Two Methods for Non-judgemental Inter-code Comparisons with Gaussian Processes

Prof. Anthony Scopatz
Dept. Mechanical Engineering
University of South Carolina
ERGS: http://www.ergs.sc.edu/

Technical Workshop on Fuel Cycle Simulations
July 6 - 9th, 2016

Problem

Sometimes, it is impossible to validate simulators in a traditional sense:

- Climate models
- Nuclear fuel cycle models
- Predictive economic models of physical systems

No simulator’s results may be assumed to be correct ahead of time.

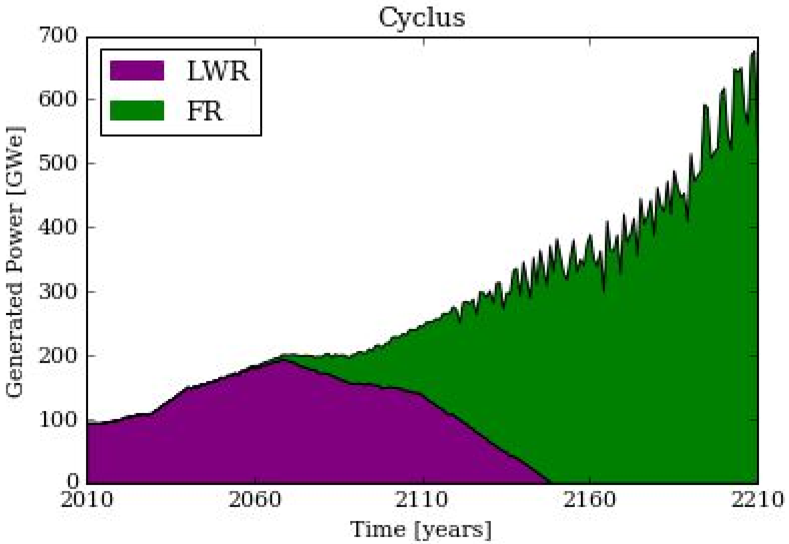
An inter-code comparison should evaluate the behaviour of its constituent simulators according to one or more metrics of interest.

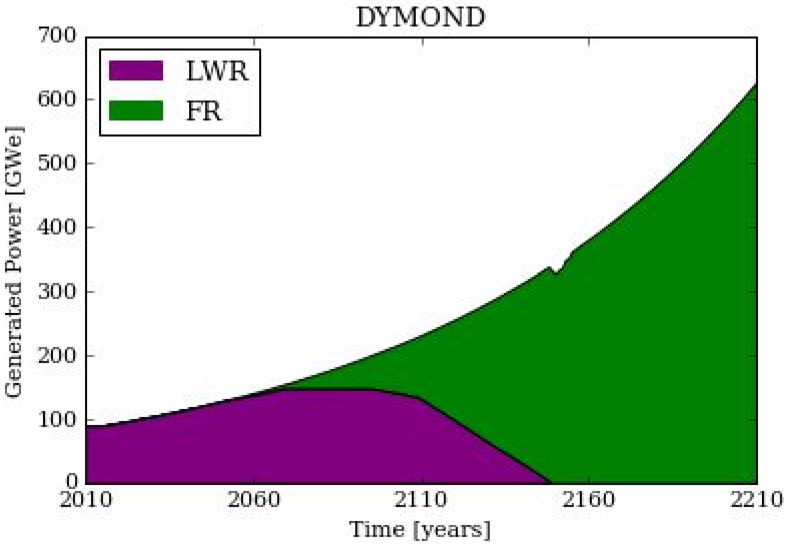
A *robust* inter-code comparison should have a rigorous and uniform approach to the above issues.

Strategy

- Use Gaussian process (GP) regression to form representative models of all simulators.
- Reduce the dimensionality to R^1 through:
    - L^2 norm of GP models
    - Dynamic time warping (DTW)

Setup EG01 -> EG23





This is a Light Water Reactor (LWR) to Fast Reactor (FR) transition scenario over 200 years. The metric displayed here is the generated electric power [GWe]. A single simulation case may have many metrics.

Gaussian Process Regression

For a metric m, a simulator s, and a feature i (e.g. LWRs vs FRs), and at time t, a Gaussian process models the expected value for the mean and covariance:

$$\mu(t) = \mathbb{E}\left[m_s^i(t)\right] = \mathbb{E}\left[m_0^i(t), m_1^i(t), \ldots\right]$$

$$k(t, t') = \mathbb{E}\left[(m_s^i(t) - \mu(t))(m_s^i(t') - \mu(t'))\right]$$

$$m_*^i(t) \approx \mathcal{GP}\left(\mu(t), k(t, t')\right) \equiv \mathcal{GP}m_s^i$$

Gaussian Process Regression

The GP then chooses a functional form (kernel) for how to model the covariances. A nominal choice is,

$$k(t,t') = \sigma^2 \exp\left[-\frac{1}{2\ell}(t-t')^2\right]$$

σ^2 and ell are called hyperparameters. Their values are determined by a typical optimization (e.g. conjugate gradient).

Gaussian Process Regression

In particular, the log-likelihood of obtaining the training set for a specific set of of hyperparameters is:

$$\log p(\mathbf{m}|T) = -\frac{1}{2}\mathbf{m}^\top \left(\mathbf{K} + u^2\mathbf{I}\right)^{-1} \mathbf{m} - \frac{1}{2}\log\left|\mathbf{K} + u^2\mathbf{I}\right| - \frac{n}{2}\log 2\pi$$

where T is the time grid, **m** is the training set, **K** is the matrix of all k(t, t’), and u is the uncertainty in the metric.

Note that this has a well-formed derivative, which allows the minimization of log(p) to use the Jacobian.

