International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 3, Issue 12 (July 2014) PP: 41-48

Distributed Web System Performance Improving Forecasting Accuracy

Shraddha B. Rahane[1], Dr. Sanjay T. Singh[2]
[1]*Research Scholar, Department of Computer, PKTC College, Chakan 410501, Pune, India.*
[2] *Professor, Department of Computer Engineering, PKTC College , Chakan 410501,Pune, India.*

Abstract: *Recently with the increasing development of distributed computer systems (DCSs) in networked industrial and manufacturing applications on the World Wide Web (WWW) platform, including service-oriented architecture and Web of Things QoS-aware systems, it has become important to predict the Web performance. In this paper, we present Web performance prediction in time by making a forecast of a Web resource downloading using the Efficient Turning Bands (TB) geostatistical simulation method. Real-life data for the research were obtained from our own website named "Distributed forecasting system". Generation of log file form website and performing monitoring of a group of Web clients from connected LAN. For better web prediction we used spatio temporal prediction method with time utility for downloading particular file from website and calculate forecasting result using Turning bands method but improving more forecasting accuracy use the efficient turning band method basically efficient turning band use Naive bays algorithm and calculate efficient result and that result is compared with Turning band and efficient turning band method. The efficient turning band method result show good forecasting quality of Web performance prediction and forecasting.*

Keyword*: Distributed Computer System, Turning Band,, Qos-Aware Systems, Geostatistics, Spatio-Temporal Prediction.*

I. Introduction

At the time of developing Internet, in which web systems play a significant role,it is worth considering an effective way of evaluating the net performance. Another important question is Internet security which is influenced by web servers reliability and the lack of communication intervals on client-server route. From the economic point of view the problem also seems serious, the interest of net tycoons and large Internet portals, for whom it is important to win and maintains many customers as possible, depends on Internet reliability and its best performance. Thus there arises a clear need for spatial-temporal forecasting of Internet performance which would take into account both the geographical location of web servers and the total downloading time of a given resource. Such need for forecast information arises for example when we have to choose from where we should download information resource which is available on many Internet nodes at various geographical locations to obtain the resource at the minimum (or known) time. Knowing the forecast of transfer throughput from these nodes to our location will allow us to choose the node from which we can receive the resource in the shortest time [1]. Analysis of network measurements plays a fundamental role in Internet knowledge discovery. The results of such analysis can help administrators as well as casual users in studying network characteristics including forecasting of network future performance behavior. In this paper we propose to use the Turning Bands method to predict Web performance that is perceived by end users while downloading resources from Web servers. The contribution of this paper is to propose this innovative application of Turning Bands method. This forecasting method has been originally employed in geostatistics [2] and has some other applications but not still in the field of Internet/Web performance. Networks can be measured using active or passive methods. Active measurement scheme uses a probing approach where a reaction is observed after sending a probe signal (traffic, request) whereas passive measurement scheme is based only on current network observation without any additional traffic injection [3,4,5]. This research is based on active measurements made by MWING system which is Internet measurement infrastructure developed in our Institute [6,7]. MWING has been developed to design and perform both kinds of experiments in a controlled way. The kind of experiment, its aim, time duration and the experiment itself are designed by the researcher using either basic or extended MWING functions. Designed experiment is performed within of MWING platform that employs multi-agent technology. MWING's agents can run on any Internet host performing designed measurement experiment, collecting data and storing it in local or central databases.

In this work we use the measurement data collected in a world-wide experiment that has been designed and performed to monitor a performance of a group of Web servers (sites) perceived from four different Internet locations [1].. The rest of the paper is organized as follows. First, we introduce the Turning Bands (TB) method.

Next we present preliminary and structural data analysis of results of network measurements. After that, we present a spatial forecast of Web server performance made by means of TB method.

