Geofeasibility scores

Team leader:

Mark Coates, McGill University

Team members:

Nando de Freitas, University of British Columbia

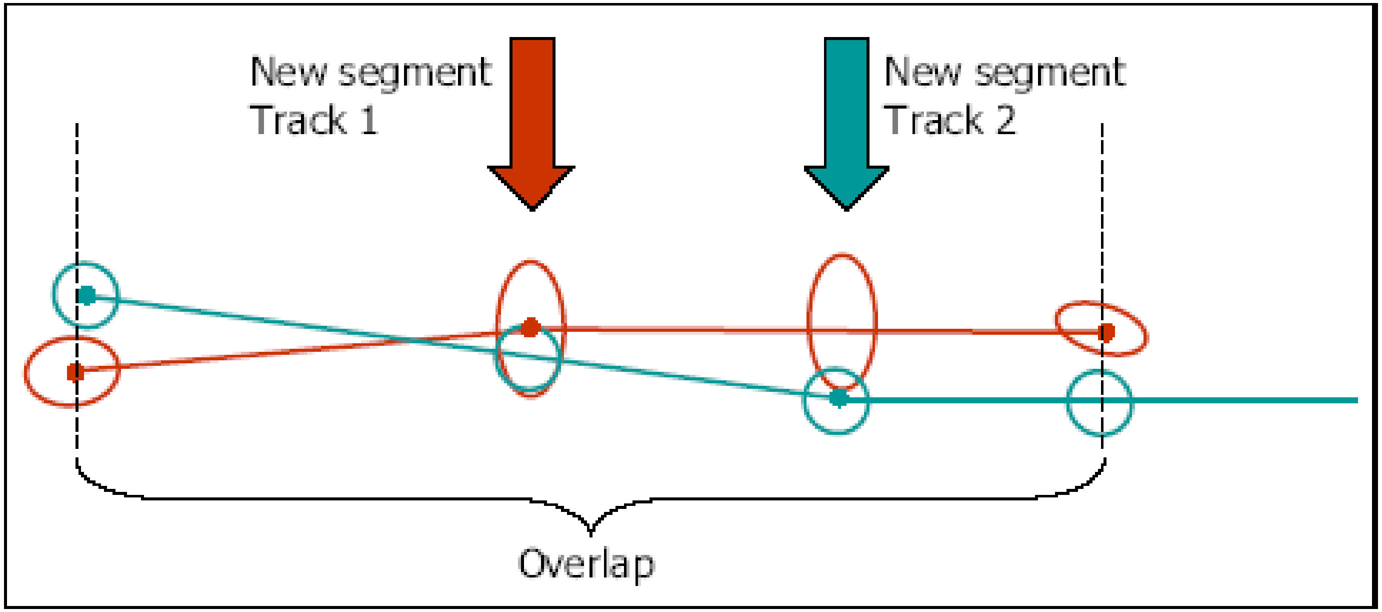
Francis Moreau, Université de Montréal

Boris Oreshkin, McGill University

Mary Pugh, University of Toronto

Problem Statement

- For each report, the source associates an area of uncertainty (AOU) of elliptic shape delimiting a 2σ probability area