Gaussian Process Regression

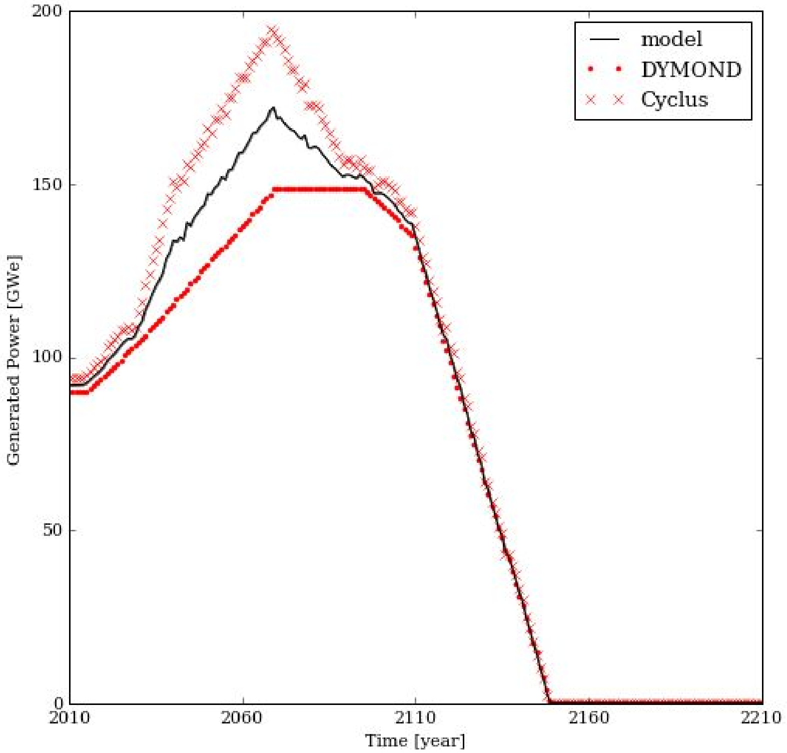
Once we have the hyperparameters, values for the model can be evaluated via the following:

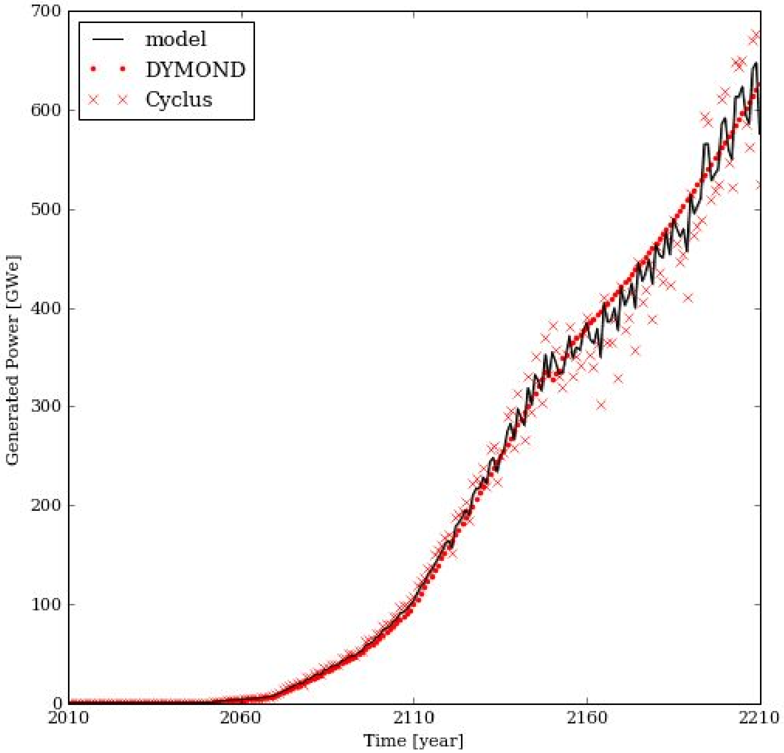
$$\mathbf{m}_*(\mathbf{t}_*) = \mathbf{k}_*^\top \left(\mathbf{K} + u^2\mathbf{I}\right)^{-1} \mathbf{m}$$

Note: Use a third party tool such as George or Scikit-Learn to handle this.

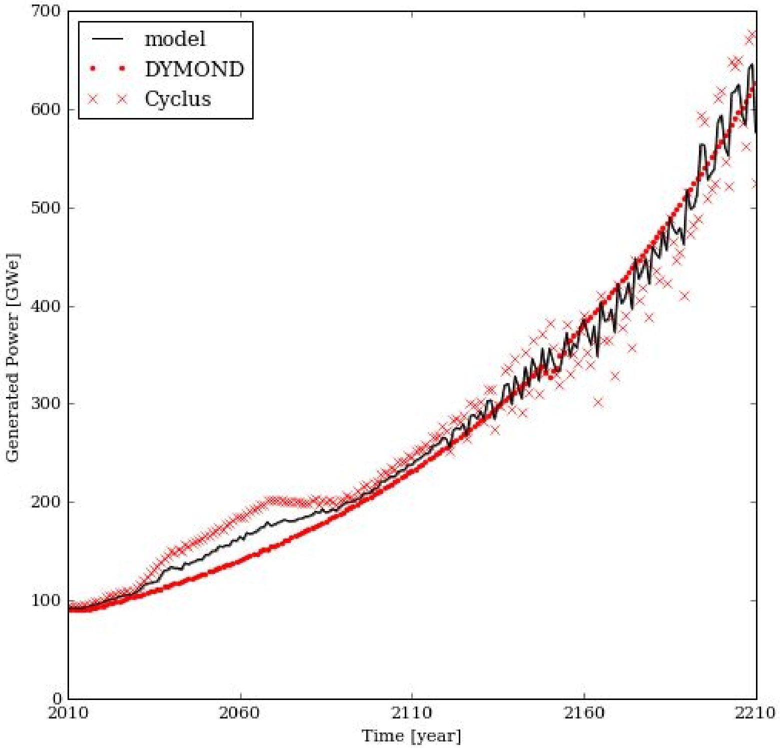
See *Rasmussen & Williams* for details: http://www.gaussianprocess.org/gpml/

