Estimating Web Service interface quality through conventional object-oriented metrics[*]

José Luis Ordiales Coscia, Marco Crasso, Cristian Mateos, Alejandro Zunino
ISISTAN Research Institute - UNICEN University,
Tandil, Buenos Aires, Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
e-mails [mcrasso|cmateos|azunino]@conicet.gov.ar

Abstract

Historically, software engineers have conceived metric suites as valuable tools to estimate the quality of their software artifacts. Recently, a fresh computing paradigm called Service-Oriented Computing (SOC) has emerged at the crossing of massively distributed and heterogeneous software. This paper presents a statistical correlation analysis showing that classic software engineering metrics can be used to predict the most relevant quality attributes of WSDL documents, the essential software artifact when materializing this novel computing paradigm with Web-based technologies. For the experiments, two recent WSDL-level metrics catalogs and 154 real world WSDL documents have been employed.

Keywords: SERVICE-ORIENTED COMPUTING, WEB SERVICES, CODE-FIRST, WEB SERVICE QUALITY, OBJECT-ORIENTED METRICS, EARLY DETECTION

1 Introduction

The Service-Oriented Computing (SOC) paradigm has been steadily gaining territory in the software industry. With SOC, the composition of loosely coupled pieces of software, called *services*, drives software construction. A key to SOC is that services may be provided by third-parties who only expose services interfaces to the outer world.

This basic idea has been present in the software industry for a long time. However, the advances in distributed system technologies nowadays encourage engineers to implement the SOC paradigm in environments with higher levels of distribution and heterogeneity. For instance, broadband and ubiquitous connections enable to reach the Internet from everywhere and at every time, enabling a global scale marketplace of software services. In such a marketplace, providers may offer their services interfaces and consumers may invoke them regardless of geographical aspects, using the current Web infrastructure as the communication channel. When services are implemented using standard Web languages and protocols –which is the most common scenario– they are called *Web Services*. Nowadays, Web Services are the leading character in diverse contexts. For example, they are the chosen technology usually employed when migrating legacy systems [1] or the technological protocol stack used when accessing remote information from smartphones [2].

Like programs or in general any other software artifact, service interface descriptions have a size, complexity and quality, all of which can be measured [3]. Previous research has emphasized on the importance of services interfaces and more specifically their non-functional concerns. Particularly, the work of [4] identifies common bad practices found in services interfaces, which impact on the understandability and discoverability of described services. In this context, understandability is the ability of a service interface description of being self-explanatory, i.e., a software engineer can effectively reason about a service purpose just by looking at its associated interface description. Discoverability, on the other hand, refers to the ability of a service of being easily retrieved, from a registry or repository, based on a partial description of its functionality, i.e., a query. At the same time, in [3] the author describes a suite of metrics to assess the complexity and quality of services interfaces.

In an attempt to study the relationships between understandability and discoverability problems associated with services interfaces, such as [4], and service implementation metrics, such as [5, 6], in [7] a statistical correlation analysis has been reported. The rationale behind the study is that in practice services interfaces are not build by hand, but instead they are automatically generated by mapping programming languages constructors onto service interface descriptions expressed in Web Service Definition Language (WSDL). WSDL is an XML-based format for describing

[*]This paper is an extended version of the paper presented at the XV Ibero-American Conference on Software Engineering - CIbSE 2012

a service as a set of operations, whose invocation is based on message exchange. For services implemented in Java, this mapping is usually achieved by tools such as Axis' Java2WSDL, Java2WS, EasyWSDL, and WSProvide.

As a complement, in this paper we study the feasibility of obtaining less *complex*, of the highest possible *quality*, and more *maintainable* services by exploiting Object-Oriented metrics (OO) values from the code implementing services. Similar to [7], the approach behind this work relies on employing these metrics as early indicators to warn software engineers about the possibility of obtaining services interfaces with high complexity, low quality or low maintainability at development time. For the purposes of this paper, the terms complexity, quality and maintainability are used as follows. Complexity refers to lingual, data, and structural aspects of WSDL documents. Furthermore, the term quality refers to Modularity, Adaptability, Reusability, Testability, Portability, and Conformity attributes. Each quality attribute will be explained later in this paper. Finally, maintainability estimates the effort required to understand the messages that are responsible for representing services operations inputs/outputs.

Interestingly, we have found that there is a statistical significant, somewhat high correlation between several traditional (source code-level) OO metrics and the catalog of (WSDL-level) service metrics described in [3] and [8]. To date, these are the most comprehensive catalog of metrics for measuring complexity/quality and maintainability, respectively, of WSDL interfaces. All in all, we argue that based on these indicators, software engineers could consider applying simple early code refactorings to avoid obtaining services interfaces with the mentioned problems.

It is worth noting that although our approach to correlation does not depend on the programming language in which services are implemented, we limit the scope of our research to Java, which is very popular in back-end and hence service development. To evaluate our approach, we performed experiments by employing a data-set of real services, and one of the most popular Java-to-WSDL generation tool, namely Java2WSDL[1].

The rest of the paper is structured as follows. Section 2 gives some background necessary to understand the goals and results of the study presented in this paper. Section 3 surveys relevant related works. Section 4 presents detailed analytical experiments that evidence the correlation of OO metrics with the Web Service metrics both proposed in [3] and [8]. Finally, Section 5 concludes the paper.

2 Background

WSDL allows providers to describe a service from two main angles, namely what it does (its functionality) and how to invoke it. Following the version 1.1 of the WSDL specification, the former part reveals the service interface that is offered to potential consumers. The latter part specifies technological aspects, such as transport protocols and network addresses. Consumers use the functional descriptions to match third-party services against their needs, and the technological details to invoke the selected service. With WSDL, service functionality is described as a *port-type* $W = \{O_0(I_0, R_0), .., O_N(I_N, R_N)\}$, which arranges different operations O_i that exchange input and return messages I_i and R_i, respectively. Port-types, operations and messages must be labeled with unique names. Optionally, these WSDL elements might contain some documentation in the form of comments.

Messages consist of *parts* that transport data between providers and consumers of services, and vice-versa. Exchanged data is represented by using XML but according to specific data-type definitions in XML Schema Definition (XSD), a language to define the structure of an XML document. XSD offers constructors for defining simple types (e.g., integer and string), restrictions, and both encapsulation and extension mechanisms to define complex constructs. XSD code might be included in a WSDL document using the *types* element, but alternatively it might be put into a separate file and imported from the WSDL document or external WSDL documents in order to achieve type reuse.