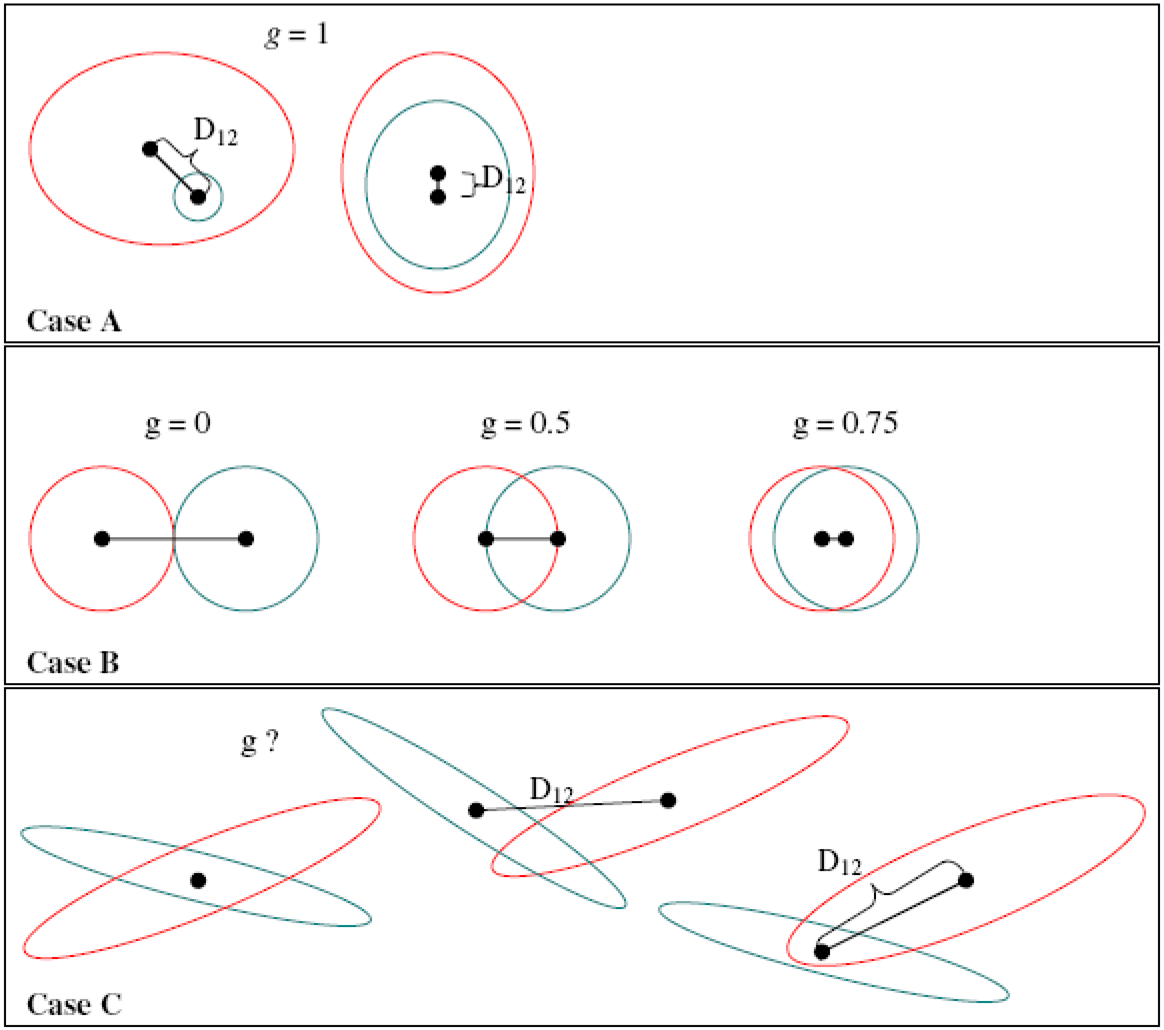
Problem Statement

- We need to define an optimal geo-feasibility score g, to quantify the overlap of two AOU.

- Here are three basic rules to define the function g:

- $g \in [0,1]$
- If ellipses do not touch, $g = 0$
- If ellipses totally overlap, $g = 1$

Problem Statement



- The geo-feasibility score has to be suitable for very elongated ellipses as well as for circles
- Reasonable approximations are possible (e.g. approximating an ellipse to a circle is NOT a reasonable approximation)

Candidate Metrics

- **The normalized area of overlap**
  - The area of overlap is normalized by the area of the smaller ellipse
  - Corresponds to human operator intuition.

Candidate Metrics

- **Statistical Approaches**

- Integrated product of two Gaussian distributions
  - Corresponds to Bayes factor.
- Symmetric KL (Kullback-Leibler) divergence
  - Distance between two distributions.
- Generalized Likelihood Ratio (GLR)
  - Find most likely position of a single boat, evaluate the likelihood that it generated the reported distributions.

Challenges

- Closed form analytical calculation of overlap area is not possible

- Numerical methods based on optimization are not fast enough

Proposed solutions

- Newton’s method to find intersection points and analytical approximation to the normalized area of the overlap
- Monte-Carlo integration to find the normalized area of overlap
- Generalized Likelihood Ratio gives a fast and meaningful approximation to the operator’s intuition

Analytical Method

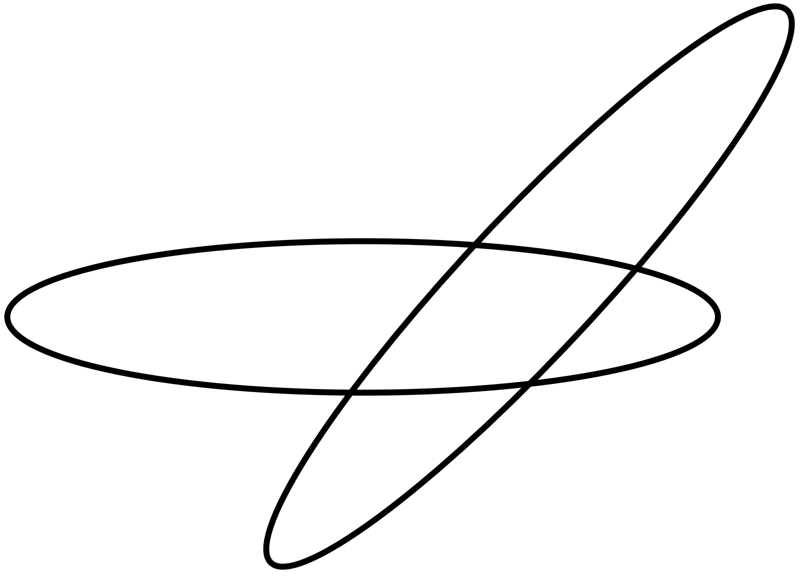
2 steps:

- Find the intersection points
  - -Too hard to find analytically
- Calculate the area
  - -Using integration in polar coordinates

Finding points of Intersection

- Using Newton's Method

  -In-and-out method for starting points



Calculating area

3 cases:

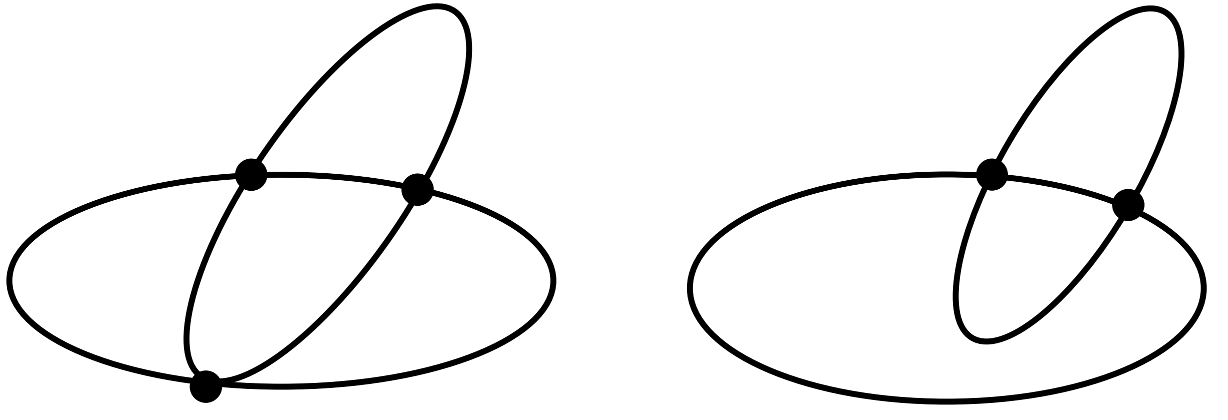
- 0 or 1 points of intersection

  -Area is zero or is equal to the area of the smaller ellipse

- 2 or 3 points of intersection
- 4 points of intersection

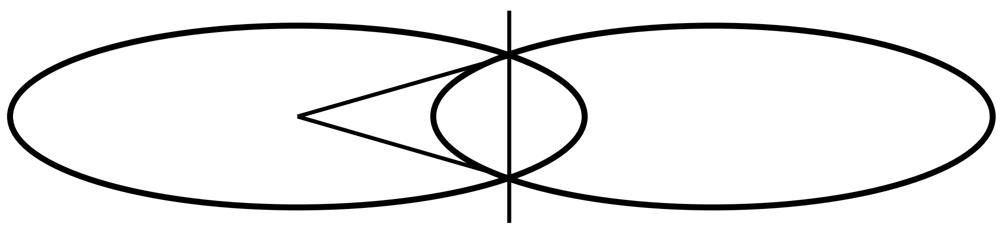
2 or 3 points of intersection

- Same case considering in-and-out technique

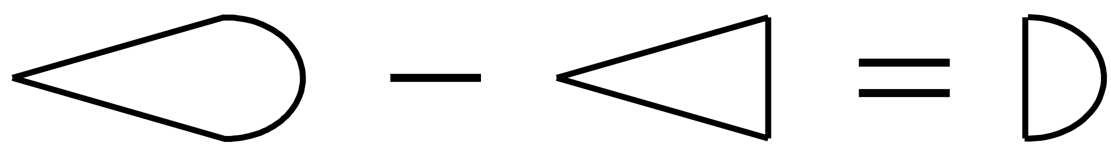


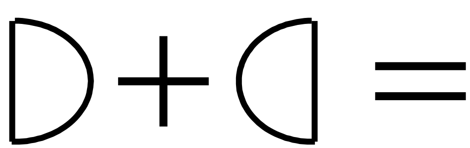
For 2 points

- Using polar coordinates



Area of ellipse’s portion – Area of the triangle



 Area of intersection

Calculation

- Ellipse in polar coordinates (R = radius)

$$R(\theta) = \frac{ab}{\sqrt{a^2 + \left(b^2 - a^2\right)\cos^2(\theta)}}$$

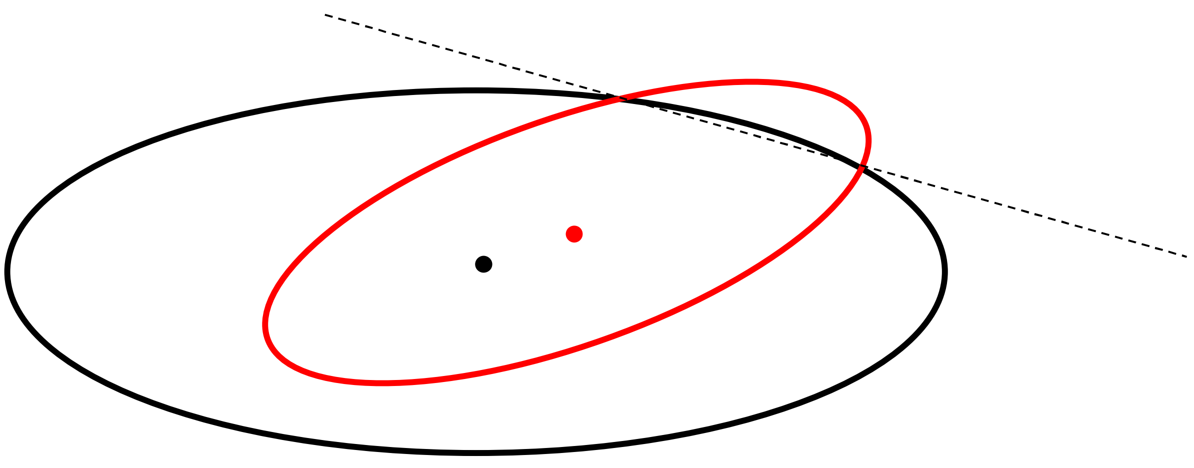
- Integral

$$\int_{\theta_1}^{\theta_2} \frac{R^2(\theta)}{2} d\theta = \left[\frac{ab}{2} \tan^{-1}\left(\frac{a \tan(\theta)}{b} \right) \right]_{\theta_1}^{\theta_2}$$

Be careful…

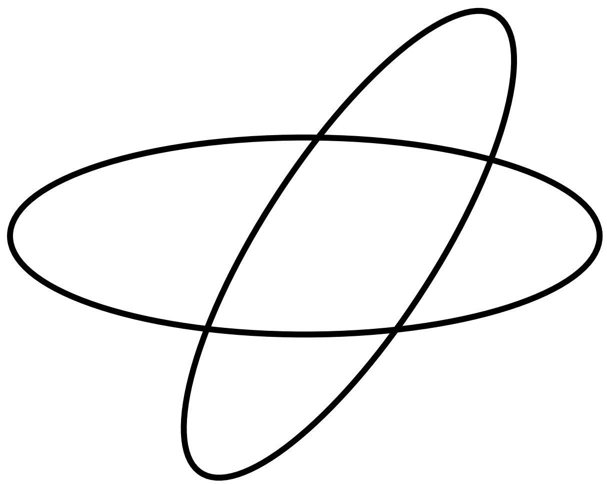
- The integral uses inverse tangent function
  - $\theta \in \left(\frac{-\pi}{2}, \frac{\pi}{2} \right)$