GP of Features



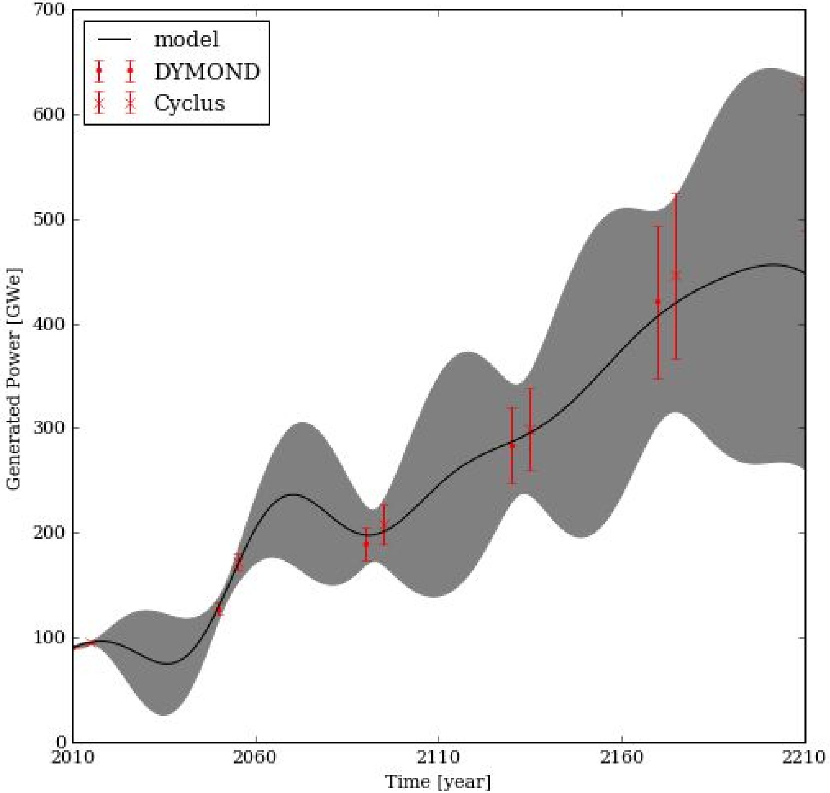


GP of Total



$$m_* \approx GP\left[\sum_i^I m_s^i\right] \neq \sum_i^I \mathcal{GP} m_s^i$$

Sparse GP Model with Uncertainty



Distance Between Two Models

Functional L^2 Norm

To compare two models, we must be able to establish the distance between them. This can be performed with the L^2 norm of the difference.

$$L^2(f,g) = \int_\Omega d\mathbf{x}\,\left(f(\mathbf{x}) - g(\mathbf{x})\right)^2$$

Using GP models instead of the ‘true’ functions, we can define a distance metric in terms of the models and their variances. This is our figure-of-merit.

$$d_{ij} = \int dt\left[k^i(t,t) + k^j(t,t) + \left(m_*^i(t) - m_*^j(t)\right)^2\right]$$

L^2 Model Distances

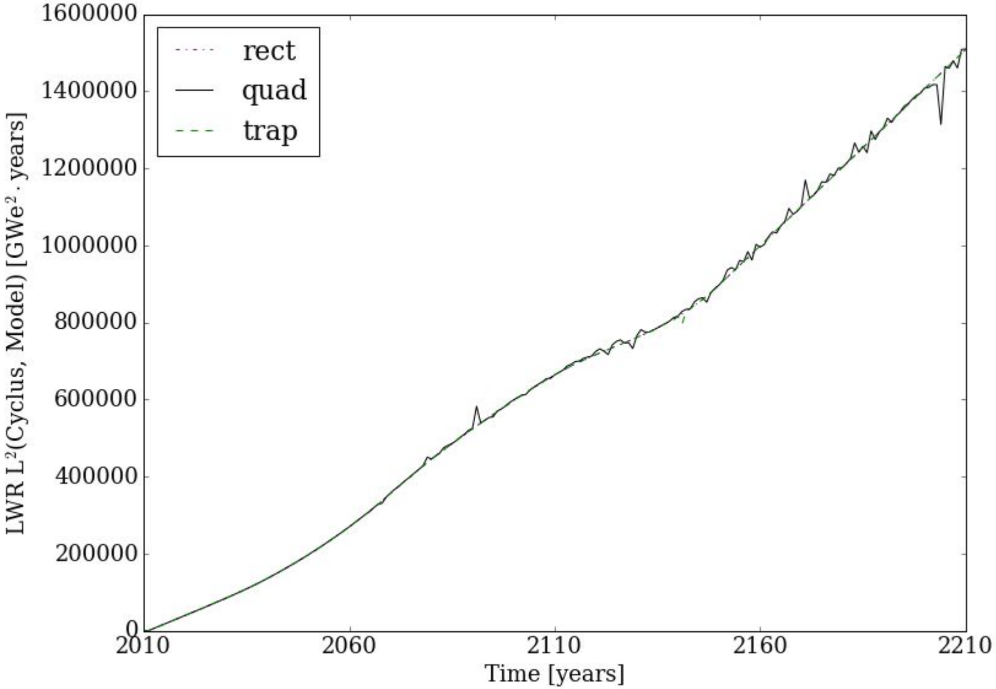
Distance of Gaussian process models for components to the full GP model:

	LWR	**FR**	**Total**
DYMOND	1.78e+04	3.27e+04	5.20e+04
Cyclus	1.52e+06	1.61e+07	5.58e+04

Integrator Comparison

Rectangular with fixed 1024 bins, adaptive quadrature, & adaptive trapezoidal

LWR Cyclus model to full model d_{ij}



Trapezoidal is the winner

Types of Comparison

What has been shown so far is sufficient for comparing any code to the model, and thus to other simulators in the study.

However, an inter-code comparison may also wish to compare different features for all codes. This may help point out how simulators differ and serve as a beacon for debugging.

Feature comparison is most useful if the features can be combined as a linear combination into a total, such as with generate power by reactor type.

To do so non-judgementaly, we compare the GP models of the features.