The most important requirements concern performance, reliability, quality of service (QoS), and security [6]. Communication issues depend on the specific distributed automation application, Internet and Web technologies, and character of the network (i.e., private or public, factory field bus, local or wide-area, or wired or wireless).The Web-based DCSs are built from DCS nodes which include either Web clients or Web servers (Web services) or both. In client-to-server communication, it is a common situation when a client has to choose a server in case of data or service replication on many servers. To make this operation predictable, Web-based DCSs need a spatio-temporal Web performance forecasting, where the temporal prediction addresses the forecasting of Web performance, ordered by one or more dimensions of time (e.g., the minute-by-minute, hour-by-hour, and day-by-day), whereas *spatial* prediction concerns forecasting of the performance behaviour across nodes of DCS (more specifically, between Web clients and Web servers).The aim of this paper is to present a robust spatio-temporal prediction method and algorithm that can provide an efficient forecasting of a Web client-perceived performance on the Web application. This may provide efficient QoS for individual nodes of Web-based DCS and enable to improve operation of the whole system. The predicted performance characteristics can be used in selection of the best performance Web server and best time. Here, we propose to make Web performance prediction with the use of the Efficient Turning Bands (TB) geostatistical method.

II. Literature Survey

In this section we are presenting the different survey paper for the use of research work.

- Kaminska-Chuchmala and A. Wilczynski describes system[22,24] where very promising results were obtained in studying load forecasting in power distribution grids, where generally forecast in transmission networks is a complex task.
- M.Mirza, J. Sommers, P. Barford, and X. Zhu[12,14], explains performance prediction is one of the key.issues of the Internet and the Web. Many works considered the Internet performance through the transmission delay (RTT), data throughput (TCP throughput). Web performance can be evaluated by the Web page view response time and Web resource download time (or HTTP throughput). Both short term and long-term performance forecasts are considered and aimed at either arbitrary or specific file sizes. Short-term forecasting needs instantaneous measuring of network performance that may cause too much additional traffic, therefore lower sampling rates are proposed but for long-term forecasting.
- L. Borzemski and A. Kaminska-Chuchmala present The work in [19] presents the TB method applied to the dataset containing measurements made by the Wroclaw agent that monitored. Web servers dispersed across the world; this paper deals with the application of the TB method for the dataset collected by the agent monitoring European servers from Gliwice.
- Kaminska-Chuchmala and A. Wilczynski persent the work in [12] employs the TB method in the analysis of a Wroclaw agent monitoring European servers.
- Cagdas Hakan Aladag1, Erol Egrioglu the work in [2] uses the Sequential Gaussian Simulation method deployed for Gliwice agent and European servers introduces a new proposal of the spatio-temporal Web performance forecasting.
- The survey of forecasting Current forecasting technologies refer to quantitative and qualitative methods. Among the quantitative methods, time-series forecasting methods are used to analyze time-dependent series data and predict the future values, brand choice models are used to calculate the probability of choice to predict choice behavior, and Bayesian models are used to infer conditional probability. Qualitative forecasting technology can be described by environment scanning, scenarios, and Delphi. A review of recent advances in technological forecasting can be found in Martino [19]. The forecasting technologies in the presented architecture focus on quantitative methods. A survey of time-series data mining was presented by Keogh & Kasetty [10]. Although some accomplishment have been achieved by the effort, in general difficulties still remain because uncertain demand variations do not follow the same pattern all the time due to consumer behavior, new technologies or products, and other environmental factors.Exploring the composite effects of these factors that influence customer demand may capture the future trend and help reduce the errors that might exist in classical time-series forecasting models.
- A survey of time-series data mining was presented by Keogh & Kasetty [10]. Although some accomplishment have been achieved by the effort, in general difficulties still remain because uncertain demand variations do not follow the same pattern all the time due to consumer behavior, new technologies or products, and other environmental factors. Exploring the composite effects of these factors that influence customer demand may capture the future trend and help reduce the errors that might exist in classical time-series forecasting models.

III. PROPOSED APPROACH FRAMEWORK AND DESIGN

3.1 Problem Definition

Improving forecasting of distributed web-system performance using efficient Turning band method. We have studied the many methods those are presented for forecasting of web system performance by considering the different real time applications.

In [1] we have studied the TB geostastical performance prediction method for distributed computer systems and web systems. Following flowchart with reference this is presented in figure 1.