A requirement inherent to manually creating and manipulating WSDL and XSD definitions is that services are built in a *contract-first* manner, a methodology that encourages designers to first specify the WSDL-based interface of a service and then supply an implementation for it. However, the most used approach to build Web Services by the industry is *code-first*, which means that one first implements a service and then generates the corresponding service interface by automatically deriving the WSDL document from the implemented code. This means that WSDL documents are not directly created by humans but are instead automatically derived via language-dependent tools. Such a tool performs a mapping T [7], formally:

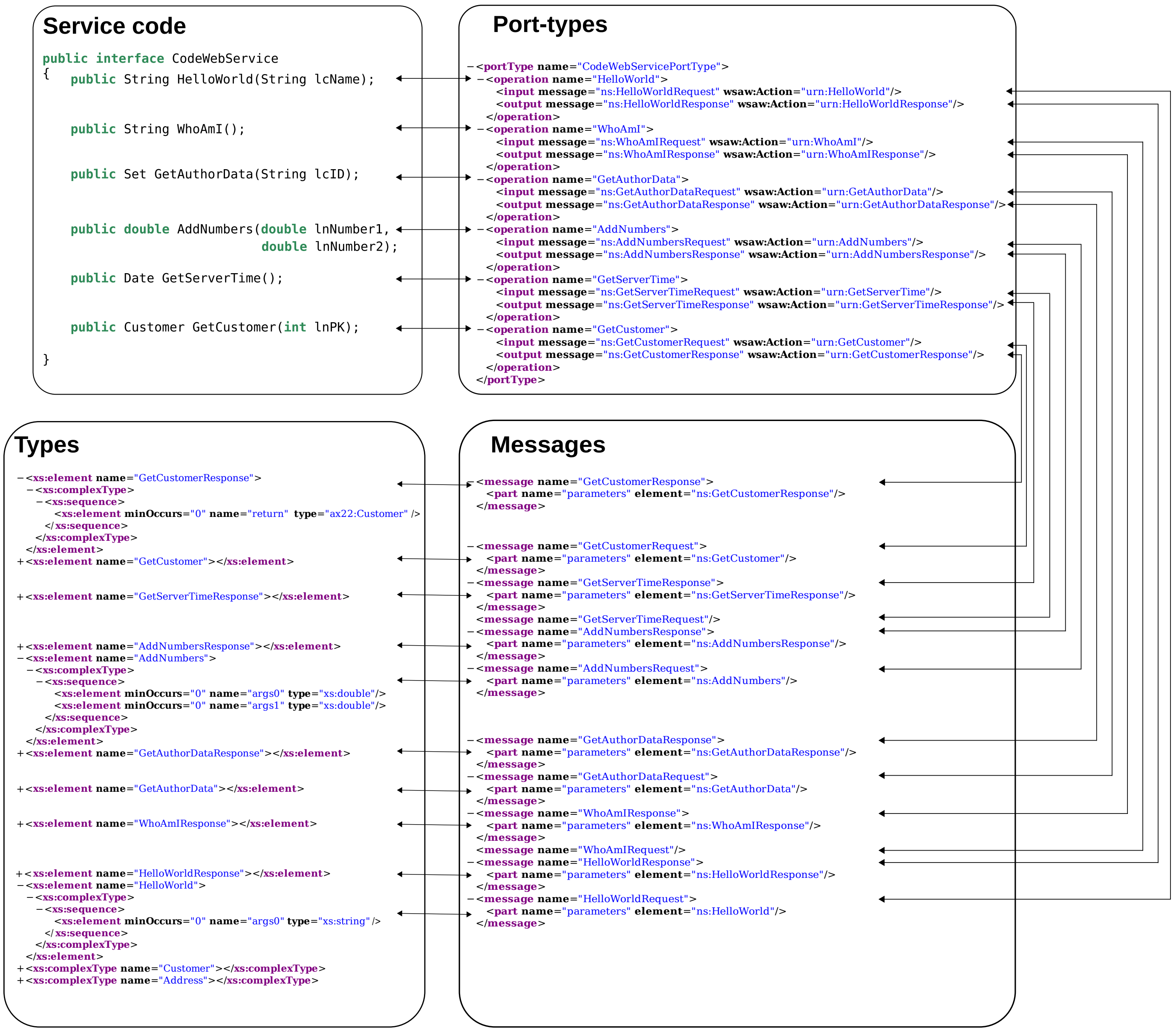
$$T : C \rightarrow W, \tag{1}$$

Mapping T from $C = \{M(I_0, R_0), .., M_N(I_N, R_N)\}$ to $W = \{O_0(I_0, R_0), .., O_N(I_N, R_N)\}$, being C the front-end class implementing a service and W the WSDL document describing the service, generates a WSDL document containing a port-type for the service implementation class, having as many operations O as public methods M are defined in the class. Moreover, each operation of W is associated with one input message I and another return message R, while each message conveys an XSD type that stands for the parameters of the corresponding class method. Tools like WSDL.exe, Java2WSDL, and gSOAP [9] are based on a mapping T for generating WSDL documents from C#, Java and C++,

[1] `http://ws.apache.org/axis/java`

respectively, though each tool implements T in a particular manner mostly because of the different characteristics of the involved programming languages.

Figure 1 shows the generation of a WSDL document using Java2WSDL. It can be noted that the mapping T maps each public method from the service code to an *operation* containing two *messages* in the WSDL document and these, in turn, are associated with an XSD type containing the parameters of that operation. As shown in [7], depending on the tool used, however, some minor differences between the generated WSDL documents may arise. For instance, for the same service Java2WSDL generates only one port-type with all the operations of the Web Service, whereas WSDL.exe generates three port-types (one for each transport protocol). As we mentioned before, these differences are a result of the implementation each tool uses when applying the mapping to the service code.



```
public interface CodeWebService
{   public String HelloWorld(String lcName);

    public String WhoAmI();

    public Set GetAuthorData(String lcID);

    public double AddNumbers(double lnNumber1,
                             double lnNumber2);

    public Date GetServerTime();

    public Customer GetCustomer(int lnPK);

}
```

```
- <portType name="CodeWebServicePortType">
  - <operation name="HelloWorld">
      <input message="ns:HelloWorldRequest" wsaw:Action="urn:HelloWorld"/>
      <output message="ns:HelloWorldResponse" wsaw:Action="urn:HelloWorldResponse"/>
    </operation>
  - <operation name="WhoAmI">
      <input message="ns:WhoAmIRequest" wsaw:Action="urn:WhoAmI"/>
      <output message="ns:WhoAmIResponse" wsaw:Action="urn:WhoAmIResponse"/>
    </operation>
  - <operation name="GetAuthorData">
      <input message="ns:GetAuthorDataRequest" wsaw:Action="urn:GetAuthorData"/>
      <output message="ns:GetAuthorDataResponse" wsaw:Action="urn:GetAuthorDataResponse"/>
    </operation>
  - <operation name="AddNumbers">
      <input message="ns:AddNumbersRequest" wsaw:Action="urn:AddNumbers"/>
      <output message="ns:AddNumbersResponse" wsaw:Action="urn:AddNumbersResponse"/>
    </operation>
  - <operation name="GetServerTime">
      <input message="ns:GetServerTimeRequest" wsaw:Action="urn:GetServerTime"/>
      <output message="ns:GetServerTimeResponse" wsaw:Action="urn:GetServerTimeResponse"/>
    </operation>
  - <operation name="GetCustomer">
      <input message="ns:GetCustomerRequest" wsaw:Action="urn:GetCustomer"/>
      <output message="ns:GetCustomerResponse" wsaw:Action="urn:GetCustomerResponse"/>
    </operation>
  </portType>
```

```
- <xs:element name="GetCustomerResponse">
  - <xs:complexType>
    - <xs:sequence>
        <xs:element minOccurs="0" name="return" type="ax22:Customer" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
+ <xs:element name="GetCustomer"></xs:element>

+ <xs:element name="GetServerTimeResponse"></xs:element>

+ <xs:element name="AddNumbersResponse"></xs:element>
- <xs:element name="AddNumbers">
  - <xs:complexType>
    - <xs:sequence>
        <xs:element minOccurs="0" name="args0" type="xs:double"/>
        <xs:element minOccurs="0" name="args1" type="xs:double"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
+ <xs:element name="GetAuthorDataResponse"></xs:element>

+ <xs:element name="GetAuthorData"></xs:element>

+ <xs:element name="WhoAmIResponse"></xs:element>

+ <xs:element name="HelloWorldResponse"></xs:element>
- <xs:element name="HelloWorld">
  - <xs:complexType>
    - <xs:sequence>
        <xs:element minOccurs="0" name="args0" type="xs:string" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
+ <xs:complexType name="Customer"></xs:complexType>
+ <xs:complexType name="Address"></xs:complexType>
```

```
- <message name="GetCustomerResponse">
    <part name="parameters" element="ns:GetCustomerResponse"/>
  </message>

- <message name="GetCustomerRequest">
    <part name="parameters" element="ns:GetCustomer"/>
  </message>
- <message name="GetServerTimeResponse">
    <part name="parameters" element="ns:GetServerTimeResponse"/>
  </message>
  <message name="GetServerTimeRequest"/>
- <message name="AddNumbersResponse">
    <part name="parameters" element="ns:AddNumbersResponse"/>
  </message>
- <message name="AddNumbersRequest">
    <part name="parameters" element="ns:AddNumbers"/>
  </message>

- <message name="GetAuthorDataResponse">
    <part name="parameters" element="ns:GetAuthorDataResponse"/>
  </message>
- <message name="GetAuthorDataRequest">
    <part name="parameters" element="ns:GetAuthorData"/>
  </message>
- <message name="WhoAmIResponse">
    <part name="parameters" element="ns:WhoAmIResponse"/>
  </message>
  <message name="WhoAmIRequest"/>
- <message name="HelloWorldResponse">
    <part name="parameters" element="ns:HelloWorldResponse"/>
  </message>
- <message name="HelloWorldRequest">
    <part name="parameters" element="ns:HelloWorld"/>
  </message>
```