- Too near ellipses center

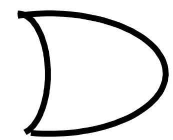


4 points of intersection

- Extension of the 2 points case



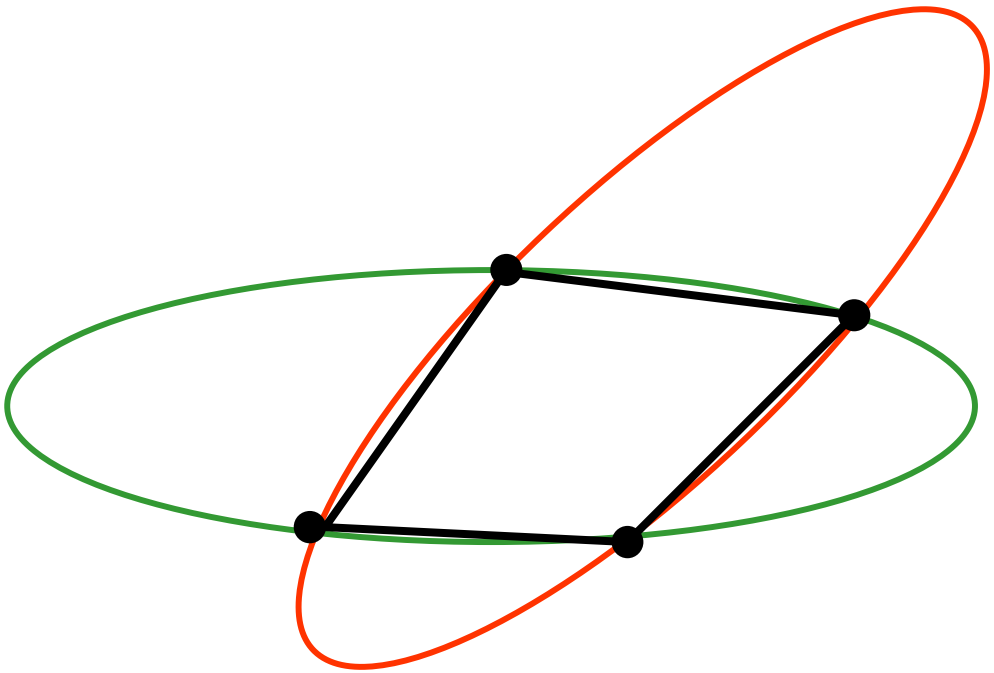
$$E_i = \text{Area of ellipse i}$$

$a_i =$ 

$$\text{Area of intersection} = \frac{E_1 + E_2 - \sum_{i=1}^{4} a_i}{2}$$

Hard to code

- Approximation by a four sided object



Performance

- The slow part of the program is if the ellipses actually intersect

- for 100,000 pairs of ellipses that intersect about 45% of the time, it takes between 640 and 710 seconds.

Monte Carlo integration

- The algorithm computes an estimate of a multidimensional integral

$$I = \iiint\limits_{x,y,z,\ldots} f(x, y, z, \ldots)\, dxdydz\ldots$$

- Naïve algorithm draws samples $(x_{i,} y_{i,} z_{i, \ldots})$ uniformly from the integration area and estimates its value as follows

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i, z_i, \ldots)$$

Monte Carlo integration

- However, in many cases it is beneficial to draw samples from some pdf $p(x,y,z,\ldots)$

$$I = \iiint\limits_{x,y,z,\ldots} \frac{f(x,y,z,\ldots)}{p(x,y,z,\ldots)} p(x,y,z,\ldots) dxdydz\ldots$$

- And use the so called **importance sampling**

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i, y_i, z_i, \ldots)}{p(x_i, y_i, z_i, \ldots)}$$

Monte Carlo integration

- **Normalized intersection area** of two ellipses defined by the regions S_1 and S_2

$$g_a = \frac{\int 1_{\mathbf{x} \in S_1 \cap S_2} \, d\mathbf{x}}{\int 1_{\mathbf{x} \in \min(S_1, S_2)} \, d\mathbf{x}}$$

- Can be estimated using Monte Carlo with importance sampling

$$g_a = \frac{\sum_{i=1}^{N} \frac{1}{p(\mathbf{x}_i)} 1_{\mathbf{x}_i \in S_1 \cap S_2}}{\sum_{i=1}^{N} \frac{1}{p(\mathbf{x}_i)} 1_{\mathbf{x}_i \in \min(S_1, S_2)}}$$

Monte Carlo integration

- **Integrated product of two Gaussian distributions**

$$g_p = \int_{\mathbf{x}} p_1(\mathbf{x} \mid \Theta_1)\, p_2(\mathbf{x} \mid \Theta_2)\, d\mathbf{x}$$

- Can be estimated using Monte Carlo even without importance sampling

$$\hat{g}_p = \frac{1}{N} \sum_{i=1}^{N/2} p_1(\mathbf{x}_i^2 \mid \Theta_1) + \frac{1}{N} \sum_{i=1}^{N/2} p_2(\mathbf{x}_i^1 \mid \Theta_2)$$

Generalized Likelihood Ratio

- We assume that a source provides us with measurements of target positions and Maximum Likelihood (ML) estimators of covariance matrices

$$\mathbf{y}_1, \quad \hat{\mathbf{R}}_1, \quad \mathbf{y}_2, \quad \hat{\mathbf{R}}_2$$

- And we choose to construct a test statistic to discriminate between the two hypotheses

$$H_1: \qquad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$H_0: \qquad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Generalized Likelihood Ratio

- In this case Uniformly Most Powerful test does not exist.
- However, we can resort to a suboptimum statistic that is called GLR

$$\Lambda(\mathbf{y}_1, \mathbf{y}_2) = \frac{\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 \mid \Theta, H_1)}{\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 \mid \Theta, H_0)}$$

