A Study on Wavelet-based Approaches for Software Reliability Assessment
（ウェーブレットに基づいたソフトウェア信頼性評価に関する研究）

Computer systems are widely used today and in many areas they serve the key function in achieving highly complicated and safety-critical missions. At the same time, the size and complexity of computer systems have continued to increase, making its performance evaluation more difficult than ever before. Software is an important element in computer systems. It has to work with hardware together successfully to complete many critical computer tasks. Different from hardware, the software does not wear-out and can be easily reproduced. Furthermore, software is usually debugged during the testing phase so that their reliability is improved over time as a result of detecting and removing software faults.

It has been widely recognized in the software reliability engineering community that quantitative assessment of software reliability is an essential issue in this area. In general, one needs several kinds of mathematical models to estimate the quantitative software reliability, which is defined as the probability that software system does not fail during a specified time period in a specified environment. Well-structured software reliability models (SRMs) can provide reliable assessment results on the quantitative software reliability, meanwhile help to address many problems, such as development cost estimation, resource planning and software release scheduling. In fact, Poisson process plays an important role in modeling a great number of phenomena in many fields such as medicine, astronomy, communications and networks. Similar to these cases, among a huge number of SRMs, non-homogeneous Poisson process (NHPP) based SRMs have gained much popularity in actual software testing phases. In software testing, two types of software fault data are commonly collected. TYPE I, software-fault count data, records the number of software faults detected on each testing date, while TYPE II, software-failure time data, records the detection time of each software fault. Since NHPP-based SRM can easily treat both types of data, it has become one of the most widely used tools in evaluating software reliability quantitatively.

In this thesis, we study NHPP-based software reliability modeling framework from two perspectives: i) the development of NHPP-based SRMs, ii) the development of estimation method for NHPP-based SRMs.

The main contribution of this thesis is the second one, of which the purpose is to develop a non-parametric estimation method for quantitative software reliability by means of wavelet technique, where the underlying SRM is described by an NHPP. It is known that the wavelet analysis has a long and rich history in the non-parametric estimation of stochastic point processes as well as the non-linear regression estimation. Nevertheless, it is surprising to know that this attractive approach has not been yet applied to the software reliability assessment. Our approach involves some advantages comparing with the commonly used techniques such as the maximum likelihood estimation: i) the wavelet-based estimation method enables to carry out the time series analysis with high speed and accuracy requirements, ii) the wavelet-based estimation method is classified into a
non-parametric estimation without specifying a parametric form of the software intensity function, which is a unique parameter to govern the probabilistic property of NHPP-based SRMs. In some cases, the computation cost may be almost similar to or greater than an effort on model selection in the parametric SRMs. However, the wavelet-based estimation method does not require solving any optimization problem, so that the implementation of the estimation algorithms is relatively easy. Apart from the development of estimation method, we are also interested in developing new parametric NHPP-based SRMs to describe the stochastic behavior of software fault-detection processes. One idea is to apply general function of probability distribution to the software fault-detection time distribution. In addition, we introduce some recent generations of Weibull-type distributions to represent the underlying software fault-detection time distribution of the NHPP-based SRMs. It is well known that the Weibull distribution plays an active part in reliability applications because of its flexibility in being able to represent various patterns of failure rate functions. We execute a comparative study on the effectiveness of Weibull-type distributions in software reliability modeling. We aim to provide much powerful candidates for evaluating software reliability quantitatively by developing these new NHPP-based SRMs.

This thesis is organized as follows. Chapter 1 is introductory. We give an introduction to the field of software reliability assessment and show the mathematical formulation of software reliability models in Section 1.1, especially concentrating on non-homogenous Poisson process, which is the fundamental stochastic process of this thesis. A brief introduction of wavelet analysis is given in Section 1.2. In Section 1.3, we show the details of all the data sets that are used in this thesis. Chapter 2 is about our first proposal. Each of Sections 2.1 to 2.2 presents ED-NHPP-based SRMs and EVD-NHPP-based SRMs, respectively, where equilibrium distribution (ED) and extreme value distribution (EVD) are studied. Section 2.3 presents Weibull-type NHPP-based SRMs. In Section 2.4, we investigate the effectiveness of the proposed models. We also discuss the model extension from NHPP to binomial process as the future research directions and show some numerical results in Section 2.5. Chapter 3 and Chapter 4 are about our main proposal. They present the non-parametric estimation method based on Haar wavelet and Daubechies wavelet, respectively. The Haar wavelet based algorithm is useful in handling software-fault count data, which can be considered as the realization of a discrete stochastic process. After introducing the fundamental Haar wavelet based denoising procedure in Section 3.1, we define discrete NHPP-based SRM without any loss of generality in Section 3.2. Section 3.3 presents the wavelet shrinkage estimation (WSE) for discrete NHPP-based SRMs. Moreover, in Section 3.4, we combine WSE with the so-called one-stage look-ahead prediction, which can be used to predict the future debugging behavior sequentially. In Section 3.5, we carry out the real project data analysis and illustrate numerical examples to examine the effectiveness of the proposed estimation methods. In Section 3.6, we discuss future research directions and show some numerical results. On the other hand, the Daubechies wavelet based algorithm is developed for analyzing software-failure time data. Section 4.1 is about the Daubechies wavelet based estimation procedure employed in this thesis. Section 4.2 is devoted to the discussions about the problems occurred when applying the Daubechies wavelet estimator to real software-failure time data. In Section 4.3, we conduct experiments on several sets of actual data to establish the credibility and usefulness of the wavelet-based approaches in practice. Finally this thesis is summarized with some comments and future work in Chapter 5.