Contribution

Let’s start by defining the maximal distance as the distance from the total model to a zero everywhere curve (i.e. the time axis itself).

$$D(m_*^{\mathrm{Total}}) = \int dt \, \left[k^{\mathrm{Total}}(t,t) + m_*^{\mathrm{Total}}(t)\right]$$

Contribution

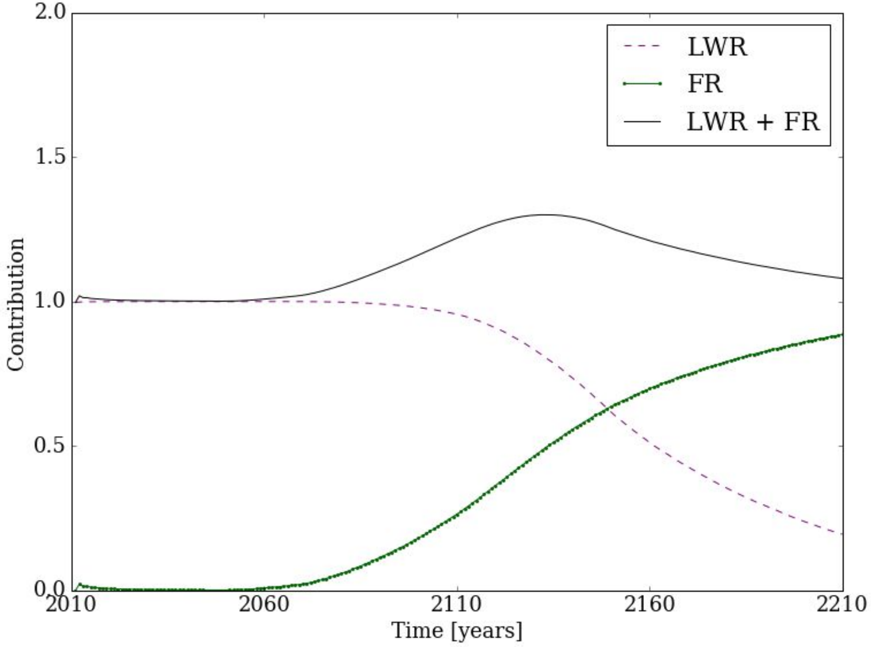
The contribution of the i-th feature is then one less the ratio of the DTW distance to the total model over D. This may also be normalized.

$$c^i = 1 - \frac{d(m_*^{\text{Total}}, m_*^i)}{D(m_*^{\text{Total}})}$$

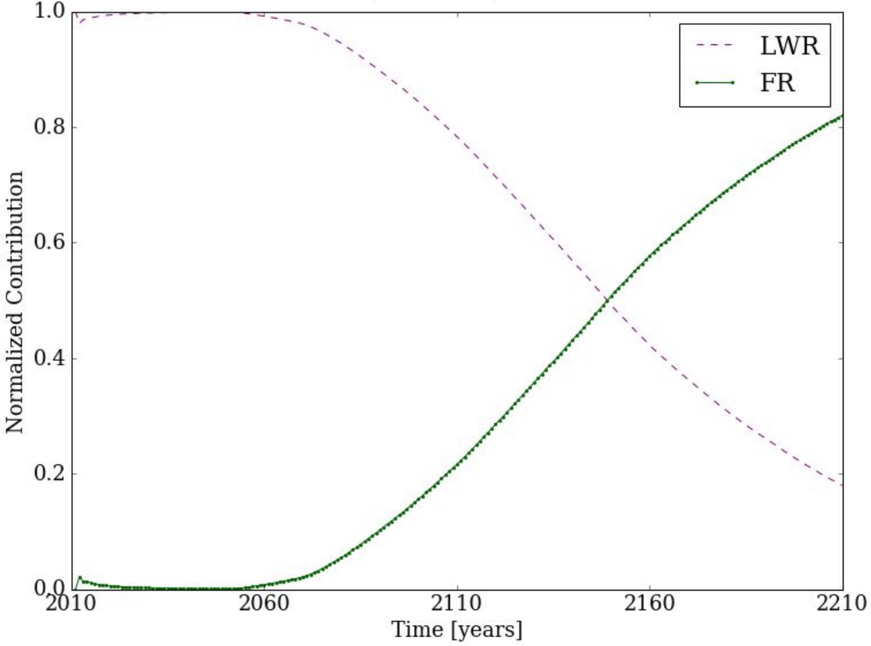
$$|c^i| = \frac{c^i}{\sum_j^I c^j}$$

Contribution

$$c^i$$



$$\left|c^i\right|$$



Dynamic Time Warping

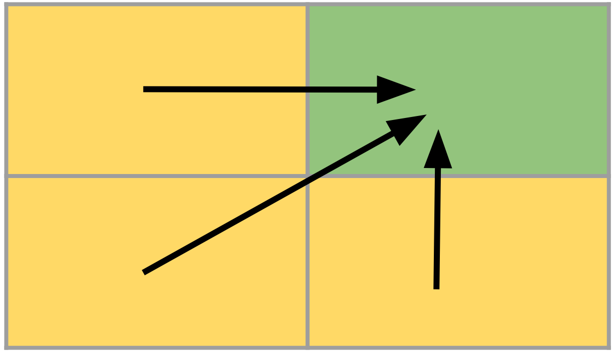
Dynamic Time Warping

DTW is a method for comparing any two time series - even if they are of different lengths. It provides a distance function that is computed from the cost that it would take to change (warp) one time series into another. For time series x and y of length A and B, say the distance between any two points is:

$$\Delta L_{a,b} = |x_a - y_b|_1$$

Dynamic Time Warping

Now, recursively define a cost matrix C:



$$C_{1,1} = \Delta L_{1,1}$$

$$C_{1,b+1} = \Delta L_{1,b} + C_{1,b}$$

$$C_{a+1,1} = \Delta L_{a,1} + C_{a,1}$$

$$C_{a+1,b+1} = \Delta L_{a,b} + \min\left[C_{a,b}, C_{a+1,b}, C_{a,b+1}\right]$$