Figure 1: WSDL generation in Java

There are metrics for assessing WSDL documents size, cohesion, understandability, discoverability, complexity and quality, just to name a few characteristics. Precisely, the fact underpinning our study is that WSDL metrics in the general sense are strongly associated with API design attributes [7, 4]. These latter have been throughly studied by the software engineering community and as a result suites of OO class-level metrics exist, such as the Chindamber and Kemerer's metric catalog [5]. Consequently, these metrics tell providers about how a service implementation conforms to specific design attributes. For example, the LCOM (Lack of Cohesion Methods) metric provides a mean to measure how well the methods of a class are semantically related to each other, or the *cohesion* design attribute.

An interesting approach is then to assess whether a desired design attribute as measured by OO metrics is maintained after WSDL generation as measured by one or more WSDL-level metrics (e.g., [3, 8]) suitable for the attribute.

As a corollary, by using well-known software engineering metrics on a service code C, a provider might have an estimation of how the resulting WSDL document W will be like in terms of understandability, discoverability, complexity and so on, since a known mapping T relates C with W. If indeed such code metric/anti-pattern relationships exist, then it would be possible to determine a wider range of metric values for C so that T generates W without undesirable WSDL-level metric values in the best case.

3 Related work

It can be found in the literature that researchers have been using traditional OO metrics at development time of conventional software to predict the number of defects or other quality attributes [10, 11, 12]. With respect to Web Services, though there has been a substantial amount of research concerning improving services interfaces quality [3, 4, 8], the approach to associate implementation metrics with services interfaces has been recently and exclusively explored in [7].

In [7] the authors collect a publicly available data-set of Web Services projects, which by itself has been a valuable contribution to the field, and analyze the statistical relationships between metrics that were taken from services implementations and metrics taken from corresponding WSDL documents. In particular, the authors employed a sub-set of the WSDL metrics presented in [4], as it subsumes those research works focused on obtaining more legible, clear and discoverable WSDL documents [13, 14, 15]. Therefore, [7] studies the relationships between service implementation metrics and the discoverability of target services interfaces in WSDL. Complementary, this paper presents the statistical relationships between Object-Oriented metrics (OO) values from the code implementing services and the WSDL metrics values of two metrics suites, namely Sneed's suite [3] and Basci & Misra's suite [8].

3.1 Harry M. Sneed's metrics

The suite of metrics proposed by H. Sneed [3] broadly comprises metrics to assess the complexity of services interfaces and to determine their size for estimating testing effort. The suite consists of different kinds of metrics, which range from common size measurements like lines of code and number of statements, to metrics for measuring the complexity and quality of Web Services. All the involved metrics can be statically computed from a service interface in WSDL, since the metric suite is purely based on WSDL schema elements occurrences. For the purposes of this work, we focus on two groups of metrics, namely 6 complexity metrics: Interface Data Complexity (Chapin's metric [16]), Interface Relation Complexity, Interface Format Complexity, Interface Structure Complexity (Henry et al.'s metric [17]), Data Flow Complexity (Elshof's Metric), Language Complexity (Halstead's Metric), and 6 quality metrics: Modularity, Adaptability, Reusability, Testability, Portability, and Conformity. It is worth noting that the terms used to denote quality metrics are known in the literature as properties, factors or characteristics [18]. However, we are here strictly following the terminology of Harry S. Sneed, which proposed these quality metrics.

Sneed defines the complexity of a service interface as the median value of its lingual complexity and its structural complexity. By basing on Halstead's metric [19], Sneed defines the lingual complexity of a service interface as $1 - \frac{statement\ types}{statement\ occurrence}$. WSDL schema operators and operands are the types of a statement that can be found in a service interface. At the same time, Sneed defines structural complexity as $1 - \frac{Entities}{Relationships}$, in which entities are the instances of the data-types, messages, operations, parameters, bindings and ports defined in the schema and occurring in the target interface, and a relationship (vertical or horizontal) refers to compositions and cross-references, respectively. An example of vertical relationship is when a complex data-type definition groups XSD built-in data-types. Another example is when a message groups several parts, or a port-type that arranges two or more operations. Instead, horizontal relationships in a WSDL schema are the cross references from one schema to another and from one data-type to another.

Regarding the service interface quality aspects that can be statically measured from a WSDL document, Sneed defines an intuitive metric suite purely based on WSDL schema elements occurrences. The notation used by Sneed for his metrics refers to group data-types to those XSD elements that are composed of smaller elements, e.g., a complex element with a sequence of built-in elements like int and string. Built-in elements are called *elementary* ones. User data-types are those elements originally defined in the WSDL types section. Table 1 summarizes these metrics.

All metrics results, except for Conformity, are expressed as a value on a scale of 0 to 1. For complexity metrics, 0 to 0.4 indicates low complexity, 0.4 to 0.6 indicates average complexity, 0.6 to 0.8 indicates high complexity and over 0.8 indicates that the code is not well designed. Instead, for quality metrics 0 to 0.2 indicates no quality, 0.2 to 0.4 indicates low quality, 0.4 to 0.6 indicates median quality, 0.6 to 0.8 indicates high quality, and 0.8 to 1.0 indicates very high quality [20]. Additionally, we employed two more metrics that represent the average of either previous groups, namely Average Interface Complexity and Average Interface Quality.

The Conformity metric result is a list of non-conformed rules. The higher the number of non-conformed rules the worse the quality of the WSDL document. Set of rules employed for the Conformity metric has been extrapolated from the work of [21], and consists of 12 rules, namely Adherence to the current standard, Use of prescribed name

Table 1: Service interface quality metrics (from [3]).