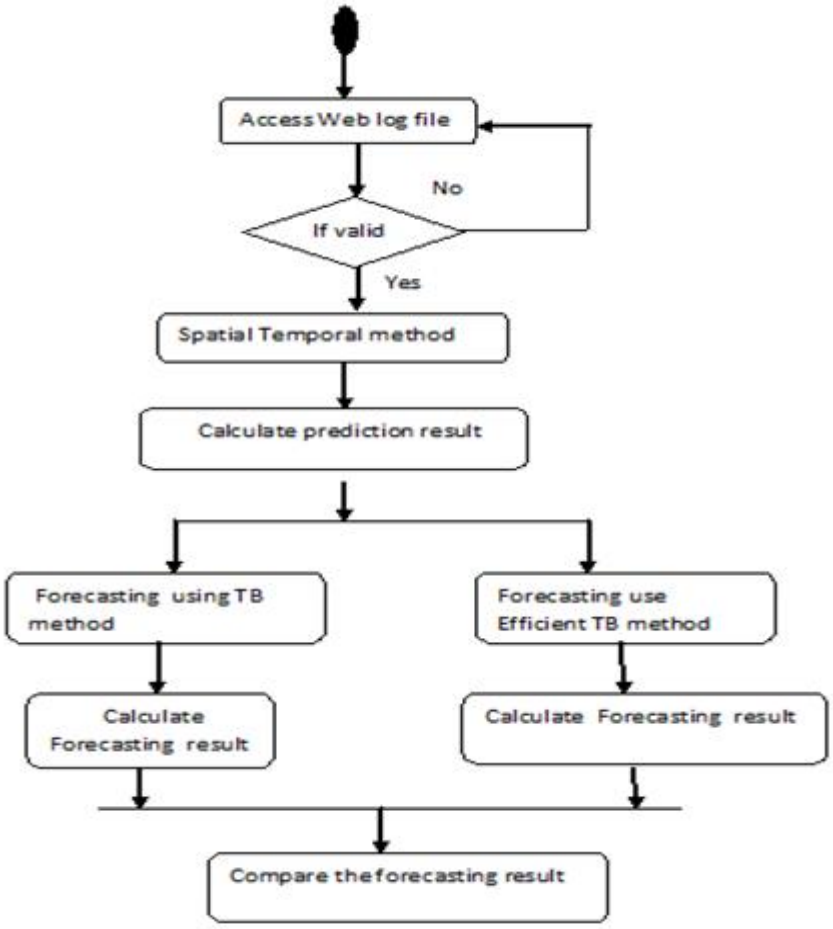

Figure 1: System Flowchart

3.2 Proposed System Architecture and Algorithms

In this we are presenting the new architecture which is based on two concepts first Turning band forecasting method which depicted in [1] and another one is Efficient Turning Band which is depicted in [2]. Based on these terminologies below is proposed architecture for predicting the performance in figure 2. As shown in figure 2, we first collect log data, then apply preprocessing steps like data cleaning, in cleaning process we remove records where status code is 404 or URL contains images like jpg, png etc. Then apply spatio temporal prediction , TB algorithm and Efficient turning band algorithm, finally we will perform analysis with comparison of TB algorithm and efficient TB algorithm. The main idea of the TB method is to perform simulations along several lines, using a unidimensional covariance function that corresponds to the given three-dimensional (3-D) one: X, Y directions—geographical coordinates and Z direction—time axis. The TB algorithm is calculated in cubic time complexity. At each point of the 3-D field, a weighted sum of the corresponding values of the line processes is assigned. In conclusion, the TB method is a multidimensional random number generator for the simulation of the spatially correlated random fields [13], [14].

