EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Comparing two location estimates of the same seismo-acoustic event
A survey of available statistical tools

R. Le Bras and E. Qorbani, CTBTO, Vienna, Austria

Ronan.LeBras@ctbto.org
Ehsan.Qorbani@ctbto.org

Outline

- Background
- Objectives
- Method – The Hellinger and Bhattacharyya distances
- Summary

Background

- Estimates of a seismic event location are usually obtained through classical least squares inversion methods yielding an epicenter, depth, and time estimate and in addition a covariance matrix that defines a probability distribution in the parametric space of these variables.

- There are many intuitive ways of measuring the difference between two estimates of the same seismic event location. One can think of simply looking at each element of the location: epicenter, depth, origin time, error ellipses, magnitude and treat each element separately. This would mean calculating the distance between the two epicenter estimates, the difference in depth estimate, the time difference between the origin time estimates, the overlap percentage of the error ellipses, the difference in magnitude estimates.

- This work proposes to apply two different methods of that a measure of the distance between two events taking into account the complete representation of a probability distribution.

Objective

Provide a method to estimate the difference between two seismic location estimates in a comprehensive way, taking into account not only the estimates of the location, depth, and origin time, but also the error estimates included in the covariance matrix that define a more complete probability distribution.

Method – The Hellinger distance

Given two probability distributions ***f*** and ***g***, the Hellinger distance is defined as:

$$D_H^2(\boldsymbol{f},\boldsymbol{g}) = \frac{1}{2}\int\left(\sqrt{\boldsymbol{f}(x)} - \sqrt{\boldsymbol{g}(x)}\right)^2 dx$$

$$D_H^2(\boldsymbol{f},\boldsymbol{g}) = 1 - \int \sqrt{\boldsymbol{f}(x)\boldsymbol{g}(x)}\, dx$$

where: $0 \leq DH \leq 1$ *since* ***f*** *and* ***g*** *are PDFs*

Hellinger, Ernst (1909), "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen", Journal für die reine und angewandte Mathematik (in German), 136: 210–271

Method – The Bhattacharyya distance

Given two probability distributions ***f*** and ***g*** , the Bhattacharyya distance is defined as:

$$\mathrm{D_B}\,(\boldsymbol{f},\boldsymbol{g}) = -\ln\ (BC(\boldsymbol{f},\boldsymbol{g}))$$

$$BC\,(\boldsymbol{f},\boldsymbol{g}) = \int \sqrt{\boldsymbol{f}(x)\boldsymbol{g}(x)}\,dx$$

$$where:\ \ 0 \leq D_B \leq\ \infty$$

where BC is the Bhattacharyya coefficient. Note the relationship between the Hollinger and Bhattacharyya distances:

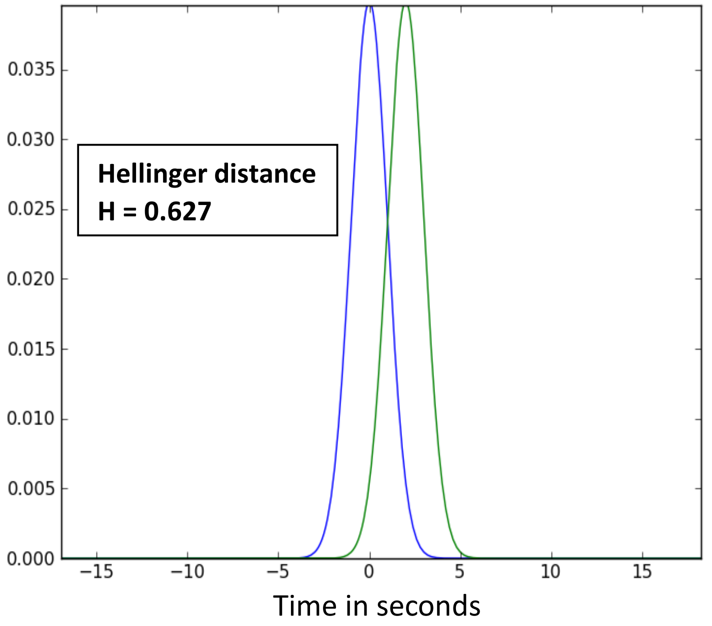
$$D_B = -\ln(1 - D_H)$$

Bhattacharyya, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions". Bulletin of the Calcutta Mathematical Society. 35: 99–109