Metric	Formally	Rationale
Usability	$(\frac{user\,datatypes}{all\,datatypes}) * (\frac{group\,datatypes}{elementary\,datatypes})$	The smaller the data units, the easier it is to adapt them to new requirements.
Reusability	$1 - \frac{external\,referenced\,datatypes}{all\,datatypes}$	The less dependent the software is on the other software around it, i.e., the fewer its external relationships, the higher is its reusability.
Testability	$1 - operations + \frac{message\,parts}{datatypes}$	The least the operations there are to test and the less the number of their parameters, the less will be the required test effort.
Modularity	$\frac{cohesion+coupling}{2}$	Modularity as a whole is the arithmetic mean of cohesion and coupling, i.e., achieving maximum cohesion and minimum coupling.
Cohesion	$\frac{operations}{datatypes}$	The less the number of data-types per operation, the higher the cohesion.
Coupling	$1 - \frac{ports}{operations}$	The number of operations invoked per port, which corresponds to the fan–out of a service request.
Conformity	$\frac{weighted\,rule\,violations}{WSDL\,statements}$	The least the broken rules for developing maintainable services in a WSDL document, the best its conformity to the rules.

spaces, Use of in source documentation, Avoidance of any data type, Adherence to the naming convention, Hiding of elementary data types, Restricting the size of data groups, Limiting interface width, Enforcing the first normal form, Minimizing the number of requests, Not exceeding a give size limit, and Limiting the use of links.

3.2 Dilek Basci & Sanjay Misra's metrics

In the same direction, this paper analyzes whether there are relations between service implementation metrics and the suite of metrics proposed by Basci & Misra [8], which comprises novel metrics for measuring Web Service maintainability by computing the complexity of the information exchanged by Web Services. These metrics can be statically computed from a service interface in WSDL, since this metric suite is purely based on WSDL and XSD schema elements occurrences.

Basci & Misra [8] define the *data complexity* of a Web Service as "the complexity of data flowed to and from the interfaces of a Web service and can be characterized by an effort required to understand the structures of the messages that are responsible for exchanging and conveying the data". The definition of the Data Weight (DW) metric is based on the above, and computes the complexity of the data-types conveyed in services messages. To the sake of brevity, we will refer to the complexity of a message $C(m)$ as an indicator of the effort required to understand, extend, adapt, and test m, by basing on its structure. $C(m)$ counts how many elements, complex types, restrictions and simple types are exchanged by messages parts, as it is deeply explained in [8]. Formally:

$$DW(wsdl) = \sum_{i=1}^{n_m} C(m_i) \tag{2}$$

where n_m is the number of messages that the WSDL document exchanges. For the purposes of this paper, we have assumed n_m to consider only those messages that are linked to an offered operation of the WSDL document, thus it does not take into account dangling messages. The DM metric always returns a positive integer. The bigger the DM of a WSDL document, the more complex its operations messages are.

The Distinct Message Ratio (DMR) metric complements DW by attenuating the impact of having similar-structured messages within a WSDL document. As the number of similar-structured messages increases the complexity of a WSDL document decreases, since it is easier to understand similar-structured messages than that of various-structured ones as a result of gained familiarity with repetitive messages [8]. Formally:

$$DMR(wsdl) = \frac{DMC(wsdl)}{n_m} \tag{3}$$

where the Distinct Message Count (DMC) metric can be defined as the number of distinct-structured messages represented by $[C(m), n_{args}]$ pair, i.e., the complexity value $C(m)$ and total number of arguments n_{args} that the message contains [8]. The DMR metric always returns a number in the range of [0,1], where 0 means that all defined messages are similar-structured, and 1 means that messages variety increases.

The Message Entropy (ME) metric exploits the probability of similar-structured messages to occur within a given WSDL document. Compared with the DMR metric, ME also bases on the fact that repetition of the same messages makes a developer more familiar with the WSDL document and results in ease of maintainability, but ME provides better differentiation among WSDL documents in terms of complexity. Formally:

$$\begin{aligned} ME(wsdl) &= -\sum_{i=1}^{DMC(wsdl)} P(m_i) * log_2 P(m_i) \\ P(m_i) &= \frac{nom_i}{n_m} \end{aligned} \tag{4}$$

where nom_i is the number of occurrences of the i^{th} message, and in turn $P(m_i)$ represents the probability that such a message occurs within the given WSDL document. The ME metric outputs values greater or equal than zero. A low ME value shows that the messages are consistent in structure, which means that data complexity of a WSDL document is lower than that of the others having equal DMR values.

The Message Repetition Scale (MRS) metric analyses variety in structures of WSDL documents. By considering frequencies of $[C(m), n_{args}]$ pairs, MRS measures the consistency of messages as follows:

$$MRS(wsdl) = \sum_{i=1}^{DMC(wsdl)} \frac{nom_i^2}{n_m} \tag{5}$$

The possible values for MRS are in the range $1 \leq MRS \leq n_m$. When comparing two or more WSDL documents, a higher MRS and lower ME show that the developer makes less effort to understand the messages structures owing to the repetition of similar-structured messages.

The next section presents the performed statistical analysis.

4 Statistical analysis and experiments

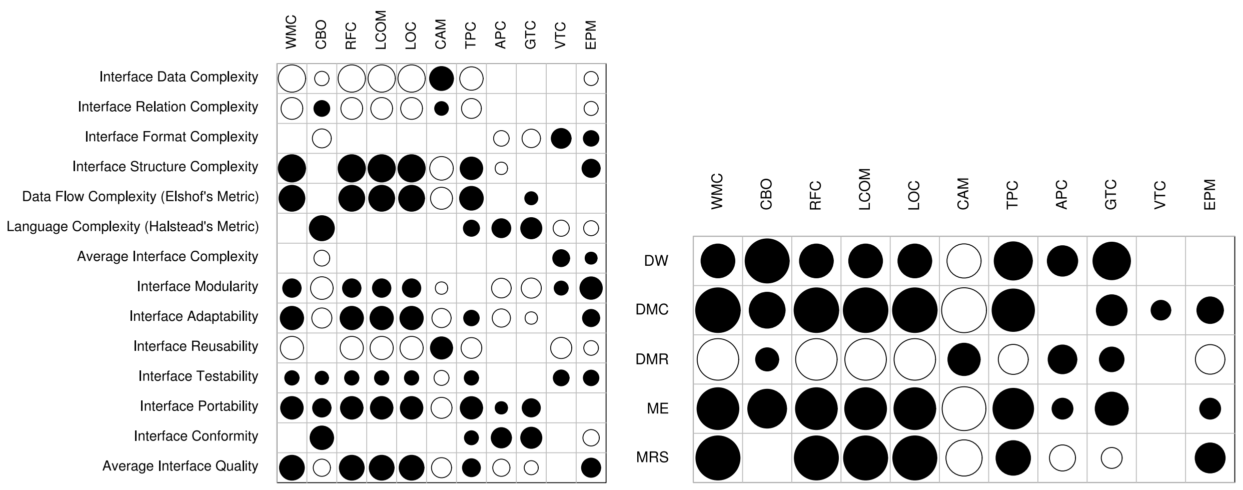
We used regression and correlation methods to analyze whether OO metrics taken on service implementations are correlated with metrics from the associated WSDL documents or not. Broadly, we gathered OO metrics from open source Web Services, calculating Sneed's complexity and quality metrics and Basci & Misra's maintainability metrics from WSDL documents, and analyzing the correlation among all pairs of metrics. To perform the analysis, we employed the newest version available of the data-set described in [7] at the time of writing this article. The criteria employed for conforming the experimental data-set consisted of 3 basis: (1) projects should be open source, because we needed to assess metrics from service implementations code; (2) projects source code should be in Java; (3) projects should be self-contained, in the sense that all the resources needed for compiling a project were available. Accordingly, this data-set consists of 154 WSDL documents from open source projects, which were collected via the Merobase component filter, the Exemplar engine and the Google Code Web site. It is worth noting that the data-set used in the experiments is available upon request. All projects are written in Java, and each project offers at least one Axis' Java2WSDL Web Service description. The data-set is self-contained in the sense that for each service, its implementation code and dependency libraries needed for compiling and generating WSDL documents are in the data-set. All in all, the generated data-set provided the means to perform a significant evaluation in the sense that the different Web Service implementations came from real-life software engineers.

