RC 21527 (97109) 20JUL99 Computer Science/Mathematics

Research Report

On Improvement of Independent Wavelet Models to Heterogeneous Network Traffic

X. Tian
Rensselaer Polytechnic Institute
110, 8 th Street
Troy, NY 12180

S. Ma
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

C. Ji
Rensselaer Polytechnic Institute
110, 8 th Street
Troy, NY 12180

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties). Copies may be requested from IBM T. J. Watson Research Center [Publications 16-220 ykt] P. O. Box 218, Yorktown Heights, NY 10598. email: reports@us.ibm.com
Some reports are available on the internet at http://domino.watson.ibm.com/library/CyberDig.nsf/home

On Improvement of Independent Wavelet Models to Heterogeneous Network Traffic

Xusheng Tian, Sheng Ma and Chuanyi Ji

Abstract

In our previous work, we showed empirically that independent (Haar) wavelet models were parsimonious, computationally efficient and accurate in modeling heterogeneous network traffic measured by both auto-covariance functions and buffer loss rate. We also proved analytically that such models were capable of capturing any decay rate of auto-covariance functions at large lags. But the simplicity of independent Haar wavelet models also results in deviations from the actual auto-covariance functions in small lags which should be improved. In this work, we focus on improving independent (Haar) wavelet models using two approaches. One still uses independent wavelet coefficients but more complex wavelet basis with higher vanishing moments. The other still uses Haar wavelet basis but incorporates major dependence among wavelet coefficients using (low-order) Markov models across time scales. Both approaches are measurement-based with parameters that can be readily estimated using traces of heterogeneous traffic. The performance in approximating auto-covariance functions at small lags is improved by both approaches at a moderate computational cost, which is $O(N)$ for a trace of length N.

I. Introduction

Modeling and analysis of heterogeneous network traffic is imperative to supporting multi-media applications with diverse statistical characteristics and Quality of Service (QoS) requirements. Traffic models should be accurate in capturing the pertinent statistics of real traffic, and in the mean time be parsimonious. The former requirement ensures that a traffic model can faithfully represent real traffic so that the resulting analysis or simulation based on the model can be reliable and convincing. The latter requirement ensures that the traffic can be described by few parameters to provide insights on key characteristics of traffic, and to be used for such applications as admission control.

Due to the heterogeneity of network traffic such as complex mixture of long-range and short-range dependence, it is difficult sometimes to use traditional models such as Markov models or FARIMA models [1] [2], since either they may not be able to capture complex dependence across time scales [1], or they may have very high computational complexity such as FARIMA [3]. Therefore, it is necessary to develop modeling approaches which can build up a model from heterogeneous network measurements, and be computationally efficient. In our previous work [3] [4] [5], we developed the independent wavelet models for modeling Gaussian traffic. We further showed that the independent wavelet models are capable of modeling a wide range of traffic both theoretically and empirically. In [6], we further extended our work to non-Gaussian traffic using traffic measurements. Related work done in applying wavelets to networks traffic modeling includes a multifractal wavelet model [7], and a wavelet-based estimator of the Hurst parameter [8].

To evaluate the capability and performance of independent wavelet models, we have shown that independent wavelet models are capable of capturing any decay rate of auto-correlation of the traffic asymptotically. Although it is more desirable to enable the wavelet model to capture the correlations at all lags, auto-correlations at small lags are not very well modeled in our previous work. Herein, we address this issue. As auto-covariance functions alone are used to assess the performance of wavelet models in this work, we only consider Gaussian traffic in this paper. Consequently, the second-order statistics, i.e. auto-correlation function, is sufficient to measure the accuracy of models.

Average auto-correlation function has been defined and developed for the independent *Haar* wavelet models in [5]. In this work, we will further study how to improve the performance of independent wavelet models measured

The work of X. Tian and C. Ji was supported by NSF grant NCR 9805338, and in part by NSF grant CAREER IRI-9502518. Part of this work was done while X.Tian was on summer co-op at IBM T.J. Watson Research Center.

X.Tian and C. Ji are with Department of Electrical, Computer and System Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, e-mail: {tianx,chuanyi}@ecse.rpi.edu.

S. Ma is with the IBM T.J. Watson Research Center, Hawthorne, NY 10532, e-mail: shengma@us.ibm.com.

by the auto-correlation function in two ways: (1) still using the independent wavelet models but considering wavelet functions with higher order vanishing moments; and (2) using *Haar* wavelets but considering dependent wavelet models. The first approach enhances the performance through using more complex wavelets. The latter approach improves the performance through considering certain correlations of wavelet coefficients across time scales.

The rest of paper is organized as follows. Section II provides background about wavelets and the independent wavelet model. In Section III, results on average auto-covariance function are presented, followed with evidence of deviation from auto-covariance function at small lags. Section IV exploring other orthogonal wavelets to improve the independent wavelet model. Section V studies dependent wavelet models. We conclude the paper in Section VI.

II. Wavelets and the Independent Wavelet Models

Let $x(t)$ be a random process generated from an independent wavelet model for discrete time t ($t \geq 0$). That is, through the inverse wavelet transform, we have

$$x(t) = \sum_{j=1}^{\infty} \sum_{m=0}^{\infty} d_j^m \psi_j^m(t) + \psi_0, \tag{1}$$

where $\psi_j^m(t)$ and d_j^m are the wavelet basis functions[1] (see [3] [9] for details) and the corresponding wavelet coefficients, respectively, at the time scale j ($j \geq 1$, and an integer) and shift m ($m \geq 0$, and an integer). The wavelet basis functions are obtained by dilating and translating a wavelet function $\psi(t)$, where

$$\psi_j^m(t) = 2^{-j/2} \cdot \psi(2^{-j} \cdot t - m),$$

and in particular for *Haar* wavelet,

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq \mathrm{t} < 1/2, \\ -1 & \text{if } 1/2 \leq \mathrm{t} < 1, \\ 0 & \text{otherwise}. \end{cases}$$

For the shape of the *Haar* wavelet, please refer to the left side of Fig. 2.

When choosing the wavelet function $\psi(t)$, one is free to select one of its important characteristics, namely, the number M of vanishing moments. M is an integer such that

$$\int t^k \psi(t) dt \equiv 0, \ \forall k = 0, 1, \ldots, M-1.$$

It is not difficult to verify that *Haar* wavelet function has 1 vanishing moment.

