AN ASSESSMENT OF RIVER WATER QUALITY: CASE STUDY OF JOHOR RIVER BASIN, MALAYSIA

PAK HUI YING

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

2020

An Assessment Of River Water Quality: Case study of Johor River Basin, Malaysia

Pak Hui Ying

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University

in partial fulfilment of the requirement for the degree of

Master of Engineering

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

12 November 2020

Date

Pak Hui Ying

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

09/12/20

. . . .

Date

. . . .

Shane Allen Snyder

Authorship Attribution Statement

(B) This thesis contains material from one paper published in the following peer-reviewed journal for which in which I am listed as

an author.

Chapter 5, section 5.1 to section 5.9 is published as Pak, H. Y., Chuah, C. J., Tan, M. L., Yong, E. L., & Snyder, S. A. (1920). A framework for assessing the adequacy of Water Quality Index–Quantifying parameter sensitivity and uncertainties in missing values distribution. Science of the Total Environment, 751, 141982. DOI: 10.1016/j.scitotenv.2020.141982

The contributions of the co-authors are as follows:

- Prof S.A Snyder provided the resources, initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts, conducted the laboratory work, and data analysis. The manuscript was revised by Dr C.J. Chuah, Dr M.L. Tan, and Dr. E.L. Yong.

12 November 2020

. . . .

Date Pak Hui Ying

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Professor Shane Snyder, who has provided me with invaluable advice and resources, and liaising with University Technology Malaysia (UTM) for seamless collaboration with other experts and researchers. I would also like thank Dr. Joon Chuah, who has assisted me with field work sampling and ensured my safety during field work. He constantly encourages me to learn new skills, which has proved to be invaluable in my learning. I am extremely appreciative and grateful of his effort in steering me in the right direction and providing me with invaluable advice.

Laboratory work in AEBC would not have been possible without the help of the analytical team of Advanced Environmental Biotechnology Centre (AEBC). I am grateful for all their assistance in the procurement of chemicals, and the operation and maintenance of analytical instruments. I would also like to acknowledge the assistance in SWAT modelling from Dr. Tan Mou Leong, who was introduced to me by Dr. Chuah. Dr. Tan introduced me to SWAT modelling and provided me with his enthusiastic encouragement and invaluable advice. I would not have delve into deterministic modelling if not for him, and I have come to appreciate the usefulness of SWAT modelling.

TABLE OF CONTENTS

ABSTRACT

Water quality monitoring is one of the most important pillars in water resource management. However, water quality monitoring can be resource intensive, especially for developing countries with limited resources. As such, Water Quality Indices (WQIs) are developed as a convenient tool to summarise general water quality into a single numerical value using a few water quality parameters. It helps to condense and communicate water quality information to policy-makers in an efficient way that could guide future water resource management. In addition, it is imperative to understand the influence of surrounding land use on water quality in order for effective water resource management to materialise. Water quality in Johor River is heavily influenced by surrounding land use within the Johor River Basin (JRB) in the state of Johor, Malaysia. In the recent years, there have been recurring reports of pollution in these rivers, which has generated concerns over the long-term sustainability of the water resources in the JRB. Specifically, this water resource is a shared commodity between two states, namely, Johor state of Malaysia and Singapore, a country located south of Johor. However, prior to this study, few research on the influence of land use configuration on water quality has been conducted in Johor. In addition, it is also unclear how point sources of pollution from wastewater treatment plants (WWTPs) influence water quality under different seasonality in the JRB. This study aims to develop a site-specific WQI for the Johor River Basin (JRB), Malaysia, and understand how external sources (i.e. land use and point sources) affect water quality. Water samples were collected from 49 sites within the JRB from March to December in 2019. Spatial analysis showed that the most polluted rivers were Chemangar, Lebam, and Tiram river, while Pelepah has the highest water quality as it consistently scored the highest WQI values across time. Results showed that influence from WWTPs on water quality was greater during the dry season, and less significant during the wet season. In particular, point source was highly positively correlated with ammoniacal-nitrogen (NH3-N). On the other hand, land use influence was greater than point source influence during the wet season. Residential and urban land use were important predictors for nutrients and organic matter (chemical oxygen demand); and forest land use were important sinks for heavy metals but a significant source of manganese.

Key words: Water Quality Index, Johor, statistical decision theory, land use configuration, water quality, point source

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

Surface water forms the interface where water interacts with both atmosphere and soil through physical and biochemical processes. It is these three essential mediums that largely determines the health of a watershed and its ecosystem functioning. However, burgeoning anthropogenic activities, especially with rapid urbanisation, has become one of the greatest external forces affecting water resources globally. Stresses to water resources includes reduction in water quantity and quality due to poor resource management and practices, climate change and water pollution (Novotny, 2002; Fennell et al., 2006; Bates et al., 2008). Water pollution is emerging as one of the largest threats to water resources, especially in urban regions (McDonald et al., 2011; Yang et al., 2018). As such, it is of imperative that water quality monitoring are undertaken to guide and improve water resource management. However, water quality monitoring efforts are usually financially taxing, especially for developing countries, due to the high analytical costs and manpower required (O'Flynn et al., 2010). As such, the Water Quality Index (WQI) is widely used by many countries (e.g. Dunnette (1979); Dojlido et al. (1994); Cude (2001); Liou et al. (2004); Abbasi & Abbasi (2012); Gazzaz et al. (2012); Alves et al. (2014); Bilgin (2018); Dutta et al. (2018); Ewaid et al. (2018); Wu et al. (2018)) as an integrated method to assess the water quality and the state of pollution in surface waters. It captures important water quality parameters and these parameters are converted into a single value, which helps to inform researchers and policy-makers on the general water quality status of surface waters (Lumb et al., 2011; Gitau et al., 2016). As such, the WQI is of great utility to developing countries as a form of convenient and integrated water quality monitoring strategy, that is ultimately essential for water resource management. This is especially relevant for rapidly developing regions like Southeast Asia, where water management practices struggle to evolve and adapt as quickly as the rate of water resource degradation (Tortajada & Islam, 2011; Chan, 2012).

Water quality assessed by WQI has to be contextualized based on the factors governing water quality, to better understand the underlying processes affecting water quality. The major factors affecting water quality includes geology, soil, atmospheric (wet/dry) deposition, land use, and point and diffuse source of pollution. Geological and soil factors are processes that has a relatively constant influence on water quality, whereas atmospheric deposition, land use, and pollution sources have a more variable influence on water quality (i.e. atmospheric pollution such as haze, industrial emissions, volcanic eruptions). However, land use and pollutant sources have a greater and consistent influence on water quality compared to other sources (Baker, 2006). Point and diffuse sources of pollution are generally associated to certain land use types, especially in an urban setting (Bowes et al., 2008; Zhou et al., 2016; Bo et al., 2017). For instance, WWTPs and industries are indispensable features in urban areas due to treatment of wastes and production of goods. As such, it may be challenging to differentiate the influence of land use and pollution sources that are inherent to urban land use. In addition, the WQI is a compound index calculated from multiple water quality parameters, and thus, associating WQI with land use variable may be challenging, unless a land use is highly associated to a sensitive water quality parameter. As such, it is important to consider and quantify how land use affects each significant water quality parameter.

Previous literature has documented the influence of land use on water quality, and it is well established that water quality is greatly influenced by surrounding land uses as they are sources of non-point source pollution (i.e. Lee et al. (2009); Bu et al. (2014); Ding et al. (2016); Yu et al. (2016); Bo et al. (2017); Cheng et al. (2018)). However, research on land use configuration effects on water quality in South East Asia has been scarce due to limited resources and lack of environmental monitoring efforts. Given that Southeast Asia is rapidly developing with increasing stress on its water resources (Tortajada & Islam, 2011; Chan, 2012), it is of imperative that such studies are undertaken to guide land use planning and improve water resource management.

Johor, the southernmost state in Peninsular Malaysia, is one of the fastest developing state in Malaysia, and is rapidly urbanising in response to the changing economic climate. The Johor River Basin (JRB) is an important source of water for drinking, industrial, and agricultural purposes in Johor, providing up to 55% (927 million litres per day (MLD)) of the total water demand in Johor (MNRE, 2011). By virtue of its close proximity with Singapore, the city state of Singapore also imports water from the Johor River. Under the 1962 Water Agreement between Malaysia and Singapore, Singapore is entitled to draw up to 1100 MLD – nearly half Singapore's daily demand for drinking water. One of the burgeoning issues affecting water resources for both parties is water pollution. Water quality in the rivers of JRB is heavily influenced by surrounding land use within the catchment. Particularly, there have been recurring reports of high ammonium content. For example, between 2017 and 2019, a total of seven cases of ammonium contamination have caused the temporary shutdown of one of Singapore's drinking water treatment plants, and the durations of these disruptions may range from several hours to several days (MEWR, 2019). On the Malaysian side of the border, such incidents have even resulted in the shutting down of five water treatment plants in 2019, and disrupting water supply to tens of thousands of residents in Johor (Shah, 2019). This in turn translates to higher treatment costs of water, and heightened water stress to both Johor and Singapore.

1.2 Purpose and scope

In light of the above issues concerning the water resource shared between Johor and Singapore, the main objective of this research thus aims to conduct spatio-temporal monitoring of Johor river within the JRB, and develop a site-specific WQI for the JRB. Thereafter, the WQI shall be assessed using statistical decision theory as an important framework which considers the optimisation of sample sizes, sampling costs, and uncertainties in water quality information in the context of proactive water quality monitoring. In view of the limited resources and where field sampling is still commonplace especially in developing regions, it is vital to optimise the number of samples collected. This is because sampling costs and uncertainties in water quality information are closely associated with sample sizes. This research also aims to quantify the impact of present land use on the water quality in Johor River, and understand the influence of point sources on water quality, on top of land use configuration.

The aims of this paper are set out as follow:

1. Analyse spatial and temporal trends in water quality within the JRB, and identify key water quality parameters that are representative of the general water quality in the JRB
2. Develop and evaluate the performance of various forms of WQIs through different methodologies, and provide various optimal sampling sizes scenario given various threshold of acceptability for changes in WQI
3. Identify the most significant land use metric and scale (i.e. reach, riparian, and sub-basin scale) affecting each important water quality parameter
4. Quantify influence of land use and point source (WWTPs) on water quality under different seasonality

2. LITERATURE REVIEW

2.1 History of WQI

WQI is widely used by many countries as an integrated method to assess the water quality and the state of pollution of surface waters. It captures important water quality parameters, and these parameters are converted into a single value, which helps to inform researchers and policy-makers on the general water quality status of surface waters (Lumb, Sharma, & Bibeault, 2011). This is particularly important for environmental management in developing countries as it greatly reduces analytical cost in view of limited resources. Further, developing a WQI serves as a convenient method for comparing water quality between various places, and it facilitates policy-makers in implementing and improving present water management strategies.

Various WQI have been developed globally, with its inception in the US by the National Sanitation Foundation (NSF) in 1965 (Horton, 1965). Parameters were selected based on the Delphi method, which is essentially a survey method where a group of water quality experts determines which water quality parameters are deemed important, and gives an assigned weight to each parameter. However, the Delphi method had its shortcomings as being an index that is not as objective (Lumb et al., 2011). The NSF-WQI model were calculated in additive and multiplicative forms (Horton, 1965; Brown, McClelland, Deininger, & Tozer, 1970). A refined form of the NSF-WQI is the Oregon WQI (OWQI), which employs harmonic averaging (harmonic mean square root formula), and removes the arbitration in assigning weights to the parameters (Dunnette, 1979). Another model of WQI developed in Spain by Bascaron (1979) normalises the concentration of the water quality parameters, and applies an additive model with its assigned weight to each parameters, similar to the mathematical form of the NSF-WQI model. Another manifestation of the NSF based WQI is the Mekong Commission (Table 2.1).

An entirely different approach in calculating a WQI was developed by the Canadian Council of Ministers of the Environment. Conceptually, the CCME WQI comprises of three factors: Scope, Frequency, and Amplitude. Scope refers to the number of water quality variables that fails to meet the water quality guideline; Frequency refers to the frequency where the observed value was above the guideline limits; and Amplitude refers to the magnitude of deviation by which the observed value exceeds the guideline limits (CCME, 2001). In another vein, biological indices were at times used in place of, or in conjunction with a WQI. Index of Biotic Integrity (IBI) was first developed by Karr (1981) to evaluate ecological health of surface waters. It characterises benthic conditions through quantifying the diversity and types of macro-invertebrates and fish species. They serve as important bio-indicators which informs the water quality. Other similar indices include Trophic State Index (TSI), Aquatic Life Index (ALI), Contamination Potential Index (CPI), and Heavy Metal Pollution Index (HPI). These indices serve as different manifestations of the WQI that were introduced for specific water quality concerns such as ecological health, and/or different end use.

The major differences between various WQIs are the statistical methods in developing the model, interpreting weightage of parameters and parameters of interests (Lumb et al., 2011). The parameters chosen to be included in the WQI is dependent on the availability of data, consideration of different water-utilisation purposes (i.e. drinking, recreational, fishery, agriculture, industry etc), and the spatial scale (i.e. national or regional level) at which the WQI will be implemented. As such, each country's WQI is unique and developed for its waters and its intended usage. The challenge in Horton (1965)'s concept, however, was the arbitrary and subjective selection of the "suitable" parameters to be included in the WQI.

There has been novel methods in recent years with the introduction of machine learning and deep learning methods in the development of WQI. For instance, approaches such as multi-layer perceptron neural networks, self-organising maps neural network, fuzzy inference neural networks, and probabilistic neural networks (Abdullah et al., 2008; Nikoo et al., 2011; Zhou et al., 2016). Multivariate linear regression (MLR) is another widely-used and popular statistical methods in developing a minimum WQI (WQI_{MIN}). For instance, Wu et al (2018) employed a multiple step-wise regression technique that reduces the number of water quality parameters that is largely representative of the general water quality. An advantage of this statistical method is the elimination of subjectivity in selecting water quality parameters to be included in the WQI. The following section shall delve into the statistical methods in developing a WQI.

2.2 Methods in developing WQI

Generally, four major steps are undertaken in developing a WQI (Abbasi & Abbasi, 2012):

1) Selection of parameters
2) Transforming parameters to a common scale (Developing sub-indices)
3) Assigning weightages to parameters
4) Aggregation of indices to produce a WQI

2.2.1 Selection of parameters

Three types of system are used to describe the selection of parameters: Fixed, open and mixed systems. Fixed system entails using a fixed set of parameters (e.g. WQI of Department of Environment (DOE), Malaysia, referred to as WQI_{MY} hereafter). While a fixed system enables convenient comparison across sites, it is too general for use in understanding the general water quality of a specific site. In addition, the WQI is inflexible to include other important additional variables to address site-specific water quality problems or future application (Swamee & Tyagi, 2007). An open system entails using a minimum number of significant parameters with specific characteristics (e.g. Dojlido et al. (1994); CCME (2001); MoEI (2003)). It offers the flexibility to include as many parameters specified by the user. A mixed system includes the common physio-chemical water quality parameters, as well as other parameters. These other parameters are often associated to toxicologic effects, albeit less frequently monitored (Hanh et al., 2011). In some cases of aggregating the final WQI, the additional parameters may be included if the 'basic' parameters exceed a certain sub-index value (Dojlido et al., 1994). The fixed and mixed system aims to select parameters that are the most representative of water quality.

Including every possible water quality parameter in the WQI makes it unwieldy, and defeats the purpose of convenience that the WQI offers. Generally, parameters should be selected based on a thorough literature review, data availability, and most importantly, parameters that are representative of a given end use (Dunnette, 1979; Pesce & Wunderlin, 2000; Hanh et al., 2011; Abbasi & Abbasi, 2012). Two main strategies are commonly employed in parameter selection, namely, Delphi method and statistical analysis of data. The Delphi method is essentially a survey method where a group of water quality experts determines which water quality parameters are deemed important, and gives an assigned weight to each parameter (Abbasi & Arya, 2000). A large number of experts are consulted to reduce the problem of subjectivity. Statistical analysis in parameter selection includes multiple step-wise regression technique, Pearson's correlation and Principal Component Analysis (PCA).

Wu et al. (2018) and Tripathi & Singal (2019) employed multiple step-wise regression and PCA techniques respectively in developing a minimum WQI (WQI_{MIN}). From a range of measured water quality parameters, the multiple step-wise regression technique captures important and significant water quality parameters. These selected water quality parameters are largely representative of the site's general water quality. PCA is often used for data reduction, and especially useful for applications involving many variables that may be highly correlated. Dimensionality reduction is conducted by considering a number (smaller than the initial number of explanatory variables) of standardized linear combinations of the explanatory variables. The dataset is broken down into sets of independent (uncorrelated) SLCs, known as principal components (PCs), which explains the variance of the entire dataset. The first few PCs would be able to capture around 95 - 99% of the dataset's total variation (Singh et al., 2004). The selected PCs are those with eigenvalues >1 (Burstyn, 2004). It should be noted that every method has its flaws and inherent uncertainties. For instance, pre-selection of parameters through the Delphi method or statistical selection of parameters cannot achieve 100% objectivity or accuracy. However, efforts should be made to reduce subjectivity within the given constraints.

2.2.2 Transforming parameters to a common scale

Transforming parameters to a common scale is necessary as various parameters have different units and range. Subindices (SIs) are parameters that have been transformed to a common scale, and there are four different methods of defining SIs – Linear, Non-linear, Segmented linear, and Segmented Non-linear. Linear SIs are computed by a simple linear equation. While it is easy to compute, it has limited flexibility in accommodating non-linear relationships between certain water quality parameters and SIs (e.g. Dissolved oxygen (DO) and pH). Non-linear SIs are defined by empirical curves or graphs obtained from studies (e.g. WQI_{MY}), albeit the empirical graphs can vary across region. For instance, an empirical graph for DO developed in the temperate region may be unsuitable for use in the tropical region where Temperatures are higher, and DO is a Temperature-dependent water quality parameter. As such, empirical graphs for creating SIs may at times be site-specific and dependent on local surface water conditions. Segmented SIs entails multiple-state functions, where a SIs is defined by values falling within specified break points. Segmented SIs offers more flexibility as water quality standards (e.g. United States Environment Protection Agency (USEPA), World Health Organisation (WHO)) can be incorporated. Non-linear segmented SIs is more advantageous over linear segmented SIs, especially for situations where increasing levels of pollution are observed. Past studies (e.g. Koçer & Sevgili (2014); Wu et al. (2018)) referred to literature for the break points (or ranges of water quality values), albeit one must be cautious when utilising the break points specified in literature. Criteria for break points (may be defined by local water quality standards) may vary from region to region, depending on the "acceptability" of certain water quality parameters (Abbasi & Abbasi, 2012). When break points are defined by water quality standards, subjectivity may still be involved when choosing from different water quality standards (e.g. United States Environment Protection Agency (USEPA), World Health Organisation (WHO)). In addition, for every type of water quality standard (local or regional), there are insufficient classes to define every unique break point, thus other break points have to be filled in with arbitrary values, which is inherently subjective and flawed.

2.2.3 Assigning weightages to parameters

Weights are often assigned to parameters, contingent on their importance to certain water quality aspects such as suitability of drinking water and ecological health. However, several indices (e.g. Oudin et al. (1999); CCME (2001); Cude (2001); Hallock (2002); Hanh et al. (2011)) assign same weights to parameters to eliminate subjectivity. Moreover, the final index may be sensitive to parameters assigned the highest weight. Parameters that are assigned lower weights, despite exceeding water quality standards, may not be explicitly exhibited in the final WQI value – a phenomenon often known as eclipsing (Abbasi & Abbasi, 2012). For instance, for an index where DO is highly weighted, and nutrients parameters are assigned low weights, high levels of nutrients may not affect the final WQI value significantly if DO level is at its optimal value (Sutadian et al., 2016). Cude (2001) recommended non-weighted WQIs if a user is interested in the general status of water quality, and a weighted WQI to assess quality of water for any specific end-use (e.g. WQI_{MY}).

2.2.4 Aggregation of indices to produce a WQI

Some of the most common methods for aggregating the sub-indices into a final WQI includes the additive and multiplicative method. The additive method; however, may mask the lower values of sub-indices when larger values of sub-indices dominates (Smith, 1990; Liou et al., 2004; Swamee & Tyagi, 2007; Juwana et al., 2012). For instance, the additive method would not affect the overall index significantly even if one of the sub-indices is zero, whereas on the other hand, the overall index would be heavily penalised for the multiplicative method (i.e. produce a zero). At the same time, this also poses a limitation for the multiplicative method, especially if many observations have sub-indices of zero. For instance, WQI_{MY} employs the additive aggregation method which takes the form of the following **Equation 1**:

$$WQI = (0.22 * SIDO) + (0.19 * SIBOD) + (0.16 * SICOD) + (0.15 * SIAN) + (0.16 * SISS) + (0.12 * SIpH) \quad (1)$$

Where SIDO = Sublndex DO (% saturation), SIBOD = Sublndex BOD, SICOD = Sublndex COD, SIAN = Sublndex NH3-N, SISS = Sublndex SS, SipH = Sublndex pH, $0 \leq WQI \leq 100$

2.3 WQI studies in state of Johor

There have been several studies on WQI within the state of Johor, Malaysia, where WQI was calculated for investigating various water quality objectives. Most studies have utilised Malaysia's current WQI (WQI_{MY}), developed by the Department of Environment (DOE), Malaysia (Shuhaimi-Othman et al., 2007) as a standardised method to calculate the WQI. The objectives of these studies based in Johor that utilised WQI_{MY} range from water quality monitoring purposes, comparative WQI studies, and water quality modelling purposes. For instance, Awang et al. (2015)

calculated the WQI for the Sembrong River, and the water quality was found to be of Class 3 (moderate quality) due to discharge of untreated sullage. Zaiha et al. (2015) also used WQI as a monitoring tool to investigate the effect of logging activities on the Berasau River. It was also found to be between Class 2 (good quality) and Class 3 (moderate quality). Arman et al. (2013) utilised various indices, including WQI_{MY}, to investigate the impact of anthropogenic influences on Mengkibol River. It was found that ammoniacal-nitrogen content in the river has exceeded the Class 4 (poor quality) guidelines. A comparative study by Nor et al. (2013) evaluated the differences between WQI_{MY} and Biological Water Quality Index (BWQI) on Melana River. Katimon et al. (2017) used the results of WQI_{MY} for water quality modelling studies to predict trends in water quality.

As observed from the previous paragraph, the applications of WQI_{MY} varied widely, and pose many more opportunities for future use. Hence, it is imperative to evaluate the effectiveness of WQI_{MY}, and address any limitations for future improvement.

2.4 Adequacy of Malaysia's current WQI

Malaysia's current WQI (WQI_{MY}) (Shuhaimi-Othman et al., 2007), is a general nation-wide WQI that entails six constituents – namely, dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), suspended solids, ammoniacal-nitrogen (NH_3-N), and pH. These six constituents are important indicators for monitoring ecological health and assessing acceptability of drinking water, while facilitating ease of use with few water quality parameters that require measurement. However, there are limitations in its applicability and utility. The water quality parameters in the WQI_{MY} are pre-identified parameters which may not accurately communicate the representativeness of the water quality for the JRB (Abbasi & Abbasi, 2012). It is imperative that the WQI developed are site-specific and representative of the site's water quality for adequate implementation of water resource practices and management. For instance, given the strong agriculture and industrial sector in Malaysia, Coliform-based indicators, heavy metals and phosphate (a major component in fertilisers), are not included in the WQI_{MY} (Naubi et al., 2016). In addition, it is doubtful whether the parameters in WQI_{MY} can communicate accurately the different water uses stipulated by the National Water Quality Standards for Malaysia (NWQS). The water uses are categorized into five different classes, each with different water uses. The classes are categorized from Class 1 to Class 5, with Class 1 having the highest water quality standard and Class 5 having the worst water quality standard (**Appendix A(c)**).

As mentioned in the aforementioned sections, it is imperative that WQIs are contextualized by their local conditions such as surrounding land use. This is because water quality is heavily influenced by surrounding land use. Thus, the next section shall delve into the influence of land use on water quality.

2.5 Influence of land use on water quality

2.5.1 Effect of land use configuration variables on water quality

Land use configurations have often been used in studying fragmentation effects in ecological studies, albeit such technique is gaining momentum especially in water quality studies (e.g. Lee et al. (2009); Abdi (2010); Ai et al. (2015); Ding et al. (2016); Zhou et al. (2016); Bo et al. (2017); Cheng et al. (2018)). Land use configuration is quantified for each land use type, and subsequently subjected to statistical analysis to investigate the relationship between land use configuration variables and water quality. Some of the commonly used land use variables in such studies include Patch Density (PD), Largest Patch Index (LPI), Edge Density (ED), Landscape shape index (LSI), Shannon’s diversity index (SHDI), and Aggregation index (AI).

Several statistical methods have been utilised to analyse connections between land use variables and water quality parameters. This includes Correlation Analysis (Lee et al., 2009; Bu et al., 2014), Partial Least Squares Regression (PLS regression) (Abdi, 2010; Ai et al., 2015; Cheng et al., 2018), Self-Organising Maps (SOM) (Zhou et al., 2016), Geographically Weighted Regression (GWR) (Tu, 2011; Huang et al., 2015), Cluster Analysis (Kandler et al., 2017), Multivariate linear regression (MLR) (Ding et al., 2016; Bo et al., 2017), Redundancy Analysis (RDA) (Ding et al., 2016; Bo et al., 2017), and Principal Component Analysis (PCA) (Paul, 2005; Amiri & Nakane, 2009). Methods such as PLS and PCA allows for high correlation between explanatory variables and is able to accommodate a wide dataset (i.e. number of explanatory variables is more than the number of observations). However, one of the drawbacks is the difficulty in the quantitative interpretation of each explanatory variable, and its influence on response variable. In addition, similar results may be produced for water quality parameters that are highly correlated to each other, which creates many redundancies in the results. As such, MLR and RDA provides greater clarity in the interpretation of land use influence water quality.

Findings from recent studies have shown that urban areas were highly correlated with pollutants, while forested areas were generally negatively correlated with pollutants. For instance, Mello et al. (2018) and Rodrigues et al. (2018) reported that forest cover was positively associated with water quality whereas urban areas are associated with poor water quality. Similarly, McBride & Booth (2005) reported that stream conditions were better when it flowed through a non-fragmented vegetated riparian buffer, and when urban pavements were absent. PLS was used in Cheng et al. (2018)'s study which entails runoff patterns, altitude, and land use as the explanatory variables. Cheng et al. (2018) reported that agricultural activities were associated with poor water quality in regions with a higher depth of runoff at lower elevation. Grassland was the most significant land use in governing the water quality in low-flow plains. Forest land use was negatively associated with poor water quality in mountainous high-flow regions. PD and ED were the most significant land use variables in influencing water quality, especially in mountainous high-flow region. This also indicates that forest fragmentation increases overland runoff and subsequent pollutant input into the river. Similarly, Lee et al. (2009) employed the use of correlation for analysis, and reported that PD, LPI, and ED were observed to be important land use variables in influencing water quality. In particular, Lee et al. (2009) reported that PD and ED were associated with degraded water quality, and also concluded that fragmentation of forest resulted in degradation of water quality, which corroborates with the findings of Cheng et al. (2018). On the other hand, previous study by Carey et al. (2011) reported that LPI for residential land use had the strongest association to nutrient loads. Interestingly, surface imperviousness (proxy for urban areas) was not found to be a significant land use variable.

2.5.2 Effect of spatial scale on water quality

Other studies such as Ding et al. (2016) and Mello et al. (2018) have also delved into spatial-scale effects on land use-water quality relationships. Ding et al. (2016) reported that land use configuration at the sub-basin scale was more useful in explaining variations in water quality compared to riparian and reach-scale. This observation is also corroborated by findings from Mello et al. (2018). On the other hand, McBride & Booth (2005) found that physical stream conditions were better explained by sub-basin and reach scale (500m buffer of the site), while Rodrigues et al. (2018) reported that the riparian zone is a significant scale at influencing the water quality.

It should be noted that while comparisons with other studies can better reinforce understanding on the relationships between spatial scale and water quality parameters, different methods were adopted in the delineation of such spatial scales. For instance, there is a lack of a benchmark in defining riparian zones, and different studies have used varying arbitrary buffer distance in their definition of riparian zones (Giri & Qiu, 2016). This is because in reality, the riparian zone varies from place to place, depending on the topography and surface conditions of a watershed. Thus, relationships between water quality and spatial scales are complex due to confounding factors such as topography, anthropogenic activity, and land use within the delineated spatial scale.

2.5.3 Effect of land use type and point sources on water quality

Few studies (e.g. Zhou et al. (2016) and Bo et al. (2017)) have investigated the influence of both land use and pollution sources on water quality. Bo et al. (2017) employed the use of Pearson correlation analysis which showed that impermeable surface (proxy for urban areas) was positively correlated with organic matter such as BOD, COD_{Mn}, as well as nutrients such as NH_4-N. On the other hand, impervious surface was associated negatively with DO. Interestingly, water bodies was found to be negatively associated to BOD and NH_4-N. Similarly, MLR analysis also showed that impermeable surface was positively associated to COD_{Mn}, NH_4-N, but negatively associated with DO. Similar to the findings from Wang et al (2014), Zhou et al. (2016) reported an interesting positive correlation between water bodies and phosphorus, and concluded that water bodies serve as a viaduct for nutrient transportation. Other interesting observations by Zhou et al. (2016) included positive correlations between COD_{Mn} and Woodland, and negative correlations with urban, bare land, and orchard land uses, which contradicts with other findings (e.g. Bu et al. (2014); Hur et al. (2014); Huang et al. (2015)).

Bo et al. (2017) also reported that water quality was negatively related to point-source pollutions of with COD_{Mn} and NH_4-N, and overall performance of model improved when point source emissions were considered in the model. On the other hand, Zhou et al. (2016) used a self-organising map (SOM) to investigate whether point-source pollution weakens the relationship between land use and water quality. Zhou et al. (2016) reported that the severity of point-source pollution produced fewer significant land use-water quality correlations, which demonstrates that point source pollution can potentially mask land use influence on water quality.

Seeboonruang (2012) is one of the few studies conducted in Southeast Asia that investigates associations between both land use and point source on water quality under different seasonality. It introduces the use of the Contamination Potential Index (CPI) concept as a method to quantify the magnitude of point source pollution on water quality. The CPI was calculated based on the amount of wastewater and chemicals (e.g. nutrients, organic content) produced for each land use type. A higher CPI indicates high contamination from the source. MLR analysis was undertaken to analyse the correlation between CPI and water quality parameters under different seasonality. Seeboonruang (2012) reported that three main land use types have high potential for pollution to the surrounding water quality, and they are: non-peak season rice agriculture, poultry agriculture, and residential effluent. The results reported by Seeboonruang (2012) were rather different from the aforementioned studies, but gave insights on land-use-water quality relationships that is specific to Nakhon Nayok Province, Thailand, the study area.

3. STUDY SITE

3.1 Johor River Basin

The JRB spans over four districts – Kota Tinggi, Kluang, Kulai Jaya, and Johor Bahru. These urbanised areas, a mosaic of institutional, commercial, industrial, and residential land-use, is congregated in major districts and towns such as Kota Tinggi, Ulu Tiram town, and Layang-Layang town. These towns are found along Johor River, Tiram River, and Sayong River, respectively. The discharge of urban runoff and wastewater effluent (e.g. sewage, sullage, grease, and sediment) from these urban areas may potentially degrade water quality downstream (MNRE, 2011; Ahmad, 2016). In addition, large clusters of residential areas have high population equivalent (PE), which are associated to high density of wastewater treatment plants (WWTPs) for the treatment of wastes. Thus, high density of WWTPs are clustered around major residential areas such as Kota Tinggi, Ulu Tiram, and Chemangan).

The Johor River is located in the south-eastern part of Johor State in Malaysia. It is approximately 123 km long, and flows southward from Belumut Mountain located in the north of JRB, into the Strait of Johor (Tan et al., 2015). The major tributaries are the Linggiu River (downstream of the Linggiu Reservoir), Sayong River, Tiram River, and Lebam Rivers. The JRB is located between latitudes 1°24′–2°10′N and longitudes 103°20′–104°20′E, and covers an area of approximately 2640 km^2 (**Figure 1a**).

Johor is a consummate commercial agricultural state where agriculture dominates the land use (Ministry of Natural Resources and Environment (MNRE), 2011). The largest land-use in the JRB is large-scale oil palm plantation (71%) (**Figure 1b**), where oil palm plantations comprise a large sector of agriculture throughout Malaysia. It is also one of the major sectors contributing to Johor's economic prosperity. However, high nutrient content in fertilisers applied on oil palm plantations can often be washed into nearby rivers during a storm event, resulting in pollution (MNRE, 2011).

Forest cover (14%) forms the second largest land-use in the JRB and is largely found in the Linggiu River sub-catchment. The Linggiu Dam was constructed in Linggiu River sub-catchment, where its outflow feeds into the main Johor River branch, where Singapore obtains her water supply from (Chew, 2011). The Linggiu dam is surrounded by a large expanse of forested area which extends down to Kota Tinggi Waterfall. However, forest cover has been declining due to expansion in commercial agricultural ventures (MNRE, 2011).

Soils are often enriched with trace metal and nutrient concentrations, making them a considerable source of trace metals and nutrients. Five major soil types exists in the JRB, namely – Dystric Histosols, Thionic Fluvisols, Eutric Fluvisols, Orthic Acrisols, and Ferric Acrisols (FAO, 1979) (**Figure 1c**). Oil palm, being a hardy and resilient crop, is able to flourish in many soil types, even in Thionic Fluvisols (acid sulfate soils) (Auxtero & Shamshuddin, 1991), which is demonstrated by the wide-scale oil palm plantation throughout the JRB.

The JRB experiences Köppen *Af* climate, and is characterised by a mean Temperatures of 27^{o}C, an annual rainfall of around 2400mm and a 7-day low flow of 0.38–1.76 m^3s^{-1}(DID, 2015; Chuah et al., 2018). Its climate is characterised as a tropical monsoon with November to January bringing about high precipitation, with frequent floods occurring during December (Tangang et al., 2012). (**Figure 1d**).

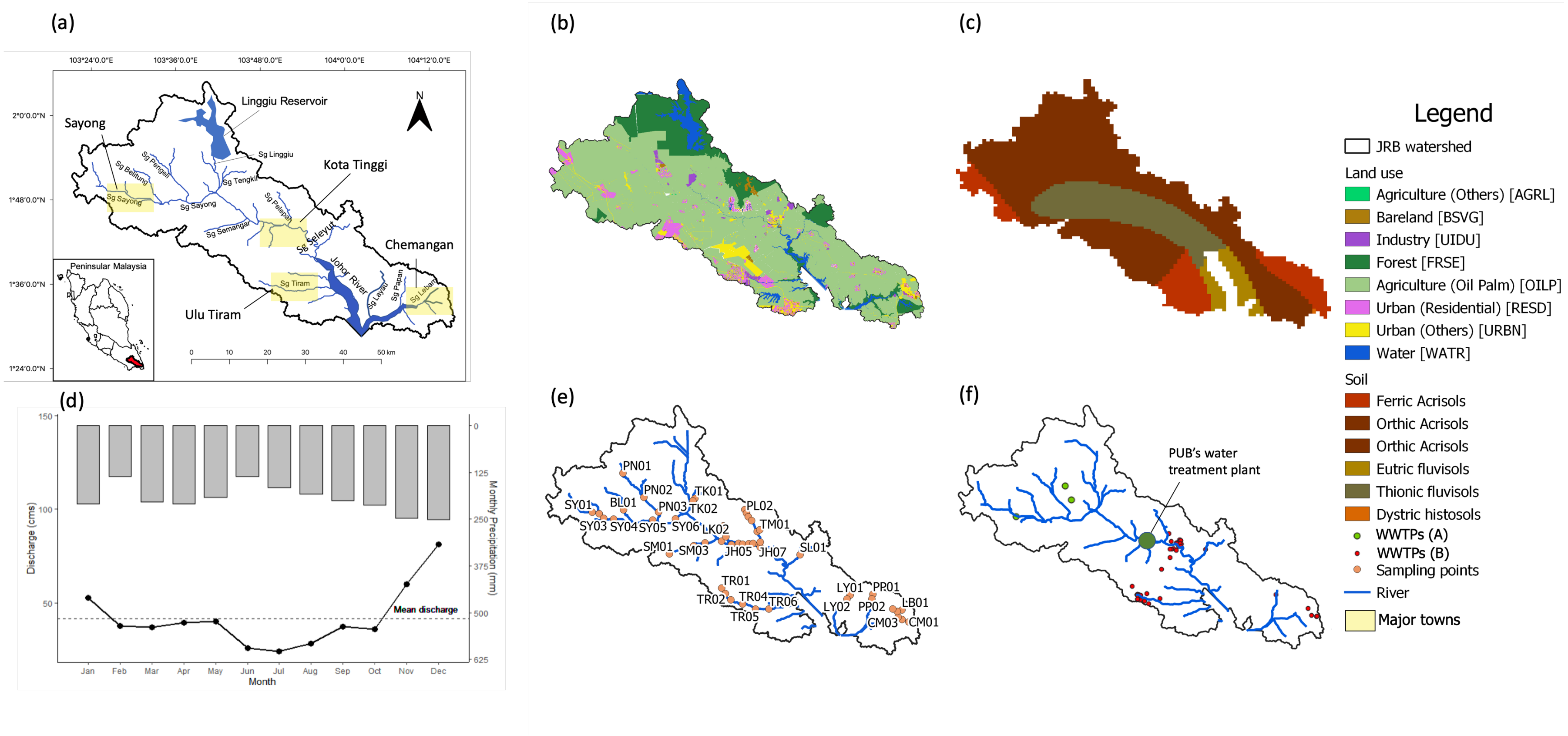


Figure 3.1. Description of the JRB

(a) Rivers within the JRB; (b) Land use map; (c) Soil map; (d) Climate of the JRB; (e) Sampling points; (f) Distribution of WWTPs with respect to PUB's water treatment plant

4. METHODS AND MATERIALS

4.1 Sampling

Active sampling was carried out to ensure the sampling sites were spatially distributed across the main rivers of JRB. Sampling was conducted along the Johor River and fifteen of its main tributaries with at least one sample collected from each river. Sites upstream of Singapore's water treatment plant were prioritised over the sites downstream **(Figure 1e & 1f)**. Sites upstream serve as reference sites as influence from point sources are generally smaller (MNRE, 2011). There are fewer WWTPs on the upstream region and they were mostly of discharge Standard A, while the downstream sites were of discharge Standard B (**Figure 1f & Appendix B**). The upstream sites are: Penggeli River (PN; n = 3), Tengkil River (TK; n = 2), Pelepah River (PL; n = 6), Johor River (JH; n = 8), Lebak River (LK; n = 2), Temenin River (TM; n = 1), Belitong River (BL; n = 1), Sayong River (SY; n = 6), Semangar River (SM; n = 3); and the downstream sites are: Layau River (LY; n = 2), Seleyut River (SL; n = 1), Papan River (PP; n = 2), Chemangar River (CM; n = 3), and Lebam River (LB; n = 3). A reconnaissance trip was conducted on October 2018 and monthly sampling began in March 2019. Samples were collected every four to six weeks from 49 sites over a period of nine months from March 2019 to December 2019 to capture seasonal variations in water quality. For this study, a total of 454 samples were collected.

4.2 Field and laboratory analysis

Dissolved Oxygen (DO), pH, and specific electrical conductivity (EC) were measured in-situ using the YSI Professional Plus Handheld Probe (YSI, Yellow Springs, OH, USA). Prior to field measurement, the probes were calibrated for pH, DO, and EC. Upon collection, the samples were temporarily placed in an icebox to inhibit biochemical reactions from altering the water quality (American Public Health Association, 1995). Samples were transferred to Nanyang Environment and Water Research Institute (NEWRI) at Nanyang Technological University (NTU) of Singapore where they were stored in a dark and cold room at 4°C. All the samples were brought to room Temperature prior to any analysis and analysed within 24 hours of collection, except for biological tests which were conducted within six hours.

Biological tests were conducted immediately after the samples were brought back to Singapore. Colilert kits and Quanti-Trays (IDEXX Laboratories, Westbrook, ME, USA) were used for the enumeration of Coliform (total) and *Escherichia Coli* (E. Coli) using the Most-Probable Number (MPN) method in (MPN/100ml). The Quanti-Trays were then incubated for 24 hours at 35°C and the number of wells were counted under UV-light at 365 nm wavelength for the measurement of E. Coli.

Chemical Oxygen Demand (COD) was measured using the dichromate closed reflux colorimetric method (American Public Health Association, 1995). Nitrite (NO_2-N) was also determined using a colorimetric method (UV-1800 UV-VIS Spectrophotometer, Shimadzu, Kyoto, Japan), using sulfanilamide and N-(1-naphthyl)- ethylenediamine (NED). The diazo dye is measured at 540 nm wavelength.

Ammoniacal-nitrogen (NH_3-N) was measured using TNT 830 (0.05 – 2 mg/l) and TNT 831 (1 – 12 mg/l) HACH kits (HACH®, Loveland, CO, USA), using a spectrophotometer (DR1900™ Portable Spectrophotometer, HACH®, Loveland, CO, USA).

