

ABSTRACT

Title of thesis: METROPOLITAN AREA NETWORK
IP GEOLOCATION
THROUGH WAVELET TECHNIQUES

Choon Yik Lee, Master of Science, 2010

Thesis directed by: Professor Richard La
Department of Electrical and Computer Engineering

IP geolocation is the process of finding the geographic locations of Internet hosts. We will focus on Internet hosts in metropolitan area network(MAN). The Internet hosts will be under the same Internet service provider(ISP). Machines in close geographic distance will share almost identical network infrastructure due to having the same ISP. We propose two MAN IP geolocation techniques that are based on wavelets, e.g. wavelet density estimation and wavelet time-frequency analysis.

Wavelet density estimation looks for similarity among RTT distributions of nearby machines. To achieve this, wavelet density estimation utilizes wavelets as orthonormal basis in $L^2(R)$ to construct estimated probability density functions(pdf) of RTT distributions. A symmetrized version of Kullback-Leibler divergence is devised to measure the similarity between two estimated pdfs. The second technique, wavelet time-frequency analysis, explores a common pattern in frequency content evolutions over time of the RTT sequences of nearby machines. Wavelet time-frequency analysis employs wavelets to analyze frequency contents of RTT sequences

over short time-intervals. Sudden rises of frequency content in RTT sequences can then be detected. We evaluate the performance of these two MAN IP geolocation techniques with data sets collected from our testbed. With these data sets, we analyze the effects of RTT sample size, RTT probing rate and landmark distribution to the performance of the techniques.

METROPOLITAN AREA NETWORK
IP GEOLOCATION
THROUGH WAVELET TECHNIQUES

by

Choon Yik Lee

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2010

Advisory Committee:
Professor Richard La, Chair/Advisor
Professor Bobby Bhattacharjee
Professor Mark Shayman

© Copyright by
Choon Yik Lee
2010

Acknowledgments

I would like to thank my colleagues, Satinder Pal Singh and Randolph Baden, for their tremendous effort in setting up the testbed and for their assistance in explaining the configurations of the testbed to me. I am grateful to Professor Mark Shayman and Professor Bobby Bhattacharjee for accepting me to this project and for their advice and suggestions during the course of this thesis. Last but not least, a special thanks to my advisor, Professor Richard La, for introducing wavelets to me and for guiding me through tackling the problem of MAN IP geolocation to produce this thesis.

Table of Contents

| | |
|--|-----|
| List of Figures | iv |
| List of Abbreviations | ix |
| 1 Introduction to MAN IP Geolocation | 1 |
| 2 Anscombe Variance-Stabilizing Transform | 5 |
| 2.1 Overview | 5 |
| 2.2 Simulation Study of Anscombe Transform | 5 |
| 3 Stein’s Unbiased Risk Estimate (SURE) | 8 |
| 3.1 Overview | 8 |
| 3.2 Stein’s Unbiased Risk Estimate | 8 |
| 3.3 Almost Differentiability of Soft Threshold Estimator | 9 |
| 3.4 Stein’s Unbiased Risk Estimate of L^2 Risk of Soft Threshold Estimator | 13 |
| 4 A Brief Introduction to Wavelet | 15 |
| 5 Wavelet Density Estimation | 20 |
| 5.1 Overview | 20 |
| 5.2 Wavelet Density Estimation | 21 |
| 5.3 Distance Metric Between Estimated PDFs | 27 |
| 5.4 Implementation Issues of Wavelet Density Estimation | 27 |
| 5.5 Estimation of Known Densities with Wavelet Density Estimation . . . | 29 |
| 5.6 Pictorial Depiction of Wavelet Density Estimation | 42 |
| 6 Wavelet Time-Frequency Analysis | 44 |
| 6.1 Overview | 44 |
| 6.2 Wavelet Time-Frequency Analysis | 45 |
| 6.3 Time-Frequency Analysis With Wavelet To Geolocate | 50 |
| 6.4 Implementation Issues of Wavelet Time-Frequency Analysis | 52 |
| 6.5 Pictorial Depiction of Wavelet Time-Frequency Analysis | 54 |
| 7 Experiments and Results | 57 |
| 7.1 Wavelet Density Estimation With Different RTT Sample Sizes | 60 |
| 7.2 Wavelet Time-Frequency Analysis With Different RTT Sample Sizes . | 77 |
| 7.3 Wavelet Time-Frequency Analysis With Different Landmark Distri- butions | 98 |
| 7.4 Wavelet Time-Frequency Analysis With Different Ping Rates | 107 |
| 8 Conclusions | 117 |
| Bibliography | 119 |

List of Figures

| | | |
|------|--|----|
| 2.1 | Comparison of cdfs of Y and $N(2\sqrt{\alpha=5}, 1)$ | 6 |
| 3.1 | Soft Threshold Function $\delta_\lambda^{soft}(\cdot)$ | 10 |
| 3.2 | Locations of Non-Differentiable Points | 13 |
| 4.1 | Examples of scaling functions | 16 |
| 4.2 | Piecewise constant function | 17 |
| 4.3 | Example of different approximation spaces, V_j and V_{j-1} | 17 |
| 4.4 | Examples of Wavelet Functions | 18 |
| 5.1 | Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 512 samples. | 33 |
| 5.2 | Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 1024 samples. | 33 |
| 5.3 | Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 2048 samples. | 34 |
| 5.4 | Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 4096 samples. | 34 |
| 5.5 | Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 512 samples. | 35 |
| 5.6 | Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 1024 samples. | 35 |
| 5.7 | Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 2048 samples. | 36 |
| 5.8 | Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 4096 samples. | 36 |
| 5.9 | Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha =$ $1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 512 samples. | 37 |
| 5.10 | Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha =$ $1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 1024 samples. | 37 |
| 5.11 | Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha =$ $1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 2048 samples. | 38 |
| 5.12 | Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha =$ $1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 4096 samples. | 38 |
| 5.13 | Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha =$ $4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 512 samples. | 39 |
| 5.14 | Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha =$ $4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 1024 samples. | 39 |
| 5.15 | Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha =$ $4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 2048 samples. | 40 |
| 5.16 | Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha =$ $4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 4096 samples. | 40 |
| 5.17 | A mixture density, $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ | 41 |

| | | |
|------|--|----|
| 5.18 | A mixture density, $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ | 41 |
| 5.19 | A 500-element RTT sequence collected from a landmark | 42 |
| 5.20 | Rescaling the RTT values into $[0,1]$ | 42 |
| 5.21 | RTT distribution of scaled RTT values | 43 |
| 5.22 | Estimated pdf constructed by wavelet density estimation | 43 |
| 6.1 | Wavelet Time-Frequency Analysis | 48 |
| 6.2 | A 500-element RTT sequence collected from a landmark | 54 |
| 6.3 | Wavelet transform of the RTT sequence at fine scale | 54 |
| 6.4 | Wavelet transform of the RTT sequence at coarser scale | 55 |
| 6.5 | Multiscale product to produce enhanced features while reducing background noise | 55 |
| 6.6 | Zero-one sequence with ones indicating sudden rises of frequency content in the original RTT sequence | 56 |
| 7.1 | Our Testbed at Baltimore-Washington D.C. Metropolitan Area | 58 |
| 7.2 | For a CP target and for all RTT sample sizes, the mean of minimum divergence from CP landmarks is the lowest followed by the mean of minimum divergence from GB landmarks. The mean of minimum divergence from GA landmark and the mean of minimum divergence from GE landmarks are significantly higher. All four lines stabilize and flatten as RTT sample size increases. | 67 |
| 7.3 | For a GB target and for all RTT sample sizes, the mean of minimum divergence from GB landmarks is the lowest followed by the mean of minimum divergence from CP landmarks. The mean of minimum divergence from GA landmark and the mean of minimum divergence from GE landmarks are significantly higher. All four lines stabilize and flatten as RTT sample size increases. | 68 |
| 7.4 | For a GA target and for all RTT sample sizes, the mean of minimum divergence from GE landmarks is the lowest. The mean of minimum divergence from CP landmarks and the mean of minimum divergence from GB landmarks are significantly higher. All three lines stabilize and flatten as RTT sample size increases. | 69 |
| 7.5 | For a GE target and for all RTT sample sizes, the mean of minimum divergence from GA landmark is the lowest followed by the mean of minimum divergence from GE landmark. The mean of minimum divergence from CP landmarks and the mean of minimum divergence from GB landmarks are significantly higher. The top two lines stabilize and flatten as RTT sample size increases. The bottom two lines go down slightly with their gap closing up as RTT sample size increases. | 70 |
| 7.6 | The general trend is that the matching percentage of a CP target to CP landmarks increases as RTT sample size increases. | 71 |
| 7.7 | The matching percentage of a CP target to CP or GB landmarks increases as RTT sample size increases. | 71 |

| | | |
|------|--|----|
| 7.8 | The matching percentage of a CP target to GA or GE landmarks decreases as RTT sample size increases. | 72 |
| 7.9 | The general trend is that the matching percentage of a GB target to GB landmarks increases as RTT sample size increases. | 72 |
| 7.10 | The matching percentage of a GB target to CP or GB landmarks increases as RTT sample size increases. | 73 |
| 7.11 | The matching percentage of a GB target to GA or GE landmarks decreases as RTT sample size increases. | 73 |
| 7.12 | The matching percentage of a GA target to GE landmarks generally increases as RTT sample size increases from 0 to 256 samples. After 256 samples, the percentage of GA matched to GE landmarks drops. | 74 |
| 7.13 | The matching percentage of a GA target to CP or GB landmarks generally decreases as RTT sample size increases from 0 to 256 samples. After 256 samples, the percentage of GA matched to CP or GB landmarks increases. | 74 |
| 7.14 | The general trend is that the matching percentage of a GE target to GE landmark increases as RTT sample size increases. | 75 |
| 7.15 | The general trend is that the matching percentage of a GE target to GA or GE landmarks increases as RTT sample size increases. | 75 |
| 7.16 | The general trend is that the matching percentage of a GE target to CP or GB landmarks decreases as RTT sample size increases. | 76 |
| 7.17 | The average number of 1's in a zero-one sequence increases as RTT sample size increases. More network activities are captured as observation period of RTT values increases. The average number of 1's matched up for inner products increases as RTT sample size increases. The difference of average number of 1's matched up between correct and incorrect city matchings increases as RTT sample size increases. | 88 |
| 7.18 | For a CP target, the mean of the maximum inner product from CP landmarks is consistently higher than those of landmarks from other cities for all RTT sample sizes. | 89 |
| 7.19 | For a GB target, the mean of the maximum inner product from GB landmarks is consistently higher than those of landmarks from other cities for all RTT sample sizes. | 89 |
| 7.20 | For a GA target, the mean of the maximum inner product from GE landmarks is higher than those of landmarks from other cities for all RTT sample sizes. | 90 |
| 7.21 | For a GE target, the mean of the maximum inner product from GE landmark is consistently lower than those of landmarks from other cities for all RTT sample sizes. However, the mean of the maximum inner product from GA landmark is consistently higher than those of landmarks from other cities. | 91 |
| 7.22 | Variation of 2% is not significant. | 92 |
| 7.23 | The matching percentage of a CP target to CP or GB landmarks increases as RTT sample size increases. | 92 |

| | | |
|------|--|-----|
| 7.24 | The matching percentage of a CP target to GA or GE landmarks decreases as RTT sample size increases. | 93 |
| 7.25 | The general trend is that the matching percentage of a GB target to GB landmarks increases as RTT sample size increases. | 93 |
| 7.26 | Variation of 2% is not significant. | 94 |
| 7.27 | Variation of 2% is not significant. | 94 |
| 7.28 | The matching percentage of a GA target to GE landmarks increases as RTT sample size increases from 0 to 300 samples. After 300 samples, the percentage of GA matched to GE landmarks drops. Wavelet density estimation registered similar observation. | 95 |
| 7.29 | The matching percentage of a GA target to CP or GB landmarks decreases as RTT sample size increases from 0 to 300 samples. After 300 samples, the percentage of GA matched to CP or GB landmarks increases. Wavelet density estimation captured similar observation. | 96 |
| 7.30 | The general trend is that the matching percentage of a GE target to GE landmark increases as RTT sample size increases. | 96 |
| 7.31 | There is no definite behavior for the matching percentage of a GE target to GA or GE landmarks as RTT sample size increases. | 97 |
| 7.32 | There is no definite behavior for the matching percentage of a GE target to CP or GB landmarks as RTT sample size increases. | 97 |
| 7.33 | As there are more landmarks in CP, a CP target is more likely to be matched to CP city. | 102 |
| 7.34 | A CP target matched to CP city or GB city more often when there are more landmarks in both cities. | 102 |
| 7.35 | As there are more landmarks in GB, a GB target is more likely to be matched to GB city. | 103 |
| 7.36 | A GB target matched to CP city or GB city more often when there are more landmarks in both cities. | 103 |
| 7.37 | With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GA to GE increases to 68%. It is 47% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs. Thus, the issue of fewer GA and GE landmarks in landmark distribution of 4CPs, 4GBs, 1GA and 2GEs does play a role in the low matching percentage of GA to GE. | 104 |
| 7.38 | With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GE to GE increases to 20%. It is 14% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs. | 105 |
| 7.39 | With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GE to GA/GE increases to 76%. It is 61% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs. Thus, fewer GA and GE landmarks in landmark distribution of 4CPs, 4GBs, 1GA and 2GEs contribute to the low matching percentage of GE to GA/GE. | 106 |

| | | |
|------|---|-----|
| 7.40 | As ping rate increases with shorter inter-packet interval, more dynamics of the network are captured as shown by the increasing average number of 1's in a 500-element zero-one sequence. | 115 |
| 7.41 | As ping rate increases with shorter inter-packet interval, the average number of 1's matched up in correct city matchings and the average number of 1's matched up in incorrect city matchings both decrease. Thus, as ping rate increases, more network dynamics are captured with more 1's in a zero-one sequence but most of these dynamics captured are not network activities experienced by nearby landmarks as suggested by the decreasing average number of 1's matched up in correct city matchings. | 116 |

List of Abbreviations

| | |
|--------------------|--|
| $1_{[a,b]}(\cdot)$ | Indicator Function of Interval $[a, b]$ |
| a.e | almost everywhere |
| CBG | Constraint-Based Geolocation |
| CDF | Cumulative Distribution Function |
| DNS | Domain Name System |
| i.i.d | independent and identically distributed |
| IP | Internet Protocol |
| ISP | Internet Service Provider |
| KL Divergence | Kullback-Leibler Divergence |
| MAN | Metropolitan Area Network |
| $N(\mu, \sigma^2)$ | Normal random variable with mean μ and variance σ^2 |
| MRA | Multiresolution Analysis |
| PDF | Probability Density Function |
| $Poisson(\alpha)$ | Poisson random variable with parameter α |
| PMF | Probability Mass Function |
| RTT | Round-Trip Time |
| SURE | Stein's Unbiased Risk Estimate |
| UMD | University of Maryland |

Chapter 1

Introduction to MAN IP Geolocation

IP geolocation is the process of finding the geographic locations of Internet hosts. This thesis focuses on Internet hosts in metropolitan area network(MAN). Potential applications of MAN IP geolocation are targeted advertising (e.g. announcement of local events and weather), selective media streaming based on territorial content policies and automatic display of nearby stores during online shopping.

Previous works on IP geolocation are[10][12]:

- (i) GeoTrack looks for possible location information from the DNS name of an Internet host. For instance, *www.comp.state.md.us* implies the state of Maryland in the U.S. and *www.weatheroffice.gc.ca* implies the country of Canada.
- (ii) Whois Database Look-up searches through Whois database for the location information of an Internet host.
- (iii) GeoPing and Constraint-Based Geolocation(CBG) exploit network delay measurement to infer the distance of an Internet host from reference hosts with known locations.

Each of the aforementioned geolocation techniques has some drawbacks. GeoTrack will fail when the DNS name does not contain location information. Whois database may not be updated with the latest information about an Internet host frequently

enough. Whois database may have just one single entry for a large group of Internet hosts. GeoPing and CBG are not suitable for MAN. In MAN, queueing delay is significant in end-to-end delay measurement. Thus, there will be less correlation between end-to-end delay and distance traveled by measurement packets.

To overcome the limitations of previous works, we introduce two MAN IP geolocation techniques, wavelet density estimation and wavelet time-frequency analysis. These two techniques involve three types of Internet hosts: targets, landmarks and probe machines. Targets are Internet hosts whose unknown locations are to be geolocated. Landmarks are Internet hosts whose exact locations are known. Probe machines are Internet hosts that send out time-synchronous ping packets(ICMP Echo Requests) to targets and landmarks. Targets and landmarks are to respond to ping packets. From the returned packets, probe machines construct time-synchronous RTT sequences for targets and landmarks. Targets and landmarks are to be under the same Internet service provider(ISP). A target and its nearby landmarks will then share almost identical network infrastructure due to having the same ISP.

The first technique, wavelet density estimation, creates RTT distributions from RTT sequences constructed by probe machines. Based on the RTT distributions, wavelet density estimation generates corresponding estimated probability density functions(pdf). A target and its nearby landmarks share a large portion of the network infrastructure under the same ISP. Hence, they have similar RTT distributions and thus similar estimated pdfs. Geolocation of a target is then based on the similarity of its estimated pdf to nearby landmarks' estimated pdfs.

The second technique, wavelet time-frequency analysis, analyzes the evolutions of frequency content over time of the time-synchronous RTT sequences of targets and landmarks. Wavelet time-frequency analysis can then detect time-synchronous sudden rises of frequency contents in those RTT sequences. Network activities cause momentary increases of frequency content in RTT sequences. Under the same ISP, a target and its nearby landmarks experience network activities at roughly the same time. Therefore, a target and its nearby landmarks record time-synchronous sudden rises of frequency contents in their RTT sequences. Wavelet time-frequency analysis geolocates a target to its nearby landmarks based on these time-synchronous sudden increases of frequency contents in their RTT sequences.

We have evaluated the two techniques, wavelet density estimation and wavelet time-frequency analysis, with a collection of data sets. Larger RTT sample size generally improves the performance of both techniques. Wavelet time-frequency analysis is more susceptible to poor performance when landmark distribution is sparse. Sufficient landmarks are crucial to the performance of both techniques. The performance of wavelet time-frequency analysis shows that high RTT probing rates capture more network dynamics which are not network activities shared by nearby landmarks.

The rest of the thesis is organized as follows. In Chapters 2 and 3, we present materials necessary for detailed explanations of wavelet density estimation. We give a brief introduction to wavelet in Chapter 4. Chapter 5 discusses the first technique, wavelet density estimation, from theoretical aspects till implementation issues. Chapter 6 does the same for the second technique, wavelet time-frequency

analysis. Performance analysis of both techniques is in Chapter 7. Finally, we present the conclusions of this thesis in Chapter 8.

Chapter 2

Anscombe Variance-Stabilizing Transform

2.1 Overview

Definition: (Anscombe Variance Stabilizing Transform[1])

Let $X = \text{Poisson}(\alpha)$, $\alpha > 0$

$$Y = 2\sqrt{X + \frac{3}{8}}$$

It was shown that $Y \approx N(2\sqrt{\alpha}, 1)$ [1]. The transform is said to be variance stabilizing because the variance of the transformed random variable Y does not depend on its mean. Variance-stabilizing transforms strive to make the transformed variance functionally free from the transformed mean. The transform is also a normalizing transform whereby the transformed random variable Y can be approximated by a normal random variable. Algorithms that expect normally-distributed input data can then be used to do further data analysis[6].

2.2 Simulation Study of Anscombe Transform

Simulation study was conducted to evaluate how well the transformed random variable Y is approximated by a normal random variable $N(2\sqrt{\alpha}, 1)$. The simulation study fixed the Poisson parameter $\alpha = 5$ for the input variable X . The cumulative

distribution functions(cdfs) of Y and $N(2\sqrt{5}, 1)$ were plotted simultaneously in Figure 2.1. Since two cdfs are close to each other, it is valid to approximate Y with normal random variable $N(2\sqrt{\alpha}, 1)$ when $\alpha = 5$.

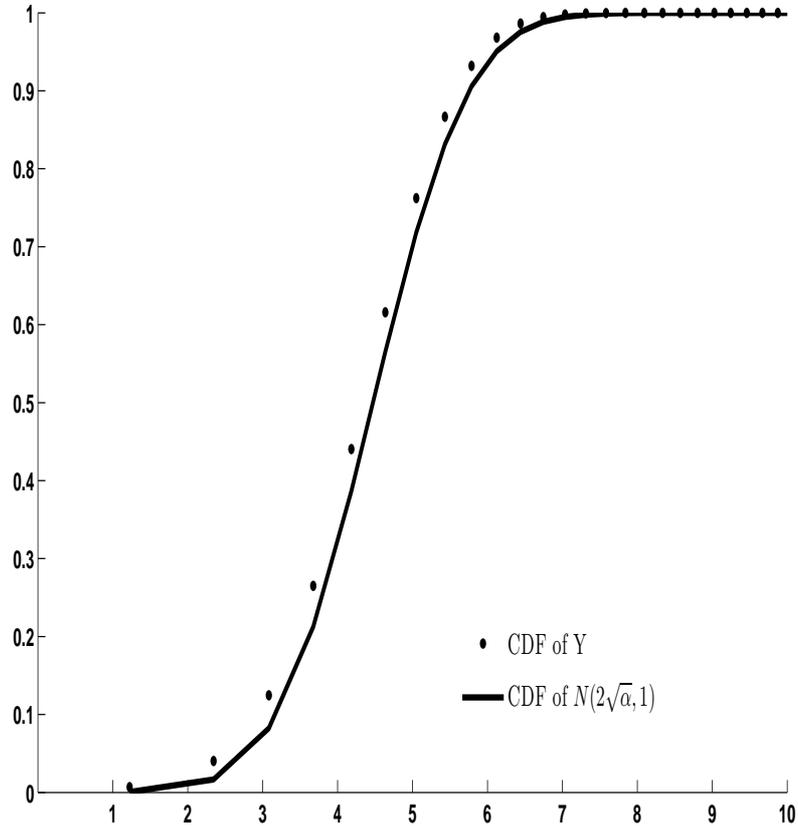


Figure 2.1: Comparison of cdfs of Y and $N(2\sqrt{\alpha} = 5, 1)$

We will tabulate the maximum absolute error(MAE) of the cdfs of Y and $N(2\sqrt{\alpha}, 1)$ as $\max\{|P(Y \leq c_i) - \int_{-\infty}^{c_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2\sqrt{\alpha})^2}{2}} dx|, c_i = 2\sqrt{i + \frac{3}{8}}, i = 0, 1, 2, \dots, 50\}$ for $\alpha = 1, 2, 3, 4, 5, 10, 15, 20$. We will also tabulate the sample mean and the sample variance of Y with sample size of 10000 Poisson(α)-distributed X samples to compare with the Gaussian mean $2\sqrt{\alpha}$ and the Gaussian variance 1 for

$\alpha = 1, 2, 3, 4, 5, 10, 15, 20.$

Numerical Investigation of Anscombe Transform

| α | MAE of CDFs Y and $N(2\sqrt{\alpha}, 1)$ | Sample Mean of Y | Gaussian Mean $2\sqrt{\alpha}$ | Sample Variance of Y | Gaussian Variance 1 |
|----------|---|------------------------|--------------------------------------|----------------------------|---------------------------|
| 1 | 0.1488 | 2.1851 | 2.0000 | 0.7204 | 1 |
| 2 | 0.0915 | 2.9319 | 2.8284 | 0.9347 | 1 |
| 3 | 0.0719 | 3.5496 | 3.4641 | 0.9764 | 1 |
| 4 | 0.0612 | 4.0635 | 4.0000 | 1.0229 | 1 |
| 5 | 0.0541 | 4.5270 | 4.4721 | 0.9808 | 1 |
| 10 | 0.0375 | 6.3706 | 6.3246 | 0.9875 | 1 |
| 15 | 0.0304 | 7.7604 | 7.7460 | 1.0109 | 1 |
| 20 | 0.0263 | 8.9673 | 8.9443 | 0.9965 | 1 |

As the magnitude of α increases, the MAE of cdfs of Y and $N(2\sqrt{\alpha}, 1)$ decreases.

With larger α , the sample mean of Y is closer to the approximation of $2\sqrt{\alpha}$ and the sample variance of Y approaches unit variance.

Chapter 3

Stein's Unbiased Risk Estimate (SURE)

3.1 Overview

Given a particular estimator of the unknown means of independent normal random variables, Stein's Unbiased Risk Estimate[14] is an unbiased estimate of the L^2 risk of that particular estimator under certain conditions. The estimator of interest here is the soft threshold estimator. We will show that soft threshold estimator fulfills certain conditions so that Stein's Unbiased Risk Estimate(SURE) can be applied to be an unbiased estimate of the L^2 risk of soft threshold estimator.

3.2 Stein's Unbiased Risk Estimate

We will present SURE and the necessary conditions to apply SURE[14]. Let $\xi = (\xi_1, \dots, \xi_p)$ be an unknown constant p-dimensional vector. Let $V = (V_1, \dots, V_p)$ be jointly Gaussian with mean ξ and with identity matrix as covariance matrix. Let $\hat{\xi} = \hat{\mu}(V)$ where $\hat{\mu}(\cdot)$ is an estimator of the unknown mean ξ and $\hat{\xi}$ is the corresponding estimate of ξ . We will define almost differentiability quoted as follows.

Definition (Almost Differentiability)[14]: For $x, y \in R^p$, we define $x \cdot y = \sum_{i=1}^p x_i y_i$ and $\|x\|^2 = x \cdot x = \sum_{i=1}^p (x_i)^2$. A function $h : R^p \rightarrow R$ will be called almost differentiable if there exists a function $\nabla h : R^p \rightarrow R^p$ such that, for all

$z \in R^p$, $h(x+z) - h(x) = \int_0^1 z \cdot \nabla h(x+tz) dt$, for almost all $x \in R^p$. A function $g : R^p \rightarrow R^p$ is almost differentiable if all its coordinate functions are.

Notationwise, let $g_i : R^p \rightarrow R$ be the i^{th} coordinate function of $g : R^p \rightarrow R^p$ to take the role of h in the quoted definition of almost differentiability. Let ∇g_i be as ∇h in the quoted definition of almost differentiability. Then, we can define $\nabla \cdot g \doteq \sum_i \nabla_i g_i = \sum_i (i^{th} \text{ component of } \nabla g_i)$. Stein showed that when $\hat{\mu}(V) = V + g(V)$ where $g : R^p \rightarrow R^p$ is an almost differentiable with $E_\xi \sum_i |\nabla_i g_i(V)| < \infty$, then the L^2 risk $E_\xi \|\hat{\xi} - \xi\|^2 = E_\xi \|\hat{\mu}(V) - \xi\|^2 = E_\xi \|V + g(V) - \xi\|^2$ is $p + E_\xi \{\|g(V)\|^2 + 2\nabla \cdot g(V)\}$ and $p + \|g(V)\|^2 + 2\nabla \cdot g(V)$ is an unbiased estimate of the L^2 risk. $p + \|g(V)\|^2 + 2\nabla \cdot g(V)$ is called Stein's Unbiased Risk Estimate(SURE).

3.3 Almost Differentiability of Soft Threshold Estimator

Let $x = (x_1, \dots, x_p) \in R^p$ as in the quoted definition of almost differentiability. x will be the multi-dimensional parameter of the following soft threshold estimator

$\hat{\mu}(\cdot)$. We define soft threshold estimator $\hat{\mu}(\cdot)$ as $\hat{\mu}(x) \doteq \begin{bmatrix} \delta_\lambda^{soft}(x_1) \\ \vdots \\ \delta_\lambda^{soft}(x_p) \end{bmatrix}$ where

$$\delta_\lambda^{soft}(u) = \begin{cases} u - \lambda & \text{if } u > \lambda \\ 0 & \text{if } -\lambda \leq u \leq \lambda \\ u + \lambda & \text{if } u < -\lambda \end{cases}$$

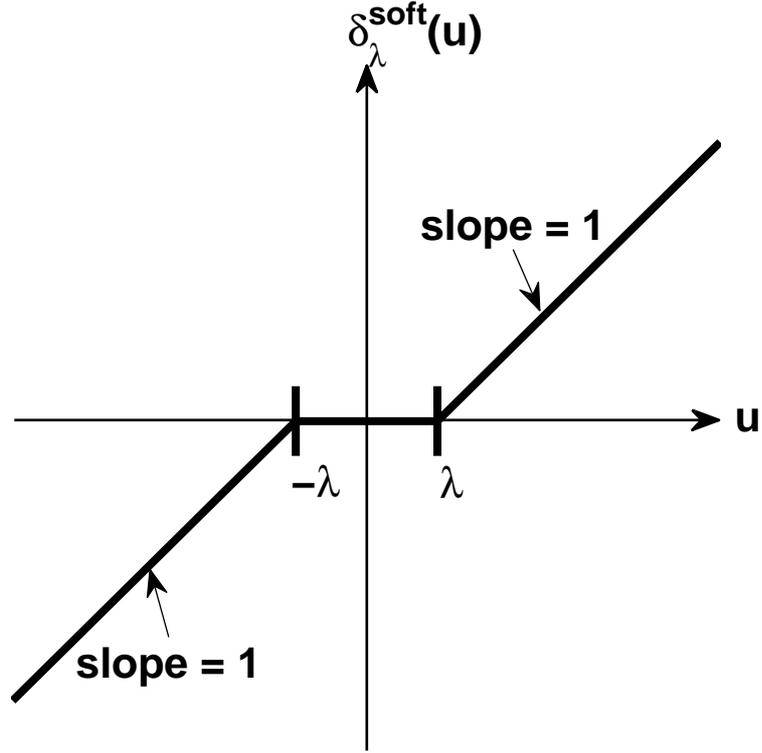


Figure 3.1: Soft Threshold Function $\delta_\lambda^{\text{soft}}(\cdot)$

Let $g(x) \doteq \begin{bmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{bmatrix} = \begin{bmatrix} \delta_\lambda^{\text{soft}}(x_1) - x_1 \\ \vdots \\ \delta_\lambda^{\text{soft}}(x_p) - x_p \end{bmatrix}$. Then, $\hat{\mu}(x) = x + g(x)$. Now, we will show that $g(x) : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is almost differentiable. According to the quoted Stein's definition of almost differentiability, we need to show that the coordinate functions $g_k(x) = \delta_\lambda^{\text{soft}}(x_k) - x_k : \mathbb{R}^p \rightarrow \mathbb{R}$ are almost differentiable for $k = 1, \dots, p$. Let $z \in \mathbb{R}^p$ and $t \in \mathbb{R}$. Then, $g_k(x + z) = \delta_\lambda^{\text{soft}}(x_k + z_k) - (x_k - z_k)$ and $g_k(x + z) - g_k(x) = \delta_\lambda^{\text{soft}}(x_k + z_k) - (x_k - z_k) - (\delta_\lambda^{\text{soft}}(x_k) - (x_k))$. Let $\nabla g_k(x + tz)$ be as

$$\nabla g_k(x + tz) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -1_{[-\lambda, \lambda]}(x_k + tz_k) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow k^{th} \text{ component}$$

Thus, $\nabla g_k : R^p \rightarrow R^p$ and $z \cdot \nabla g_k(x + tz) = z_k \cdot -1_{[-\lambda, \lambda]}(x_k + tz_k)$. Taking integration from 0 to 1, we find that $\int_0^1 z \cdot \nabla g_k(x + tz) dt = \int_0^1 z_k \cdot -1_{[-\lambda, \lambda]}(x_k + tz_k) dt = \int_{x_k}^{x_k + z_k} -1_{[-\lambda, \lambda]}(y) dy$. Now, we need to use one form of Fundamental Theorem of Calculus. We will define it first and we will later verify that all the necessary conditions of Fundamental Theorem of Calculus are satisfied in our case.