- Here Θ stands for all the unknown parameters

Generalized Likelihood Ratio

- Given the Gaussian and independence assumptions we have

$$p(\mathbf{y}_1, \mathbf{y}_2 \mid \Theta, H) = \frac{1}{2\pi\sqrt{\det(\mathbf{R}_1)\det(\mathbf{R}_2)}} \exp$$

$$\left[-\frac{1}{2}(\mathbf{y}_1 - \boldsymbol{\mu}_1)^T \mathbf{R}_1^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{y}_2 - \boldsymbol{\mu}_2)^T \mathbf{R}_2^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2) \right]$$

- We deduce immediately that

$$\max_{\Theta} p(\mathbf{y}_1, \mathbf{y}_2 \mid \Theta, H_0) = \frac{1}{2\pi\sqrt{\det(\hat{\mathbf{R}}_1)\det(\hat{\mathbf{R}}_2)}}$$

Generalized Likelihood Ratio

- To find the ML estimator of the mean under H_1 we use $\mu_1 = \mu_2 = \mu$ and equate partial derivative to 0:

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln p(\mathbf{y}_1, \mathbf{y}_2 \mid \Theta, H_1) = \mathbf{R}_1^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}) + \mathbf{R}_2^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}) = \overline{0}$$

- Which results in the following expression for the ML estimator of the mean

$$\hat{\boldsymbol{\mu}} = \left[\mathbf{R}_1^{-1} + \mathbf{R}_2^{-1}\right]^{-1}\left[\mathbf{R}_1^{-1}\mathbf{y}_1 + \mathbf{R}_2^{-1}\mathbf{y}_2\right]$$

Generalized Likelihood Ratio

- Substituting this estimator into likelihood ratio we get statistic of the form

$$\Lambda(\mathbf{y}_1, \mathbf{y}_2) = \exp\left[-\frac{1}{2}\Delta^T \hat{\mathbf{R}}_\Delta^{-1} \Delta\right]$$

- Where Δ is the difference between measurements:

$$\Delta = \mathbf{y}_1 - \mathbf{y}_2$$

Generalized Likelihood Ratio

- And the estimate of the covariance inverse has the following "nice" expression

$$\hat{\mathbf{R}}_{\Delta}^{-1} = \hat{\mathbf{R}}_{2}^{-1}\left[\hat{\mathbf{R}}_{1}^{-1} + \hat{\mathbf{R}}_{2}^{-1}\right]^{-1}\hat{\mathbf{R}}_{1}^{-1}\left[\hat{\mathbf{R}}_{1}^{-1} + \hat{\mathbf{R}}_{2}^{-1}\right]^{-1}\hat{\mathbf{R}}_{2}^{-1} + $$

$$+\hat{\mathbf{R}}_{1}^{-1}\left[\hat{\mathbf{R}}_{1}^{-1} + \hat{\mathbf{R}}_{2}^{-1}\right]^{-1}\hat{\mathbf{R}}_{2}^{-1}\left[\hat{\mathbf{R}}_{1}^{-1} + \hat{\mathbf{R}}_{2}^{-1}\right]^{-1}\hat{\mathbf{R}}_{1}^{-1}$$

- However, using the rule "the inverse of the product is equal to the product of inverses in reversed order" we can show that

$$\hat{\mathbf{R}}_{\Delta}^{-1} = \left[\hat{\mathbf{R}}_{1} + \hat{\mathbf{R}}_{2}\right]^{-1}$$

Generalized Likelihood Ratio

- Thus the geofeasibility score based on GLR admits the following simple and intuitive form

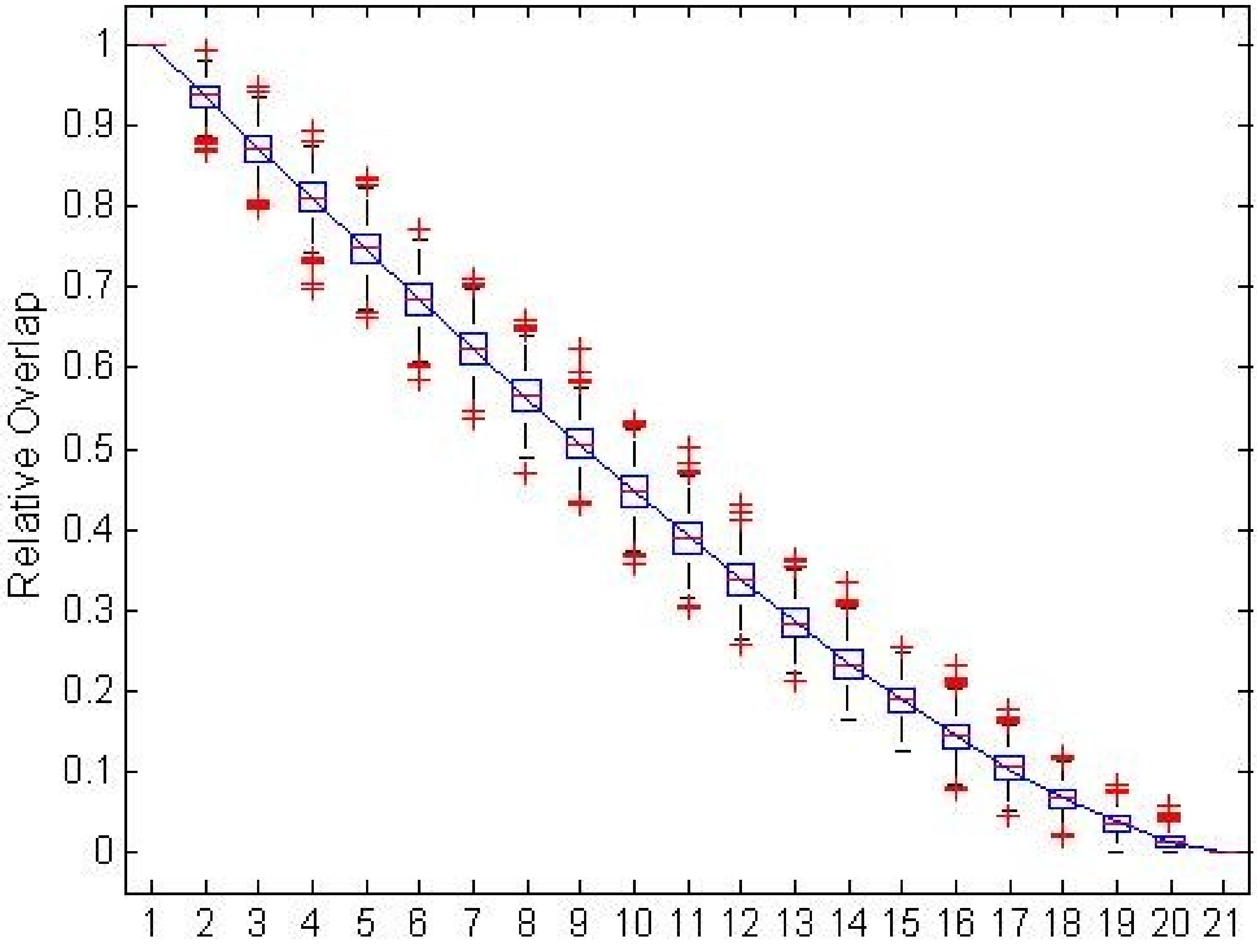
$$g_G \equiv \Lambda(\mathbf{y}_1, \mathbf{y}_2)$$