One also has the freedom to choose orthogonal, bi-orthogonal, or other non-orthogonal wavelets. Orthogonal wavelets were considered in our work because their forward and inverse wavelet transforms have computational complexity of $O(N)$ using filter bank algorithm. We will discuss the performance of independent wavelet models with different orthogonal wavelet functions later.

The wavelet coefficients, d_j^m's, can be obtained through the wavelet transform

$$d_j^m = \sum_{t=0}^{+\infty} x(t) \psi_j^m(t). \tag{2}$$

[1] ψ_0 represents the mean of $x(t)$, for the *Haar* wavelet. In general, ψ_0 is the projection of $x(t)$ to the coarsest time scale. Without loss of generosity, $\psi_0 = 0$ is assumed in the rest of the paper.

In the framework of independent wavelet models, the wavelet coefficients, d_j^m's, are assumed to be independent random variables. For a given j ($j \geq 1$), d_j^m's are independent and identically distributed (i.i.d.) random variables with a zero mean and a variance σ_j^2. That is,

$$\mathbf{E}[d_{j_1}^{m_1} d_{j_2}^{m_2}] = \begin{cases} \sigma_j^2 & m_1 = m_2 \,\text{and}\, \mathrm{j}_1 = \mathrm{j}_2 = \mathrm{j}, \\ 0 & \text{otherwise}. \end{cases} \tag{3}$$

To facilitate the discussion of the various wavelet domain models, we use the following acronyms,
IWM The independent wavelet models in general.
IWM-i The independent wavelet model with wavelet functions having i vanishing moments.
MWM-i The i-th order Markov wavelet model.

III. Auto-covariance of the IWM

A. Average auto-covariance of the IWM

Auto-covariance function is the criterion we use in this work to evaluate the performance of the IWM.

Although an auto-covariance[2] function is well defined for a wide-sense-stationary process, the existing definition does not apply directly to that of the IWM. This is because the time signal $x(t)$ generated from the IWM given in (1) is non-stationary [10]. To overcome the difficulty, an average auto-covariance function was defined in [5]. The average auto-covariance is in agreement to the time-averaged spectrum defined by Wornell [11]. In fact, the so-defined average auto-covariance function has a rather simple relation with the parameters of the IWM as summarized in the following theorem.

Theorem 1: For IWM's, the average auto-covariance function $\bar{R}(\tau)$ can be obtained by

$$\bar{R}(\tau) = \sum_{j=1}^{+\infty} \sigma_j^2 \, 2^{-j} \, h_j(\tau), \tag{4}$$

where,

$$h_j(\tau) = \sum_{t=0}^{T_j - 1} \psi_j^0(t) \psi_j^0(t+\tau), \tag{5}$$

is the temporal correlation of wavelet functions at time scale j, and $T_j = 2^j$. σ_j^2 is the variance of wavelet coefficient as defined in (3).

The proof of the theorem is contained in Appendix A. It should be emphasized that the above theorem is valid not only for the IWM using the *Haar* wavelet discussed in our previous work but also for IWM's using other orthogonal wavelet functions. In fact, it is valid for all IWM's regardless of wavelet functions used.

For the *Haar* wavelet functions, the temporal correlation of wavelet functions, $h_j(\tau)$, can be easily shown to possess a simple form:

$$h_j(\tau) = \begin{cases} 1 - \frac{3\tau}{T_j} & 0 \leq \tau < T_j/2, \\ \frac{\tau}{T_j} - 1 & T_j/2 < \tau < T_j, \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

Such a $h_j(\tau)$ is plotted in Fig. 1, which shows that $h_j(\tau)$ is piece-wise linear and has limited support. The support $[0, T_j)$ increases as a function of the time scale j. A set $h_j(\tau)$ at all time scales forms a set of basis functions for

[2]Since for the independent wavelet models, the wavelet coefficients are zero-mean, auto-correlation and auto-covariance functions are the same. We will use both terms interchangeably.

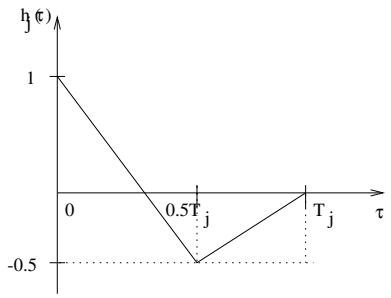

Fig. 1. $h_j(\tau)$ versus τ.

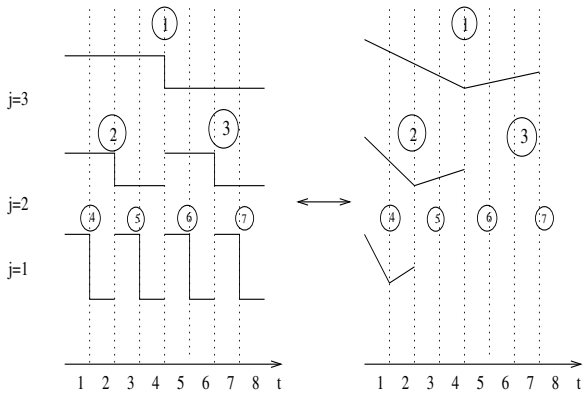

Fig. 2. Left side shows the *Haar* wavelet basis functions. Right side illustrates the corresponding basis $h_j(\tau)$ of the auto-covariance function.

the average auto-covariance function $\bar{R}(\tau)$. The average auto-covariance function given by (4) is therefore a linear combination of the basis functions, where $h_j(\tau)$ is weighted by the variance, σ_j^2, of the wavelet coefficient d_j^m at the same time scale. The *Haar* wavelet basis functions and the corresponding temporal correlation $h_j(\tau)$ are illustrated side-by-side in Fig. 2.

B. Approximation Capability of Independent (Haar) Wavelet Models

Since *Haar* wavelets are the simplest, we first investigated the capability of independent *Haar* wavelet models [5]. Intuitively, capabilities of independent *Haar* wavelet models to approximate auto-covariance functions of heterogeneous traffic can be understood through Fig. 2. Since (*Haar*) wavelet basis themselves capture most of the (temporal) dependence in original heterogeneous traffic, statistical dependence among wavelet coefficients are weak enough to be neglected. As a result, independent wavelet models were shown to be able to capture various decay rates of auto-covariance functions as described in a theorem given in [5].