Definition (Fundamental Theorem of Calculus)[3]: Suppose there is a finite set E in $[a, b]$ and functions $f, F : [a, b] \rightarrow R$ such that:

- (a) F is continuous on $[a, b]$,
- (b) $F'(x) = f(x)$ for all $x \in [a, b] \setminus E$,
- (c) f belongs to $R[a, b]$.

Then we have $\int_a^b f = F(b) - F(a)$.

We will verify the conditions of Fundamental Theorem of Calculus are satisfied in our case, $\int_{x_k}^{x_k + z_k} -1_{[-\lambda, \lambda]}(y) dy$. Let $a = x_k$ and $b = x_k + z_k$. Let $f(u) = -1_{[-\lambda, \lambda]}(u) : [a, b] \rightarrow R$ and $F(u) = \delta_\lambda^{soft}(u) - u : [a, b] \rightarrow R$. Since $F(u) = \delta_\lambda^{soft}(u) - u$ is continuous on R and thus on $[a, b]$, condition (a) is satisfied. $F(u) = \delta_\lambda^{soft}(u) - u$

is not differentiable at $u = -\lambda$ and at $u = \lambda$ as shown in Figure 3.2. Hence, $F'(u) = \frac{d}{du} \left[\delta_\lambda^{soft}(u) - u \right] \stackrel{a.e.}{=} -1_{[-\lambda, \lambda]}(u) = f(u)$, $\forall u \in [a, b] \setminus \{-\lambda, \lambda\}$. Condition (b) is satisfied as $\{-\lambda, \lambda\}$ is a finite set. $f(u) = -1_{[-\lambda, \lambda]}(u)$ is Riemann-integrable as indicator function is a simple and nice function. Therefore, condition (c) is satisfied.

Since our case, $\int_{x_k}^{x_k+z_k} -1_{[-\lambda, \lambda]}(y) dy$, satisfies all conditions of Fundamental Theorem of Calculus, we can proceed as follows: $\int_{x_k}^{x_k+z_k} -1_{[-\lambda, \lambda]}(y) dt = F(b) - F(a) = F(x_k + z_k) - F(x_k) = \delta_\lambda^{soft}(x_k + z_k) - (x_k + z_k) - \left(\delta_\lambda^{soft}(x_k) - (x_k) \right) = g_k(x + z) - g_k(x)$. Finally, we have shown that $\int_0^1 z \cdot \nabla g_k(x + tz) dt = g_k(x + z) - g_k(x)$, $\forall x, z \in R^p$. By Stein's definition, $g_k(x) = \delta_\lambda^{soft}(x_k) - x_k$ is almost differentiable. Since the previous proof is true for g_k 's, $k = 1, \dots, p$, $g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{bmatrix} = \begin{bmatrix} \delta_\lambda^{soft}(x_1) - x_1 \\ \vdots \\ \delta_\lambda^{soft}(x_p) - x_p \end{bmatrix}$ is almost differentiable as all its coordinate functions g_k 's are almost differentiable. Hence, we have shown that soft threshold estimator $\hat{\mu}(V) = \begin{bmatrix} \delta_\lambda^{soft}(V_1) \\ \vdots \\ \delta_\lambda^{soft}(V_p) \end{bmatrix} = V + g(V)$ where $g(V) = \begin{bmatrix} g_1(V) \\ \vdots \\ g_p(V) \end{bmatrix} = \begin{bmatrix} \delta_\lambda^{soft}(V_1) - V_1 \\ \vdots \\ \delta_\lambda^{soft}(V_p) - V_p \end{bmatrix}$ is almost differentiable.

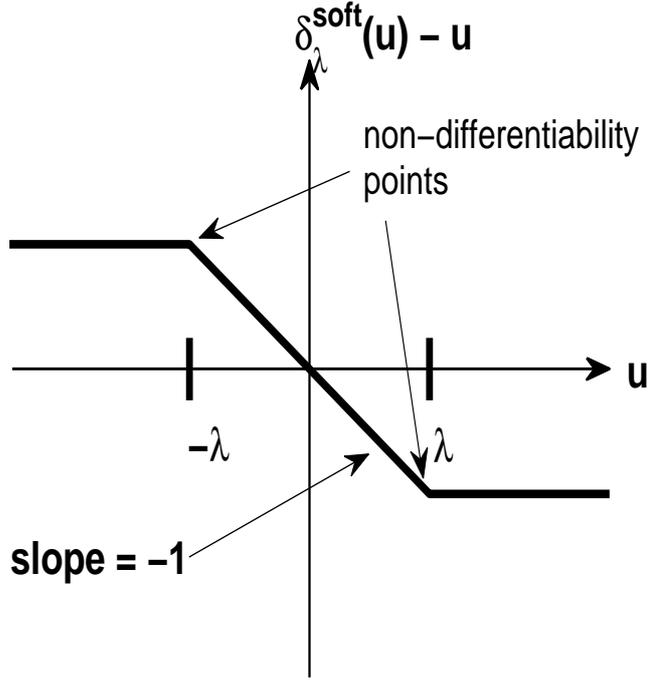


Figure 3.2: Locations of Non-Differentiable Points

3.4 Stein's Unbiased Risk Estimate of L^2 Risk of Soft Threshold Estimator

We have demonstrated that soft threshold estimator $\hat{\mu}(V) = \begin{bmatrix} \delta_{\lambda}^{\text{soft}}(V_1) \\ \vdots \\ \delta_{\lambda}^{\text{soft}}(V_p) \end{bmatrix} = V + g(V)$ where $g(V) = \begin{bmatrix} g_1(V) \\ \vdots \\ g_p(V) \end{bmatrix} = \begin{bmatrix} \delta_{\lambda}^{\text{soft}}(V_1) - V_1 \\ \vdots \\ \delta_{\lambda}^{\text{soft}}(V_p) - V_p \end{bmatrix}$ is almost differentiable. It is also true that $E_{\xi} \sum_i |\nabla_i g_i(V)| = E_{\xi} \sum_i |-1_{[-\lambda, \lambda]}(V_i)| = E_{\xi} \sum_i 1_{[-\lambda, \lambda]}(V_i) = \sum_i E_{\xi} 1_{[-\lambda, \lambda]}(V_i) = \sum_i P(-\lambda \leq V_i < \lambda) < \infty$. Thus, we can use Stein's Unbiased Risk Estimate to estimate the L^2 risk of soft threshold estimator as follows:

$$\|g(V)\|^2 = \sum_{i=1}^p (g_i(V))^2 = \sum_{i=1}^p (\delta_{\lambda}^{\text{soft}}(V_i) - V_i)^2 = \sum_{i=1}^p (\min(|V_i|, \lambda))^2$$

$$\nabla \cdot g(V) = \sum_{i=1}^p \nabla_i g_i = \sum_{i=1}^p -1_{[-\lambda, \lambda]}(V_i)$$

$$\begin{aligned}
SURE(\lambda; V) &= p + \|g(V)\|^2 + 2\nabla \cdot g(V) \\
&= p + \sum_{i=1}^p (\min(|V_i|, \lambda))^2 + 2 \sum_{i=1}^p -1_{[-\lambda, \lambda]}(V_i) \\
&= p - 2 \cdot \#\{i : |V_i| \leq \lambda\} + \sum_{i=1}^p (\min(|V_i|, \lambda))^2
\end{aligned}$$

where $\#A$ for some set A denotes the cardinality of the set A .

Chapter 4

A Brief Introduction to Wavelet

We will present wavelet and some of its properties in this chapter. Chui[7], Boggess and Narcowich[4] and Ogden[11] are good introductory texts on wavelet. We borrow materials from the manuscript by Boggess and Narcowich[4] in the following descriptions of wavelet. We start with the definition of multiresolution analysis(MRA).

Definition (Multiresolution Analysis)[4]: Let $V_j, j = \dots, -2, -1, 0, 1, 2, \dots$ be a sequence of subspaces of functions of $L^2(\mathbb{R})$. The collection of spaces $\{V_j, j \in \mathbb{Z}\}$ is called a multiresolution analysis with scaling function ϕ if the following conditions hold:

1. (nested) $V_j \subset V_{j+1}$
2. (density) $\overline{\cup V_j} = L^2(\mathbb{R})$
3. (separation) $\cap V_j = \{0\}$
4. (scaling) The function $f(x)$ belongs to V_j if and only if the function $f(2^{-j}x)$ belongs to V_0 .
5. (orthonormal basis) The function ϕ belongs to V_0 and the set $\{\phi(x - k), k \in \mathbb{Z}\}$ is an orthonormal basis (using the L^2 inner product) for V_0 .

From the definition of MRA, distinct scaling functions ϕ 's give birth to distinct

multiresolution analyzes. Usually, scaling functions have finite support and are identically zero outside the finite support. Examples of scaling functions are Haar scaling function and Daubechies scaling function as plotted in Figure 4.1.

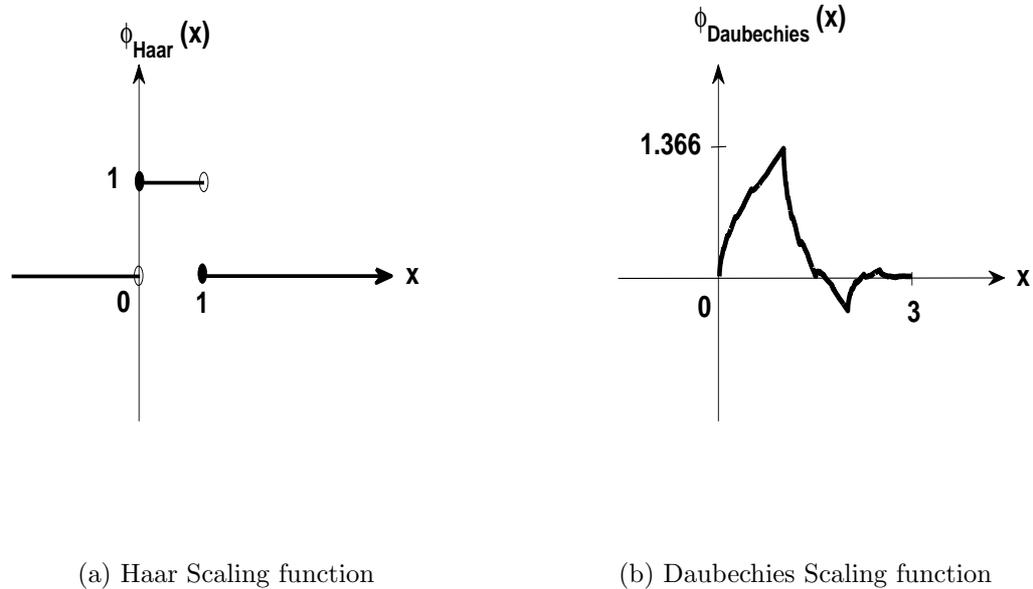


Figure 4.1: Examples of scaling functions

The V_j 's are called approximation spaces. It can be shown that for any $j \in \mathbb{Z}$, $\{\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), k \in \mathbb{Z}\}$ is an orthonormal basis for V_j . Thus, V_j consists of functions constructed from the scaled and translated scaling functions, $2^{j/2}\phi\left(\frac{x-k2^{-j}}{2^{-j}}\right)$. For instance, if $\phi(x)$ is the Haar scaling function and $f(x) \in V_j$, $f(x)$ is piecewise constant as shown in Figure 4.2. If j is sufficiently large, we can use a function $\hat{g}_j(x) \in V_j$ to approximate some arbitrary function $g(x)$ as in Figure 4.3(a). By changing j , we can approximate the function $g(x)$ at various resolutions as in Figure 4.3. Thus, V_j 's are aptly called the approximation spaces.

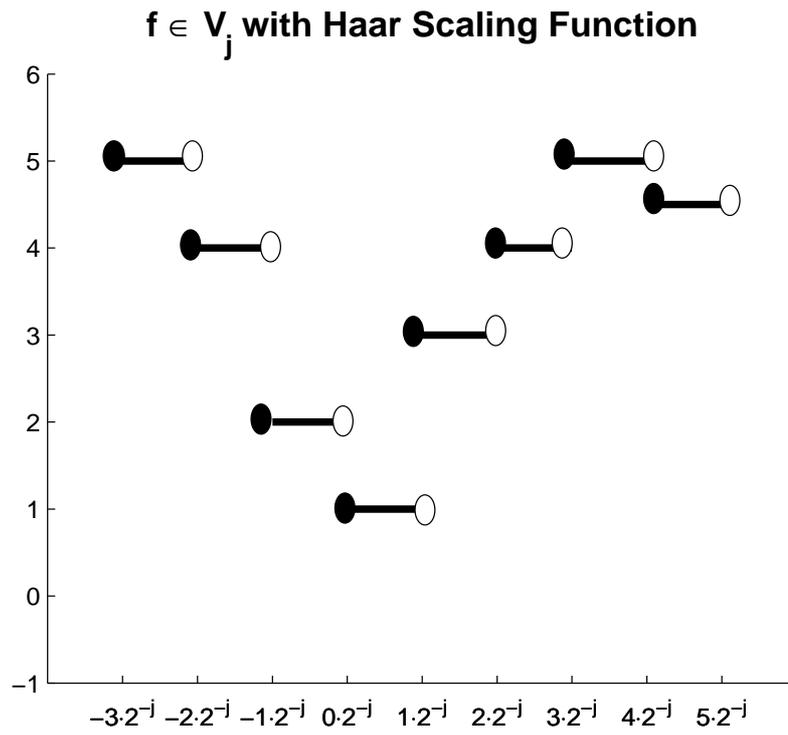
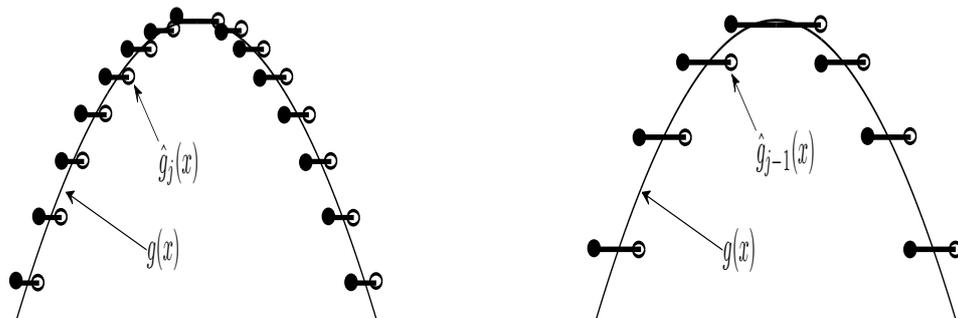


Figure 4.2: Piecewise constant function

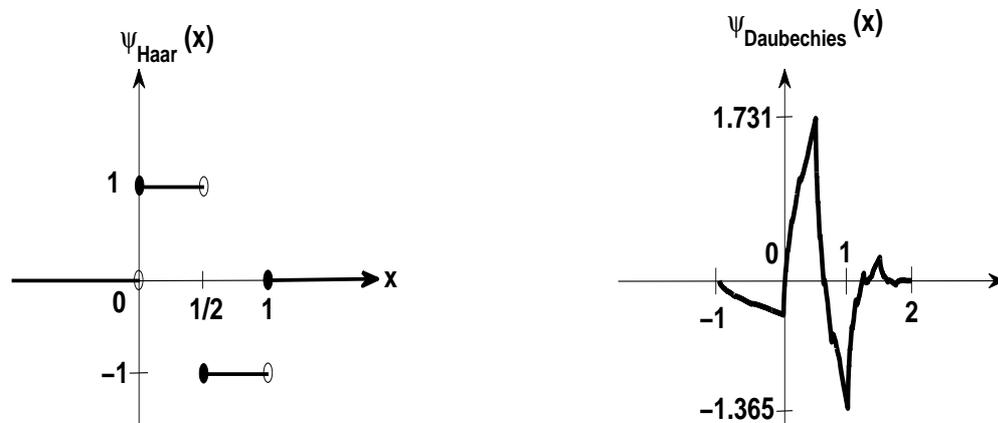


(a) Approximation in V_j

(b) Approximation in V_{j-1}

Figure 4.3: Example of different approximation spaces, V_j and V_{j-1} .

Next, we introduce wavelet function $\psi(x)$. For every scaling function $\phi(x)$, there is a corresponding wavelet function $\psi(x)$. For example, the Haar wavelet and the Daubechies wavelet are plotted in Figure 4.4. A wavelet function $\psi(x)$ has most of its energy localized on a finite support. A wavelet decays to zero as x moves away from the finite support. A wavelet function $\psi(x)$ oscillates or vibrates around zero on the finite support with $\int_{-\infty}^{\infty} \psi(x) dx = 0$. Due to its localized and oscillatory properties, a “wave”let function looks like a small wave as in the Haar and Daubechies wavelets. When the x-axis is considered as time axis and with the oscillatory part representing some frequency content, a wavelet function $\psi(x)$ is said to possess the property of time-frequency localization when energy of the wavelet function is localized in both time domain and frequency domain.



(a) Haar Wavelet function

(b) Daubechies Wavelet function

Figure 4.4: Examples of Wavelet Functions

Lastly, we define the dilated and translated wavelet functions $\psi_{j,k}(x)$'s as fol-

lows:

$$\psi_{j,k}(x) \doteq 2^{j/2} \psi(2^j x - k) = 2^{j/2} \psi\left(\frac{x - k2^{-j}}{2^{-j}}\right)$$

It can be shown that the set of wavelets $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. Notationwise, we denote the inner product of two functions $h_1(x)$ and $h_2(x)$ as $\langle h_1(x), h_2(x) \rangle \doteq \int_{-\infty}^{\infty} h_1(x) \overline{h_2(x)}$. Thus, for any function $q(x) \in L^2(\mathbb{R})$, $q(x) = \sum_{j,k \in \mathbb{Z}} \langle q(x), \psi_{j,k}(x) \rangle \psi_{j,k}(x)$.

Chapter 5

Wavelet Density Estimation

5.1 Overview

We will describe the first geolocation technique, wavelet density estimation, in this chapter. The discussion of wavelet density estimation will involve materials in Chapter 2 and Chapter 3, which are about Anscombe variance-stabilizing transform and Stein's Unbiased Risk Estimate(SURE). As explained in the introduction, a probe machine sends out ping packets to the target and all the landmarks. From the returned packets, the probe machine constructs a RTT sequence for each machine. For example, if we send out n ping packets to each machine, we will have a RTT sequence $(X_1^{target}, X_2^{target}, \dots, X_n^{target})$ for the target and a RTT sequence $(X_1^{landmark}, X_2^{landmark}, \dots, X_n^{landmark})$ for some landmark.

We assume that the RTT of a machine is a random variable and the elements in the machine's RTT sequence are independent and identically distributed(i.i.d.) samples of the RTT random variable. Wavelet density estimation creates a RTT distribution from these i.i.d. samples in the RTT sequence. Next, wavelet density estimation utilizes wavelets $\psi_{j,k}$'s as orthonormal basis in $L^2(R)$ to construct estimated probability density function(pdf) from the RTT distribution. For instance, wavelet density estimation will construct an estimated pdf of target's RTT from $(X_1^{target}, X_2^{target}, \dots, X_n^{target})$. After estimated pdfs of target and landmarks have

been constructed, we will use a symmetrized version of Kullback-Leibler(KL) divergence to measure how similar two estimated pdfs are.

If two estimated pdfs of two machines' RTTs are similar under the aforementioned KL divergence, we assume that the two machines are geographically close to each other. Two nearby machines share a large portion of network infrastructure under the same Internet service provider(ISP). Almost identical network infrastructure generates similar RTT distributions for the two nearby machines. Similar RTT distributions will result in similar estimated pdfs. Thus, the physical closeness of two machines can be inferred from the degree of similarity between two estimated pdfs. Therefore, wavelet density estimation geolocates target to the landmark that has the most similar estimated pdf with respect to target's estimated pdf under the symmetrized KL divergence.

5.2 Wavelet Density Estimation

We will show how to construct an estimated pdf from a RTT sequence. We follow the ideas proposed in [8] and [9]. Let X_1, \dots, X_n be i.i.d. samples of pdf $f(x)$ with support $[0, 1]$. X_1, \dots, X_n are assumed to be the normalized elements of a RTT sequence. Partition $[0, 1]$ into $M = 2^{\lceil \log_2 n \rceil - 2}$ equally spaced subintervals where $\lceil \cdot \rceil$ denotes the closest integer function. For n close to a power-of-two number, $\lceil \log_2 n \rceil \approx \log_2 n$ and thus $2^{\lceil \log_2 n \rceil - 2} \approx 2^{\log_2 n - 2} = \frac{n}{4}$. Let N_i be the number of samples X_k 's that fall into i^{th} interval, for $i = 1, 2, \dots, M$. Let $1_{[a,b]}(\cdot)$ be the

indicator function for interval $[a, b]$. Then, we can express N_i as follows:

$$N_i \doteq \sum_{k=1}^n 1_{[\frac{i-1}{M}, \frac{i}{M}]}(X_k), \quad i = 1, 2, \dots, M.$$

N_i is *Binomial* (n, p_i) where $p_i = \int_{\frac{i-1}{M}}^{\frac{i}{M}} f(x) dx$ for $i = 1, \dots, M$. As $n \rightarrow \infty$, N_i 's are approximately independent. As $n \rightarrow \infty$, each subinterval's length, $\frac{1}{M} = \frac{4}{n}$, goes to zero. Thus, $p_i = \int_{\frac{i-1}{M}}^{\frac{i}{M}} f(x) dx \approx f(\frac{i}{M}) \cdot \frac{1}{M} \rightarrow 0$ as $n \rightarrow \infty$. But, $\alpha_i \doteq np_i \approx n \cdot f(\frac{i}{M}) \cdot \frac{1}{M} = n \cdot f(\frac{i}{M}) \cdot \frac{4}{n} = 4 \cdot f(\frac{i}{M})$ is fixed as $n \rightarrow \infty$. So, we can approximate such binomial random variable with Poisson random variable. $N_i \sim \text{Binomial}(n, p_i) \approx \text{Poisson}(\alpha_i = np_i = 4f(\frac{i}{M}))$. Thus, as $n \rightarrow \infty$, N_i is approximately *Poisson* $(\alpha_i = 4f(\frac{i}{M}))$.

Next, we will use Anscombe variance-stabilizing transform to transform a Poisson random variable into a Gaussian random variable. By using Anscombe transform, let $Y_i = 2\sqrt{N_i + \frac{3}{8}}$. Then, for large sample size n , $Y_i \approx N(2\sqrt{4f(\frac{i}{M})}, 1) = N(4\sqrt{f(\frac{i}{M})}, 1)$ and Y_i 's are approximately independent. So, $Y_i \approx 4\sqrt{f(\frac{i}{M})} + z_i$, where z_i 's are i.i.d. $N(0,1)$. For notational simplicity, we will ignore the scaling and the square-root operations. We will consider $Y_i \approx f(\frac{i}{M}) + z_i$, where z_i 's are i.i.d. $N(0,1)$. Once we obtain the estimates of means of Y_i 's, we can divide the estimates by four and square them to compensate the scaling and the square-root operations. Therefore, Y_i 's can be viewed as the sampled values of density $f(x)$ with noise components z_i 's.

Now, we will project the noisy samples of density $f(x)$ onto wavelet basis $\psi_{j,k}$'s

through wavelet transform. Let

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix} = \underbrace{\begin{bmatrix} f(\frac{1}{M}) \\ \vdots \\ f(\frac{M}{M}) \end{bmatrix}}_{\doteq \mathbf{F}} + \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_M \end{bmatrix}}_{\doteq \mathbf{Z}}$$

$$\tilde{\Theta} = \begin{bmatrix} \tilde{\theta}_{J-1,2^{J-1}-1} \\ \vdots \\ \tilde{\theta}_{-1,0} \end{bmatrix} = WY$$

where W is an orthogonal matrix implementing the wavelet transform of orthonormal wavelet basis $\psi_{j,k}$'s[9]. $\tilde{\theta}_{j,k}$'s are the observed wavelet coefficients of basis $\psi_{j,k}$'s. $J = \log_2 M$ as this is the finest resolution on density $f(x)$. $\tilde{\theta}_{-1,0}$ denotes the scaling coefficient at the coarsest level. Since $Y = F + Z$, $\tilde{\Theta} = WF + WZ$. WF represents the true wavelet coefficients and WZ represents the noise components. Let $\theta_{j,k}$'s denote the true wavelet coefficients and let $u_{j,k}$'s denote the noise components, i.e.,

$$\Theta = \begin{bmatrix} \theta_{J-1,2^{J-1}-1} \\ \vdots \\ \theta_{-1,0} \end{bmatrix} \doteq WF$$

$$U = \begin{bmatrix} u_{J-1,2^{J-1}-1} \\ \vdots \\ u_{-1,0} \end{bmatrix} \doteq WZ$$

Since W is an orthogonal matrix and z_i 's are i.i.d. Gaussian $N(0,1)$, noise components $u_{j,k}$'s are jointly Gaussian with mean $E(U) = E(WZ) = WE(Z) = W \cdot \bar{0} = \bar{0}$

and covariance $COV(U) = COV(WZ) = E[(WZ)(WZ)^T] = E[WZZ^TW^T] = WE[ZZ^T]W^T = WIW^T = WW^T = WW^{-1} = I$. So, the transformed noise components $u_{j,k}$'s are also i.i.d. Gaussian $N(0,1)$.

Donoho and Johnstone[9] proposed using soft threshold estimator $\hat{\theta}_{j,k} = \delta_{\lambda_j}^{soft}(\tilde{\theta}_{j,k})$, $k = 0, 1, \dots, 2^j - 1$ to estimate the true coefficients $\theta_{j,k}$'s at level j with a still-to-be-determined level-dependent threshold λ_j . The L^2 risk of this soft threshold estimator is

$$\begin{aligned} E_{\Theta}[\sum_j \sum_k (\hat{\theta}_{j,k} - \theta_{j,k})^2] &= E_{\Theta}[\sum_j \sum_k (\delta_{\lambda_j}^{soft}(\tilde{\theta}_{j,k}) - \theta_{j,k})^2] \\ &= \sum_j E_{\Theta}[\sum_k (\delta_{\lambda_j}^{soft}(\tilde{\theta}_{j,k}) - \theta_{j,k})^2]. \end{aligned}$$

The L^2 risk at level j is $E_{\Theta}[\sum_k (\delta_{\lambda_j}^{soft}(\tilde{\theta}_{j,k}) - \theta_{j,k})^2]$. We want to choose the level-dependent threshold λ_j to minimize the L^2 risk at level j . To do this, we will find an unbiased estimate of the L^2 risk at level j and then will minimize the unbiased estimate by choosing the proper λ_j . Recall from Chapter 3 that the Stein's Unbiased Risk Estimate(SURE) of the L^2 risk of soft threshold estimator with threshold t is

$$SURE(t; \tilde{\theta}_{j,k}'s) = 2^j - 2 \cdot \#\{k : |\tilde{\theta}_{j,k}| \leq t\} + \sum_{k=1}^{2^j} [\min(|\tilde{\theta}_{j,k}|, t)]^2$$

Now, the threshold t is selected to minimize the unbiased estimate $SURE(t; \tilde{\theta}_{j,k}'s)$:

$$\begin{aligned} \lambda_j &= \arg \min_{t \geq 0} SURE(t; \tilde{\theta}_{j,k}'s) \\ &= \arg \min_{t \geq 0} 2^j - 2 \cdot \#\{k : |\tilde{\theta}_{j,k}| \leq t\} + \sum_{k=1}^{2^j} [\min(|\tilde{\theta}_{j,k}|, t)]^2 \end{aligned}$$

We will show that there is not much computation in finding λ_j . Without loss of

generality, assume that $\tilde{\theta}_{j,k}$'s are arranged such that $|\tilde{\theta}_{j,0}| \leq |\tilde{\theta}_{j,1}| \leq \dots \leq |\tilde{\theta}_{j,2^j-1}|$. $SURE(t; \tilde{\theta}_{j,k}'s)$ is strictly increasing when t goes from $|\tilde{\theta}_{j,i}|$ to $|\tilde{\theta}_{j,i+1}|$ excluding both ends $|\tilde{\theta}_{j,i}|$ and $|\tilde{\theta}_{j,i+1}|$ because the term $\sum_{k=1}^{2^j} [\min(|\tilde{\theta}_{j,k}|, t)]^2$ increases while the remaining terms stay constant. Similarly, $SURE(t; \tilde{\theta}_{j,k}'s)$ is strictly increasing when t goes from 0 to $|\tilde{\theta}_{j,0}|$ excluding both ends 0 and $|\tilde{\theta}_{j,0}|$. Also, $SURE(t; \tilde{\theta}_{j,k}'s)$ stays constant for $t \geq |\tilde{\theta}_{j,2^j-1}|$. So, minimum of $SURE(t; \tilde{\theta}_{j,k}'s)$ must occur at 0 or at one of $|\tilde{\theta}_{j,k}|$'s.

After the soft threshold λ_j for level j is determined through $SURE(t; \tilde{\theta}_{j,k}'s)$, the observed wavelet coefficients at level j , $\tilde{\theta}_{j,k}$'s, are soft thresholded accordingly. This process is done for all the levels $j = 0, \dots, J-1$.

As pointed out by Donoho and Johnstone[9], the SURE-based selection of soft threshold does not perform well when most of the true wavelet coefficients are zero. This is because the contribution to SURE estimate from the noise of most coordinates of the multivariate Gaussian with zero means will “swamp” the little information provided by the few coordinates with non-zero means. The situation in which most wavelet coefficients are zeros is called the sparsity of wavelet coefficients. Donoho and Johnstone then proposed to check the sparsity of wavelet coefficients $\tilde{\theta}_{j,k}$'s at a fixed level j as follows:

Let $d = 2^j$ (the total number of wavelet coefficients at level j)

$$s^2 = \frac{\sum_{k=0}^{d-1} (\tilde{\theta}_{j,k}^2 - 1)}{d}$$

$$\gamma = \frac{(\log_2 d)^{\frac{3}{2}}}{\sqrt{d}} \text{ (the critical value of sparsity)}$$

If $s^2 \leq \gamma$, then the wavelet coefficients at level j are determined to be sparse and the soft threshold will be defaulted to a fixed value $\sqrt{2\log(d)} = \sqrt{2\log(2^j)}$.

If $s^2 > \gamma$, then the wavelet coefficients at level j are determined to be not sparse and the soft threshold will be determined through $SURE(t; \tilde{\theta}_{j,k}'s)$.