Phosphate (PO_4^{3-}), nitrate (NO_3^-), bromide (Br^-), chloride (Cl^-), sulfate (SO_4^{2-}) and fluoride (F^-) were measured using Ion Chromatography (Dionex™ ICS-2100, Thermo Scientific™, Sunnyvale, CA, USA), whereas cations (Na^+, Mg^{2+}, Ca^{2+}, K^+) were measured using Inductively Coupled Plasma – Optical Emission Spectrometry (Optima™ 8300 ICP-OES, PerkinElmer, Waltham, Massachusetts, USA). Trace metals (Pb, Cd, Cr, Mn, As, Ag, Se, Al, Ba, Cu, Fe, Ni, Zn) were measured using Inductively Coupled Plasma – Mass Spectrometry (iCAP™ Q ICP-MS, Thermo Scientific™, Sunnyvale, CA, USA). Turbidity was measured using a turbidimeter (2100N™ Turbidimeter, HACH, Loveland, CO, USA).

4.3 WQI development and evaluation

4.3.1 Selection of key parameters

A reference WQI, referred to as WQI_{AVG} hereafter, was computed using all measured water quality parameters to serve as a reference for the evaluation of WQI_{MIN} performance **(Figure 4.1)**. After which, both MLR and PCA techniques were conducted independently of each other in the selection of key water quality parameters, and as a form of parameter reduction to create a WQI with the least number of key water quality parameters (WQI_{MIN}). The WQI_{MIN} created from these two techniques will be referred to as WQI_{MLR} and WQI_{PCA}, respectively hereafter. As parameter selection generally involves some ambiguity and inherent uncertainties, two of the most common statistical parameter reduction methods (as described in Chapter 2 – Literature Review) – MLR and PCA were used to reduce subjectivity. Prior to stepwise MLR, data was log(1+x) transformed to ensure normality and homogeneity of variance. Different scenarios were created for each method, due to the difference in their inherent statistical procedure. Stepwise MLR selects significant parameters primarily through relationships between explanatory variables (water quality variables) and each response variable ($WQI_{WEIGHTED}$ or $WQI_{NON\text{-}WEIGHTED}$), while PCA conducts dimension reduction without the need for a response variable (Harrell, 2015). PCA was conducted separately on dataset collected during the dry and wet seasons. MLR was not conducted separately on the dry and wet datasets as MLR is more sensitive to the number of observations that are associated with its corresponding response variable (Harrell, 2015). As such, two different sets of parameters were selected under these two scenarios for each method. MLR was carried out using the base package in R (Core, 2017); and PCA was carried out using *FactoMineR* package in R (Husson et al., 2020).

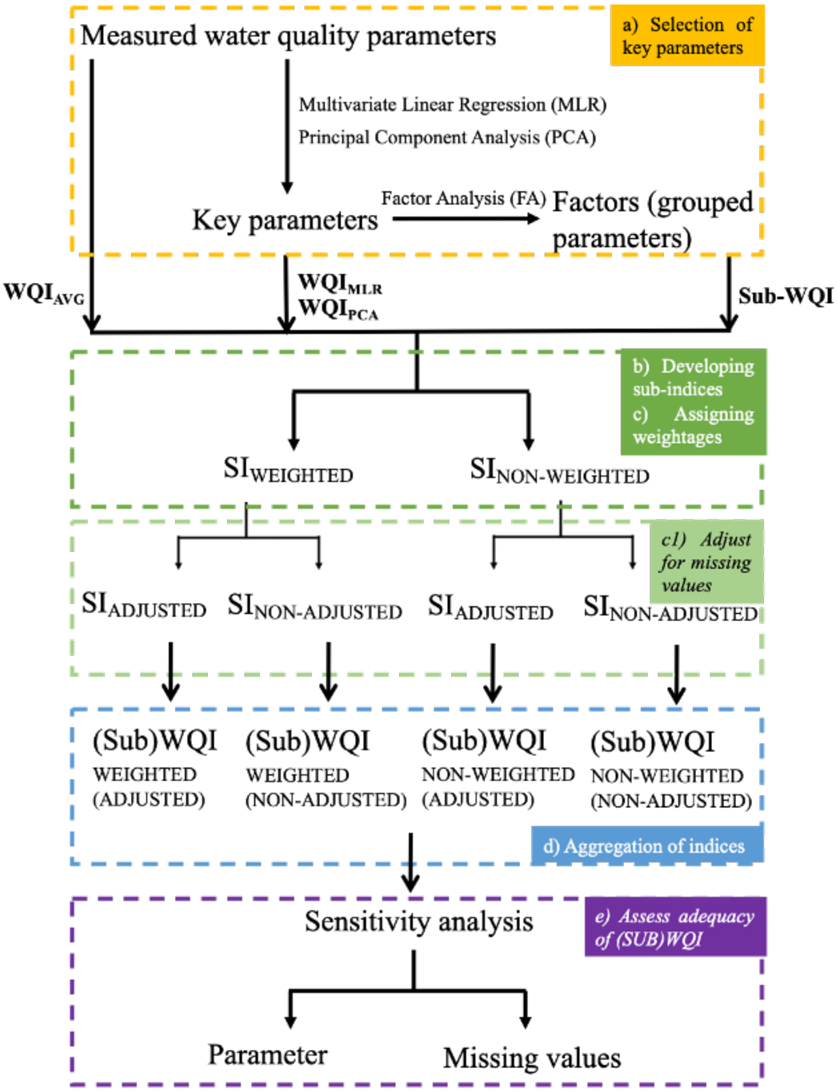


Figure 4.1. Conceptual framework

Notes: Conceptual framework. *Notes*: Weights for each parameter is found in **Table 2** and summary of parameters selected by each method is found in **Table 4. Bolded words** indicates the nomenclature that will be assigned to each path/process. *Italicised words* indicates new processes added to the traditional WQI framework paradigm. (Sub)WQI indicates either Sub-WQI or WQI.

4.3.1 Transforming key parameters to a common scale

The non-linear segmented approach was taken to define the functions for the assignment of SIs **(Table 4.1).** The following method focuses on the dynamic creation of unique threshold breaks and SIs based on the sample distribution of each water quality parameter. In addition, it reduces the limitation of censored data as SIs can still be assigned to censored data given that it falls within the range of the breaks. While it is not a perfect method in handling censored data due to the fact that actual values may be much lower than detection limit, an underestimation of WQI is more useful than an overestimation of WQI (i.e. water quality may in reality be better than what is calculated or represented by the WQI). This is because it will be of greater utility in the context of cost-benefit analysis – where more cost will be incurred if water quality is worse in reality, versus low to no cost if water quality is better in reality. Potential underestimation of WQI (as a result of censored data) – an artefact of representing the worse-case scenario, can better improve water resource management. In comparison, an empirical function would not be able to compute SIs from a censored data.

Table 4.1. Defining subindices

SI_i	Functions
100	$observation_i < x_{i,1}$;
	$x_{i,1} \leq observation_i < x_{i,2}$
90	$x_{i,2} \leq observation_i < x_{i,3}$
80	$x_{i,3} \leq observation_i < x_{i,4}$
70	$x_{i,4} \leq observation_i < x_{i,5}$
60	$x_{i,5} \leq observation_i < x_{i,6}$
50	$x_{i,6} \leq observation_i < x_{i,7}$
40	$x_{i,7} \leq observation_i < x_{i,8}$
30	$x_{i,8} \leq observation_i < x_{i,9}$
20	$x_{i,9} \leq observation_i < x_{i,10}$
10	$x_{i,10} \leq observation_i < x_{i,11}$
0	$x_{i,11} \leq observation_i < x_{i,12}$;
	$observation_i \geq x_{i,12}$

A logarithmic arithmetic progression was used to create the breaks as it achieves a normalised SI where extreme values of observations are captured, and observations are proportionally represented by its SIs throughout the entire distribution. This is especially useful for situations where increasing levels of pollution are observed for certain parameters. The minimum and maximum values were moved away by 0.1% of the range in order to capture the lowest and highest values observed **(Equation 2 – 4)**. There are in total, eleven unique breaks, separated by a constant logarithmic difference as shown in **Equation 5**. The breaks were then exponentiated to scale it back to its original magnitude **(Equation 6)**. It must be noted that if a WQI is calculated for the same watershed (i.e. JRB in this case), the same breaks should be used to ensure that comparisons across space and time are valid **(Appendix C)**.

$$range = \frac{0.1}{100}[\log(b_i) - \log(a_i)] \quad (2)$$

$$min_i = \log(a_i) - range \quad (3)$$

$$max_i = \log(b_i) + range \quad (4)$$

$$d_i = \frac{(max_i - min_i)}{11} \quad (5)$$

$$x_{i,n} = \exp[min_i + (n-1)d_i] \quad (6)$$

Where i = a water quality parameter, a_i = minimum (smallest) value of i, b_i = maximum (largest) value of i, d_i = constant interval value for i, n = number of unique breaks, 1 ≤ integer ≤12, r = number of water quality parameters, x_i = break value of n^{th} sequence, SI_i = subindex of i, W_i = assigned weight of i

4.3.2 Assigning weightages to parameters

Weights were assigned to each water quality parameter ranging from one (having the least impact to water quality) to five (having the most impact to water quality). A weight of five was assigned to water quality parameters that have known toxicity effects to human health and/or aquatic ecosystem health should the levels exceed water quality guidelines (Yidana & Yidana, 2010; Varol & Davraz, 2015) **(Appendix D)**. On the other end of the spectrum, a weight of one was assigned to water quality parameters that have the least significance to water quality assessment. All the weightages are non-zero as setting the weight as 0 would completely discount the effect of the parameter on the water quality even at high concentrations, which may not completely reflect the general status of water quality of the water body. For some water quality parameters, such as sodium, sulfate ions etc, they have small effect on the general water quality at low concentrations. However, at high concentrations, they do have an impact on the ecological health of water and portability of drinking water as demonstrated in the guidelines stipulated by EPA and WHO in **Appendix D**. These effects cannot be ignored, and hence, to err on the side of caution, a non-zero weightage should still be assigned for "less significant" water quality parameters. Both weighted ($WQI_{WEIGHTED}$) and non-weighted ($WQI_{NON-WEIGHTED}$) variations of WQI were also computed in this study, where all the parameters in $WQI_{NON-WEIGHTED}$ have a weight of one **(Figure 4.1)**.

An adjusted form of WQI ($WQI_{ADJUSTED}$) was also created to adjust the WQI for missing values in the data. Where observations are missing, the weight assigned to the missing observation was assigned zero, and would not be calculated in the WQI equation. This is because, in practical and real-life situations, human errors or instrument errors can inevitably lead to missing values. Typically and traditionally, all parameters are needed (i.e. no missing values) to produce a WQI value, which makes the method of calculation undoubtedly rigid, not to mention the resources that may have already been invested in measuring other parameters will regrettably be lost. However, the most glaring limitation of $WQI_{ADJUSTED}$ is the discrepancy between $WQI_{ADJUSTED}$ and the actual WQI. This brings about the importance of assessing adequacy of WQI, while rarely explored and conducted, it is essential in answering the uncertainties associated with the development of $WQI_{ADJUSTED}$ **(Figure 4.1)**.

4.3.3 Aggregation of indices to produce a WQI

An additive aggregation method was used to aggregate the subindices, which takes the form of Conesa Fdez-Vitoria (1995) and Pesce & Wunderlin (2000)'s WQI formula **(Equation 7).** The magnitude of the WQI is proportional to the state of water quality, with the WQI ranging from 0 – 100. The qualitative classification for each variation of WQI is found in **Appendix A**.

$$WQI = \frac{\sum_{i=1}^{r} SI_i * W_i}{\sum_{i=1}^{r} W_i} \tag{7}$$

Where i = a water quality parameter, r = number of water quality parameters, SI_i = subindex of i, W_i = assigned weight of i

While the additive method has the limitation of 'eclipsing' (as described in the aforementioned section), this is circumvented with the introduction of Sub-WQI. If there is concern regarding specific groups of parameters (e.g. nutrients, heavy metals etc), water quality can be evaluated based on the Sub-WQI, which offers a more flexible and less resource-intensive form of WQI.

The Sub-WQIs were created by parameters grouped by FA – a method similar to PCA, which is often utilised for grouping water quality parameters with similar characteristics. Specifically, FA is a technique that groups variables based on their associations with an underlying latent variable known as factors (Liou et al., 2004; Hanh et al., 2011; Kline, 2014). The grouped parameters are especially useful for communicating water quality for a specified end use of interest (e.g. nutrients, heavy metals etc). Principal Axis Factor Analysis using the *psych* package in R was used to conduct the FA (Revelle & Revelle, 2015). Parallel analysis in the *psych* package was first conducted to optimise the number of factors extracted. The FA solution was then transformed obliquely (oblimin rotation method) to produce more interpretable factors (Kieffer, 1998; Osborne, 2015). An absolute threshold value of 0.6 in factor loadings was used to identify significant parameters with high factor loadings.

4.3.4 Assessing adequacy of WQI

This study aimed to introduce additional steps to the above framework paradigm as documented in **Chapter 2.2**. Assessment of WQI's performance with respect to WQI_{AVG} (WQI that entails all measured water quality parameters) and WQI_{MY} was quantified using determination of coefficient (R^2), root-mean-squared-error (RMSE) and Pearson's correlation coefficient. Parameter Sensitivity Analysis (SA) was conducted, where water quality parameters were removed one-at-a-time, and the change in WQI (absolute difference between mean WQI_{AVG} and mean WQI computed with the removal of a parameter) was quantified. Monte Carlo sampling was conducted using a bootstrap resampling approach to understand how different random sample size (conducted on sample data after all missing observations were removed) changes the mean of WQI. Bootstrapping is a convenient method that can simulate "field sampling" to obtain groups of sample data similar to the distribution of the original data that eliminates the need of repeating field sampling (McPeek &

Kalisz, 1993). This helps to provide a theoretical understanding of how WQI changes with sample size, and optimise the sampling frequency if field sampling for routine water quality monitoring within the JRB were to be conducted.

A percentage change between the mean of WQI from the full dataset ($WQI_{population}$) and the mean of WQI from the random samples (WQI_{sample}) was used as a metric of change in WQI. One-thousand simulations were conducted for each random sampling of specified sample size, and a probability of percentage change exceeding a specified threshold was estimated. This helps a user or a decision-maker to decide the amount of samples to collect, based on their available resources and their threshold of acceptability (i.e. the percentage change in WQI that one can tolerate). In addition, to understand how missing values affect the WQI_{MIN}, specified percentage of each water quality parameters in WQI_{MIN} (one-at-a-time) were randomly replaced with missing values to quantify changes between the mean of $WQI_{ADJUSTED}$ and WQI_{MIN}, and bootstrapping of one-hundred simulations were used. Similarly, a probability of percentage change exceeding a specified threshold was estimated.

The water quality standards – WQI classes (Class 1 to Class 4, with Class 5 implicitly declared as values lower than Class 4) were calculated using the aforementioned WQI method, where parameter values were defined by the threshold limits stipulated by the respective class categories specified by the National Water Quality Standards For Malaysia (NWQS). Where water quality standard values are lower than the minimum value of the breaks or water quality parameters are not specified (i.e. parameters with no known toxicity effects regardless of its concentration e.g. Na, Mg, Ca, K), its SI will be assigned 100, the converse is true for values higher than the maximum value of the breaks. Weighted and non-weighted variations of WQI classes were also considered. Where parameters are absent from the list of parameters measured, they are replaced with the water quality standards by USEPA **(Appendix A(a)).**

4.4 Land use-water quality analysis

Land use data was obtained from GeoJohor (PLANMalaysia@Johor, 2017). The land use categories were aggregated and reclassified to eight main categories – Agriculture (others) (AGRL), Agriculture (Oil Palm) (OILP), Bare land and shrubs vegetation (BSVG), Forest (FRSE), Urban (Residential) (RESD), Urban (others) (URBN), Urban (Industry) (UIDU), and Water bodies (WATR) (**Table** 4.2).

Table 4.2. Description of land use classes

Land use category	Land use	Abbreviation of land use	Description
1	Agriculture (others)	AGRL	Agriculture activities other than oil palm and animal agriculture/husbandry
2	Bare land and shrubs vegetation	BSVG	Bare land, empty land, shrubs
3	Urban (Industry)	UIDU	Low to high activity industries
4	Forest	FRSE	Natural forest
5	Agriculture (Oil Palm)	OILP	Oil palm agriculture (small-scale, private and commercial plots)
6	Urban (Residential)	RESD	Low, medium and high density residential unites
7	Urban (others)	URBN	Urban infrastructure including commercial, institutional, public facilities and roads
8	Water bodies	WATR	Water bodies including river and reservoirs

Five different land use metrics - percentage of land use types (PLAND); landscape shape index (LSI); percentage of like adjacencies (PLADJ); aggregation index (AI) of each land use type, and Shannon's Diversity Index (SHDI) were calculated to quantify land use configuration for each sampling site under three different spatial scales – Reach, Riparian, and Sub-basin scale. These metrics were computed for each landscape scale using FRAGSTATS 4.0 – a spatial analysis software for analysing landscape configuration (McGarigal & Marks, 1995; McGarigal, 2015) (**Table** 4.3). Sub-basins within the JRB were obtained using QSWAT (Version 1.9), a QGIS (Brighton 2.6) interface of Soil and Water Assessment Tool (SWAT), derived using a Digital Elevation Model (DEM) of one arc second (Earth Explorer, USGS). The sub-basins were created using sampling points as the outlet of the sub-basins. The riparian and reach scale are defined as the 500m buffer zone around the river and the sampling point that is enclosed by the sub-basin respectively (**Figure** 4.2). The buffering was conducted using QGIS buffering tool.

Table 4.3. Description of each land use metric and WWTP variables. Modified from McGarigal & Marks (1995); McGarigal (2015)

Metric type	Variables	Description	Equation
Class metrics	PLAND (%)	Percentage of land use types. It is a measure of landscape composition	$PLAND = P_i = \frac{\sum_{j=1}^{n} a_{ij}}{A}(100)$ P_i = Ratio of the landscape occupied by land use i. a_{ij} = Area (m^2) of land use ij. A = Total area of landscape (m^2).
	LSI (unitless)	Landscape Shape Index LSI = 1 when the landscape consists of a single square patch of a land use type; LSI increases as irregularity of landscape increases.	$LSI = \frac{0.25 \sum_{k=1}^{m} e_{ik}^{*}}{\sqrt{A}}$ e_{ik} * = Total length (m) of edge in landscape between land use types i and k; A = Total area of landscape (m^2).
	PLADJ (%)	Percentage of Like Adjacencies The proportion of adjacent cells belonging to the same class. PLADJ equals 0 when there is no adjacent patch of the same land use type.	$PLADJ = \left(\frac{\sum_{i=1}^{m}(g_{ii})}{\sum_{i=1}^{m}\sum_{k=1}^{m}(g_{ik})}\right)(100)$ g_{ii} = Number of same adjacent land use type i cells (double count method) g_{ik} = Number of adjacent cells of same land use type i and k (double count method)
	AI (%)	Aggregation Index AI equals the number of same adjacent land use type, divided by the maximum possible number of same corresponding adjacent land use type.	$AI = \left[\sum_{i=1}^{m}\left(\frac{g_{ii}}{max \rightarrow g_{ii}}\right)P_i\right](100)$ g_{ii} = Number of same adjacent land use type i cells (single-count method)

		AI equals 100 when the same land use is aggregated into a single patch	$max\text{-}g_{ii}$ = Maximum number of same adjacent land use type i cells (single-count method) P_i = Ratio of the landscape occupied by land use i.
Landscape metric	SHDI (unitless)	Shannon's Diversity Index SHDI = 0 when the landscape contains a single land use. SHDI increases as the number of different land use increases.	$SHDI = -\sum_{i=1}^{m}(P_i * lnP_i)$ P_i = Ratio of the landscape occupied by land use i
Point source	NUM (unitless)	Number of WWTPs within a sub-basin	$NUM = \sum WWTP$ WWTP = Number of WWTP in a sub-basin
	DESIGNPE (unitless)	Design Population Equivalent of WWTP	$DESIGNPE = \sum_{i=1}^{n} PE_i$ n = Number of WWTPs in a sub-basin PE_i= Population equivalent of each WWTP
	NUM.den.wt (number/m^2)	Weighted Number of WWTP per unit area of sub-basin	$NUM = \frac{1}{A}\sum WWTP * wt$ WWTP = Number of WWTP in a sub-basin Wt = Weightage corresponding to discharge standards of WWTP. Discharge standard A= 1; Discharge standard B= 2 A = Area of sub-basin
	DESIGNPE.den.wt (m^{-2})	Weighted sum of Population Equivalent of WWTP per unit area of sub-basin	$DESIGNPE = \frac{1}{A}\sum_{i=1}^{n} PE_i * wt$ n = number of WWTPs in a sub-basin

	PE_i= Population equivalent of each WWTP wt = weightage corresponding to discharge standards of WWTP. Discharge standard A= 1; Discharge standard B= 2 A = Area of sub-basin

Notes: Cell = Pixel of a raster; Patch = homoegenous areas within the landscape; Class = categories of land use type

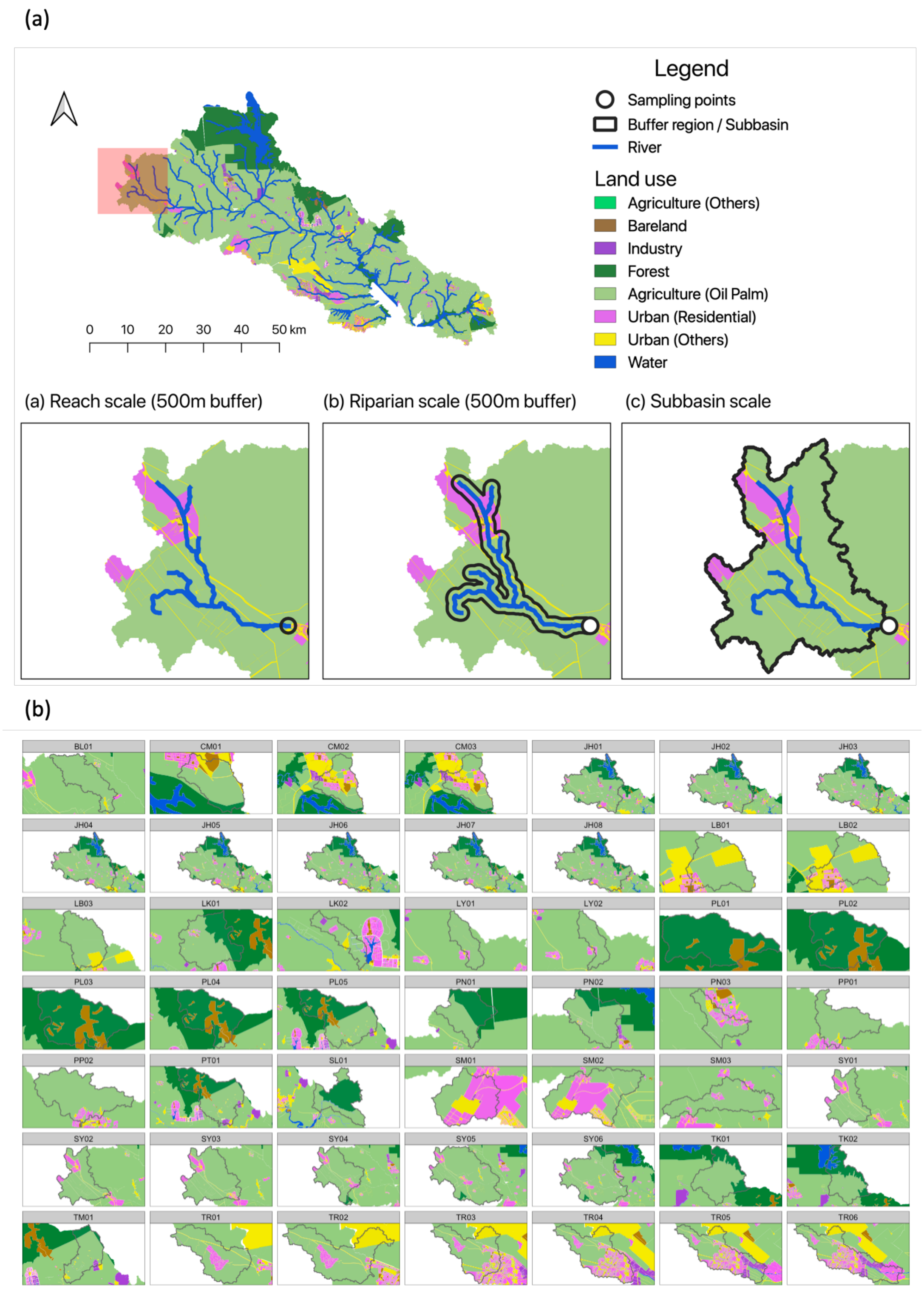


Figure 4.2. (a) Extraction of land use variables at three different scale (b) Land use within each sub-basin

Two separate scenarios were created: (i) models with only land use metrics as explanatory variables (referred to as land use only scenario hereafter) (ii) models with land use metrics and point source variables as explanatory variables (referred to as combined scenario hereafter). The scenarios were compared to quantify the influence of point sources on water quality, on top of land use influence.

Point source variables were obtained from known point sources – Waste Water Treatment Plants (WWTPs) operated by the Indah Water Konsortium (IWK), a state-owned company that manages the sewerage system for Peninsular Malaysia. Two main point source variables were obtained – the number of WWTPs found within each sub-basin (It is not necessary to compute point source variables for each scale as the WWTPs are all discharged directly into the river), and the Design Population Equivalent (PE) of each WWTP. The WWTPs were also categorized by their discharged standards (A or B, with B having poorer water quality in the effluent). The two variables (number of WWTP and Design PE) were transformed to account for the density of WWTPs per unit area of the sub-basin and type of discharge standards (discharge standard A and B were given weights one and two respectively) (**Table** 4.3). The WWTPs in Malaysia goes through preliminary, primary, and secondary sewage treatment where suspended solids and organic matter are removed through sedimentation and biological unit processes. However, tertiary sewage treatment which entails the removal of nutrients and toxic substances like heavy metals have yet to be implemented in Malaysia (IWK, 2019c).

Two different seasonality, namely, the dry and wet season are defined as river discharge having lower and higher flow than mean discharge respectively (**Figure 3.1 d**). As such, samples collected from February to October, and November to December, were classified as samples collected during the dry season and wet season, respectively. As water quality varies temporally from month to month, the water quality parameters are averaged for each season to reduce the effect of monthly variation on water quality.

Monthly water quality trends were evaluated by calculating exceedance rate – $E = \frac{n_i}{N_i} * 100\%$, where E = exceedance rate, n_i = number of sampling sites exceeding National Water Quality Standards For Malaysia (Grade IIA/IIB), N_i = total number of sites.

In summary, five major sequence was conducted at each scale and scenario to identify the corresponding significant land use variables for each water quality parameters:

1. Stepwise Multivariate Linear Regression (MLR) was conducted to identify significant land use variables
2. Variance Inflation Factor (VIF) was used to assess multicollinearity and highly correlated variables were removed
3. Model is refitted after removing correlated variables
4. Redundancy Analysis (RDA), a constrained ordination of Principle Component Analysis (PCA), was conducted to examine relationships between significant land use metrics (from the MLR results), water quality parameters and sampling sites
5. Permutation test was carried out on the RDA models to test for the validity of the models

Stepwise MLR models were used to model the relationship between each water quality parameter (response) and the land use metrics (predictors). Prior to applying the stepwise MLR, the data was log(x+1) transformed to normalise the residuals and to reduce heteroscedascity. Performance of the models was determined using the adjusted coefficient of determination to penalise models with more predictor variables (adjusted R^2). Multicollinearity for each model was assessed using the VIF. A threshold value of 10 for VIF was used to assess multicollinearity within variables and the robustness of response variable. A value below 10 for VIF is deemed to be inconsequential in collinearity (Hair et al., 1998; O'Brien, 2007). In the case where the VIF for the variables exceeded 10, the variable with the highest VIF value was subsequently removed, and the multivariate linear regression was refitted until the VIF for all variables were below 10. This eliminates the need of entirely removing a land use metric due to correlation with other metrics of the same land use, while simultaneously accommodates multiple types of land use metrics (Carey et al., 2011).

Standard partial coefficients (B) for the significant variables was also obtained to aid in the quantitative interpretation of land use influence on water quality.

The stepwise MLR was repeated for all landscape scales, and the best MLR was selected based on the highest R^2 for that landscape scale. To constrain the number of variables used in RDA, the significant predictors (p-value<0.01) from the results of MLR were used as the explanatory variables in the RDA. Scatterplots (predicted against observed) were used to evaluate the performance of the MLR models **(Appendix I).**

RDA combines MLR and PCA by computing axes that are linear combinations of the explanatory variables (land use metrics). It is a constrained form of PCA as only the significant explanatory variables are tested. RDA function from the *vegan* package in R was used to create the triplots (Oksanen et al., 2019). The triplots aid in the interpretation of relationships between significant land use metrics, water quality variables and sampling sites. Variables are correlated when their arrows point in the same direction. The axes of the triplots provide information on the variance explained by the explanatory variables. A Monte Carlo permutation test (999 permutations) was conducted using *vegan* package from R to assess the significance of all terms and validity of the constrained ordination using RDA. The permutation involves random permutation of the data and refitting the model. When permutated constrained inertia is lower than observed constrained inertia, the constraints are deemed to be significant (Oksanen et al., 2013).

5. RESULTS AND DISCUSSION

5.1 Characterisation of water quality in the JRB

Based on the mean and median values of the measured water quality parameters from the sixteen river branches within the JRB, most of the measured values were within NWQS's water quality standards **(Table 5.1)**. The NWQS's Class 2 water quality limits were used as a reference for calculating each water quality parameters' exceedance rate. However, the exceedance rate is merely a water quality parameter-specific indicator that is benchmarked against its corresponding stipulated water quality standard, and may not necessarily be representative of the general state of water quality. The general state of water quality is assessed based on the calculated WQI – a numerical value calculated from the composite of significant water quality parameters, each benchmarked against NWQS's water quality limits. The specific river classes of the study sites will be classified based on the WQI in the subsequent sections.

Twenty out of thirty-four measured water quality parameters have an exceedance rate of more than 0%. In particular, the parameters of interest are Cu, DO, and E.coli, with more than half of the observations exceeding water quality standards (66.1%, 85.4%, and 66.1%, respectively). Also noteworthy are COD (11.3%), Coliform (total) (49.4%), NH3-N (36.2%), pH (30.9%), Temperature (20.4%), and Turbidity (21.6%). With the exception of Mn (8.9%), all the other parameters have an exceedance rate of under 3%.

These water quality parameters have varying effects on water quality and its intended usage. Cu may originate from anthropogenic and natural sources. Anthropogenic sources of Cu includes industrial emissions, effluent from wastewater treatment plants, runoff from roads, and agricultural pesticides (Boulay & Edwards, 2000; Shrivastava, 2009). The high Cu concentration observed may be associated with high input of fertilisers due to the extensive oil palm plantation land use in the JRB as described in the findings by Rakib et al. (2017). It is unclear the source of copper within the JRB but based on literature, there are few key sources. Copper compounds such as copper sulfate pentahydrate, copper enolate, and copper citrate may be added to waters as they serve as a biocide for preventing algae growth

(National Research Council, 1977). This is often employed in eutrophic waters when algae and plankton causes problems such as eutrophication, tastes and odour, which affects the acceptability of drinking water produced from such waters (Boulay & Edwards, 2000). While copper is also an essential micronutrient to human and aquatic organisms, copper toxicity at elevated levels may result in deleterious ecological impacts especially for sensitive species. For instance, at a concentration of 0.01 – 0.02 mg/l, Cu toxicity effects were observed in some fish species (National Research Council, 1977). NWQS describes that rivers of Class 2 (e.g. Johor River) are suitable for sensitive fishery aquatic species **(Appendix A(c)).** With some of the observations exceeding the standard stipulated by Class 2, more research will be required to investigate Cu toxicity tolerance for freshwater fishes within the JRB.

More than 80% of the observations have low DO values, which is typical of tropical rivers due to its high Temperatures and productivity (Davies Jr et al., 2008). More than 60% of the observations have elevated E.coli levels and Coliform (total) values. This phenomenon indicates relatively extensive sources (point and/or diffuse) of faecal contamination. This could be attributed to wastewater treatment plants (WWTPs) from urban and residential areas, and/or animal agriculture which at times exists as livestock-oil palm integrated system in Johor (MNRE, 2011; Gabdo & Abdlatif, 2013).

Apart from E.coli and Coliform (total), Mn, NH_3-N, PO_4^{3-}, and Turbidity are parameters of concern which have maximum values exceeding the NWQS's Class 2 water quality standard by several orders of magnitude. High levels of NH_3-N and PO_4^{3-} may be attributed to diffuse sources of nutrient run-off, originating from fertilisers applied on surrounding agricultural lands or point sources such as WWTPs (Pocernich & Litke, 1997; Berka et al., 2001). In addition, sources of high Turbidity may originate from plantation or animal agriculture, which can increase soil erosion due to improper agricultural practices such as ploughing or grazing by animals (Sidle et al., 2006; Giri & Qiu, 2016; Mello et al., 2018). On the other hand, elevated levels of Mn may be of anthropogenic or natural origin (Richardson, 2017).

Table 5.1. Summary of statistics for each water quality parameters of the studied rivers in the JRB

Parameter	N_{obs}	Missing values (%)	Censored values (%)	Detection limit[a]	Median[b]	Mean[b]	SD[b]	Min	Max	Class 2 Standard[c]	Exceedance rate[d] (%)
Ag (ppb)	425	6.39	26.40	0.05	0.07	0.23	0.64	BDL	8.92	50	0.00
Al (ppb)	425	6.39	8.71	1.00	3.67	71.90	515.00	BDL	8930.00	1500	1.18
As (ppb)	425	6.39	0.00	0.005	0.68	1.08	1.23	0.04	11.80	50	0.00
Ba (ppb)	425	6.39	0.00	0.05	84.30	98.80	72.10	1.44	299.00	1000	0.00
Br (ppm)	449	1.10	70.40	0.10	0.09	0.19	1.11	BDL	21.60	-	-
Ca (ppm)	450	0.88	0.00	0.05	5.35	6.92	7.69	0.45	115.00	400	0.00
Cd (ppb)	425	6.39	3.29	0.02	0.02	0.12	0.71	BDL	8.81	10	0.00
Cl (ppm)	450	0.88	0.00	0.10	13.40	38.80	232.00	1.60	3820.00	200	1.56
Co (ppb)	425	6.39	0.00	0.005	0.11	0.83	4.53	0.01	72.60	-	-
COD (ppm)	425	6.39	7.06	0.50	11.30	13.70	13.00	BDL	89.20	25	11.30
Coliform (total) (MPN/100ml)*	387	14.80	49.90	>2419.6 >24196 >241960	-	-	-	41.00	242000	5000	49.40
Cr (ppb)	425	6.39	5.18	0.01	0.11	0.27	1.20	0.01	18.90	50	0.00
Cu (ppb)	425	6.39	0.00	0.50	32.40	41.00	45.30	1.01	460.00	20	66.10
DO (mg/l)	426	6.17	0.00	0.01	3.97	3.86	1.13	0.20	6.57	6	85.40
E.coli (MPN/100ml)*	387	14.80	11.40	<10 >2419.6 >24196 >241960	-	-	-	1.00	242000	400	66.10
Fe (ppb)	425	6.39	2.12	1.00	4.33	70.50	269.00	BDL	3380.00	1000	1.88
F^- (ppm)	450	0.88	97.60	0.10	0.07	0.08	0.37	BDL	4.41	1.5	0.22
K (ppm)	450	0.88	0.00	0.05	3.12	4.54	5.73	0.43	58.30	-	-
Mg (ppm)	450	0.88	0.00	0.05	1.26	2.83	15.10	0.23	263.00	-	-
Mn (ppb)	425	6.39	0.00	0.01	26.10	52.40	96.20	0.11	749.00	100	8.94

Na (ppm)	447	1.54	0.00	0.05	3.59	13.90	97.60	1.07	1840.00	-	-
NH_3-N (ppm)	426	6.17	1.41	0.015	0.15	1.28	4.87	0.015	80.40	0.3	36.20
Ni (ppb)	425	6.39	0.00	0.01	0.44	3.40	23.80	0.04	414.00	50	0.94
NO_2-N (ppm)	426	6.17	23.90	0.01	0.01	0.04	0.10	BDL	0.66	0.4	1.88
NO_3^- (ppm)	446	1.76	5.61	0.10	2.24	2.81	10.50	BDL	219.00	7	2.91
Pb (ppb)	425	6.39	1.41	0.10	0.08	0.21	0.67	BDL	8.75	50	0.00
pH	450	0.88	0.00	n.a.	6.37	6.29	0.60	3.67	8.21	6	30.90
PO_4^{3-} (ppm)	450	0.88	92.40	0.10	0.03	0.08	0.25	BDL	2.08	0.7	2.22
Se (ppb)	425	6.39	0.71	0.10	0.05	0.10	0.42	BDL	6.72	10	0.00
SEC (uS/cm)	426	6.17	0.00	0.10	90.20	200.00	774.00	21.30	11600.00	1000	1.88
SO_4^{2-} (ppm)	450	0.88	0.00	0.10	4.80	9.77	31.50	0.24	520.00	250	0.44
Temperature (°C)	426	6.17	0.00	n.a.	27.90	27.90	1.78	24.50	33.70	29	20.40
Turbidity (NTU)	445	1.98	0.00	0.30	19.30	64.30	201.00	1.23	2680.00	50	21.60
Zn (ppb)	425	6.39	0.00	0.50	35.60	45.40	87.80	1.80	1630.00	5000	0.00

Notes: N_{obs} = number of observations, SD=standard deviation
[a] Detection limit reported for parameters with censored value
[b] Kaplan-Meier model with a confidence limit of 0.95 was used to calculate summary statistics for left-censored data
[c] Refers to NWQS's Class 2 water quality standard as Johor river is classified as a Class 2 river (Ahmad, 2016). More information on NWQS's water quality standard is found in **Appendix A**
[d] Exceedance rate is the percentage of observations exceeding NWQS's Class 2 water quality standard
BDL = Below Detection Limit
- Water quality standard is not specified for the water quality parameter
*Mean, Median and SD statistics were unable to be calculated reliably due to multiple left and right censored data (due to different levels of dilution)

5.2 Temporal and spatial patterns in WQI

It is not surprising to find that, generally, different sites from the same river demonstrates similarity in water quality as shown in the hierarchical clustering **(Figure 5.1a).** Spatial clustering of sites also shows seasonal variability **(Figures 5.1b and 5.1c)**. Sites that show deviance outside of the clustering could indicate potential point or diffuse sources of pollution. As an example, CM03 shows consistent distinct deviation from the rest of the sites, which may indicate strong external forcing (e.g. pollution, surface change, environmental conditions) on the in-situ water quality. TR01 and TR02 also show consistent deviation from the rest of the TR sites, which shows explicitly the location where water quality begins to differ, and indicates sites that experience point or diffuse source of pollution.

This spatial pattern is also exhibited in **Figure 5.2** where WQI for TR01 and TR02 is vastly different from TR03 onwards, and WQI generally decreases downstream where TR05 and TR06 exhibits poor WQI. CM sites also exhibit the poorest WQI, which is corroborated by the spatial clustering in **Figure 5.1**.