As shown by the theorem, average auto-covariance function from the independent *Haar* wavelets model has the same decay rate as the temporal correlation in heterogeneous network traffic. Therefore, the IWM based on *Haar* wavelets could be considered to have the so-called universal approximation capability of auto-covariance functions in large lags.

C. Evidence of Insufficient Modeling with IWM-1

How well do independent *Haar* wavelet models approximate an auto-covariance function at small lags? To provide an answers to this question is important since capturing short-range dependence in heterogeneous traffic is important to many applications [12].

To clearly demonstrate the modeling accuracy of IWM-1 in terms of the auto-covariance function at all lags, we

define the following two mean-squared sense error functions. The first one is the mean squared error (MSE) defined as

$$MSE(k) = \frac{1}{k}\sum_{i=1}^{k}\left[\hat{R}(i) - R(i)\right]^2, \tag{7}$$

where $R(i)$ is the sample auto-covariance function directly obtained from the trace and $\hat{R}(i)$ is the sample auto-covariance function from the wavelet models. The second one is the segmented mean squared error (SMSE) defined as

$$SMSE(k;s) = \frac{1}{s}\sum_{i=ks-s+1}^{ks}\left[\hat{R}(i) - R(i)\right]^2, \tag{8}$$

where s is the segment size, and k is the index of segment.

The first error function (MSE) enables us to observe how the mean squared error evolves with additional samples and the second error function (SMSE) enables us to figure out which part of the auto-covariance function contributes the most to the errors.

The MSE to the auto-correlation of the IWM-1 for an AR(1) process is shown in solid line from the top in Fig. 7. The SMSE of the same process is shown in left most bar in Fig. 8. It is observed from the figures that IWM-1 can capture the correlations at large lags very accurately. At relatively small lags, the errors are more significant. The same phenomenon can be observed from the error functions for Bellcore trace pAug in Fig. 9 and Fig. 10.

IV. Improving the Independent wavelet models

One of the possible reasons, which lead to the modeling inaccuracy especially in small lags, is that IWM's neglect the possible correlations between wavelet coefficients. To improve the modeling accuracy, the neglected correlations have to be taken into account. One way to accomplish that is to use orthogonal wavelet with higher vanishing moments. As will be shown later, by using orthogonal wavelet with higher vanishing moments, more correlations can be embedded in the model, leaving the ignored correlations small. The other way is to model the correlations directly. Markov models in wavelet domain fall into this category. We will discuss using other wavelet first. Then, we will discuss using Markov models in wavelet domain in Section V.

A. Motivation on Using Wavelets with Higher Vanishing Moments

Our intuition for using orthogonal wavelets with higher vanishing moments is based on the following theorem developed by Tewfik and Kim [13]. Similar result for a particular time scale j was also given by Flandrin [14].

Theorem 2 (Tewfik and Kim) Let $B_H(t)$ be fractal Brownian motion (fBm) with a parameter H, $0 < H < 1$. Further more, let $\psi(t)$ be a discrete orthogonal wavelet with M vanishing moments and supported over the interval $[-K_1, K_2]$, $K_1, K_2 > 0$. Define the stochastic process $b(j;m)$ as

$$b(j;m) = \sqrt{2^j}\int_{-\infty}^{\infty} B_H(t)\psi(2^j t - m)dt,$$

where the integral holds in the mean-square sense and where the cross-correlation $R_b(j,k;m,n)$ between $b(j;m)$ and $b(k;n)$ is defined as

$$R_b(j,k;m,n) = \mathbf{E}[b(j;m)b(k;n)].$$

Then,

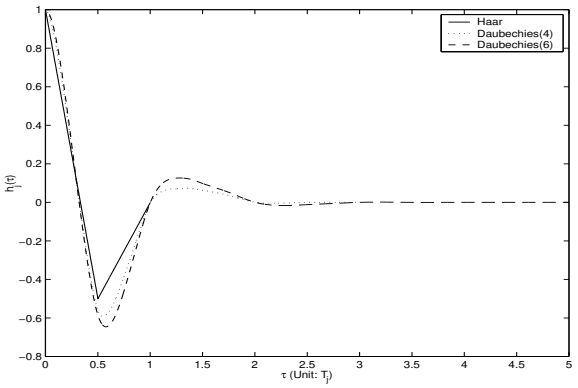

Fig. 3. $h_j(\tau)$ versus τ for different wavelet functions. Solid line: *Haar* wavelet; dotted line: *Daubechies* with vanishing moment 2; dashed line: *Daubechies* with vanishing moment 3.

1. for fixed j the 1-D process $2^{j(H+0.5)}\, b(j;m)$ are stationary and have the same statistics;
2. for fixed j and k, $R_b(j,k;m,n)$ decays as $O(|2^{-j}m - 2^{-k}n|^{2(H-M)}$ for all m, n such that $|2^{-j}m - 2^{-k}n| > \max(2^{-j}K_1 + 2^{-k}K_2, 2^{-k}K_1 + 2^{-j}K_2)$.

The first statement of Theorem 2 is in accord to our development of IWM's. The second statement shows that as the vanishing moments M increase, the decay rate of the cross-correlation, $R_b(j,k;m,n)$, increases accordingly. In another word, the higher the vanishing moments of the wavelet function, the less correlations among wavelet coefficients. As the IWM ignoring the possible correlations among wavelet coefficients, the higher vanishing moments the less ignored correlations. Hence, intuitively, the better the modeling performance.

However, the theorem is only for fBm, which is a long-range dependent process. In addition, it is still not clear that how the vanishing moments and the correlations of wavelet coefficients relate to auto-correlation function. In the following section, we address this issue, and then provide empirical results to compare the performance of various models with different vanishing moments.

B. Independent wavelet models with higher vanishing moments

In the family of orthogonal wavelets, *Daubechies* wavelets[3] are chosen due to three reasons. First of all, the wavelet functions are compactly supported, which provides the time localization property. The second reason is that the support in the time domain is not expanded too rapidly with the increase of vanishing moments. For instance, *Daubechies* wavelet functions have support size of $2M-1$ given the vanishing moment is M [15]. In fact, *Daubechies* wavelets have a support of minimum size for any vanishing moment M. The last reason is backward compatibility. That is, *Daubechies* wavelet collapses to *Haar* wavelet when $M = 1$.