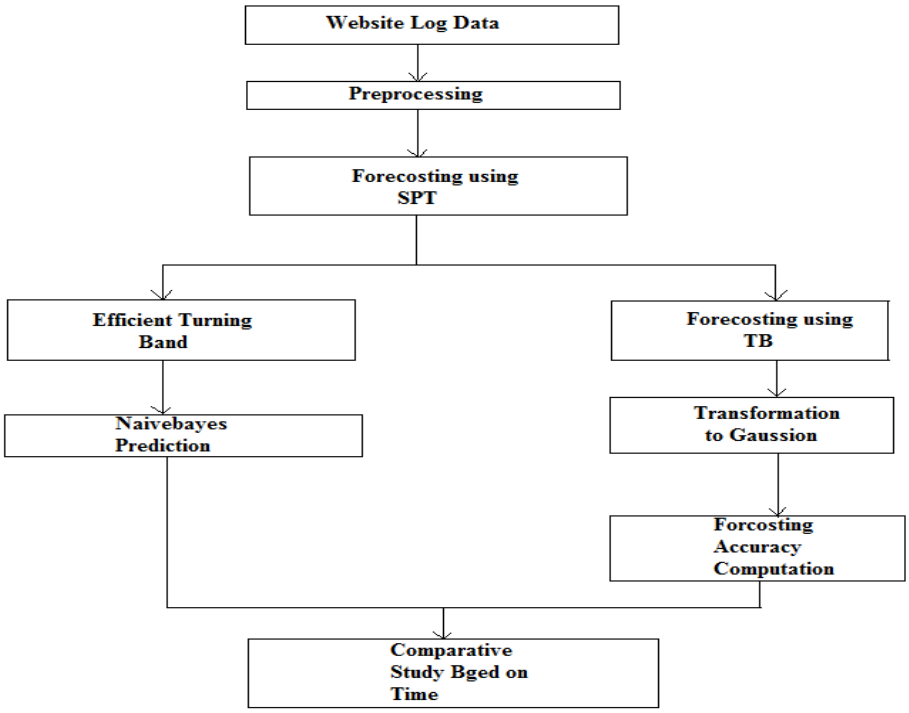

Figure 2: Proposed architecture for predicting the performance

3.3 Turning Bands

The basic assumption of the TB method is as follows: the field to be simulated is second-order stationary and isotropic; at each point, the values of the field are normally distributed and have zero mean. In other cases, the transformation to Gaussian with subsequent subtraction of the mean could be applied. The next assumption is the knowledge of the covariance C(r) of the field
which is to be simulated.

Here, we present the methodology of the proposed approach and the algorithm of the TB method, which will be used for spatio-temporal forecasting of Web system performance (WSP).The basic assumption of the TB method is as follows: the field to be simulated is second-order stationary and isotropic; at each point, the values of the field are normally distributed and have zero mean. In other cases, the transformation to Gaussian with subsequent subtraction of the mean could be applied. The next assumption is the knowledge of the covariance C(r) of the field which is to be simulated. The main idea of the TB method is to perform simulations along several lines, using a uni dimensional covariance function that corresponds to the given three-dimensional (3-D) one: X, Y directions—geographical coordinates; and Z direction—time axis. The TB algorithm is calculated in cubic time complexity. At each point of the 3-D field, a weighted sum of the corre-sponding values of the line processes is assigned. In conclusion, the TB method is a multidimensional random number generator for the simulation of the spatially correlated random fields.

Let (ɷn, n £ N) be a sequence of directions $S + d$ and let $(Xn, n \ £ N)$ be a sequence of independenta stochastic process of covariance. The algorithm of TB method takes the following from CƱn.

TB Algorithm:

1. Directions Selection: $\theta_{1\ldots}\ \theta_n$ so that $1/n \sum_{k=1}^{n} \delta\ \theta_k$ is weakly convergent to ϖ

   θ_n=($\cos(2\Pi un)\sqrt{1-v\hat{}2n}$, $\sin(2\,\Pi un)\sqrt{1-v\hat{}2n}$, vn).

   Un= binary expansion of each integer n=1,2,…….;

   Vn= ternary expansion of each integer n=1,2……..;

2. Convariance function

   C (n)(h)= $1/n + \sum_{k=1}^{n}(C\theta k\ (< h, \theta k >))$ ~ ϖ

   $\theta_{n=}\ \theta_k$

3. Calculate x Ɛ D

$$1/\mathrm{sqrt}(n) + \sum_{k=1}^{n} (C\theta k\ (< h, \theta k >))$$

X= forecasted download time;

3.5 Naive Forecasts Efficient Algorithm

The simplest forecasting technique is termed the naive method. A naive forecast for any period simply projects the previous period's actual value. For example, if demand for a particular health service was 100 units last week, the naive forecast for the upcoming week is 100 units. If demand in the upcoming week turns out to be 75 units, then the forecast for the following week would be 75 units. The naive forecast can also be applied to a data set that exhibits seasonality or a trend. For example, if the seasonal demand in October is 100 units, then the naive forecast for *next October* would equal the actual demand for *October of this year.*

Although this technique may seem too simplistic, its advantages are low cost and ease of preparation and comprehension. Its major weakness, of course, is its inability to make highly accurate forecasts. Another weakness is that it simply replicates the actual data, with a lag of one period; it does not smooth the data.