This method, called “SureShrink”, will first test the sparsity of wavelet coefficients at each level j to determine whether the subsequent threshold selection is to be based on SURE estimate or to use a fixed threshold $\sqrt{2\log(2^j)}$.

The “SureShrink” method is summarized as follows:

1. From the noisy observable samples of the unknown density $f(x)$, $\{Y_i \sim N(f(\frac{i}{M}), 1), i = 1, \dots, M\}$, we will perform wavelet transform to get the noisy observable wavelet coefficients $\tilde{\theta}_{j,k}$, $j = 0, \dots, J - 1$, $k = 0, \dots, 2^j - 1$, where $J = \log_2 M$.
2. For each fixed level j , we will soft threshold the wavelet coefficients $\tilde{\theta}_{j,k}$'s at that level j with soft threshold λ_j selected as follows:

$$\begin{aligned} \lambda_j &= \sqrt{2\log(2^j)} && \text{if } \frac{\sum_{k=0}^{2^j-1} (\tilde{\theta}_{j,k}^2 - 1)}{2^j} \leq \frac{(\log_2 2^j)^{\frac{3}{2}}}{\sqrt{2^j}} \\ &= \arg \min_{t \geq 0} SURE(t; \tilde{\theta}_{j,k}'s) && \text{if } \frac{\sum_{k=0}^{2^j-1} (\tilde{\theta}_{j,k}^2 - 1)}{2^j} > \frac{(\log_2 2^j)^{\frac{3}{2}}}{\sqrt{2^j}} \\ \hat{\theta}_{j,k} &= \delta_{\lambda_j}^{soft}(\tilde{\theta}_{j,k}), \quad k = 0, \dots, 2^j - 1. \end{aligned}$$

3. From the estimated wavelet coefficients $\hat{\theta}_{j,k}$'s, we will perform inverse wavelet transform to get an estimated pdf $\hat{f}(x)$ of the unknown pdf $f(x)$.

5.3 Distance Metric Between Estimated PDFs

After getting the samples of two estimate densities $\hat{f}(\frac{i}{M})$ and $\hat{g}(\frac{i}{M})$, $i = 1, \dots, M$, we will measure their similarity with a symmetrized version of Kullback-Leibler divergence $D(\hat{f}||\hat{g})$,

$$D(\hat{f}||\hat{g}) \doteq \int_0^1 \hat{f}(x) \log_2\left(\frac{\hat{f}(x)}{\hat{g}(x)}\right) dx + \int_0^1 \hat{g}(x) \log_2\left(\frac{\hat{g}(x)}{\hat{f}(x)}\right) dx.$$

Since we only have discrete samples of \hat{f} and \hat{g} , we will use Simpson's rule[2] to do numerical integration for $D(\hat{f}||\hat{g})$. Simpson's rule offers better numerical integration result than midpoint and trapezoidal numerical integrations. By Simpson's rule, we have

$$\begin{aligned} \int_0^1 \hat{f} \log_2 \frac{\hat{f}}{\hat{g}} \approx & \frac{1}{3M} \left[4\hat{f}\left(\frac{1}{M}\right) \log_2 \frac{\hat{f}\left(\frac{1}{M}\right)}{\hat{g}\left(\frac{1}{M}\right)} + 2\hat{f}\left(\frac{2}{M}\right) \log_2 \frac{\hat{f}\left(\frac{2}{M}\right)}{\hat{g}\left(\frac{2}{M}\right)} \right. \\ & + \dots + \\ & \left. + 4\hat{f}\left(\frac{M-1}{M}\right) \log_2 \frac{\hat{f}\left(\frac{M-1}{M}\right)}{\hat{g}\left(\frac{M-1}{M}\right)} + \hat{f}(1) \log_2 \frac{\hat{f}(1)}{\hat{g}(1)} \right] \end{aligned}$$

5.4 Implementation Issues of Wavelet Density Estimation

To fit the theoretical framework, certain modifications and configurations are necessary to implement wavelet density estimation.

1. The elements of the RTT sequences acquired from machines in one experiment

will be scaled to $[0, 1]$ as follows:

$$\text{Let } X^{IP_1} = (X_1^{IP_1}, X_2^{IP_1}, \dots, X_n^{IP_1})$$

$$X^{IP_2} = (X_1^{IP_2}, X_2^{IP_2}, \dots, X_n^{IP_2})$$

\vdots

$$X^{IP_k} = (X_1^{IP_k}, X_2^{IP_k}, \dots, X_n^{IP_k})$$

where n is the total ping packets sent for each machine and k is the number of machines being probed in the experiment.

$$\text{Let } p = \max_{\substack{1 \leq s \leq n \\ 1 \leq r \leq k}} \{X_s^{IP_r}\}$$

$$q = \min_{\substack{1 \leq s \leq n \\ 1 \leq r \leq k}} \{X_s^{IP_r}\}$$

Then, the rescaled RTT values are:

$$X_l^{IP_{m,rescaled}} = \frac{X_l^{IP_m} - q}{p - q} \quad \forall l \in \{1, \dots, n\}, \forall m \in \{1, \dots, k\}$$

2. Dropped packets will be ignored. However, machines that have more than 20% drop rate will be discarded for geolocation purpose. This is to protect the integrity of raw data before further processing.
3. We will use periodic extension to alleviate boundary distortion when taking wavelet transform.
4. Daubechies 4 is selected as the wavelet basis for wavelet density estimation due to its orthonormality and its smoothness. Daubechies 4 wavelets are orthonormal basis in $L^2(\mathbb{R})$. Projecting noisy samples of a density function onto Daubechies

4 wavelets will transform original i.i.d. Gaussian noise into i.i.d. Gaussian noise in wavelet domain. Daubechies 4 wavelets are relatively smooth. The smoothness of Daubechies 4 will result in the smoothness of the reconstructed functions which are the estimated pdfs in our case.

5.5 Estimation of Known Densities with Wavelet Density Estimation

After presenting wavelet density estimation, we will use it to estimate a known density. The estimated pdf and the actual pdf will then be compared to evaluate the performance of wavelet density estimation. After some literature survey on RTT distributions, it appears that most RTT distributions are like gamma distribution[5]. This concurs with our observations that most of the RTT distributions collected from our testbed have gamma shape. Thus, we will use a known gamma density to generate the corresponding i.i.d. samples which will be used to construct an estimated pdf through wavelet density estimation.

The pdf of a *gamma*(α, λ) random variable is given by:

$$f(x) = \frac{\lambda (\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad \alpha > 0, \quad \lambda > 0, \quad 0 < x < \infty$$

$$where \quad \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Figures 5.1, 5.2, 5.3 and 5.4 contain the actual pdf of a gamma density with $\alpha = 1.5$ and $\lambda = 1$ and the estimated pdfs based on different sample sizes, i.e. 512, 1024, 2048 and 4096 samples. We employ mean absolute percentage error(MAPE): $\frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{f}_i - f_i|}{|f_i|}$ and mean absolute error(MAE): $\frac{1}{n} \sum_{i=1}^n |\hat{f}_i - f_i|$ to measure the similarity between the actual pdf and the estimated pdf. As shown, the four esti-

mated pdfs have gamma shape. As sample size increases through 512, 1024, 2048 and 4096 samples, the MAPE decreases through 51%, 42%, 10% and 5% and the MAE decreases through 0.0721, 0.0534, 0.0112 and 0.0091.

Next, wavelet density estimation will estimate a different gamma density with $\alpha = 4$ and $\lambda = 0.5$. Figures 5.5, 5.6, 5.7 and 5.8 contain the actual pdf $gamma(\alpha = 4, \lambda = 0.5)$ and the estimated pdfs based on different sample sizes, i.e. 512, 1024, 2048 and 4096 samples. The MAPE between the actual pdf and the estimated pdf is shown on respective plots. As in previous gamma density estimation with different gamma parameters, an increase in sample size leads to smaller estimation error in terms of MAPE. In detail, the MAPE decreases through 17%, 15%, 14% and 13% and the MAE decreases through 0.0059, 0.0051, 0.0037 and 0.0024 as sample size increases through 512, 1024, 2048 and 4096.

Some RTT distributions have gamma shape with Gaussian lobe at the tail. Therefore, wavelet density estimation will estimate the following mixture density $0.7 * gamma(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ where the Gaussian density $N(5, 1)$ will simulate a Gaussian lobe at the tail of the gamma density $gamma(\alpha = 1.5, \lambda = 1)$. Figure 5.17 plots out the mixture density. Figures 5.9, 5.10, 5.11 and 5.12 contain the actual mixture pdf and the estimated pdfs based on different sample sizes, i.e. 512, 1024, 2048 and 4096 samples. The MAPE decreases through 28%, 16%, 10% and 6% and the MAE decreases through 0.0288, 0.0157, 0.0105 and 0.0086 as sample size increases through 512, 1024, 2048 and 4096.

Wavelet density estimation will estimate another mixture density $0.9 * gamma(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ where the Gaussian density $N(17, 1)$ will simulate a Gaus-

sian lobe. Figure 5.18 shows the mixture density. Figures 5.13, 5.14, 5.15 and 5.16 show the actual mixture pdf and the estimated pdfs based on sample sizes of 512, 1024, 2048 and 4096 samples. The MAPE decreases through 19%, 14%, 13% and 11% and the MAE decreases through 0.0072, 0.0053, 0.0035 and 0.0028 as sample size increases through 512, 1024, 2048 and 4096.

The MAPEs and the MAEs of three different densities at various sample sizes are tabulated as follows:

| Sample Size | 512 | 1024 | 2048 | 4096 |
|--|-----|------|------|------|
| MAPE of $Gamma(\alpha = 1.5, \lambda = 1)$ | 51% | 42% | 10% | 5% |
| MAPE of $Gamma(\alpha = 4, \lambda = 0.5)$ | 17% | 15% | 14% | 13% |
| MAPE of $0.7 * Gamma(\alpha = 1.5, \lambda = 1)$ $+0.3 * N(5, 1)$ | 28% | 16% | 10% | 6% |
| MAPE of $0.9 * Gamma(\alpha = 4, \lambda = 0.5)$ $+0.1 * N(17, 1)$ | 19% | 14% | 13% | 11% |

| Sample Size | 512 | 1024 | 2048 | 4096 |
|---|---------|--------|--------|--------|
| MAE of $Gamma(\alpha = 1.5, \lambda = 1)$ | 0.0721 | 0.0534 | 0.0112 | 0.0091 |
| MAE of $Gamma(\alpha = 4, \lambda = 0.5)$ | 0.0059 | 0.0051 | 0.0037 | 0.0024 |
| MAE of $0.7 * Gamma(\alpha = 1.5, \lambda = 1)$ $+0.3 * N(5, 1)$ | 0.02888 | 0.0157 | 0.0105 | 0.0086 |
| MAE of $0.9 * Gamma(\alpha = 4, \lambda = 0.5)$ $+0.1 * N(17, 1)$ | 0.0072 | 0.0053 | 0.0035 | 0.0028 |

If we accept MAPE of less than 15% as satisfactory estimation based on visual inspection of the estimated pdfs, the minimum sample size to adequately estimate the four densities is about 2048 samples.

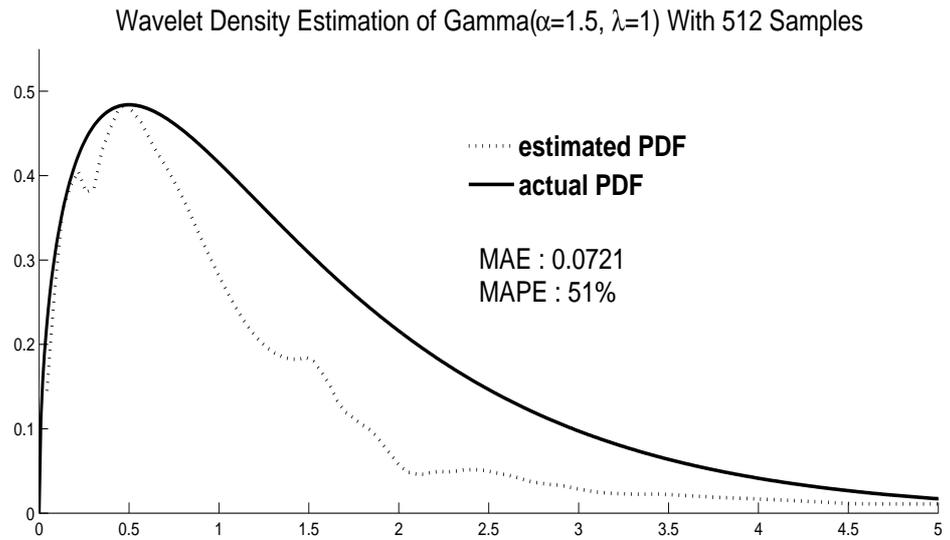


Figure 5.1: Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 512 samples.

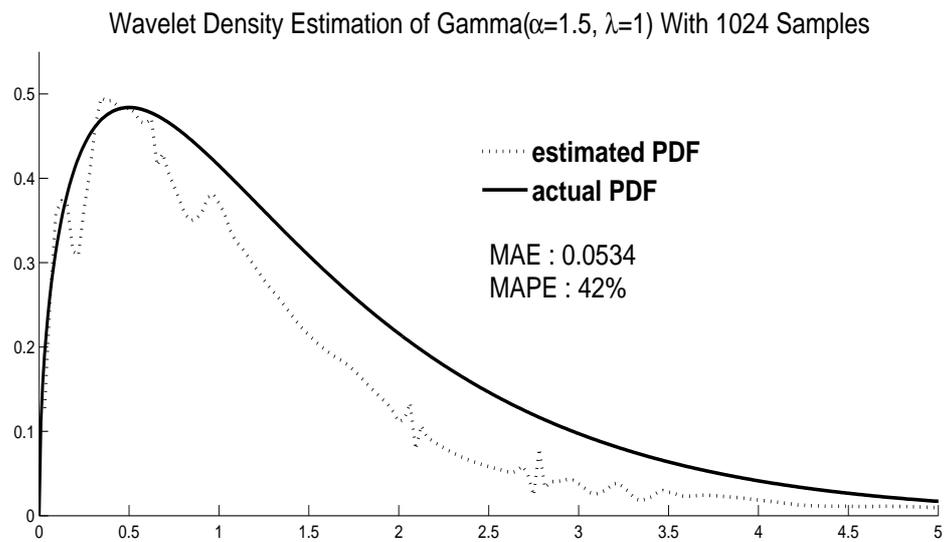


Figure 5.2: Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 1024 samples.

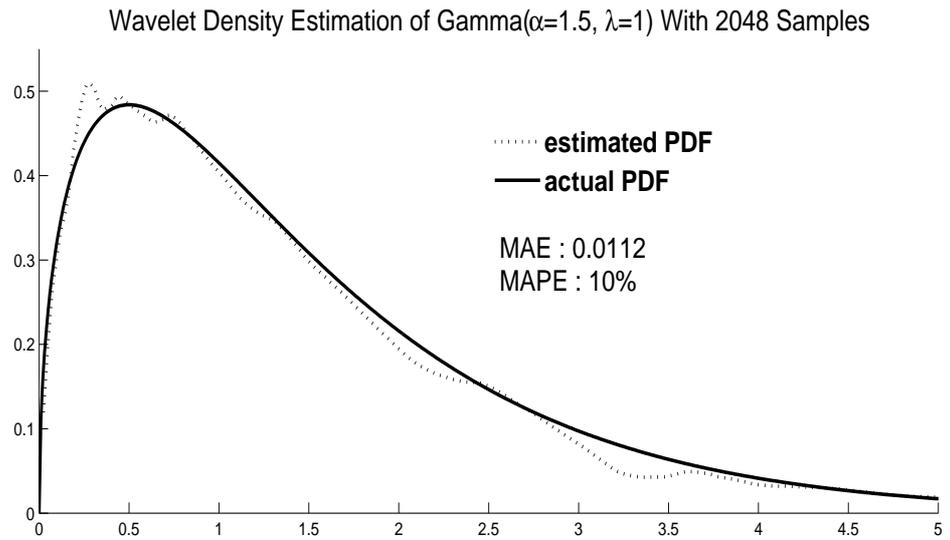


Figure 5.3: Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 2048 samples.

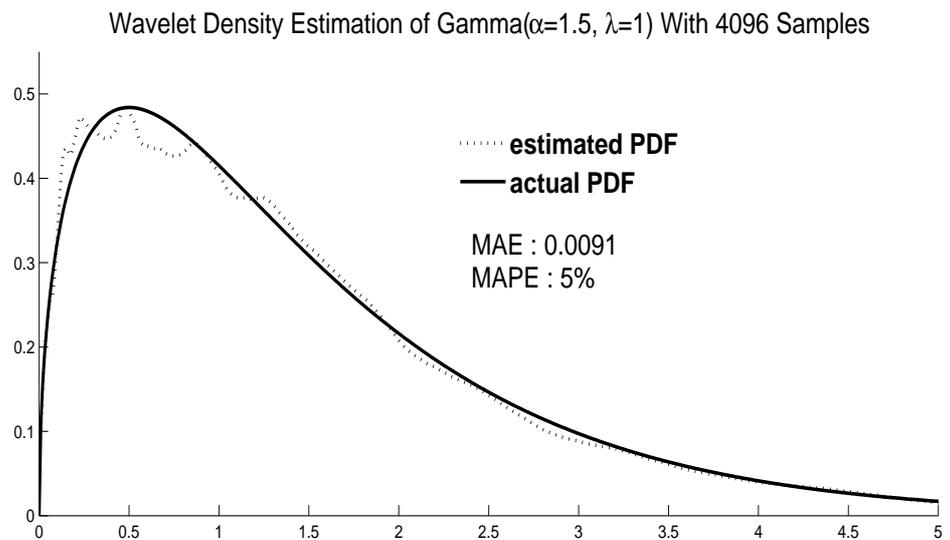


Figure 5.4: Wavelet density estimation of a gamma density of $\alpha = 1.5$ and $\lambda = 1$ with 4096 samples.

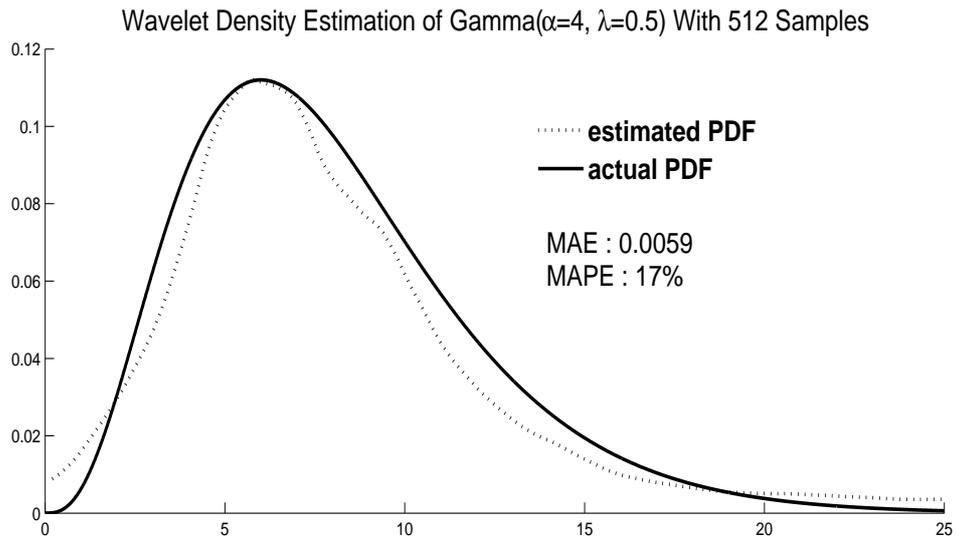


Figure 5.5: Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 512 samples.

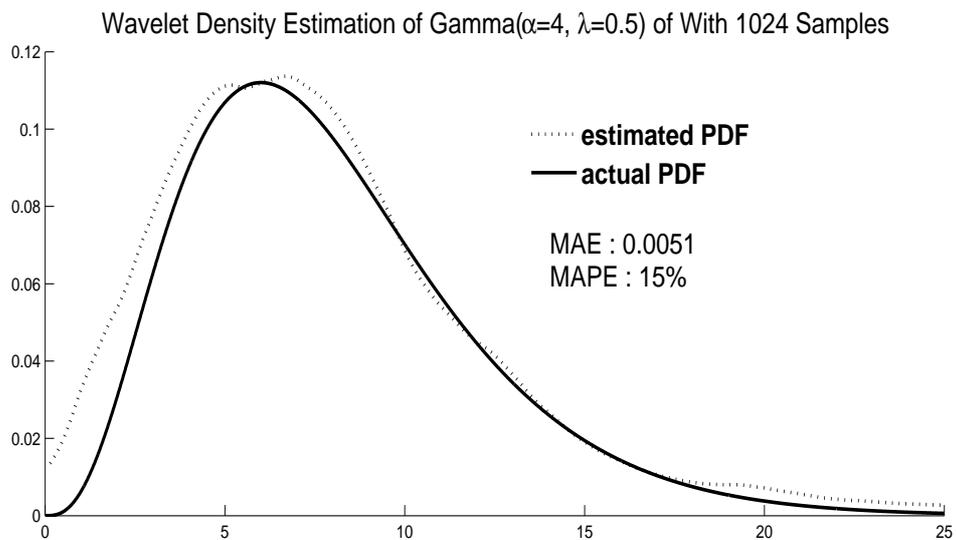


Figure 5.6: Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 1024 samples.

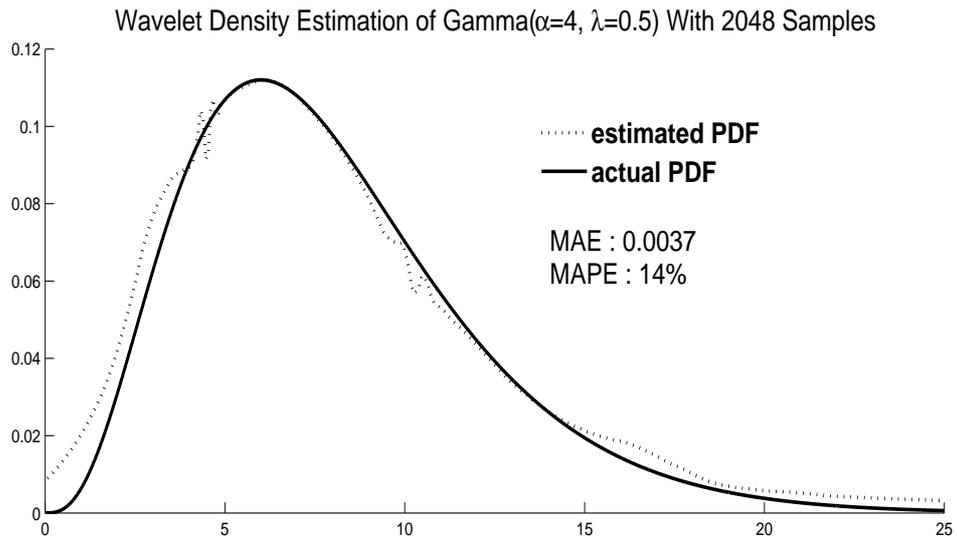


Figure 5.7: Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 2048 samples.

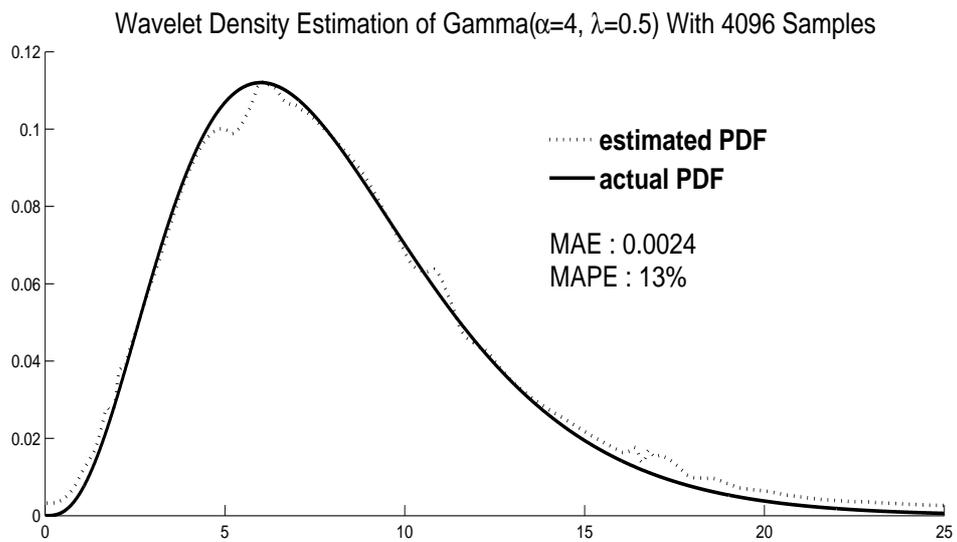


Figure 5.8: Wavelet density estimation of a gamma density of $\alpha = 4$ and $\lambda = 0.5$ with 4096 samples.

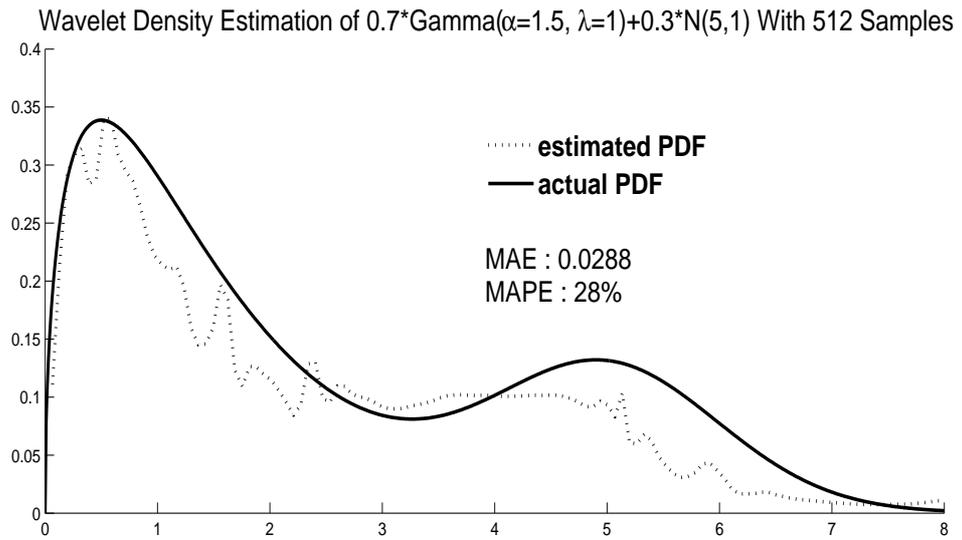


Figure 5.9: Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 512 samples.

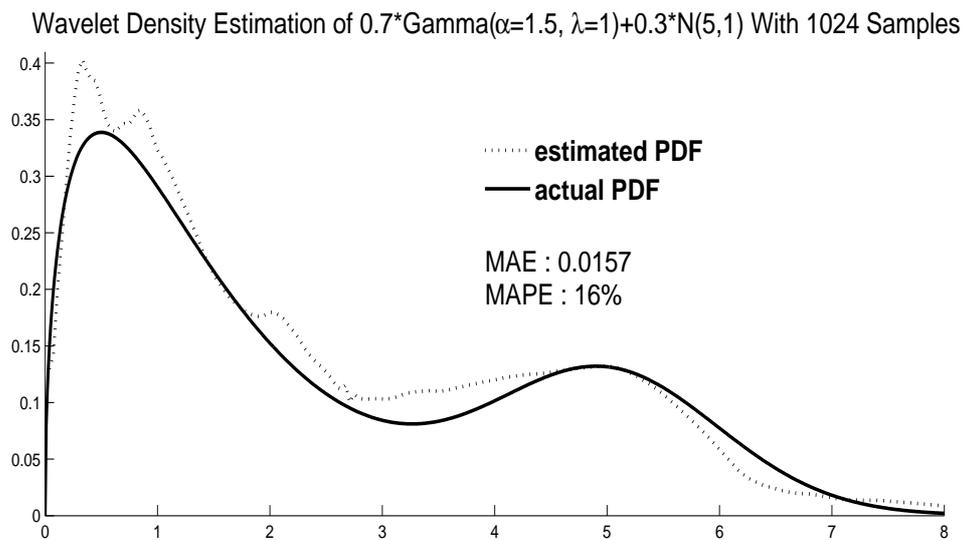


Figure 5.10: Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 1024 samples.

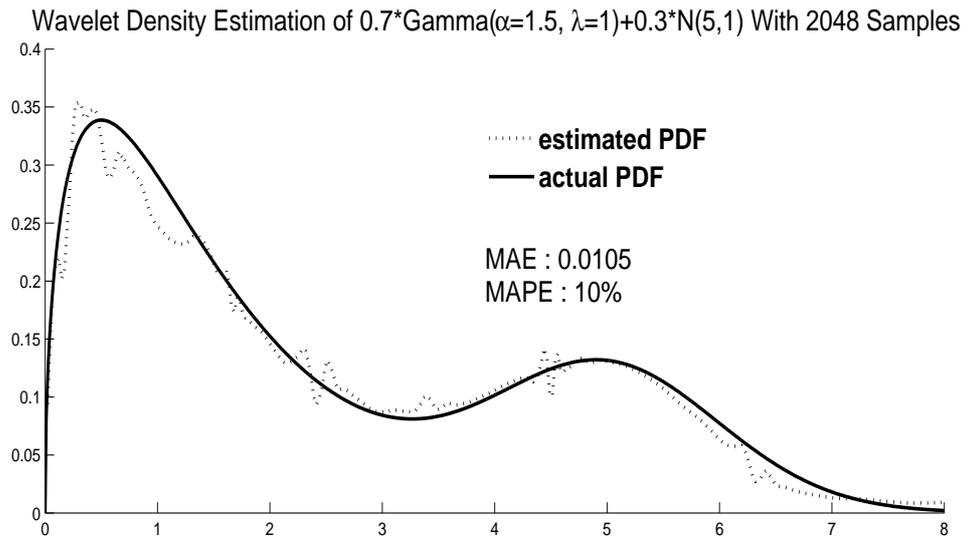


Figure 5.11: Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 2048 samples.

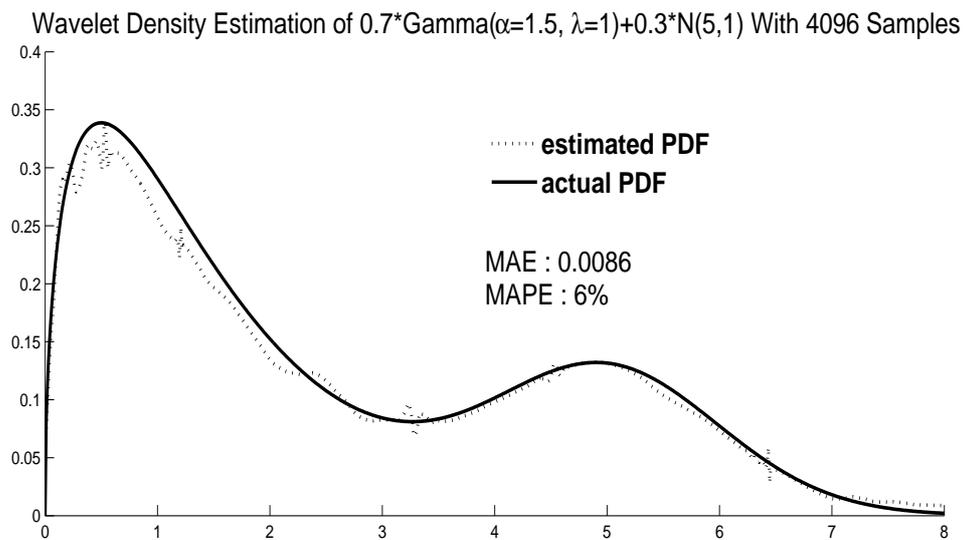


Figure 5.12: Wavelet density estimation of a mixture density $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$ with 4096 samples.