The average auto-covariance function obtained in Theorem 1 is still valid in this case. Different wavelet functions have different $h_j(\tau)$. Unlike *Haar* wavelets, *Daubechies* wavelets with the vanishing moments M greater than 1 do not have closed form expressions of $h_j(\tau)$. However, $h_j(\tau)$ for different wavelet functions can be drawn in Fig. 3. As the figure illustrates, the support of $h_j(\tau)$ increases as the increase of vanishing moments M, and so does its smoothness. In specific, when the vanishing moments $M = 1$, $h_j(\tau)$ is a piece-wise linear functions. When $M = 2$, $h_j(\tau)$'s derivative becomes continuous. When $M = 3$, second order derivative of $h_j(\tau)$ is continuous.

C. Empirical Results

To further validate our intuition, experiments are conducted. In the experiments, the following traces were used: SRD traffic generated from AR(1), LRD traffic from FARIMA(0,0.4,0), a mixture of SRD and LRD from FARIMA(1,0.4,0),

[3]We consider *Daubechies* wavelets with minimum phase factorization.

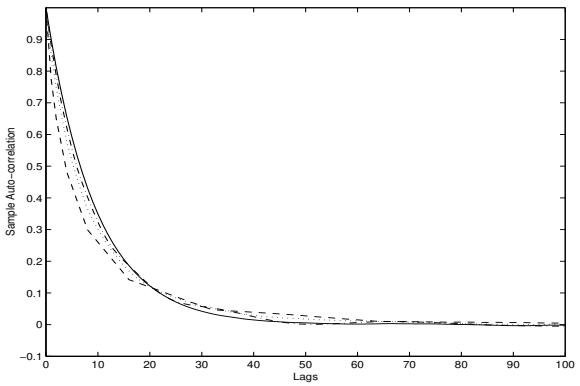

Fig. 4. Autocorrelation for an AR(1). Solid line: sample correlation; dashed line: sample correlation from IWM-1; dotted line: sample correlation from IWM-2; dashed-dotted line: sample correlation from IWM-5.

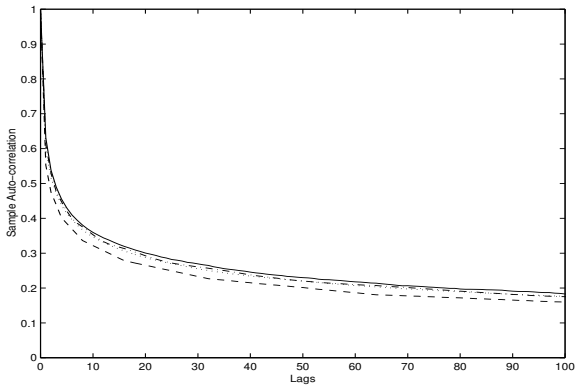

Fig. 5. Autocorrelation for an FARIMA(0, 0.4, 0). Solid line: sample correlation; dashed line: sample correlation from IWM-1; dotted line: sample correlation from IWM-2; dash-dotted line: sample correlation from IWM-5.

and an *Ethernet* data trace (pAug[4]).

Sample correlations for an AR(1) are compared with those from IWM-1, IWM-2 and IWM-5 in Fig. 4. We can clearly observe from the figure that the sample correlations are better matched as the vanishing moments of IWM increase. Fig. 5 and Fig. 6 illustrate similar results for an FARIMA(0, 0.4, 0) and Bellcore trace pAug. A possible explanation for such a better match is that auto-correlation functions are smooth, as the vanishing moments increase, so does the smoothness of the corresponding $h_j(\tau)$. In the case of *Haar* wavelet, $h_j(\tau)$ is a piece-wise linear function. It can only approximate piece-wise linear functions.

The MSE function for the AR(1) is shown in Fig. 7. It is clear that, as the vanishing moments of wavelet function increase, the MSE measure decreases accordingly. Moreover, the errors are reduced significantly as the vanishing moments increased from 1 to 2. Fig. 8 is the SMSE function for the AR(1) with segment size of 100. In other words, the mean squared errors are obtained every 100 lags. The figure shows that the errors at small lags are overwhelming as opposed to the errors at relatively large lags. Again, IWM-2 has the largest gain in terms of reducing the errors. Similarly, the MSE function and SMSE function for Bellcore trace pAug are illustrated in Fig. 9 and Fig. 10, respectively. Fig. 9 shares the same features with Fig. 7. Fig. 10 has a little difference as opposed to Fig. 8. The errors at small lags in Fig. 10 are not that dominant as observed in Fig. 7. It is also interesting to observe from Fig. 10 that the SMSE does not always decrease with the increase of vanishing moments.

[4]See [16] for detailed description of this trace.

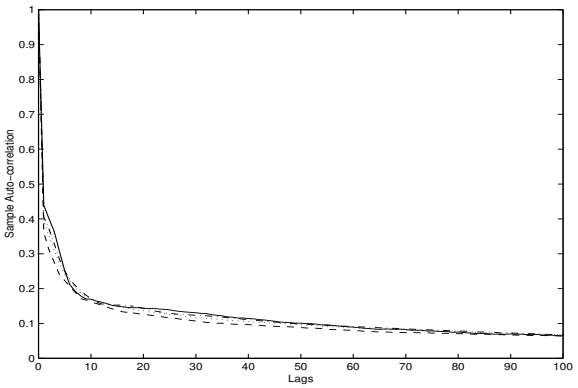

Fig. 6. Autocorrelation for Bellcore trace pAug. Solid line: sample correlation; dashed line: sample correlation from IWM-1; dotted line: sample correlation from IWM-2; dash-dotted line: sample correlation from IWM-5.

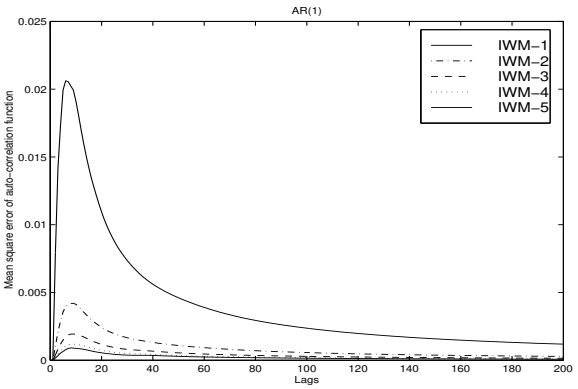

Fig. 7. MSE for AR(1). (From top down) Solid line: IWM-1; dash-dotted line: IWM-2; dashed line: IWM-3; dotted line: IWM-4; solid line: IWM-5.