However, the decision to use naive forecasts certainly has merit if the results *experienced* are relatively close to the forecast (if the resulting accuracy is deemed acceptable). The accuracy of a naive forecast can serve as a standard against which to judge the cost and accuracy of other techniques; the health care manager can decide whether or not the increase in accuracy of another method is worth its additional cost.

Here follow the following step of this algorithm for the system.

1. Generate log file with extra parameters
2. For efficiency and accuracy, here we consider more parameters to forecast a accurate download time for particular ip address
3. Delay: it shows delay in seconds to download time as compare to previous value
4. Speed: We also consider speed of router to recognize exact speed of download for future connections.
5. Distance: For accuracy parameter we also consider distance of server from client on the basis of calculation of a download time.
6. The Microsoft Naive Bayes algorithm is a classification algorithm based on Bayes' theorems, and provided by Microsoft SQL Server Analysis Services for use in predictive modeling. The word naïve in the name Naïve Bayes derives from the fact that the algorithm uses Bayesian techniques but does not take into account dependencies that may exist.
7. This algorithm is less computationally intense than other Microsoft algorithms, and therefore is useful for quickly generating mining models to discover relationships between input columns and predictable columns. You can use this algorithm to do initial exploration of data, and then later you can apply the results to create additional mining models with other algorithms that are more computationally intense and more accurate.

IV. Implementation Details

In this section we represent the input, result of practical work and environment used for implementation.

- First take as input Web log file database and the show log file details and log file is convert into XML format.
- Log file is preprocess using Preprocessing technique and preprocessing log file apply spatial temporal prediction algorithm .The temporal prediction depending on historical database that is using downloading time taken and shows prediction in graphically.
- The prediction result apply tuning bands algorithm The tuning bands result shows forecasting output
- The Efficient turning band method show the accuracy forecasting result the efficient turning band use Navie bays algorithm and calculate efficient and reliable forecasting result .
- Analysis and compare result with exact downloading time using Turning bands algorithm and Efficient turning band algorithm forecasting technique and show the good accuracy result.

V. Results of Practical Work

As dataset we use our own web applications log file generated by IIS server, for that we design one web application named as "distributed web forecasting. Figure 3 shows some content of Log file which we used for forecasting a download time. It contains Date-Time, server IP address, http method, URL, port number, client IP address, browser used, status code, time taken to load page.As we required only few data, so only extract needed data like client IP, URL, status code, time taken.

```
<?xml version="1.0" encoding="us-ascii" standalone="yes" ?>
- <Web-log>
  - <record>
      <Date>2013-11-30</Date>
      <Time>07:04:12</Time>
      <Server_IP>::1</Server_IP>
      <Method>GET</Method>
      <URL>/Web_Performance_Prediction/Home.aspx</URL>
      <URI_Query>-</URI_Query>
      <Port>80</Port>
      <User_Name>-</User_Name>
      <Client_IP>::1</Client_IP>
      <Browser_Used>Mozilla/5.0+(Windows+NT+6.1;+rv:26.0)+Gecko/20100101+Firefox/26.0</Browser_Used>
      <Status>200</Status>
      <Sub_Status>0</Sub_Status>
      <Win_Status>0</Win_Status>
      <Time_Taken>1225</Time_Taken>
    </record>
```

Figure 3: Content of LOG file generated by IIS

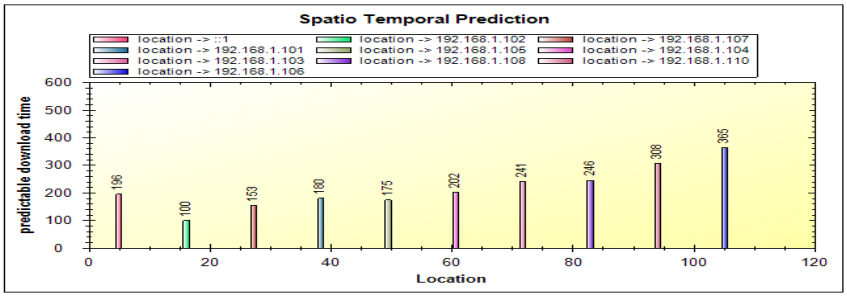

Figure 4: Spatial temporal Web Performance Prediction Result.