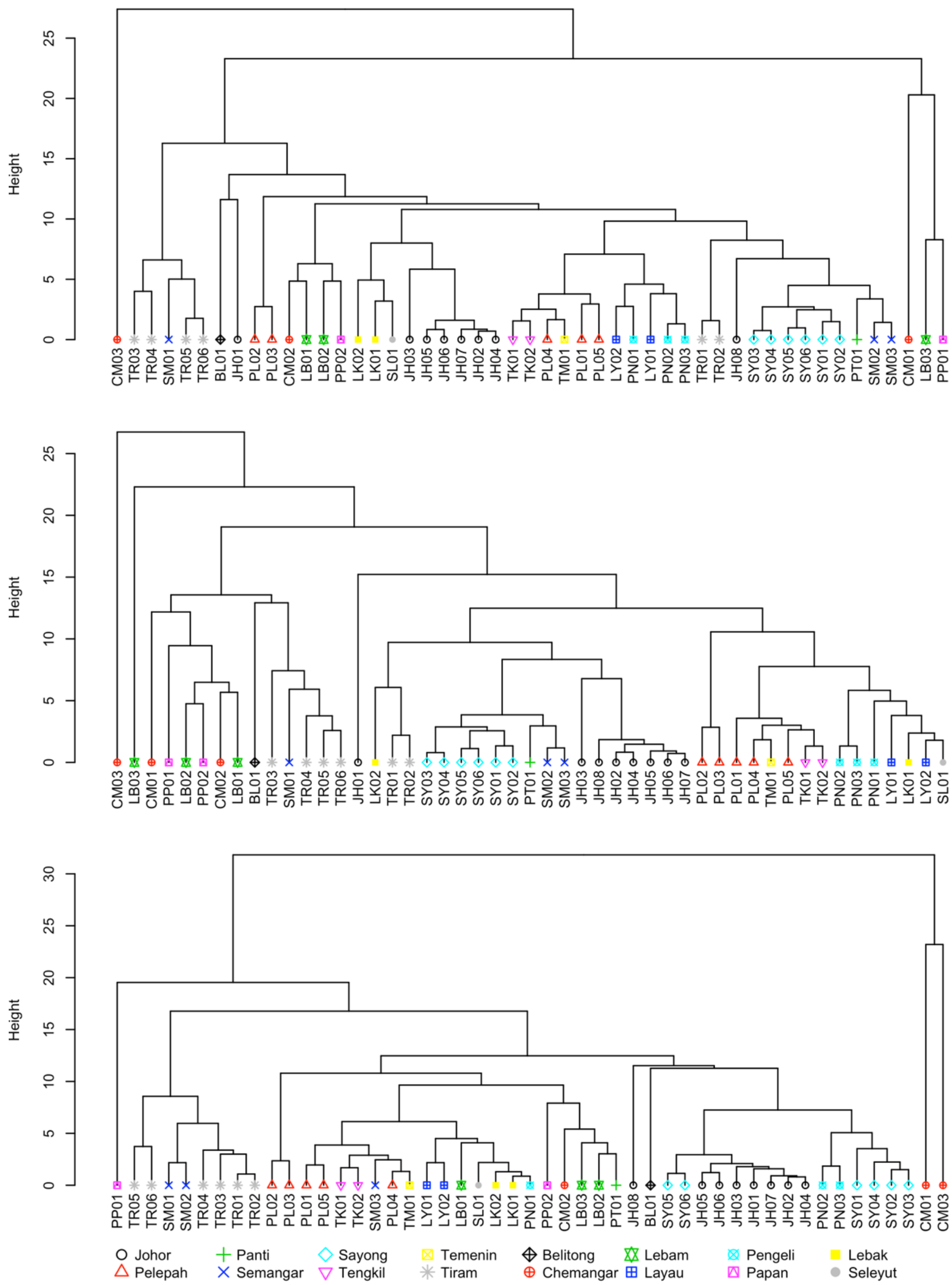


Figure 5.1. Hierarchical clustering of sites based on Euclidean dissimilarity in all measured water quality parameters. (a) Full set of data (b) Dry season (c) Wet season.

Notes: Height (y-axis) is a measure of similarity between individual points or clusters

Figure 5.2 and **5.3** show the spatio-temporal variations in WQI_{AVG}. Most of the observations lie within Class 2, which corroborates with the current classification of Johor River. More than 99% of the observations are classified as Class 2 and less than five observations are classified as Class 4 and above. $WQI_{NON\text{-}WEIGHTED}$ has more sites classified as Class 4 and above as compared to $WQI_{WEIGHTED}$, although there is not much difference to the relative proportion **(Appendix E)**. Chemangar (CM) River is the only river that has observations exceeding Class 2, while the rest of the sites were well within the range of Class 1 and Class 2. Overall, sites from Chemangar River, located downstream of the Singapore's Water Treatment Plant, has the lowest WQI, which corroborates the observations in **Figure 5.1**, indicating strong pollution source at CM. Differences in water quality parameter between wet and dry seasons are not distinct. However, the variability of WQI for each site is relatively consistent across all months **(Figure 5.3)**. For instance, Chemangar (CM), Lebak (LK), Lebam LB), Pengeli (PN), Semangar (SM), and Tiram (TR) exhibit relatively large variation in WQI throughout the months although the number of sites for each river are relatively few (except for Tiram). This is in contrast with the small variation in WQI exhibited in sites such as Johor (JH), Pelepah (PL), and Sayong (SY) rivers. As such, variation in WQI also reveals the differences in water quality between sites belonging to the same river. A large variation in WQI suggests that at least one of the sites in the same river branch has distinctly different water quality from the rest of the sites, which may indicate evidence of pollution. This is also corroborated by the spatial variability in water quality shown in **Figure 5.2**.

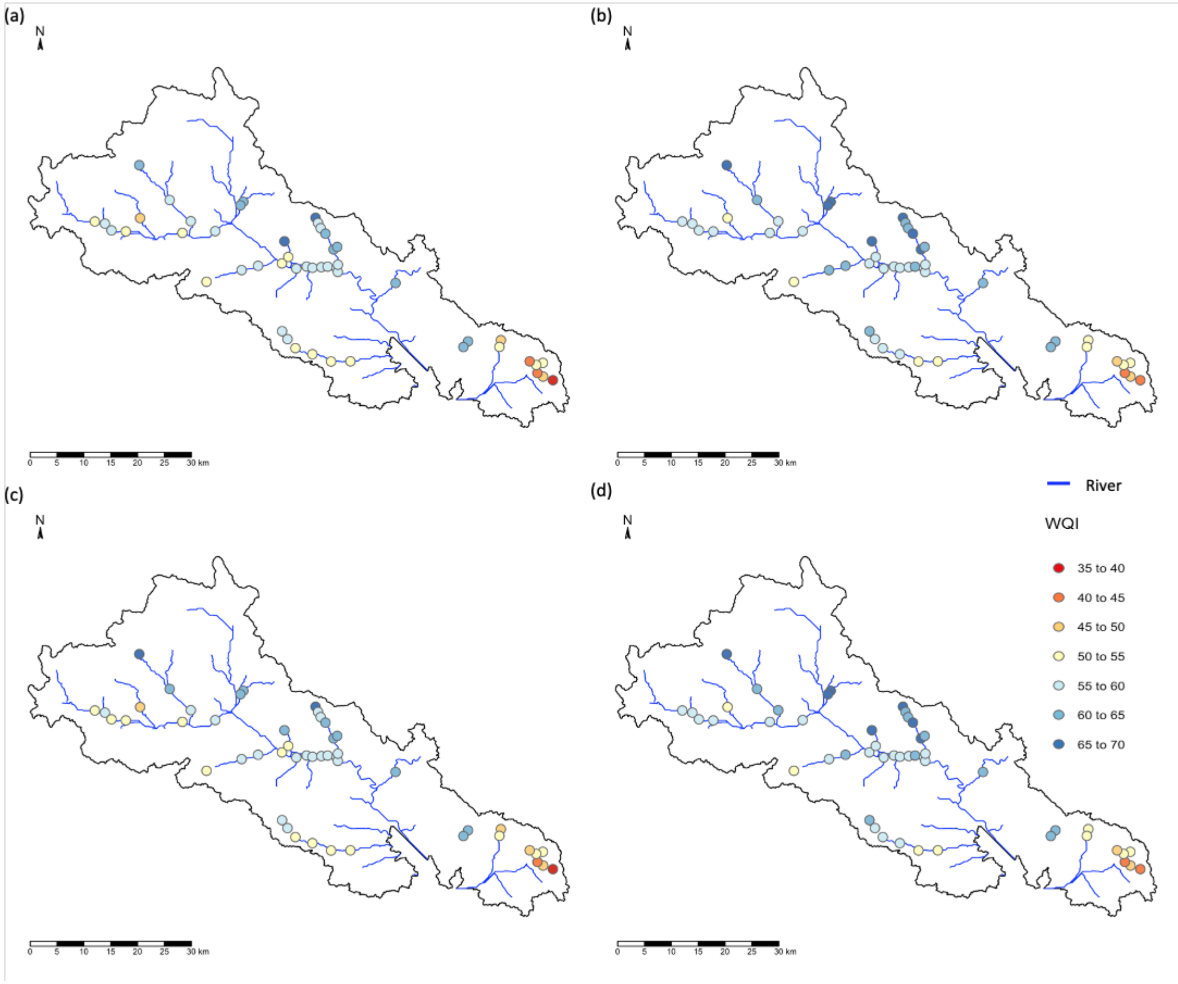


Figure 5.2. Spatial variability of the study sites' water quality based on WQI (a) $WQI_{WEIGHTED}$ (b) $WQI_{NON\text{-}WEIGHTED}$ (c) $WQI_{WEIGHTED\ (ADJUSTED)}$ (d) $WQI_{NON\text{-}WEIGHTED\ (ADJUSTED)}$

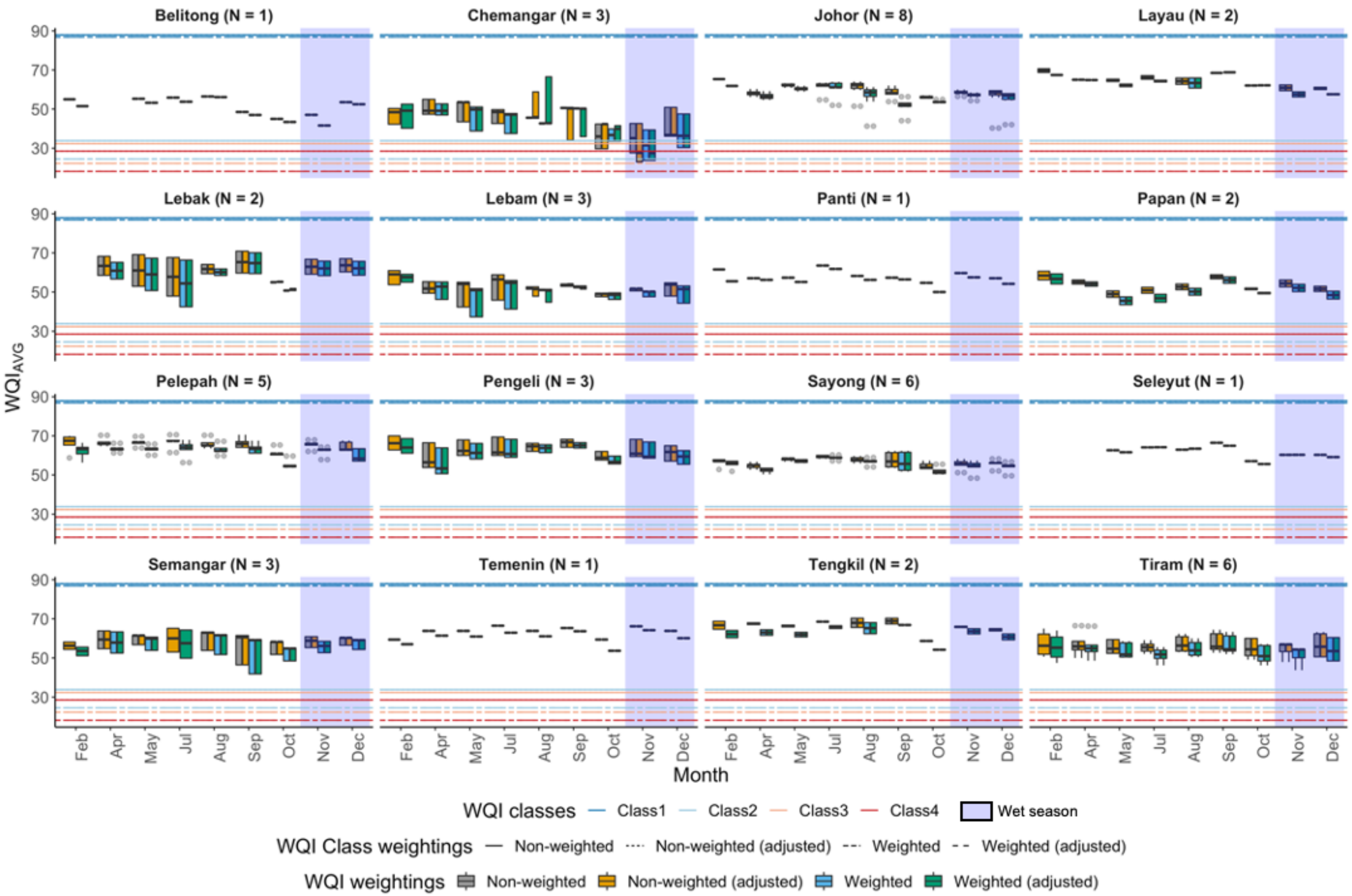


Figure 5.3. Spatio-temporal changes in WQI_{AVG} under different water quality standards.

Notes: Boxplots without adjusted WQI have no missing values. N refers to the number of sites collected for each river per month

5.3 Key water quality parameters

Multivariate Linear Regression (MLR) and Principal Component Analysis (PCA) were employed as methods to elucidate key parameters, and serves as parameter reduction techniques for constructing a WQI using minimum number of parameters (WQI_{MIN}). Through these two methods, MLR and PCA produces WQI_{MLR} and WQI_{PCA}, respectively.

It was interesting to observe that different significant water quality parameters were selected through different methods. The key parameters selected by MLR were more varied while the PCA method selects key parameters which were more inclined towards heavy metals. The common key parameters selected by both methods were Ca, Mg, Fe, Coliform (total), E.coli, NH_3-N, Cu, Cr **(Table 5.2)**. These parameters explain most of the spatio-temporal variation in water quality compared to other parameters.

The key factors selected by MLR and PCA were subsequently grouped together by Factor Analysis (FA), a technique that groups variables based on their associations with an underlying latent variable known as factors (Kline, 2014).

Some key parameters grouped by FA appear to be logical and fit what is commonly observed. For instance, common ions (e.g. Na, Mg, K, Ca), and E.coli and Coliform (total) are grouped together for both MLR and PCA methods **(Table 5.2)**. However, key parameters grouped by FA are different for both MLR and PCA methods, which may call for a different interpretation for the grouped parameters, despite some commonalities observed. For instance, the MLR method only selects Ca and Mg (Factor 4_{MLR}) which are good indicators for water hardness. On the other hand, the PCA method selects other ions, including Ca and Mg (Factor 4_{PCA}), which are good indicators for salinity or conductivity. Another curious observation is the grouping of Fe with NH_3-N and NO_2-N using the MLR method (Factor 1_{MLR}). One possible explanation could be the function of Fe in the redox processes of nitrogen transformation regulated by the pH in soil environments. These compounds in soil inevitably ends up in surrounding water bodies, as soil acts as a medium that interacts with surface water bodies such as rivers (Huang et al., 2016). COD, Cu, Zn are grouped together as Factor 3_{MLR}, which could be attributed to the complexation or adsorption of Cu and Zn to organic matter (McGrath et al., 1988; Fageria et al., 2002; Lim et al., 2012; Rakib et al., 2017). Organic matter subsequently enters surface water primarily through surface runoff and erosion. As such, Factor 3_{MLR} could possibly indicate leaching of micronutrients through erosion and even identify surrounding regions with deficiency in these micronutrients.

Table 5.2. Factor Analysis (FA) of key parameters selected by Multivariate linear regression (MLR) and Principle Components Analysis (PCA)

Method	Key parameters grouped by Factor Analysis								Key Parameters
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Combined parameters
MLR	Fe, NH3-N, NO2-N	Coliform (total), E.coli	COD, Cu, Zn	Ca, Mg	pH	NO3	Cr, Se		Fe, NH3-N, NO2-N, Coliform (total), E.coli, Ca, Mg, Cr, pH, COD, Cu, Zn, NO3
PCA	NH3-N	Coliform (total), E.coli	Cu	Na, Ca, Mg, Cl, K	As	Al, Fe	Cr	Co, Cd	Na, Ca, Mg, Cl, K, Co, Cd, Coliform (total), E.coli, Cu, NH3-N, As, Al, Fe, Cr

Notes: Details of each key parameters selected by each method (MLR and PCA) under different scenario can be found in **Appendix F**

5.4 Evaluating Sub-WQI and WQI_{MIN}

The Sub-WQIs serve as an alternative form of WQI that quantifies the water quality represented by various groups of key parameters that explains specific aspects of water quality. As such, key parameters grouped by FA serves as an important precedent and underpins the creation of Sub-WQI. The Sub-WQI is of great utility in the context of obtaining water quality information of a specific end-use or investigating a certain phenomenon with respect to the in-situ water quality. It can also reveal aspects of water quality that may be masked in WQI_{MIN} due to the 'eclipsing' effect – an inherent limitation in the additive aggregation method.

The factors exhibit variations in Sub-WQI values traversing across different water quality classes (Class 1 to Class 4), with some factors having Sub-WQI values ranging from 3 to 100 and others ranging from 30 to 100. Generally, the Sub-WQI_{MLR} of the factors are within the range of Class 1 and Class 4, with Factor 7_{MLR} (Cr, Se) as the only factor with Sub-WQI_{MLR} within the range of Class 1 and Class 2 **(Figure 5.4a).** Similarly, the Sub-WQI_{PCA} for Factor 3 (Cu), 5 (As), 6 (Al, Fe), 7 (Cr), and 8 (Co, Cd) are within Class 1 and Class 2, with the exception of Factor 3 (Cu) having values between Class 1 and Class 4. This demonstrates that heavy metal pollution is not a major problem in rivers of the JRB.

Some factors of Sub-WQI exhibit distinct spatial and temporal variability **(Figure 5.4)**. For instance, Factor 1_{MLR} and Factor 1_{PCA}; Factor 2_{MLR} and Factor 2_{PCA} (which includes NH3-N; E. coli, and Coliform (total), respectively) consistently shows low Sub-WQI_{MLR} and Sub-WQI_{PCA} values for Chemangar, Lebam, and Tiram rivers, where these three rivers are all located downstream of Singapore's Water Treatment Plant (for both dry and wet season). The low Sub-WQI values are reflected in the low WQI_{MIN} values for these sites as well **(Figure 5.4 & Figure 5.5)**. This shows that these sites experience high ammonium pollution and faecal contamination, with most observations exceeding Class 4. Factor 4_{MLR} (Ca and Mg) and Factor 4_{PCA} (Na, Ca, Mg, K, and Cl) are indicators for hardness and ions, respectively, and also show similar spatial variability for Sub-WQI_{MLR} and Sub-WQI_{PCA}, which indicates that hardness may not be a reliable indicator when the Sub-WQI_{PCA} is high or the water sample has higher salinity (Adey & Loveland, 2007).

Seasonal variation in water quality can also be observed for some sites and factors, especially for Factor 7_{MLR} (Cr, Se) **(Figure 5.4a)** and Factor 7_{PCA} (Cr) **(Figure 5.4b)** in sites such as Belitong and Chemangar rivers, where wet season resulted in significantly lower Sub-WQI values. Sub-WQI_{PCA} values were also lower for Factor 6_{PCA} (Al, Fe) and Factor 8_{PCA} (Co, Cd) during wet season for Chemangar and Papan rivers **(Figure 5.4b)**. It appears that Chemangar River is the common denominator in sites that were observed to have low Sub-WQI values for factors associated with heavy metals, which could indicate stronger point or diffuse sources of heavy metals at Chemangar compared to other sites. This is also supported from field observations at Chemangar sites where one of the sites – CM02 is situated near an industrial factory.

It is also interesting to observe that Factor 3_{PCA} (Cu) concentration is vastly different across all sites during the wet season compared to the dry season, with most values falling between Class 3 and Class 4 during the wet season **(Figure 5.4a)**. Factor 2 (Coliform (total) and E. coli) showed lower Sub-WQI values during the wet season for some of the rivers such as Tiram, Chemangar, and Lebam **(Figure 5.4a & Figure 5.4b)**. This is because intense rain and large volumes of runoff during the wet season are important processes in loading of pollutants, particulates, and nutrients into surrounding water bodies (Heinonen-Tanski & Uusi-Kämppä, 2001; Lee et al., 2002; Li et al., 2007; Nazahiyah et al., 2007; Chuah & Ziegler, 2018).

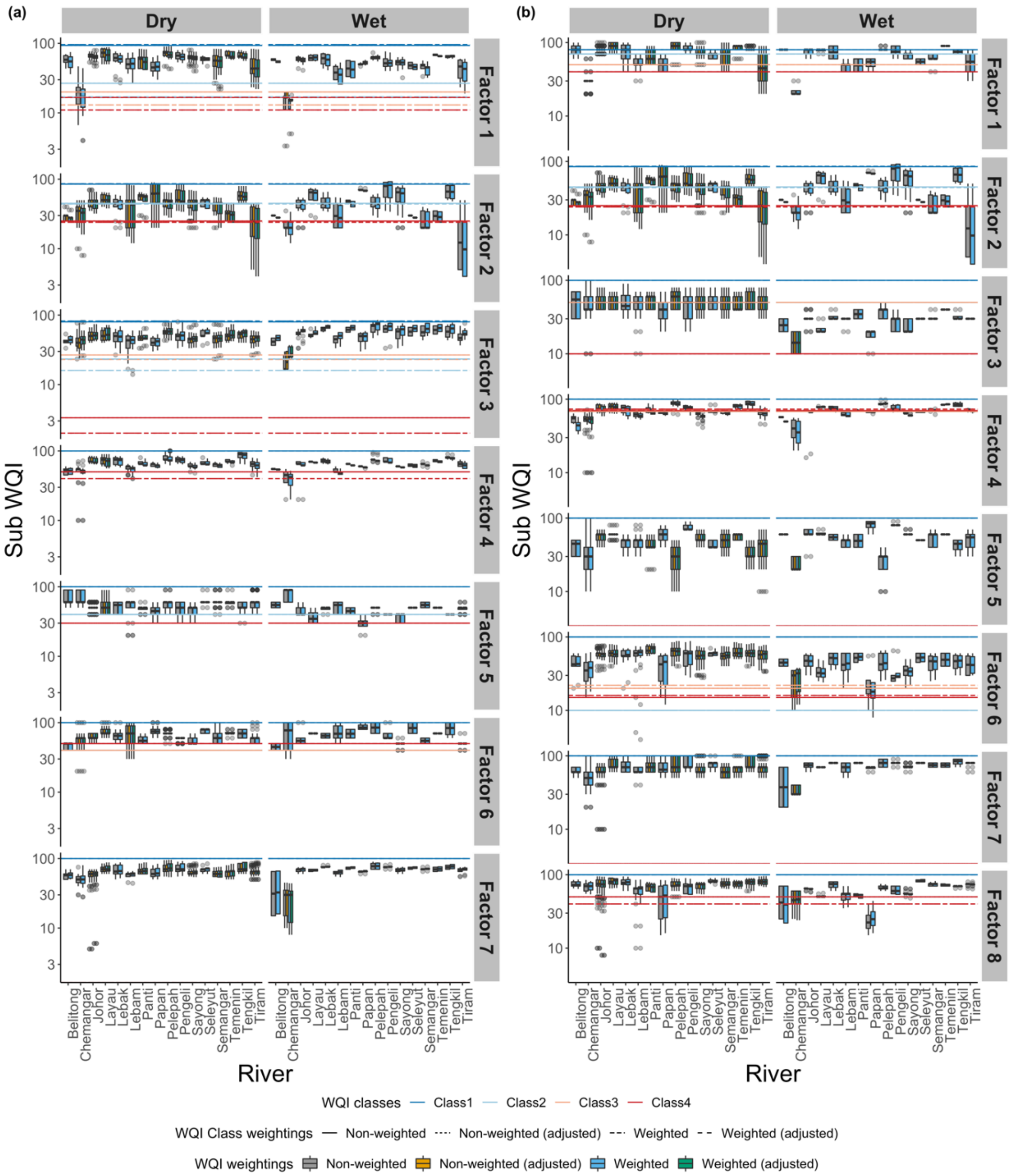


Figure 5.4.Spatial variation in SUB-WQI for each factor under different seasonality (a) SUB-WQI$_{MLR}$ (b) SUB-WQI$_{PCA}$.

Notes: boxplots without adjusted WQI means that observation has no missing values

WQI$_{MIN}$ shows the composite effect of all the Sub-WQI factors **(Figure 5.5)**. Both WQI$_{MIN}$ demonstrate similar spatial variations, but extreme values in WQI were more apparent in the case of WQI$_{PCA,}$ with sites at Chemangar River having WQI values far exceeding Class 4 **(Figure 5.5b & Appendix E)**. Additionally, Pelepah exhibits the highest water quality for both WQI$_{MLR}$ and WQI$_{PCA}$, with WQI$_{PCA}$ showing slightly better water quality for Pelepah than WQI$_{MLR}$. This brings about some decision making implications associated with the choice of statistical methods used in elucidating key parameters. For instance, during the wet season, WQL_{MLR} values for Chemangar is within Class 4 but WQI_{PCA} for Chemangar exceeds Class 4, which according to the class description by NWQS, deems it unsuitable for any use **(Appendix A)**. As such, this shows the importance of considering and comparing other statistical methods which may otherwise lead to different decision making outcomes.

Furthermore, it is also noteworthy to point out that weighted and non-weighted variations of water quality standards (Classes) for WQL_{MLR} differs significantly as observed from the overlaps between different water quality classes. Non-weighted classes have significantly higher threshold values than the weighted classes (except for Class 1) for WQL_{MLR} **(Figure 5.5a)**. However, that is not observed for WQI_{PCA} **(Figure 5.5b)**. This also brings about some implications associated with decision making regarding the use of non-weighted or weighted WQI as different water quality classes represent different end use according to the class description by NWQS **(Appendix A(c))**.

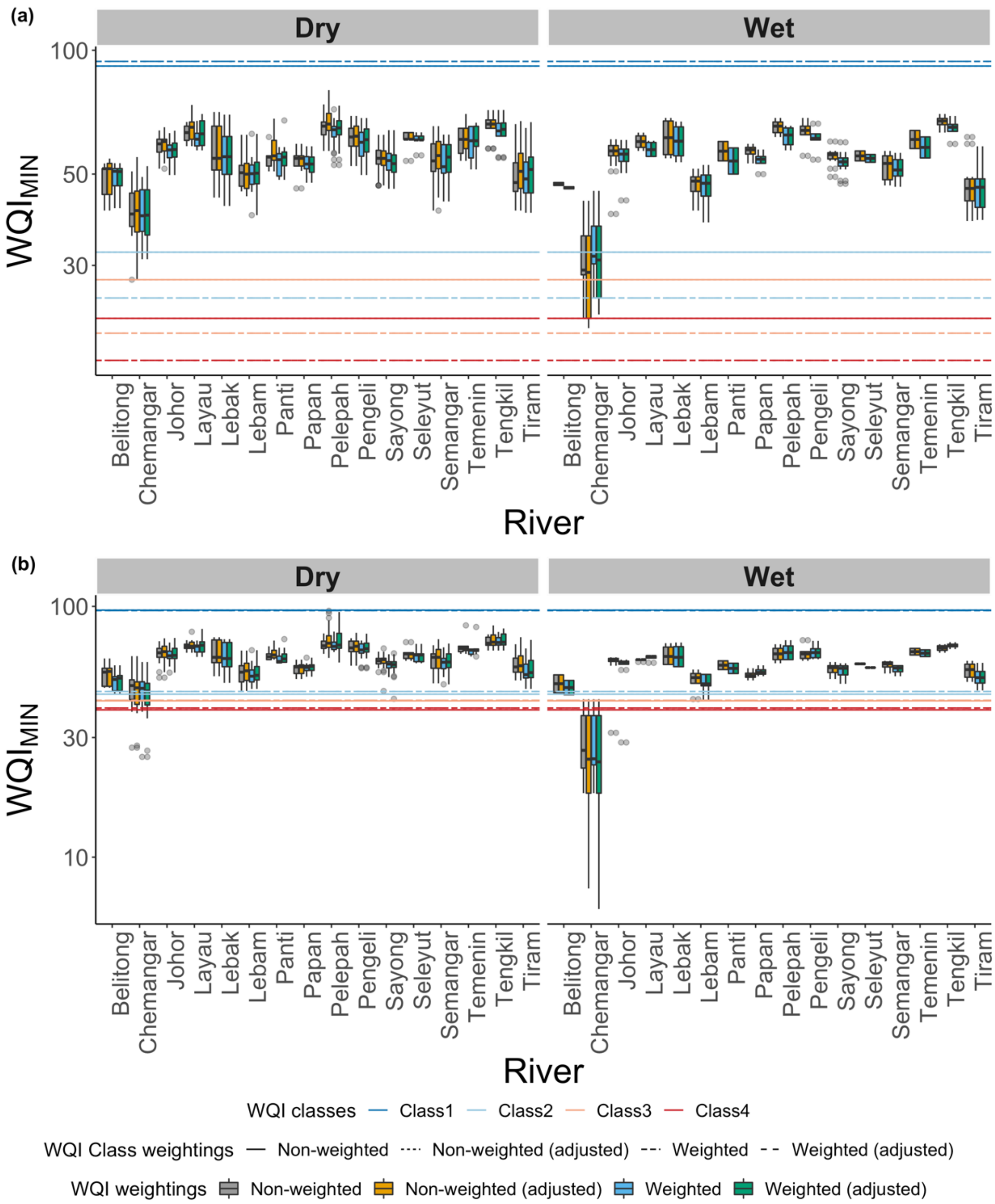


Figure 5.5.Spatial variation in WQI_{MIN} under different seasonality (a) WQI_{MLR} (b) WQI_{PCA}

5.5 Performance between WQI_{MIN}, WQI_{AVG} and WQI_{MY}

The WQI_{MIN} results show high linear correlation with WQI_{AVG}, in agreement with past studies (e.g. Sánchez et al. (2007); Akkoyunlu & Akiner (2012); Wu et al. (2018)), demonstrating that few significant water quality parameters can largely explain the variation in water quality. The performance of both WQI_{MIN} (WQI_{MLR} and WQI_{PCA}) does not deviate too far from WQI_{AVG}, albeit a clear pattern can be observed in their distribution **(Figure 5.6).** WQI_{MLR} and WQI_{PCA} values are generally distributed below and above WQI_{AVG}, respectively, representing underestimation and overestimation, respectively. $WQI_{NON\text{-}WEIGHTED}$ exhibited stronger correlation with WQI_{AVG} compared to $WQI_{WEIGHTED}$. $WQI_{ADJUSTED}$ performed poorer in comparison to $WQI_{NON\text{-}ADJUSTED}$, which could be attributed to the difference is the distribution of missing values, resulting in a greater deviance between $WQI_{AVG\ (ADJUSTED)}$ and $WQI_{MIN\ (ADJUSTED)}$.

For the non-weighted variation of WQI_{MIN}, $WQI_{PCA\ (NON\text{-}WEIGHTED)}$ performs better than $WQI_{MLR\ (NON\text{-}WEIGHTED)}$ in both R^2 and correlation coefficient (cor). RMSE of $WQI_{PCA\ (NON\text{-}WEIGHTED)}$ is slightly higher than $WQI_{MLR\ (NON\text{-}WEIGHTED)}$ as the former has more parameters in its WQI_{MIN}. However, $WQI_{MLR\ (WEIGHTED)}$ performs better than $WQI_{PCA\ (WEIGHTED)}$ in all performance metrics for the weighted variation of WQI_{MIN}. Similarly, WQI_{MLR} showed the greatest correlation with WQI_{MY} amongst all the other WQIs **(Figure 5.7).** In particular, $WQI_{MLR\ (WEIGHTED)}$ showed the best performance (R^2 = 0.634) against WQI_{MY}. As observed from **Figure 5.7**, the regression line for WQI_{PCA} deviates the furthest from WQI_{MY}, which demonstrates that WQI_{PCA} would not be as suitable as WQI_{MLR} to quantify the water quality that WQI_{MY} assesses. Thus, general water quality is best represented by WQI_{PCA} while water quality for specified purposes such as drinking water or ecological well-being is best represented by WQI_{MLR}.

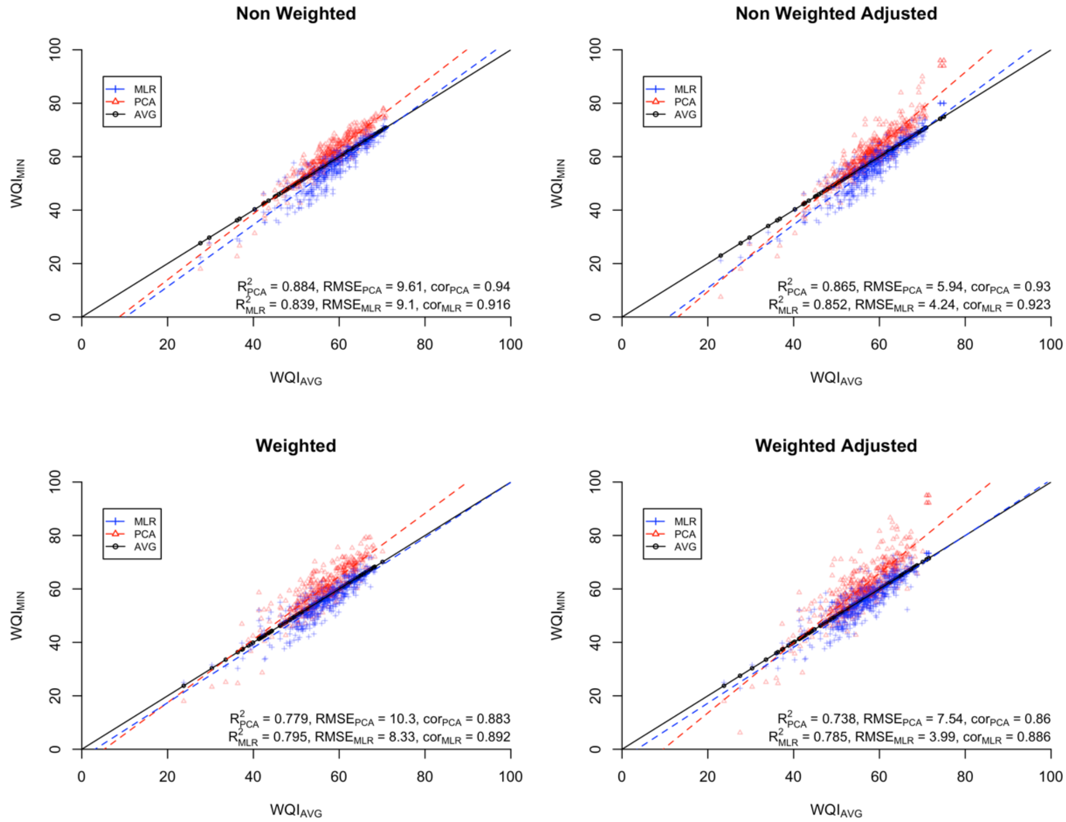


Figure 5.6.Performance of WQI_{MLR} and WQI_{PCA} with respect to WQI_{AVG}

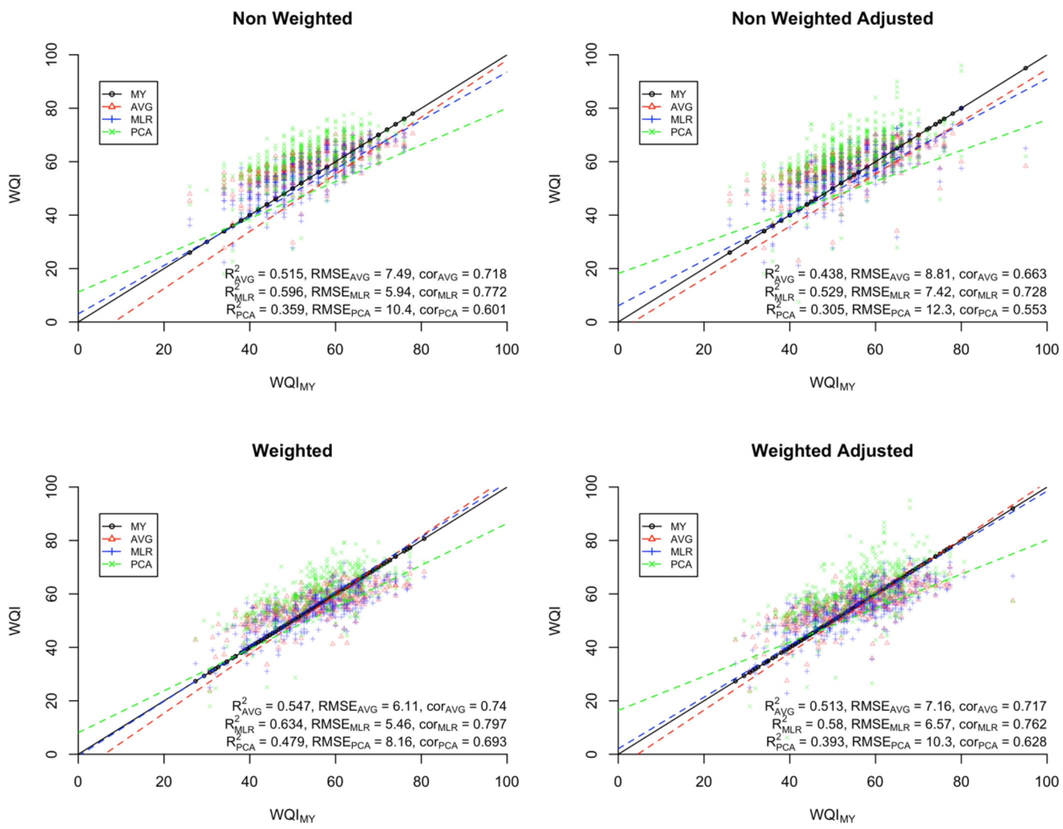


Figure 5.7. Performance of WQI_{MY} against WQI.

Notes: WQI_{MY} was represented with only five parameters (COD, DO, NH3, Turbidity, and pH), the common parameters measured in this study

5.6 Optimal sampling sizes scenario

Various optimal sampling sizes were obtained through Monte Carlo simulation, a technique to simulate the effect of sample size (i.e. number of samples collected) on WQI through bootstrap resampling method **(Figure 5.8)**. The x-axis of the plots simulates the number of samples, represented by the proportion of the data (%) used in this study. The y-axis represents the probability of WQI_{MIN} deviating from WQI_{AVG} (referred to as WQI_{CHANGE} hereafter) for a given percentage change (represented by various groups of colour). This lends to statistical decision theory in facilitating decision-makers or end users to decide the optimal sample size given the uncertainties (threshold of WQI_{CHANGE} they find acceptable) and available resources (Pratt et al., 1995).