D. Discussion

Why do wavelets with higher vanishing moments lead to better models in terms of auto-covariance functions? As observed through Fig. 3, the basis of average auto-correlation functions, $h_j(\tau)$, are smoother for wavelets with higher vanishing moments than that of *Haar* wavelets. Since auto-correlations are smooth functions in our experiments, wavelets with higher vanishing moments can thus provide a better modeling capability. However, extra computational time is needed, and proportional to the number of the vanishing moments. In addition, as the vanishing moment increases, the support of wavelets increases as well, and in turn the edge-effect, i.e., the situation that a support can not be fully covered by the data, will be more pronounced. As a result, a wavelet with a moderate high vanishing moment (for example, 2) can be used at the cost of the extra computation. In fact, this extra computational cost is moderate, since the computational cost is still $O(N)$, where N is the length of a trace.

It should be noted that since *Haar* wavelets are essentially rectangular window functions with finite supports, *Haar* wavelet coefficients can be related directly to cumulative processes [4]. This is an important property which makes it possible to use wavelet models to predict buffer loss probabilities. The wavelets of higher vanishing moments, however, can not be directly related to cumulative processes.

V. Beyond the independent wavelet models

If *Haar* wavelets are still being used but the corresponding average auto-correlation functions need to be more accurate at small lags, an alternative approach is to use the so-called Markov wavelet models (MWM). The essence

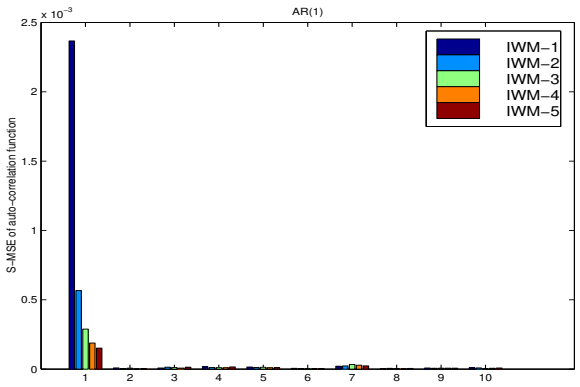

Fig. 8. SMSE for AR(1), segment size $= 100$. From left to right: IWM-1, IWM-2, IWM-3, IWM-4, IWM-5.

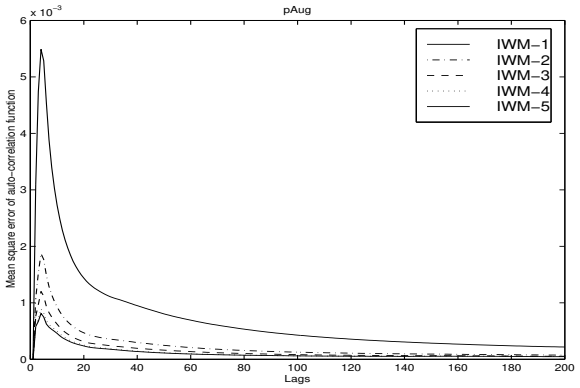

Fig. 9. MSE for Bellcore trace pAug. (From top down) Solid line: IWM-1; dash-dotted line: IWM-2; dashed line: IWM-3; dotted line: IWM-4; solid line: IWM-5.

of MWM is to incorporate the correlations of the wavelet coefficients through Markov models in the wavelet domain across time scales. In what follows, we will describe key characteristics of MWM. More detailed descriptions can be found in [17] [3].

A. Markov wavelet models

To define Markov wavelet models, we first need to define the equivalent one-dimensional indices for wavelets coefficients through the tree diagram [18] [3]. The tree diagram is illustrated by the right side of Fig. 11, in which the circled numbers, assigned sequentially to wavelet coefficients from top to down and left to right, represent the one-dimensional indices of the wavelet basis functions. It can be easily seen that the one-dimensional index s can be one-to-one mapped onto the two dimensional index $(j(s), m(s))$, where $j(s)$ and $m(s)$ represent the scale and the shift index of the s-th wavelet, respectively. Therefore, the equivalent notation of d_s is $d_{j(s)}^{m(s)}$. For example, d_6 is d_1^2 (the shift index, m, starts from 0.) in the given example. Furthermore, we define the parent and the neighbor wavelets of a wavelet through the tree diagram. As illustrated in Fig. 11, $\gamma(s)$ is defined to be the parent node of node s. $\nu(s)$ is defined to be the left neighbor of node s.

Let d_s be the wavelet coefficient with one-dimensional index. In general, dependence among wavelet coefficients can be captured by Markov models. Such Markov models can be implemented through a linear model which relate a wavelet coefficient with others [19].

Assuming causal relations between wavelet coefficients, we have

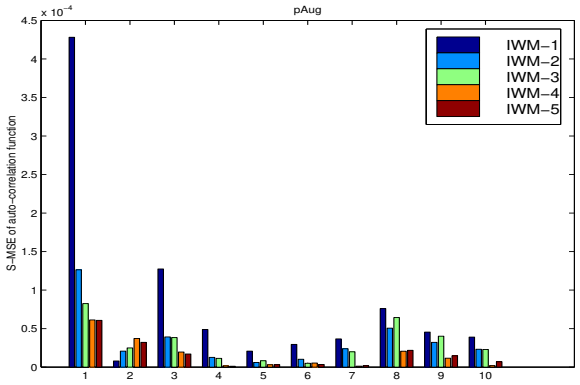

Fig. 10. SMSE for Bellcore trace pAug, segment size $= 100$. From left to right: IWM-1, IWM-2, IWM-3, IWM-4, IWM-5.

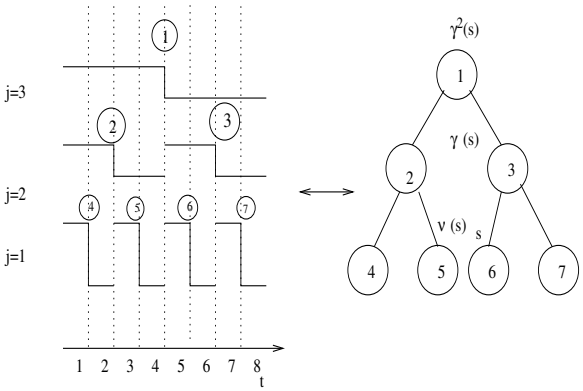

Fig. 11. Left figure shows the *Haar* wavelet functions. Right figure illustrates the corresponding tree diagram and two types of operations. $\gamma(s)$ is defined as the parent node of s. $\nu(s)$ is defined as the left neighbor of node s. The number in the circle represents the one dimensional index of the wavelet functions. s, $\nu(s)$, and $\gamma(s)$ are one dimensional indices of wavelet functions.

$$d_s = \sum_{l=1}^{s-1} a_s(l) d_l + b_s w_s, \tag{9}$$

where $a_s(l)$, for $1 \leq l \leq N$, and b_s are weighting factors depending on the one-dimensional index s (see Fig. 11). w_s is i.i.d Gaussian noise with unit variance.