Dynamic Time Warping

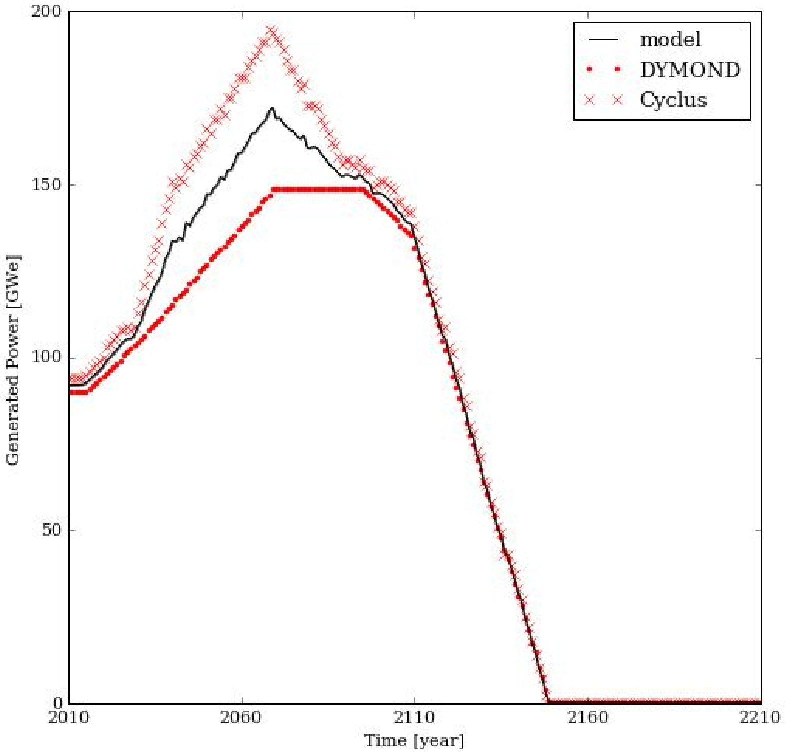
The distance is then defined as the last entry in the cost matrix divided by the maximal length of the warp path

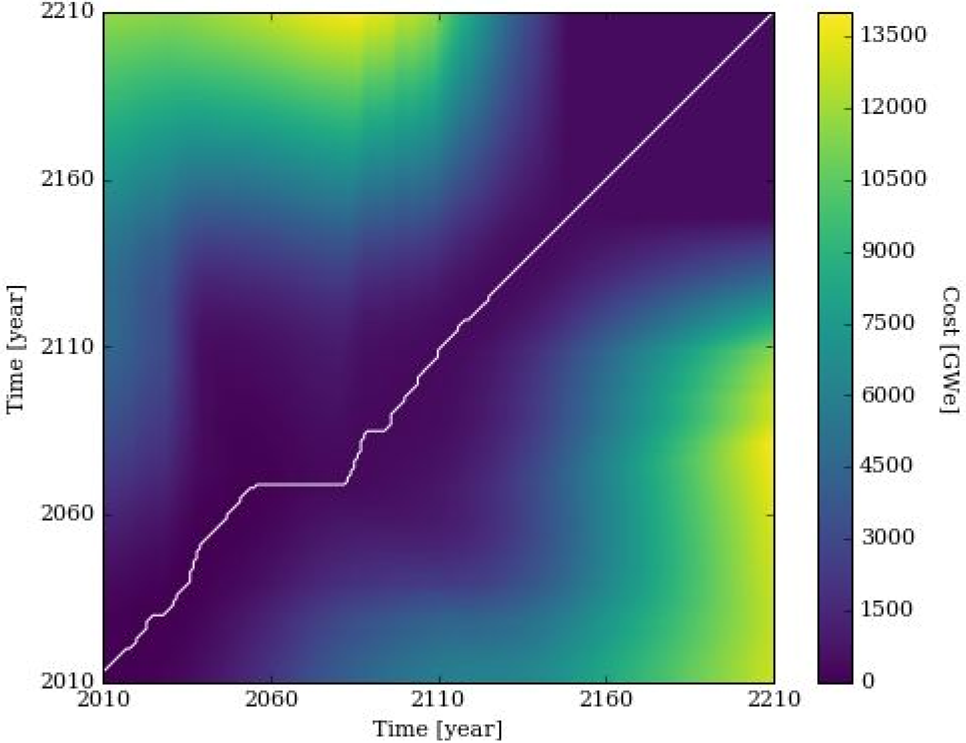
$$d = \frac{C_{A,B}}{A+B}$$

This is our figure-of-merit.

Dynamic Time Warping

LWR Model to LWR Cyclus





Dynamic Time Warping

Distances to Gaussian process model

	LWR	**FR**	**Total**
DYMOND	1.452	2.783	3.022
Cyclus	1.053	3.732	3.984

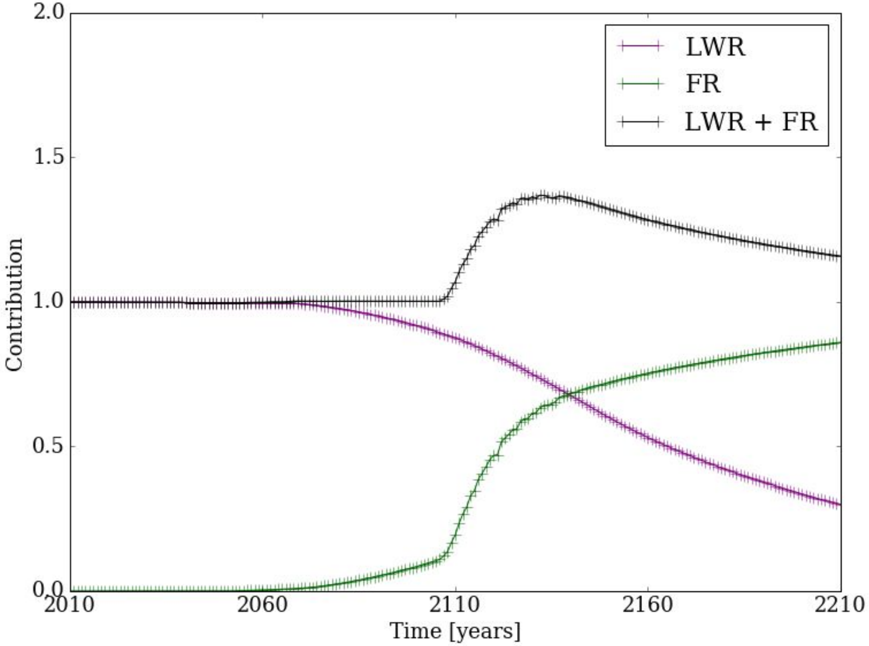
DTW Contribution

For DTW, the maximal distance as the distance from the total model to a zero everywhere curve is as follows:

