> print(neighbour)
[0.0967085, 0.194408]
```
New features of the 1.10 & 1.11 releases

Spatial lookup (cont’ed)

✓ New classes to find the enclosing simplex of a Point in a Mesh:
  EnclosingSimplexAlgorithm (interface class)
  EnclosingSimplexMonotonic1D, RegularGridEnclosingSimplex (1-d meshes)
  NaiveEnclosingSimplex, BoundingVolumeHierarchy

✓ These classes provide a query method to find the index of the enclosing simplex

```python
>>> import openturns as ot
>>> mesh = ot.IntervalMesher([100]*2).build(ot.Interval([0.0]*2, [10.0]*2))
>>> locator = ot.EnclosingSimplexAlgorithm(mesh.getVertices(), mesh.getSimplices())
>>> simplex = locator.query([0.1, 0.2])
>>> print(mesh.getVertices().select(mesh.getSimplex(simplex)))
    [ v0  v1  ]
 0 : [ 0.1 0.2 ]
 1 : [ 0.2 0.2 ]
 2 : [ 0.1 0.3 ]
```
New features of the 1.10 & 1.11 releases

- **Spatial lookup (cont’ed)**
  - Used by P1LagrangeEvaluation (to create a Function) and new class P1LagrangianInterpolation (to create a FieldFunction)
  - Mesh interpolation is computed for each evaluation of P1LagrangeEvaluation
  - Mesh interpolation is computed only once in P1LagrangeInterpolation, this is preferred when interpolating several fields

```python
>>> import openturns as ot
>>> inputMesh = ot.RegularGrid(0.0, 1.0, 4)
>>> outputMesh = ot.Mesh([[2.3]])
>>> interpolation = ot.P1LagrangianInterpolation(inputMesh, outputMesh, 1)
>>> field = ot.Field(inputMesh, [[0.5], [1.5], [1.0], [-0.5]])
>>> print(interpolation(field))
0 : [ 0.55 ]
```
Perspectives 2018-2019: main ongoing works

- Increase the coherence in the use of parameters
  - Modèles de covariance: notion de paramètres actifs
  - Fonctions: notion de ParametrizedFunctions
  - Distributions: notion de paramétrage alternatif.

- Complete Python Projection?
  - Use of a dictionary to declare the arguments of a method
  - Integration in pypi (python packaging system, currently: conda)
  - Enhanced the classes Sample and Matrix with functionalities similar to numpy

- Install OpenTURNS under the windows environment without Administrator privilege nor Internet connection

- Performance
  - Parallel evaluations of functions
  - Gérer de manière globale les threads lancés par OpenTURNS provenant de TBB ou d'OpenBlas
  - Ongoing work on H-Mat représentation compressée de matrices)

- Documentation
  - Finalize the transformation to dynamic doc
  - More and more complete…
Perspectives 2018-2019: main ongoing works

**Meta models**

- Validation tests
  - QQ-Plot, residuals graph, ...
  - Mutualize as much as possible the validation methods among meta models

- « Easy chaos »
  - Simplify as much as possible the parametrization of the algorithms

**Statistical inference**

- Distribution of the estimators
  - Quantile, probabilities, SRC indices, Sobol indices, moments, ...
  - Useful to establish confidence intervals

**Sensitivity analysis**

- Csiszar indices
  - Distance between \((Y,X_i)\) copula and the independent copula

\[ d_i = \mathbb{E} [D_f(F_Y|F_Y|X_i)] \quad \Rightarrow \quad d_i = \mathbb{E} [D_f(C_Y,X_i||\Pi)] \]
Perspectives 2018-2019: main ongoing works

- **System models**
  - Systems Events: union and intersection of elementary events
    - Probability estimation
  - Modulus otfmi
    - Fmi: standard interface model
    - Enables to insert within a modelica description a connection to an external composant
    - Enhance the modulus with all the OpenTURNS connectors: vector / field

- **Maintenance - Validation**
  - Unit validation python test
    - Distance from a reference result
    - Infrastructure of the environment test
    - A good example: Scikit-Learn ....

- **Diffusion**
  - Python communities
    - EuroScipy
    - Scipy
  - Uncertainty communities
    - SIAMUQ

- **Publications**
  - Dedicated to a particular method: multivariate Kriging, Karhunen Loeve, ... + notebooks
  - Journal of Statistical Software
« Graph approach for high dimensional probabilistic model – otagrum module »

PH. Wuillemin (Sorbonne University, LIP6)
R. Lebrun (Airbus)
Studies with Open TURNS

Study 1: « Anomaly detection in aircraft sensor data »: N.Huet (IMACS), J. Sen Gupta (Airbus)

Study 2: « ANATOLE: 3D tolerancing and sensitivity analysis »: B. Chaigne (IMACS)

Study 3: « New features of the GUI of OpenTURNS: system modelisations, dependence, validation»: M. Baudin (EDF), J. Schueller (Phimeca), A. Ladier (Phimeca)

Study 4: « Marges for the conception of complex systems »: A. Touboul (IRT)

Break

Study 5: « Estimation of penstock pipes reliability using system reliability methods»: E. Ardillon (EDF), A. Dumas (Phimeca)

Study 6: « Modelisation of the maximum annual magnitude of earthquakes with extreme value distributions»: A. Dutfoy (EDF)
Thanks for your participation ...
and see you next year!