We used 11 metrics for measuring services implementations, which played the role of independent variables. WMC, CBO, RFC, and LCOM have been selected from the work of Chindamber and Kemerer [5]. The WMC (Weighted Methods Per Class) metric counts the methods of a class. CBO (Coupling Between Objects) counts how many methods or instance variables defined by other classes are accessed by a given class. RFC (Response for Class) counts the methods that can potentially be executed in response to a message received by an object of a given class. LCOM (Lack of Cohesion Methods) provides a mean to measure how well the methods of a class are related to each other. From the work of Bansiya and Davis [6] we picked CAM (Cohesion Among Methods of Class) metric. CAM computes the relatedness among methods based upon the parameter list of these methods. Additionally, we used a number of ad-hoc measures we thought could be related to the WSDL metrics, namely TPC (Total Parameter Count), APC (Average Parameter Count), ATC (Abstract Type Count), VTC (Void Type Count), and EPM (Empty Parameters Methods). The last employed metric was the well-known lines of code (LOC) metric.

On the other hand, the dependent variables were the metrics at the WSDL-level, i.e., those proposed in [3] and [8]. Table 2 statistically describes both independent and dependent variables. As shown in Std. Dev. column all dependent variables values were concentrated around the central tendency (i.e., Mean column), while this phenomenon persisted for independent variables values with some exceptions namely LCOM, WMC, and TPC. The Skewness column characterizes the location for metrics values distribution, in particular these results showed that most distributions are skewed to the right since their means were higher than their median. Furthermore, the skewness of three metrics was very near to 0, which is the skewness of the Normal distribution. The Kurtosis column shows that most metrics values

Table 2: Descriptive statistics

Metric	**Min**	**Max**	**Mean**	**Std. Dev.**	**Kurtosis**	**Skewness**
Interface Data Complexity	0.10	0.90	0.82	0.09	27.91	-4.53
Interface Relation Complexity	0.10	0.90	0.88	0.13	34.52	-6.00
Interface Format Complexity	0.10	0.90	0.23	0.14	14.18	3.34
Interface Structure Complexity	0.10	0.75	0.13	0.08	28.21	5.10
Data Flow Complexity (Elshof's Metric)	0.10	0.90	0.23	0.07	51.04	6.18
Language Complexity (Halstead's Metric)	0.16	0.87	0.19	0.06	81.00	8.02
Average Interface Complexity	0.38	0.61	0.41	0.02	40.55	5.26
Interface Modularity	0.10	0.82	0.14	0.12	27.72	5.26
Interface Adaptability	0.10	0.90	0.60	0.30	-1.23	-0.49
Interface Reusability	0.10	0.90	0.31	0.13	9.97	1.93
Interface Testability	0.10	0.90	0.12	0.13	33.94	5.93
Interface Portability	0.33	0.50	0.49	0.03	33.70	-5.90
Interface Conformity	0.79	1.00	0.92	0.03	5.37	-0.41
Average Interface Quality	0.33	0.68	0.43	0.06	4.61	1.34
DW	2.00	1148.00	108.53	165.12	14.63	3.42
DMC	1.00	21.00	3.65	2.74	10.81	2.48
DMR	0.07	1.00	0.61	0.29	-1.25	0.14
ME	0.00	3.25	1.38	0.82	-0.71	-0.06
MRS	1.00	47.18	3.52	5.62	37.44	5.63
WMC	1.00	97.00	5.73	11.14	35.87	5.47
CBO	0.00	27.00	2.03	2.92	37.60	5.05
RFC	1.00	97.00	5.73	11.14	35.87	5.47
LCOM	0.00	4656.00	75.21	427.43	90.28	9.00
LOC	1.00	97.00	5.73	11.14	35.87	5.47
CAM	0.13	1.00	0.78	0.23	-0.86	-0.61
TPC	0.00	228.00	10.92	24.23	46.77	6.16
APC	0.00	17.00	2.04	1.83	30.61	4.45
GTC	0.00	20.00	1.09	2.26	34.15	5.01
VTC	0.00	25.00	1.05	3.54	26.65	4.96
EPM	0.00	11.00	0.57	1.64	16.52	3.85



(a) Sneed's WSDL metrics

(b) Basci & Misra's WSDL metrics

Figure 2: Graphical representation of the correlation between OO and WSDL metrics

did not tend to have a flat top near the mean rather than a distinct peak near the mean, decline rather rapidly, and have heavy tails. These results and the correlation analysis were obtained by using Apache's Commons Math library[2].

It is worth noting that all the employed OO metrics have been automatically gathered by using an extended version of the *ckjm* [22] tool. At the same time, we use Sneed's SoftAUDIT tool [3] to automate the recollection of the WSDL metrics proposed in [3]. SoftAUDIT receives a given WSDL document as input, and returns a metric report in a structured text file. Moreover, to compute DW, DMC, DMR, ME and MRS metrics [8] we used a software library of our own[3] that automates the recollection of these WSDL metrics.

The next section presents the correlation analysis results.

4.1 Correlation analysis between OO and WSDL metrics

We used the Spearman's rank correlation coefficient in order to establish the existing relations between the two kind of variables of our model, i.e., the OO metrics (independent variables) and the WSDL metrics (dependent variables). Table 3 shows the correlation between the employed OO and WSDL metrics.

The cell values in bold are those coefficients which are statistically significant at the 5% level, i.e., p-value < 0.05, which is a common choice when performing statistical studies [23]. The sign of the correlation coefficients defines the direction of the relationship, i.e., positive or negative. A positive relation means that when the independent variable grows, the dependent variable grows too, and when the independent variable falls the dependent goes down as well. Instead, a negative relation means that when independent variables grow, the dependent metrics fall, and vice versa. The absolute value, or correlation factor, indicates the intensiveness of the relation regardless of its sign.

Additionally, for the sake of readability, we have employed a different approach to depict the correlation matrix, which is shown in Figure 2. In the Figure, blank cells stand for not statistically significant correlations, whereas cells with circles represent correlation factors at the 5% level. The diameter of a circle represents a correlation factor, i.e., the bigger the correlation factor the bigger the diameter. The color of a circle stands for the correlation sign, being black used for positive correlations and white for negative ones.

4.2 Exploiting correlation results for improving WSDL documents

From both Table 3 and Figure 2, it can be observed that there is a somewhat high statistical correlation between a sub-set of the analyzed OO metrics and some of the Sneed's and Basci & Misra's WSDL metrics. Although correlation between two variables does not necessarily imply that one causes the other, for the purposes of this paper we may assume that this fact allows us to determine which OO metrics could be somehow "controlled" by a software engineer attempting to assure the quality of his/her target WSDL documents. However, for the scope of this paper we will focus on determining a minimalist sub-set of OO metrics in each case, because determining the best set of metrics would deserve a deeper analysis.