B. First Order MWM

It was shown [3] that Markov models based on parent-child relation had better performance as opposed to the ones based on neighbors. First order Markov model (MWM-1) can therefore be established to capture the parent-child correlation. Using the one-dimensional index, the first order MWM can be represented as

$$d_s = a'(j(s)) d_{\gamma(s)} + b'(j(s)) w_s,$$

where $a'(j(s))$ and $b'(j(s))$, which only depend on the scale index $j(s)$, are the parameters to be determined from data. s is the one dimensional index of a wavelet coefficient. w_s is Gaussian noise with zero mean[5] and unit variance.

[5]It can be shown easily that the mean of wavelet coefficients of a stationary process is zero.

This is a special case of (9) for all $a_s(l)$'s to be zero except $a_s(\gamma(s))$. $\gamma(s)$ represents the parent of s.

Furthermore, the wavelet correlations of MWM-1 can be derived as

$$\mathbf{E}[d_{j_1}^{m_1} d_{j_2}^{m_2}] = \begin{cases} \sigma_j^2 & m_1 = m_2 \,\text{and}\, \mathrm{j_1 = j_2 = j}, \\ \Pi_{j=j_1}^{j_2-1} a_j \sigma_{j_1}^2 & j_2 > j_1 \,\text{and}\, \mathrm{m_2} = \left\lfloor \frac{\mathrm{m_1}}{2^{\mathrm{j_2 - j_1}}} \right\rfloor, \\ \Pi_{j=j_2}^{j_1-1} a_j \sigma_{j_2}^2 & j_1 > j_2 \,\text{and}\, \mathrm{m_1} = \left\lfloor \frac{\mathrm{m_2}}{2^{\mathrm{j_1 - j_2}}} \right\rfloor, \\ 0 & \text{otherwise.} \end{cases}$$

$\lfloor x \rfloor$ represents the largest integer smaller or equal to x.

For simplicity, we focus on the correlation between a wavelet and its parent, and ignore others. That is, we assume that

$$\mathbf{E}[d_{j_1}^{m_1} d_{j_2}^{m_2}] = \begin{cases} \sigma_j^2 & m_1 = m_2 \,\text{and}\, \mathrm{j_1 = j_2 = j}, \\ \rho_{j_1} \sigma_{j_1} \sigma_{j_2} & j_2 = j_1 + 1 \,\text{and}\, \mathrm{m_2} = \left\lfloor \frac{\mathrm{m_1}}{2} \right\rfloor, \\ \rho_{j_2} \sigma_{j_1} \sigma_{j_2} & j_1 = j_2 + 1 \,\text{and}\, \mathrm{m_1} = \left\lfloor \frac{\mathrm{m_2}}{2} \right\rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

where $\rho_j = a_j \sigma_j / \sigma_{j-1}$ is the normalized correlation.

Based on (10), we have the following theorem.

Theorem 3: For the first order Markov wavelet models, with correlation structure expressed in (10), the average auto-covariance function $\bar{R}(\tau)$ can be expressed as

$$\bar{R}(\tau) = \sum_{j=1}^{+\infty} \sigma_j^2 \, 2^{-j} \, h_j(\tau) + \sum_{j=1}^{+\infty} \sigma_j \sigma_{j+1} \, 2^{-j} \, \rho_j h_j^{(p)}(\tau), \tag{11}$$

where $h_j(\tau)$ is the temporal correlation between wavelet functions at time scale j as defined in (5), and $h_j^{(p)}(\tau)$ is the temporal correlation between parent and children defined as

$$h_j^{(p)}(\tau) = \sum_{t=0}^{T_{j+1}-1} \psi_{j+1}^0(t) \left[\psi_j^0(t) + \psi_j^1(t+\tau) \right]. \tag{12}$$

Proof of the theorem can be found in Appendix B.

Compared with (4) for IWM, the average auto-correlation function of MWM-1 has an extra term determined by the correlation strength ρ_j and the temporal correlation between a wavelet and its parent. Because of this extra term with a free parameter, the average auto-correlation function of MWM-1 is more accurate in general than that of IWM.

Similar to the parent-child correlation, if more correlations, such as grandparent-child and grand-grandparent-child, are considered, the corresponding average auto-correlation functions can be derived using the same approach, and is ignored in this paper. Instead, empirical results are given to further study the performance of MWM's with different orders in next section.

C. Empirical Results

Sample correlations for AR(1) are compared with those from MWM-4, MWM-1 and IWM in Fig. 13. It is clear that as more and more correlations are incorporated in the wavelet models the sample correlations are better matched. With fourth-order Markov model (MWM-4), the approximation errors are merely seen. Fig. 14 illustrates a similar result for FARIMA(1, 0.4, 0). The MSE functions for AR(1) and FARIMA(1, 0.4, 0) are shown in Fig. 15 and Fig. 16, respectively.

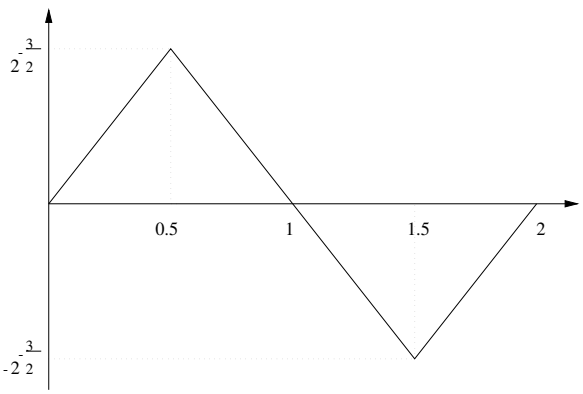

Fig. 12. $h_j^{(p)}(\tau)$ versus τ. X-axis: τ (Unit: T_j); Y-axis: $h_j^{(p)}(\tau)$

