Specifying the right research question(s) is very important for an SR in order to present detailed and suitable findings of a given domain, and to help identify and scope future research activities [17]. The research questions for this SR were structured with the help of the Population, Intervention, Comparison, Outcome, Context (PICOC) criteria [25]. However, as the objective of this SR was not to find evidence about the comparison of approaches, methods or models; the ‘Comparison’ part of the PICOC was excluded. The population considered is that of software applications; our intervention was methods/techniques to predict maintainability and maintainability predictors; the outcomes of interest were accuracy of maintainability prediction methods/techniques and successful maintainability predictors; and all kinds of empirical studies and study settings encompassed the context. The research questions addressed by our SR are: Question 1: What evidence is there for maintainability forecasting techniques/methods for software applications? Question 2: What forecasting techniques/methods have been used to predict maintainability for software applications? Question 2 (a): What measures have been used for measuring accuracy of the maintainability prediction models/techniques for software applications? Question 2 (b): What were the numeric values obtained for the prediction accuracy of the maintainability prediction models/techniques for software applications? Question 2 (c): Which cross-validation method was used to find the accuracy of the prediction techniques for maintainability in software applications? Question 2 (d): Was there a forecasting technique/method reported to be significantly superior for predicting maintainability for software applications, and if so which one? Question 3: What factors and metrics have been investigated as maintainability predictors for software applications? Which of these are considered successful predictors? Question 3 (a): At what stage of the project are these predictors gathered? Question 3 (b): If previous studies predict various types of maintenance and facets of maintainability, what were the different types of predictors used for each of them? Question 4: How is maintainability understood and measured in the context of software applications and which quality model(s) does it comply with? The first research question is our primary research question whereas the last three and their associated sub-questions are the secondary research questions. 2.2. Search strategy used for primary studies The search terms used in our SR were developed using the following steps [16], [17]: 1. Derive major search terms from the research questions by identifying Population, Intervention, Outcome, and Context. 2. Identify keywords in the relevant papers. 3. Identify alternative spellings and synonyms for the search terms with the help of a thesaurus. 4. Use Boolean OR to construct search strings from the search terms with similar meanings. 5. Use Boolean AND to concatenate the search terms and restrict the research. The resulting search string is given below: (software application OR software OR system OR software system OR application) AND (method OR Process OR technique OR system OR practice OR procedure) AND (factor OR feature) AND (metric OR measure OR measurement) AND (predict OR forecast OR calculate OR prediction OR estimation) AND (maintainability OR analysability OR analyzability OR changeability OR stability OR testability OR compliance OR maintenance OR maintenance process OR enhancive maintenance OR adaptive maintenance OR corrective maintenance OR preventive maintenance) AND (quality model) AND (successful)

Primary search process. The primary search process was directed towards searching online databases, search engines, electronic journals, conference proceedings, and grey literature using the derived search string. The resources used were chosen because (1) they were used by previous SRs in software engineering, (2) they were suggested by the University of Auckland Library website as relevant to the subject, and (3) the venues contained publications that were relevant to our areas of interest. Only the online databases that publish peer-reviewed articles were considered. As a result, Agile Alliance was removed from the list after a confirmation from Agile Alliance that it contains publications that are not peer reviewed. Also, IEEE Computer Society Digital Library and Lecture Notes in Computer Science were removed from the list after confirmation from the respective publishers that these provide the same results as IEEE Xplore and SpringerLink, respectively. In order to make the search process comprehensive, our searches included literature published from (and inclusive of) year 1985 to Nov. 11th, 2008. The year 1985 was chosen as a preliminary search found no results previous to it. The original comprehensive search string given above did not yield any results. To address this, we broadened the search criteria and tailored the strings to each database on the recommendations of the subject librarian of the University of Auckland. Using broader search criteria on full text searches lead to thousands of results, most of which were irrelevant. In the end, searches were restricted to titles and abstracts only. Table 1 summarizes the results of the searches and the sources on which they were ran.