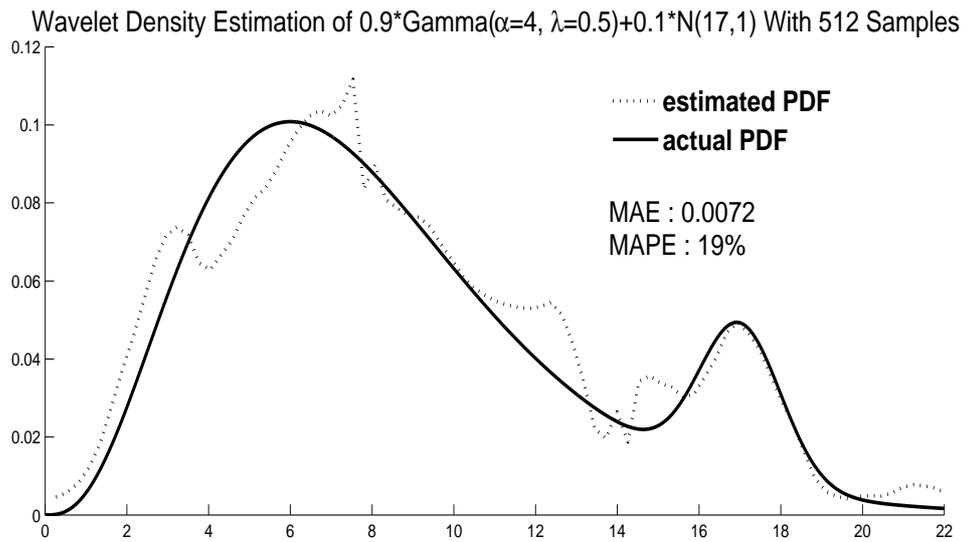


Figure 5.13: Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 512 samples.

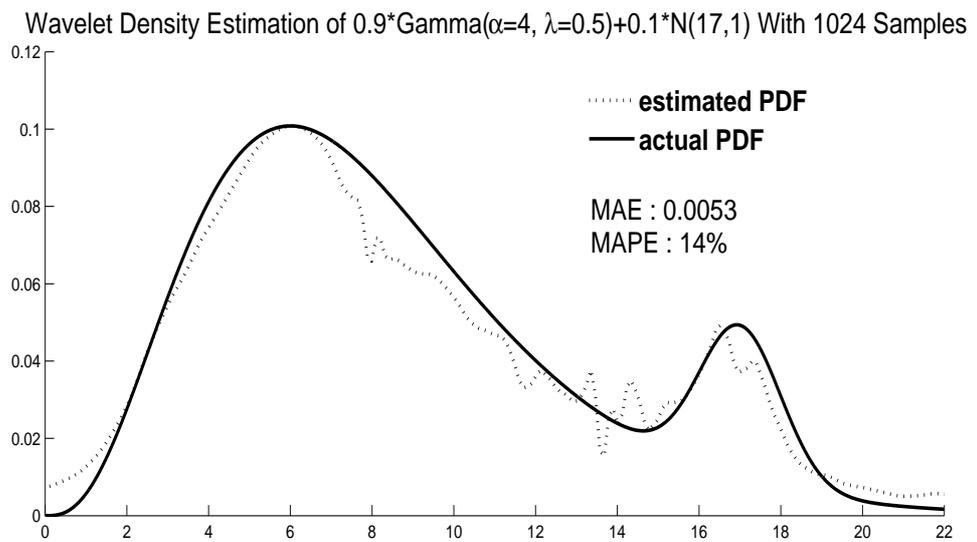


Figure 5.14: Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 1024 samples.

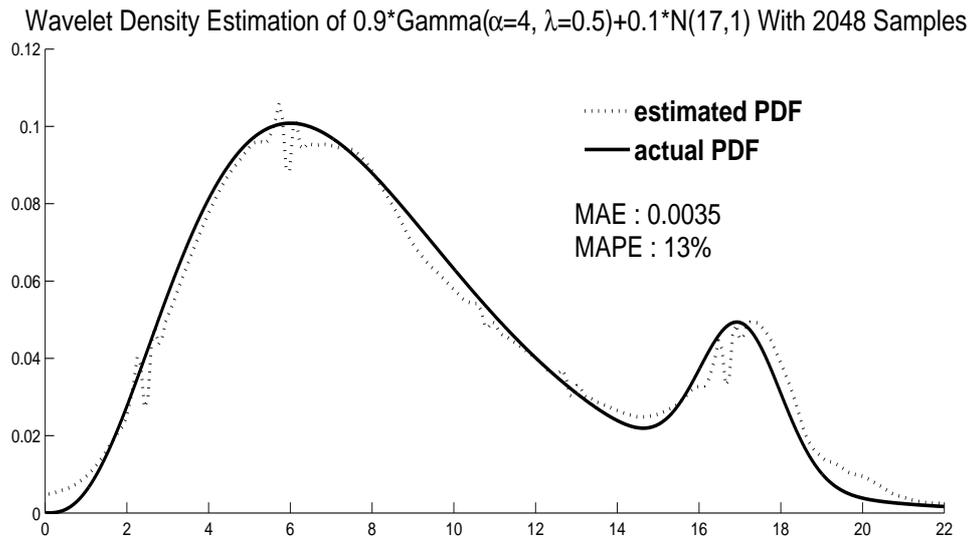


Figure 5.15: Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 2048 samples.

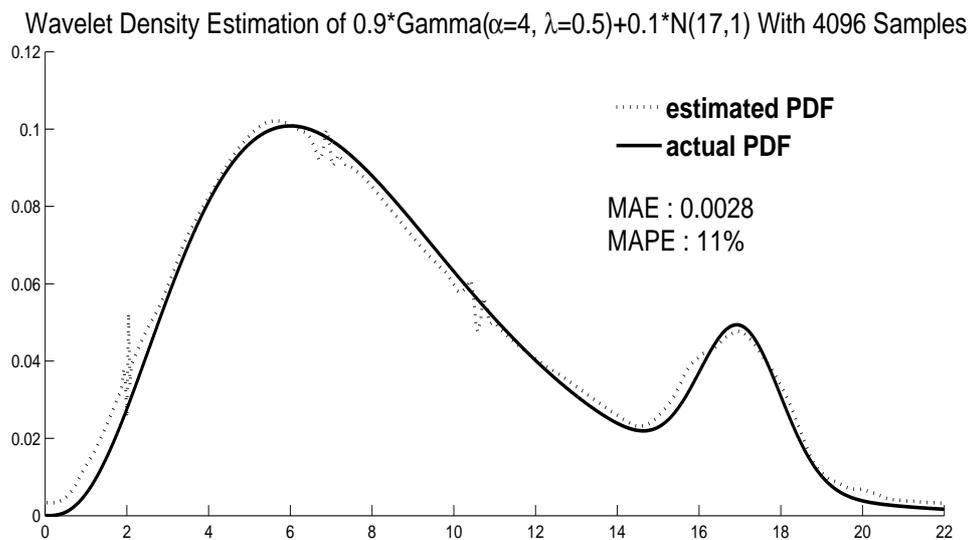


Figure 5.16: Wavelet density estimation of a mixture density $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$ with 4096 samples.

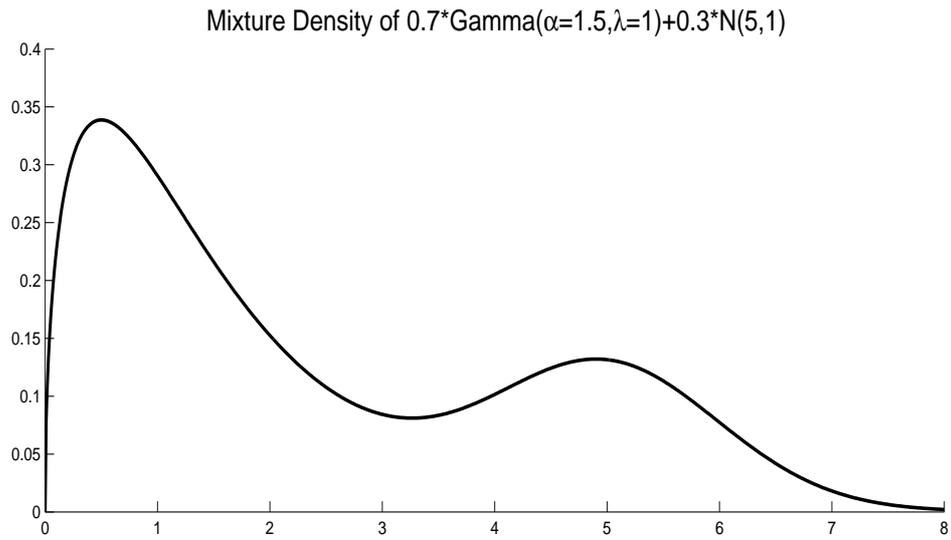


Figure 5.17: A mixture density, $0.7 * \text{gamma}(\alpha = 1.5, \lambda = 1) + 0.3 * N(5, 1)$

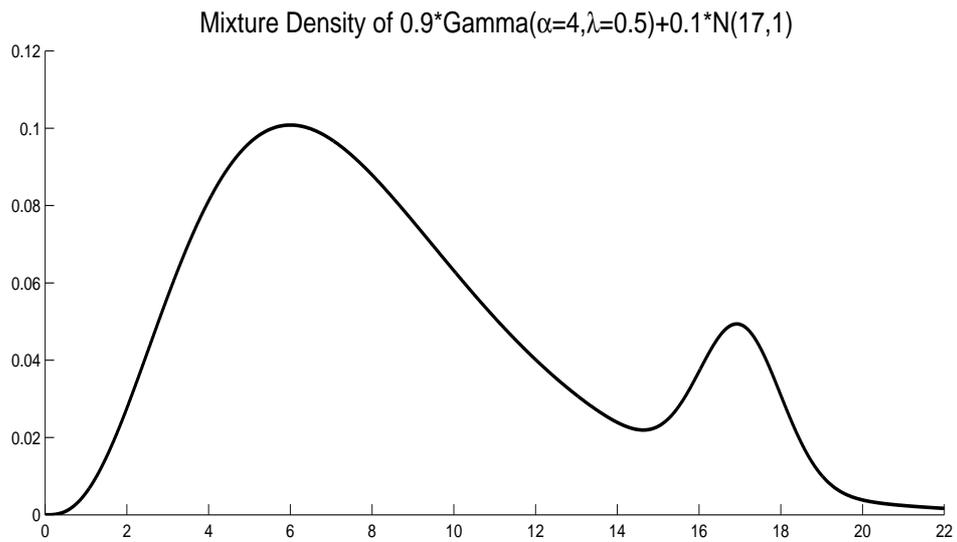


Figure 5.18: A mixture density, $0.9 * \text{gamma}(\alpha = 4, \lambda = 0.5) + 0.1 * N(17, 1)$

5.6 Pictorial Depiction of Wavelet Density Estimation

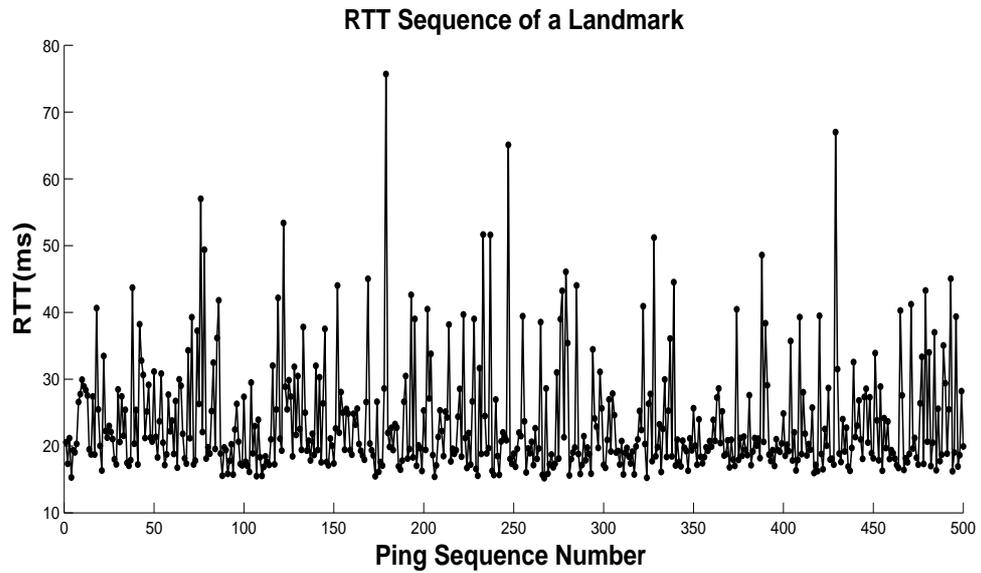


Figure 5.19: A 500-element RTT sequence collected from a landmark

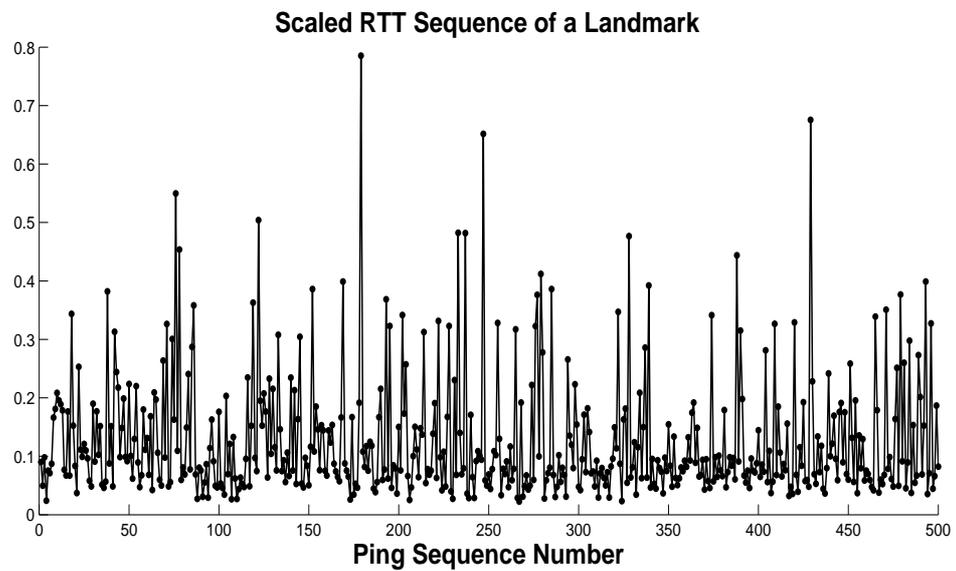


Figure 5.20: Rescaling the RTT values into $[0,1]$

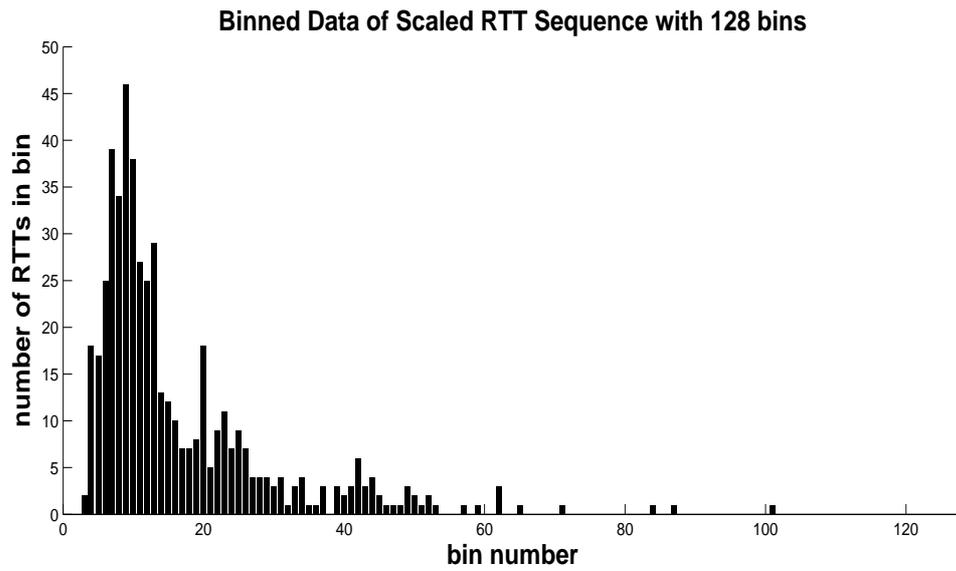


Figure 5.21: RTT distribution of scaled RTT values

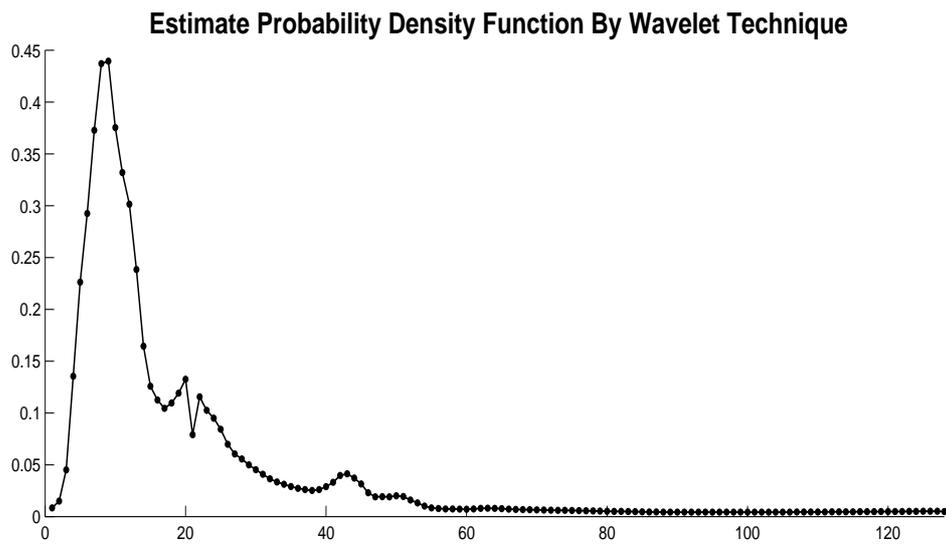


Figure 5.22: Estimated pdf constructed by wavelet density estimation

Chapter 6

Wavelet Time-Frequency Analysis

6.1 Overview

We discuss the second geolocation technique, wavelet time-frequency analysis, in this chapter. We will start with explaining the time synchronization among RTT sequences. A probe machine sends out time-synchronized ping packets to all machines. Thus, the elements of the RTT sequences collected for all the machines are synchronized in time. All RTT sequences share a common time axis. We want to analyze the frequency contents of these RTT sequences over short time intervals along the common time axis. For some signal $s(t)$ where t represents time, the analysis of its frequency content as a function of time t is called time-frequency analysis. Time-frequency analysis allows us to know how frequency content of $s(t)$ changes over time t .

Geographically-close machines under the same ISP share almost identical network infrastructure. These machines in close vicinity experience similar network activities at roughly the same time. Network activities cause significant variations in RTT values during short time intervals. These momentarily significant variations of RTT values will translate into an increase in frequency content at those time intervals. Wavelet time-frequency analysis employs wavelets to analyze frequency content over short time intervals. Wavelet time-frequency analysis can then ge-

locate a target to its nearby landmarks by detecting time-synchronized increases of frequency contents in target's RTT sequence and its nearby landmarks' RTT sequences.

6.2 Wavelet Time-Frequency Analysis

Fourier transform could analyze the frequency content of a signal $s(t)$. Let $s(t)$ be relatively smooth except with a sharp spike during some short time interval. The Fourier transform of $s(t)$ will have high frequency component to indicate the existence of the sharp spike. But, Fourier transform of $s(t)$ could not provide time information about when the sharp spike happens along the time axis of the signal $s(t)$.

To locate the sharp burst along the time axis, we need to analyze the frequency content of the signal $s(t)$ as a function of time t . This is called time-frequency analysis. Wavelet can perform time-frequency analysis due to its time-frequency localization property. The following explanations of time-frequency analysis with wavelet could be found in [11]. We will define the following notations:

| | | |
|---------------------|--|---|
| Inner Product | $\langle f, g \rangle$ | $\doteq \int_{-\infty}^{\infty} f(x)\overline{g(x)} dx$ |
| Fourier Transform | $\hat{f}(w)$ | $\doteq \int_{-\infty}^{\infty} f(x)e^{-jwx} dx$ |
| Parseval's Identity | $\langle f, g \rangle$ | $= \frac{1}{2\pi} \langle \hat{f}, \hat{g} \rangle$ |
| | $\int_{-\infty}^{\infty} f(x)\overline{g(x)} dx$ | $= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w)\overline{\hat{g}(w)} dw$ |
| Wavelet | $\psi_{(a,b)}(x)$ | $\doteq a^{-\frac{1}{2}}\psi\left(\frac{x-b}{a}\right)$ |

(Real-Valued) where b = translation index to

translate/shift wavelet

a = dilation index to

dilate/scale wavelet

| | | |
|------------|--------------------|---|
| Continuous | $(W_\psi f)(a, b)$ | $\doteq \langle f, \psi_{(a,b)} \rangle$ |
| Wavelet | | $= a^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx$ |
| Transform | | |

Suppose wavelet $\psi(t)$ is localized in time and frequency which means its energy is localized in time domain and in frequency domain. Such wavelet is said to possess the property of time-frequency localization. Assume the effective support of $\psi(t)$ has center t^* and radius(=distance from center) Δ_ψ on the x-axis. Then, $\psi_{(a,b)} = a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)$ has effective support with center $b + at^*$ and radius $a\Delta_\psi$.

Continuous wavelet transform $(W_\psi f)(a, b) = a^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$ will produce a windowing effect on $f(t)$ with time window $[b + at^* - a\Delta_\psi, b + at^* + a\Delta_\psi]$. Thus, only the portion of $f(t)$ at $t \in [b + at^* - a\Delta_\psi, b + at^* + a\Delta_\psi]$ is analyzed. Notice that the window $[b + at^* - a\Delta_\psi, b + at^* + a\Delta_\psi]$ can be shifted left or right by shifting wavelet $\psi_{(a,b)}(t)$ with its translation index b and can be shrunk or stretched by scaling wavelet $\psi_{(a,b)}(t)$ with its dilation index a. Thus, by shifting and scaling wavelet $\psi_{(a,b)}(t)$, we can analyze function $f(t)$ at any interval with any width. In short, we can do a time-localized analysis on $f(t)$.

To see the corresponding frequency-localized analysis in frequency domain, we will derive the Fourier transform of $\psi_{(a,b)}(t)$:

$$\begin{aligned}
\hat{\psi}_{(a,b)}(w) &= \int_{-\infty}^{\infty} a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) e^{-jw t} dt \\
&\quad \text{Let } x = \frac{t-b}{a} \\
&= \int_{-\infty}^{\infty} a^{-\frac{1}{2}} \psi(x) e^{-jw(ax+b)} a dx \\
&= a^{-\frac{1}{2}} a e^{-jwb} \int_{-\infty}^{\infty} \psi(x) e^{-j(wa)x} dx \\
&= \sqrt{a} e^{-jwb} \hat{\psi}(aw)
\end{aligned}$$

By Parseval's identity,

$$\begin{aligned}
(W_{\psi} f)(a, b) &= \int_{-\infty}^{\infty} f(x) \psi_{(a,b)}(x) dx \stackrel{\text{Parseval's}}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w) \overline{\hat{\psi}_{(a,b)}(w)} dw \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w) \overline{\sqrt{a} e^{-jwb} \hat{\psi}(aw)} dw \\
&= \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} \hat{f}(w) \overline{\hat{\psi}(aw)} e^{jwb} dw
\end{aligned}$$

Assume the effective support of $\hat{\psi}(w)$ has center w^* and radius $\Delta_{\hat{\psi}}$ on the frequency w -axis. Then, $\hat{\psi}_{(a,b)}(w) = \sqrt{a} e^{-jwb} \hat{\psi}(aw)$ has effective support with center $\frac{w^*}{a}$ and radius $\frac{\Delta_{\hat{\psi}}}{a}$. Thus, continuous wavelet transform $\frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} (\hat{f}(w) \overline{\hat{\psi}(aw)}) e^{jwb} dw$ will produce a windowing effect in frequency domain with frequency window $[\frac{w^*}{a} - \frac{\Delta_{\hat{\psi}}}{a}, \frac{w^*}{a} + \frac{\Delta_{\hat{\psi}}}{a}]$. Notice that the frequency window stretches out for small a and shrinks for large a . Thus, by adjusting a , we can analyze different frequency contents at different frequency windows. In short, we can do a frequency-localized analysis.

By changing a and b , time-frequency localization of wavelet $\psi_{(a,b)}$ permits analyzing a portion of the function $f(t)$ at time $t \in [b + at^* - a\Delta_{\psi}, b + at^* + a\Delta_{\psi}]$ for frequency content at frequency $w \in [\frac{w^*}{a} - \frac{\Delta_{\hat{\psi}}}{a}, \frac{w^*}{a} + \frac{\Delta_{\hat{\psi}}}{a}]$. Translation index b is responsible for moving wavelet $\psi_{(a,b)}$ and thus the time window $[b + at^* - a\Delta_{\psi}, b +$

$at^* + a\Delta_\psi]$ along the time axis so that portion of the function $f(t)$ inside the time window could be analyzed regardless of the rest of the function $f(t)$ outside the time window. Dilation index a is responsible for stretching or compressing wavelet $\psi_{(a,b)}(t)$ and thus the time window $[b + at^* - a\Delta_\psi, b + at^* + a\Delta_\psi]$ but inversely compressing or stretching frequency window $[\frac{w^*}{a} - \frac{\Delta\hat{\psi}}{a}, \frac{w^*}{a} + \frac{\Delta\hat{\psi}}{a}]$ so that different frequency contents could be analyzed. By changing the translation index b and the dilation index a , we could analyze the frequency content as a function of time. This means that due to the time-frequency localization of wavelet, we could use wavelet to do time-frequency analysis.

For an illustrative purpose, consider the following plot:

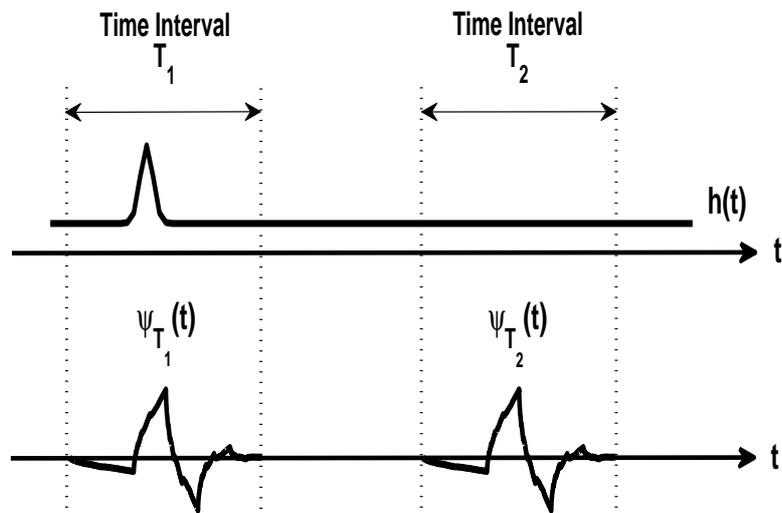


Figure 6.1: Wavelet Time-Frequency Analysis

A signal $h(t)$ is an relatively smooth except with a significant variation over time interval T_1 . The signal $\psi_{T_1}(t)$ is a wavelet time-localized at time interval T_1 and the $\psi_{T_2}(t)$ is identical to $\psi_{T_1}(t)$ except for being shifted to time interval T_2 . So, $\psi_{T_2}(t)$ is also a wavelet time-localized at time interval T_2 . Let $h_{T_1}(t)$ be the truncated version of $h(t)$ to time interval T_1 and $h_{T_2}(t)$ be the truncated version of $h(t)$ to time interval T_2 .

$$\begin{aligned} W_{\psi_{T_1}} h &= \int_{-\infty}^{\infty} h(t) \psi_{T_1}(t) dt \\ &\approx \int_{-\infty}^{\infty} h_{T_1}(t) \psi_{T_1}(t) dt \stackrel{Parseval's}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{h}_{T_1}(w) \overline{\hat{\psi}_{T_1}(w)} dw \end{aligned}$$

Due to significant variations of $h_{T_1}(t)$, $\hat{h}_{T_1}(w)$ is localized at high frequency. Due to oscillatory nature of $\psi_{T_1}(t)$, $\hat{\psi}_{T_1}(w)$ is also localized at high frequency. Thus, $\int_{-\infty}^{\infty} \hat{h}_{T_1}(w) \hat{\psi}_{T_1}(w) dw > 0$. And, $W_{\psi_{T_1}} h > 0$. This could also be derived in time domain as $h_{T_1}(t)$ and $\psi_{T_1}(t)$ have good correlation in time domain with their peaks almost aligned in time.

Now, we will turn our focus to wavelet $\psi_{T_2}(t)$.

$$\begin{aligned} W_{\psi_{T_2}} h &= \int_{-\infty}^{\infty} h(t) \psi_{T_2}(t) dt \\ &\approx \int_{-\infty}^{\infty} h_{T_2}(t) \psi_{T_2}(t) dt \stackrel{Parseval's}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{h}_{T_2}(w) \overline{\hat{\psi}_{T_2}(w)} dw \end{aligned}$$

Due to smoothness of $h_{T_2}(t)$, $\hat{h}_{T_2}(w)$ is localized at low frequency. Due to oscillatory nature of $\psi_{T_2}(t)$, $\hat{\psi}_{T_2}(w)$ is localized at high frequency. Thus, $\int_{-\infty}^{\infty} \hat{h}_{T_2}(w) \hat{\psi}_{T_2}(w) dw \approx 0$. And, $W_{\psi_{T_2}} h \approx 0$. This could also be derived in time domain as $h_{T_2}(t)$ and $\psi_{T_2}(t)$ are hardly correlated in time domain.

In summary, by using wavelets, we can analyze the frequency content of $h(t)$ over two time intervals T_1 and T_2 . We can detect $h(t)$ to have high frequency content at T_1 and to have low frequency content at T_2 .