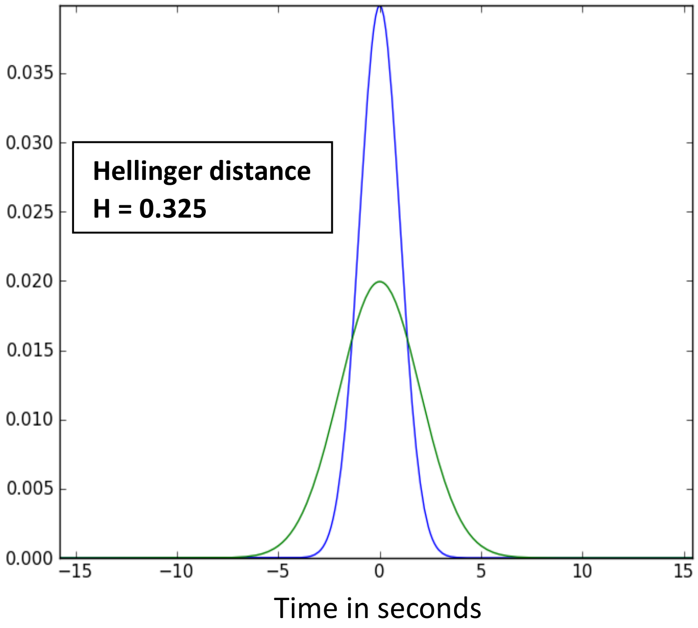
Method – The Hellinger distance

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Illustration on comparing two one-dimensional Gaussian distributions for origin time



Distance for estimates which have different averages (0s and 2s) and the same standard deviation.

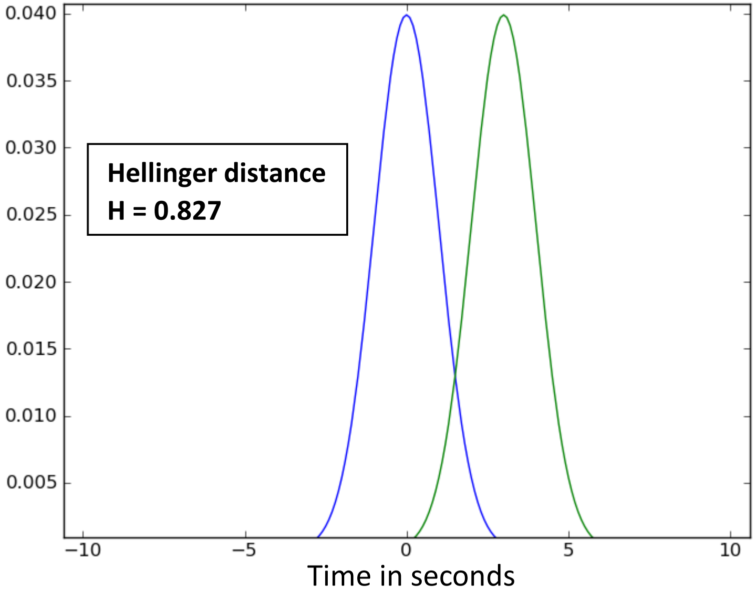


Distance for estimates which have the same average (0s) and two different standard deviations (1s and 2s)

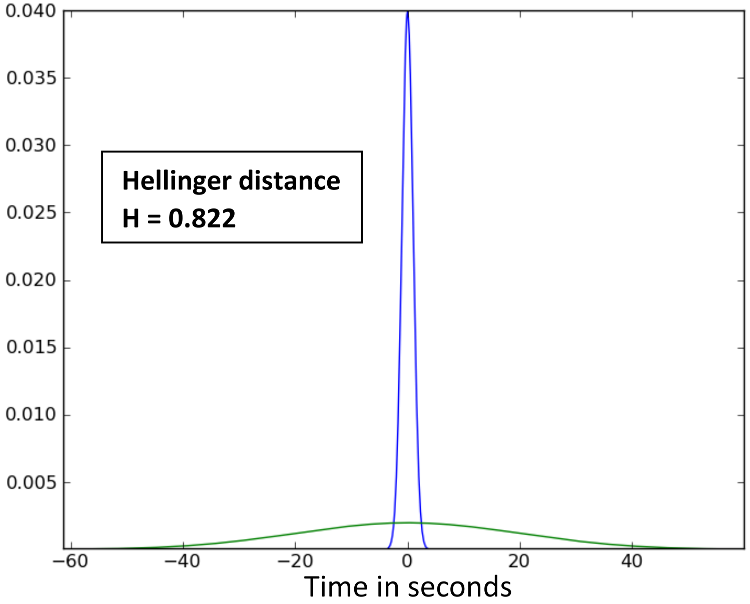
Method – The Hellinger distance

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Illustration on comparing two one-dimensional Gaussian distributions for origin time



Distance for estimates which have different averages (0s and 3s) and the same standard deviation.