Spatio temporal prediction of data with graphical representation is shown in figure 4.Spatio temporal prediction method works with time series, as historical time series defined particular time of instance, and then we calculate prediction time using historical time series of data.

Table 1: Spatio prediction for download time prediction from particular location

IP Address	Predicted time to download (ms)
192:168:1:101	331
192.168:1:104	342
192:168:1:106	456
192:168:1:109	224

Above table contains IP address and predicted time to download file from that IP address. We predict this time using spatio temporal prediction method which uses historical data to predict next download time to download file from particular server.

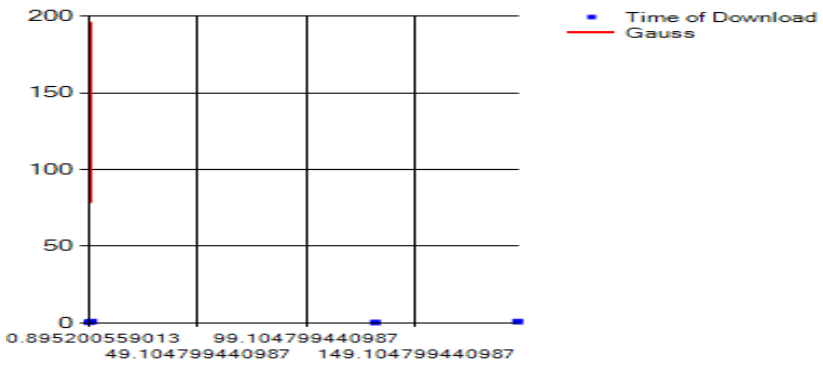

Figure 5: Turning band Gaussian Result

Here we calculate Turning band Transformation Gaussian.

Table 2: Forecasting Table TB method and Efficient TB method

Turning Band- Forecasted ...	Effective Turning Band- Forec
1140.34482758621	1140.34482758621
1140.34482758621	1140.34482758621
997.031388152079	100.67
942.782830196624	1025.69407603891
791.056776556777	960.798213759422
727.505632360471	824.114415182519
595.694222648071	747.802163881725
510.748699454975	625.234061800366
433.809733761171	534.328374579214

Above table 2 show the forecasting value Turning bands method and Efficient turning bands method efficient turning bands method use naive forecast algorithm show the accuracy result.

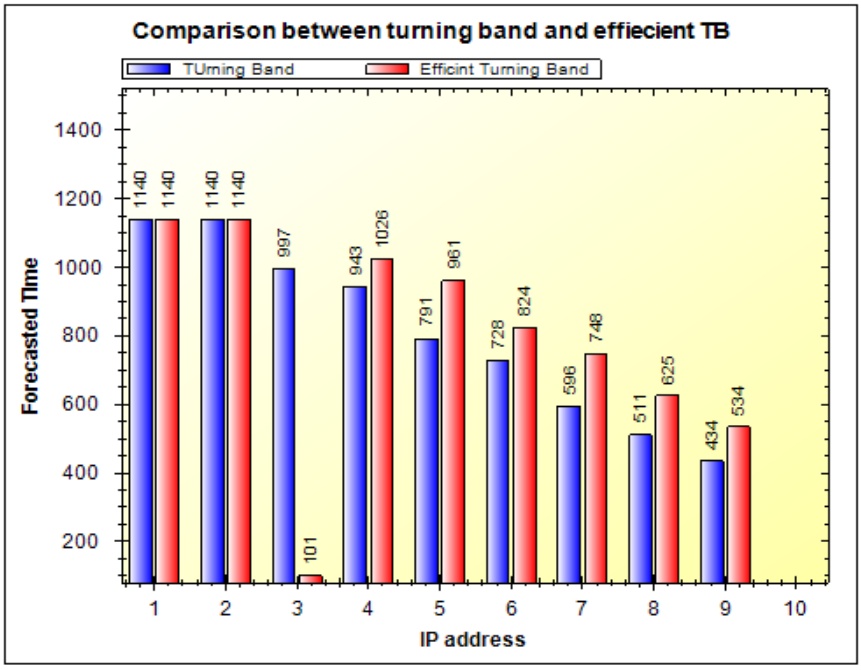

Figure 7: Comparison graph TB method and Efficient TB method

Above figure 5 shows the comparison between Turning bands algorithm and Efficient turning bands algorithm Efficient turning bands algorithm shows the accuracy result.