6.3 Time-Frequency Analysis With Wavelet To Geolocate

To do time-frequency analysis on a RTT sequence $\tilde{Y} = (Y_1, Y_2, \dots, Y_n)$ where Y_i 's are RTT values at equally spaced time instants, we will perform continuous wavelet transform on the RTT sequence \tilde{Y} . Let $Y(t)$ be the continuous RTT waveform of the discretized RTT sequence \tilde{Y} . Let $\tilde{\psi}_{(a,b)}$ be the discretized version of continuous-time wavelet $\psi_{(a,b)}(t)$. In computers, continuous wavelet transform $\int_{-\infty}^{\infty} Y(t)\psi_{(a,b)}(t) dt$ is performed by taking elementwise multiplication of \tilde{Y} and $\tilde{\psi}_{(a,b)}$ and then conducting numerical integration.

We will perform continuous wavelet transform at two scales, fine scale and coarser scale to conduct multiscale product analysis as introduced in [13].

Let $\tilde{V}^f = (V_1^f, V_2^f, \dots, V_n^f)$ be the wavelet coefficients of fine scale.

$\tilde{V}^c = (V_1^c, V_2^c, \dots, V_n^c)$ be the wavelet coefficients of coarser scale.

We will construct the multiscale product \tilde{V} as follows:

$$\tilde{V} = \tilde{V}^f \cdot (\tilde{V}^c)^T$$

$$(V_1, V_2, \dots, V_n) = (V_1^f \cdot V_1^c, V_2^f \cdot V_2^c, \dots, V_n^f \cdot V_n^c)$$

At fine scale, wavelets are time-localized to small finite support. These compressed wavelets could detect fine features of some sample function $f(t)$. At the same

time, the compressed wavelets will be more susceptible to pick up fine, tiny noise from the sample function $f(t)$. At coarser scale, wavelets are time-localized to larger support. The stretched-out wavelets will slightly smoothen out the fine features of the sample function $f(t)$. However, these stretched-out wavelets will smoothen out fine, tiny noise. Thus, the multiscale product of the fine-scale coefficients and the coarser-scale coefficients will amplify the features of the sample function $f(t)$ while reducing noise.

The absolute values of the multiscale product coefficients in \tilde{V} are thresholded into zeros and ones. Thresholding schemes that have been studied are standard deviation, mean and simple compression of certain percentage. Standard deviation as the threshold produced the best result and is decided to be the finalized thresholding scheme. Note that the ones in a zero-one sequence indicate sudden rises of frequency content at those portions of the RTT sequence.

After RTT sequences have been transformed into zero-one sequences, inner products will be taken as follows:

$$\frac{\langle x_l, x_t \rangle}{\langle x_t, x_t \rangle}, \quad \text{where } x_l \text{ denotes the zero-one sequence of a landmark.}$$

x_t denotes the zero-one sequence of target.

The inner product will have value between 0 and 1. If inner product is close to 1, that means the sudden rises of frequency contents of the target's RTT sequence and the landmark's RTT sequence are highly correlated in time. This will imply the network activities experienced by the target and the landmark are time-synchronized. This

will happen when the target and the landmark are geographically close. If inner product is close to 0, the frequency contents of the target's RTT sequence and the landmark's RTT sequence are not correlated in time.

6.4 Implementation Issues of Wavelet Time-Frequency Analysis

1. Machines that have more than a 5% drop rate of total packets sent will be discarded for geolocation purpose. This is to protect the integrity of raw data before further processing. As compared to the 20% drop rate allowed in wavelet density estimation, we choose a smaller drop rate for wavelet time-frequency analysis. Wavelet time-frequency analysis looks for temporal information in terms of frequency content evolution over time. High drop rate will have serious repercussion to this temporal information.
2. Missing RTT values in RTT sequences due to dropped packets will be interpolated with cubic spline.
3. We will use odd-symmetric extension to avoid creating artificial high frequency content at boundaries when taking wavelet transform.
4. We choose Daubechies 2 wavelet to perform time-frequency analysis due to its time-frequency localization property. Daubechies 2 also offered better performance for wavelet time-frequency analysis during the testing phase with other types of wavelets.
5. A landmark could produce a zero-one sequence with large number of 1's and few

0's. Then, a lot of targets irrespective of their geographic locations will be matched to this landmark. To solve this issue, we reason that a landmark close to the target should have almost the same number of 1's in its zero-one sequence as in target's zero-one sequence. Thus, we will compare the number of 1's in every landmark's zero-one sequence. Assume that target has p 1's in its zero-one sequence. If there are q 1's in a landmark's zero-one sequence with $q > p$, we will retain p 1's that correspond to multiscale product coefficients of the largest absolute values among the q coefficients. The remaining $(q-p)$ 1's will be converted to 0's.

6.5 Pictorial Depiction of Wavelet Time-Frequency Analysis

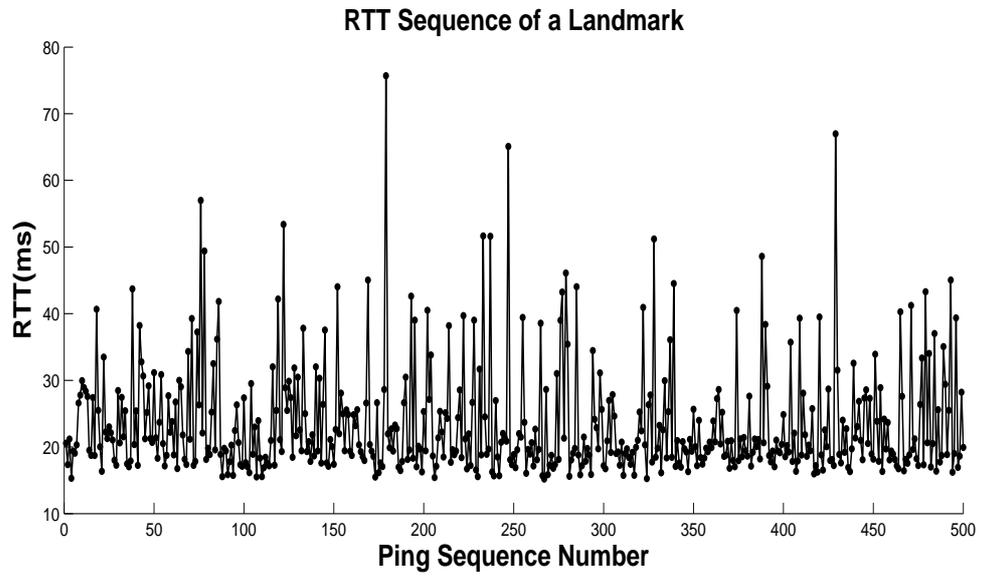


Figure 6.2: A 500-element RTT sequence collected from a landmark

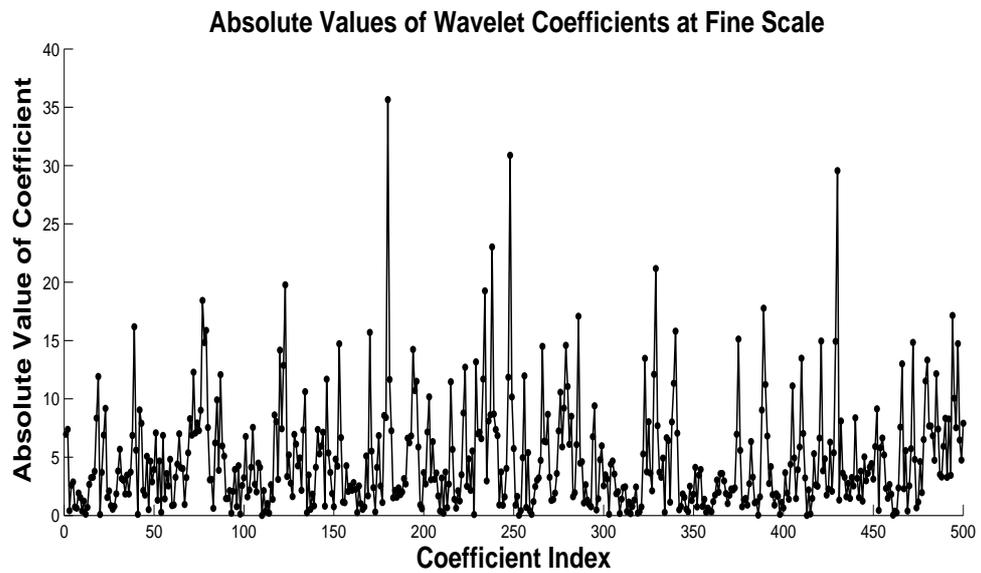


Figure 6.3: Wavelet transform of the RTT sequence at fine scale

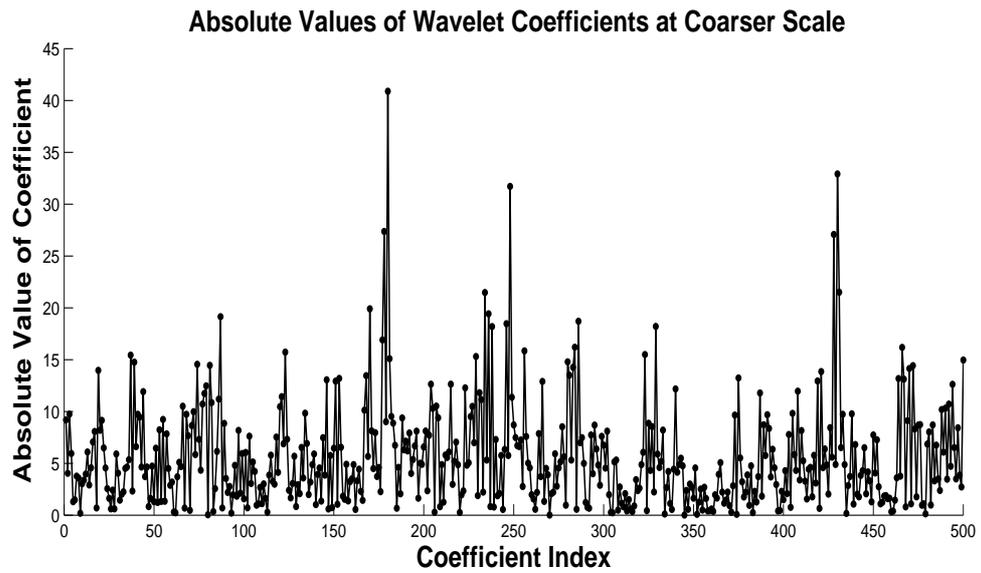


Figure 6.4: Wavelet transform of the RTT sequence at coarser scale

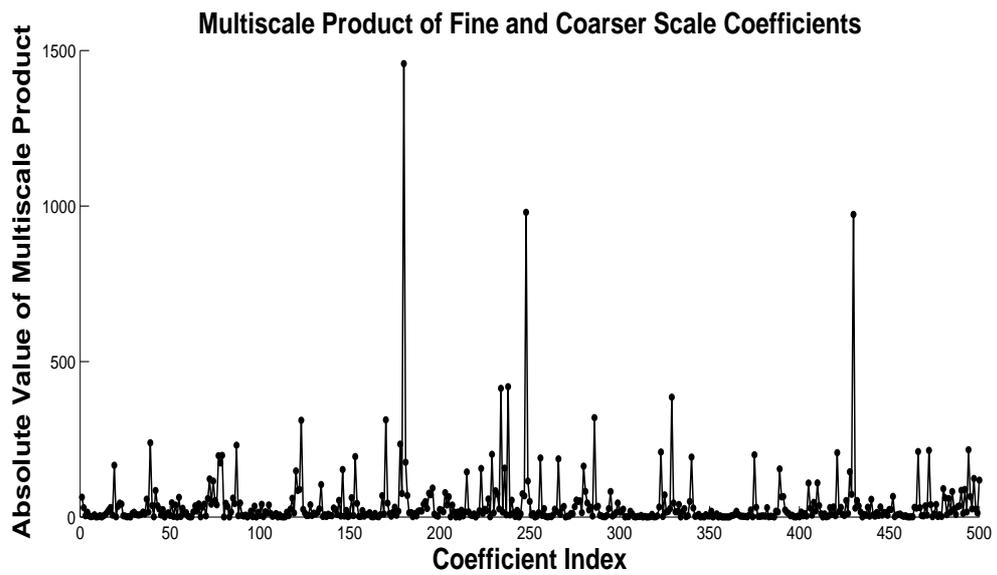


Figure 6.5: Multiscale product to produce enhanced features while reducing background noise

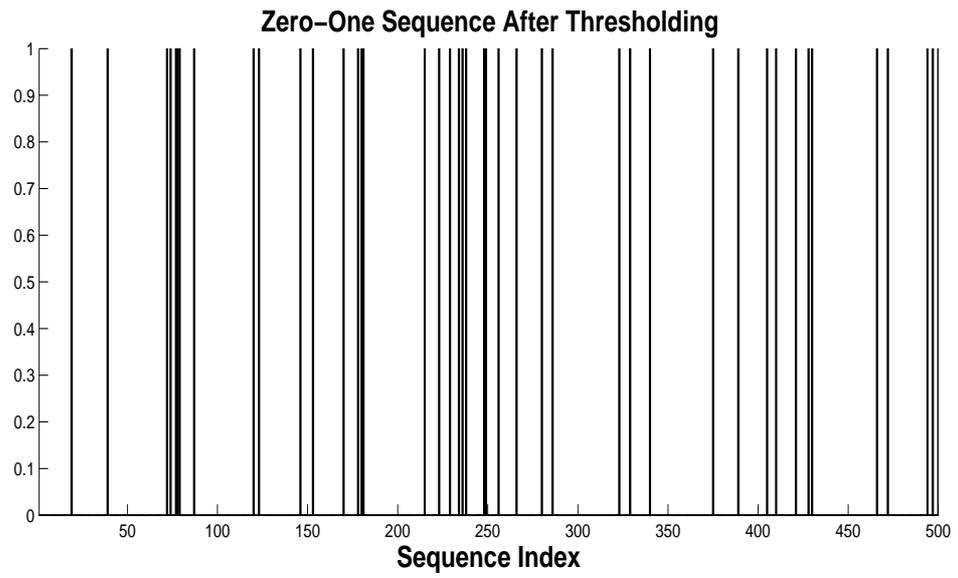


Figure 6.6: Zero-one sequence with ones indicating sudden rises of frequency content in the original RTT sequence

Chapter 7

Experiments and Results

We now analyze the performance of both geolocation techniques, wavelet density estimation(WDE) and wavelet time-frequency analysis(WTA). To do this, we will use data sets collected from our testbed described as follows:

| | |
|---------------------------------|--|
| Testbed | : Comcast network in the Baltimore-Washington D.C. Metropolitan Area |
| Probe machine | : Shuttle PC running the 2.6.27-9 revision of the Linux kernel connected to UMD network |
| Landmark Distribution (Comcast) | : One landmark in Gaithersburg(GA) Two landmarks in Germantown(GE) Four landmarks in College Park(CP) Four landmarks in Greenbelt(GB) |



Figure 7.1: Our Testbed at Baltimore-Washington D.C. Metropolitan Area

For one data set, probe machine will send out 500 time-synchronous ping packets to each landmark with 200ms interval between successive packets. Thus, each data set contains 11 RTT sequences of 500 RTT values as there are 11 landmarks in our testbed. 100 such data sets were collected to analyze the two geolocation techniques.

We will conduct several tests on both geolocation techniques with the 100 data sets. After each test, there will be a city-to-city matching percentage result calculated as follows. First, we choose a geolocation technique. We pick a landmark as target and try to geolocate the target to one of the remaining landmarks with the chosen technique for each of the 100 data sets. Next, we switch to another landmark as target and try to geolocate this new target to one of the remaining landmarks with the chosen technique for each of the 100 data sets. This process is repeated until all the landmarks have been picked once as a target. The matching percentage

of City A to City B under the chosen geolocation technique is obtained as follows:

the number of matchings = the number of landmarks in City A \times 100 data sets

the number of City B matchings = the number of geolocations to some landmark in City B for targets in City A

the matching percentage of City A to City B = $\frac{\text{the number of City B matchings}}{\text{the number of matchings}} \times 100\%$

We conduct several tests on the two geolocation techniques to characterize different aspects of the techniques. We will test the performance of wavelet density estimation and wavelet time-frequency analysis with different RTT sample sizes. We want to investigate possible impacts of different RTT sample sizes to both techniques. We will also analyze the performance of wavelet time-frequency analysis for different landmark distributions to explore the effects of landmark distribution to the technique. We will conduct wavelet time-frequency analysis on additional data sets collected with different ping rates to examine possible consequences.

7.1 Wavelet Density Estimation With Different RTT Sample Sizes

We want to explore possible impacts of RTT sample size to wavelet density estimation. We will conduct wavelet density estimation with the 100 RTT data sets collected from our testbed. We will use the first 16 elements, 32 elements, 64 elements, 128 elements, 256 elements and 500 elements of the 500-element RTT sequences in the 100 RTT data sets to evaluate this geolocation technique with different RTT sample sizes. The reason of choosing power-of-two numbers for RTT sample sizes is to alleviate the effect of taking the closest integer of $\log_2(\text{sample size})$, i.e. $\lceil \log_2 n \rceil$ of Section 5.2.

We will tabulate the city-to-city matching percentages for each of the RTT sample sizes. In addition, we will describe how to obtain the mean of the minimum divergence from each city for targets in a city. Wavelet density estimation produces KL divergence metric for each pair of target(TG) and a landmark(LM), denoted as $\text{div}(\text{TG}, \text{LM})$. The mean of the minimum divergence from City A for targets in City B is $\text{mean}\{\min\{\text{div}(\text{TG}, \text{LM}), \text{LM} \in \text{City A}\}, \text{TG} \in \text{City B}, \text{all 100 data sets}\}$. Thus, the mean of the minimum divergence from City A for targets in City B is the average of the minimum divergence in City A over all targets in City B and over all 100 data sets.

City-to-City Matching Percentages of WDE with 500 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 61% | 34% | 0% | 0% |
| | Greenbelt | 39% | 65% | 3% | 0% |
| | Gaithersburg | 0% | 0% | Nil | 60% |
| | Germantown | 0% | 1% | 97% | 40% |

Note: 'Nil' is due to having only one landmark in Gaithersburg.

City-to-City Matching Percentages of WDE with 256 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 58% | 37% | 0% | 0% |
| | Greenbelt | 42% | 62% | 0% | 7% |
| | Gaithersburg | 0% | 0% | Nil | 65% |
| | Germantown | 0% | 1% | 100% | 28% |

City-to-City Matching Percentages of WDE with 128 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 61% | 35% | 0% | 0% |
| | Greenbelt | 39% | 64% | 0% | 22% |
| | Gaithersburg | 0% | 0% | Nil | 52% |
| | Germantown | 0% | 1% | 100% | 26% |

City-to-City Matching Percentages of WDE with 64 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 61% | 37% | 0% | 0% |
| | Greenbelt | 39% | 61% | 7% | 18% |
| | Gaithersburg | 0% | 1% | Nil | 52% |
| | Germantown | 0% | 1% | 93% | 30% |

City-to-City Matching Percentages of WDE with 32 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 59% | 40% | 0% | 3% |
| | Greenbelt | 40% | 58% | 3% | 7% |
| | Gaithersburg | 0% | 0% | Nil | 48% |
| | Germantown | 1% | 2% | 97% | 42% |

City-to-City Matching Percentages of WDE with 16 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 50% | 49% | 7% | 14% |
| | Greenbelt | 46% | 44% | 14% | 16% |
| | Gaithersburg | 2% | 3% | Nil | 47% |
| | Germantown | 2% | 4% | 79% | 23% |

Mean of Minimum Divergence From Each City with 500 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 2.0308 | 1.4662 | 4.8985 | 4.5682 |
| | Greenbelt | 2.7108 | 0.7806 | 3.3345 | 2.9168 |
| | Gaithersburg | 6.9545 | 5.7149 | Nil | 0.8715 |
| | Germantown | 6.2121 | 4.8086 | 0.6757 | 1.0494 |

Mean of Minimum Divergence From Each City with 256 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 2.0176 | 1.4147 | 4.7755 | 4.4720 |
| | Greenbelt | 2.6644 | 0.8090 | 3.2498 | 2.7369 |
| | Gaithersburg | 6.8259 | 5.6278 | Nil | 1.0770 |
| | Germantown | 6.0906 | 4.7044 | 0.7702 | 1.3985 |

Mean of Minimum Divergence From Each City with 128 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 1.9429 | 1.2928 | 4.2058 | 3.9847 |
| | Greenbelt | 2.5213 | 0.7534 | 2.8826 | 2.4473 |
| | Gaithersburg | 6.4415 | 5.1789 | Nil | 1.0893 |
| | Germantown | 5.7014 | 4.3260 | 0.6595 | 1.4738 |

Mean of Minimum Divergence From Each City with 64 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 1.7503 | 1.1482 | 3.5729 | 3.3175 |
| | Greenbelt | 2.3828 | 0.7045 | 2.3394 | 1.9353 |
| | Gaithersburg | 6.4370 | 4.9381 | Nil | 0.8795 |
| | Germantown | 5.3404 | 3.9296 | 0.5393 | 1.1145 |

Mean of Minimum Divergence From Each City with 32 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 1.4969 | 0.8947 | 2.7462 | 2.2319 |
| | Greenbelt | 2.0455 | 0.5447 | 1.9191 | 1.4208 |
| | Gaithersburg | 5.9523 | 4.5000 | Nil | 0.4961 |
| | Germantown | 5.1998 | 3.6614 | 0.2426 | 0.5801 |

Mean of Minimum Divergence From Each City with 16 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Min Divergence From | Collge Park | 0.7462 | 0.5118 | 1.8118 | 1.5243 |
| | Greenbelt | 0.6292 | 0.4437 | 0.9470 | 0.8085 |
| | Gaithersburg | 4.2499 | 3.3307 | Nil | 0.4801 |
| | Germantown | 3.4592 | 2.5666 | 0.2145 | 0.6603 |

Mean of Minimum Divergence of Landmarks in a City for a CP Target Versus Different Sample Sizes

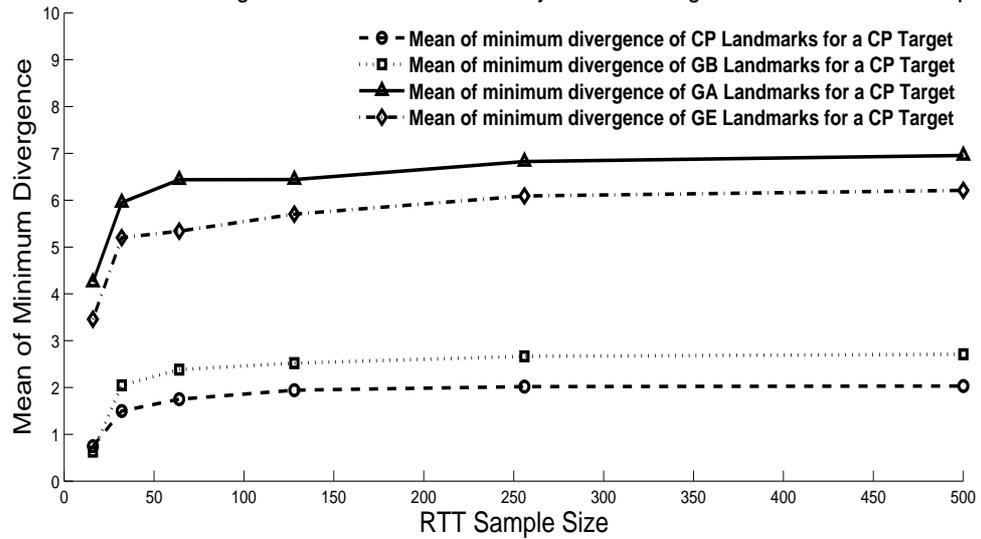


Figure 7.2: For a CP target and for all RTT sample sizes, the mean of minimum divergence from CP landmarks is the lowest followed by the mean of minimum divergence from GB landmarks. The mean of minimum divergence from GA landmark and the mean of minimum divergence from GE landmarks are significantly higher. All four lines stabilize and flatten as RTT sample size increases.

Mean of Minimum Divergence of Landmarks in a City for a GB Target Versus Different Sample Sizes

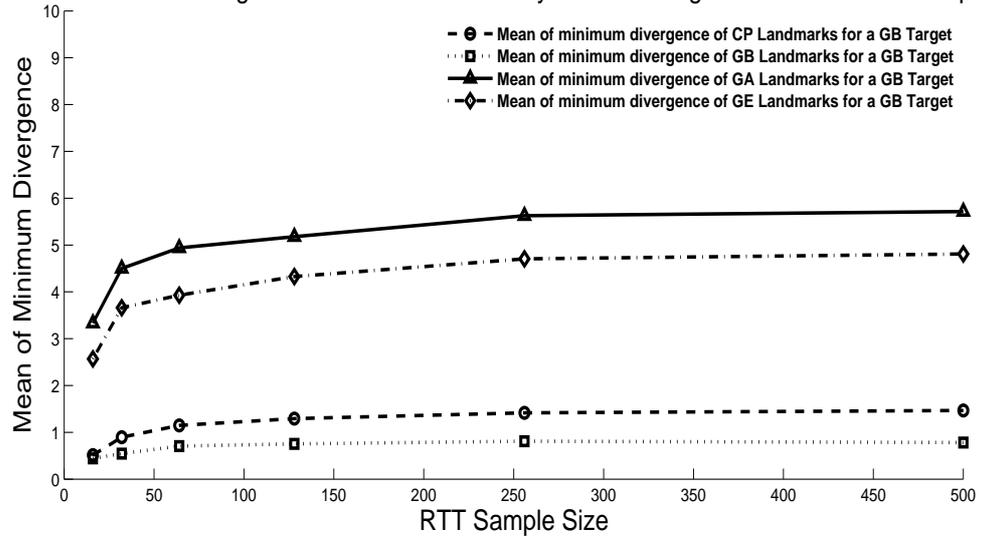


Figure 7.3: For a GB target and for all RTT sample sizes, the mean of minimum divergence from GB landmarks is the lowest followed by the mean of minimum divergence from CP landmarks. The mean of minimum divergence from GA landmark and the mean of minimum divergence from GE landmarks are significantly higher. All four lines stabilize and flatten as RTT sample size increases.

Mean of Minimum Divergence of Landmarks in a City for a GA Target Versus Different Sample Sizes

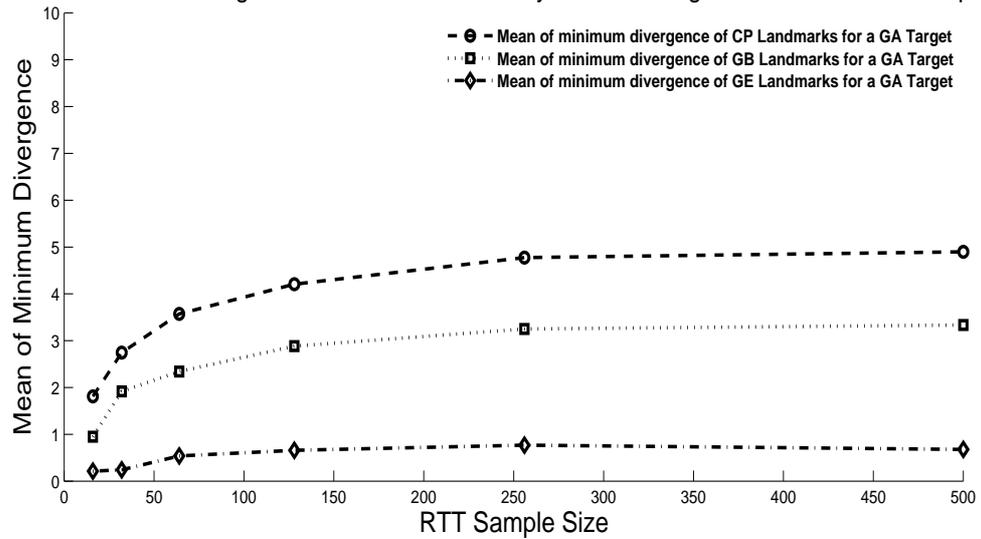


Figure 7.4: For a GA target and for all RTT sample sizes, the mean of minimum divergence from GE landmarks is the lowest. The mean of minimum divergence from CP landmarks and the mean of minimum divergence from GB landmarks are significantly higher. All three lines stabilize and flatten as RTT sample size increases.

Mean of Minimum Divergence of Landmarks in a City for a GE Target Versus Different Sample Sizes

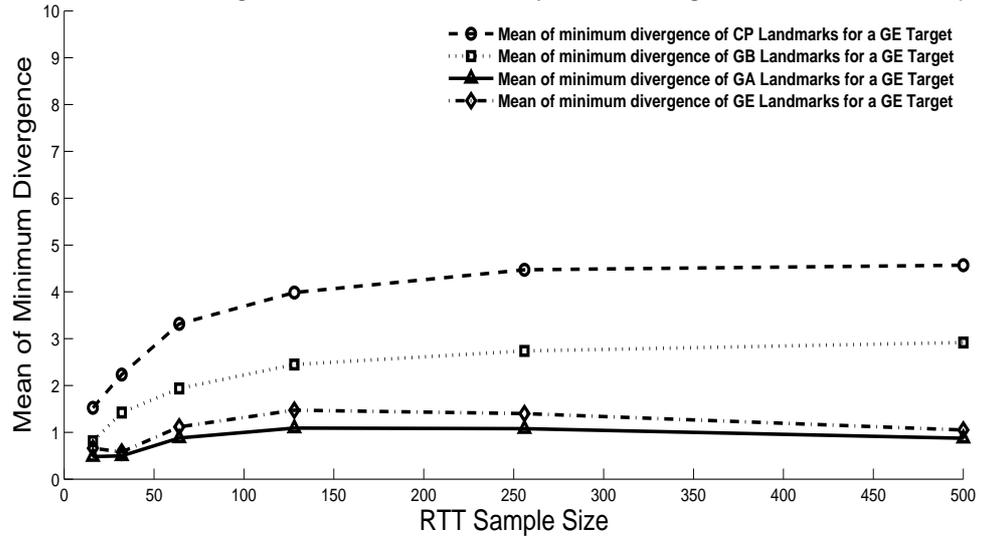


Figure 7.5: For a GE target and for all RTT sample sizes, the mean of minimum divergence from GA landmark is the lowest followed by the mean of minimum divergence from GE landmark. The mean of minimum divergence from CP landmarks and the mean of minimum divergence from GB landmarks are significantly higher. The top two lines stabilize and flatten as RTT sample size increases. The bottom two lines go down slightly with their gap closing up as RTT sample size increases.