$$g_G = \exp\left[-\frac{1}{2}\Delta^T\left[\hat{\mathbf{R}}_1 + \hat{\mathbf{R}}_2\right]^{-1}\Delta\right]$$

Simulation results

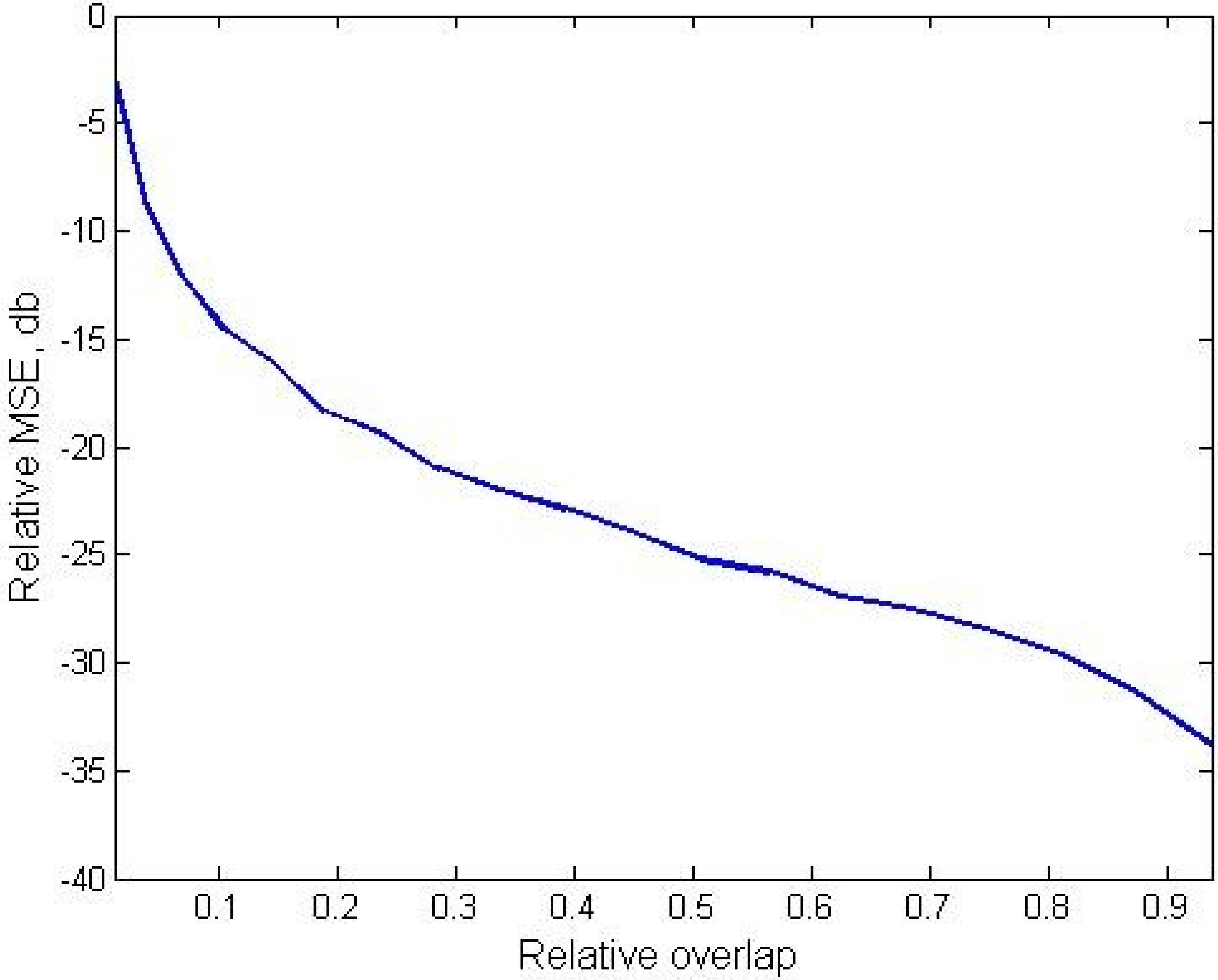
Error bars, Monte-Carlo integration of the overlap area