The Monte Carlo simulation for $WQI_{AVG\ (NON\text{-}WEIGHTED)}$ and $WQI_{AVG\ (WEIGHTED)}$ displayed the same trends as percentage of data sampling increases. The simulation showed that at least 35% of the data will be required (around 130 samples within the JRB) if an average of 2% change from the mean of WQI_{AVG} can be tolerated by the end user. At least 10% of the data will be required (around 30 samples within the JRB) if an average of 4% change is tolerated. An 8-10% of change from the mean of WQI_{AVG} may not be expected regardless of the number of samples collected within the JRB. On the other hand, one can always expect at least 1% of change from the mean of WQI_{AVG} **(Figure 5.8).**

In the case of computing WQI_{MLR} using the parameters in **Table 5.2**, 0.5 – 2% change from the average of WQI_{AVG} can always be expected. At least 85% of the data will be required if a 4% change ($WQI_{MLR\ (WEIGHTED)}$) from the average of WQI_{AVG} can be tolerated. On the other hand, the $WQI_{MLR\ (NON\text{-}WEIGHTED)}$ will always expect a 4% change from the average of WQI_{AVG}. This also shows that $WQI_{MLR\ (NON\text{-}WEIGHTED)}$ is more sensitive to number of samples compared to $WQI_{MLR\ (WEIGHTED)}$. At least 15% of the data is required if 8% change (WQI_{MLR}) from the average of $WQI_{AVG\ (NON\text{-}WEIGHTED)}$ can be tolerated, and a 10% change from the average of WQI_{AVG} may not be expected regardless of the number of samples collected **(Figure 5.8).**

In the case of computing WQI_{PCA} using the parameters in **Table 5.2**, 0.5 – 8% change from the average of WQI_{AVG} can always be expected, albeit $WQI_{PCA\ (WEIGHTED)}$ shows a lower probability of it exceeding 8%. At least 15% of the data will be required if a 8% and 10% change from the average of WQI_{AVG} can be tolerated for the $WQI_{PCA\ (WEIGHTED)}$ and $WQI_{PCA\ (NON\text{-}WEIGHTED)}$, respectively. In comparison to WQI_{MLR}, WQI_{PCA} appears to be a WQI_{MIN} that is more sensitive to the number of samples, and thus have a higher tendency of deviating from WQI_{AVG} **(Figure 5.8).**

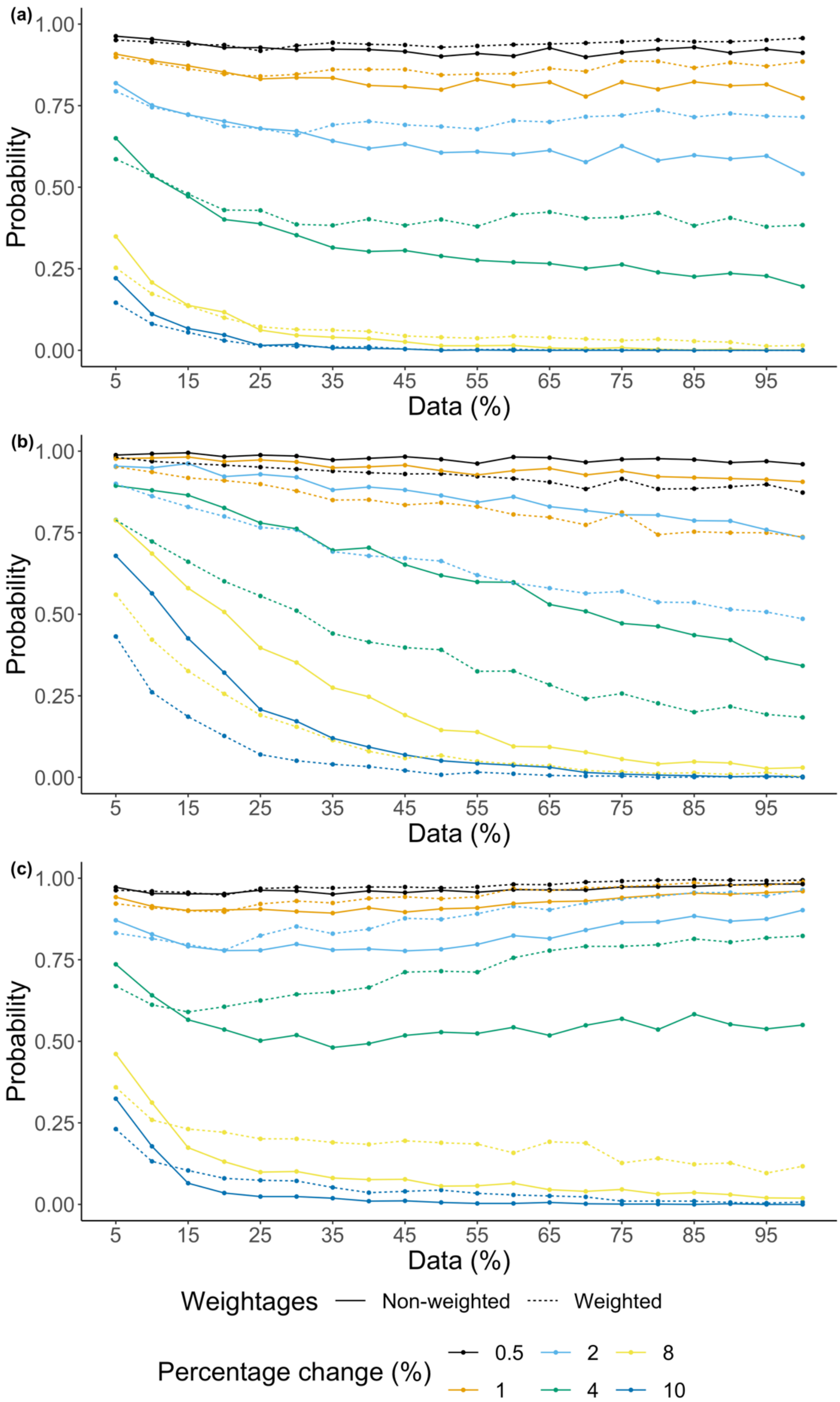


Figure 5.8.Relationship between percentage of data and probability of WQI change exceeding a specified percentage

5.7 Sensitivity Analysis

5.7.1 Parameter sensitivity analysis

The sensitivity of parameters differs for WQI_{AVG}, WQI_{MLR}, and WQI_{PCA} **(Figure 5.9)**. In addition, **Figure 5.9** shows the distribution of most sensitive parameters across all the sites within JRB, and the effect of each WQI variation (i.e. $WQI_{WEIGHTED}$ and $WQI_{NON\text{-}WEIGHTED}$) on parameter sensitivity. The most sensitive parameter is Ba for all WQI_{AVG} variations. Upon obtaining the key parameters through MLR and PCA, the most sensitive parameter is Coliform (total) and Mg for $WQI_{MLR\ (WEIGHTED)}$ and $WQI_{MLR\ (NON\text{-}WEIGHTED)}$, respectively. It is interesting to note that for $WQI_{MLR\ (NON\text{-}WEIGHTED)}$, only one occurrence of Coliform (total) is observed, while for $WQI_{MLR\ (WEIGHTED)}$, Coliform (total) has the highest occurrence. Understandably, this is because Coliform (total) has a higher weightage assigned to it than Mg. As such, parameters with higher weightage tend to be more sensitive, and this effect place more emphasis to parameters' role in certain water quality aspects such as ecological health and toxicity risks. Since Mg only has a weightage of 1, it does not occur at all in the plots of $WQI_{MLR\ (WEIGHTED)}$. On the other hand, COD occurs as the second most frequent parameter in all the plots of WQI_{MLR}, indicating that COD is an important parameter for all variations of WQI_{MLR}. Arsenic (As) and Coliform (total) are the most sensitive parameter for $WQI_{PCA\ (WEIGHTED)}$ and $WQI_{PCA\ (NON\text{-}WEIGHTED)}$, respectively, albeit the highest occurring parameter for $WQI_{PCA\ (NON\text{-}WEIGHTED\ ADJUSTED)}$ are Coliform (total) and Na **(Figure 5.8)**. Coliform (total) is the most frequently occurring sensitive parameter for WQI_{MLR} and WQI_{PCA}, and thus shows that it is the most sensitive parameter. Generally, not much difference can be observed in the distribution of sensitive parameters between adjusted and non-adjusted WQIs.

Parameter sensitivity analysis is dependent on weightages and differs across WQI_{AVG}, WQI_{MLR}, and WQI_{PCA}, as demonstrated in **Figure 5.9**. This shows that different statistical methods (i.e. MLR, PCA) of obtaining key parameters yields different parameter sensitivity. This is because sensitivity of parameter is dependent on the other parameters that are selected through the statistical process as well as characteristics of the water quality for certain sites **(Appendix G)**.

5.7.2 Parameter and missing values sensitivity analysis

While parameter sensitivity analysis demonstrates the sensitivity of parameters with respect to changes in WQI, the sensitivity of parameters and missing values with respect to changes in WQI have yet to be analysed in most literature. **Figure 5.10** demonstrates the relationship between probability of WQI_{CHANGE} exceeding a given percentage and percentage of missing values. The gradient of the empirical graphs gives an idea of the sensitivity to missing values for each parameter. The steeper the gradient, the greater the rise in probability of WQI_{CHANGE} exceeding a given percentage per unit increase of missing values. Generally, the sensitivity of parameters to missing values is lower when the threshold of percentage change in WQI is higher. This is because at higher threshold of percentage change, the increase in probability per unit change of missing values is smaller.

Coliform (total) is the most sensitive to missing values for both WQI_{MLR} and WQI_{PCA}, and the $WQI_{MLR\ (WEIGHTED)}$ and $WQI_{PCA\ (WEIGHTED)}$ is more sensitive. With 35% of missing values for Coliform (total), a 1% change in WQI can be expected. For WQI_{MLR}, Coliform (total), Mg, COD, and Cr have high sensitivity to missing values, while NO_3^-, NH_3-N, and E.coli have moderate sensitivity to missing values **(Figure 5.10a)**. Copper (Cu) appears to have the lowest sensitivity to missing values as it has the gentlest gradient. As for WQI_{PCA}, Coliform (total), Na, and Mg have high sensitivity to missing values while the rest of the parameters have moderate sensitivity **(Figure 5.10b)**.

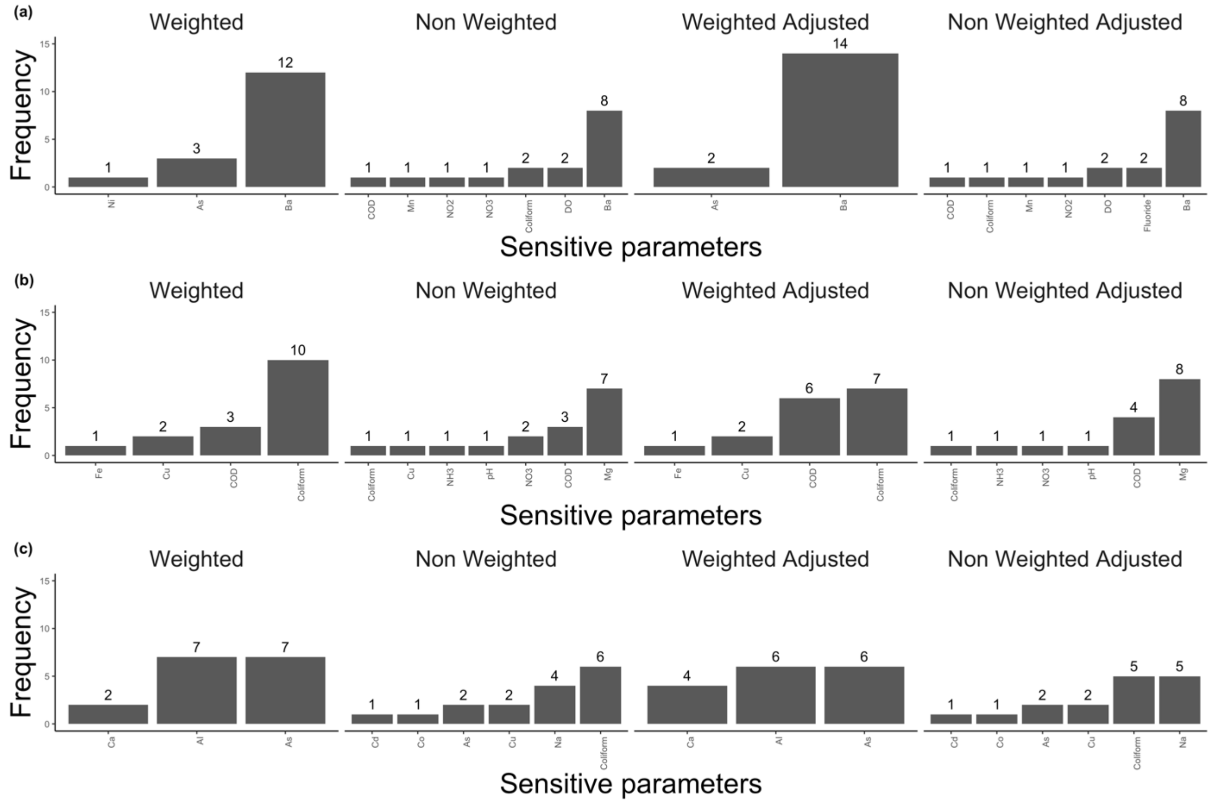


Figure 5.9. Frequency plot of the most sensitive parameters aggregated across all sampling sites. (a) WQI_{AVG} (b) WQI_{MLR} (c) WQI_{PCA}

Notes: Frequency refers to the total number of times a parameter appears as the most sensitive parameter across the rivers (i.e. 16 unique rivers within the JRB). As such, the total frequency value should add up to 16 in each of the plots. Details of sensitive parameters of each site is found in **Appendix G**

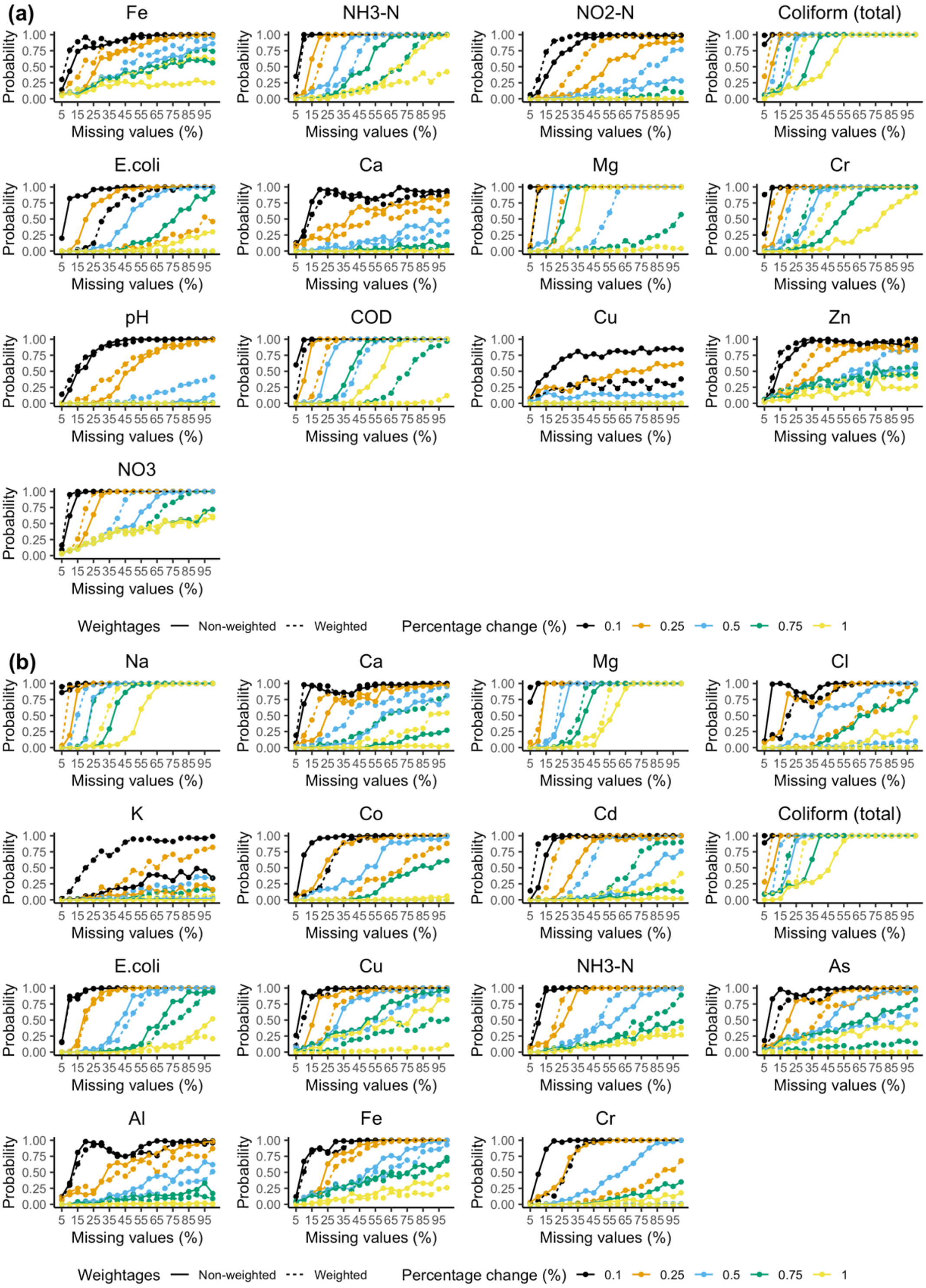


Figure 5.10. Sensitivity of parameters to missing values (a) WQI_{MLR} (b) WQI_{PCA}

5.8 Adequacy of WQIs

Three WQIs were developed for this study – WQI_{AVG}, WQI_{MLR}, and WQI_{PCA}, where WQI_{AVG} computes all 34 measured water quality variables while WQI_{MIN} (WQI_{MLR} and WQI_{PCA}) computes the reduced form of WQI_{AVG} using MLR and PCA techniques to select key parameters. WQI_{MLR} performs better than WQI_{PCA} for the weighted variation of WQI while WQI_{PCA} performs better for the non-weighted variation of WQI (**Figure 5.6).** Despite WQI_{MLR} (13 parameters) having fewer parameters than WQI_{PCA} (15 parameters), WQI_{MLR} generally serves as a more well-rounded WQI. WQI_{MLR} covers various groups of water quality parameters (i.e. heavy metals, Coliform (total), and nutrients) while WQI_{PCA} is heavily inclined towards heavy metals parameters. In addition, WQI_{MLR} has better performance, exhibited in both **Figure 5.6 and Figure 5.7**, and can represent WQI_{AVG} better when fewer data is available. This is because WQI_{PCA} appears to be more sensitive to sample size, and there is a higher probability of WQI_{CHANGE} exceeding a given threshold (**Figure 5.8)**.

Traditionally, WQI is computed without any room for missing values for any water quality parameters. This holds true when a one-time sampling of water quality for a site is conducted, and thus a $WQI_{ADJUSTED}$ may not be valid in the interpretation of water quality when there are missing values for a single-site measurement. Nonetheless, a single measurement is often not representative of the general water quality due to the inherent variabilities. On the other hand, acquiring more data may increase the chances of errors (e.g. human or analytical) that can result in missing values. Dealing with missing data is one of the most important steps in data analysis. Previous studies (e.g. Srebotnjak et al. (2012)) have explored methods such as Hot-Deck imputation to replace missing values with other values, which may lead to erroneous results and conclusions. With the $WQI_{ADJUSTED}$, no assumptions on missing values are implied, and it aims to continue capitalising on the available data (provided that a few sets of samples are collected) while producing a WQI value that can still contribute to the basic understanding of general water quality.

To put things into perspective, environmental data such as water quality varies according to environmental conditions (e.g. discharge conditions, climate, and land use change). Therefore, the difference between WQI and $WQI_{ADJUSTED}$ may not be too different from the natural variability observed in environmental conditions. In addition, there are always uncertainties associated with each process (i.e. instrument uncertainty, variability of natural conditions), and thus spatio-temporal replicates are necessary to reduce uncertainties as much as possible. While missing values contributes to greater uncertainty, and its validity may be questionable, it is therefore essential to evaluate and assess the adequacy of $WQI_{ADJUSTED}$ based on the tolerance of user-defined acceptable difference which plays into the importance of decision-making. As such, $WQI_{ADJUSTED}$ is assessed by the distribution of its missing data, as shown in **Figure 5.10,** as $WQI_{ADJUSTED}$ is constructed for the very purpose of salvaging important information in the face of incomplete data. $WQI_{ADJUSTED}$ provides an alternative for computing WQI with missing values in the dataset, which would be not computable based on traditional aggregation methods. However, $WQI_{ADJUSTED}$ should be used with caution as the parameter sensitivity to missing values differs for each water quality parameter. As such, one should take note of the water quality parameters containing missing values, and evaluate the usefulness of $WQI_{ADJUSTED}$, based on the threshold of acceptability for WQI_{CHANGE} **(Figure 5.10)**. Moreover, sensitivity of parameters to missing values also has to be evaluated with the parameters' weightages. **Figure 5.10** shows that weighted parameters are more sensitive to missing values. Thus, $WQI_{ADJUSTED}$ values should be carefully interpreted when missing values are present for that particular water quality parameter as it can undermine the usefulness of $WQI_{ADJUSTED}$.

5.9 Comparison with Malaysia's WQI

Malaysia's WQI – WQI_{MY} has six main parameters, namely DO, BOD, COD, Suspended Solids, ammoniacal-nitrogen, and pH. On the other hand, WQI_{MLR} and WQI_{PCA} has thirteen and fifteen parameters, respectively **(Table 5.2)**. While more parameters can better represent the general water quality, having many parameters in the WQI can make it unwieldy. However, the biggest drawback for DOE's WQI is its generality for different water usage, and it being not a site-specific WQI. On the other hand, the WQI developed in this study selects significant water quality parameters that are representative of the total measured water quality parameters for the JRB. It is essential that a site-specific WQI is developed on the watershed scale. That is, an adequate coverage of water quality assessment within the watershed should be conducted in the development of the WQI as in-situ water quality is affected by water quality changes and environmental conditions upstream. Extrapolating a general WQI, in the case of WQI_{MY}, to monitor a site-specific water quality issue may not be adequate.

We recognise that WQI_{MY} has parameters that were not measured in this study – namely BOD_5 and TSS. Instead, Turbidity was used as a proxy for TSS. BOD_5 is an indicator for biodegradable organic compounds, and is usually measured to assess the efficiency of processes in wastewater treatment plants. However, it is often challenging to culture a consistent viable microbial community to acquire accurate measurements (Hudson et al., 2008), and BOD_5 measurements may yield an uncertainty of 15%~20% (Kwak et al., 2012). In addition, toxic compounds such as heavy metals, which were also measured in this particular WQI, may stunt bacterial oxidation of organic compounds, and thus influence the reliability of BOD_5 measurements.

5.10 Summary and recommendations

There is an urgent need for better water quality monitoring within the JRB to improve water management strategies through developing an effective WQI that optimises both water quality information and resources. Measuring a large number of parameters is often not practical, and is extremely financially taxing for developing countries with limited resources. Not only does this translate to inefficient use of resources, measuring water quality parameters that are not representative to the site-of-interest fails to help water managers make adequate implementations for a specific water quality concern. Moreover, a site-specific WQI containing only representative water quality parameters helps water managers in obtaining the most important information about the local water quality at the lowest cost. Concretely, rather than having to measure all water quality parameters which would incur a huge cost, only a selected set of water quality parameters needs to be measured, and this is represented by WQI_{MLR} and WQI_{PCA}.

Spatio-temporal replicates are often required for a better representation of water quality, and by doing so, reduces uncertainties for interpretation especially when decision-making regarding water resource is concerned. Ultimately, the very purpose of a WQI aims to condense water quality information to bridge science and decision-making, which makes the WQI an important user-based tool. The flexibility, utility, and practicality of a WQI were demonstrated in the proposed new framework in this study. The flexibility aspect of the WQI was introduced through Sub-WQIs, where each Sub-WQIs can serve different end-use or address certain water quality aspects in the context of local water quality conditions. The utility aspect can be improved through optimising the number of samples required to build a robust WQI, based on user's pre-defined threshold of acceptability for WQI_{CHANGE}. Specifically, around 130 samples will be required if a 2% change in WQI_{CHANGE} can be tolerated. WQI_{MLR} serves as a better model than WQI_{PCA} when there is a smaller sample size. Similarly, the practicality aspect of the WQI is enhanced through overcoming the limitations of missing values using $WQI_{ADJUSTED}$, provided that the user has taken into consideration the threshold of WQI_{CHANGE} based on the water quality parameter's proportion of missing values. Concretely, we recommend that missing values should not exceed 10% of the total observations, as above this range, WQI_{CHANGE} is expected to increase steeply. Thus, future studies on WQI are encouraged to further delve into other applications of assessing WQI, which can ultimately guide decision-making.

The next following sections shall delve into the influence of land use configuration on each water quality parameter. This helps to better contextualises results of WQI within the JRB, with respect to the local conditions such as surrounding land use, known point sources, and seasonality.

5.11 Patterns in land use metrics

The JRB is largely dominated by oil palm land use, as shown in many of the sites. Pelepah (PL sites) have a larger proportion of Bare land and Forest compared to the other sites (**Figure** 5.11). Tiram sub-basin (TR sites), Chemangan (CM sites), Sayong (SY sites), Lebam (LB sites), and Semangar (SM sites) have a higher proportion of Urban (others) and Urban (Residential) land use. TR sites in particular, have a significant proportion of Agriculture (others) land use for the riparian scale, which demonstrates that other forms of agriculture such as animal agriculture are largely found in riparian regions.

Other land use metrics (AI, LS, PLADJ, and SHDI) showed similar trends across different spatial scales, with the exception of reach scale not having any agriculture land use. Reach has the lowest SHDI due it being the smallest scale. Agriculture (Oil palm) has the highest value for AI and PLADJ, which suggests that Agriculture (Oil palm) land use is largely cohesive, followed by Urban (others) and Urban (Residential) land use.

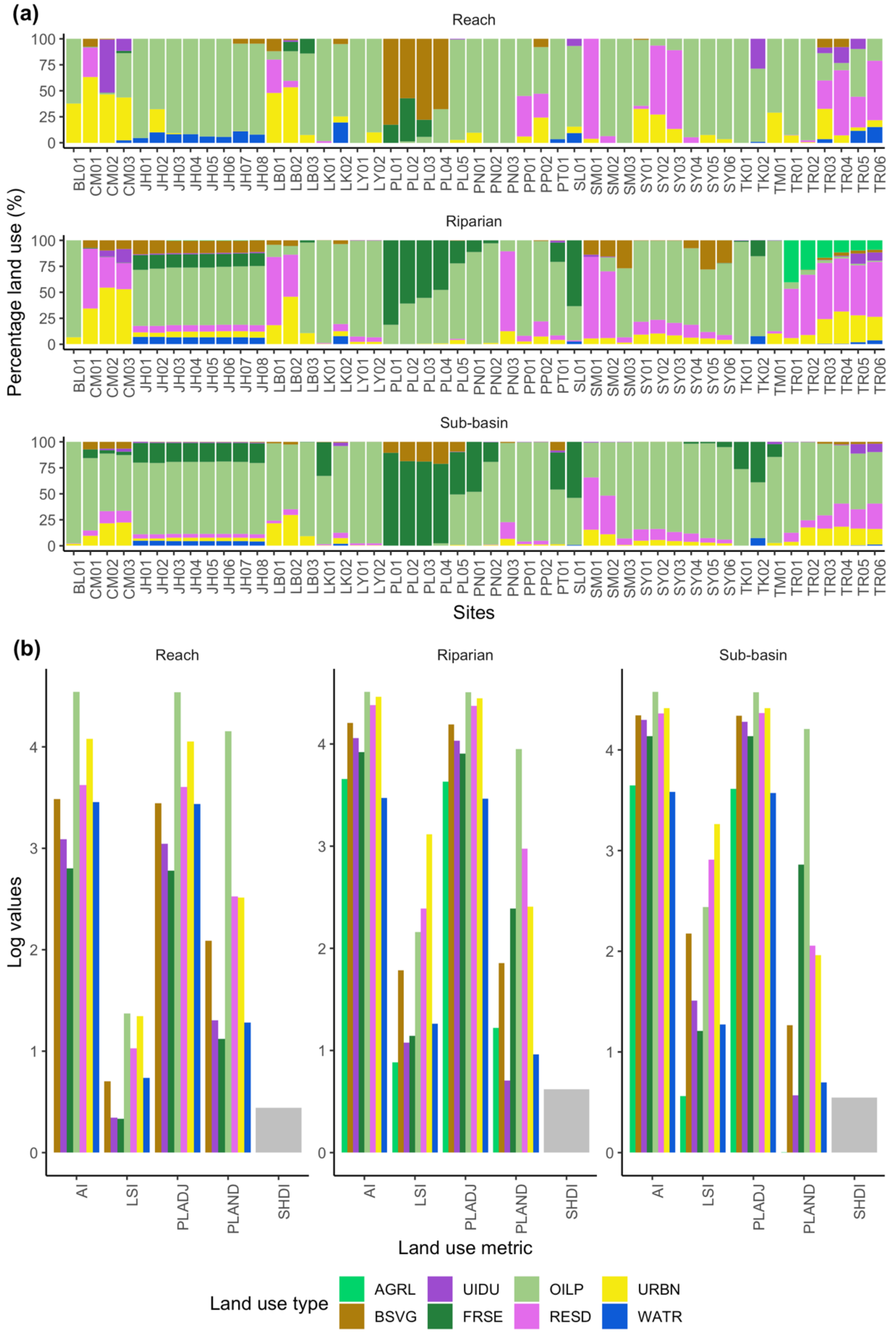


Figure 5.11. Land use metrics under three different scales. (a) Percentage of land use for each site (b) Mean AI, LS, PLADJ, and SHDI across sites

5.12 Influence of spatial scale on water quality under different seasonality

The significance of spatial scale on water quality gives information on the processes or activities taking place within each scale, which may change during different seasonality, depending on surface conditions and activities. Cd, Fluoride (F), and Se were poorly explained by the land use models, as demonstrated by the low R^2 values (**Figure 5.12**). This suggests that land use variables alone are insufficient to explain the variations in Cd, Fluoride (F), and Se.

Generally, the reach scale has higher R^2 values than riparian and sub-basin scale. Reach scale was the most significant scale under dry and wet season for water quality parameters such as Cr, Mn, Ni, SEC, Na, Br, Co, SEC, and NO2-N, which indicates that nearby sources at the reach scale largely governs the water quality (**Figure 5.12**). R^2 values for heavy metals including Cr, Ni, and Mn are significantly higher during the wet season (**Figure 5.12**), which indicates that nearby sources of heavy metals are washed into the river through overland flow during the wet season, while during the dry season, these heavy metals were not mobilised.

On the other hand, the riparian scale was the most significant scale under dry and wet season for E.coli, NH3-N, Cu, Fe, and Turbidity, with higher R^2 values during the dry season **(Figure 5.12)**. Buck et al. (2004) reported that nutrients and faecal Coliform (total) concentrations were better explained by land use within the riparian zone for a basin dominated by pastures in New Zealand, which concurred with the observations. This could suggest that processes in the riparian region were more significant during the dry season, and these processes were masked during the wet season, possibly due to dilution effects in the river.

Sub-basin scale was the most significant scale for only Temperature and NO3 for both dry and wet season, which suggests that processes affecting these parameters operate at a larger spatial scale **(Figure 5.12 & Table 5.4).** However, the frequency of sub-basin scale occurring as a significant scale for both dry and wet season was significantly smaller than that of reach or riparian scale. This suggests that the sub-basin scale is perhaps too large of a scale to have a significant or dominant process governing the water quality under different seasonality. The plethora of processes occurring at a sub-basin scale could have antagonistic or 'diluted' effects on water quality due to the large spatial scale at which these processes interact. For instance, Pan et al. (2004) and McBride & Booth (2005) found that water quality generally had poor correlations with large spatial scales such as sub-basins.

No significant models were found for Ag, Cd, NO3, Se, and Pb during the dry season, albeit some models were found to be significant during the wet season. Significance of spatial scale varied for some water quality parameters across different seasons. For instance, As (Sub-basin$_{dry}$, Reach$_{wet}$), COD (Sub-basin$_{dry}$, Reach$_{wet}$), DO (Reach$_{dry}$, Riparian$_{wet}$), and PO4 (Riparian$_{dry}$, Reach$_{wet}$) have significant spatial scales varying across different seasons, which could be attributed to combinations of processes occurring during each season. Whereas, if the spatial scale is consistent across different seasonality, a dominant process within that scale may be a significant source or sink for the water quality parameter. For instance, significant land use predictors remained consistent under both dry and wet seasons for Cr, Mn, NO2-N, and E.coli (**Table 5.4**).

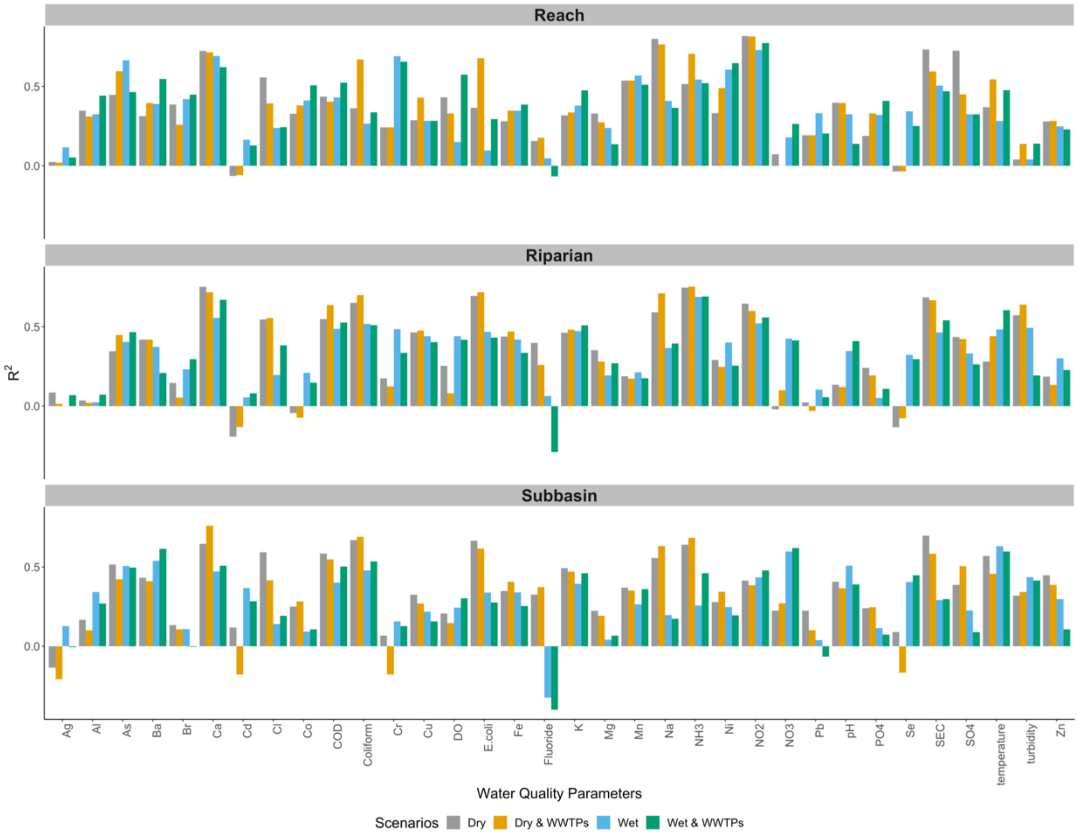


Figure 5.12. Performance of multivariate regression conducted for each water quality parameter under different spatial scale

Notes: Details of R^2 values are found in **Appendix H**.

Table 5.3. Significant variables from its corresponding significant scale

	Dry			Wet		
Parameter	**Reach**	**Riparian**	**Sub-basin**	**Reach**	**Riparian**	**Sub-basin**
(a) Land use only scenario						
Al	lsi_3 (-1.77); pland_2 (-1.29); pland_8 (-0.977); **shdi (4.89)**	-	-	-	-	-
As	-	-	-	ai_5 (0.178); pladj_4 (0.205); **pland_5 (-0.303)**	-	-
Ba	-	-	-	-	-	-
Br	**lsi_4 (0.4)**	-	-	**lsi_4 (0.86);** pland_4 (-0.252)	-	-
Ca	-	**pland_7 (0.344)**	-	**lsi_4 (1.91);** pland_4 (-0.479); pland_6 (0.294)	-	-
Cd	-	-	-	-	-	ai_2 (-0.0526); **ai_6 (0.109)**
Cl	-	-	ai_8 (0.291); lsi_7 (0.471); pland_7 (0.404); **pland_8 (-0.64)**	**lsi_4 (2.52);** pland_4 (-0.828)	-	-
Co	-	-	ai_8 (0.256); pladj_3 (-0.258); pland_7 (0.504); **pland_8 (-0.915)**	-	**pland_1 (1.18)**	-
COD	-	-	**pland_7 (0.241)**	-	-	-

Coliform (total)	-	-	ai_8 (0.256); pladj_3 (-0.258); pland_7 (0.504); **pland_8 (-0.915)**	-	**pland_1 (1.18)**	-
Cr	ai_4 (0.105); **pland_4 (-0.17)**	-	-	ai_4 (0.398); **pland_4 (-0.703)**	-	-
Cu	-	ai_3 (-0.0724); ai_5 (-0.19); pland_4 (-0.216); **pland_5 (0.218)**	-	-	ai_5 (-0.239); **pland_4 (-0.266)**	-
DO	pland_5 (-0.112); **pland_6 (-0.153)**	-	-	-	**ai_4 (-0.0758)**	-
E.coli	-	**pland_1 (0.898);** pland_7 (0.631)	-	-	**pland_1 (1.85);** pland_2 (1.21)	-
Fe	-	-	-	-	-	-
Fluoride	-	pladj_8 (-0.0142); **pland_5 (-0.0312)**	-	-	-	-
K	-	-	**pland_7 (0.262)**	-	ai_5 (-0.282); **pland_4 (-0.346)**	-
Mg	-	**pland_7 (0.327)**	-	ai_4 (0.567); **pland_4 (-0.74)**	-	-
Mn	**pland_4 (0.854)**	-	-	pladj_6 (-0.193); **pland_4 (0.849)**	-	-
Na	**lsi_4 (2.15);** pland_3 (0.314); pland_4 (-0.497); pland_5 (0.27); pland_6 (0.296)	-	-	ai_4 (0.985); **pland_4 (-1.22)**	-	-
NH3	-	**lsi_8 (0.375)**	-	-	**pladj_5 (-0.354);** pland_7 (0.346)	-
Ni	**ai_4 (0.311)**	-	-	ai_2 (0.225); ai_4 (0.504); **lsi_3 (-0.713);** pladj_6 (-0.217);	-	-

				pland_4 (-0.541); pland_6 (0.423); pland_8 (-0.253)		
NO2	**lsi_4 (0.218);** pland_4 (-0.0835)	-	-	**lsi_4 (0.222);** pland_4 (-0.0836)	-	-
NO3	-	-	-	-	-	ai_7 (-0.331); **lsi_7 (0.402);** pladj_3 (-0.139)
Pb	-	-	ai_7 (0.122); pland_7 (0.122); **shdi (-0.815)**	ai_2 (0.105); **pland_2 (-0.177)**	-	-
pH	-	-	ai_3 (-0.0203); **pladj_6 (0.0278)**	-	-	-
PO4	-	-	-	-	-	-
Se	-	-	-	-	-	ai_2 (-0.0414); ai_6 (0.068); pladj_3 (-0.0443); **pland_2 (0.114)**
SEC	**lsi_4 (1.9);** lsi_6 (-0.661); pland_4 (-0.381); pland_5 (0.437); pland_6 (0.553); pland_7 (0.25)	-	-	ai_4 (0.936); **pland_4 (-1.16)**	-	-
SO4	lsi_3 (-0.605); **lsi_4 (1.45);** lsi_7 (0.237); pland_3 (0.295); pland_5 (0.416); pland_6 (0.438)	-	-	-	**pland_4 (-0.441)**	-
Temperature	-	-	**lsi_5 (-0.0275);** lsi_7 (0.0198); pladj_5 (0.0174);	-	-	**lsi_4 (0.0334)**

			pland_7 (0.0119); pland_8 (0.0248)			
Turbidity	-	**pland_1 (0.702);** pland_2 (0.555)	-	-	**pland_5 (-0.916)**	-
Zn	-	-	pland_7 (0.322); **shdi (-2.44)**	-	**pland_1 (-0.418)**	-
(b) Combined scenario						
Al	**lsi_3 (-1.75);** pland_2 (-1.28)	-	-	**lsi_3 (-1.57);** pland_2 (-0.685); pland_6 (0.515)	-	-
As	pladj_3 (0.0914); **pland_5 (-0.24)**	-	-	-	-	pland_2 (0.398)
Ba	-	-	-	-	-	**lsi_5 (0.574)**
Br	**lsi_4 (0.32)**	-	-	**lsi_4 (1.19);** pland_4 (-0.322)	-	-
Ca	-	-	ai_5 (0.217); pland_3 (0.309); **pland_7 (0.398)**	-	**pland_7 (0.417)**	-
Cd	-	-	-	-	-	ai_2 (-0.0541); **shdi (-0.657)**
Cl	-	-	-	-	-	-
Co	-	**pland_7 (0.559)**	-	-	-	**lsi_4 (-1.46)**
COD	-	**pladj_5 (-0.163);** pladj_8 (0.115)	-	-	pland_4 (-0.381); **pland_5 (-0.752)**	-
Coliform (total)	-	**pland_7 (0.559)**	-	-	-	**lsi_4 (-1.46)**
Cr	ai_4 (0.105); **pland_4 (-0.17)**	-	-	ai_4 (0.409); **pland_4 (-0.684)**	-	-

Cu	-	**pland_5 (0.211);** pland_7 (0.183)	-	-	ai_5 (-0.227); **pland_7 (0.326)**	-
DO	**pland_6 (-0.127)**	-	-	ai_4 (-0.135); pland_5 (-0.122); **pland_6 (-0.2);** pland_7 (-0.118)	-	-
E.coli	-	**pland_1 (0.764)**	-	-	**pland_1 (1.79);** pland_2 (1.28)	-
Fe	-	ai_3 (-0.318); **ai_5 (-0.609)**	-	-	-	-
Fluoride	-	-	**pland_8 (-0.083)**	-	-	-
K	-	-	-	-	-	-
Mg	-	-	-	-	-	-
Mn	**pland_4 (0.854)**	-	-	**lsi_7 (0.834); pland_4 (0.476)**	-	-
Na	**ai_4 (0.476);** pland_5 (0.263); pland_6 (0.349)	-	-	-	-	-
NH3	-	**NUM (0.302);** pland_5 (-0.232)	-	-	**ai_5 (-0.35)**	-
Ni	**ai_4 (0.669);** ai_6 (-0.253); **pland_5 (0.669);** pland_6 (0.57)	-	-	ai_3 (-0.157); ai_4 (0.497); pladj_2 (0.23); pladj_6 (-0.234); **pland_4 (-0.508);** pland_5 (0.368); pland_6 (0.499)	-	-
NO2	**lsi_4 (0.213);** pland_4 (-0.0797); pland_6 (0.0209)	-	-	**lsi_4 (0.209);** pland_4 (-0.0786)	-	-

NO3	-	-	ai_2 (-0.22); **ai_6 (0.427)**	-	-	ai_6 (0.22); ai_7 (-0.303); **lsi_7 (0.319)**
Pb	-	-	-	-	-	-
pH	pland_2 (0.0366); **pland_4 (-0.0524)**	-	-	-	**pland_5 (-0.0561)**	-
PO4	-	-	-	**DESIGNPE (-0.0238)**	-	-
Se	-	-	-	-	-	lsi_2 (-0.112); pladj_3 (-0.0484); **pland_2 (0.155);** pland_5 (0.112)
SEC	-	**pland_7 (0.337)**	-	-	**pland_7 (0.608)**	-
SO4	-	-	pland_7 (0.597); **shdi (-2.56)**	**lsi_4 (2.01)**	-	-
Temperature	DESIGNPE (0.00636); **DESIGNPE.den.wt (-0.00861)**	-	-	-	**pland_1 (-0.0218);** pland_4 (-0.0178)	-
Turbidity	-	**pland_1 (0.713);** pland_2 (0.484); pland_7 (0.364)	-	-	-	-
Zn	-	-	pland_7 (0.411); **shdi (-2.02)**	-	-	-

Notes: only significant land use variables (p<0.05) are listed; ( ) standard partial regression coefficient (B); [ ] VIF values; **Bolded**: Variable with the greatest absolute standard partial regression coefficient (B); *Adj.* R^2 means the adjusted coefficient of multiple determination; metrics include the percentage of landscape (PLAND), landscape shape index (LSI), aggregation index (AI). Metric values for land use category: 1 Agriculture (Others), 2 (Bare land and shrubs), 3 Urban (Industry), 4 (Forest), 5 Agriculture (Oil palm), 6 Urban (Residential), 7 Urban (Others) and 8 (Water Bodies).