VI. Conclusion and Future Work

In this research, an approach for predicting Web performance by the innovative application of the Efficient Turning Band geostatistical simulation method proposed. A large-scale measurement experiment was performed in the real-life Internet to gather the data characterizing performance of over LAN and perceived from different clients installed in different Internet locations. The efficient TB based method is for improving the performance forecasting accuracy of whole system. The limitation of TB method has been overcome by using this method. proposed approach is more efficient as compared to the existing method. For the future work, there is still chance to improve this hybrid method and present approach which can handle the different kinds of measurements and data. and try to implement this research in World wide web and give efficiency result.

Acknowledgements

We the author gives special thanks goes to Leszek Borzemski and Anna Kaminska- Chuchmala to contributed this paper, Fellow for their valuable comments and sharing their knowledge.

REFERENCES

[1] A. Kaminska-Chuchmala and A. Wilczynski, "Application of simulation methods to spatial electric load forecasting," Rynek Energii, vol. 80, no. 1, pp. 2–9, 2009.

[2] A. Kaminska-Chuchmala and A. Wilczynski, "Spatial electric load forecasting in transmission networks with sequential Gaussian simulation method," Rynek Energii, vol. 92, no. 1, pp. 35–40, 2011.

[3] A. Kaminska-Chuchmala and A. Wilczynski, "Analysis of different methodological factors on accuracy of spatial electric load forecast performed with turning bands method," Rynek Energii, vol. 87, no. 2, pp. 54–59, 2010

[4] M. Ulieru and S. Grobbelaar, "Engineering industrial ecosystems in a networked world," in Proc. 5th Int. IEEE Conf. Ind. Informat., Vienna, Austria, Jul. 23–27, 2007, keynote address.

[5] Internet-based Control Systems: Design and Applications, Advances in Industrial Control, S-H. Yang, Ed. London, U.K.: Springer-Verlag, 2011.

[6] F. Tao, D. Zhao, Y. Hu, and Z. Zhou, "Resource service composition and its optimal-selection based on particle swarm optimization in manufacturing grid system," IEEE Trans. Ind. Inform., vol. 4, no. 4, pp. 315–327, Nov. 2008.

[7] T. Cucinotta, A. Mancina, G. F. Anastasi, G. Lipari, L. Mangeruca, R. Checcozzo, and F. Rusina, "A real-time service-oriented architecture for industrial automation," IEEE Trans. Ind. Inform., vol. 5, no. 3, pp.267–277, Aug. 2009.

[8] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the Internet of things to the web of things: Resource oriented architecture and best practices," in Architecting the Internet of Things, D. Uckelmann, M. Harrison, and F. Michahelles, Eds. Berlin, Germany: Springer, 2011, pp. 97–129.

[9] Borzemski l., cichocki l., kliber m., fras m., nowak z., mwing: a multiagent system for web site measurements, in: lecture notes in computer science, 4496, 2002, 278–287.

[10] Borzemski l., cichocki l., kliber m., a distributed system to measure the internet based on agent architecture, in: information systems architecture and technology, web-age information systems, eds leszek borzemski [et al.], oficyna wydawnicza politechniki wrocławskiej, 2009.

[11] Borzemski l., nowak z., empirical web performance evaluation with using a mwing system, in: information systems architecture and technology: advances in web-age information systems, eds: leszek borzemski [et al.], oficyna wydawnicza politechniki wroclawskiej, 2009, 25–34.

[12] Lee Y.S., Tong L.I., "Forecasting time series using a me-thodology based on autoregressive integrated moving average and genetic programming," Knowledge-Based Systems 24(1) (2011) pp:66-72.

[13] L. Borzemski, L. Cichocki, andM. Kliber, "Architecture of multiagent Internet measurement system MWING release 2," LNCS, vol. 5559, pp. 410–419, 2009.

[14] L. Borzemski, "The experimental design for data mining to discover web performance issues in a wide area network," Cyber. Syst., vol. 41, no. 1, pp. 31–45, 2010.

[15] Y.-M. Lee and J. H. Ellis, "Turning bands method and its application to geostatistical analysis of groundwater contaminant concentration fields," J. Chin. Inst. Environ. Eng., vol. 10, no. 1, pp. 1–12, 2000.