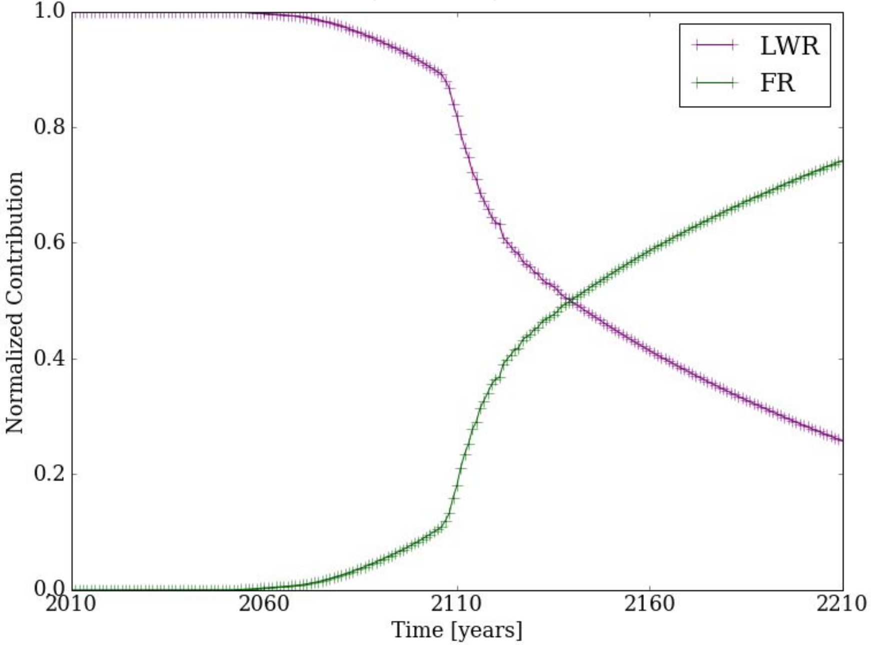
$$D(m_*^{\mathrm{Total}}) = \frac{\left|m_*^{\mathrm{Total}}\right|_1}{2A}$$

DTW Contribution

$$c^i$$



$$\left|c^i\right|$$



Comparison

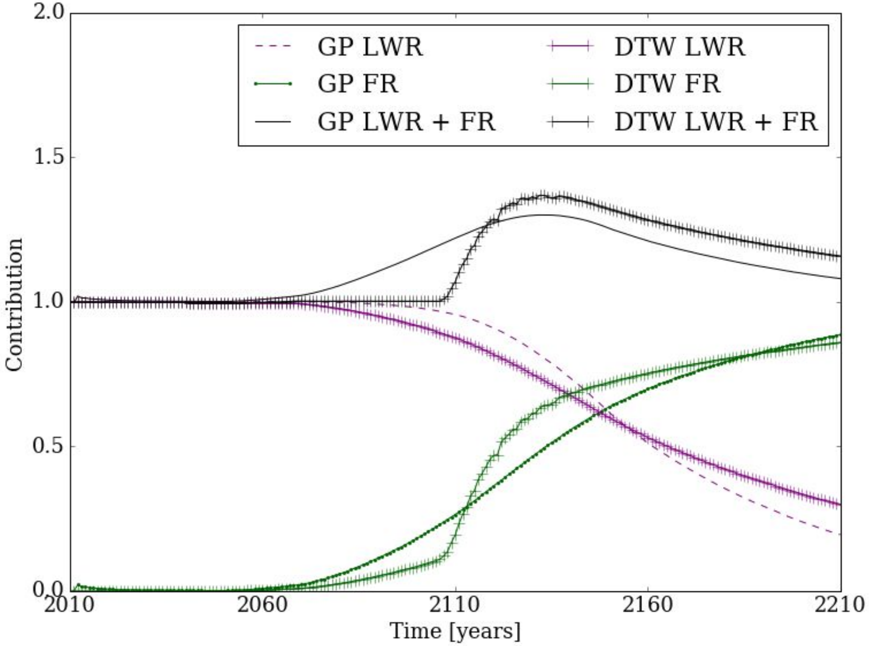
Distance Comparison

Distance of Gaussian process models for components to the full GP model:

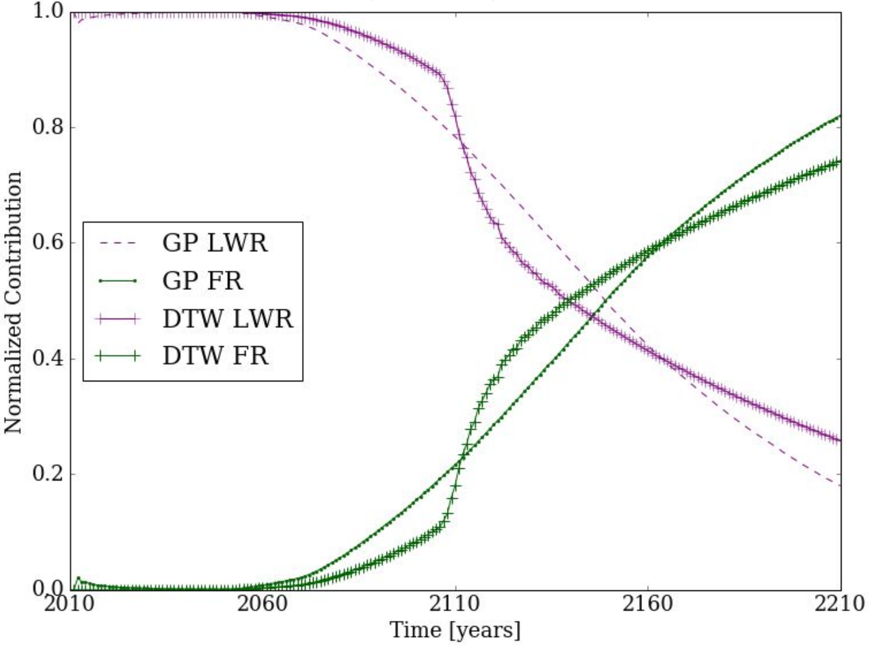
	LWR	**FR**	**Total**
GP DYMOND	1.78e+04	3.27e+04	5.20e+04
GP Cyclus	1.52e+06	1.61e+07	5.58e+04
DTW DYMOND	1.452	2.783	3.022
DTW Cyclus	1.053	3.732	3.984

Contribution Comparison

$$c^i$$



$$|c^i|$$



Conclusions

- May handle as many simulators as desired.
- Does not choose a “correct” simulator.
- Method is easily extensible to higher dimensional constructs
- Similar ideas could be used for non-time series data. The notion of a cost matrix as mechanism to create a figure-of-merit is useful in other areas.
- A variety of kernel functions have yet to be explored.
- Can handle comparisons for simulators with different time grids.
- Incorporates model uncertainties.
- Do not filter the data (see supplemental).
- L^2 norm of GPs better at separating functions, comes at cost of odd units.
- No edge effects with DTW.