5.13 Influence of land use on water quality

This section shall discuss the corresponding results for the most significant scale, as described in the aforementioned section. Performance of the regression models for each water quality parameter varied widely, with a range of R^2 values from 0.212 to 0.818 (**Figure 5.12 & Appendix H**).

Agriculture (Others), Urban (Others), Urban (Residential), Forest, and Agriculture (Oil palm) are the most common land use that were found to be correlated with water quality parameters, while Urban (Industry) land use variables are less frequently correlated with water quality parameters **(Table 5.4)**. The water quality parameters that are significantly correlated with Urban (Industry) land use metrics were: Co, Coliform (total), Cu, Na, NO3, Pb, and SO4 (land use only scenario); As, Ca, and Fe (combined scenario); Al, Ni, and Se (both scenarios). The water quality parameters that are significant under both scenarios are all heavy metals (i.e. Al, Ni, and Se), which suggests that the sources of these heavy metals were likely from Urban (Industry) land use.

During the dry season, Urban (Residential) PLAND **(B=-0.153)** has the greatest negative impact on DO, while during the wet season, Forest AI **(B=-0.0758)** has the greatest negative impact on DO **(Table 5.4)**. Ding et al. (2016) reported that DO had the best model at the riparian scale and it was negatively related to Forest AI, which is consistent with this study's observations. Low DO values in tropical forest are characteristic of tropical regions due to high organic content from leaf litter and higher water temperatures (Manahan, 2017). This was also supported by studies conducted in Kuala Selangor and Sabah, Malaysia by Irvine et al. (2012) and Harun et al. (2014), respectively. During the wet season, erosion of the soil or leaf litter washed into nearby streams from large forest patch may also increase organic matter in rivers, leading to lower DO values (Yule & Gomez, 2009). However, Urban (Residential) PLAND has a greater negative impact on DO than Forest AI during the dry season **(Table 5.4)**. Wastes from residential areas contribute to high loadings of organic waste which, through their oxidation, consumes a lot of oxygen, resulting in low DO values (Ding et al., 2016; Bo et al., 2017).

Agriculture (Oil palm) PLAND and Bare land shrubs PLAND are the most significant predictors for As **(B=-0.303)** and Pb **(B=-0.177)**, respectively **(Table 5.4).** This may indicate that vegetation of any type could serve as sinks for As and Pb through assimilation and attenuating their transport to nearby rivers (Walsh & Kunapo, 2009; Mirza et al., 2014).

Forest PLAND is the most significant predictor for Cr and Mn for both dry and wet season, albeit Cr is negatively related to Forest PLAND, while Mn is positively related to Forest PLAND **(Table 5.4)**, indicating different dominant process occurring for each compound. Mn is likely of natural origin than anthropogenic origin. Natural sources include igneous rocks that contains minerals with Mn, and JRB's underlying geology is mostly igneous rock (Czarnowska, 1996; Ahmad et al., 2004). Enriched sources of Mn may come from the forest's foliage and forest floor (Fujii et al., 2013; Richardson, 2017). On the other hand, forest serves as effective sinks for Cr. Partial regression coefficient of Forest PLAND is also significantly higher during the wet season for Cr **(Table 5.4)**, which could be attributed to downward leaching of Cr into the soil, resulting in less Cr entering the river. This is because naturally occurring Cr exists as chromium salts, some of which are soluble in water (Bartlett & James, 1979). NO2-N is also positively related to Forest LSI for both dry and wet season, which is observed under reducing conditions in wetland forest that oxidises $NH4^{+}$ to NO2-N during the process of nitrification (Shrestha et al., 2009).

Agriculture (Others) PLAND **(B=0.702)** and Agriculture (Oil palm) PLAND **(B=-0.916)** are the most significant predictors for Turbidity during the dry and wet season respectively **(Table 5.4)**. Agriculture (plantation of animal agriculture) can increase soil erosion due to improper agricultural practices such as ploughing or grazing by animals (Sidle et al., 2006; Giri & Qiu, 2016; Mello et al., 2018). However, oil palm plantation in most parts of Malaysia practices reduce soil erosion by placing oil palm leaves on gullies to reduce erosion (Hartemink, 2006; Sahat et al., 2016).

Agriculture (Others) is the most important predictor for E.coli for both dry **(B=0.898)** and wet season **(B=1.85)**, with the wet season having a much higher partial coefficient. This could be attributed to animal agriculture or fertilisers applied on plantations which enters nearby rivers easily through overland flow during the wet season. Such effects of land use on E.coli concentration was observed in Southeast Asia by Rochelle-Newall et al. (2016). E.coli largely originates in the faecal matter of warm blooded animals, and may also be present in fertilisers applied on plantations (Heinonen-Tanski & Uusi-Kämppä, 2001; Johnson et al., 2003).

NH3-N is positively related to Water Bodies LSI **(B=0.375)** during the dry season, and negatively related to Agriculture (Oil palm) PLADJ **(B=-0.354)** during the wet season **(Table 5.4)**. However, this observation contrasted against previous studies which reported that surface water was negatively correlated with pollutants such as NH4-N, and argued that it may be attributed to dilution and degradation of the pollutants (Shen et al., 2015). It was also suggested that wetlands including swamps near water bodies assimilate such pollutants (Bo et al., 2017). However, we argue that the correlation with water bodies was observed because of effluent discharge into the river by WWTPs. Complex configurations of rivers resulting in high LSI of Water Bodies land use can abate the transportation and dispersion of NH3-N loading from nearby WWTPs (Cardenas, 2008).

COD is positively related with Urban (Others) PLAND **(B=0.241)** during the dry season **(Table 5.4)**. One possible reason could be the high loadings of organic wastes from industries, commercial, and institutions that contribute to the COD loadings in urban regions. This is consistent with the findings of numerous studies (i.e. Lee et al. (2009); Bu et al. (2014); Yu et al. (2016); Bo et al. (2017); Cheng et al. (2018)) that reported positive relationships between COD and urban areas. Cd is positively related with Urban (Residential) AI **(B=0.109),** and Cd could originate from cadmium pigments in paint flakes, chips, and dust in continuous patch of residential areas, as demonstrated by a study conducted on the residential areas of Lagos (Adeyi & Babalola, 2017).

5.14 Influence of point source on water quality

Performance of the regression models for each water quality parameter varied widely, with a range of R^2 values from 0.247 to 0.816 for combined scenarios respectively (**Figure 5.11 & Appendix H**). Generally, the land use only scenario is better in explaining the overall water quality under both dry and wet seasons for most water quality parameters. However, some water quality parameters such as Co, COD, Coliform (total), NH3-N, Ni, and PO4 displayed higher R^2 values for the combined scenario than the land use only scenario under both dry and wet season (**Figure 5.11 & Appendix H**). This demonstrates that including point source variables in the MLR models can better explain the variations for these water quality parameters, and point source plays a significant role in influencing the aforementioned water quality parameters. In particular, COD, Coliform (total), and NH3-N have higher R^2 values for the riparian scale during the dry season (**Figure 5.11 & Appendix H**). This could be due to the dispersion of pollutants from the WWTPs down the river.

In addition, E.coli, Na, and Turbidity have significantly higher R^2 values during the dry season than the wet season under the combined scenario (**Figure 5.11 & Appendix H**). While WWTPs variables did not appear as important predictors for these parameters **(Table 5.4)**, the large increase in R^2 values when WWTPs variables are included in the MLR models provides some evidence that processes associated to point sources could have contributed to the levels of these water quality parameters, and they are more dominant during the dry season than wet season.

Point source variables are important predictors for only PO4, NH3-N, and Temperature, albeit Temperature showed considerably small correlation with point source variables **(Table 5.4)**. Design PE is negatively related to PO4 during the wet season, which could suggest that PO4 are possibly removed from larger WWTPs with greater PE design. On the other hand, number of WWTPs **(B=0.302)** is positively related to NH3-N during the dry season, while Agriculture (Oil palm) AI is negatively related to NH3-N during the wet season **(B=-0.35)** **(Table 5.4)**. This suggests that influence of WWTPs on NH3-N levels is greater during the dry season, possibly attributed to dilution effects during the wet season. Under the combined scenario, it also concurs with the finding that NH3-N is positively related to Water Bodies LSI under the land use only scenario as effluent from WWTP are discharged into rivers. In addition, the existing WWTPs are not developed for the removal of NH3-N, and they are discharged into waterways until they are gradually diluted to less harmful compounds (IWK, 2019a). The finding also demonstrates that NH3-N is highly positively related to WWTPs compared to other water quality parameters.

Findings for the wet season also suggests that run off from land use during the wet season was more important in influencing NH3-N. In particular, negative relation between NH3-N and Agriculture (Oil palm) AI suggests that more fragmented oil palm give rise to higher NH3-N, likely attributed to greater erosion of fertilisers enriched with ammonium sulfate, a compound used commonly in oil palm plantations in Malaysia (Zakaria & Tarmizi, 2007).

Table 5.4. Redundancy Analysis results

Scale	Dry season: Explained Variation (%) Axis 1	Axis 2	Total axes	pseudo-F	p-value	Parameters selected by MLR	Wet season: Explained Variation (%) Axis 1	Axis 2	Total axes	pseudo-F	p-value	Parameters selected by MLR
(a) Land use only scenario												
Reach	21.0	10.7	52.1	3.267	0.001	lsi_3, pland_2, pland_8, shdi, lsi_4, ai_4, pland_4, pland_5,pland_6, pland_3, lsi_6, pland_7, lsi_7	23.6	11.4	53.0	3.380	0.003	ai_5, pladj_4, pland_5, lsi_4, pland_4, pland_6, ai_4, pladj_6,ai_2, lsi_3, pland_8, pland_2
Riparian	18.8	6.51	39.0	2.768	0.001	pland_7, ai_3, ai_5, pland_4, pland_5, pland_1, pladj_8, lsi_8, pland_2	21.7	7.49	38.9	3.176	0.004	pland_1, ai_5, pland_4, ai_4, pland_2, pladj_5, pland_7, pland_5
Sub-basin	14.8	9.29	41.4	2.372	0.001	ai_8, lsi_7, pland_7, pland_8, pladj_3, ai_7, shdi, ai_3, pladj_6+ lsi_5, pladj_5	7.40	5.08	21.5	1.602	0.050	ai_2, ai_6, ai_7, lsi_7, pladj_3, pland_2, lsi_4
(b) Combined scenario												
Reach	18.7	8.63	47.1	2.991	0.004	lsi_3, pland_2, pladj_3, pland_5, lsi_4, ai_4, pland_4, pland_6,ai_6, DESIGNPE, DESIGNPE.den.wt	21.5	7.55	44.4	2.149	0.024	lsi_3, pland_2, pland_6, lsi_4, pland_4, ai_4, pland_5, pland_7,lsi_7, ai_3, pladj_2, pladj_6, DESIGNPE

Riparian	19.0	7.23	39.4	2.817	0.003	pland_7, pladj_5, pladj_8, pland_5, pland_1, ai_3, ai_5, NUM, pland_2	21.5	6.44	35.6	3.863	0.002	pland_7, pland_4, pland_5, ai_5, pland_1, pland_2
Sub-basin	13.3	5.50	27.6	2.227	0.025	ai_5, pland_3, pland_7, pland_8, ai_2, ai_6, shdi	17.3	8.39	41.0	2.340	0.003	pland_2, lsi_5, ai_2, shdi, lsi_4, ai_6, ai_7, lsi_7, lsi_2,pladj_3, pland_5

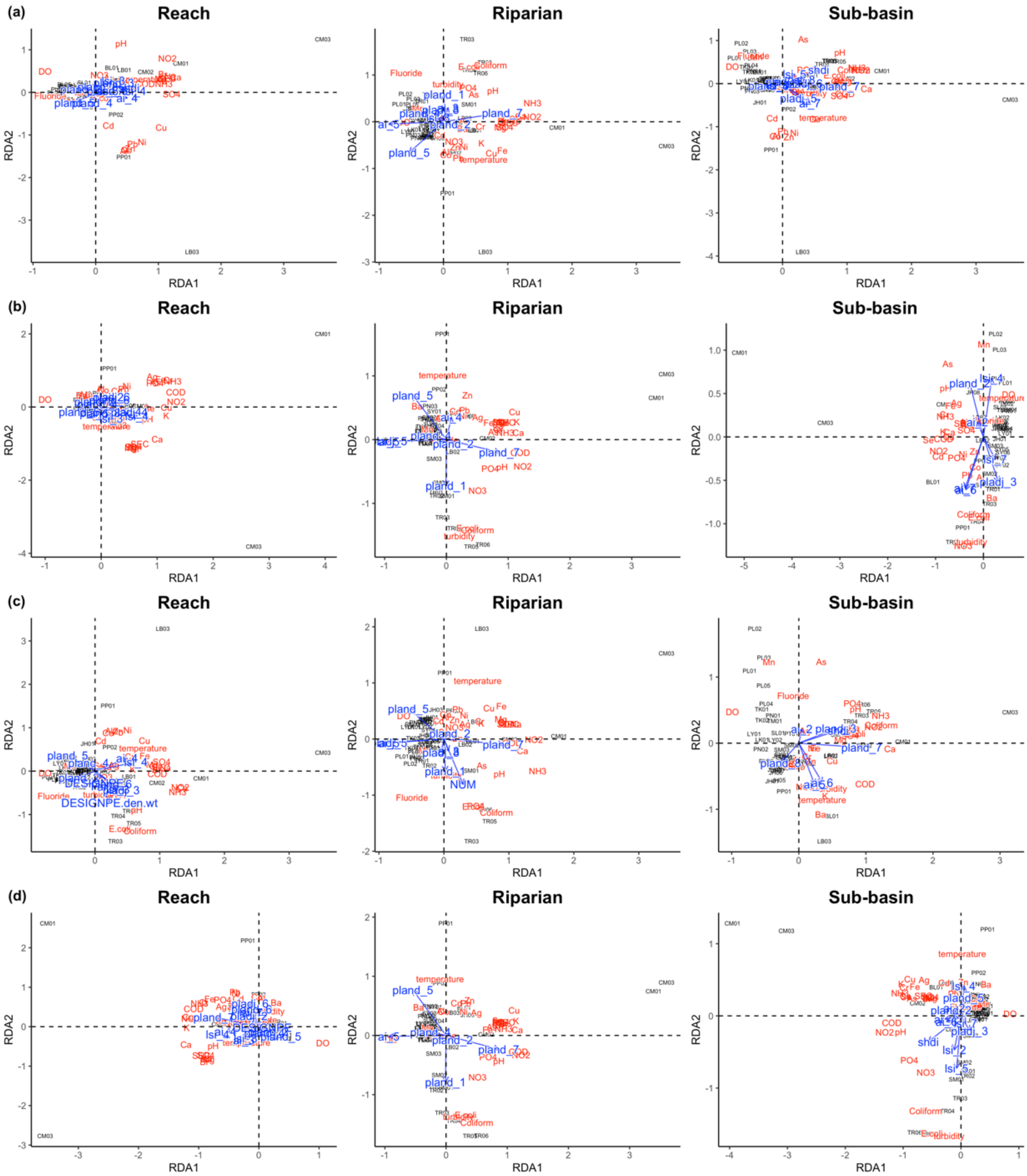


Figure 5.13. RDA results across seasons and landscape scales (a) Land use only scenario during the dry season (b) Land use only scenario during the wet season scales (c) Combined scenario during the dry season (d) Combined scenario during the wet season

Notes: only significant variables ($p<0.01$) from **Table** 5.4 are shown; metrics include the percentage of landscape (PLAND), landscape shape index (LSI), aggregation index (AI). Metric values for land use category: 1 Agriculture (Others), 2 (Bare land and shrubs), 3 Urban (Industry), 4 (Forest), 5 Agriculture (Oil palm), 6 Urban (Residential), 7 Urban (Others) and 8 (Water Bodies).

5.15 Relationship between sites, water quality and land use

RDA triplots explicitly display the relationship between sites, water quality, and land use. The RDA triplot's canonical axes explained greater variation for the land use only scenario (**Figure** 5.13). This might be due to the fact that only few water quality parameters (e.g. PO4, NH3-N, and Temperature out of 34 total water quality parameters) were correlated to point source variables, while other water quality parameters were correlated to land use variables. Therefore, a greater proportion of the total variation was better explained by land use variables. Under both scenarios, RDA triplot's canonical axes explained more of the variation for the wet season, which demonstrates that land use (non-point source) influence on water quality largely dominates during the wet season. This phenomenon aligns with observations from other studies which concluded that non-point source pollution (such as land use influence) is dependent on surface runoff, which predominates during the wet season (Bowes et al., 2008; Zhou et al., 2016).

E.coli, Coliform (total), NH3-N, PO4, NO2-N, and COD were most frequently clustered together in the RDA triplots, demonstrating high correlations between these water quality parameters **(Figure 5.13)**. In addition, they are most frequently correlated with Urban (Residential) and Urban (Others) metrics, suggesting that Urban (Residential) and Urban (Others) land use are significant sources for these water quality parameters. TR, LB, and CM sites are highly correlated with these parameters and land use metrics, and these sites coincide with being major residential and urban areas with the JRB. These water quality parameters are usually associated with the inevitable production of waste from anthropogenic activities in urban areas which can enter into nearby rivers through storm runoff (Lee et al., 2002; Li et al., 2007; Memon et al., 2017). Further, human waste from residential areas, food and beverage processing plants, and detergents all contribute to high levels of Coliform (total), nutrients, and organic matter including E.coli, NH3-N, PO4, NO2-N, and COD (Kundu et al., 2015; Rochelle-Newall et al., 2016; Zhou et al., 2016).

RDA triplots also show high positive correlations between PL sites, forest metrics, and Mn, which agrees with the spatial distribution of Mn (**Figure** 5.13). The

RDA triplots have consistently show positive correlations between As and Forest metrics for the reach scale, which may imply that the source of As is likely of natural origin, such as soil. The RDA triplots also show that Cd and Pb are positively correlated, especially during the wet season.

5.16 Limitations and Recommendations

5.16.1 Development of a WQI

Censored data and its limitations are rarely studied with regards to the formulation of WQI due to many uncertainties associated with it. The segmented approach in creating SIs is more favourable for data with left-censored values as SIs are defined by observations falling within a range of values. Although the method is rather coarse at handling left-censored data, the WQI will not be affected greatly as long as the censored data (detection limit) is much lower than the water quality standard. On the other hand, if SIs were computed using non-segmented approaches, one will need to ensure that the data does not contain any censored values. Such approaches are challenging especially for certain water quality parameters which are found in naturally low concentrations unless external inputs (e.g. geology and pollution) results in elevated concentrations. Otherwise, more resources have to be invested into expensive analytical instruments with very low detection limit capabilities, which is often challenging for developing countries with limited resources. Thus, future studies can delve into the utility of censored data in the formulation of WQI.

Sensitivity analysis on weights can be further investigated as weights assigned to parameters are usually based on literature values or subjectivity. Some studies (e.g. Dutta et al. (2018)) calculate weights for each parameter based on its corresponding water quality standards, albeit such method has its limitations when a water quality parameter has a non-linear relationship with water quality (e.g. pH and Temperature). Parameter sensitivity analysis and sensitivity of parameters to missing values could only be conducted one-at-a-time in this study. The vast number of different parameter combinations of parameters were not considered in this study due to computational constraints. A global parameter sensitivity analysis could be considered for future studies as well.

It is always important to report $WQI_{WEIGHTED}$ and $WQI_{NON-WEIGHTED}$, where $WQI_{NON-WEIGHTED}$ informs user on the state of general water quality, while $WQI_{WEIGHTED}$ is better suited for informing water quality on specified usage (e.g. ecological health of water, drinking water, or for specified concerns such as nutrients and heavy metals). Other water quality parameters such as contaminants of emerging concern (CEC) (e.g. personal care products, pesticides, and pharmaceutical drugs) may also be included in WQI if resources allow. In addition, other forms of indices such as bioindicators (e.g. aquatic macroinvertebrates) can be used to assess ecological health of rivers, albeit such assessments can be time consuming and requires knowledge from aquatic ecology experts. Such findings can be useful to corroborate the results from WQI.

5.16.2 Land use-water quality relationships

Point sources variables were only obtained from WWTP, and thus other point sources that could affect water quality were not considered. For instance, Johor is largely a consummate agricultural state, with much of its land use consisting of palm oil agriculture. Therefore, a significant point source could be Palm Oil Mill Effluent (POME). With more knowledge in its discharge and spatial distribution, accounting for POME could enhance the performance of the models and aid in understanding the influence of POME on surface water quality. In addition, industrial effluents as point sources, if available, should be included to better explain variations in heavy metals.

Many of the studies conducted have simplified land use categories for easy statistical analysis. This is especially so for MLR statistical methods that requires the number of explanatory variables to be smaller than the number of observations to ensure the reliability of models. It should be noted that aggregating land use may mask specific relationships between certain water quality parameters and land use. For instance, urban land use consists of multiple sub-land uses such as institutions, commercial shops, and roads, and each sub-land use produce varying composition of pollutants. This limitation will always be present when conducting such analyses.

As only ten months of data was analysed, there exist uncertainties in whether land use relationships with water quality can hold true for other years due to potential land use change and variations in climatic variables. Collection of future water quality data and land use information may be needed to validate the findings of this study.

In this study, it was observed that point source variables (i.e. WWTPs) are highly positively correlated to NH3-N, which demonstrates that a major source of NH3-N comes from the effluent of WWTPs. In addition, E.coli, NH3-N, PO4, NO2-N, and COD are highly correlated to Urban (Others) and Urban (Residential) land use. We recommend that soft-engineering measures such as increasing vegetation buffer zones along the river can be employed at the riparian scale to attenuate the transport of pollutants including sediments, untreated manure and nutrients. This is because the riparian region is the most important spatial scale for E.coli and NH3-N. While nutrient removal (tertiary treatment) for WWTPs has yet to be implemented, we suggests prioritizing nutrient removal process for WWTPs in urban areas with complex river configurations to prevent high NH3-N concentrations in these area.

6. CONCLUSION

Reoccurring river pollution and water scarcity in light of climate change has escalated concerns regarding long-term yield and sustainable consumption of water resources. Moreover, land use change with increasing urbanisation, coupled with the threats of climate change could alter hydrological parameters, and exacerbate water pollution in the Johor River. As such, there is an urgent need for better water quality monitoring within the JRB to better improve water management strategies through developing an effective WQI that optimises both water quality information and resources. Ultimately, the very purpose of a WQI aims to condense water quality information to bridge science and decision-making, which makes the WQI an important user-based tool. In particular, the sensitivity analysis has shown that an optimal sample of around 130 samples will be required if a 2% change in WQI_{CHANGE} can be tolerated. The performance of $WQI_{ADJUSTED}$ is comparable to $WQI_{NON\text{-}ADJUSTED}$ given that missing values do not exceed 10%.

The WQI results were contextualized based on the relationship with its surrounding land use. Results from Multivariate Linear Regression (MLR) showed that most water quality parameters were generally well-explained by percentage land use (PLAND) metrics, given the high frequency of PLAND metrics in the MLR results for both dry and wet seasons. In particular, agriculture and forest land use are more commonly associated with faecal coliforms and heavy metals respectively. In addition, Redundancy Analysis (RDA) results generally showed that proportion of explained variance are greater for the land use only scenario (52.1% for dry season and 53% for wet season) than the combined scenario (47.1% for dry season and 44.4% for wet season). As such, most water quality parameters can be largely explained by land use variables. However, it is worthy to note that certain water quality parameters like Na, NH3-N, Coliform (total), E.coli, Turbidity, and COD showed relatively high correlation under combined scenario, indicating that point source influence on these water quality parameters are significant, especially during the dry season. This also indicates that point sources (effluent from WWTPs) are highly associated to these water quality parameters, and point source influence on these water quality was more dominant than that of land use, especially during the dry season. On the other hand, land use impact on these water quality was more pronounced during the wet season due to the dominance of overland flow during the wet season.

This study aims to serve as a framework and preliminary study for coupling water quality information with decision making grounded by statistical analysis; provide greater insight and understanding on the influence of land use on surface water quality, in hopes to better inform policy makers on land use management, environmental monitoring, and water management strategies.

REFERENCES

Abbasi, S. A., & Arya, D. (2000). *Environmental Impact Assessment: Available Techniques, Emerging Trends*: Discovery Publishing House.

Abbasi, T., & Abbasi, S. A. (2012). *Water Quality Indices*: Elsevier.

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(1), 97-106.

Abdullah, M. P., Waseem, S., Bai, V. R., & Mohsin, I.-u. (2008). Development of new water quality model using fuzzy logic system for Malaysia. *Open Environmental Sciences, 2*(1), 101-106.

Adey, W., & Loveland, K. (2007). Chapter 4: Water Composition: Management of Salinity, Hardness, and Evaporation. Dynamic Aquaria: London: Academic Press.

Adeyi, A. A., & Babalola, B. A. (2017). Lead and cadmium levels in residential soils of Lagos and Ibadan, Nigeria. *Journal of Health and Pollution, 7*(13), 42-55.

Ahmad, N. (2016). Water samples from Johor rivers show high levels of chemical waste, says rep. Retrieved from https://www.thestar.com.my/metro/community/2016/09/05/do-something-about-pollution-water-samples-from-johor-rivers-show-high-levels-of-chemical-waste-says/

Ahmad, T. A. R., Ahmad, T. R., & Abd, K. W. (2004). Analysis of the concentrations of natural Radionuclides in rivers in Kota Tinggi district, Malaysia. *Journal of Nuclear and Related Technologies, 1*(1), 41-52.

Ai, L., Shi, Z., Yin, W., & Huang, X. (2015). Spatial and seasonal patterns in stream water contamination across mountainous watersheds: linkage with landscape characteristics. *Journal of Hydrology, 523*, 398-408.

Akkoyunlu, A., & Akiner, M. E. (2012). Pollution evaluation in streams using water quality indices: A case study from Turkey's Sapanca Lake Basin. *Ecological Indicators, 18*, 501-511.

Alves, M. T. R., Teresa, F. B., & Nabout, J. C. (2014). A global scientific literature of research on water quality indices: trends, biases and future directions. *Acta Limnologica Brasiliensia, 26*(3), 245-253.

American Public Health Association, A. (1995). *Standard methods for the examination of water and wastewater* (Vol. 21): American public health association Washington, DC.

Amiri, B. J., & Nakane, K. (2009). Modeling the linkage between river water quality and landscape metrics in the Chugoku district of Japan. *Water Resources Management, 23*(5), 931-956.

Arman, N. Z., Said, M. I. M., & Azman, S. (2013). Anthropogenic influences on aquatic life community and water quality status in Mengkibol River, Kluang, Johor, Malaysia. *Journal of Applied Sciences in Environmental Sanitation, 8*(3).

Auxtero, E. A., & Shamshuddin, J. (1991). Growth of oil palm (Elaeis guineensis) seedlings on acid sulfate soils as affected by water regime and aluminium. *Plant and Soil, 137*(2), 243-257.

Awang, H., Daud, Z., & Hatta, M. Z. M. (2015). Hydrology properties and water quality assessment of the Sembrong Dam, Johor, Malaysia. *Procedia-Social and Behavioral Sciences, 195*, 2868-2873.

Baker, A. (2006). Land use and water quality. *Encyclopedia of Hydrological Sciences*.

Bartlett, R., & James, B. (1979). Behavior of chromium in soils: III. Oxidation 1. *Journal of Environmental Quality, 8*(1), 31-35.

Bates, B., Kundzewicz, Z., & Wu, S. (2008). *Climate change and water*: Intergovernmental Panel on Climate Change Secretariat.

Berka, C., Schreier, H., & Hall, K. (2001). Linking water quality with agricultural intensification in a rural watershed. *Water, Air and Soil Pollution, 127*(1-4), 389-401. doi:10.1023/a:1005233005364

Bilgin, A. (2018). Evaluation of surface water quality by using Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) method and discriminant analysis method: a case study Coruh River Basin. *Environmental Monitoring Assessment, 190*(9), 554. doi:10.1007/s10661-018-6927-5

Bo, W., Wang, X., Zhang, Q., Xiao, Y., & Ouyang, Z. (2017). Influence of Land Use and Point Source Pollution on Water Quality in a Developed Region: A Case Study in Shunde, China. *International Journal of Environmental Research and Public Health, 15*(1). doi:10.3390/ijerph15010051

Boulay, N., & Edwards, M. (2000). Copper in the urban water cycle. *Critical Reviews in Environmental Science and Technology, 30*(3), 297-326.

Bowes, M. J., Smith, J. T., Jarvie, H. P., & Neal, C. (2008). Modelling of phosphorus inputs to rivers from diffuse and point sources. *Science of the Total Environment, 395*(2-3), 125-138.

Bu, H., Meng, W., Zhang, Y., & Wan, J. (2014). Relationships between land use patterns and water quality in the Taizi River basin, China. *Ecological Indicators, 41*, 187-197.

Buck, O., Niyogi, D. K., & Townsend, C. R. (2004). Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environmental Pollution, 130*(2), 287-299.

Burstyn, I. (2004). Principal component analysis is a powerful instrument in occupational hygiene inquiries. *Annals of Occupational Hygiene, 48*(8), 655-661.

Cardenas, M. B. (2008). The effect of river bend morphology on flow and timescales of surface water–groundwater exchange across pointbars. *Journal of Hydrology, 362*(1-2), 134-141.

Carey, R. O., Migliaccio, K. W., Li, Y., Schaffer, B., Kiker, G. A., & Brown, M. T. (2011). Land use disturbance indicators and water quality variability in the Biscayne Bay Watershed, Florida. *Ecological Indicators, 11*(5), 1093-1104. doi:10.1016/j.ecolind.2010.12.009

CCME. (2001). *Canadian water quality guidelines for the protection of aquatic life. CCME water quality index 1.0, User's Manual, Winnipeg, Manitoba, Canada.* . Retrieved from

Chan, N. W. (2012). Managing urban rivers and water quality in Malaysia for sustainable water resources. *International Journal of Water Resources Development, 28*(2), 343-354.

Cheng, P., Meng, F., Wang, Y., Zhang, L., Yang, Q., & Jiang, M. (2018). The Impacts of Land Use Patterns on Water Quality in a Trans-Boundary River Basin in Northeast China Based on Eco-Functional Regionalization. *International Journal of Environmental Research and Public Health, 15*(9). doi:10.3390/ijerph15091872

Chew, V. (2011). Singapore-Malaysia water agreements Retrieved from Singapore Infopedia website:

Chuah, C. J., Ho, B. H., & Chow, W. T. (2018). Trans-boundary variations of urban drought vulnerability and its impact on water resource management in Singapore and Johor, Malaysia. *Environmental Research Letters, 13*(7), 074011.

Chuah, C. J., & Ziegler, A. D. (2018). Temporal variability of faecal contamination from on-site sanitation systems in the groundwater of Northern Thailand. *Environmental Management, 61*(6), 939-953.

Conesa Fdez-Vitoria, V. (1995). Methodological Guide for Environmental Impact Evaluation (Guía Metodológica para la Evaluación del Impacto Ambiental), Madrid. marine waters using phytoplankton communities. *Marine Pollution Bulletin, 55*, 91-103.

Core, T. R. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL https://www. R-project. org.*

Cude, C. G. (2001). Oregon water quality index a tool for evaluating water quality management effectiveness 1. *Journal of the American Water Resources Association, 37*(1), 125-137.

Czarnowska, K. (1996). Total content of heavy metals in parent rocks as reference background levels of soils. *Soil Science Annual*.

Davies Jr, P. M., Bunn Jr, S. E., & Hamilton Jr, S. K. (2008). Primary production in tropical streams and rivers *Tropical Stream Ecology* (pp. 23-42): Elsevier.

DID. (2015). *Water Resources and Hydrology Publication—Hydrological Procedures*. Retrieved from (www.water.gov.my/resource-centre-mainmenu-255/publications/waterresources-management-and-hydrology-mainmenu-295?lang=en:

Ding, J., Jiang, Y., Liu, Q., Hou, Z., Liao, J., Fu, L., & Peng, Q. (2016). Influences of the land use pattern on water quality in low-order streams of the Dongjiang River basin, China: A multi-scale analysis. *Science of the Total Environment, 551-552*, 205-216. doi:10.1016/j.scitotenv.2016.01.162

Dojlido, J., Raniszewski, J., & Woyciechowska, J. (1994). Water quality index applied to rivers in the Vistula river basin in Poland. *Environmental Monitoring and Assessment, 33*(1), 33-42.

Dunnette, D. (1979). A geographically variable water quality index used in Oregon. *Journal (Water Pollution Control Federation)*, 53-61.

Dutta, S., Dwivedi, A., & Suresh Kumar, M. (2018). Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. *Environmental Monitoring Assessment, 190*(12), 718. doi:10.1007/s10661-018-7100-x

Ewaid, S. H., Abed, S. A., & Kadhum, S. A. (2018). Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis. *Environmental Technology & Innovation, 11*, 390-398. doi:10.1016/j.eti.2018.06.013

Fageria, N., Baligar, V., & Clark, R. (2002). Micronutrients in crop production *Advances in Agronomy* (Vol. 77, pp. 185-268): Elsevier.

FAO. (1979). *Soil map of the world* Vol. IX Southeast Asia. F. a. A. O. (FAO) (Ed.)

Fennell, J., Zawadzki, A., & Cadman, C. (2006). Influence of Natural vs. Anthropogenic stresses on water resource sustainability: A case study. *Water Science and Technology, 53*(10), 21-27.

Fujii, K., Uemura, M., Hayakawa, C., Funakawa, S., & Kosaki, T. (2013). Environmental control of lignin peroxidase, manganese peroxidase, and laccase activities in forest floor layers in humid Asia. *Soil Biology and Biochemistry, 57*, 109-115.

Gabdo, B., & Abdlatif, I. B. (2013). Analysis of the benefits of livestock to oil palm in an integrated system: Evidence from selected districts in Johor, Malaysia. *Journal of Agricultural Science, 5*(12), 47.

Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin, 64*(11), 2409-2420.

Giri, S., & Qiu, Z. (2016). Understanding the relationship of land uses and water quality in Twenty First Century: A review. *Journal of Environmental Management, 173*, 41-48. doi:10.1016/j.jenvman.2016.02.029

Gitau, M. W., Chen, J., & Ma, Z. (2016). Water quality indices as tools for decision making and management. *Water Resources Management, 30*(8), 2591-2610.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5): Prentice hall Upper Saddle River, NJ.

Hallock, D. (2002). *A water quality index for ecology's stream monitoring program*: Washington State Department of Ecology Olympia.

Hanh, T. M., Sthiannopkao, P., Ba, S. T., Kim, D., & Kyoung-Woong. (2011). Development of water quality indexes to identify pollutants in Vietnam's surface water. *Journal of Environmental Engineering, 137*(4), 273-283.

Harrell, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*: Springer.

Hartemink, A. E. (2006). Soil erosion: perennial crop plantations. *Encyclopedia of Soil Science*, 1613-1617.

Harun, S., Dambul, R., Abdullah, M. H., & Mohamed, M. (2014). Spatial and seasonal variations in surface water quality of the Lower Kinabatangan River Catchment, Sabah, Malaysia. *Journal of Tropical Biology & Conservation (JTBC), 11*.

Heinonen-Tanski, H., & Uusi-Kämppä, J. (2001). Runoff of faecal microorganisms and nutrients from perennial grass ley after application of slurry and mineral fertiliser. *Water Science and Technology, 43*(12), 143-146.

Huang, J., Huang, Y., Pontius Jr, R. G., & Zhang, Z. (2015). Geographically weighted regression to measure spatial variations in correlations between water pollution versus land use in a coastal watershed. *Ocean & Coastal Management, 103*, 14-24.

Huang, X., Zhu-Barker, X., Horwath, W. R., Faeflen, S. J., Luo, H., Xin, X., & Jiang, X. (2016). Effect of iron oxide on nitrification in two agricultural soils with different pH. *Biogeosciences, 13*(19), 5609-5617.

Hudson, N., Baker, A., Ward, D., Reynolds, D. M., Brunsdon, C., Carliell-Marquet, C., & Browning, S. (2008). Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. *Science of the Total Environment, 391*(1), 149-158.

Hur, J., Nguyen, H. V.-M., & Lee, B.-M. (2014). Influence of upstream land use on dissolved organic matter and trihalomethane formation potential in watersheds for two different seasons. *Environmental Science and Pollution Research, 21*(12), 7489-7500.

Husson, F., Josse, J., Le, S., Mazet, J., & Husson, M. F. (2020). Package 'FactoMineR'. *Package FactorMineR*.

Irvine, K., Vermette, S., & Mustafa, F. B. (2012). *The "black waters" of Malaysia: tracking water quality from the peat swamp forest to the sea.* Paper presented at the 2012 International Symposium on Geomatics for Integrated Water Resource Management.

IWK, I. W. K. (2019a). Ammonia. Retrieved from https://www.iwk.com.my/do-you-know/ammonia

IWK, I. W. K. (2019b). Effluent Standards. Retrieved from https://www.iwk.com.my/do-you-know/effluent-standards

IWK, I. W. K. (2019c). Sewerage Treatment Methods. Retrieved from https://www.iwk.com.my/do-you-know/sewage-treatment-methods

Johnson, J. Y., Thomas, J., Graham, T., Townshend, I., Byrne, J., Selinger, L., & Gannon, V. P. (2003). Prevalence of Escherichia coli O157: H7 and Salmonella spp. in surface waters of southern Alberta and its relation to manure sources. *Canadian Journal of Microbiology, 49*(5), 326-335.

Juwana, I., Muttil, N., & Perera, B. (2012). Indicator-based water sustainability assessment—A review. *Science of the Total Environment, 438*, 357-371.

Kandler, M., Blechinger, K., Seidler, C., Pavlu, V., Sanda, M., Dostal, T., . . . Stich, M. (2017). Impact of land use on water quality in the upper Nisa catchment in the Czech Republic and in Germany. *Science of the Total Environment, 586*, 1316-1325. doi:10.1016/j.scitotenv.2016.10.221

Kannel, P. R., Lee, S., Lee, Y.-S., Kanel, S. R., & Khan, S. P. (2007). Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. *Environmental Monitoring and Assessment, 132*(1-3), 93-110.

Katimon, A., Shahid, S., & Mohsenipour, M. (2017). Modeling water quality and hydrological variables using ARIMA: a case study of Johor River, Malaysia. *Sustainable Water Resources Management, 4*(4), 991-998. doi:10.1007/s40899-017-0202-8

Kieffer, K. M. (1998). Orthogonal versus Oblique Factor Rotation: A Review of the Literature regarding the Pros and Cons.

Kline, P. (2014). *An easy guide to factor analysis*: Routledge.

Koçer, M. A. T., & Sevgili, H. (2014). Parameters selection for water quality index in the assessment of the environmental impacts of land-based trout farms. *Ecological Indicators, 36*, 672-681. doi:10.1016/j.ecolind.2013.09.034

Kundu, S., Coumar, M. V., Rajendiran, S., Rao, A., & Rao, A. S. (2015). Phosphates from detergents and eutrophication of surface water ecosystem in India. *Current Science*, 1320-1325.

Kwak, J., Khang, B., Kim, E., & Kim, H. (2012). Estimation of biochemical oxygen demand based on dissolved organic carbon, UV absorption, and fluorescence measurements. *Journal of Chemistry, 2013*.

Lee, J., Bang, K., Ketchum Jr, L., Choe, J., & Yu, M. (2002). First flush analysis of urban storm runoff. *Science of the Total Environment, 293*(1-3), 163-175.

Lee, S.-W., Hwang, S.-J., Lee, S.-B., Hwang, H.-S., & Sung, H.-C. (2009). Landscape ecological approach to the relationships of land use patterns in watersheds to water quality characteristics. *Landscape and Urban Planning, 92*(2), 80-89. doi:10.1016/j.landurbplan.2009.02.008

Li, L.-Q., Yin, C.-Q., He, Q.-C., & Kong, L.-L. (2007). First flush of storm runoff pollution from an urban catchment in China. *Journal of Environmental Sciences, 19*(3), 295-299.

Lim, K. H., Lim, S. S., Parish, F., & Suharto, R. (2012). RSPO manual on best management practices (BMPs) for existing oil palm cultivation on peat: RSPO Secretariat Sdn Bhd.