Distance for estimates which have the same average (0s) and two different standard deviations (1s and 20s)

Method – The Hellinger distance

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Hellinger distance in the case of epicentre estimates

An epicentre is defined by the two coordinates latitude and longitude. A subset of the covariance matrix defines the 2D distribution function for these and there will be two 2D distributions corresponding to the two different estimates. In this section, we will show some examples of the application of the Hellinger distance between the two estimates, on real examples.

The first real example is of an automatic SEL3 event close to the corresponding analyst-reviewed REB event. The parameters and covariance matrices for the two events are shown in the two tables below.

	orid	lat	lon	smajax	sminax	strike	depth	sdepth	time	stime
REB	18122123	13.47	145.34	15.87	10.34	84.06	57.8	8.44	1574034947	1.08
SEL3	18101274	13.47	145.23	19.29	13.85	89.62	65.5	9.61	1574034948	1.17

	sxx	syy	szz	stt	sxy	sxz	syz	stx	sty	stz
REB	54.44	23.61	26.33	.43	3.24	19.44	.18	-2.97	0.48	-3.16
SEL3	80.90	41.72	34.10	.51	0.26	24.27	.19	-3.55	1.23	-3.62

Hellinger distance for epicentres and error ellipses

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

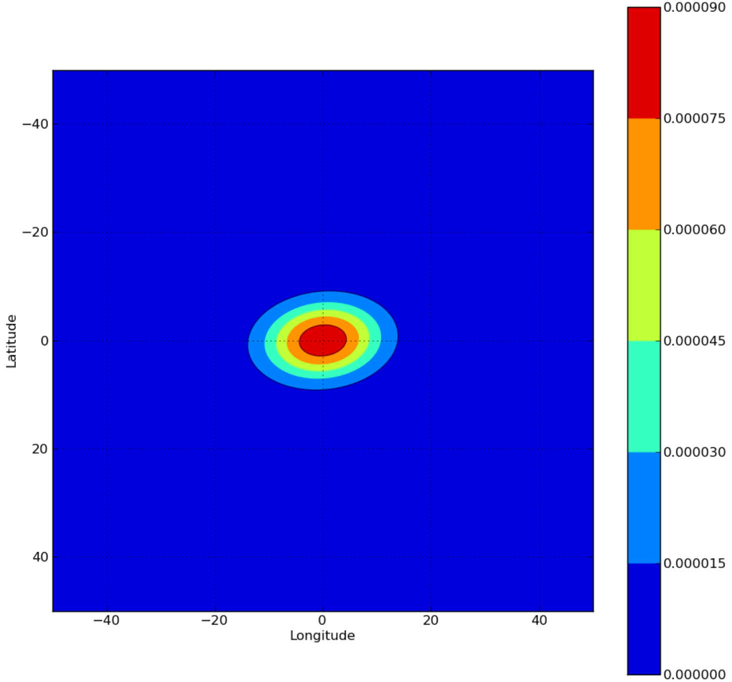
In the following slides, we will show examples of Hellinger and Bhattacharyya distances computed for two dimensional probability distribution functions (PDFs) derived from the covariance matrices of actual examples. The following definitions are used to derive the probability distribution functions. The differences in latitude (Δx) and longitude (Δy) are with respect to the location of the REB event.

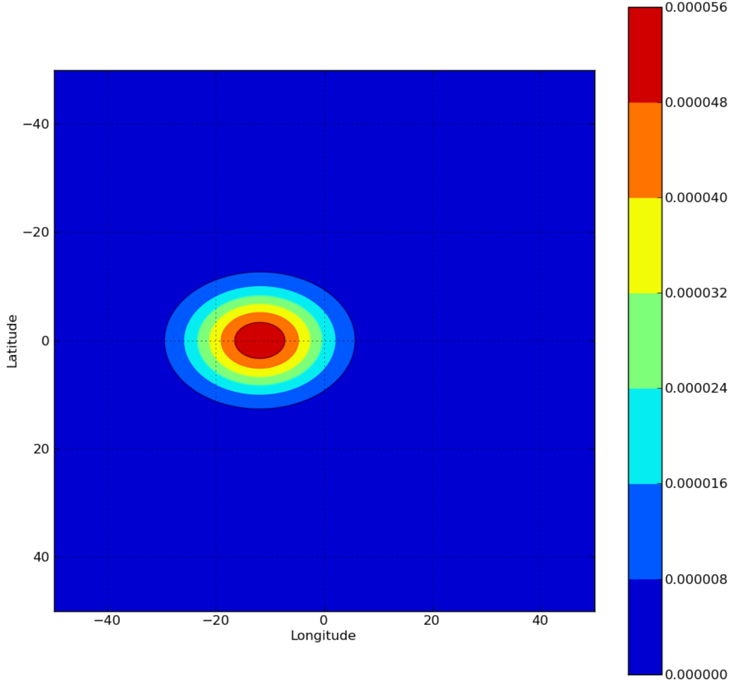
$$\Delta X = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \qquad \text{Cov} = \begin{pmatrix} sxx & sxy \\ sxy & syy \end{pmatrix}$$