Future Work

- Use functional L^1 norm and/or use
- Deployment schedule optimization using presented tools.
- More fuel cycle data needed!

Acknowledgments

Special thanks to Dr. Carlo Graziani of the FLASH Center at the University of Chicago for his efforts on L^2 norm of difference of Gaussian processes.

Special thanks to Dr. Bo Feng for providing the DYMOND data.

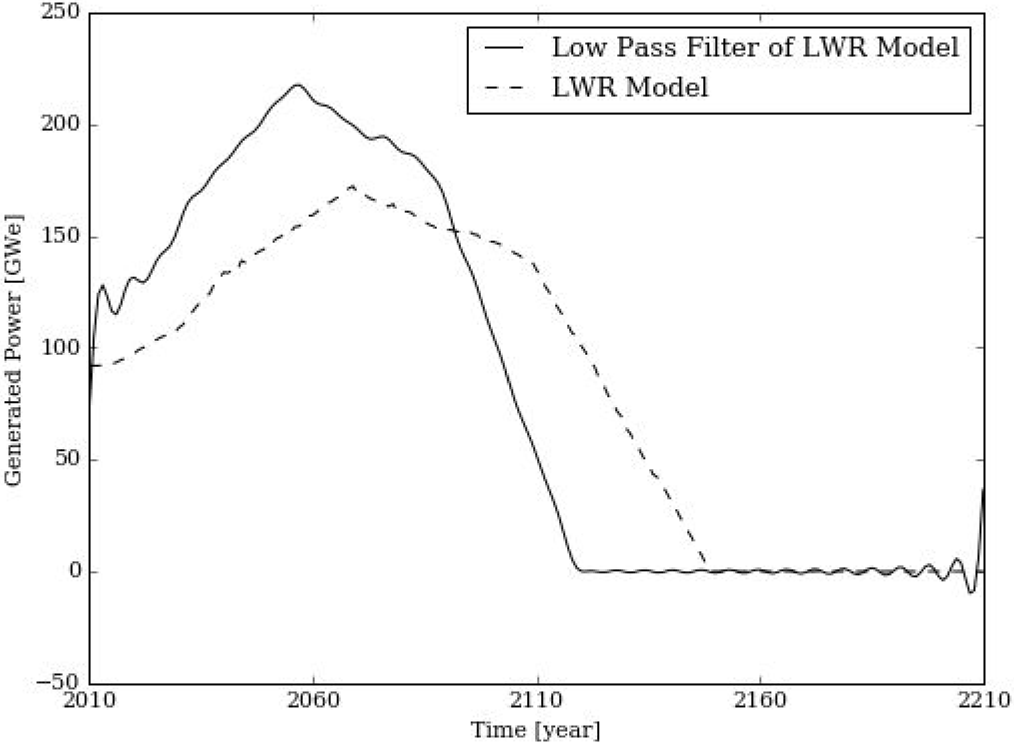
Supplemental

Terminology

~~Benchmark~~

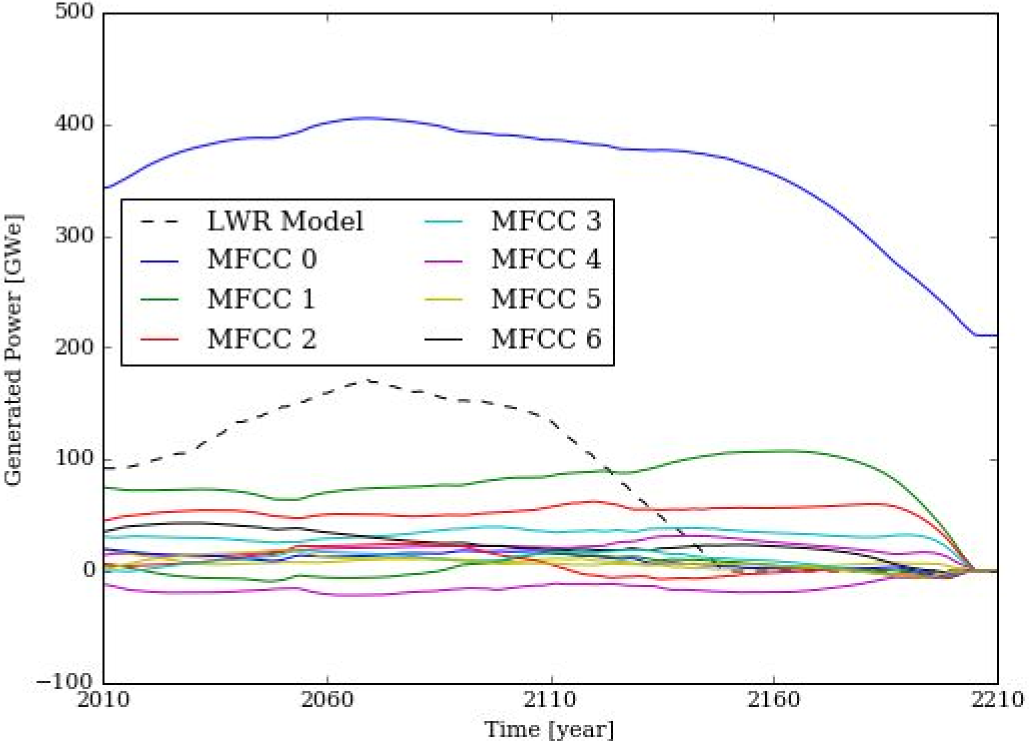
Inter-code Comparison

Filtering with FFT is a Bad Idea



- Low pass filter with Used 256 channel FFT, keeping lowest 32 channels.
- Peak has wrong amplitude and location.
- Added sinusoidal behaviour at end points and to the flat region at large times.