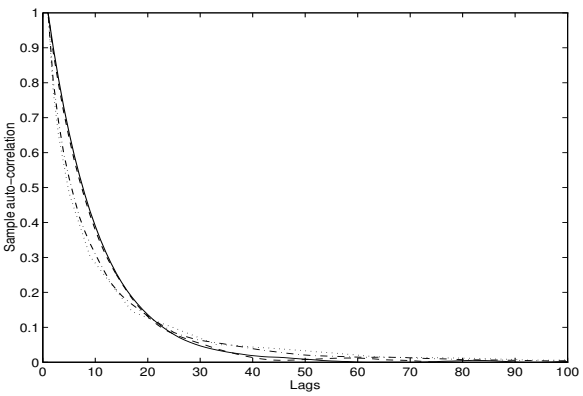

Fig. 13. Autocorrelation function for an AR(1). Solid line: sample correlation; dashed line: sample correlation from MWM-4; dash-dotted line: sample correlation from MWM-1; dotted line: sample correlation from IWM.

From Fig. 13 to Fig. 16, it is concluded that Markov wavelet models can give a better match in terms of sample correlation functions, especially at small lags. The cost includes more parameters, and computational time. But compared with independent (*Haar*) wavelet model, the increase in the cost is moderate. That is, when a trace of length N is used, the computational cost of (low-order) Markov wavelet models is still $O(N)$. In addition, the parameters of Markov wavelet models can be easily obtained in the wavelet domain using measurements.

VI. Conclusions

Independent wavelet models based on *Haar* wavelets have shown to be able to approximate various decay rates of auto-covariance functions at large lags in heterogeneous traffic. However, the information lost due to negligence of (short-range) dependence among wavelet coefficients results in deviations from the original auto-covariance functions at small lags.

The first approach used to improve upon independent (*Haar*) wavelet models employs orthogonal wavelets with higher vanishing moments. This approach improves the performance through providing a smoother temporal correlation function of the wavelet function than that of *Haar* wavelets. Analytical average auto-correlation functions have been derived for independent wavelet models. From the analytical expression, we found that the high-order vanishing moments provide the smoother temporal correlation function of the wavelet function than that of the *Haar* wavelet, and thus improve the performance on modeling a smooth autocorrelation functions.

The second approach explores a correlation model based on Markov models in the wavelet domain. Then major dependence among wavelet coefficients are captured through (low-order) Markov models used across time scales. For

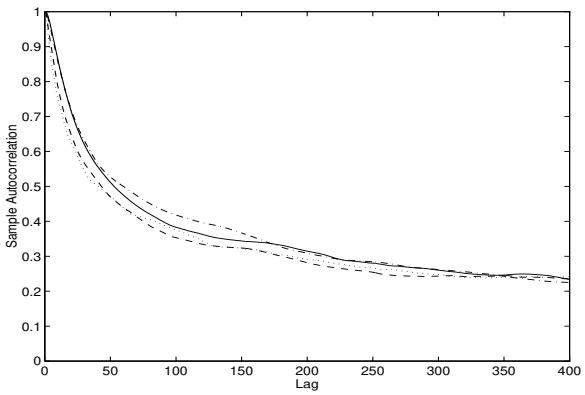

Fig. 14. Autocorrelation function for an FARIMA(1, 0.4, 0). Solid line: sample correlation; dashed line: sample correlation from MWM-4; dash-dotted line: sample correlation from MWM-1; dotted line: sample correlation from IWM.

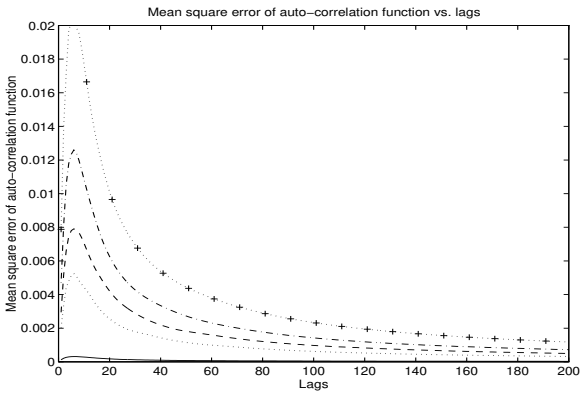

Fig. 15. Mean squared error for an AR(1). Dotted line with cross: IWM; dash-dotted line: MWM-1; dotted line: MWM-3; solid line: MWM-4.

instance, the first-order Markov wavelet model improves performance through providing two more free parameters, and utilizes the statistical correlations between a wavelet and its parent.

To compare these two approaches in improving the performance, we note that using wavelets with low vanishing moments (e.g. $M = 2$) can improve the performance substantially. The first-order Markov model only improves marginally. But by increasing the order of Markov wavelet models, the performance can be improved systematically. In terms of computational costs, the computation complexity of either approaches increases proportionally as the order of vanishing moments increases, or as the order of a Markov model. Therefore, these two approaches have roughly similar computational complexity, which is higher than that of independent (*Haar*) wavelet models. But the computational complexity is $O(N)$ for both approaches, when a trace of length N is used. In addition, both approaches are measurement-based with parameters that can be readily estimated from network measurements.

ACKNOWLEDGMENTS

The authors wish to thank Thomas Yu for helping us generate curves of autocorrelation of *Daubechies* wavelet functions.

REFERENCES

[1] Mark W. Garrett and Walter Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proc. ACM SIGCOMM'94*, London, England, UK, 1994, pp. 269–280.

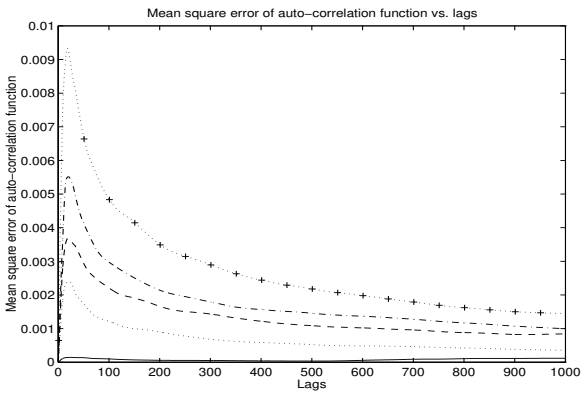

Fig. 16. Mean squared error for an FARIMA(1, 0.4, 0). Dotted line with cross: IWM; dash-dotted line: MWM-1; dotted line: MWM-3; solid line: MWM-4.