Liou, S.-M., Lo, S.-L., & Wang, S.-H. (2004). A generalized water quality index for Taiwan. *Environmental Monitoring and Assessment, 96*(1-3), 35-52.

Lumb, A., Sharma, T. C., & Bibeault, J.-F. (2011). A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. *Water Quality, Exposure and Health, 3*(1), 11-24. doi:10.1007/s12403-011-0040-0

Manahan, S. (2017). *Environmental chemistry*: CRC press.

McBride, M., & Booth, D. B. (2005). Urban impacts on physical stream condition: effects of spatial scale, connectivity, and longitudinal trends. *Journal of the American Water Resources Association, 41*(3), 565-580.

McDonald, R. I., Douglas, I., Revenga, C., Hale, R., Grimm, N., Grönwall, J., & Fekete, B. (2011). Global urban growth and the geography of water availability, quality, and delivery. *Ambio, 40*(5), 437-446.

McGarigal, K. (2015). FRAGSTATS help. *University of Massachusetts: Amherst, MA, USA*.

McGarigal, K., & Marks, B. J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. *Gen. Tech. Rep., 351*.

McGrath, S., Sanders, J., & Shalaby, M. (1988). The effects of soil organic matter levels on soil solution concentrations and extractabilities of manganese, zinc and copper. *Geoderma, 42*(2), 177-188.

McPeek, M. A., & Kalisz, S. (1993). Population sampling and bootstrapping in complex designs: demographic analysis. *Design and Analysis of Ecological Experiments*, 232-252.

Mello, K. d., Valente, R. A., Randhir, T. O., dos Santos, A. C. A., & Vettorazzi, C. A. (2018). Effects of land use and land cover on water quality of low-order streams in Southeastern Brazil: Watershed versus riparian zone. *Catena, 167*, 130-138. doi:10.1016/j.catena.2018.04.027

Memon, S., Paule, M. C., Yoo, S., Umer, R., Lee, B.-Y., Sukhbaatar, C., & Lee, C.-H. (2017). Trend of storm water runoff pollutants temporal variability from different land use sites in Korea. *Desalination Water Treatment, 63*, 433-441.

MEWR, M. o. t. E. a. W. R. (2019). Oral reply by Masagos Zulkifli, Minister for the Environment and Water Resources, to Parliamentary Questions on Johor River Waterworks on 7 May 2019 [Press release]. Retrieved from https://www.mewr.gov.sg/news/oral-reply-by-masagos-zulkifli--minister-for-the-environment-and-water-resources--to-parliamentary-questions-on-johor-river-waterworks-on-7-may-2019

Mirza, N., Mahmood, Q., Maroof Shah, M., Pervez, A., & Sultan, S. (2014). Plants as useful vectors to reduce environmental toxic arsenic content. *The Scientific World Journal, 2014*.

MNRE. (2011). *Review of national water resources study (2000–2050) and formulation of national water resources policy*. Retrieved from http://www.water.gov.my/images/Hidrologi/NationalWaterResourcesStudy/Vol17Johor.pdf

MoEI. (2003). *The guidance of water quality status in Indonesia. Decree No. 115/2003* Ministry of the Environment of Indonesia).

National Research Council, N. (1977). Copper: medical and biological effects of environmental pollutants. *National Academy of Sciences, Washington, DC*.

Naubi, I., Zardari, N. H., Shirazi, S. M., Ibrahim, N. F. B., & Baloo, L. (2016). Effectiveness of Water Quality Index for Monitoring Malaysian River Water Quality. *Polish Journal of Environmental Studies, 25*(1).

Nazahiyah, R., Yusop, Z., & Abustan, I. (2007). Stormwater quality and pollution loading from an urban residential catchment in Johor, Malaysia. *Water Science and Technology, 56*(7), 1-9.

Nikoo, M. R., Kerachian, R., Malakpour-Estalaki, S., Bashi-Azghadi, S. N., & Azimi-Ghadikolaee, M. M. (2011). A probabilistic water quality index for river water quality assessment: a case study. *Environmental Monitoring and Assessment, 181*(1-4), 465-478.

Nor, Z. A., Mohd, I. M. S., Shamila, A., & Muhammad, H. M. H. (2013). Comparison between water quality index (WQI) and biological water quality index (BWQI) for water quality assessment: case study of Melana River, Johor. *Malaysian Journal of Analytical Sciences, 17*(2), 224-229.

Novotny, V. (2002). *Water quality: diffuse pollution and watershed management*: John Wiley & Sons.

O'Flynn, B., Regan, F., Lawlor, A., Wallace, J., Torres, J., & O'mathuna, C. (2010). Experiences and recommendations in deploying a real-time, water quality monitoring system. *Measurement Science and Technology, 21*(12), 124004.

O’Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2019). vegan: Community Ecology Package (Version R package version 2.5-6.). https://CRAN.R-project.org/package=vegan.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’hara, R., . . . Wagner, H. (2013). Package ‘vegan’. *Community Ecology Package, 2*(9), 1-295.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research, and Evaluation, 20*(1), 2.

Oudin, L., Meybeck, M., & Roussel, P. (1999). Système d’évaluation de la qualité de l’eau des cours d’eau. *Rapport de presentation SEQ-Eau (version 1), Agence de l’eau Loire-Bretagne, France*.

Pan, Y., Herlihy, A., Kaufmann, P., Wigington, J., Van Sickle, J., & Moser, T. (2004). Linkages among land-use, water quality, physical habitat conditions and lotic diatom assemblages: a multi-spatial scale assessment. *Hydrobiologia, 515*(1-3), 59-73.

Paul, S. (2005). *Bacterial Total Maximum Daily Load (TMDL): Development and evaluation of a new classification scheme for impaired waterbodies of Texas.* Texas A&M University.

Pesce, S. F., & Wunderlin, D. A. (2000). Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquía River. *Water Research, 34*(11), 2915-2926.

PLANMalaysia@Johor (Producer). (2017, 26/5/2019). Johor Landuse Portal. Retrieved from http://geoportal.johor.gov.my/petaawam/gunatanahsemasa/jbahru

Pocernich, M., & Litke, D. W. (1997). Nutrient concentrations in wastewater treatment plant effluents, South Platte River Basin. *Journal of the American Water Resources Association, 33*(1), 205-214. doi:10.1111/j.1752-1688.1997.tb04096.x

Pratt, J. W., Raiffa, H., Schlaifer, R. O., & Schlaifer, R. (1995). *Introduction to statistical decision theory*: MIT press.

Rakib, M. R. M., Bong, C. F. J., Khairulmazmi, A., Idris, A. S., Jalloh, M. B., & Ahmed, O. H. (2017). Association of Copper and Zinc levels in oil palm (Elaeis guineensis) to the spatial distribution of ganoderma species in the plantations on peat. *Journal of Phytopathology, 165*(4), 276-282.

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The Comprehensive R Archive Network.*

Richardson, J. (2017). Manganese and Mn/Ca ratios in soil and vegetation in forests across the northeastern US: Insights on spatial Mn enrichment. *Science of the Total Environment, 581*, 612-620.

Rochelle-Newall, E. J., Ribolzi, O., Viguier, M., Thammahacksa, C., Silvera, N., Latsachack, K., . . . Soulileuth, B. (2016). Effect of land use and hydrological processes on Escherichia coli concentrations in streams of tropical, humid headwater catchments. *Scientific Reports, 6*, 32974.

Rodrigues, V., Estrany, J., Ranzini, M., de Cicco, V., Martin-Benito, J. M. T., Hedo, J., & Lucas-Borja, M. E. (2018). Effects of land use and seasonality on stream water quality in a small tropical catchment: The headwater of Corrego Agua Limpa, Sao Paulo (Brazil). *Science of the Total Environment, 622-623*, 1553-1561. doi:10.1016/j.scitotenv.2017.10.028

Sahat, S., Yusop, Z., Askari, M., & Ziegler, A. (2016). *Estimation of soil erosion rates in oil palm plantation with different land cover.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Sánchez, E., Colmenarejo, M. F., Vicente, J., Rubio, A., García, M. G., Travieso, L., & Borja, R. (2007). Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators, 7*(2), 315-328.

Seeboonruang, U. (2012). A statistical assessment of the impact of land uses on surface water quality indexes. *Journal of Environmental Management, 101*, 134-142. doi:10.1016/j.jenvman.2011.10.019

Shah, M. F. (2019). 17,000 households in Kulai go dry after ammonia pollution in Sungai Sayong. Retrieved from https://www.thestar.com.my/news/nation/2019/04/04/17000-households-in-kulai-go-dry-after-ammonia-pollution-in-sungai-sayong/

Shen, Z., Hou, X., Li, W., Aini, G., Chen, L., & Gong, Y. (2015). Impact of landscape pattern at multiple spatial scales on water quality: A case study in a typical urbanised watershed in China. *Ecological Indicators, 48*, 417-427.

Shrestha, J., Rich, J. J., Ehrenfeld, J. G., & Jaffe, P. R. (2009). Oxidation of ammonium to nitrite under iron-reducing conditions in wetland soils: Laboratory, field demonstrations, and push-pull rate determination. *Soil Science, 174*(3), 156-164.

Shrivastava, A. (2009). A review on copper pollution and its removal from water bodies by pollution control technologies. *Indian Journal of Environmental Protection, 29*(6), 552-560.

Shuhaimi-Othman, M., Lim, E. C., & Mushrifah, I. (2007). Water quality changes in Chini Lake, Pahang, West Malaysia. *Environmental Monitoring Assessment, 131*(1-3), 279-292. doi:10.1007/s10661-006-9475-3

Sidle, R. C., Ziegler, A. D., Negishi, J. N., Nik, A. R., Siew, R., & Turkelboom, F. (2006). Erosion processes in steep terrain—truths, myths, and uncertainties related to forest management in Southeast Asia. *Forest Ecology and Management, 224*(1-2), 199-225.

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research, 38*(18), 3980-3992.

Smith, D. G. (1990). A better water quality indexing system for rivers and streams. *Water Research, 24*(10), 1237-1244.

Srebotnjak, T., Carr, G., de Sherbinin, A., & Rickwood, C. (2012). A global Water Quality Index and hot-deck imputation of missing data. *Ecological Indicators, 17*, 108-119.

Sutadian, A. D., Muttil, N., Yilmaz, A. G., & Perera, B. (2016). Development of river water quality indices—a review. *Environmental Monitoring and Assessment, 188*(1), 58.

Swamee, P. K., & Tyagi, A. (2007). Improved method for aggregation of water quality subindices. *Journal of Environmental Engineering, 133*(2), 220-225.

Tan, M. L., Ibrahim, A. L., Yusop, Z., Duan, Z., & Ling, L. (2015). Impacts of land-use and climate variability on hydrological components in the Johor River basin, Malaysia. *Hydrological Sciences Journal, 60*(5), 873-889.

Tangang, F. T., Juneng, L., Salimun, E., Sei, K., & Halimatun, M. (2012). Climate change and variability over Malaysia: gaps in science and research information. *Sains Malaysiana, 41*(11), 1355-1366.

Tortajada, C., & Islam, S. (2011). Governance in urban water quality and water disasters: a focus on Asia. *Water International, 36*(6), 764-766.

Tripathi, M., & Singal, S. K. (2019). Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India. *Ecological Indicators, 96*, 430-436.

Tu, J. (2011). Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography, 31*(1), 376-392.

Varol, S., & Davraz, A. (2015). Evaluation of the groundwater quality with WQI (Water Quality Index) and multivariate analysis: a case study of the Tefenni plain (Burdur/Turkey). *Environmental Earth Sciences, 73*(4), 1725-1744.

Walsh, C. J., & Kunapo, J. (2009). The importance of upland flow paths in determining urban effects on stream ecosystems. *Journal of the North American Benthological Society, 28*(4), 977-990.

Wu, Z., Wang, X., Chen, Y., Cai, Y., & Deng, J. (2018). Assessing river water quality using water quality index in Lake Taihu Basin, China. *Science of the Total Environment, 612*, 914-922. doi:10.1016/j.scitotenv.2017.08.293

Yang, L. E., Chan, F. K. S., & Scheffran, J. (2018). Climate change, water management and stakeholder analysis in the Dongjiang River basin in South China. *International Journal of Water Resources Development, 34*(2), 166-191.

Yidana, S. M., & Yidana, A. (2010). Assessing water quality using water quality index and multivariate analysis. *Environmental Earth Sciences, 59*(7), 1461-1473.

Yu, S., Xu, Z., Wu, W., & Zuo, D. (2016). Effect of land use types on stream water quality under seasonal variation and topographic characteristics in the Wei River basin, China. *Ecological Indicators, 60*, 202-212.

Yule, C. M., & Gomez, L. N. (2009). Leaf litter decomposition in a tropical peat swamp forest in Peninsular Malaysia. *Wetlands Ecology and Management, 17*(3), 231-241.

Zaiha, A. N., Ismid, M. M., & Azri, M. S. (2015). Effects of logging activities on ecological water quality indicators in the Berasau River, Johor, Malaysia. *Environmental Monitoring and Assessment, 187*(8), 493.

Zakaria, Z., & Tarmizi, A. (2007). Efficient use of urea as nitrogen fertilizer for mature oil palm in Malaysia. *RAE (Urea), 10*, 100.

Zhou, P., Huang, J., Pontius, R. G., Jr., & Hong, H. (2016). New insight into the correlations between land use and water quality in a coastal watershed of China: Does point source pollution weaken it? *Science of the Total Environment, 543*(Pt A), 591-600. doi:10.1016/j.scitotenv.2015.11.063

APPENDICES

Appendix A. Water quality standards for National Water Quality Standard (NWQS) Malaysia

(a) Water quality standards

Parameters[a]	National Water Quality Standards (Malaysia)			
	Class 1	**Class 2**	**Class 3**	**Class 4**
Ag (ppb)	0	50	0.2	0.2
Al (ppb)	0	1500	60	500
As (ppb)	0	50	400	100
Ba (ppb)	0	1000	1000*	1000*
Br (ppm)	0	-	-	-
Ca (ppm)	0	400	400	400
Cd (ppb)	0	10	10	10
Cl (ppm)	0	200	200	80
Co (ppb)	0	-	-	-
COD (mg/l)	10	25	50	100
Coliform (total)	100	5000	50000	50000
Cr (ppb)	0	50	3900	100
Cu (ppb)	0	20	20	200
DO (mg/l)	7	6	4	2
E.coli (MPN/100ml)	10	400	5000	5000
Fe (ppb)	0	1000	1000	1000
F^- (ppm)	0	1.5	10	1
K (ppm)	0	-	-	-
Mg (ppm)	0	-	-	-
Mn (ppb)	0	100	100	200
Na (ppm)	0	-	-	-
NH_3-N (ppm)	0.1	0.3	0.9	2.7
Ni (ppb)	0	50	900	200
NO_2-N (ppm)	0	0.4	0.4	0.4
NO_3^- (ppm)	0	7	7	5
Pb (ppb)	0	50	20	5000
pH (unitless)	7	6	5	5
PO_4^{3-} (ppm)	0.5*	0.7*	0.7*	0.7*
Se (ppb)	0	10	250	20
SEC (uS/cm)	1000	1000	1000*	6000
SO_4^{2-} (ppm)	0	250	250	250
Temperature (oC)	27	29	29	29
Turbidity (NTU)	5	50	50	50
Zn (ppb)	0	5000	400	2000

(b) WQI classes

Classes	Weighted	Non-weighted
Class 1	92.26	90.88
Class 2	24.52	33.82
Class 3	22.26	32.35

Class 4	18.17	28.53

(c) Description of each classes

Class 1	Conservation of natural environment.
	Water Supply I - Practically no treatment necessary.
	Fishery I - Very sensitive aquatic species.
Class 2	Water Supply II - Conventional treatment.
	Fishery II - Sensitive aquatic species.
	Recreational use body contact.
Class 3	Water Supply III - Extensive treatment required.
	Fishery III - Common,of economic value and tolerant species; livestock drinking.
Class 4	Irrigation

Notes:

[a] Parameters measured in this study, which is not exhaustive of the total water quality parameters measured by NWQS

- Water quality standards for the particular water quality parameter is not indicated
* Water quality standards for the particular class is not indicated by NWQS and is replaced with USEPA's standard of a similar class

Appendix B. Effluent discharge of WWTPs with discharge standards A and B

Parameter	**Standard A (mg/L)**	**Standard B (mg/L)**
Temperature (Celcius)	40	40
pH Value	6.0 - 9.0	5.5 - 9.0
Biological Oxygen Demand	20	50
Chemical Oxygen Demand	50	100
Suspended Solids	50	100
Ammoniacal Nitrogen	15	25

Source: (IWK, 2019b)

Appendix C. Subindices breaks for each parameter

Parameters	**SI**											
	100		**90**	**80**	**70**	**60**	**50**	**40**	**30**	**20**	**10**	**0**
Ag (ppb)	0.00099	0.00227	0.0052	0.0119	0.0273	0.0624	0.143	0.327	0.75	1.72	3.93	9
Al (ppb)	0.00592	0.0216	0.0788	0.288	1.05	3.83	14	51	186	680	2480	9060
As (ppb)	0.0398	0.0668	0.112	0.188	0.316	0.531	0.891	1.5	2.51	4.22	7.08	11.9
Ba (ppb)	1.44	2.33	3.79	6.16	10	16.3	26.5	43	70	114	185	301
Br (ppm)	0.0215	0.0404	0.0757	0.142	0.266	0.499	0.936	1.76	3.29	6.17	11.6	21.7
Ca (ppm)	0.449	0.743	1.23	2.04	3.38	5.61	9.29	15.4	25.5	42.3	70.1	116
Cd (ppb)	0.00099	0.00227	0.00519	0.0119	0.0271	0.0621	0.142	0.325	0.743	1.7	3.89	8.89
Cl (ppm)	1.59	3.23	6.56	13.3	27	54.9	112	226	460	934	1900	3850
Co (ppb)	0.00694	0.0161	0.0374	0.0868	0.201	0.468	1.09	2.52	5.85	13.6	31.6	73.3
COD (ppm)	0.336	0.559	0.928	1.54	2.56	4.26	7.08	11.8	19.6	32.5	54	89.7
Coliform (total) (MPN/100ml)	40.6	89.6	198	436	962	2120	4680	10300	22800	50200	111000	244000
Cr (ppb)	0.00794	0.0161	0.0327	0.0663	0.134	0.273	0.553	1.12	2.28	4.62	9.37	19
Cu (ppb)	1	1.75	3.05	5.33	9.31	16.3	28.4	49.7	86.8	152	265	463
DO (mg/l)*		>7.5	7	6.5	6	5	4	3.5	3	2	1	<1
E.coli (MPN/100ml)	0.988	3.06	9.45	29.2	90.4	280	865	2680	8280	25600	79200	245000
Fe (ppb)	0.0168	0.051	0.155	0.471	1.43	4.35	13.2	40.2	122	371	1130	3420
Fluoride (ppm)	0.00073	0.00162	0.00358	0.0079	0.0174	0.0385	0.0849	0.187	0.413	0.913	2.01	4.45
K (ppm)	0.429	0.671	1.05	1.64	2.56	4.01	6.27	9.8	15.3	24	37.5	58.6
Mg (ppm)	0.225	0.429	0.815	1.55	2.95	5.6	10.7	20.3	38.5	73.2	139	265
Mn (ppb)	0.105	0.236	0.528	1.18	2.65	5.95	13.3	29.9	67	150	337	755

Na (ppm)	1.07	2.1	4.14	8.15	16.1	31.7	62.4	123	242	478	942	1860
NH3-N (ppm)	0.0149	0.0325	0.0711	0.155	0.34	0.743	1.62	3.55	7.76	17	37.1	81.1
Ni (ppb)	0.0376	0.0878	0.205	0.477	1.11	2.6	6.06	14.1	32.9	76.8	179	418
NO2-N (ppm)	0.00033	0.00066	0.00132	0.00263	0.00525	0.0105	0.0209	0.0418	0.0835	0.167	0.333	0.664
NO3 (ppm)	0.0731	0.152	0.314	0.651	1.35	2.79	5.79	12	24.9	51.5	107	221
Pb (ppb)	0.00198	0.00426	0.00913	0.0196	0.0421	0.0903	0.194	0.416	0.892	1.92	4.11	8.82
pH (unitless)*		7	7-8	8-8.5	8.5-9	6.5-7	6-6.5 or 9-9.5	5-6 or 9.5-10	4-5 or 10-11	3-4 or 11-12	2-3 or 12-13	Otherwise
PO4 (ppm)	0.0303	0.0446	0.0655	0.0963	0.141	0.208	0.305	0.449	0.659	0.969	1.42	2.09
Se (ppb)	0.00198	0.00416	0.00871	0.0182	0.0382	0.0801	0.168	0.352	0.737	1.54	3.23	6.77
SEC (µS/cm)	21.2	37.6	66.7	118	210	373	662	1180	2090	3700	6570	11700
SO4 (ppm)	0.242	0.486	0.977	1.96	3.95	7.94	16	32.1	64.5	130	261	524
Temperature (°C)	24.5	25.2	26	26.7	27.5	28.3	29.2	30	30.9	31.8	32.7	33.7
Turbidity (NTU)	1.22	2.46	4.95	9.98	20.1	40.5	81.5	164	331	666	1340	2700
Zn (ppb)	1.79	3.32	6.18	11.5	21.4	39.7	73.9	138	256	476	885	1650

*Nonlinear relationships with SI have breaks adapted from Pesce and Wunderlin (2000), Kannel, Lee, Lee, et al. (2007) and Koçer and Sevgili (2014)

Appendix D. The relative weights and descriptions of each parameters

Parameters	Weights (W_i)	EPA Description	EPA Guidelines	WHO drinking-water guidelines
DO	4	Essential for ecological functioning.	SWR recommends at least 30% oxygen saturation.	(only applicable for raw water)
Specific conductance	1	Indicates the amount of dissolved ions in the water, and could indicate alkalinity or water hardness.	SWR recommends 1000µS/cm.	(only applicable for raw water)
Temperature	1	Affects the air-water partitioning of compounds (i.e. DO, NH4) as compound's solubility is dependent on temperature.	SWR recommends around 25^{o}C	-
pH	1	Varies from 4.5 (in acid peaty land) to 10 (intense photosynthetic activity by algae). pH influences health of fisheries,	Optimal range of pH lies between 6.5-8.5.	(only applicable for operational monitoring)
Turbidity	4	Affects consumer acceptability, and water treatment and disinfection process as particles can trap pathogens.	-	(only applicable for raw water and operational monitoring)
Ca	1	An essential nutrient, but high levels can result in hard water.	-	-
K	1	An essential nutrient and an important component of artificial fertiliser. An essential parameter when investigating nutrient input into surface waters.	-	-
Mg	1	An essential nutrient, but high levels result in hard water.	-	-
Na	1	An essential nutrient, but excessive consumption may result in hypertension.	DWD indicates a limit of 200mg/l.	-
Cr	5	Both chromium species (tri and hexavalent) are toxic, and could be carcinogenic at high levels. It is listed under the DSD.	SWR indicates a limit of 0.05mg/l.	0.05 mg/l
Mn	3	Common constituent of soil and groundwater. Does not pose any toxicity issues, although it may affect palatability of water to consumers.	SWR and DWD limit the levels at 1.0mg/l and 50µg/l respectively.	0.4 mg/l
Co	2	Low level of significance due to its rare natural occurrence in surface	-	-

		waters, but its compounds are hazardous as solids or in concentrated levels. It is listed under DSD.		
Ni	5	Potential carcinogen in humans, and varying levels of harmful impacts on aquatic life, especially to fishes in high levels. Listed under DSD.	DWD limits the level at 20 µg/l.	0.07 mg/l
Cu	3	Serves as an algal toxicant, but may result in fish kills in high concentration.	FFD and SWR limit the levels at 0.04mg/l and 1mg/l respectively.	2.0 mg/l
Zn	3	An essential nutrient, but ingestion in high levels has emetic effect. Its toxicity is more significant for aquatic life.	FFD and SWR limit the levels at 1.0 mg/l and 5.0 mg/l respectively.	-
As	5	High toxicity to humans, and some compounds are known carcinogens. Listed under DSD.	DWD and SWR limit the levels at 10 µg/l and 0.1 mg/l respectively.	0.01 mg/l
Fe	3	A naturally occurring element, but possess no toxicity to humans. However, toxicity is more significant for aquatic life.	SWR limits the level at 2.0 mg/l	-
Pb	5	Toxic heavy metal with potential for bioaccumulation in body tissues.	SWR and DWD limit the levels at 0.05 mg/l and 10 µg/l respectively.	0.01 mg/l
Se	5	It has natural and anthropogenic sources. An essential nutrient but possesses toxicity in small concentrations that can result in various illnesses in humans. Listed under DSD.	SWR and DWD limit the levels at 0.01 mg/l and 10 µg/l respectively.	0.01 mg/l
Ag	3	Particularly toxic to micro-organisms, although it does not possess toxicity to humans. Naturally found in very low concentrations. Listed under DSD.	-	-
Cd	5	Sources in water likely originates from industrial processes. Possesses high toxicity for humans and aquatic life with potential for bioaccumulation. Listed under DSD.	SWR and DWD limit the levels at 0.005 mg/l and 5.0 µg/l respectively.	0.003 mg/l
Al	2	Commonly used in water treatment process, and does not pose as a significant health hazard.	DWD limits the value at 200 µg/l.	0.1 – 0.2 mg/l

Ba	4	Naturally occurring element, and naturally found in low concentration in surface waters. However, high concentrations can cause muscular, cardiovascular and renal damage. Listed under DSD.	SWR limits the level at 1.0mg/l.	0.7 mg/l
COD	3	Oxygen demand from the decomposition of organic matter which affects DO.	SWR limits the level at 40 mg/l.	(only applicable for raw water and operational monitoring)
E. Coli	3	Indication of sewage pollution (has human and/or animal origin). Poses infection risk to consumers from drinking contaminated waters.	SWR and DWD limit the levels at 40,000 no/100ml and 0 no/100ml respectively.	0 no/100ml
Total coliform	3	Indication of faecal and non-faecal bacteria.	SWR and DWD limit the levels at 100,000 no/100ml and 0 no/100ml respectively.	0 no/100ml
F^-	3	Provides dental benefits at low concentrations, but high concentrations can result in dental and skeletal fluorosis.	SWR and DWD limit the levels at 1.7 mg/l and 1.5 mg/l respectively.	1.5 mg/l
Cl^-	1	Does not possess health hazards, although higher concentrations affect drinking water palatability. High concentration in freshwater may indicate sewage pollution.	SWR and DWD limit levels at 250 mg/l.	-
SO_4^{2-}	2	Not a significant health hazard, but high concentration in combination with Na or Mg may enhance laxative effect; Reduction process to sulfide causes odours.	SWR and DWD limit levels at 200 mg/l and 250 mg/l respectively.	-
Br^-	1	Not a health hazard and may be nutritionally beneficial. However, bromate may be formed from the oxidation of bromide when ozone is used.	-	-
NO_3^-	2	High concentration indicates fertiliser or sewage pollution, and may cause eutrophication. It may be reduced to nitrite which can result in health hazard e.g. "blue baby syndrome".	SWR and DWD limit the levels at 50 mg/l.	50 mg/l (short-term exposure)

PO_4^{3-}	2	Widely used as a constituent in agricultural fertiliser and detergents. High concentrations may result in eutrophication. Listed under DSD.	SWR limits the level at 0.7 mg/l.	-
NO_2-N	2	High concentration can result in health hazard e.g. "blue baby syndrome", and could indicate recent sewage pollution. Listed under DSD.	FFD and DWD limit the levels at 0.03 mg/l and 0.5 mg/l.	3.0 mg/l (short-term exposure); 0.4 mg/l (long-term exposure)
NH_3-N	3	High concentrations may indicate sewage pollution and the presence of potential pathogenic micro-organisms. Listed under DSD.	SWR, FFD and DWD limit the levels at 4.0 mg/l, 0.2 mg/l and 0.5 mg/l respectively.	-

[a]Values of weight are modified from Pesce & Wunderlin (2000); Kannel et al. (2007); Koçer & Sevgili (2014)

Notes: (-): No recommended limit due to rare occurrence, or occurs at a level well below which possesses toxicity, but may affect acceptability of drinking-water;
SWD: Surface Water Regulations; FFD: Freshwater Fish Directive; DWD: Drinking Water Directive; DSD: Dangerous Substance Directive.

Appendix E. Water quality classes of observations

Classes	WQI Non-weight	Weight	Non-weight adj	Weight adj
(a) WQI_{AVG}				
Class 5	1 (0.27%)	0 (0.00%)	2 (0.47%)	0 (0.00%)
Class 4	1 (0.27%)	0 (0.00%)	1 (0.24%)	0 (0.00%)
Class 3	0 (0.00%)	1 (0.27%)	0 (0.00%)	1 (0.24%)
Class 2	372 (99.5%)	373 (99.7%)	423 (99.3%)	425 (99.8%)
Class 1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
(b) WQI_{MLR}				
Class 5	1 (0.26%)	0 (0.00%)	2 (0.44%)	0 (0.00%)
Class 4	1 (0.26%)	0 (0.00%)	1 (0.22%)	0 (0.00%)
Class 3	2 (0.53%)	1 (0.26%)	3 (0.67%)	2 (0.44%)
Class 2	377 (99.0%)	380 (99.7%)	444 (98.7%)	448 (99.6%)
Class 1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
(c) WQI_{PCA}				
Class 5	7 (1.82%)	7 (1.82%)	10 (2.22%)	11 (2.44%)
Class 4	2 (0.52%)	3 (0.78%)	3 (0.67%)	3 (0.67%)
Class 3	4 (1.04%)	5 (1.30%)	4 (0.89%)	7 (1.56%)
Class 2	371 (96.6%)	369 (96.1%)	433 (96.2%)	429 (95.3%)
Class 1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

Notes: Number of sites for non-adjusted and adjusted WQI is different as WQI cannot be computed where missing values are present for non-adjusted WQI

Appendix F. Factor Analysis (FA) of key parameters selected by Multivariate linear regression (MLR) and Principle Components Analysis (PCA) under different scenarios

	Key parameters grouped by Factor Analysis							**Key Parameters**
(a) FA_{MLR}								
Scenario	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Factor 7**	**Combined parameters**
Weighted	Fe (0.578), NH3-N (0.838), NO2-N (0.636)	Coliform (0.895), E.coli (0.763)	COD (0.591), Cu (-0.593), Zn (0.610)	Ca (0.750), Mg (0.694)	pH (-0.806)	NO3 (0.712)	Cr (0.816), Se (0.915)	Fe, NH3-N, NO2-N, Coliform, E.coli, COD, Cu, Zn, Cr, Se, pH, NO3,Ca, Mg
Non-weighted	Cr (0.588), Fe (0.526), NH3-N (0.792), NO2-N (0.744),	Coliform (0.912), E.coli (0.775)	COD (0.592), Cu (-0.539), Zn (0.652)	Ca (0.664), Cl (0.963), Mg (0.865), Na (0.750)	Co (0.631), pH (-0.812)	NO3 (0.742)	n.a.	Ca, Cl, Mg, Na, Coliform, E.coli, COD, Cu, Zn, Co, pH,NO3,Cr, Fe, NH3-N, NO2-N, PO4

PO4 (0.517)

(b) FA$_{PCA}$

Scenario	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Factor 7**	**Factor 8**	**Combined parameters**
Year	NH3-N (0.594), NO2-N (0.554)	Coliform (0.945), E.coli (0.737)	Ba (-0.581), Cu (0.635), COD (-0.520)	Na (0.751), Ca (0.605), F (0.506), Mg (0.942), Cl (0.806), K (0.593)	As (0.955)	Al (0.777), Fe (0.677)	Cr (0.651)	Co (0.949), Ni (0.795), Cd (0.771)	Na, Ca, F, Mg, Cl, K, Co, Ni, Cd, As, Ba, Cu, COD, Coliform, E.coli, Cr, Al, Fe, NH3-N, NO2-N
Dry	NH3-N (0.723)	Coliform (0.978), E.coli (0.718)	Ba (-0.573), Cu (0.664)	Mg (0.960), K (0.605), Ca (0.621), Cl (0.804),	As (0.942)	Al (0.827), Fe (0.638)	Cr (0.862)	Cd (0.706), Co (0.924)	Mg, K, Ca, Cl, Na, Cd, Co, As, Ba, Cu, Coliform, E.coli, Cr, Al, Fe, NH3-N

				Na (0.751)					
Wet	NH3-N (0.527), F (-0.53), NO2-N (0.532)	Coliform (0.965), E.coli (0.738)	COD (-0.514), Cu (0.743)	Cl (0.850), F (0.560), Na (0.804), Mg (0.889), Ca (0.617), K (0.575)	As (-0.587), NO3 (0.659)	Fe (0.765), Al (0.622)	Cr (0.715)	Cd (0.789), Ni (0.782), Co (0.956)	Cl, F, Na, Mg, Ca, K, Cd, Ni, Co, As, NO3, COD, Cu, Coliform, E.coli, Cr, NH3-N, F, NO2-N, Fe, Al

Notes: Underlined: Common parameters selected by both methods; Values in bracket (): Loadings from factor analysis (Only loadings with absolute value more than 0.6 are shown).

Appendix G. Parameter sensitivity across sites

River	Maximum absolute WQI_{CHANGE}			
	Weighted	**Non-weighted**	**Weighted adjusted**	**Non-weighted adjusted**
(a) WQI_{AVG}				
Belitong	Ba (2.97)	Ba (1.13)	Ba (2.97)	Ba (1.17)
Chemangar	As (2.91)	NO2 (0.97)	As (3.38)	NO2 (1.04)
Johor	Ba (3.68)	Ba (1.24)	Ba (3.81)	Ba (1.20)
Layau	Ba (2.66)	Ba (0.96)	Ba (2.45)	Ba (0.98)
Lebak	Ba (1.63)	Ba (1.10)	Ba (1.85)	Ba (1.14)
Lebam	Ba (3.08)	DO (0.84)	Ba (2.83)	DO (0.90)
Panti	Ni (2.26)	DO (1.23)	Ba (2.58)	DO (1.12)
Papan	Ba (4.13)	NO3 (0.92)	Ba (4.32)	Ba (0.91)
Pelepah	**As (4.50)**	**Mn (1.53)**	**As (4.47)**	**Mn (1.33)**
Pengeli	Ba (2.92)	Ba (1.25)	Ba (2.87)	Ba (1.30)
Sayong	Ba (2.41)	Ba (1.34)	Ba (2.42)	Ba (1.23)
Seleyut	Ba (2.01)	COD (0.88)	Ba (2.01)	COD (0.88)
Semangar	Ba (2.44)	Ba (1.11)	Ba (3.08)	Ba (1.08)
Temenin	Ba (3.49)	Coliform (total) (1.11)	Ba (3.95)	Fluoride (1.19)
Tengkil	As (3.48)	Ba (0.97)	Ba (3.47)	Fluoride (1.07)
Tiram	Ba (2.32)	Coliform (total) (1.17)	Ba (2.39)	Coliform (total) (0.94)
(b) WQI_{MLR}				
Belitong	Coliform (total) (3.14)	NH3 (2.43)	COD (3.39)	NH3 (2.69)
Chemangar	COD (7.24)	**pH (3.11)**	COD (7.67)	pH (3.02)
Johor	Coliform (total) (4.28)	Mg (1.97)	Coliform (total) (4.33)	Mg (2.08)
Layau	Fe (4.65)	Mg (2.00)	Cu (4.07)	COD (2.12)
Lebak	Coliform (total) (5.95)	Mg (1.85)	Coliform (total) (5.95)	Mg (1.85)
Lebam	Coliform (total) (5.18)	NO3 (1.94)	COD (4.92)	Mg (1.88)
Panti	Coliform (total) (5.98)	Mg (2.04)	COD (5.65)	Mg (2.52)
Papan	Cu (5.25)	NO3 (2.27)	Fe (5.46)	NO3 (2.15)
Pelepah	Coliform (total) (6.38)	Mg (2.16)	Coliform (total) (6.07)	Mg (2.55)
Pengeli	Cu (5.15)	Cu (1.89)	Cu (4.97)	COD (2.00)
Sayong	COD (2.97)	COD (2.10)	COD (3.28)	COD (1.89)
Seleyut	Coliform (total) (5.64)	COD (2.24)	Coliform (total) (5.64)	COD (2.24)

Semangar	Coliform (total) (5.26)	COD (1.89)	Coliform (total) (5.01)	Mg (2.22)
Temenin	Coliform (total) (6.00)	Coliform (total) (2.73)	Coliform (total) (5.28)	Coliform (total) (2.19)
Tengkil	COD (6.16)	Mg (2.28)	COD (6.13)	Mg (2.62)
Tiram	**Coliform (total) (8.88)**	Mg (2.70)	**Coliform (total) (8.08)**	**Mg (3.09)**
(c) WQI_{PCA}				
Belitong	Al (6.64)	Na (2.48)	Al (6.43)	Na (2.53)
Chemangar	As (4.37)	Cd (1.77)	As (4.30)	Cd (1.85)
Johor	Al (5.35)	Na (1.92)	Al (5.26)	Na (1.97)
Layau	Al (5.64)	Cu (1.64)	Al (5.16)	Cu (1.70)
Lebak	As (6.96)	Na (1.85)	As (6.96)	Na (1.85)
Lebam	Ca (6.79)	Coliform (total) (1.62)	Ca (6.66)	Coliform (total) (1.44)
Panti	As (5.85)	Coliform (total) (1.60)	As (5.75)	Na (1.79)
Papan	Al (3.96)	Co (1.96)	Al (3.98)	Co (1.96)
Pelepah	As (7.84)	**As (3.04)**	As (7.12)	**As (2.77)**
Pengeli	Al (5.65)	Cu (1.82)	Al (5.36)	Cu (1.85)
Sayong	Al (5.85)	Na (1.77)	Al (5.61)	Na (2.04)
Seleyut	As (6.36)	Coliform (total) (1.97)	As (6.36)	Coliform (total) (1.97)
Semangar	Ca (6.68)	Coliform (total) (1.94)	Ca (6.31)	Coliform (total) (1.61)
Temenin	Al (7.65)	Coliform (total) (2.83)	Ca (6.67)	Coliform (total) (2.27)
Tengkil	As (5.92)	As (2.54)	As (5.48)	As (2.46)
Tiram	**As (9.49)**	Coliform (total) (2.89)	**Ca (8.96)**	Coliform (total) (2.31)

Notes: Only most sensitive parameters for each site are shown. Values in brackets (): Absolute change in WQI upon removal of the corresponding parameter. **Bold:** Parameter removal that produces the largest WQI_{CHANGE}

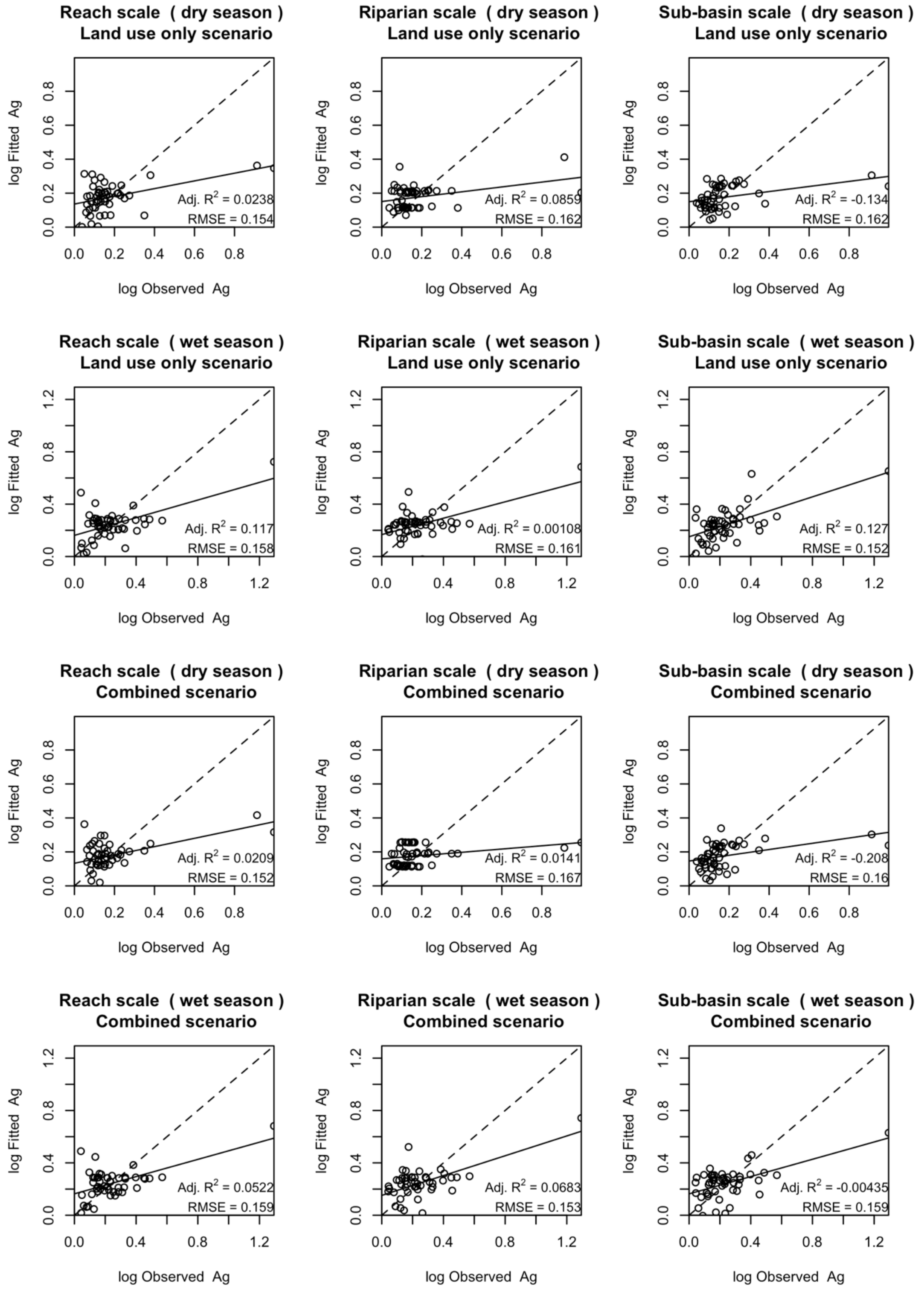
Appendix H. Adjusted R^2 values for all MLR models