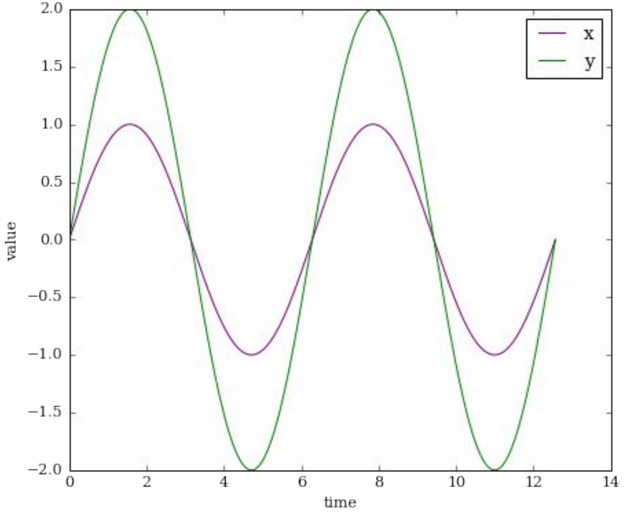
Filtering with MFCCs is a Bad Idea



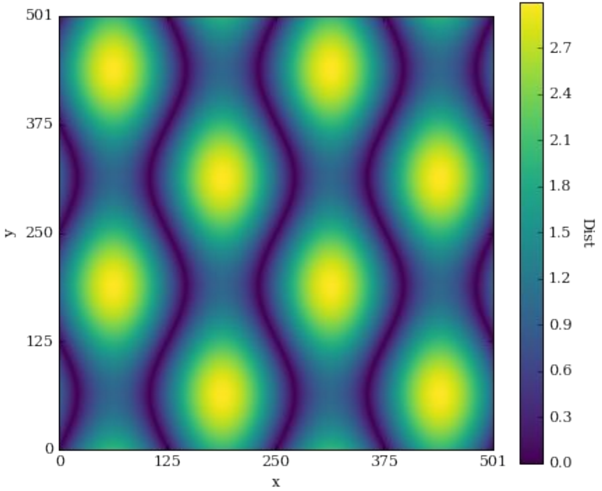
- Uses 13 mel frequency cepstral coefficients.
- Only zero coefficient has right-ish shape, but wrong amplitude.
- Other MFCCs are wrong, some are negative.
- Meant for high frequency data, which this is not.

DTW Intuition - Sine with Two Amplitudes

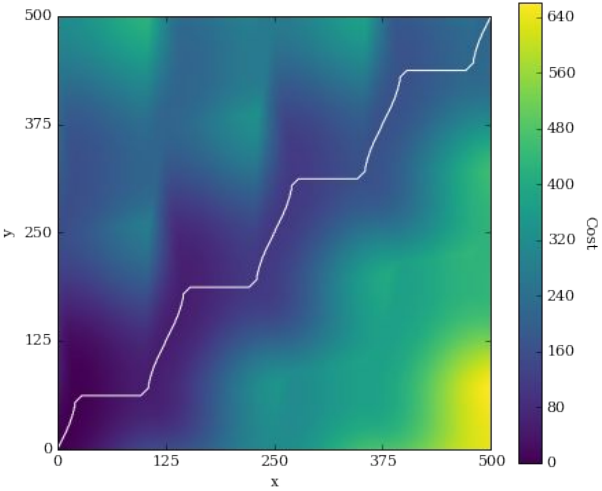
Some people have had a hard time visualizing what ΔL and DTW costs are doing. This is a gallery of examples, containing the following sine example: https://github.com/scopatz/fc-bench-method/blob/master/dtw-intuition.ipynb



Time Series

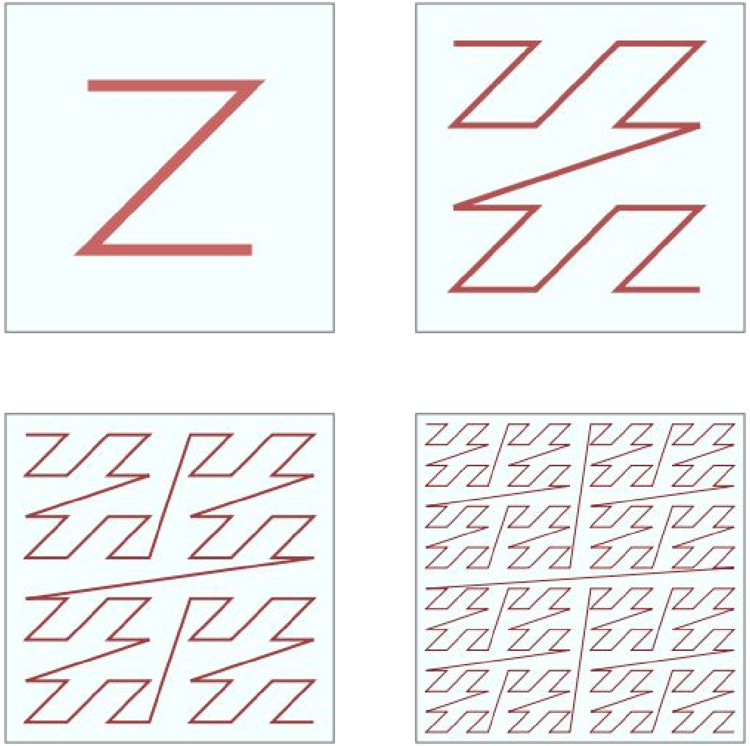


L1 Norm



Cost

Spatial Warping Idea: AMR



- The DTW method can effectively be used without modification.
- Sort based on Morton order and set A and B as the lengths of the curves.
- ΔL calculation remains the same.

Spatial Warping Idea: Unstructured Meshes

First sort the points in each mesh individually according to some canonical ordering of axes.

Add cost to points that are far away from each other spatially by modifying ΔL:

$$\Delta L_{a,b} = |x_a - y_b|_1 \cdot \left(1 + \frac{||r_a - r_b||}{\max\left(||R_A||, ||R_B||\right)}\right)$$

r_a and r_b are the locations of the a^th and b^th points in the A and B meshes.

$||R_A||$ and $||R_B||$ represent the maximum distance possible in the meshes.