N=500



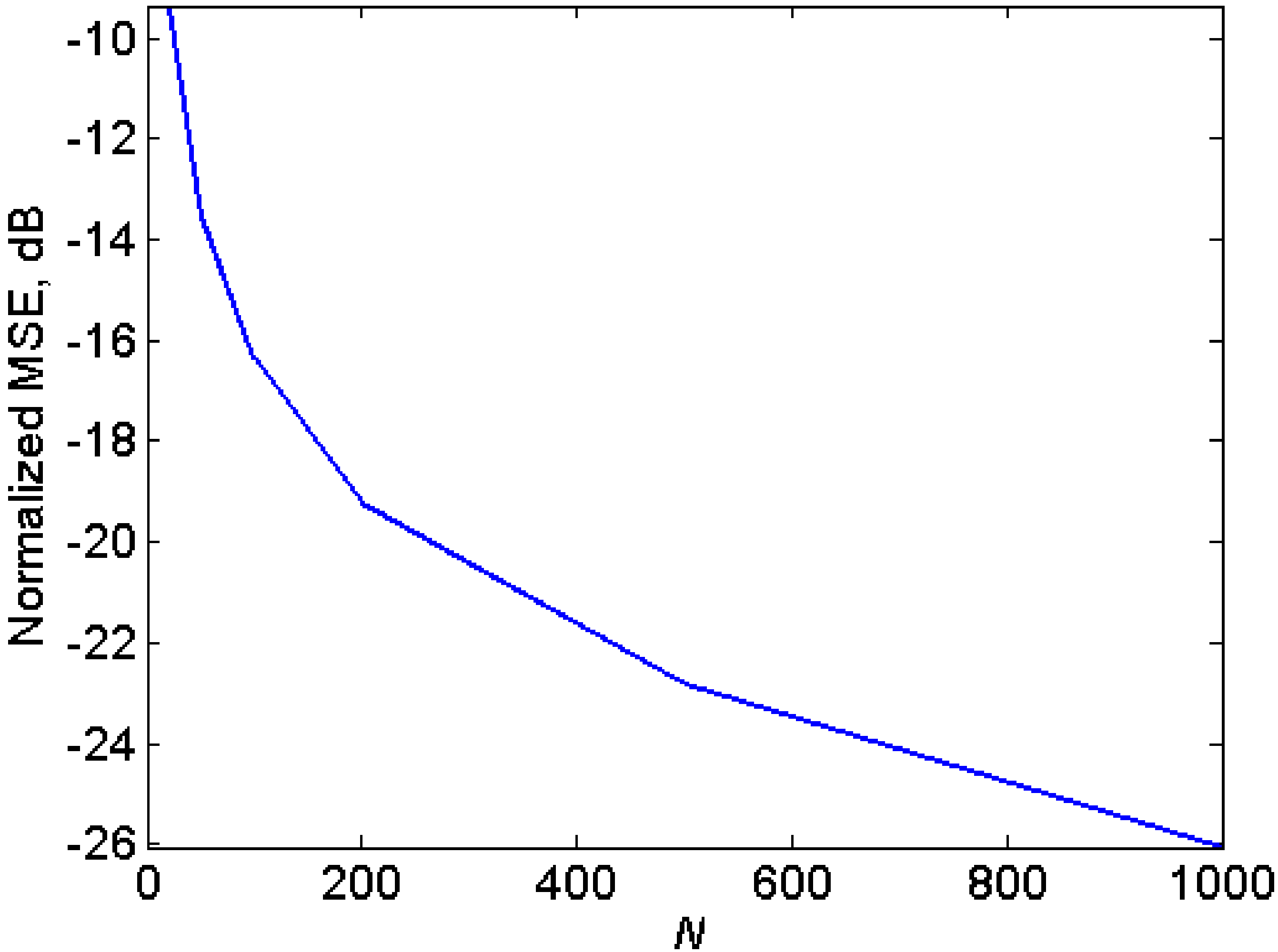
Simulation results

Monte Carlo Mean Squared Error, N=500



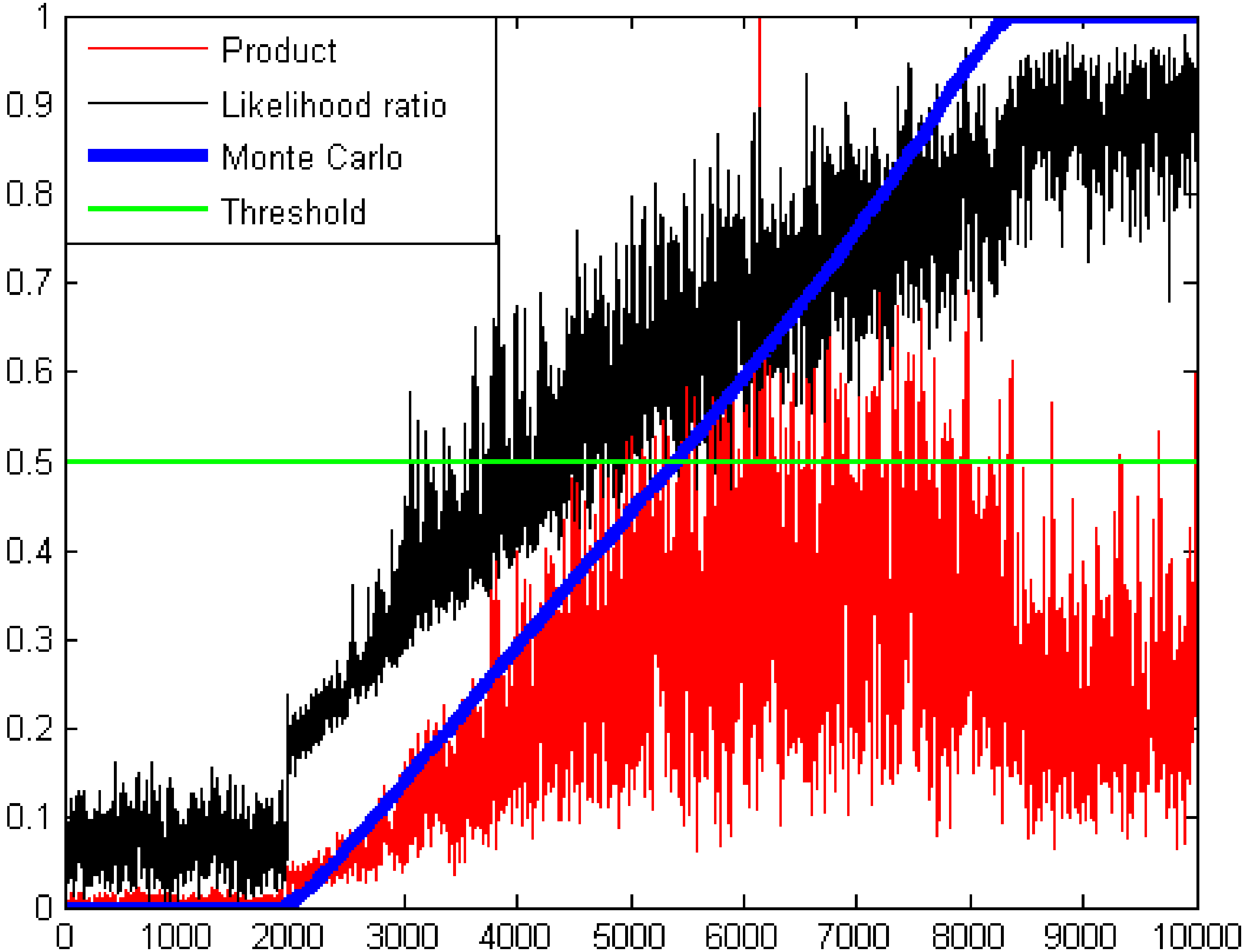
Simulation results

Monte-Carlo error for a fixed value of relative overlap area equal to 0.391