[2] Changcheng Huang, Michael Devetsikiotis, Ioannis Lambadaris, and A.Roger Kaye, "Modeling and simulation of self-similar variable bit rate compressed video: A unified approach," in *Proc. ACM SIGCOMM'95*, Cambridge, MA, 1995, pp. 114–125.

[3] Sheng Ma and Chuanyi Ji, "Modeling video traffic in the wavelet domain," in *Proc. IEEE INFOCOM'98*, San Francisco, CA, April 1998, vol. 1, pp. 201–208.

[4] Sheng Ma and Chuanyi Ji, "Modeling video traffic using wavelets," *IEEE Communication Letters*, vol. 2, no. 4, pp. 100–103, April 1998.

[5] Chunayi Ji, Sheng Ma, and Xusheng Tian, "Approximation capability of independent wavelet models to heterogeneous network traffic," in *Proc. IEEE INFOCOM'99*, New York, NY, March 1999, vol. 1, pp. 170–177.

[6] Sheng Ma and Chuanyi Ji, "Time-scale wavelet modeling for non-Gaussian network traffic," unpublished.

[7] Rudolf H. Riedi, Mattew S. Crouse, Vinay J. Ribeiro, and Richard G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 992–1018, April 1999.

[8] Patrice Abry and Darryl Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 2–15, January 1998.

[9] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

[10] Gregory W. Wornell and Alan V. Oppenheim, "Wavelet-based representations for a class of self-similar signals with application to fractal modulation," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 785–800, March 1992.

[11] Gregory W. Wornell, "A karhunen-loeve-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 859–861, July 1990.

[12] Bong K. Ryu and Anwar Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," in *Proc. ACM SIGCOMM'96*, San Francisco, CA, 1996, pp. 3–14.

[13] A.H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 904–909, March 1992.

[14] Patrick Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 910–917, March 1992.

[15] Stephane G. Mallat, *A wavelet tour of signal processing*, Academic Press, 1997.

[16] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, February 1994.

[17] Sheng Ma, *Network Traffic Modeling and Analysis*, Ph.D. thesis, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechinc Institute, Troy, NY 12180, May 1998.

[18] Mark R. Luettgen, William C. Karl, and Alan S Willsky, "Multiscale representations of markov random fields," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3377–3396, December 1993.

[19] Michele Basseville, Albert Benvenists, Kenneth C. Chou, Stuart A. Golden, Ramine Nikoukhah, and Alan S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 766–784, March 1992.

APPENDIX

I. PROOF OF THEOREM 1

Let $x_K(t)$ be the resolution-limited approximation of $x(t)$, and $\bar{R}_K(\tau)$ be the resolution-limited approximation of $\bar{R}(\tau)$, i.e.

$$x_K(t) = \sum_{j=1}^{K} \sum_{m=0}^{2^j-1} d_j^m(t)\psi_j^m(t), \tag{13}$$

$$\bar{R}_K(\tau) = \frac{1}{T_k} \sum_{t=0}^{T_K-1} \mathbf{E}[x_K(t)x_K(t+\tau)], \tag{14}$$

and

$$x(t) = \lim_{K\to\infty} x_K(t), \tag{15}$$

$$\bar{R}(\tau) = \lim_{K\to\infty} \bar{R}_K(\tau). \tag{16}$$

By plugging (3) and (13) into (14) and manipulating the expression for $\bar{R}_K(\tau)$, we have

$$\begin{aligned} \bar{R}_K(\tau) &= \sum_{j=1}^{K} \sigma_j^2 \sum_{m=0}^{T_{K-j}-1} \frac{1}{T_K} \sum_{t=0}^{T_k-1} \psi_j^m(t)\,\psi_j^m(t+\tau) \\ &= \sum_{j=1}^{K} \sigma_j^2 \frac{1}{T_j} \sum_{t=0}^{T_j-1} \psi_j^0(t)\,\psi_j^0(t+\tau), \end{aligned} \tag{17}$$

where $T_j = 2^j$. The last step in the above equation is obtained by using the fact that:

$$\sum_{t=0}^{T_j-1} \psi_j^0(t)\psi_j^0(t+\tau) = \sum_{t=0}^{T_j-1} \psi_j^m(t)\,\psi_j^m(t+\tau),$$

for $0 \le m \le 2^{K-j} - 1$. Define the (deterministic) temporal correlation, $h_j(\tau)$, of wavelet functions $\psi_j^m(t)$ as in (5). Then the average auto-correlation function in (17) can be rewritten as:

$$\bar{R}_K(\tau) = \sum_{j=1}^{K} \sigma_j^2\, 2^{-j}\, h_j(\tau).$$

Therefore the average auto-correlation function of an independent wavelet model with un-limited resolution can be expressed as in (4).

II. Proof of Theorem 3

The definitions of $x_K(t)$, $\bar{R}_K(\tau)$, $x(t)$ and $\bar{R}(\tau)$ are shown in equations (13) (14) (15) and (16), respectively. Instead of using the independent model expressed in (3), parent-child type Markov model expressed in (10) is used to derive the average auto-covariance function for MWM-1.

Plugging (10) and (13) into (14) and manipulating the expression for $\bar{R}_K(\tau)$, we have

$$\bar{R}_K(\tau) = \sum_{j=1}^{K} \sigma_j^2 \frac{1}{T_j} \sum_{t=0}^{T_j-1} \psi_j^0(t)\, \psi_j^0(t+\tau) + \sum_{j=1}^{K-1} \sigma_j \sigma_{j+1} \frac{1}{T_j} \rho_j \sum_{t=0}^{T_{j+1}-1} \psi_{j+1}^0(t) \left[\psi_j^0(t) + \psi_j^1(t)\right], \tag{18}$$

where $T_j = 2^j$. Define the (deterministic) temporal correlation, $h_j(\tau)$ and $h_j^{(p)}(\tau)$ as expressed in (5) and (12). Then the average auto-correlation function in (18) can be rewritten as:

$$\bar{R}_K(\tau) = \sum_{j=1}^{K} \sigma_j^2\, 2^{-j}\, h_j(\tau) + \sum_{j=1}^{K-1} \sigma_j \sigma_{j+1}\, 2^{-j}\, \rho_j h_j^{(p)}(\tau).$$

Therefore the average auto-correlation function of an first order Markov wavelet model with un-limited resolution can be expressed as in (11).