$$PDF(\Delta X) = \frac{1}{2\pi\sqrt{\det(Cov)}} e^{-\left[\!\left[\frac{\Delta X^T Cov\, \Delta X}{\det(\text{Cov})} \right]\!\right]}$$

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Probability distribution functions for the SEL3(right) and REB(left) solutions for REB orid 18122123. The Hellinger distance in this case is 0.50 and the Bhattacharyya distance is 0.29. Note that the ellipses centres are not coincident but overlap substantially.

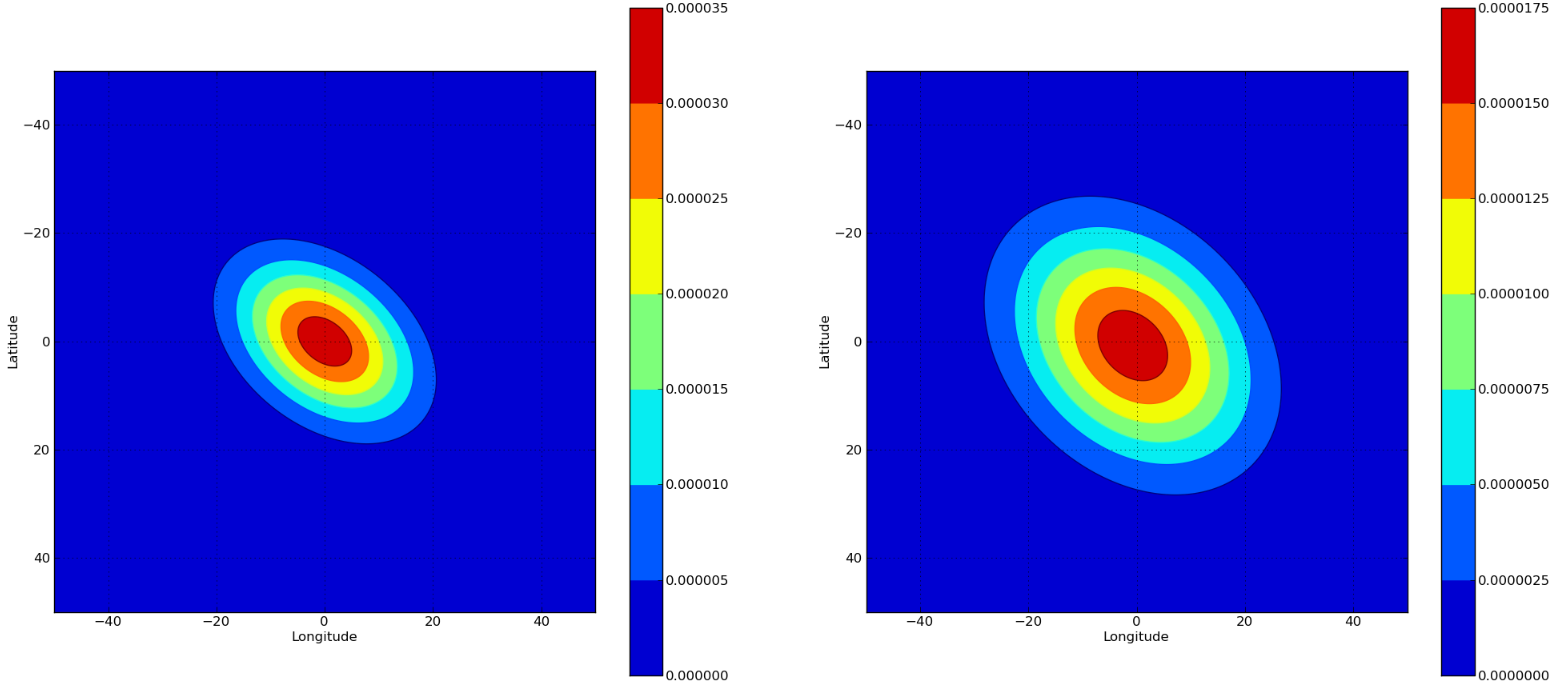




COMPREHENSIVE NUCLEAR-TEST-BAN TREATY ORGANIZATION

CTBTO PREPARATORY COMMISSION

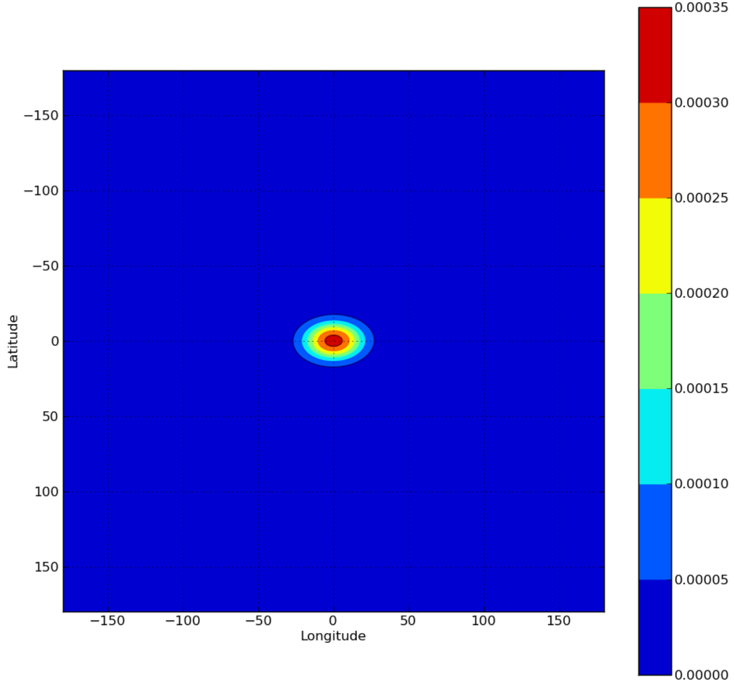
Distributions for the SEL3(right) and REB(left) solutions for REB orid 18090754. The Hellinger distance in this case is 0.25 and the Bhattacharyya distance is 0.06. Note that the ellipses centres are nearly co-located.

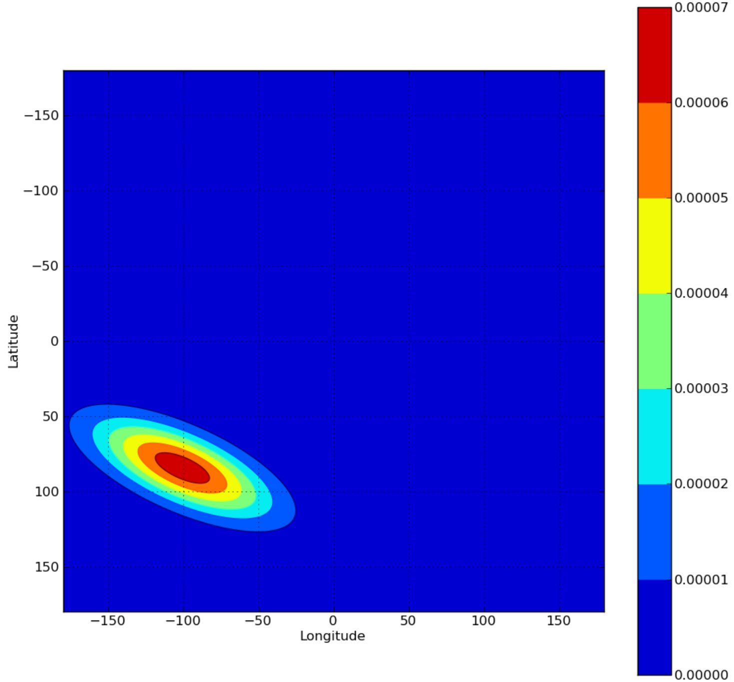


CTBTO PREPARATORY COMMISSION
COMPREHENSIVE NUCLEAR-TEST-BAN TREATY ORGANIZATION

EGU2020: Sharing Geoscience Online
Session: ITS1.7/SM3.5
Abstract: EGU2020-3340

Distributions for the SEL3 (right) and REB(left) solutions for REB orid 18147141. The Hellinger distance in this case is 1.0 and the Bhattacharyya distance is 11.16. Note the large distance between the centres of the ellipses.





Summary

- The Hellinger distance can provide an objective way to measure the difference between two estimates of the same seismo-acoustic location. It is defined in the interval [0 , 1]
- The Bhattacharyya distance is very closely related to the Hellinger distance and varies monotonically with it, on interval [0 , ∞]
- The advantage of using this approach is that it includes not only the estimates of the parameters themselves but also the PDFs of the parameters considered as random variables
- The computational cost for the complete 4 dimensional covariance matrix is expansive, but would provide an objective number including all four parameters included in the computation of the location.