By basing on the previous assumption and since all OO metrics correlate to at least one WSDL metric, one could argue that every independent variable should be *controllable*. However, the study presented in [7] shows that some OO metrics are not statistically independent among them and, therefore, capture redundant information. In other words,

[2]Apache's Commons Math library, `http://commons.apache.org/math`

[3]WSDL maintainability metric suite, `http://code.google.com/p/wsdl-metrics-soc-course/`

Table 3: Correlation between OO metrics and WSDL ones

Metric	**WMC**	**CBO**	**RFC**	**LCOM**	**LOC**	**CAM**	**TPC**	**APC**	**GTC**	**VTC**	**EPM**
Interface Data Complexity	**-0.83**	**-0.23**	**-0.83**	**-0.83**	**-0.83**	**0.64**	**-0.61**	0.12	-0.12	-0.01	**-0.22**
Interface Relation Complexity	**-0.53**	**0.28**	**-0.53**	**-0.53**	**-0.53**	**0.21**	**-0.44**	0.01	0.12	-0.12	**-0.22**
Interface Format Complexity	0.07	**-0.40**	0.07	0.07	0.07	-0.00	-0.09	**-0.27**	**-0.38**	**0.44**	**0.27**
Interface Structure Complexity	**0.83**	0.01	**0.83**	**0.83**	**0.83**	**-0.62**	**0.57**	**-0.17**	-0.04	0.15	**0.37**
Data Flow Complexity (Elshof's metric)	**0.75**	0.08	**0.75**	**0.75**	**0.75**	**-0.54**	**0.63**	0.04	**0.19**	-0.01	0.15
Language Complexity (Halstead's metric)	0.09	**0.70**	0.09	0.09	0.09	-0.12	**0.29**	**0.40**	**0.50**	**-0.29**	**-0.25**
Average Interface Complexity	-0.03	**-0.28**	-0.03	-0.03	-0.03	0.06	-0.07	-0.07	-0.15	**0.32**	**0.16**
Interface Modularity	**0.38**	**-0.57**	**0.38**	**0.38**	**0.38**	**-0.17**	0.03	**-0.43**	**-0.45**	**0.22**	**0.56**
Interface Adaptability	**0.61**	**-0.45**	**0.61**	**0.61**	**0.61**	**-0.41**	**0.27**	**-0.37**	**-0.17**	0.15	**0.33**
Interface Reusability	**-0.60**	0.12	**-0.60**	**-0.60**	**-0.60**	**0.55**	**-0.48**	0.04	-0.16	**-0.50**	**-0.24**
Interface Testability	**0.23**	**0.20**	**0.23**	**0.23**	**0.23**	**-0.25**	**0.23**	0.09	-0.09	**0.28**	**0.27**
Interface Portability	**0.57**	**0.38**	**0.57**	**0.57**	**0.57**	**-0.49**	**0.56**	**0.17**	**0.37**	-0.13	0.05
Interface Conformity	-0.05	**0.62**	-0.05	-0.05	-0.05	0.05	**0.22**	**0.46**	**0.50**	-0.12	**-0.30**
Average Interface Quality	**0.69**	**-0.33**	**0.69**	**0.69**	**0.69**	**-0.45**	**0.36**	**-0.33**	**-0.22**	0.07	**0.41**
DW	**0.47**	**0.80**	**0.47**	**0.47**	**0.47**	**-0.48**	**0.60**	**0.38**	**0.58**	0.05	-0.03
DMC	**0.82**	**0.53**	**0.82**	**0.82**	**0.82**	**-0.83**	**0.74**	0.14	**0.39**	**0.16**	**0.29**
DMR	**-0.71**	**0.22**	**-0.71**	**-0.71**	**-0.71**	**0.43**	**-0.37**	**0.34**	**0.25**	-0.10	**-0.36**
ME	**0.72**	**0.62**	**0.72**	**0.72**	**0.72**	**-0.79**	**0.68**	**0.18**	**0.45**	0.13	**0.18**
MRS	**0.80**	-0.13	**0.80**	**0.80**	**0.80**	**-0.54**	**0.49**	**-0.28**	**-0.18**	0.10	**0.37**

Table 4: Correlation among OO metrics

Metric	**WMC**	**CBO**	**RFC**	**LCOM**	**LOC**	**CAM**	**TPC**	**APC**	**GTC**	**VTC**	**EPM**
WMC	**1.00**	**0.20**	**1.00**	**1.00**	**1.00**	**-0.84**	**0.76**	-0.07	**0.17**	**0.28**	**0.41**
CBO	**0.20**	**1.00**	**0.20**	**0.20**	**0.20**	**-0.37**	**0.29**	**0.26**	**0.41**	-0.07	-0.15
RFC	**1.00**	**0.20**	**1.00**	**1.00**	**1.00**	**-0.84**	**0.76**	-0.07	**0.17**	**0.28**	**0.41**
LCOM	**1.00**	**0.20**	**1.00**	**1.00**	**1.00**	**-0.84**	**0.76**	-0.07	**0.17**	0.28	**0.41**
LOC	**1.00**	**0.20**	**1.00**	**1.00**	**1.00**	**-0.84**	**0.76**	-0.07	**0.17**	**0.28**	**0.41**
CAM	**-0.84**	**-0.37**	**-0.84**	**-0.84**	**-0.84**	**1.00**	**-0.63**	0.08	**-0.24**	**-0.35**	**-0.36**
TPC	**0.76**	**0.29**	**0.76**	**0.76**	**0.76**	**-0.63**	**1.00**	**0.55**	**0.33**	**0.28**	0.08
APC	-0.07	**0.26**	-0.07	-0.07	-0.07	0.08	**0.55**	**1.00**	**0.30**	0.04	**-0.33**
GTC	**0.17**	**0.41**	**0.17**	**0.17**	**0.17**	**-0.24**	**0.33**	**0.30**	**1.00**	0.03	**-0.18**
VTC	**0.28**	-0.07	**0.28**	**0.28**	**0.28**	**-0.35**	**0.28**	0.04	0.03	**1.00**	**0.38**
EPM	**0.41**	-0.15	**0.41**	**0.41**	**0.41**	**-0.36**	0.08	**-0.33**	**-0.18**	**0.38**	**1.00**

if a group of variables in a data-set are strongly correlated, these variables are likely to measure the same underlying dimension (i.e., cohesion, complexity, coupling, etc.). Because of this, we also calculated the statistical correlation among OO metrics for the employed version of the data-set. Table 4 shows the results, which confirm that OO metrics are not statistically independent. Then, RFC, LCOM and LOC can be removed while keeping WMC, since these are statistically correlated at the 5% level with the maximum correlation factor. To understand why this happens it is worth considering that, in most cases, the WSDL generation tool input is a code facade containing only the signatures of the services' operations. Therefore, each new method added to the facade contributes with exactly one new line of code, which explains the perfect correlation between WMC and LOC. Similarly, since LCOM measures the cohesion of a class based on its instance variables but the services' facades contain no such variables, the value of these metrics increases in the same proportion as WMC does. Finally, RFC is defined as the set of all methods and constructors that can be invoked as a result of a message sent to an object of the class. This set includes the methods in the class and methods that can be invoked on other objects. Since methods signatures contain no logic, no invocations are made from the class methods. Therefore, for these facades, the value of RFC is exactly the same of WMC, thus resulting in a maximum correlation factor between the two.

Once the number of independent metrics has been reduced the next two subsections focus on describing the correlation relationships found.

4.2.1 Sneed's WSDL metrics

By analyzing the correlation between OO and Sneed's WSDL metrics, it can be seen that the values of two WSDL metrics can be computed based on the values of a different sub-set of WSDL metrics. Average Interface Complexity is the average value of complexity metrics, and Average Interface Quality is the average value of quality metrics. Then, these two metrics may be removed by basing on the fact that their statistical relationships are represented by base metrics. Furthermore, if we want to keep only highest correlated pairs of variables, the correlation factors below $|0.6|$ at the 5% level can be removed. Then, the OO metrics APC, GTC, and VTC can be removed. Then, from Table 3, a new 4x7 table (see Table 5) can be obtained, which represents a minimalist sub-set of correlated metrics.