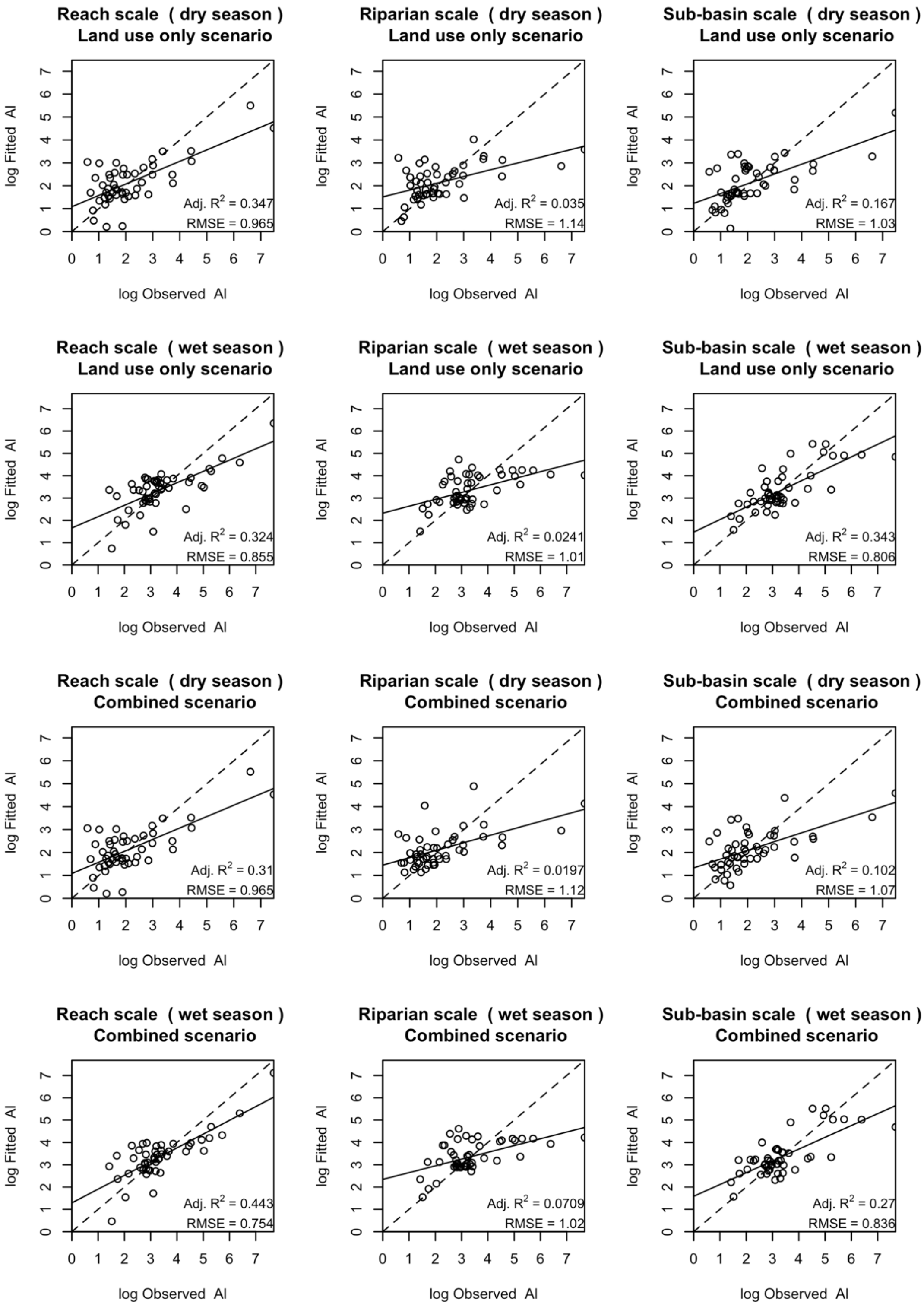
	Dry			Wet		
Parameter	**Reach**	**Riparian**	**Sub-basin**	**Reach**	**Riparian**	**Sub-basin**
(a) Land use only scenario						
Ag	-	-	-	-	-	-
Al	**0.347**	-	-	0.324	-	**0.343**
As	0.448	0.346	**0.515**	**0.666**	0.405	0.506
Ba	0.313	0.419	**0.432**	0.390	0.373	**0.539**
Br	**0.386**	-	-	**0.420**	0.232	-
Ca	0.724	**0.752**	0.646	**0.692**	0.556	0.472
Cd	-	-	-	-	-	**0.367**
Cl	0.558	0.546	**0.592**	**0.239**	0.196	0.140
Co	**0.328**	-	0.250	**0.411**	-	-
COD	0.436	0.548	**0.585**	0.431	**0.487**	0.401
Coliform (total)	0.363	0.651	**0.669**	0.266	**0.518**	0.479
Cr	**0.243**	-	-	**0.691**	0.485	-
Cu	0.287	**0.463**	0.324	0.283	**0.441**	-
DO	**0.432**	0.253	-	-	**0.440**	0.244
E.coli	0.366	**0.695**	0.666	-	**0.468**	0.339
Fe	0.281	**0.438**	0.349	0.348	**0.419**	0.340
Fluoride	-	**0.398**	0.326	-	-	-
K	0.318	0.462	**0.493**	0.380	**0.472**	0.394
Mg	0.330	**0.352**	-	**0.238**	-	-
Mn	**0.537**	-	0.369	**0.569**	0.212	0.264
Na	**0.800**	0.591	0.557	**0.409**	0.366	-
NH3	0.516	**0.747**	0.640	0.544	**0.688**	0.257
Ni	**0.332**	0.291	0.279	**0.606**	0.401	0.248
NO2	**0.818**	0.646	0.414	**0.729**	0.521	0.435
NO3	-	-	-	-	0.425	**0.598**
Pb	-	-	**0.224**	**0.332**	-	-
pH	0.398	-	**0.407**	0.324	0.347	**0.508**
PO4	-	**0.241**	-	**0.321**	-	-
Se	-	-	-	0.344	0.322	**0.405**
SEC	**0.734**	0.685	0.698	**0.505**	0.464	0.291
SO4	**0.726**	0.435	0.388	0.324	**0.331**	-
Temperature	0.369	0.280	**0.570**	0.282	0.483	**0.631**
Turbidity	-	**0.572**	0.320	-	**0.493**	0.436
Zn	0.280	-	**0.447**	0.249	**0.301**	0.298
(b) Combined scenario						
Ag	-	-	-	-	-	-

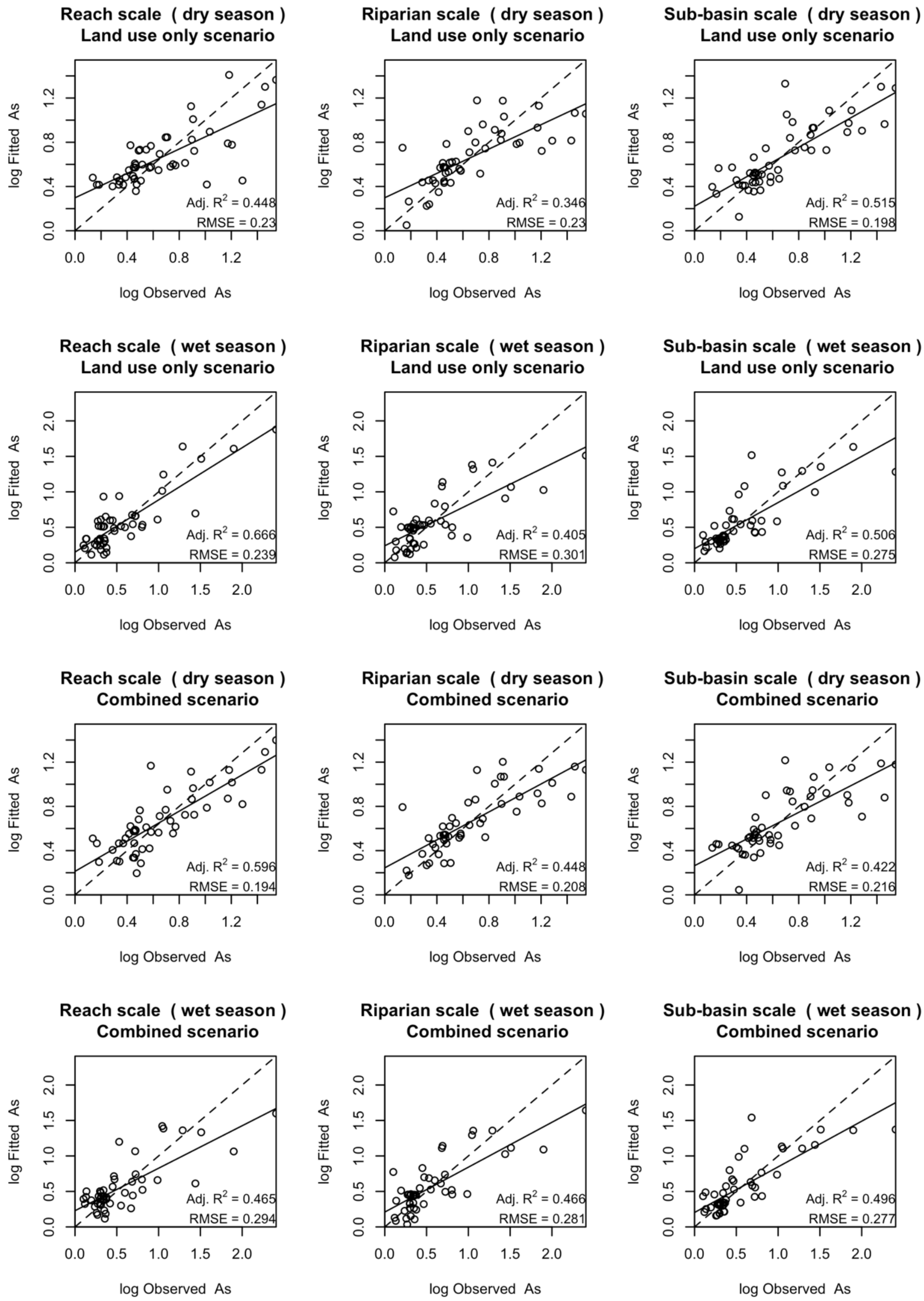
Al	**0.310**	-	-	**0.443**	-	0.270
As	**0.596**	0.448	0.422	0.465	0.466	**0.496**
Ba	0.396	**0.419**	0.410	0.547	-	**0.614**
Br	**0.259**	-	-	**0.449**	0.296	-
Ca	0.715	0.718	**0.760**	0.622	**0.670**	0.508
Cd	-	-	-	-	-	**0.283**
Cl	0.394	**0.555**	0.415	0.244	**0.383**	-
Co	**0.381**	-	0.282	**0.508**	-	-
COD	0.404	**0.637**	0.547	0.525	**0.527**	0.503
Coliform (total)	0.670	**0.699**	0.690	0.337	0.510	**0.535**
Cr	**0.243**	-	-	**0.656**	0.335	-
Cu	0.430	**0.477**	0.269	0.284	**0.403**	-
DO	**0.329**	-	-	**0.574**	0.417	0.303
E.coli	0.678	**0.717**	0.615	0.295	**0.432**	0.276
Fe	0.347	**0.470**	0.407	**0.386**	0.335	-
Fluoride	-	0.260	**0.373**	-	-	-
K	0.336	**0.481**	0.471	0.476	**0.509**	0.460
Mg	0.275	**0.280**	-	-	**0.270**	-
Mn	**0.537**	-	0.352	**0.512**	-	0.360
Na	**0.765**	0.711	0.632	0.365	**0.395**	-
NH3	0.706	**0.753**	0.683	0.520	**0.690**	0.461
Ni	**0.490**	0.247	0.344	**0.648**	0.254	-
NO2	**0.816**	0.599	0.385	**0.774**	0.558	0.479
NO3	-	-	**0.271**	0.264	0.414	**0.620**
Pb	-	-	-	-	-	-
pH	**0.396**	-	0.366	-	**0.410**	0.390
PO4	**0.331**	-	0.246	**0.409**	-	-
Se	-	-	-	0.251	0.296	**0.447**
SEC	0.595	**0.667**	0.583	0.470	**0.540**	0.297
SO4	0.450	0.423	**0.505**	**0.325**	0.262	-
Temperature	**0.545**	0.441	0.457	0.477	**0.604**	0.598
Turbidity	-	**0.639**	0.343	-	-	**0.414**
Zn	0.284	-	**0.388**	**0.229**	-	-

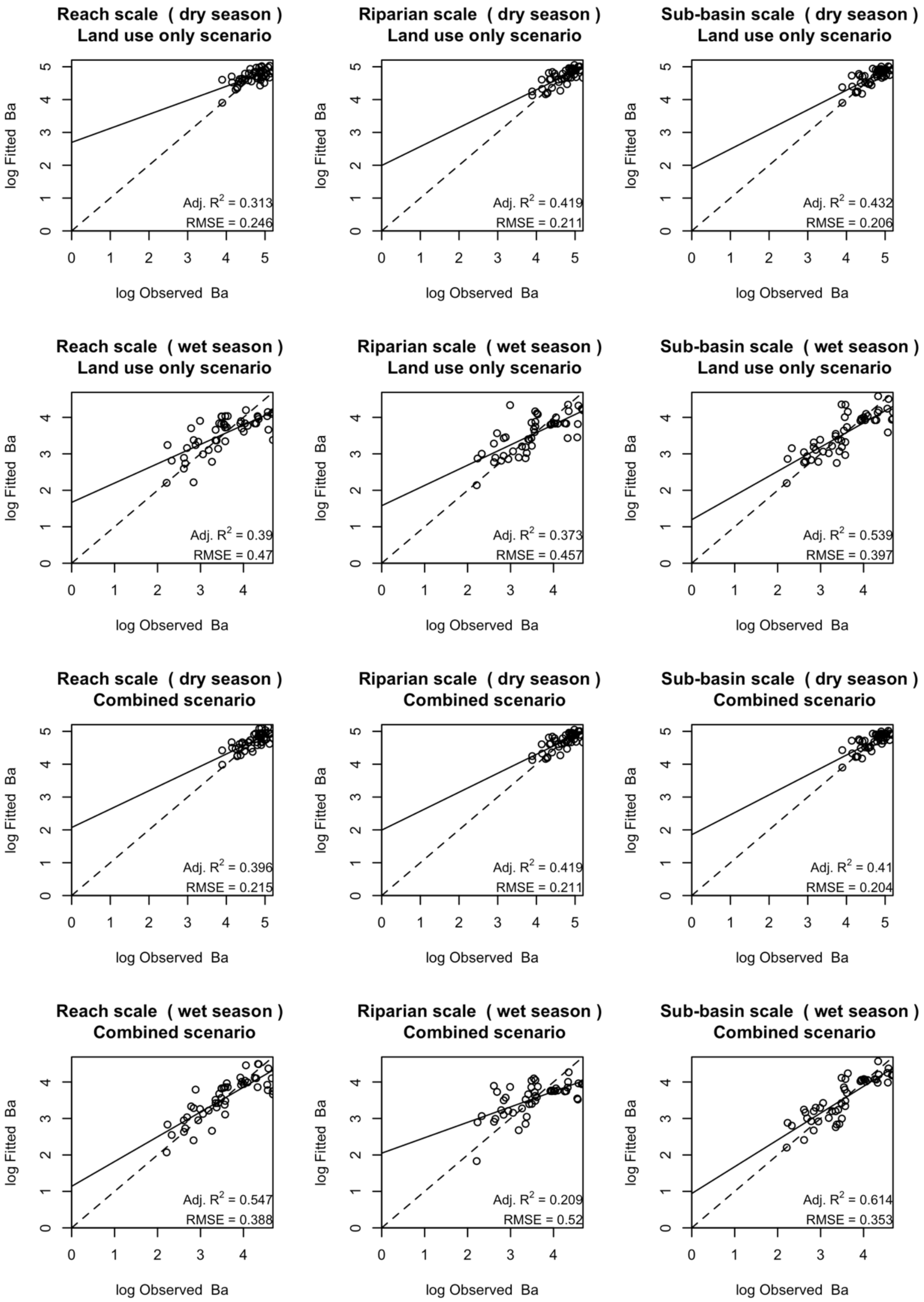
Notes: - No significant models (p-value >0.05); *Adj.* R^2 means the adjusted coefficient of multiple determination; **Bolded:** Model with the highest adjusted R^2 value

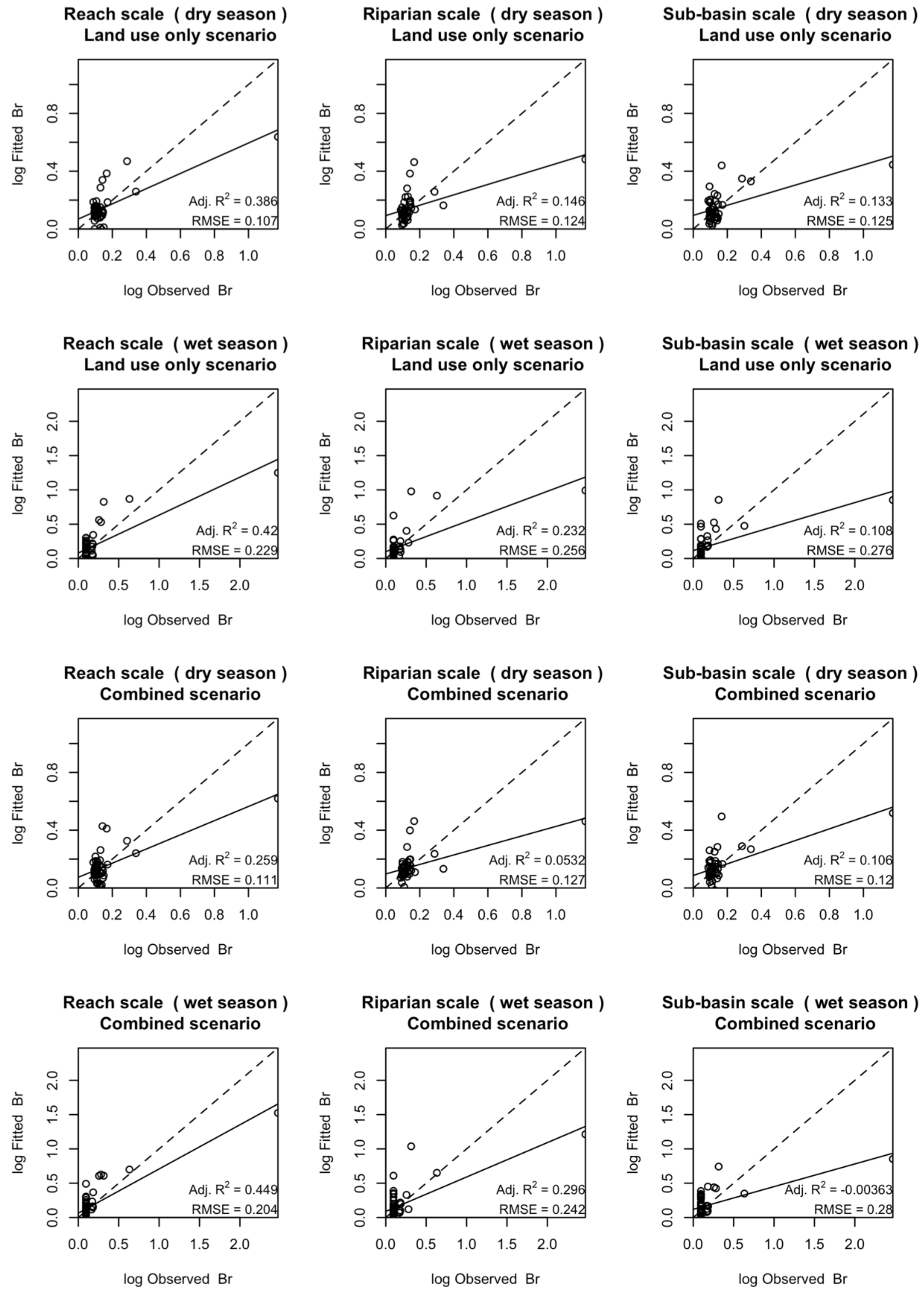
Appendix I. Scatterplots for all measured water quality parameters

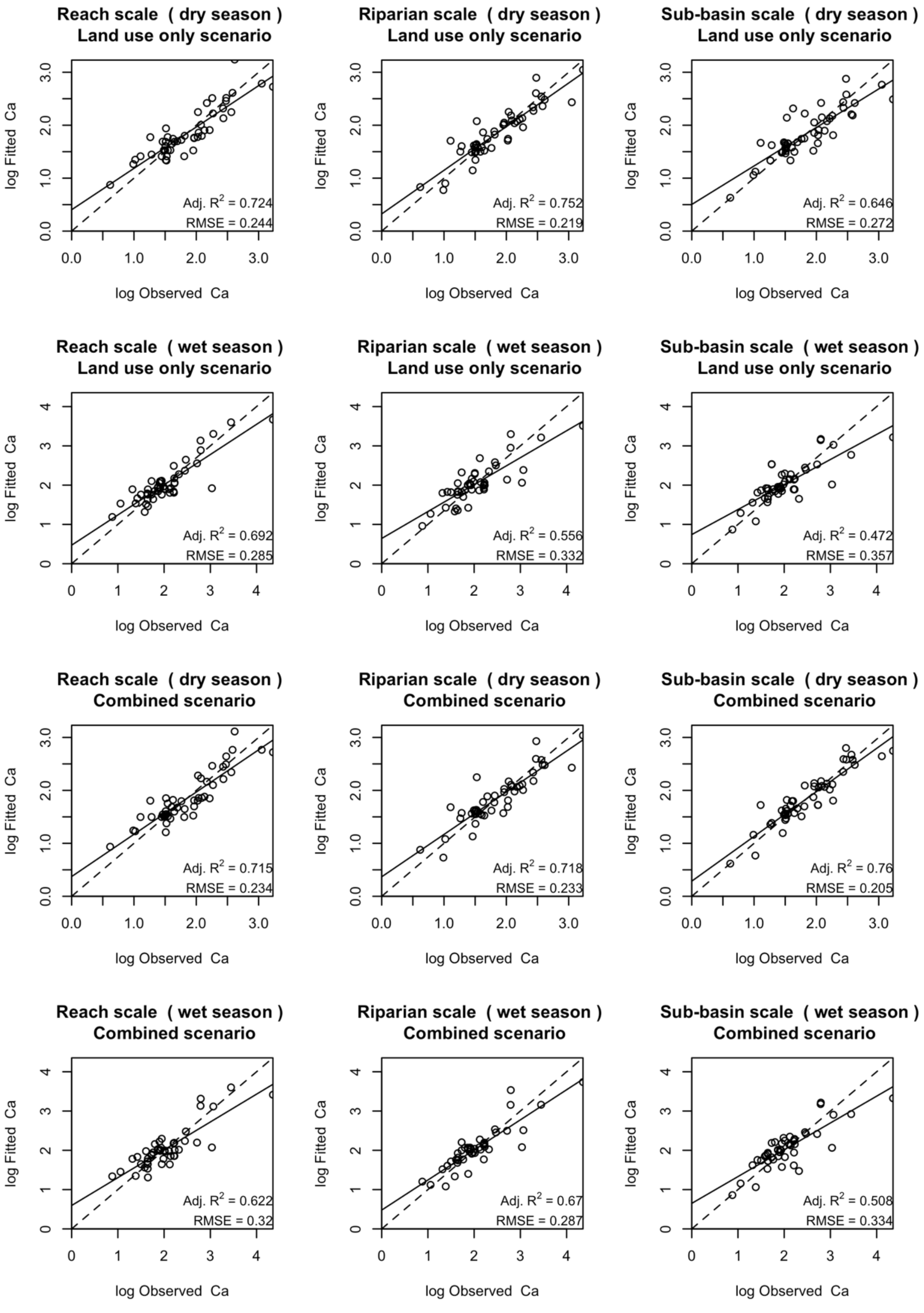


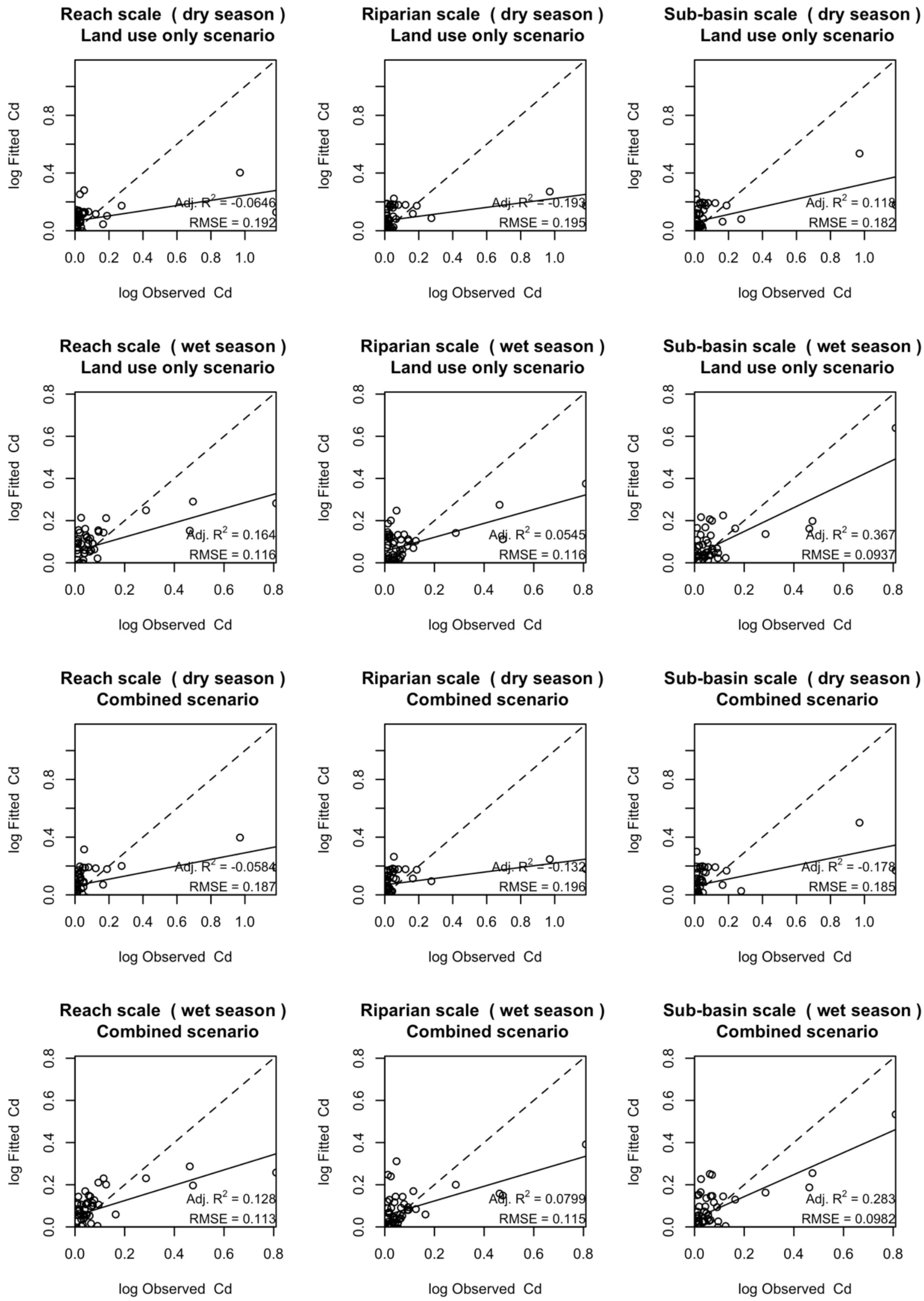

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.347
RMSE = 0.965
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.035
RMSE = 1.14
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.167
RMSE = 1.03
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.324
RMSE = 0.855
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.0241
RMSE = 1.01
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.343
RMSE = 0.806
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.31
RMSE = 0.965
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.0197
RMSE = 1.12
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.102
RMSE = 1.07
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.443
RMSE = 0.754
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.0709
RMSE = 1.02
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.27
RMSE = 0.836
log Fitted AI
log Observed AI

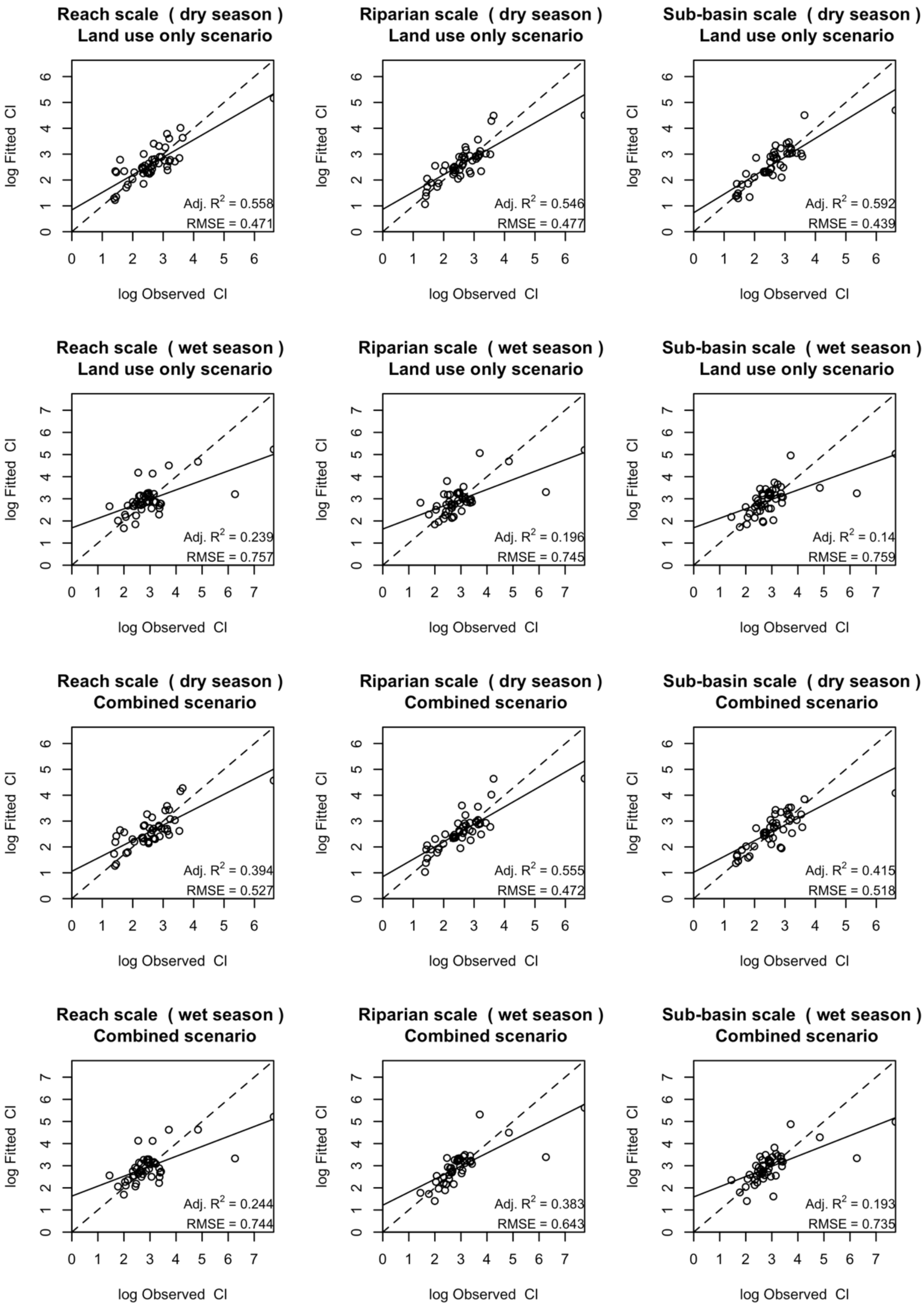

Reach scale (dry season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.448
RMSE = 0.23
Riparian scale (dry season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.346
RMSE = 0.23
Sub-basin scale (dry season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.515
RMSE = 0.198
Reach scale (wet season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.666
RMSE = 0.239
Riparian scale (wet season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.405
RMSE = 0.301
Sub-basin scale (wet season)
Land use only scenario
log Fitted As
log Observed As
Adj. R^2 = 0.506
RMSE = 0.275
Reach scale (dry season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.596
RMSE = 0.194
Riparian scale (dry season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.448
RMSE = 0.208
Sub-basin scale (dry season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.422
RMSE = 0.216
Reach scale (wet season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.465
RMSE = 0.294
Riparian scale (wet season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.466
RMSE = 0.281
Sub-basin scale (wet season)
Combined scenario
log Fitted As
log Observed As
Adj. R^2 = 0.496
RMSE = 0.277

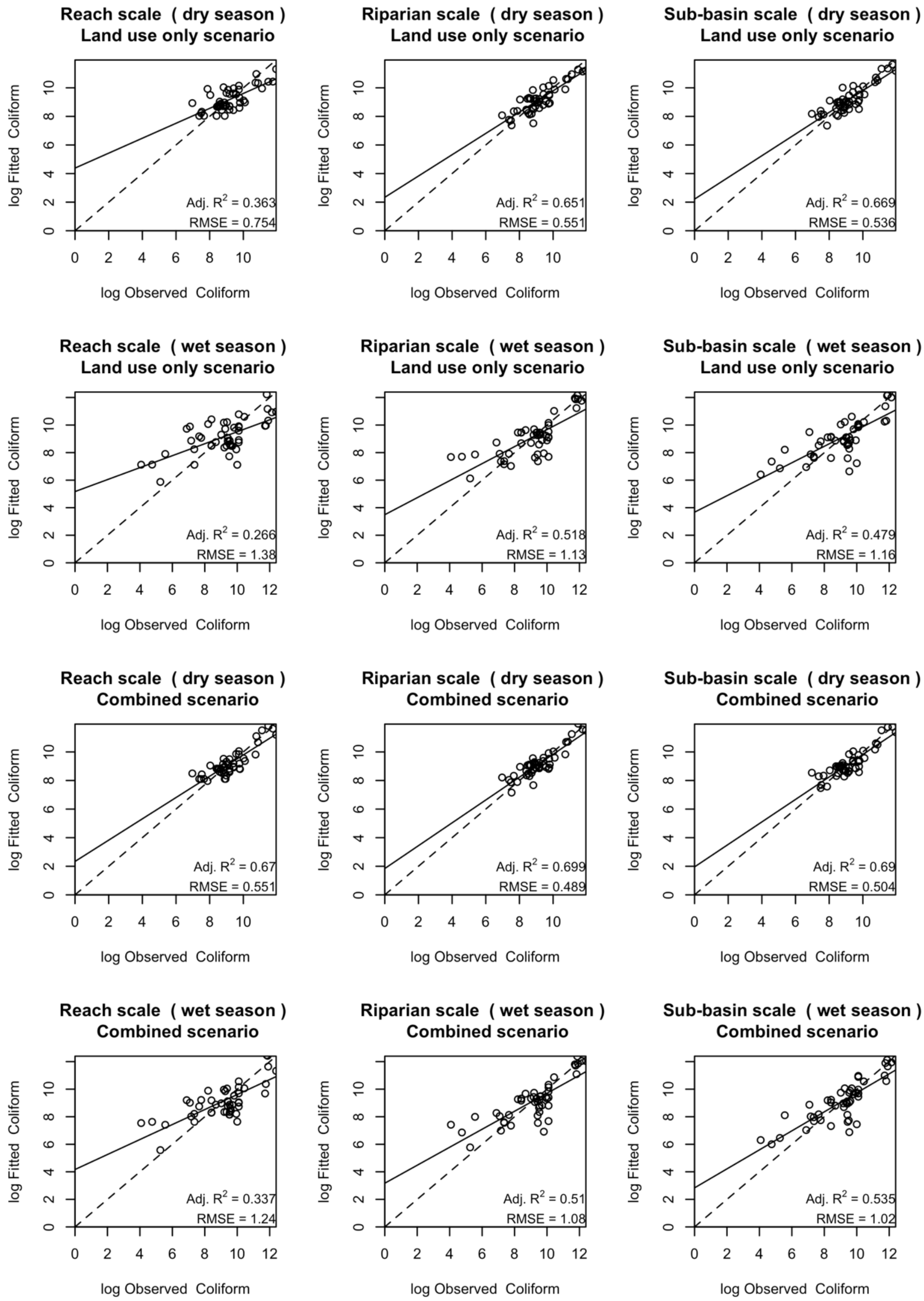

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.313
RMSE = 0.246
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.419
RMSE = 0.211
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.432
RMSE = 0.206
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.39
RMSE = 0.47
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.373
RMSE = 0.457
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.539
RMSE = 0.397
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.396
RMSE = 0.215
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.419
RMSE = 0.211
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.41
RMSE = 0.204
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.547
RMSE = 0.388
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.209
RMSE = 0.52
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.614
RMSE = 0.353
log Fitted Ba
log Observed Ba


Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.386
RMSE = 0.107
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.146
RMSE = 0.124
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.133
RMSE = 0.125
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.42
RMSE = 0.229
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.232
RMSE = 0.256
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.108
RMSE = 0.276
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.259
RMSE = 0.111
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.0532
RMSE = 0.127
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.106
RMSE = 0.12
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.449
RMSE = 0.204
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.296
RMSE = 0.242
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = -0.00363
RMSE = 0.28
log Fitted Br
log Observed Br

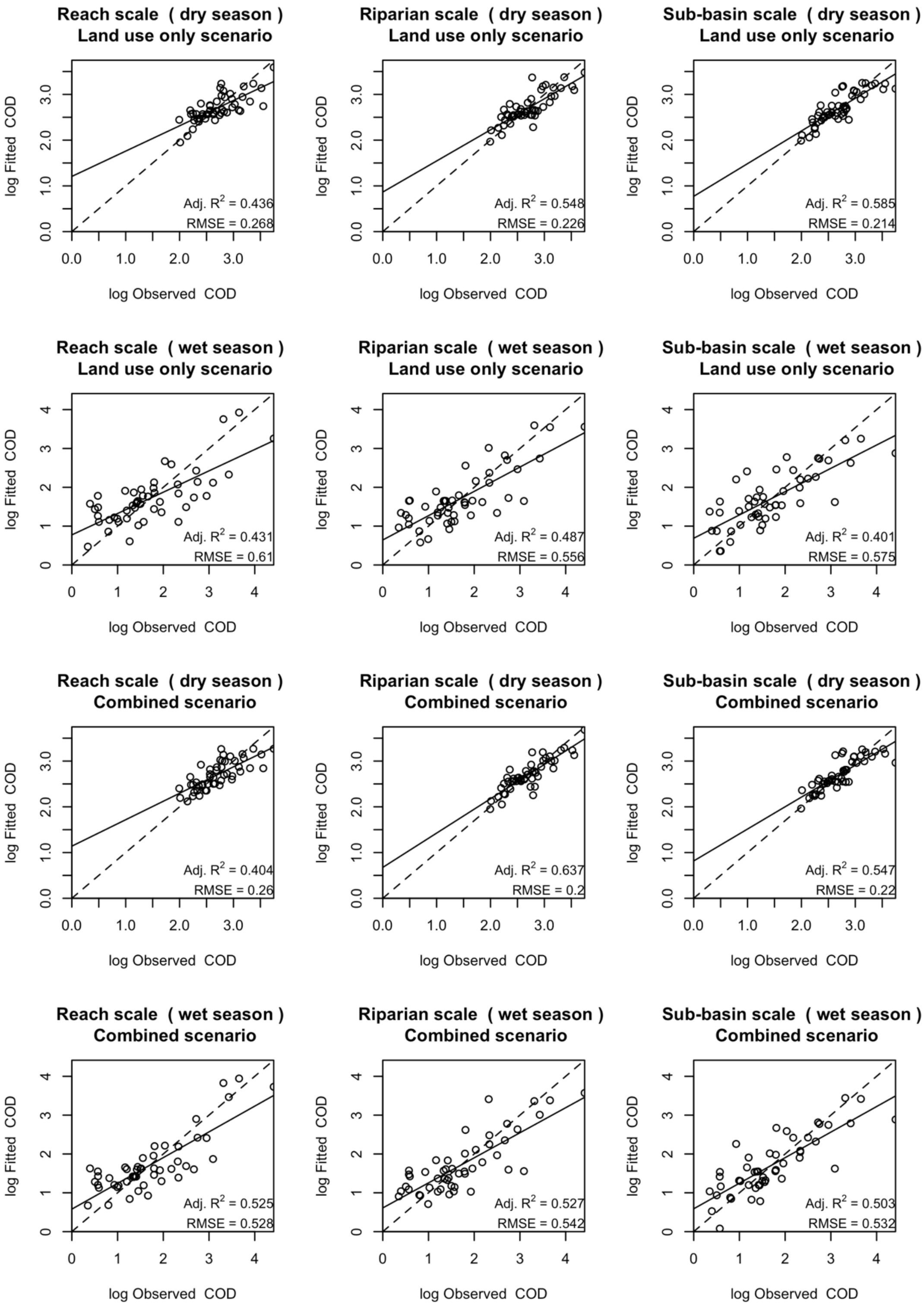

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.724
RMSE = 0.244
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.752
RMSE = 0.219
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.646
RMSE = 0.272
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.692
RMSE = 0.285
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.556
RMSE = 0.332
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.472
RMSE = 0.357
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.715
RMSE = 0.234
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.718
RMSE = 0.233
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.76
RMSE = 0.205
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.622
RMSE = 0.32
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.67
RMSE = 0.287
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.508
RMSE = 0.334
log Fitted Ca
log Observed Ca