Simulation results

Comparison of statistics

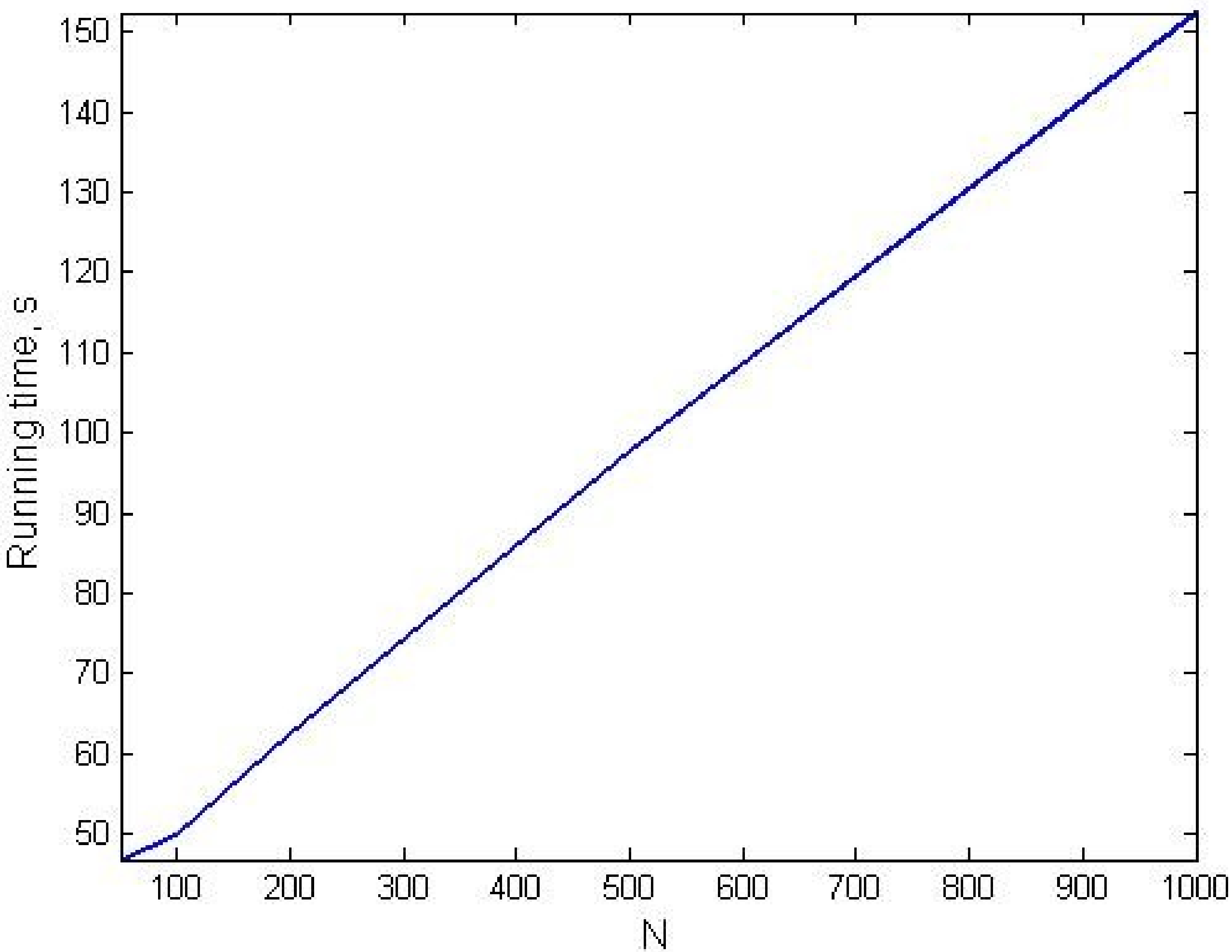


Simulation results

Calculation time for 100,000 evaluations, Monte-Carlo.

Calculation time for GLR is 4 seconds

(0.04 ms/evaluation)



Questions