In order to analyze the minimalist sub-set shown in Table 5, it is worth remembering the importance of correlation factor signs, since the higher the value of a WSDL metric the higher its complexity or the best its quality is. To clarify this, let us take for example the case of the TPC metric, which is positively correlated to Data Flow Complexity (Elshof's metric), but negatively correlated to Interface Data Complexity. This means that if a software engineer modifies his/her service implementation and in turn this produces an increment in the TPC metric, such an increment will cause an increment in Data Flow Complexity (Elshof's metric), but a fall in Interface Data Complexity. Clearly, incrementing Data Flow Complexity (Elshof's metric) would be undesirable. However the presented evidence shows that this could be the side-effect of pursuing an improvement in the Interface Data Complexity metric. From now on, we will refer to this kind of situations as *trade-offs*. As in software literature in general, here a trade-off represents a situation in which the software engineer should analyze and select among different metric-driven implementation alternatives.

As shown in Table 5, the resulting OO metrics are correlated with at least two WSDL metrics, and the signs of the correlations within one metric are always opposite, with the exception of the CBO metric. Therefore, some

Table 5: Minimalist sub-set of correlated OO and Sneed's metrics

Metric	**WMC**	**CBO**	**CAM**	**TPC**
Interface Conformity	-	**0.62**	-	-
Interface Data Complexity	**-0.83**	-	**0.64**	**-0.61**
Interface Structure Complexity	**0.83**	-	**-0.62**	-
Data Flow Complexity (Elshof's metric)	**0.75**	-	-	**0.63**
Language Complexity (Halstead's metric)	-	**0.70**	-	-
Interface Adaptability	**0.61**	-	-	-
Interface Reusability	**-0.60**	-	-	-

Table 6: Minimalist sub-set of correlated OO and Basci & Misra's metrics

Metric	WMC	CBO	RFC	LCOM	LOC	CAM	TPC
DW	-	**0.80**	-	-	-	-	**0.74**
DMC	**0.82**	-	**0.82**	**0.82**	**0.82**	**-0.83**	**0.68**
DMR	**-0.71**	-	**-0.71**	**-0.71**	**-0.71**	-	**0.60**
ME	**0.72**	**0.62**	**0.72**	**0.72**	**0.72**	**-0.79**	-
MRS	**0.80**	-	**0.80**	**0.80**	**0.80**	-	-

metrics represent trade-off opportunities. For example, the CAM metric is negatively related to Interface Structure Complexity, but positively related to Interface Data Complexity. Then, by incrementing the CAM metric, resulting WSDL documents will present a better value for Interface Structure Complexity metric than the original WSDL document. However, this will cause a deterioration of Interface Data Complexity metric, since the latter and CAM statistically grow together.

Controlling the CBO metric is safe, in the sense that it does not present trade-off situations and by modifying its value no undesired collateral effects will be generated. This metric is positively correlated to Interface Conformity and Interface Data Complexity. Therefore, a WSDL document may be improved in terms of conformity and data complexity alike by decrementing the CBO metric from its associated service implementation.

Finally, the WMC metric is correlated to 5 of the 7 WSDL metrics in Table 5. The correlation factor signs indicate that the metric is positively related to Interface Adaptability, Interface Structure Complexity, and Data Flow Complexity. This means that by decrementing WMC on a service implementation, the corresponding WSDL document may present improvements with regard to Interface Adaptability, Interface Structure Complexity, and Data Flow Complexity metrics. However, there are two negative correlations, i.e., with Interface Reusability and Interface Data Complexity metrics, respectively. Therefore, WMC presents different 3x2 trade-offs.

4.2.2 Basci & Misra's

We also employed two criteria for reducing the number of OO metrics that particularly explain these WSDL metrics. First, by basing on the data of Table 4 that shows the existence of groups of statistically dependent OO metrics, we select one representative metric for each group for measuring the same underlying dimension. This is, RFC, LCOM and LOC can be removed while keeping WMC, since these are statistically correlated at the 5% level with correlation factor of 1. Second, we removed the APC, GTC, VTC, and EPM metrics because they do not have at least one relationship with its correlation factor above |0.6| at the 5% level. The rationale of this criterion reduction was, again, to keep only the highest correlated pairs of variables.

Table 6 represents a minimalist sub-set of correlated metrics, and shows that the DW metric depends on two OO metrics, i.e., CBO and TPC. Then, the DW of a service may be influenced by the number of classes coupled to its implementation, and by the number of parameters their operations exchange. The results also show that DW is not highly influenced by cohesion related metrics, such as LCOM and CAM, neither by how many methods its implementation invokes (RFC) or methods complexity (WMC).

The DMC and ME metrics may be decreased by reducing the complexity of service implementation methods. This means that if a software developer modifies his/her service implementation and in turn this reduces the WMC, RFC and LOC metrics, such a fall will cause a decrement in DMC and ME. At the same time, DMC and ME may be reduced by improving the cohesion of services implementations. This is because when the cohesion of services implementations is improved, LCOM falls and CAM rises, which may produce a lower value for DMC and ME, by basing on the signs of their respective statistical relations. ME is influenced by CBO as well.

The DMR metric has negative correlations with WMC, RFC, and LOC. Surprisingly, this means that when the complexity of services implementations methods grows, the ratio of distinct messages falls. Also, DMR has a positive correlation with TPC, which means that the higher the number of operations parameters the higher the ratio of distinct messages. The MRS metric presents high correlations with 3 complexity and 1 cohesion service implementation metrics.

The presented evidence shows that pursuing an improvement in the DMR metric may conflict with other metric-driven goals. This means that if a software developer modifies his/her service implementation and in turn this produces an increment in the WMC, RFC, LOC or LCOM metrics, such an increment will cause that DMC, ME, and MRS alert about an increment in the WSDL document complexity, however the DMR will fall. Clearly, incrementing DMC and ME would be undesirable, but this could be the side-effect of a gaining in the DMR metric. As mentioned before, this kind of situations could be treated as *trade-offs*, in which the software engineer should analyze and select among different metric-driven implementation alternatives for his/her Web Services.

5 Conclusions

In this article, we have empirically shown that when developing code-first Web Services, there is a significant statistical correlation between several traditional OO metrics and some WSDL-related metrics that measure complexity/quality and maintainability at the service interface (i.e., WSDL) level. It is worth noting that a correlation between two variables does not imply causation. In other words, former variable values are not a sufficient condition for latter variable ones. However, such a correlation means that the former variable values are a necessary condition for latter ones [24]. In other words, we are not stating that certain levels of WSDL complexity, quality and maintainability are strictly caused by OO metrics values in their associated implementations, but we state that the mentioned WSDL qualitative aspects require certain OO metrics values to be present on services implementations. This enforces the findings reported in [7], in which a correlation between some OO metrics and metrics that measure service discoverability was also found. Interestingly, these facts allow software engineers to early adjust OO metrics, for example via M. Fowler's code refactorings, so that resulting WSDL documents and hence services are less complex, have better quality and are more maintainable.

However, these findings open the door to extra research questions regarding how to increase/reduce a single OO metric, and how to deal with the various trade-offs that arise. With respect to the former issue, modern IDEs provide specific refactorings that can be employed to adjust the value of metrics such as WMC, CBO and RFC. For metrics which are not that popular among developers/IDEs and therefore have not associated a refactoring (e.g., CAM), indirect refactorings may be performed. For example, from Table 4, CAM is negatively correlated to WMC, so refactoring for WMC means indirect refactoring for CAM. In principle, a complete catalog of refactoring actions for each metric regardless popularity levels could thus be built.