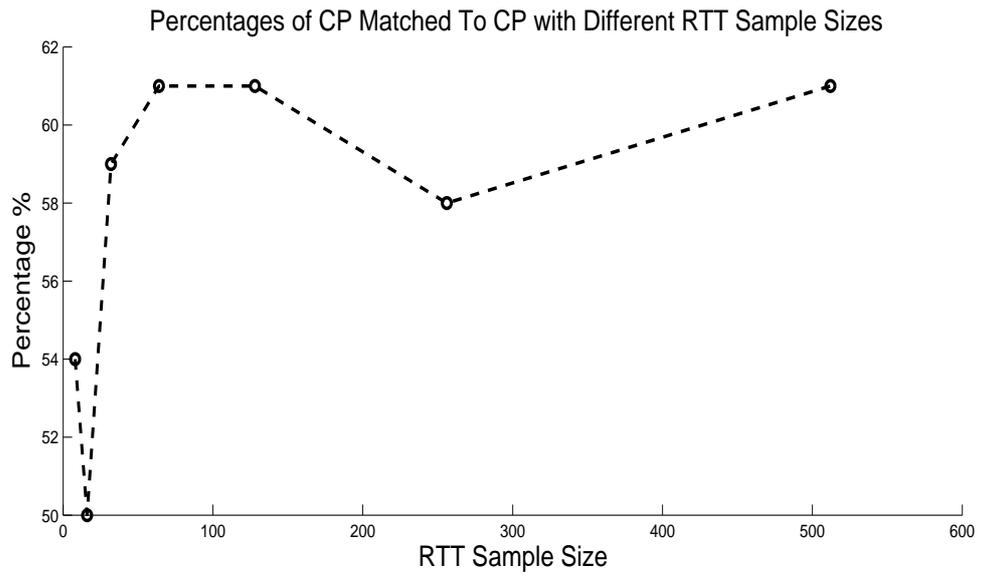


Figure 7.6: The general trend is that the matching percentage of a CP target to CP landmarks increases as RTT sample size increases.

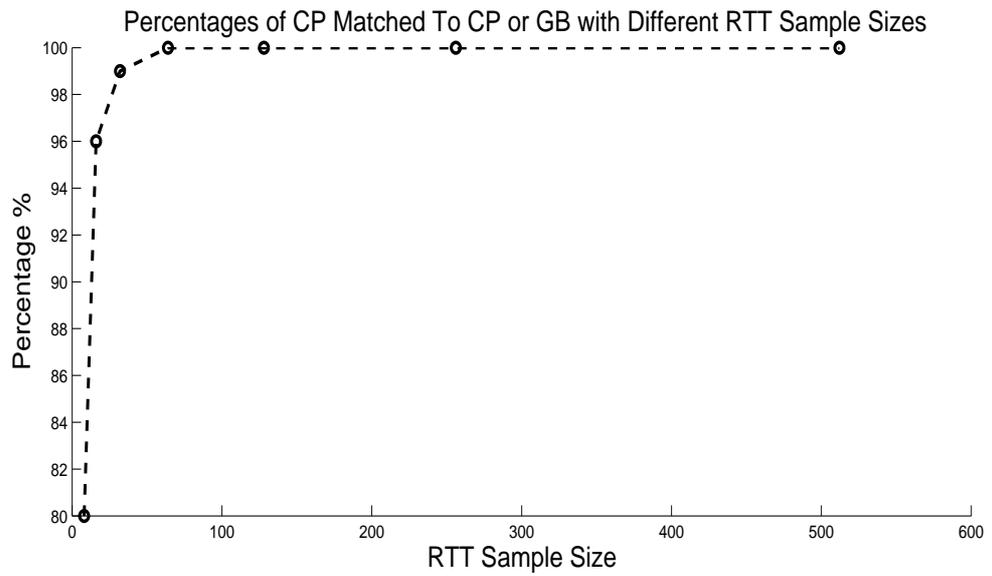


Figure 7.7: The matching percentage of a CP target to CP or GB landmarks increases as RTT sample size increases.

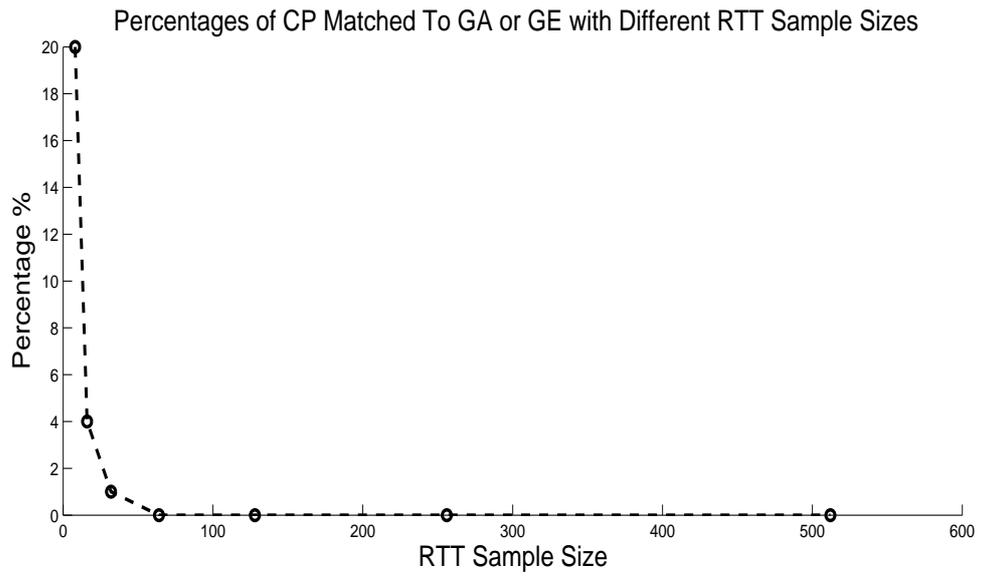


Figure 7.8: The matching percentage of a CP target to GA or GE landmarks decreases as RTT sample size increases.

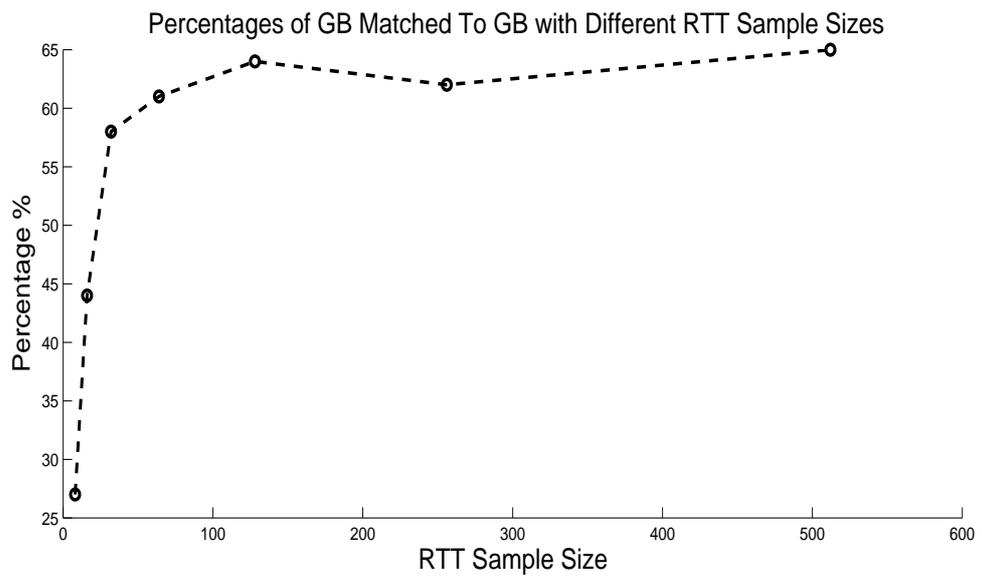


Figure 7.9: The general trend is that the matching percentage of a GB target to GB landmarks increases as RTT sample size increases.

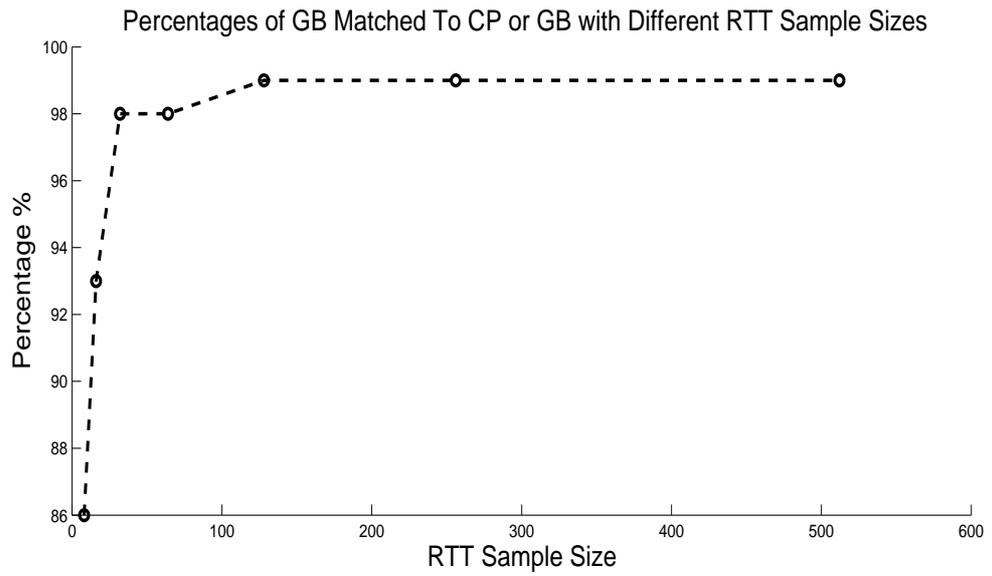


Figure 7.10: The matching percentage of a GB target to CP or GB landmarks increases as RTT sample size increases.

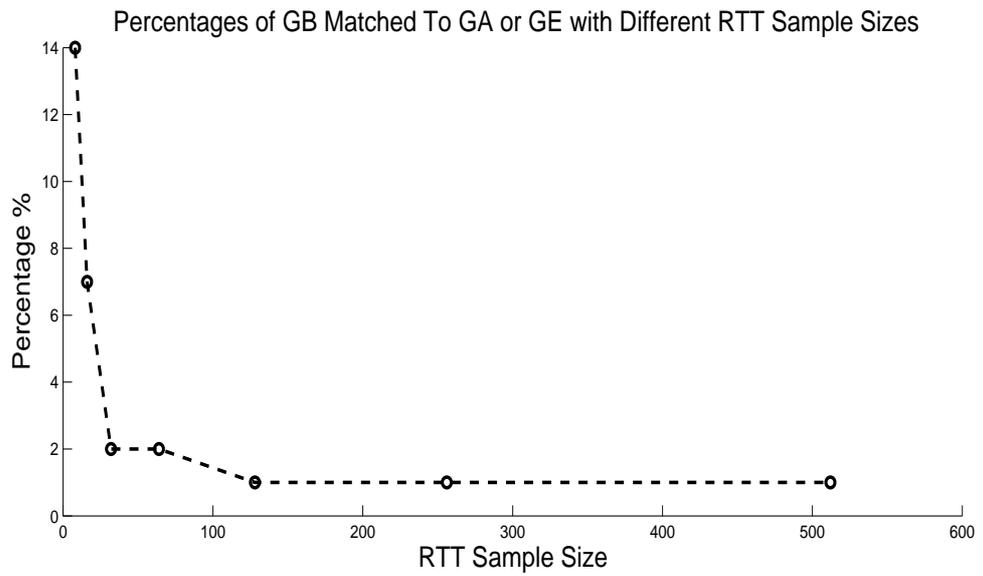


Figure 7.11: The matching percentage of a GB target to GA or GE landmarks decreases as RTT sample size increases.

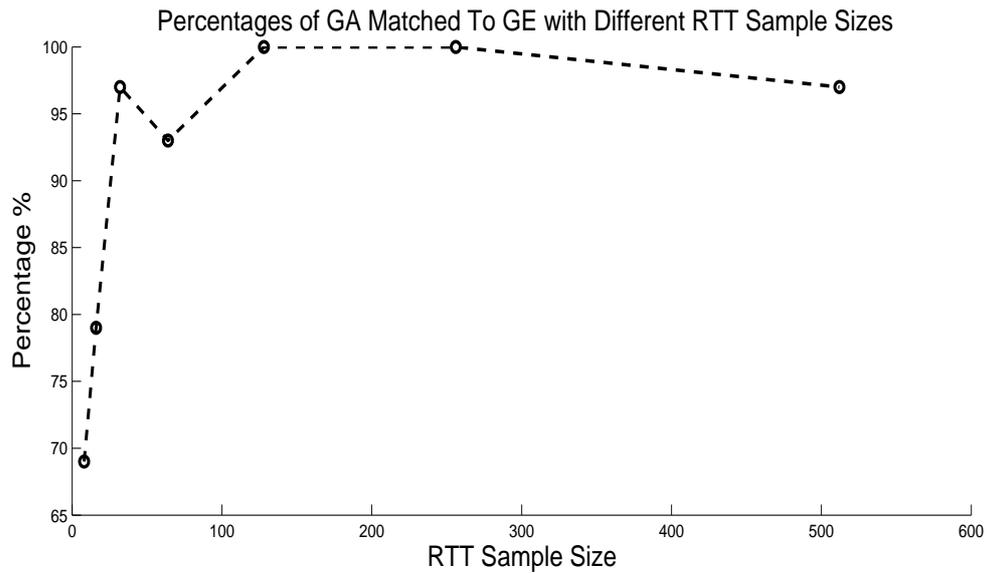


Figure 7.12: The matching percentage of a GA target to GE landmarks generally increases as RTT sample size increases from 0 to 256 samples. After 256 samples, the percentage of GA matched to GE landmarks drops.

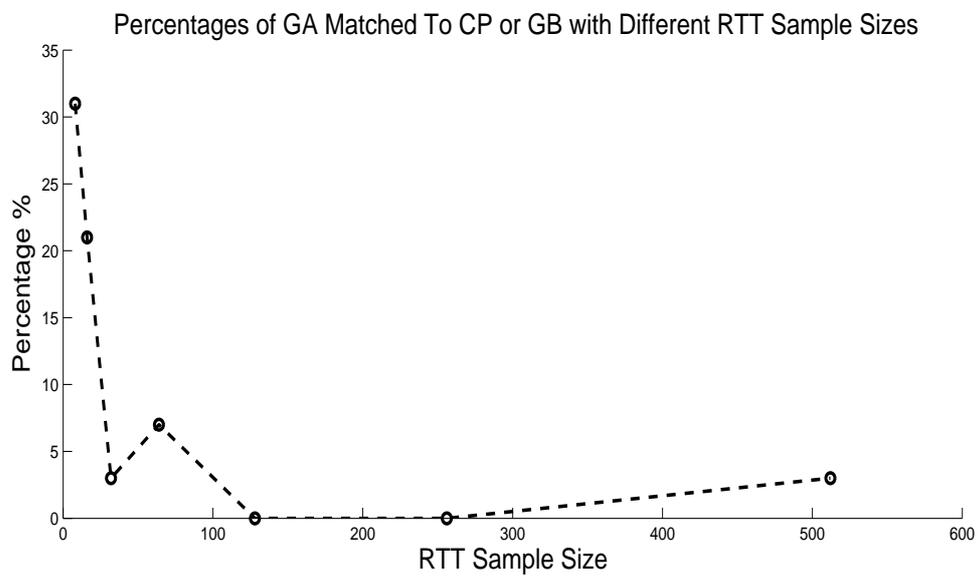


Figure 7.13: The matching percentage of a GA target to CP or GB landmarks generally decreases as RTT sample size increases from 0 to 256 samples. After 256 samples, the percentage of GA matched to CP or GB landmarks increases.

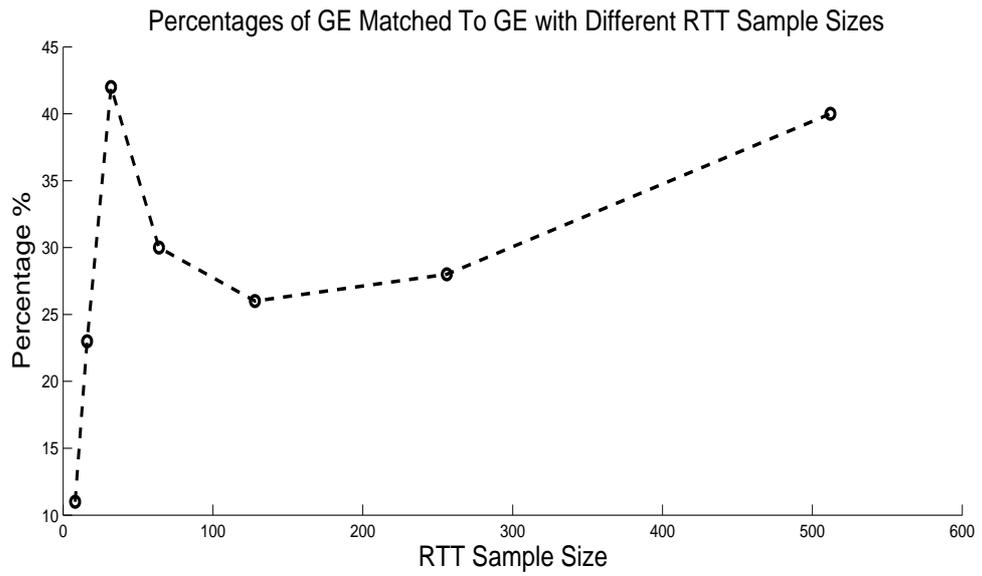


Figure 7.14: The general trend is that the matching percentage of a GE target to GE landmark increases as RTT sample size increases.

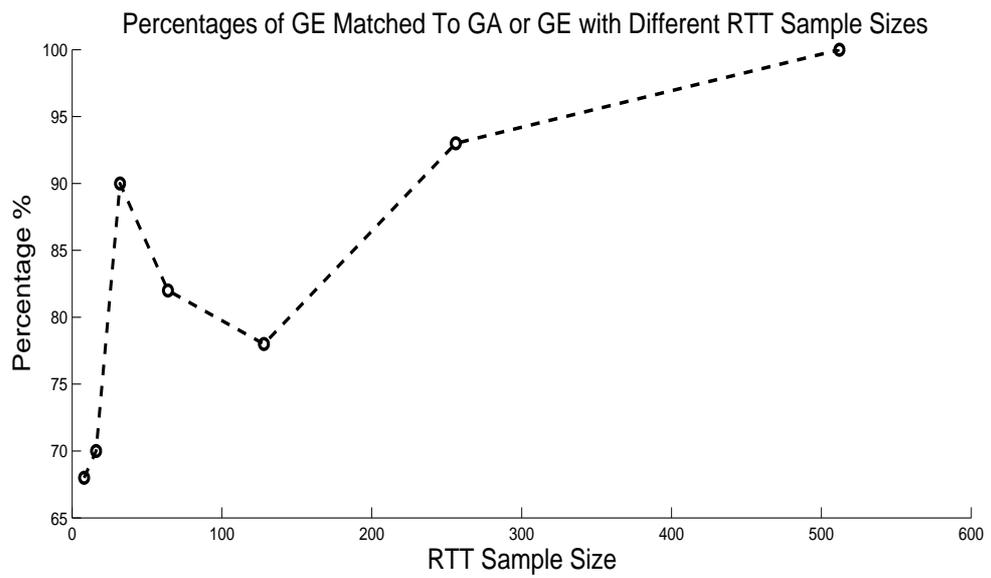


Figure 7.15: The general trend is that the matching percentage of a GE target to GA or GE landmarks increases as RTT sample size increases.

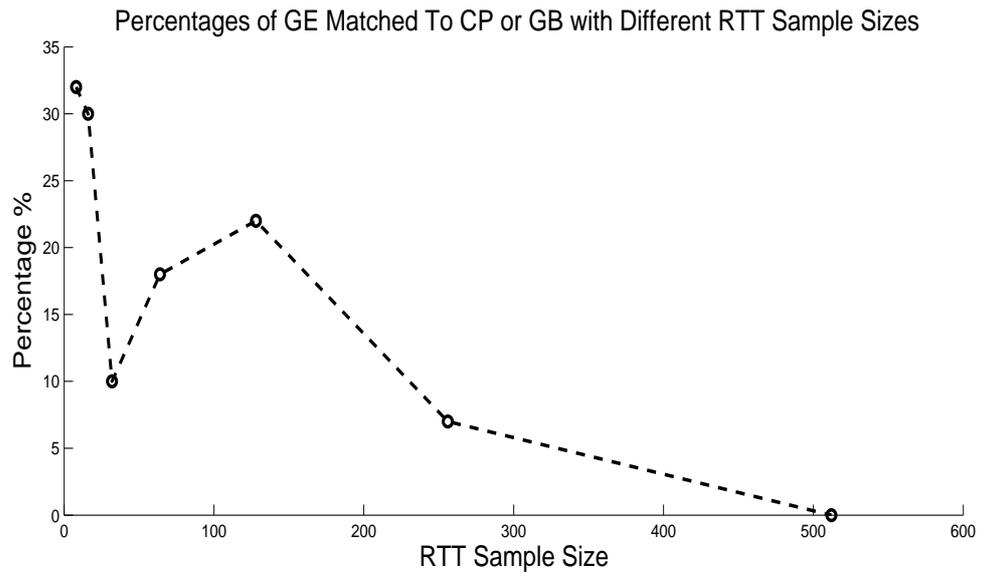


Figure 7.16: The general trend is that the matching percentage of a GE target to CP or GB landmarks decreases as RTT sample size increases.

7.2 Wavelet Time-Frequency Analysis With Different RTT Sample Sizes

As with wavelet density estimation, we will explore possible impacts of RTT sample size to wavelet time-frequency analysis. We will perform wavelet time-frequency analysis with the 100 RTT data sets. We will use the first 100 elements, 200 elements, 300 elements, 400 elements and 500 elements of the 500-element RTT sequences in the 100 RTT data sets to analyze the technique with different RTT sample sizes. We will present the city-to-city matching percentages for each of the RTT sample sizes. In place of minimum divergence, we will explain how to calculate the mean of the maximum inner product from each city for targets in a city. Wavelet time-frequency analysis produces inner product metric for each pair of target(TG) and a landmark(LM), denoted as $\text{inProd}(\text{TG}, \text{LM})$. The mean of the maximum inner product from City A for targets in City B is $\text{mean}\{\max\{\text{inProd}(\text{TG}, \text{LM}), \text{LM} \in \text{City A}\}, \text{TG} \in \text{City B}, \text{all } 100 \text{ data sets}\}$. Thus, the mean of the maximum inner product from City A for targets in City B is the average of the maximum inner product in City A over all targets in City B and over all 100 data sets.

We will present several statistics regarding the zero-one sequences for different RTT sample sizes. One of the statistics is the average number of 1's in a zero-one sequence for different RTT sample sizes. The average number of 1's that are matched up for inner products is obtained by multiplying the average number of 1's in a zero-one sequence with the average inner product. The average number of 1's matched up in correct city matchings is obtained by multiplying the average number of 1's

in a zero-one sequence with the average inner product for correct city matchings. The average number of 1's matched up in incorrect city matchings is obtained by multiplying the average number of 1's in a zero-one sequence with the average inner product for incorrect city matchings. We will also look at the difference between the average number of 1's matched up in correct city matchings and the average number of 1's matched up in incorrect city matchings. The calculations of these statistics are shown after the following table.

Mean of Maximum Inner Product From Each City with 500 RTT Sample Size

| | | Targets | | | |
|--------------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Max Inner Product From | Collge Park | 0.3605 | 0.2075 | 0.2448 | 0.2321 |
| | Greenbelt | 0.2183 | 0.2826 | 0.2726 | 0.2459 |
| | Gaithersburg | 0.1705 | 0.1689 | Nil | 0.2497 |
| | Germantown | 0.1883 | 0.1702 | 0.2787 | 0.2219 |

Various statistics of zero-one sequences with 500 RTT sample size are calculated as follows:

- The average number of 1's in a zero-one sequence = 34.1615 out of 500
- The average number of 1's matched up for inner products = $(1/15) * (0.3605 + 0.2183 + 0.1705 + 0.1883 + 0.2075 + 0.2826 + 0.1689 + 0.1702 + 0.2448 + 0.2726 + 0.2787 + 0.2321 + 0.2459 + 0.2497 + 0.2219) * 34.1615 = 7.9995$
- The average number of 1's matched up in correct city matchings = $(1/4) * (0.3605 + 0.2826 + 0.2787 + 0.2219) * 34.1615 = 9.7676$
- The average number of 1's matched up in incorrect city matchings = $(1/11) * (0.2183 + 0.1705 + 0.1883 + 0.2075 + 0.1689 + 0.1702 + 0.2448 + 0.2726 + 0.2321 + 0.2459 + 0.2497) * 34.1615 = 7.3565$

- The difference of average number of 1's matched up between correct and incorrect city matchings = $9.7676 - 7.3565 = 2.4111$

Mean of Maximum Inner Product From Each City with 400 RTT Sample Size

| | | Targets | | | |
|--------------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Max Inner Product From | Collge Park | 0.3374 | 0.1966 | 0.2897 | 0.2602 |
| | Greenbelt | 0.2008 | 0.2711 | 0.2944 | 0.2610 |
| | Gaithersburg | 0.1815 | 0.1772 | Nil | 0.2836 |
| | Germantown | 0.2012 | 0.1863 | 0.3357 | 0.2385 |

Various statistics of zero-one sequences with 400 RTT sample size are as follows:

- The average number of 1's in a zero-one sequence = 28.7550 out of 400
- The average number of 1's matched up for inner products = 7.1220
- The average number of 1's matched up in correct city matchings = 8.5021
- The average number of 1's matched up in incorrect city matchings = 6.6202
- The difference of average number of 1's matched up between correct and incorrect city matchings = 1.8819

Mean of Maximum Inner Product From Each City with 300 RTT Sample Size

| | | Targets | | | |
|--------------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Max Inner Product From | Collge Park | 0.3484 | 0.2128 | 0.3266 | 0.2940 |
| | Greenbelt | 0.2198 | 0.2884 | 0.3469 | 0.3041 |
| | Gaithersburg | 0.2129 | 0.1970 | Nil | 0.3146 |
| | Germantown | 0.2394 | 0.2128 | 0.4151 | 0.2665 |

Various statistics of zero-one sequences with 300 RTT sample size are as follows:

- The average number of 1's in a zero-one sequence = 22.7992 out of 300
- The average number of 1's matched up for inner products = 6.3827
- The average number of 1's matched up in correct city matchings = 7.5146
- The average number of 1's matched up in incorrect city matchings = 5.9711
- The difference of average number of 1's matched up between correct and incorrect city matchings = 1.5435

Mean of Maximum Inner Product From Each City with 200 RTT Sample Size

| | | Targets | | | |
|--------------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Max Inner Product From | Collge Park | 0.3673 | 0.2405 | 0.4495 | 0.3696 |
| | Greenbelt | 0.2354 | 0.3065 | 0.4250 | 0.3547 |
| | Gaithersburg | 0.2277 | 0.2229 | Nil | 0.3782 |
| | Germantown | 0.2548 | 0.2523 | 0.5243 | 0.2983 |

Various statistics of zero-one sequences with 200 RTT sample size are as follows:

- The average number of 1's in a zero-one sequence = 16.8265 out of 200
- The average number of 1's matched up for inner products = 5.5045
- The average number of 1's matched up in correct city matchings = 6.2948
- The average number of 1's matched up in incorrect city matchings = 5.2171
- The difference of average number of 1's matched up between correct and incorrect city matchings = 1.0777

Mean of Maximum Inner Product From Each City with 100 RTT Sample Size

| | | Targets | | | |
|--------------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Mean of Max Inner Product From | Collge Park | 0.3855 | 0.2970 | 0.5702 | 0.4881 |
| | Greenbelt | 0.2850 | 0.3705 | 0.5119 | 0.4664 |
| | Gaithersburg | 0.2810 | 0.2807 | Nil | 0.5368 |
| | Germantown | 0.3052 | 0.3129 | 0.6901 | 0.4612 |

Various statistics of zero-one sequences with 100 RTT sample size are as follows:

- The average number of 1's in a zero-one sequence = 9.8575 out of 100
- The average number of 1's matched up for inner products = 4.1024
- The average number of 1's matched up in correct city matchings = 4.7003
- The average number of 1's matched up in incorrect city matchings = 3.8849
- The difference of average number of 1's matched up between correct and incorrect city matchings = 0.8154

City-to-City Matching Percentages of WTA with 500 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 67% | 30% | 23% | 17% |
| | Greenbelt | 24% | 54% | 30% | 22% |
| | Gaithersburg | 5% | 9% | Nil | 47% |
| | Germantown | 4% | 7% | 47% | 14% |

City-to-City Matching Percentages of WTA with 400 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 65% | 31% | 20% | 12% |
| | Greenbelt | 25% | 55% | 17% | 20% |
| | Gaithersburg | 5% | 9% | Nil | 57% |
| | Germantown | 5% | 5% | 63% | 11% |

City-to-City Matching Percentages of WTA with 300 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 65% | 30% | 10% | 22% |
| | Greenbelt | 24% | 54% | 7% | 23% |
| | Gaithersburg | 7% | 11% | Nil | 47% |
| | Germantown | 4% | 5% | 83% | 8% |

City-to-City Matching Percentages of WTA with 200 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 66% | 36% | 20% | 23% |
| | Greenbelt | 20% | 48% | 3% | 18% |
| | Gaithersburg | 8% | 8% | Nil | 55% |
| | Germantown | 6% | 8% | 77% | 4% |

City-to-City Matching Percentages of WTA with 100 RTT Sample Size

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 66% | 40% | 24% | 31% |
| | Greenbelt | 17% | 44% | 3% | 3% |
| | Gaithersburg | 11% | 11% | Nil | 60% |
| | Germantown | 6% | 5% | 73% | 6% |

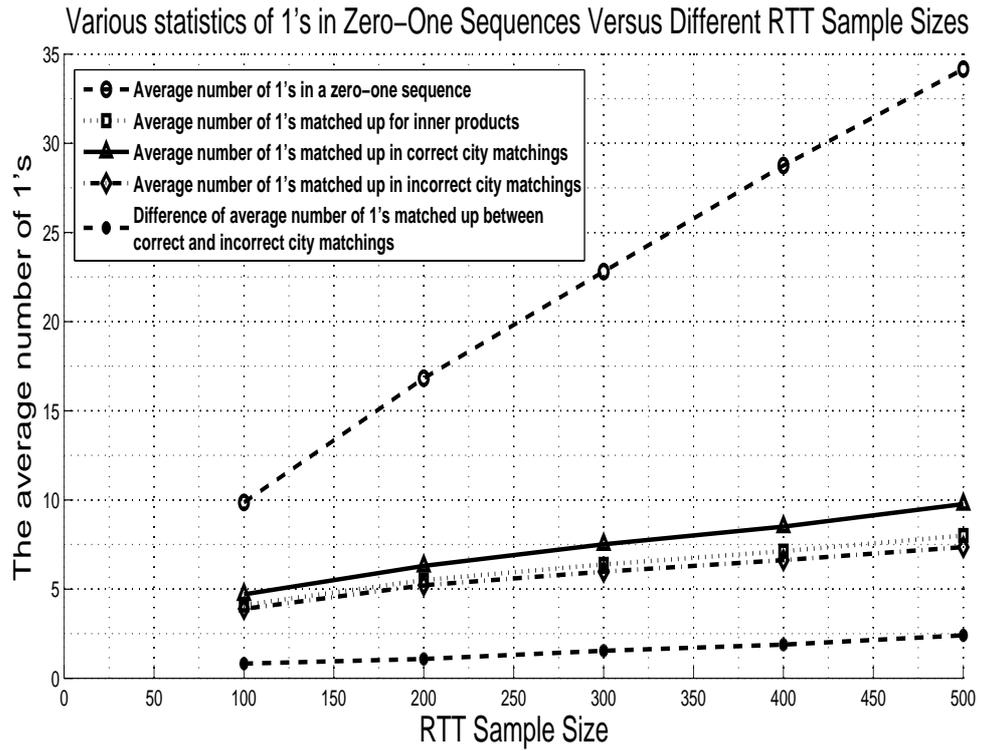


Figure 7.17: The average number of 1's in a zero-one sequence increases as RTT sample size increases. More network activities are captured as observation period of RTT values increases. The average number of 1's matched up for inner products increases as RTT sample size increases. The difference of average number of 1's matched up between correct and incorrect city matchings increases as RTT sample size increases.