The trade-off problem is on the other hand largely more challenging as we have found that refactoring for a particular OO metric may yield good results as measured by some WSDL metrics but bad results with respect to other WSDL metrics in the same target catalog. For example, for the metrics catalog of [3], it can be seen from Table 5 that reducing the WMC metric positively impacts on Interface Reusability, but negatively affects Interface Structure Complexity. However, the absolute correlation between WMC and Interface Reusability is weaker than the correlation between WMC and Interface Structure Complexity (i.e., 0.60 against 0.83). It is then necessary to determine to what extent WMC is modified upon code refactoring so that a balance with respect to (ideally) all WSDL metrics in a catalog is achieved. Such trade-off situations also apply to the WSDL maintainability metrics studied. The empirical evidence shows that pursuing an improvement in the DMR metric may conflict with other metric-driven goals. This means that an increment in the WMC, RFC, LOC or LCOM metrics will cause that DMC, ME, and MRS alert about a decrement in the WSDL document maintainability, however the DMR will fall. Then, incrementing DMC and ME would be undesirable, but this could be the side-effect of having gains with respect to the DMR metric.

Note that this problem is even more difficult when trying to balance the values of metrics of different catalogs, i.e., when attempting to build a service that is of better quality, and lower complexity, but also maintainable. Therefore, addressing this issue is subject of further research.

The present study has at least one treat to validity, which can be classified as "Internal validity" according to [25]. This is because there can be one more factor that has influenced the correlation between services implementations metrics and services interfaces ones, and we have not measured this factor until now. This factor is the tool employed for mapping from services implementations onto WSDL documents. Changing the code-first tools is not always a smooth process, and we have to place effort on preparing each project of the data-set for the specific requirements of each selected tool. For example, for using WSProvide a developer requires to place proper annotations on a service implementation. This is the main reason because in this study code-first tools influence has not been measured.

Therefore, in a follow-up study, we will analyze the correlation when WSDL documents are generated via code-first tools other than Java2WSDL, particularly Java2WS, EasyWSDL, and WSProvide. This will allow us to assess the effect of the associated code-WSDL mappings in the correlation tables reported so far.

Acknowledgments

We acknowledge the financial support provided by ANPCyT through grant PAE-PICT 2007-02311. We thank the reviewers for their helpful comments to improve this paper, and the CIbSE 2012 reviewers for their suggestions regarding the conference version of this paper. We deeply thank Harry M. Sneed and Sanjay Misra for their predisposition and valuable help towards building this article, whom provided us with several relevant documents and articles. Moreover, Harry M. Sneed facilitated us a working copy of the SoftAUDIT tool used in our experiments. We also thank Taiyun Wei, the author of the *R* script for drawing the graphical correlation matrixes.

References

[1] J. M. Rodriguez, M. Crasso, C. Mateos, A. Zunino, and M. Campo, "Bottom-up and top-down COBOL system migration to Web Services: An experience report," *IEEE Internet Computing*, 2011, to appear.

[2] G. Ortiz and A. G. D. Prado, "Improving device-aware Web Services and their mobile clients through an aspect-oriented, model-driven approach," *Information and Software Technology*, vol. 52, no. 10, pp. 1080–1093, 2010.

[3] H. M. Sneed, "Measuring Web Service interfaces," in *12th IEEE International Symposium on Web Systems Evolution (WSE), 2010*, Sep. 2010, pp. 111 –115.

[4] J. M. Rodriguez, M. Crasso, A. Zunino, and M. Campo, "Improving Web Service descriptions for effective service discovery," *Science of Computer Programming*, vol. 75, no. 11, pp. 1001–1021, 2010.

[5] S. Chidamber and C. Kemerer, "A metrics suite for Object Oriented design," *IEEE Transactions on Software Engineering*, vol. 20, no. 6, pp. 476–493, 1994.

[6] J. Bansiya and C. G. Davis, "A hierarchical model for Object-Oriented design quality assessment," *IEEE Transactions on Software Engineering*, vol. 28, pp. 4–17, January 2002.

[7] C. Mateos, M. Crasso, A. Zunino, and J. L. O. Coscia, "Detecting WSDL bad practices in code-first Web Services," *International Journal of Web and Grid Services*, vol. 7, pp. 357–387, 2011.

[8] D. Baski and S. Misra, "Metrics suite for maintainability of extensible markup language Web Services," *IET Software*, vol. 5, no. 3, pp. 320–341, 2011.

[9] R. A. Van Engelen and K. A. Gallivan, "The gsoap toolkit for web services and peer-to-peer computing networks," in *2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID '02)*. Washington, DC, USA: IEEE Computer Society, 2002, pp. 128–135.

[10] R. Subramanyam and M. S. Krishnan, "Empirical analysis of ck metrics for Object-Oriented design complexity: Implications for software defects," *IEEE Transactions on Software Engineering*, vol. 29, no. 4, pp. 297–310, 2003.

[11] T. Gyimothy, R. Ferenc, and I. Siket, "Empirical validation of Object-Oriented metrics on open source software for fault prediction," *IEEE Transactions on Software Engineering*, vol. 31, no. 10, pp. 897–910, 2005.

[12] P. Meirelles, C. S. Jr., J. Miranda, F. Kon, A. Terceiro, and C. Chavez, "A study of the relationships between source code metrics and attractiveness in free software projects," in *Brazilian Symposium on Software Engineering (SBES '10)*, vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 11–20.

[13] J. Fan and S. Kambhampati, "A snapshot of public Web Services," *SIGMOD Record*, vol. 34, no. 1, pp. 24–32, 2005.

[14] M. B. Blake and M. F. Nowlan, "Taming Web Services from the wild," *IEEE Internet Computing*, vol. 12, pp. 62–69, September 2008.

[15] J. Pasley, "Avoid XML schema wildcards for Web Service interfaces," *IEEE Internet Computing*, vol. 10, pp. 72–79, 2006.

[16] N. Chapin, "A measure of software complexity," *Managing Requirements Knowledge, International Workshop on*, vol. 0, p. 995, 1979.

[17] S. Henry and D. Kafura, "Software structure metrics based on information flow," *Software Engineering, IEEE Transactions on*, vol. SE-7, no. 5, pp. 510 – 518, sept. 1981.

[18] IEEE Computer Society, "610.12-1990 - IEEE standard glossary of software engineering terminology," http://standards.ieee.org/findstds/standard/610.12-1990.html, 1990.

[19] M. H. Halstead, *Elements of Software Science (Operating and programming systems series)*. New York, NY, USA: Elsevier Science Inc., 1977.

[20] H. M. Sneed, "Applying size complexity and quality metrics to an Object-Oriented application," in *Proceedings of the European Software Control and Metrics Conference*, 1999.

[21] G. K. Tayi and D. P. Ballou, "Examining data quality," *Commun. ACM*, vol. 41, pp. 54–57, February 1998.

[22] D. Spinellis, "Tool writing: A forgotten art?" *IEEE Software*, vol. 22, pp. 9–11, 2005.

[23] S. Stigler, "Fisher and the 5% level," *Chance*, vol. 21, pp. 12–12, 2008.

[24] J. Aldrich, "Correlations genuine and spurious in pearson and yule," *Statistical Science*, no. 10, pp. 364–376, 1995.

[25] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering: an introduction*. Norwell, MA, USA: Kluwer Academic Publishers, 2000.