Mean of Maximum Inner Product of Landmarks in a City for a CP Target Versus Different Sample Sizes

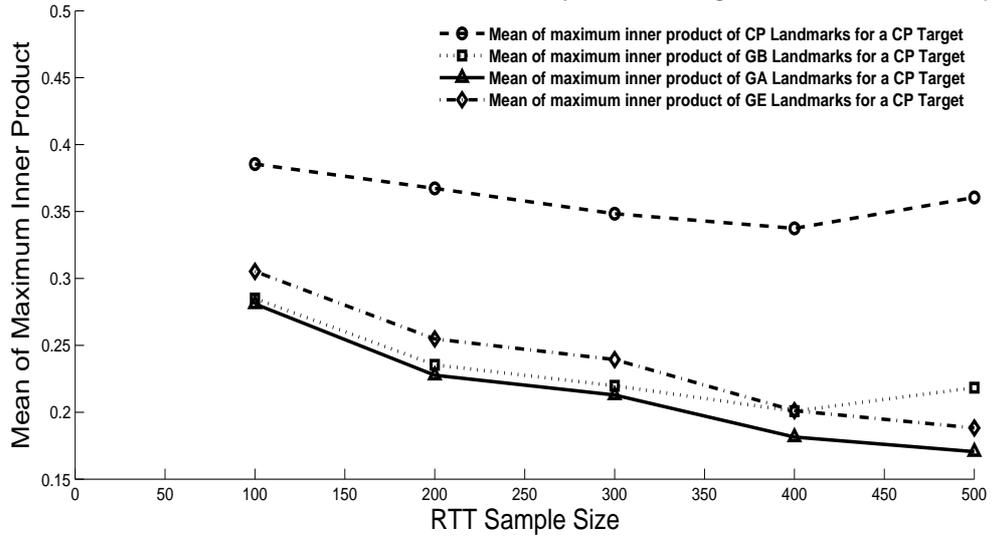


Figure 7.18: For a CP target, the mean of the maximum inner product from CP landmarks is consistently higher than those of landmarks from other cities for all RTT sample sizes.

Mean of Maximum Inner Product of Landmarks in a City for a GB Target Versus Different Sample Sizes

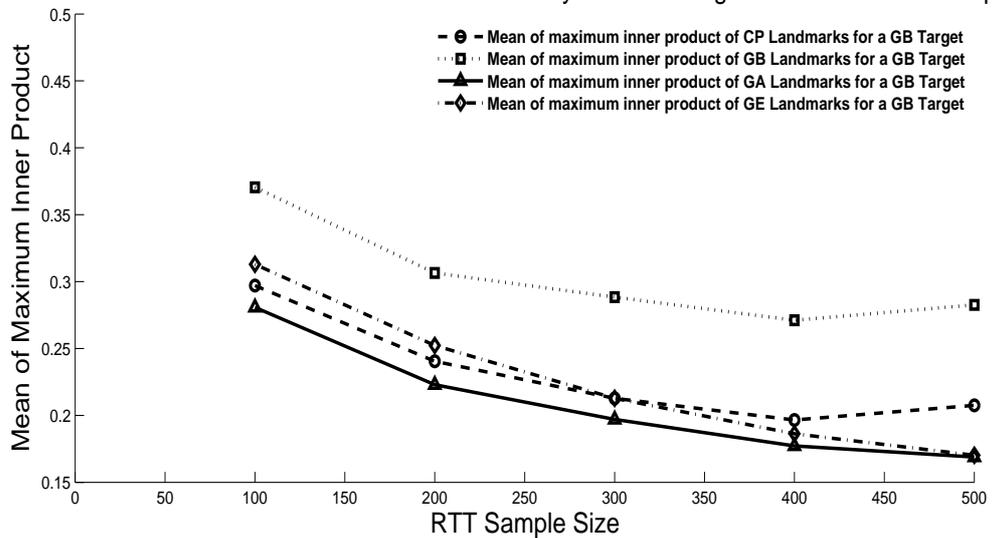


Figure 7.19: For a GB target, the mean of the maximum inner product from GB landmarks is consistently higher than those of landmarks from other cities for all RTT sample sizes.

Mean of Maximum Inner Product of Landmarks in a City for a GA Target Versus Different Sample Sizes

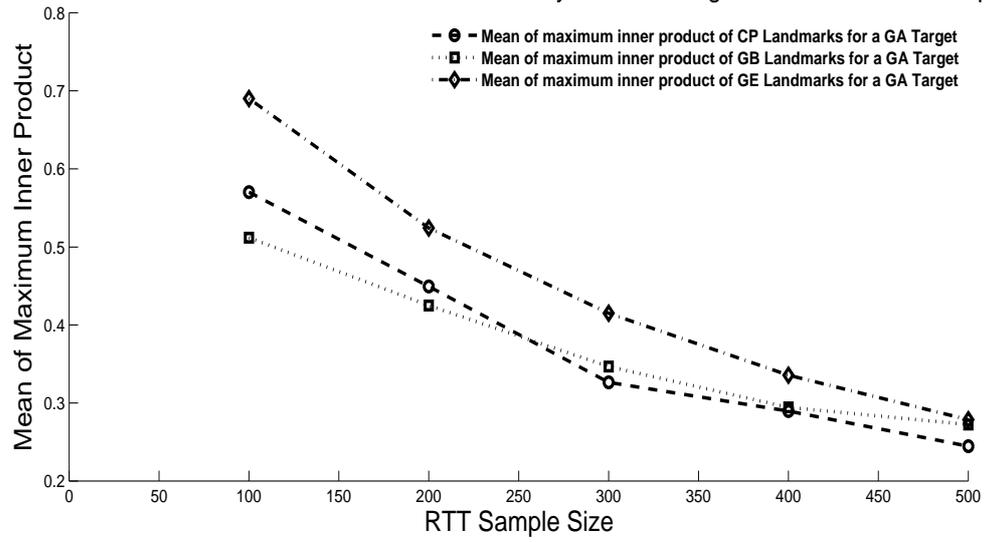


Figure 7.20: For a GA target, the mean of the maximum inner product from GE landmarks is higher than those of landmarks from other cities for all RTT sample sizes.

Mean of Maximum Inner Product of Landmarks in a City for a GE Target Versus Different Sample Sizes

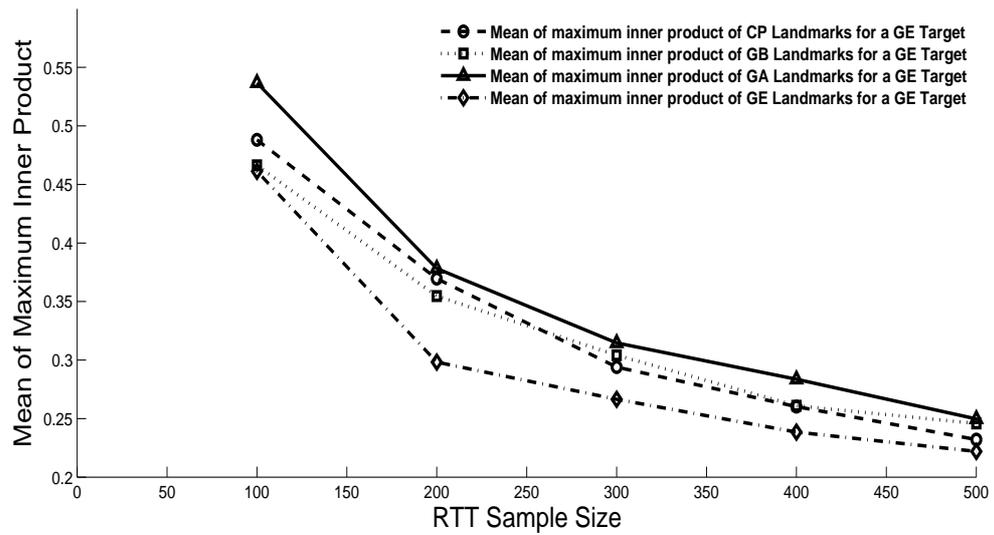


Figure 7.21: For a GE target, the mean of the maximum inner product from GE landmark is consistently lower than those of landmarks from other cities for all RTT sample sizes. However, the mean of the maximum inner product from GA landmark is consistently higher than those of landmarks from other cities.

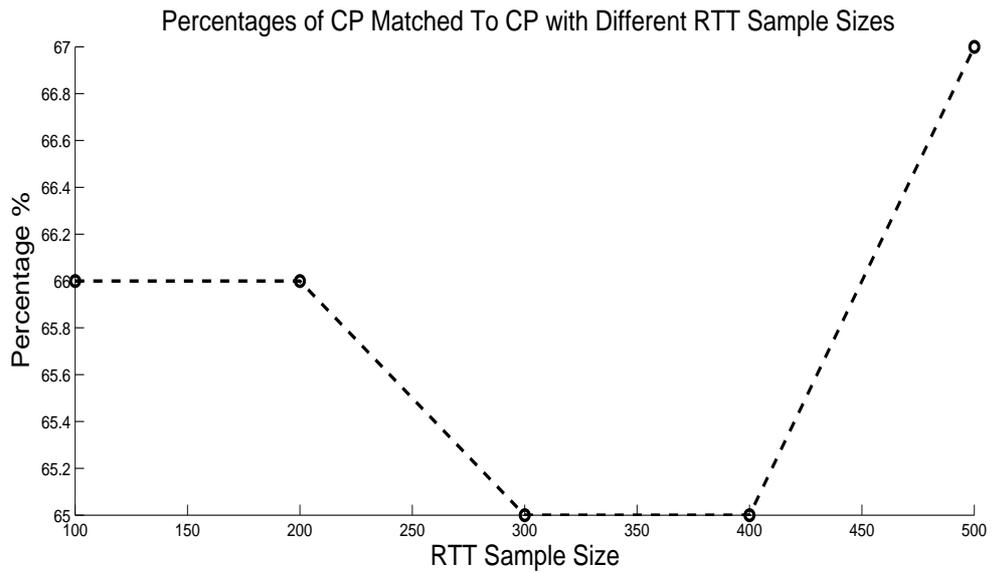


Figure 7.22: Variation of 2% is not significant.

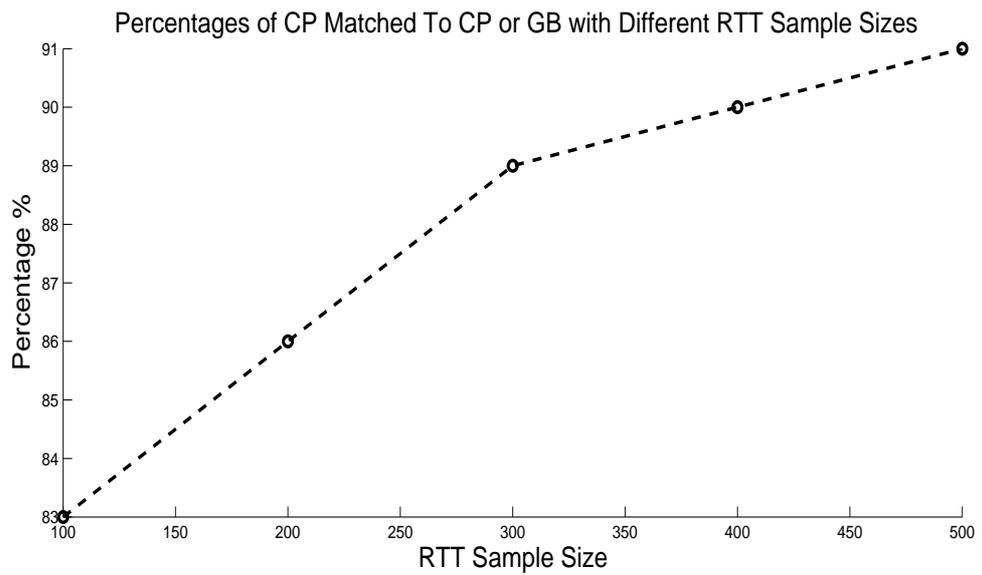


Figure 7.23: The matching percentage of a CP target to CP or GB landmarks increases as RTT sample size increases.

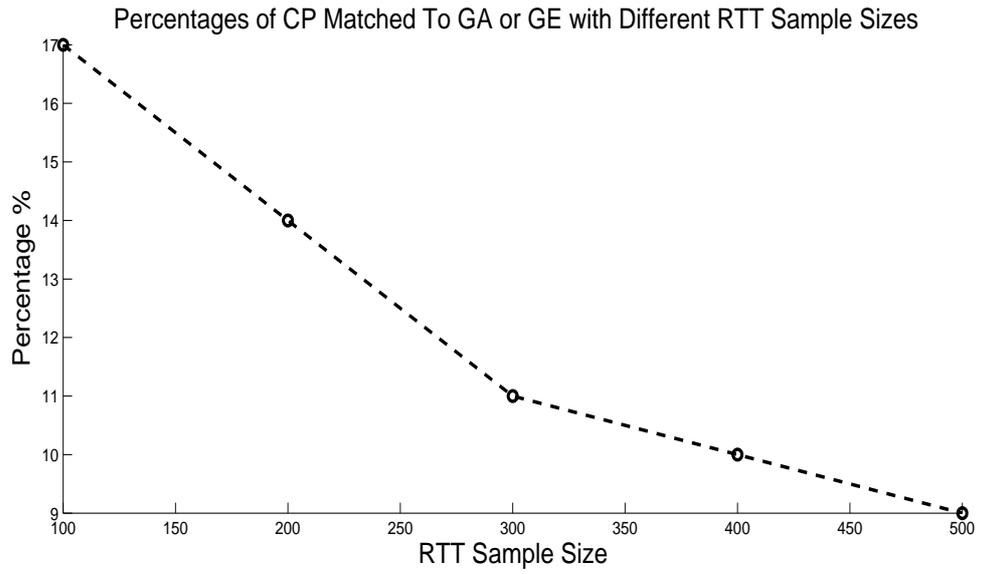


Figure 7.24: The matching percentage of a CP target to GA or GE landmarks decreases as RTT sample size increases.

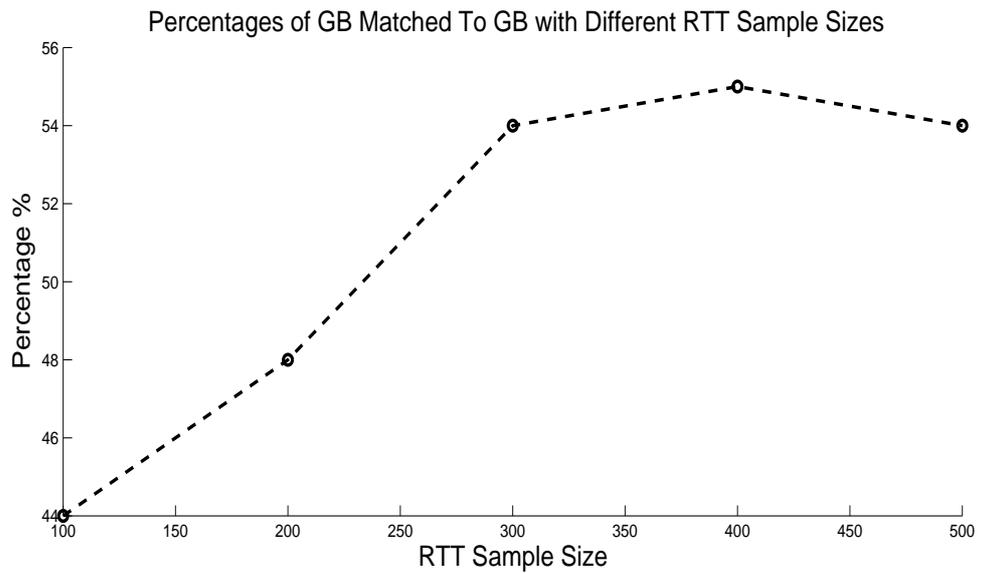


Figure 7.25: The general trend is that the matching percentage of a GB target to GB landmarks increases as RTT sample size increases.

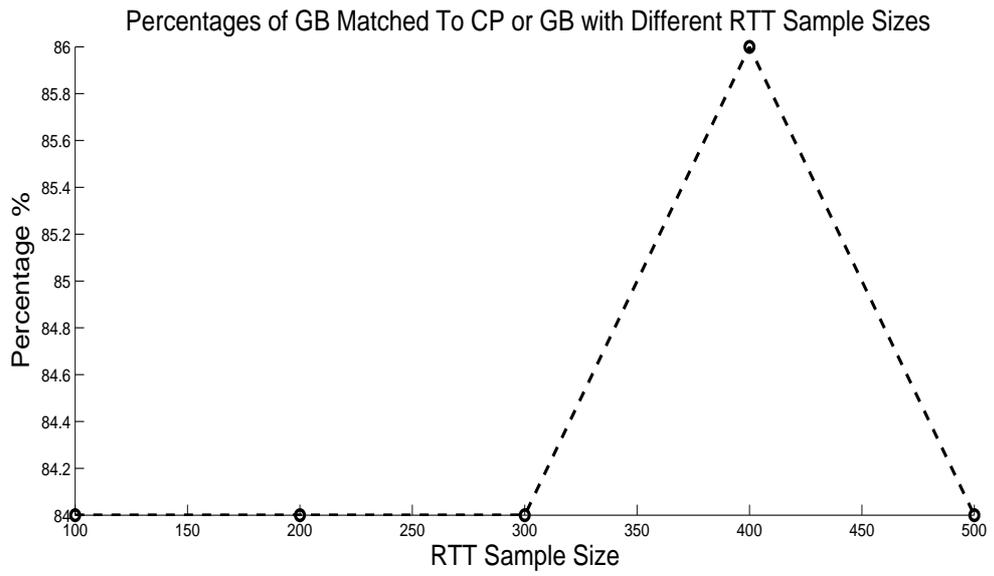


Figure 7.26: Variation of 2% is not significant.

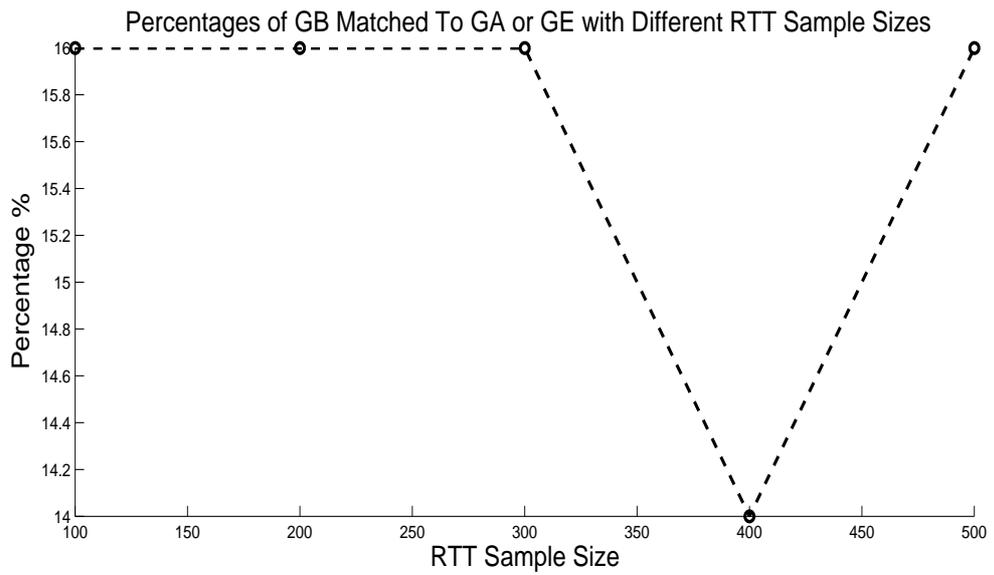


Figure 7.27: Variation of 2% is not significant.

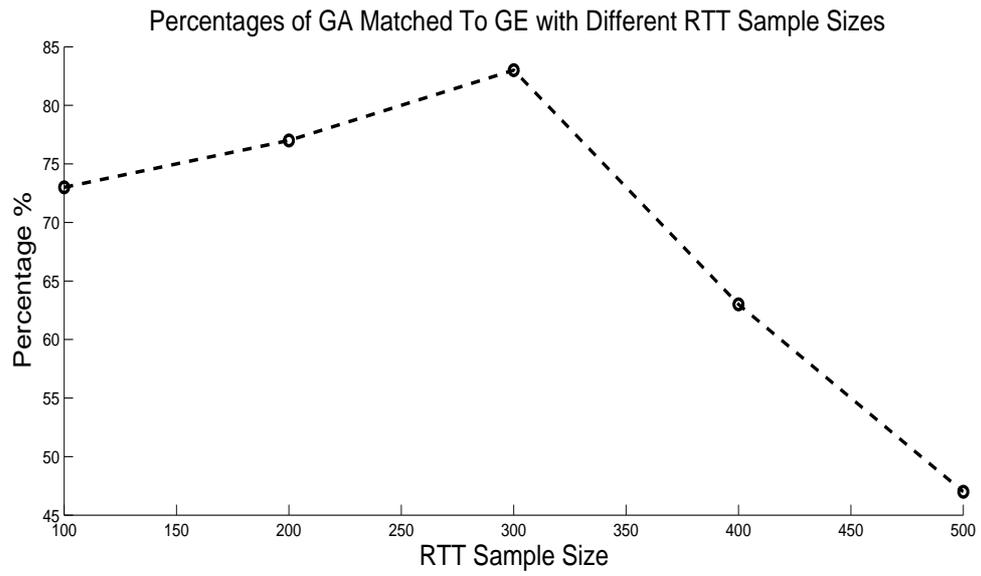


Figure 7.28: The matching percentage of a GA target to GE landmarks increases as RTT sample size increases from 0 to 300 samples. After 300 samples, the percentage of GA matched to GE landmarks drops. Wavelet density estimation registered similar observation.

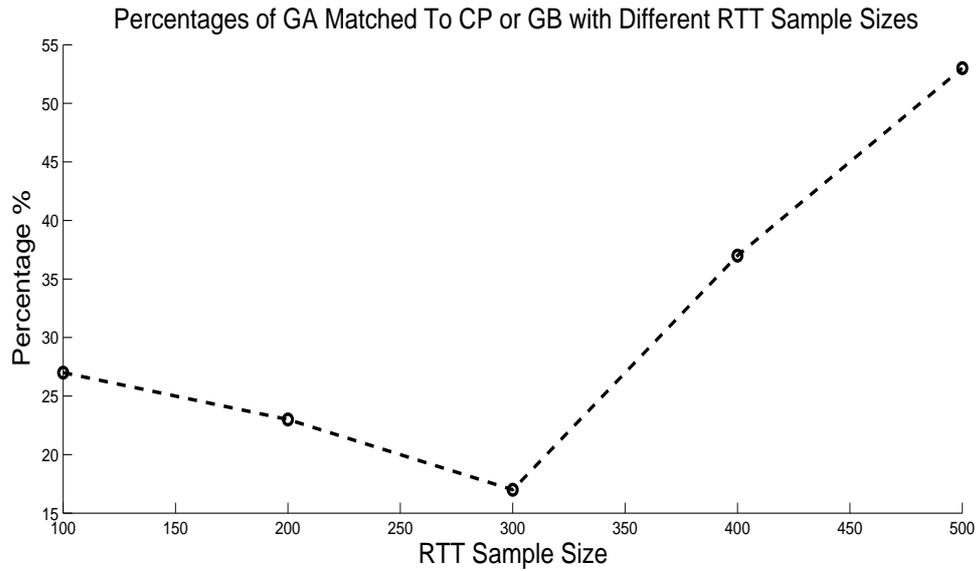


Figure 7.29: The matching percentage of a GA target to CP or GB landmarks decreases as RTT sample size increases from 0 to 300 samples. After 300 samples, the percentage of GA matched to CP or GB landmarks increases. Wavelet density estimation captured similar observation.

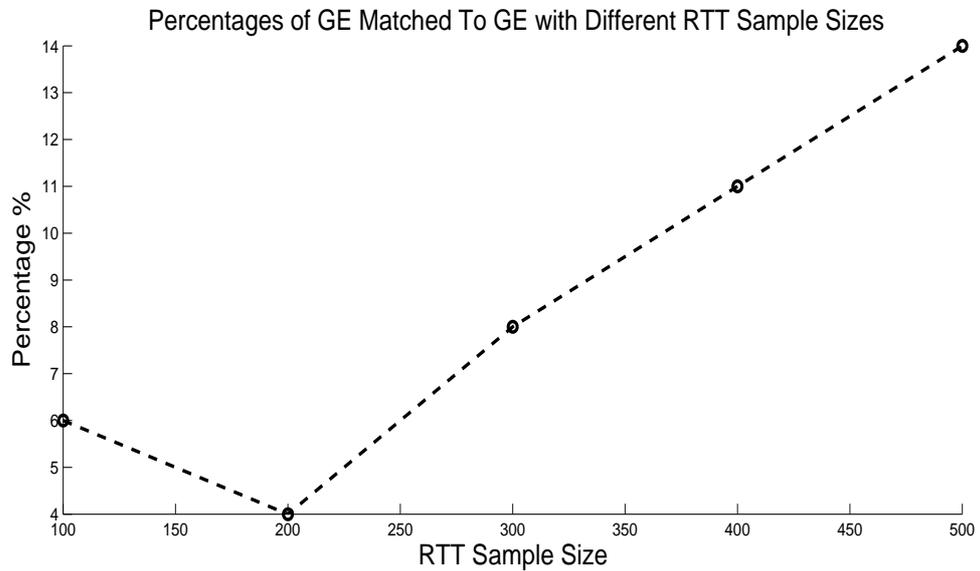


Figure 7.30: The general trend is that the matching percentage of a GE target to GE landmark increases as RTT sample size increases.

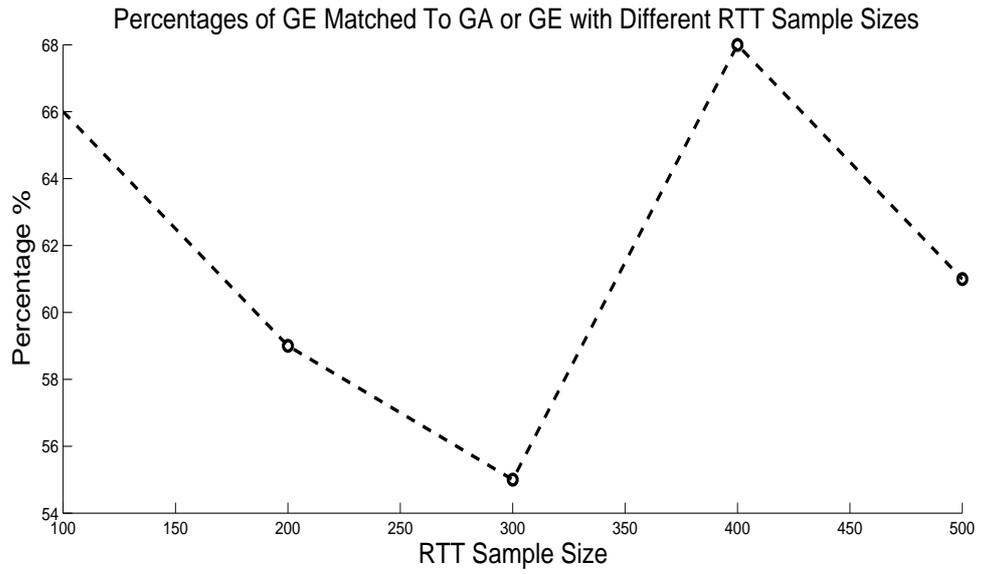


Figure 7.31: There is no definite behavior for the matching percentage of a GE target to GA or GE landmarks as RTT sample size increases.

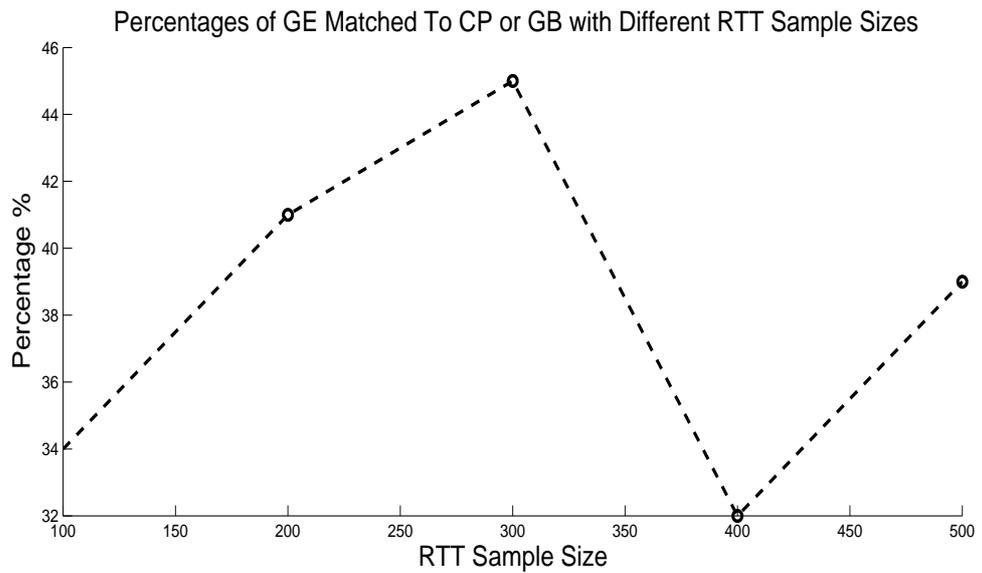


Figure 7.32: There is no definite behavior for the matching percentage of a GE target to CP or GB landmarks as RTT sample size increases.

7.3 Wavelet Time-Frequency Analysis With Different Landmark Distributions

For wavelet time-frequency analysis with 500 RTT sample size, the city-to-city matching percentages of GA to GE and GE to GA/GE are 47% and 61% respectively. These matching percentages are significantly lower than those of CP to CP/GB and GB to CP/GB, which are 91% and 84% respectively. Nonetheless, wavelet density estimation does not record the same discrepancy between the two groups of matching percentages. There are 1 GA landmark and 2 GE landmarks compared to 4 CP landmarks and 4 GB landmarks. We want to investigate if the issue of fewer GA and GE landmarks causes the low matching percentages of GA to GE and GE to GA/GE in wavelet time-frequency analysis. To do so, we will evaluate wavelet time-frequency analysis with the 100 RTT data sets collected from our testbed. First, we will analyze wavelet time-frequency analysis with the original landmark distribution, 4GBs, 4CPs, 1GA and 2GEs. We will execute wavelet time-frequency analysis with the 500-element RTT sequences of all the landmarks, 4CPs, 4GBs, 1GA and 2GEs, in the 100 RTT data sets. We then tabulate the city-to-city matching percentage result for this landmark distribution.

Next, we will analyze wavelet time-frequency analysis with a slightly reduced landmark distribution, 3CPs, 3GBs, 1GA and 2GEs. There are 4 ways of taking out one CP landmark from four CP landmarks. Similarly, there are 4 ways of taking out one GB landmark from four GB landmarks. Thus, there are $4 \times 4 = 16$ ways of obtaining a reduced landmark distribution of 3CPs, 3GBs and 1GA and 2GEs from

the original landmark distribution of 4CPs, 4GBs, 1GA and 2GEs. We will run wavelet time-frequency analysis for each of the 16 reduced landmark distributions of 3CPs, 3GBs, 1GA and 2GEs. We will put a weight factor of $\frac{1}{16}$ on the city-to-city matching percentage result of each of the 16 reduced landmark distributions. We then sum up the 16 sets of weighted city-to-city matching percentage result to produce a representative city-to-city matching percentage result for the reduced landmark distribution of 3CPs, 3GBs, 1GA and 2GEs.

We will analyze wavelet time-frequency analysis with a further reduced landmark distribution, 2CPs, 2GBs, 1GA and 2GEs. There are 6 ways of taking out two CP landmarks from four CP landmarks. Similarly, there are 6 ways of taking out two GB landmarks from four GB landmarks. Therefore, there are $6 \times 6 = 36$ ways of obtaining a reduced landmark distribution of 2CPs, 2GBs and 1GA and 2GEs from the original landmark distribution of 4CPs, 4GBs, 1GA and 2GEs. We will execute wavelet time-frequency analysis for each of the 36 reduced landmark distributions of 2CPs, 2GBs, 1GA and 2GEs. We will now put a weight factor of $\frac{1}{36}$ on the city-to-city matching percentage result of each of the 36 reduced landmark distributions. We then add up the 36 sets of weighted city-to-city matching percentage result to produce a representative city-to-city matching percentage result for the reduced landmark distribution of 2CPs, 2GBs, 1GA and 2GEs.

City-to-City Matching Percentages of WTA with 4CPs, 4GBs, 1GA and 2GEs

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 67% | 30% | 23% | 17% |
| | Greenbelt | 24% | 54% | 30% | 22% |
| | Gaithersburg | 5% | 9% | Nil | 47% |
| | Germantown | 4% | 7% | 47% | 14% |

City-to-City Matching Percentages of WTA with 3CPs, 3GBs, 1GA and 2GEs

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 50% | 33% | 16% | 13% |
| | Greenbelt | 39% | 49% | 26% | 18% |
| | Gaithersburg | 5% | 10% | Nil | 51% |
| | Germantown | 6% | 8% | 58% | 18% |

City-to-City Matching Percentages of WTA with 2CPs, 2GBs, 1GA and 2GEs

| | | Targets | | | |
|-----------------------------|--------------|--------------|-----------|--------------|------------|
| | | College Park | Greenbelt | Gaithersburg | Germantown |
| Matching Percentage To City | Collge Park | 27% | 41% | 11% | 10% |
| | Greenbelt | 59% | 38% | 21% | 14% |
| | Gaithersburg | 6% | 13% | Nil | 56% |
| | Germantown | 8% | 8% | 68% | 20% |

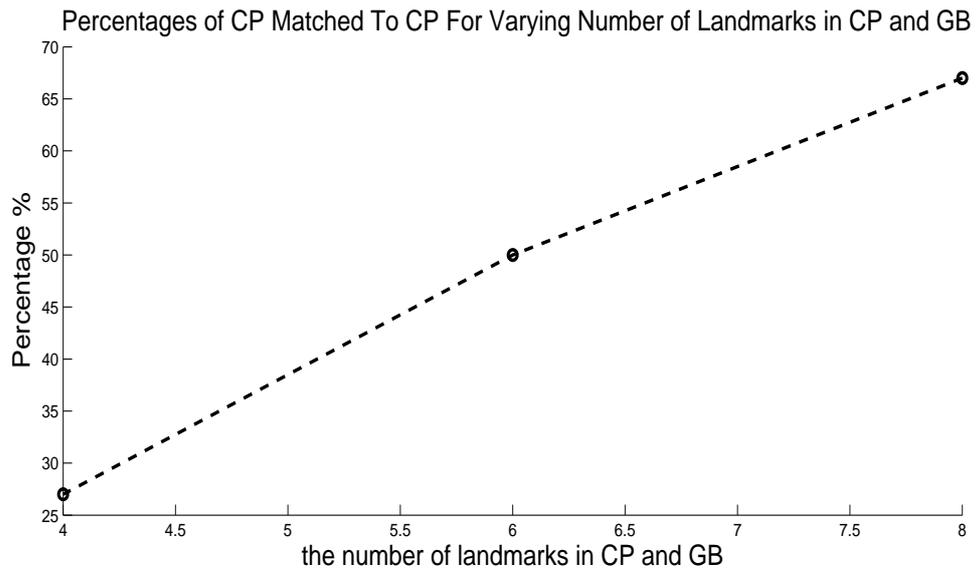


Figure 7.33: As there are more landmarks in CP, a CP target is more likely to be matched to CP city.

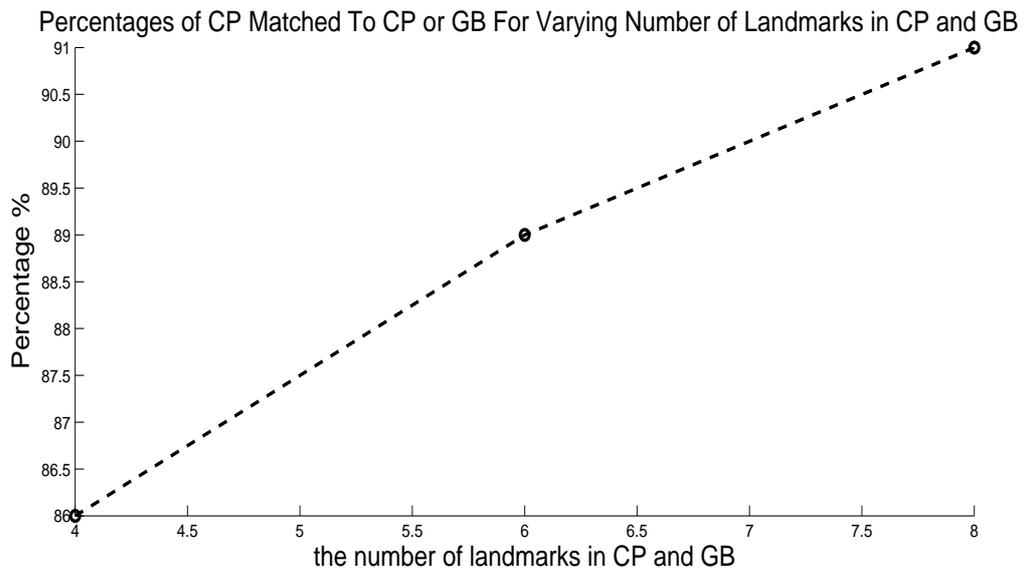


Figure 7.34: A CP target matched to CP city or GB city more often when there are more landmarks in both cities.

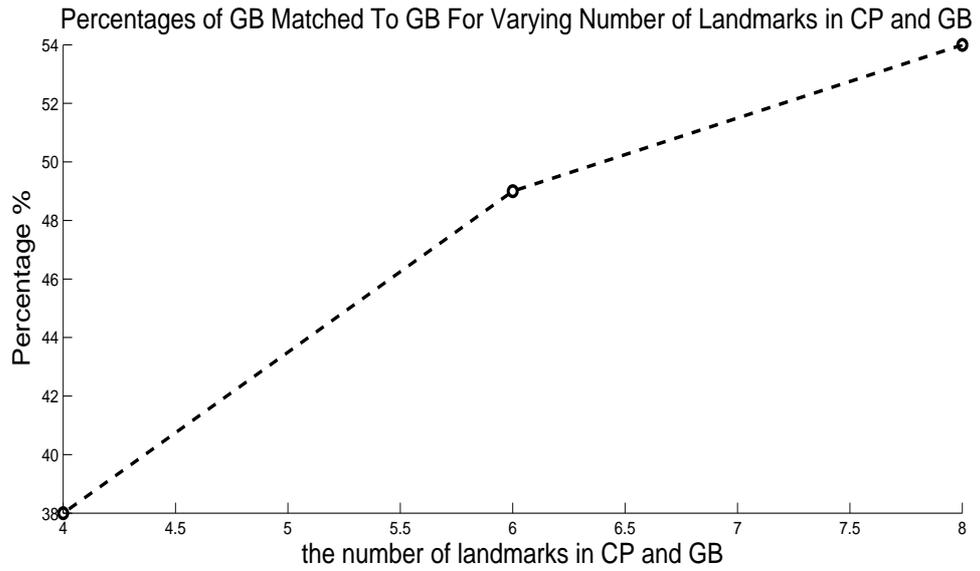


Figure 7.35: As there are more landmarks in GB, a GB target is more likely to be matched to GB city.

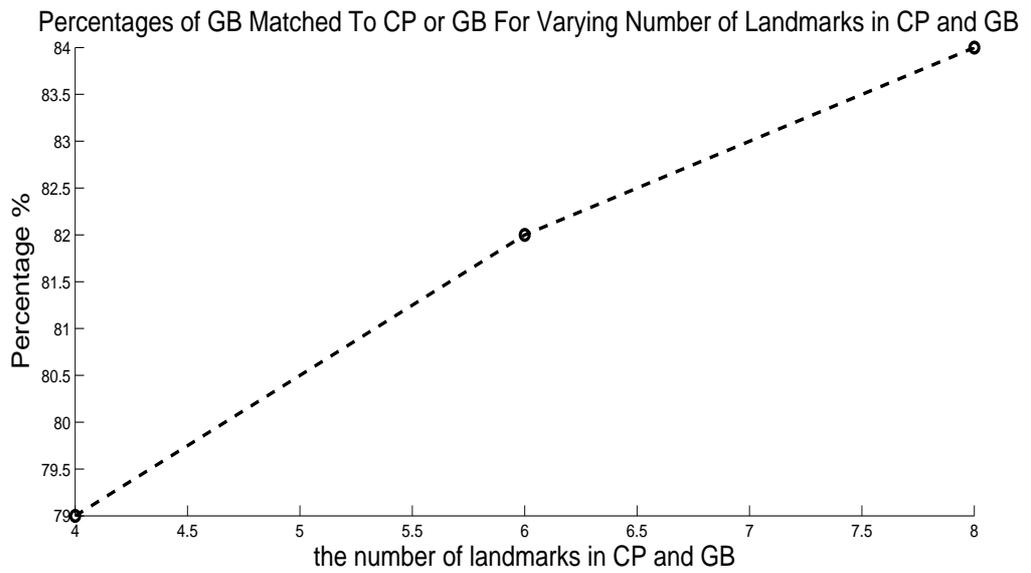


Figure 7.36: A GB target matched to CP city or GB city more often when there are more landmarks in both cities.

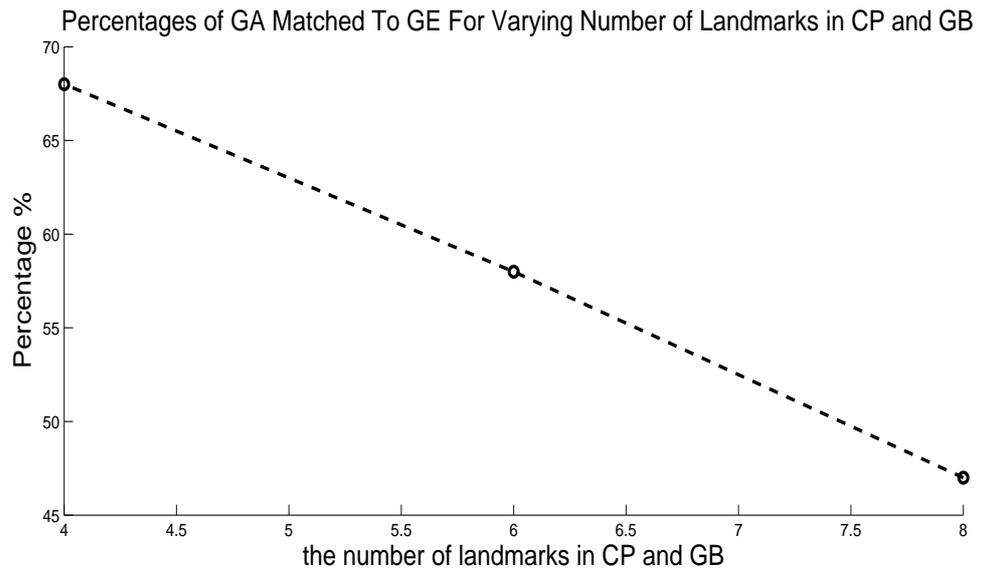


Figure 7.37: With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GA to GE increases to 68%. It is 47% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs. Thus, the issue of fewer GA and GE landmarks in landmark distribution of 4CPs, 4GBs, 1GA and 2GEs does play a role in the low matching percentage of GA to GE.

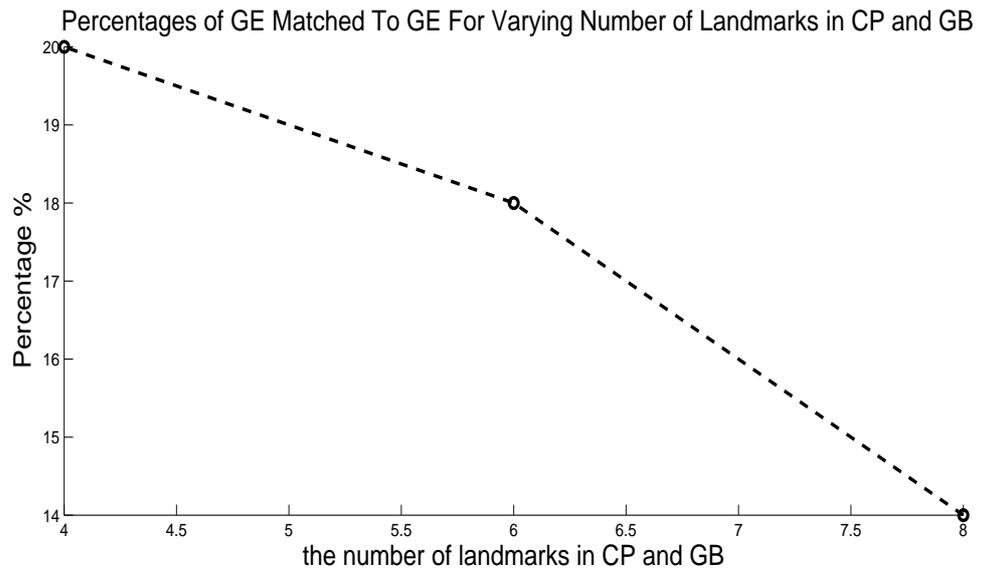


Figure 7.38: With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GE to GE increases to 20%. It is 14% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs.

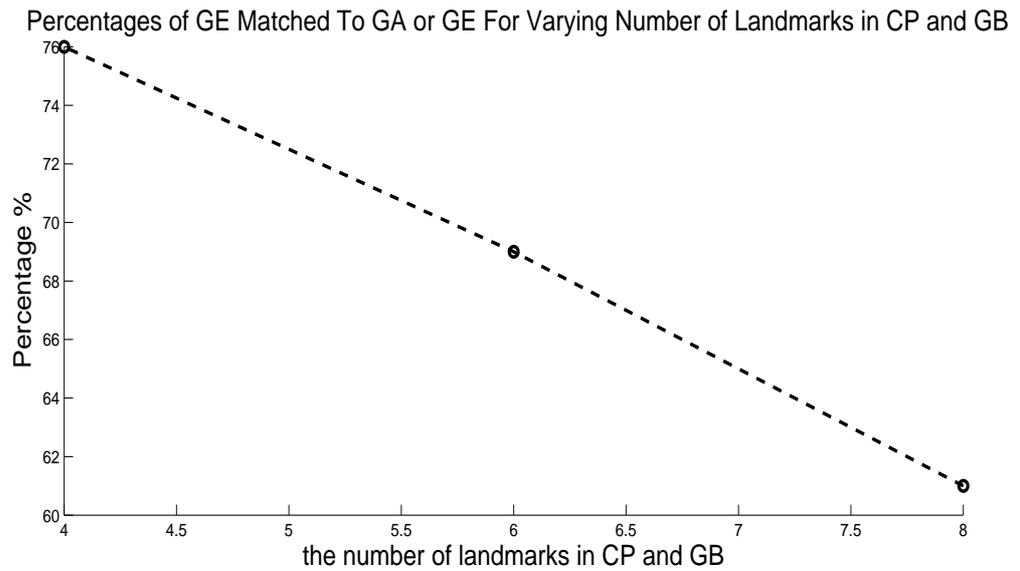


Figure 7.39: With landmark distribution of 2CPs, 2GBs, 1GA and 2GEs, the matching percentage of GE to GA/GE increases to 76%. It is 61% when landmark distribution is 4CPs, 4GBs, 1GA and 2GEs. Thus, fewer GA and GE landmarks in landmark distribution of 4CPs, 4GBs, 1GA and 2GEs contribute to the low matching percentage of GE to GA/GE.

7.4 Wavelet Time-Frequency Analysis With Different Ping Rates

Different ping rates will mean different probing rates of network dynamics. Capturing network dynamics at different rates in the form of RTT values might have possible consequences to wavelet time-frequency analysis that exploits time-synchronous network activities shared among nearby landmarks. To examine possible consequences of different ping rates to wavelet time-frequency analysis, we collected the following additional RTT data sets. Landmark distribution was 2CPs, 3GBs and 1GE. For experiment 1, probe machine sent out 500 pings with 400ms interval between successive pings to all landmarks. For experiment 2, probe machine sent out 500 pings with 200ms interval between successive pings to all landmarks. For experiment 3, probe machine sent out 500 pings with 100ms interval between successive pings to all landmarks. For experiment 4, probe machine sent out 500 pings with 50ms interval between successive pings to all landmarks. The process was repeated. For experiment 5, probe machine sent out 500 pings with 400ms interval between successive pings to all landmarks. For experiment 6, probe machine sent out 500 pings with 200ms interval between successive pings to all landmarks. The process was continued until experiment 120. We also imposed a 60-second pause between experiments.

Thus, we have 30 data sets for each of four different ping rates, e.g. 400ms, 200ms, 100ms and 50ms. We will conduct wavelet time-frequency analysis through the 30 data sets of a specific ping rate. After that, we tabulate the corresponding city-to-city matching percentage result and the various statistics of zero-one se-

quences and the mean of the maximum inner product as in wavelet time-frequency analysis with different RTT sample sizes.

City-to-City Matching Percentages of WTA with 400ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Matching Percentage To City | Collge Park | 35% | 53% | 35% |
| | Greenbelt | 55% | 35% | 65% |
| | Germantown | 10% | 12% | Nil |

Note: ‘Nil’ is due to having only one landmark in Germantown.

City-to-City Matching Percentages of WTA with 200ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Matching Percentage To City | Collge Park | 22% | 42% | 28% |
| | Greenbelt | 67% | 47% | 72% |
| | Germantown | 11% | 11% | Nil |

For 200ms ping rate, GB is matched to GB with 47%. This percentage is very close to 49% of wavelet time-frequency analysis simulated with landmark distribution of 3CPs, 3GBs, 1GA and 2GEs in previous section. Similarly, CP is matched to CP with 22%, which is also close to 27% of wavelet time-frequency analysis simulated with landmark distribution of 2CPs, 2GBs, 1GA and 2GEs.

City-to-City Matching Percentages of WTA with 100ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Matching Percentage To City | Collge Park | 22% | 48% | 21% |
| | Greenbelt | 71% | 33% | 79% |
| | Germantown | 7% | 19% | Nil |

City-to-City Matching Percentages of WTA with 50ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Matching Percentage To City | Collge Park | 14% | 48% | 33% |
| | Greenbelt | 66% | 42% | 67% |
| | Germantown | 20% | 10% | Nil |

Mean of Maximum Inner Product From
Each City with 400ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Mean of Max Inner Prod From | Collge Park | 0.1556 | 0.2704 | 0.3582 |
| | Greenbelt | 0.2730 | 0.2793 | 0.4662 |
| | Germantown | 0.2314 | 0.2618 | Nil |

Various statistics of zero-one sequences with 400ms ping rate are as follows:

- The average number of 1's in a zero-one sequence = 22.2324 out of 500
- The average number of 1's matched up for inner products = 6.3804
- The average number of 1's matched up in correct city matchings = 4.8344
- The average number of 1's matched up in incorrect city matchings = 6.8957
- The difference of average number of 1's matched up between correct and incorrect city matchings = -2.0613

Mean of Maximum Inner Product From
Each City with 200ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Mean of Max Inner Prod From | Collge Park | 0.1214 | 0.1720 | 0.2439 |
| | Greenbelt | 0.2048 | 0.2057 | 0.3379 |
| | Germantown | 0.1760 | 0.1968 | Nil |

Various statistics of zero-one sequences with 200ms ping rate are as follows:

- The average number of 1's in a zero-one sequence = 25.2428 out of 500
- The average number of 1's matched up for inner products = 5.2331
- The average number of 1's matched up in correct city matchings = 4.1285
- The average number of 1's matched up in incorrect city matchings = 5.6014
- The difference of average number of 1's matched up between correct and incorrect city matchings = -1.4729

Mean of Maximum Inner Product From
Each City with 100ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Mean of Max Inner Prod From | Collge Park | 0.1073 | 0.2030 | 0.2312 |
| | Greenbelt | 0.1804 | 0.2137 | 0.3217 |
| | Germantown | 0.1507 | 0.2107 | Nil |

Various statistics of zero-one sequences with 100ms ping rate are as follows:

- The average number of 1's in a zero-one sequence = 26.8372 out of 500
- The average number of 1's matched up for inner products = 5.4302
- The average number of 1's matched up in correct city matchings = 4.3074
- The average number of 1's matched up in incorrect city matchings = 5.8044
- The difference of average number of 1's matched up between correct and incorrect city matchings = -1.4971

Mean of Maximum Inner Product From

Each City with 50ms Ping Rate

| | | Targets | | |
|-----------------------------|-------------|--------------|-----------|------------|
| | | College Park | Greenbelt | Germantown |
| Mean of Max Inner Prod From | Collge Park | 0.0878 | 0.1522 | 0.2182 |
| | Greenbelt | 0.1652 | 0.1722 | 0.2928 |
| | Germantown | 0.1369 | 0.1645 | Nil |

Various statistics of zero-one sequences with 50ms ping rate are as follows:

- The average number of 1's in a zero-one sequence = 26.7232 out of 500
- The average number of 1's matched up for inner products = 4.6425
- The average number of 1's matched up in correct city matchings = 3.4740
- The average number of 1's matched up in incorrect city matchings = 5.0320
- The difference of average number of 1's matched up between correct and incorrect city matchings = -1.5580

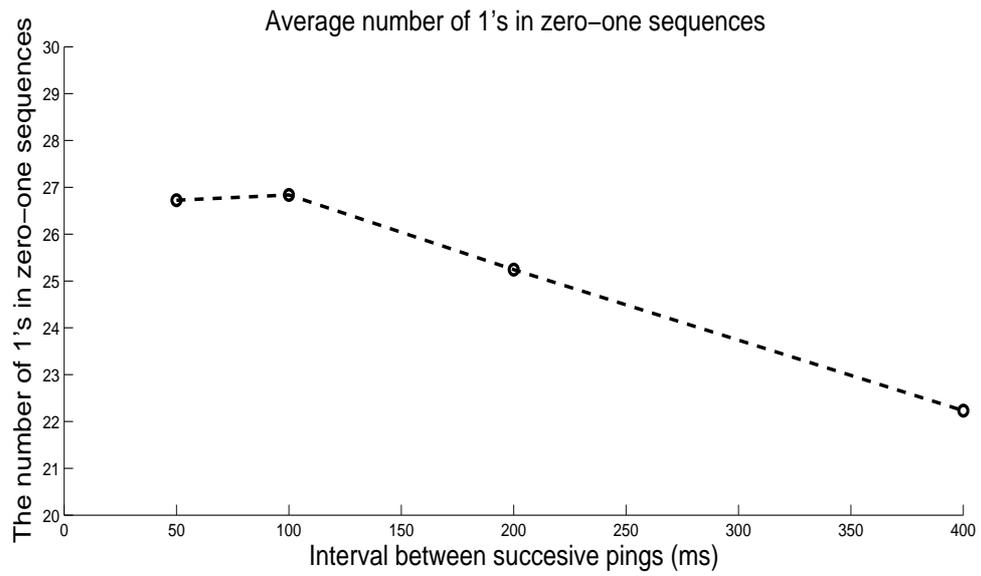


Figure 7.40: As ping rate increases with shorter inter-packet interval, more dynamics of the network are captured as shown by the increasing average number of 1's in a 500-element zero-one sequence.

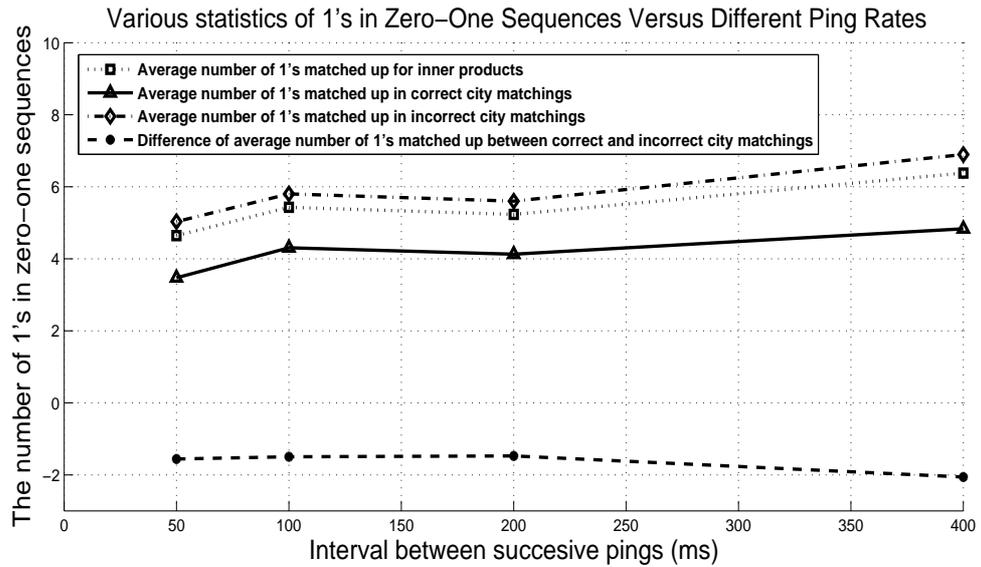


Figure 7.41: As ping rate increases with shorter inter-packet interval, the average number of 1's matched up in correct city matchings and the average number of 1's matched up in incorrect city matchings both decrease. Thus, as ping rate increases, more network dynamics are captured with more 1's in a zero-one sequence but most of these dynamics captured are not network activities experienced by nearby landmarks as suggested by the decreasing average number of 1's matched up in correct city matchings.

Chapter 8

Conclusions

We have attempted to geolocate an Internet host in a metropolitan area network (MAN) within city-level granularity. We explored two wavelet-based techniques to perform MAN IP geolocation, i.e. wavelet density estimation and wavelet time-frequency analysis.

For wavelet density estimation, we used the fact that wavelets are orthonormal basis in $L^2(R)$ to construct the estimated pdf of a RTT distribution. We devised a symmetrized version of Kullback-Leibler divergence to measure the similarity of two estimated pdfs. Thus, geolocation of a target is based on the similarity of its RTT distribution to some landmark's RTT distribution. We evaluated this technique with data sets collected from our testbed. As RTT sample size increases, the matching percentages to the correct cities are more likely to increase. There are good separations in the means of the minimum divergence from different cities.

For wavelet time-frequency analysis, we utilized the time-frequency localization property of wavelets to analyze how the frequency content of a RTT sequence changes with time. Therefore, geolocation of target is based on the temporal frequency content. We geolocate a target to a landmark that displays similar pattern in its frequency content changes over time. From our evaluation of this technique with data sets collected, we found that wavelet time-frequency analysis generally per-

forms better with larger RTT sample size. This geolocation technique can achieve better matching percentages to correct cities when there are sufficient landmarks in those cities. Wavelet time-frequency analysis reveals that increased RTT probing rates tend to capture more network dynamics that are not network activities experienced by nearby landmarks.

In summary, MAN IP geolocation can be conducted through wavelet density estimation and wavelet time-frequency analysis with their performance influenced by RTT sample size and landmark distribution.

Bibliography

- [1] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, Vol. 35, pp. 246-254, 1948.
- [2] H. Anton. *Calculus: A New Horizon (Sixth Edition)*. John Wiley & Sons, Inc, 1999.
- [3] R. G. Bartle and D. R. Sherbert. *Introduction to Real Analysis (Third Edition)*. John Wiley & Sons, Inc, 2000.
- [4] A. Boggess and F. J. Narcowich. *A First Course in Wavelets with Fourier Analysis*. Prentice Hall, 2001.
- [5] C. J. Bovy, H. T. Mertodimedjo, G. Hooghiemstra, H. Uijterwaal, and P. van Mieghem, "Analysis of end-to-end delay measurements in Internet," in *Proc. Passive and Active Measurement Workshop*, Fort Collins, CO, Mar. 2002.
- [6] R. Christensen. *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. Chapman and Hall/CRC, 1996.
- [7] C. K. Chui. *An Introduction to Wavelets*. Academic Press, 1992.
- [8] D. L. Donoho, "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," in *Proc. of Symposia in Applied Mathematics*, Vol. 47, pp. 173-205, 1993.
- [9] D. L. Donoho and I. M. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of the American Statistical Association*, Vol. 90, No. 432, pp. 1200-1224, December 1995.
- [10] B. Gueye, A. Ziviani, M. Crovella and S. Fdida, "Constraint-Based Geolocation of Internet Hosts," *IEEE/ACM Transactions on Networking*, Vol. 14, No. 6, pp. 1219-1232, December 2006.
- [11] T. Ogden. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, 1997.
- [12] V. N. Padmanabhan and L. Subramanian, "An Investigation of Geographic Mapping Techniques for Internet Hosts," in *Proc. ACM SIGCOMM*, San Diego, CA, Aug. 2001, pp. 173-185.

- [13] B. M. Sadler and A. Swami, "Analysis of Multiscale Products for Step Detection and Estimation," IEEE Transactions on Information Theory, Vol. 45, No. 3, pp. 1043-1051, April 1999.
- [14] C. M. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," The Annals of Statistics, Vol. 9, No. 6, pp. 1135-1151, 1981.