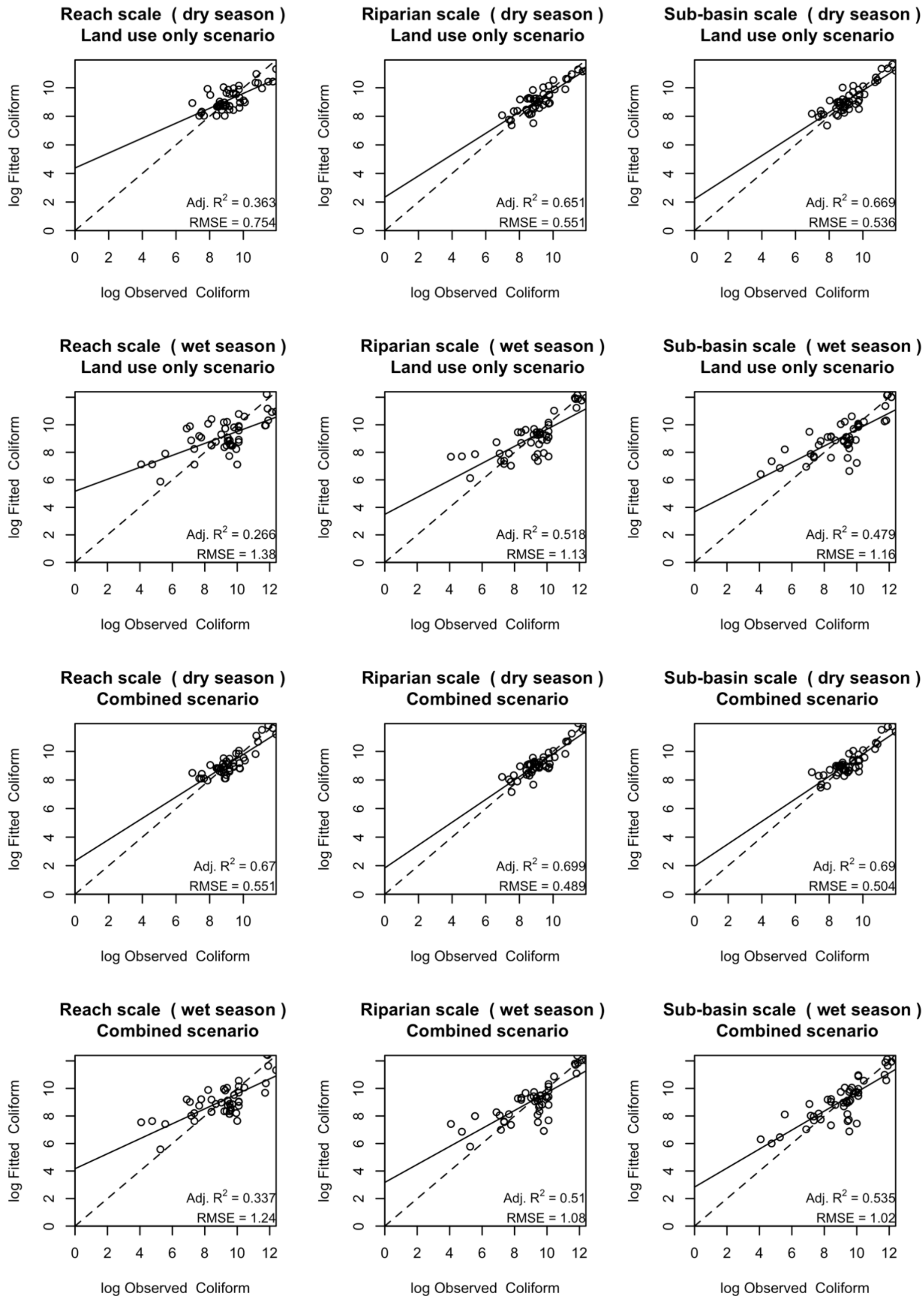

Reach scale (dry season)
Land use only scenario
Adj. R2 = -0.0646
RMSE = 0.192
Riparian scale (dry season)
Land use only scenario
Adj. R2 = -0.193
RMSE = 0.195
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.118
RMSE = 0.182
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.164
RMSE = 0.116
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.0545
RMSE = 0.116
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.367
RMSE = 0.0937
Reach scale (dry season)
Combined scenario
Adj. R2 = -0.0584
RMSE = 0.187
Riparian scale (dry season)
Combined scenario
Adj. R2 = -0.132
RMSE = 0.196
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = -0.178
RMSE = 0.185
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.128
RMSE = 0.113
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.0799
RMSE = 0.115
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.283
RMSE = 0.0982
log Fitted Cd
log Observed Cd

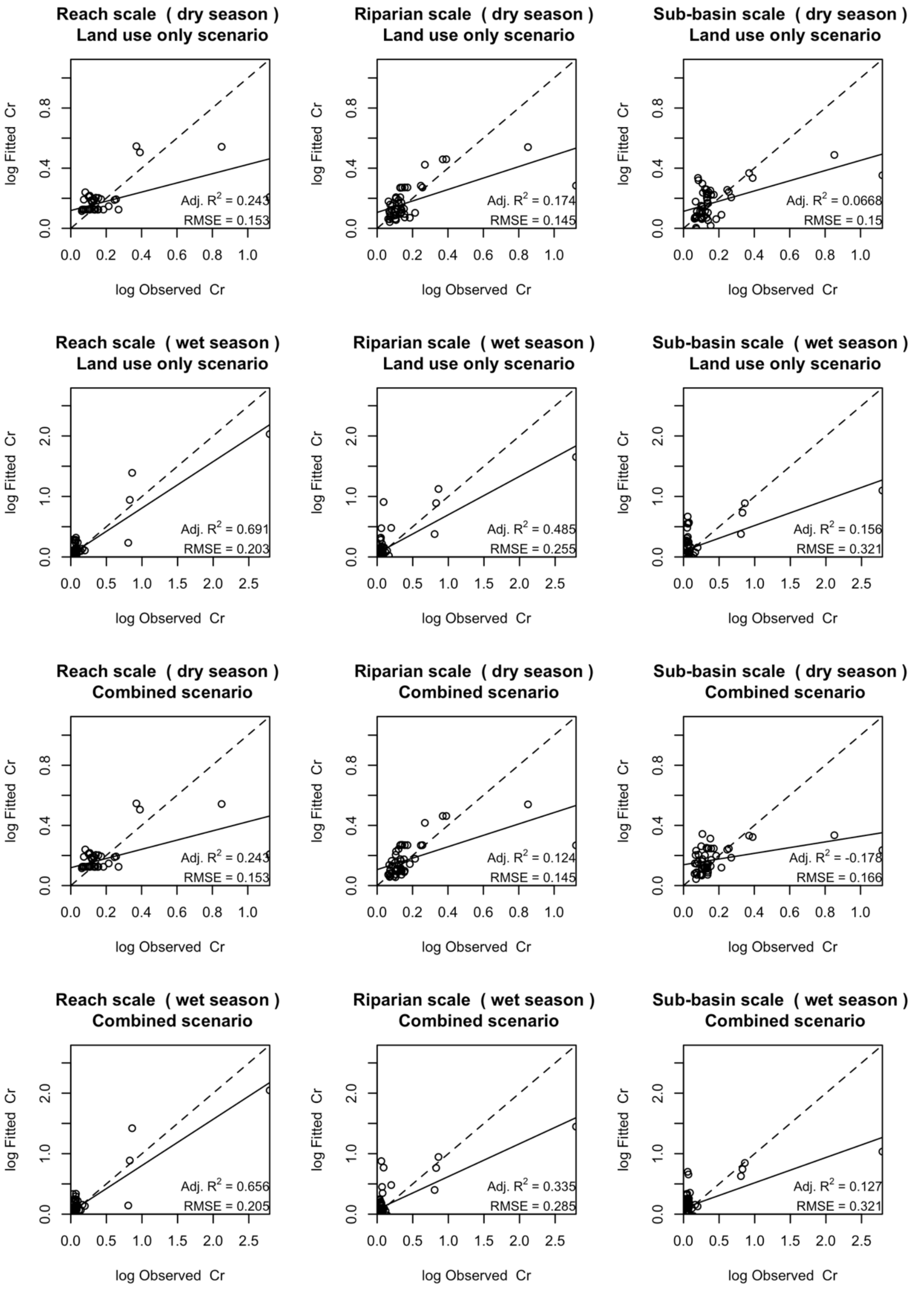

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.558
RMSE = 0.471
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.546
RMSE = 0.477
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.592
RMSE = 0.439
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.239
RMSE = 0.757
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.196
RMSE = 0.745
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.14
RMSE = 0.759
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.394
RMSE = 0.527
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.555
RMSE = 0.472
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.415
RMSE = 0.518
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.244
RMSE = 0.744
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.383
RMSE = 0.643
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.193
RMSE = 0.735
log Fitted Cl
log Observed Cl

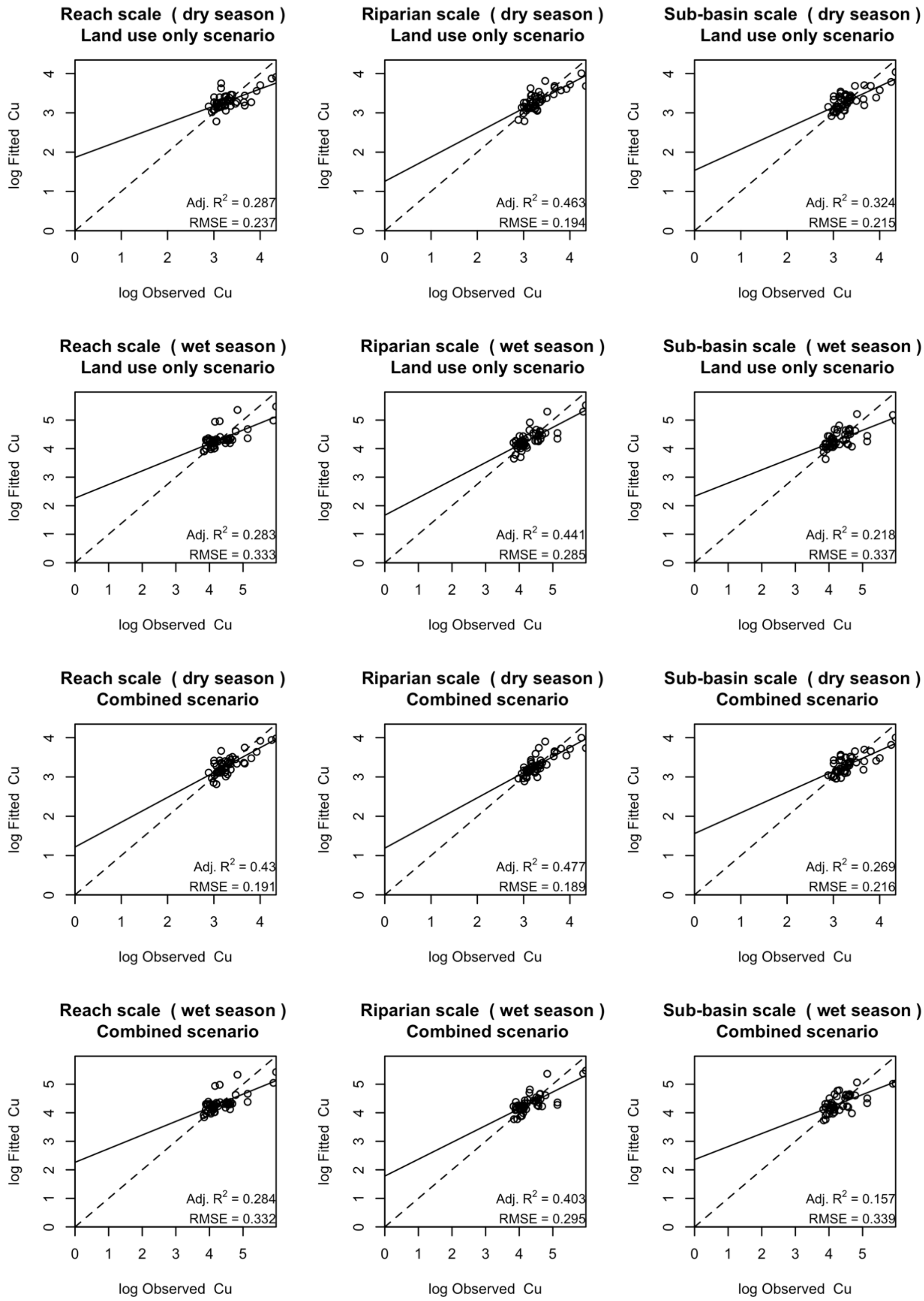


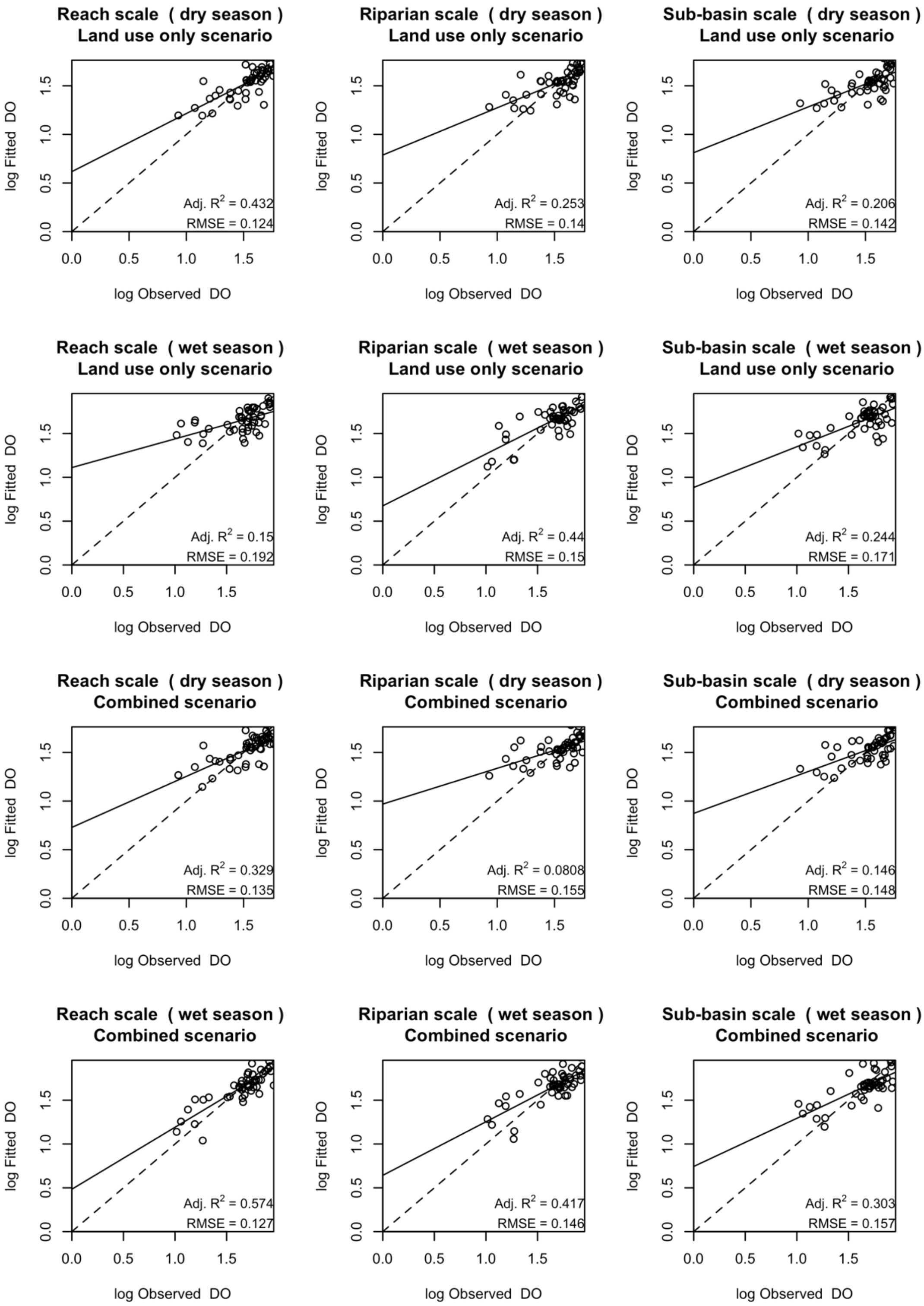
Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.363
RMSE = 0.754
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.651
RMSE = 0.551
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.669
RMSE = 0.536
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.266
RMSE = 1.38
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.518
RMSE = 1.13
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.479
RMSE = 1.16
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.67
RMSE = 0.551
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.699
RMSE = 0.489
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.69
RMSE = 0.504
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.337
RMSE = 1.24
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.51
RMSE = 1.08
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.535
RMSE = 1.02
log Fitted Coliform
log Observed Coliform

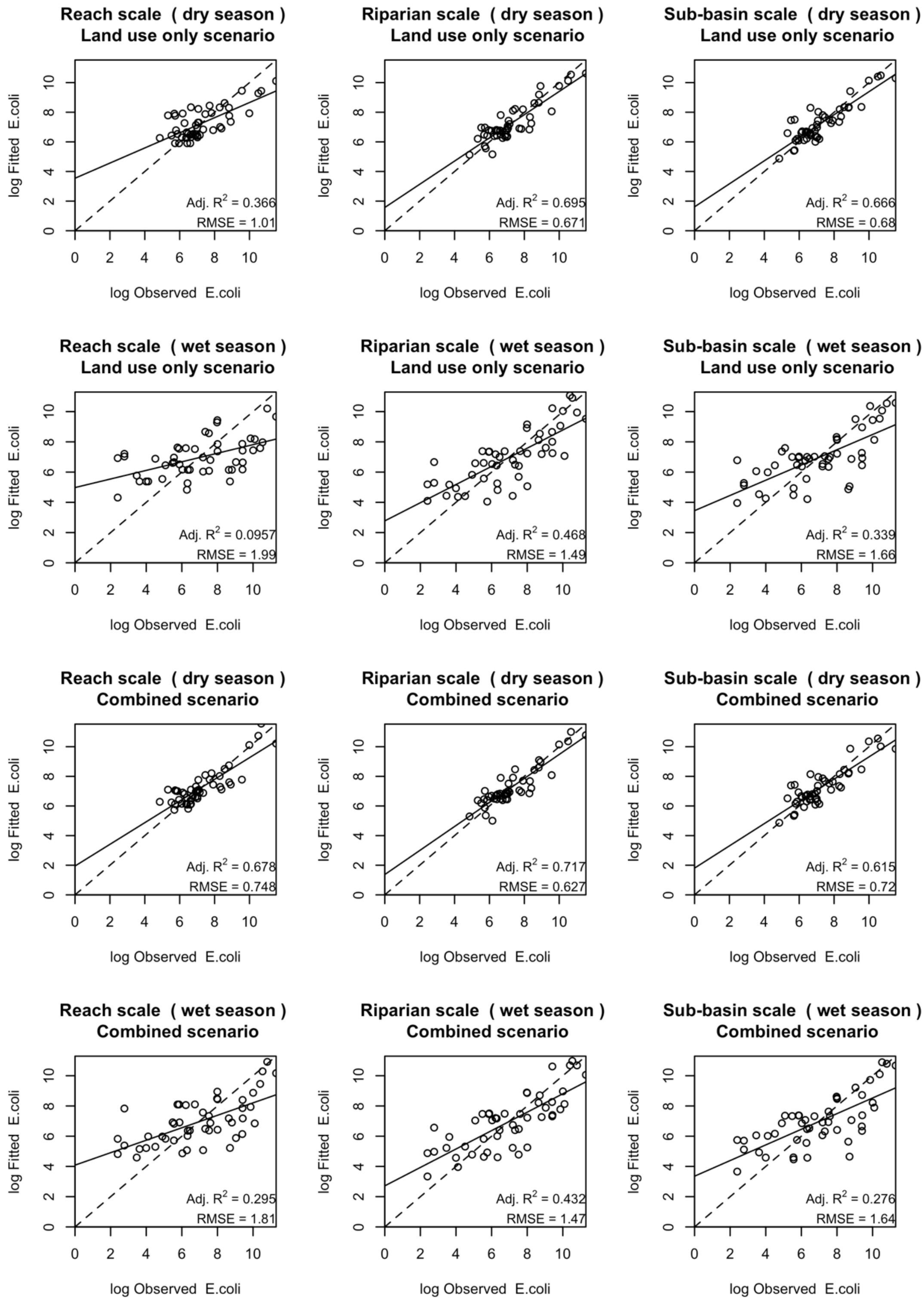

Reach scale (dry season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.436
RMSE = 0.268
Riparian scale (dry season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.548
RMSE = 0.226
Sub-basin scale (dry season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.585
RMSE = 0.214
Reach scale (wet season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.431
RMSE = 0.61
Riparian scale (wet season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.487
RMSE = 0.556
Sub-basin scale (wet season)
Land use only scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.401
RMSE = 0.575
Reach scale (dry season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.404
RMSE = 0.26
Riparian scale (dry season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.637
RMSE = 0.2
Sub-basin scale (dry season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.547
RMSE = 0.22
Reach scale (wet season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.525
RMSE = 0.528
Riparian scale (wet season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.527
RMSE = 0.542
Sub-basin scale (wet season)
Combined scenario
log Fitted COD
log Observed COD
Adj. R^2 = 0.503
RMSE = 0.532

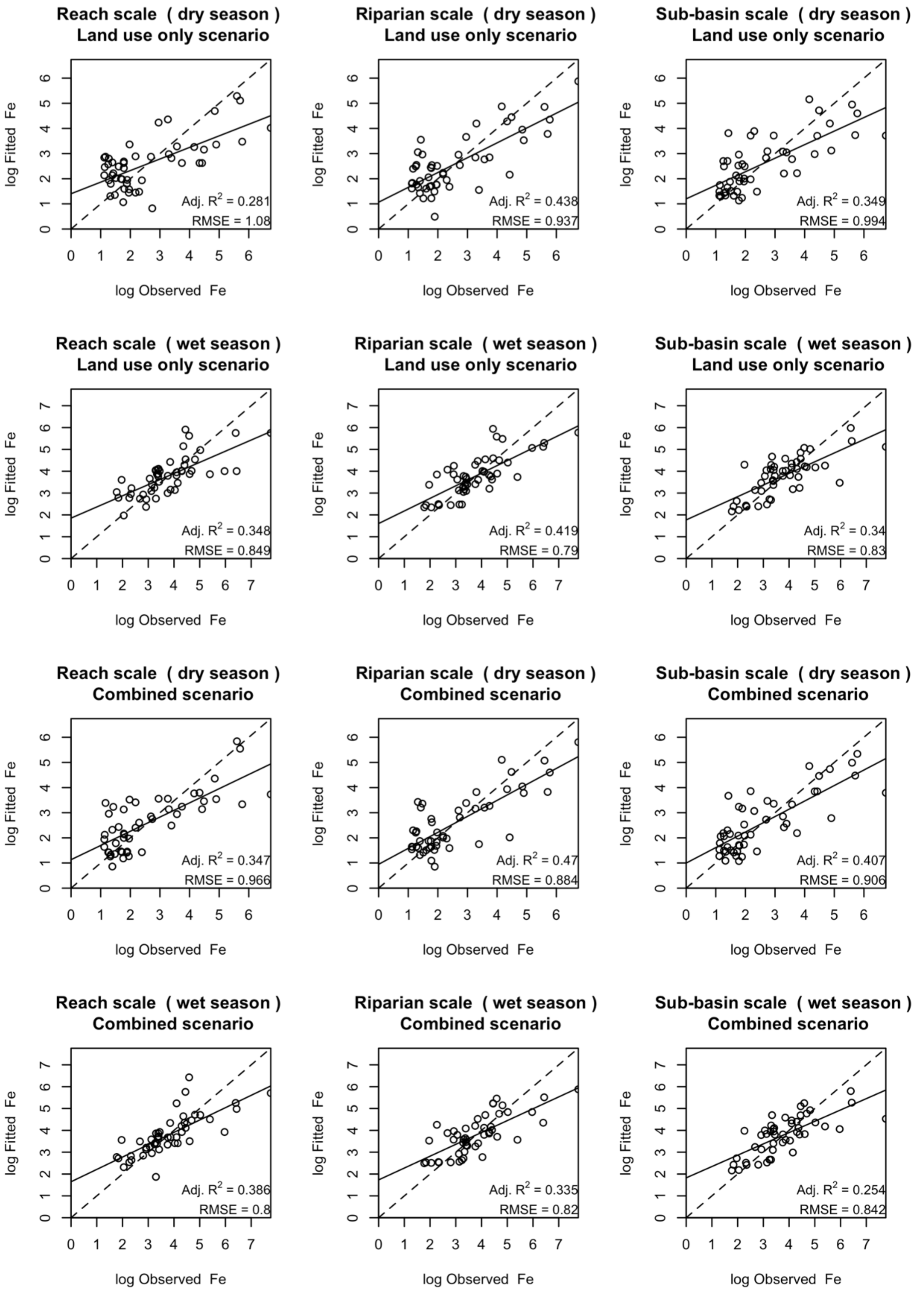

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.363
RMSE = 0.754
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.651
RMSE = 0.551
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.669
RMSE = 0.536
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.266
RMSE = 1.38
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.518
RMSE = 1.13
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.479
RMSE = 1.16
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.67
RMSE = 0.551
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.699
RMSE = 0.489
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.69
RMSE = 0.504
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.337
RMSE = 1.24
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.51
RMSE = 1.08
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.535
RMSE = 1.02
log Fitted Coliform
log Observed Coliform

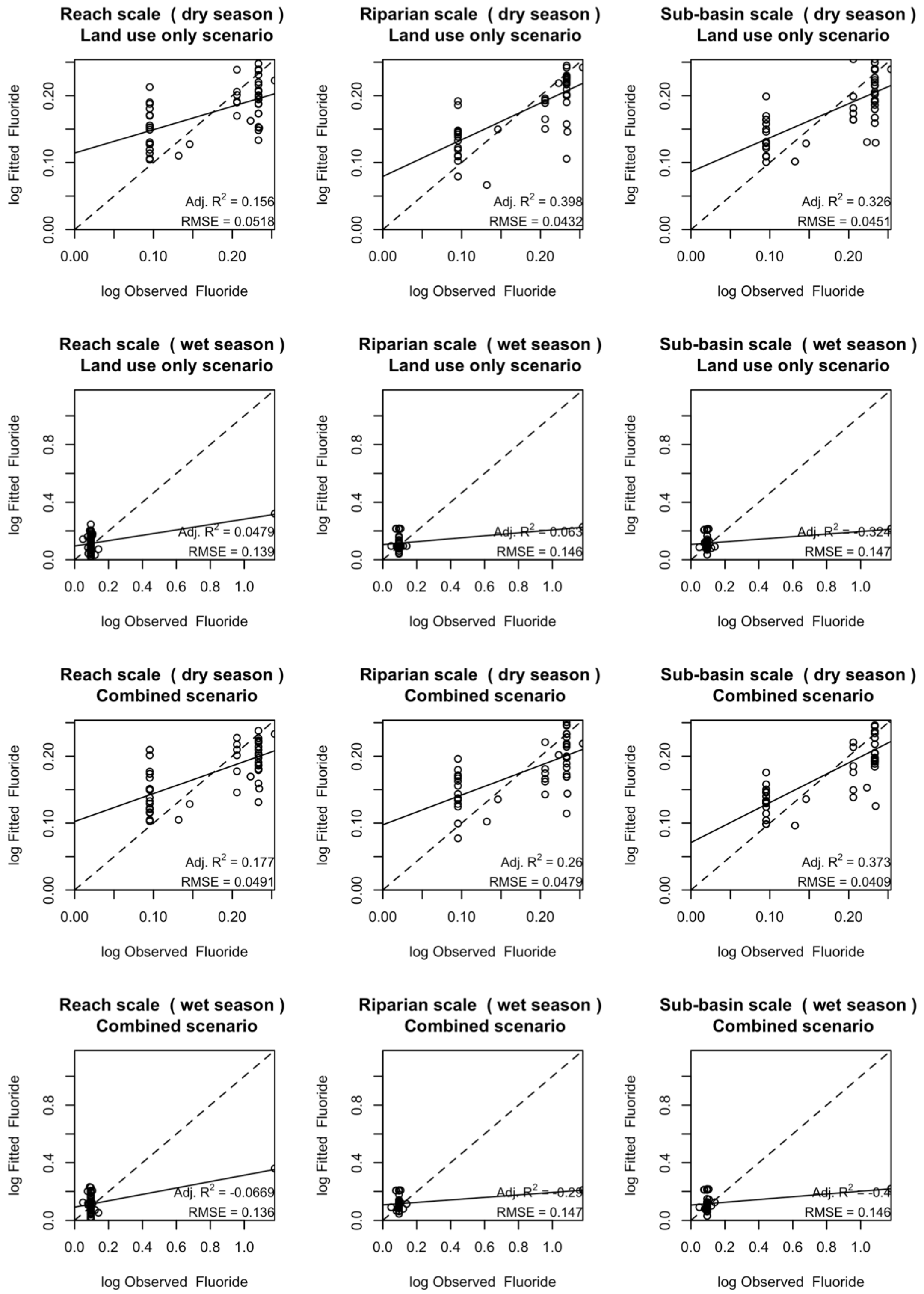

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.243
RMSE = 0.153
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.174
RMSE = 0.145
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.0668
RMSE = 0.15
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.691
RMSE = 0.203
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.485
RMSE = 0.255
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.156
RMSE = 0.321
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.243
RMSE = 0.153
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.124
RMSE = 0.145
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = -0.178
RMSE = 0.166
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.656
RMSE = 0.205
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.335
RMSE = 0.285
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.127
RMSE = 0.321
log Fitted Cr
log Observed Cr


Reach scale (dry season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.287
RMSE = 0.237
Riparian scale (dry season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.463
RMSE = 0.194
Sub-basin scale (dry season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.324
RMSE = 0.215
Reach scale (wet season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.283
RMSE = 0.333
Riparian scale (wet season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.441
RMSE = 0.285
Sub-basin scale (wet season)
Land use only scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.218
RMSE = 0.337
Reach scale (dry season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.43
RMSE = 0.191
Riparian scale (dry season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.477
RMSE = 0.189
Sub-basin scale (dry season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.269
RMSE = 0.216
Reach scale (wet season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.284
RMSE = 0.332
Riparian scale (wet season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.403
RMSE = 0.295
Sub-basin scale (wet season)
Combined scenario
log Fitted Cu
log Observed Cu
Adj. R2 = 0.157
RMSE = 0.339

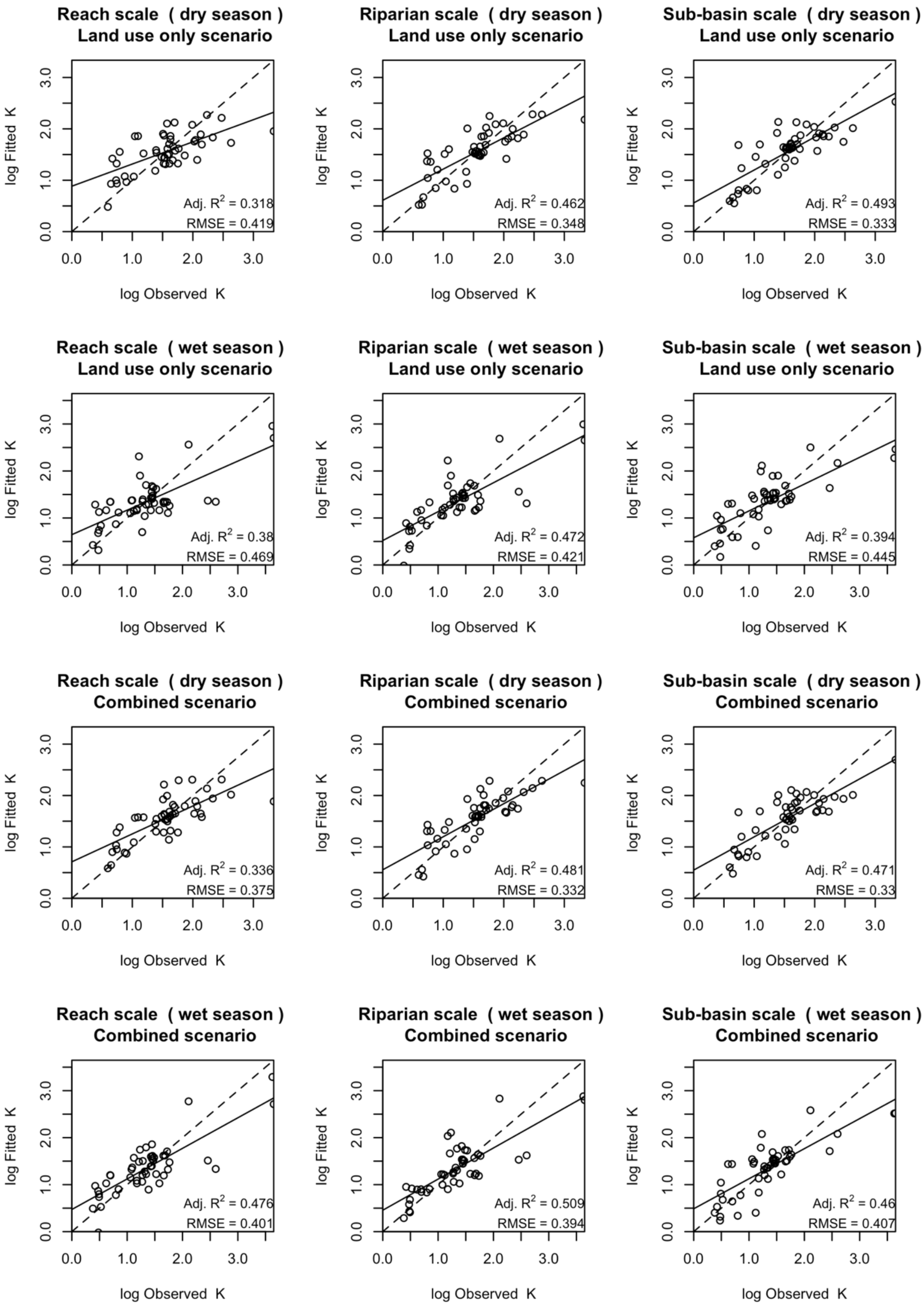

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.432
RMSE = 0.124
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.253
RMSE = 0.14
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.206
RMSE = 0.142
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.15
RMSE = 0.192
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.44
RMSE = 0.15
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.244
RMSE = 0.171
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.329
RMSE = 0.135
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.0808
RMSE = 0.155
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.146
RMSE = 0.148
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.574
RMSE = 0.127
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.417
RMSE = 0.146
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.303
RMSE = 0.157
log Fitted DO
log Observed DO


Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.366
RMSE = 1.01
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.695
RMSE = 0.671
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.666
RMSE = 0.68
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.0957
RMSE = 1.99
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.468
RMSE = 1.49
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.339
RMSE = 1.66
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.678
RMSE = 0.748
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.717
RMSE = 0.627
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.615
RMSE = 0.72
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.295
RMSE = 1.81
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.432
RMSE = 1.47
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.276
RMSE = 1.64
log Fitted E.coli
log Observed E.coli

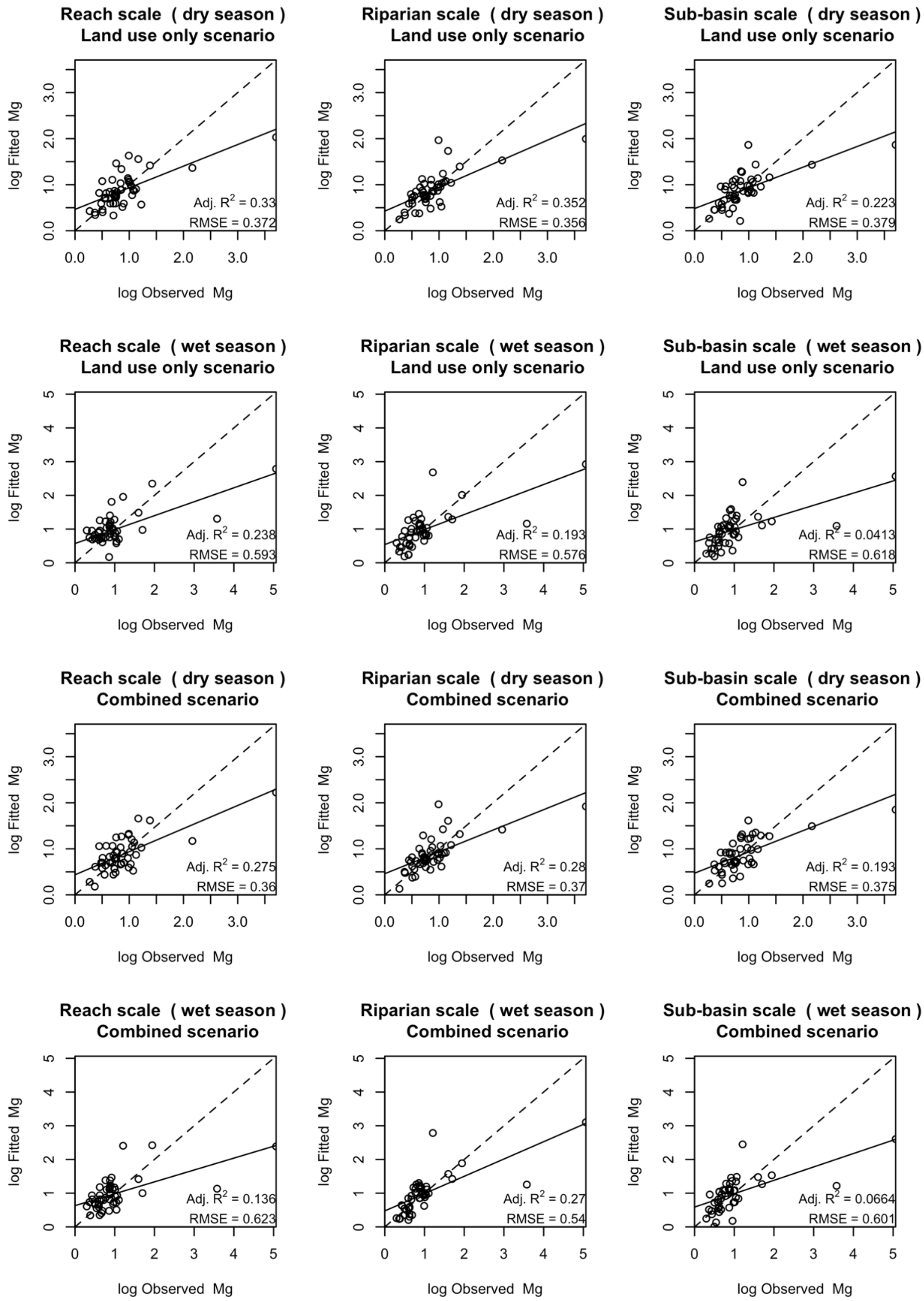

Reach scale (dry season)
Land use only scenario
log Fitted Fe
log Observed Fe
Adj. R^2 = 0.281
RMSE = 1.08
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.438
RMSE = 0.937
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.349
RMSE = 0.994
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.348
RMSE = 0.849
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.419
RMSE = 0.79
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.34
RMSE = 0.83
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.347
RMSE = 0.966
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.47
RMSE = 0.884
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.407
RMSE = 0.906
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.386
RMSE = 0.8
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.335
RMSE = 0.82
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.254
RMSE = 0.842

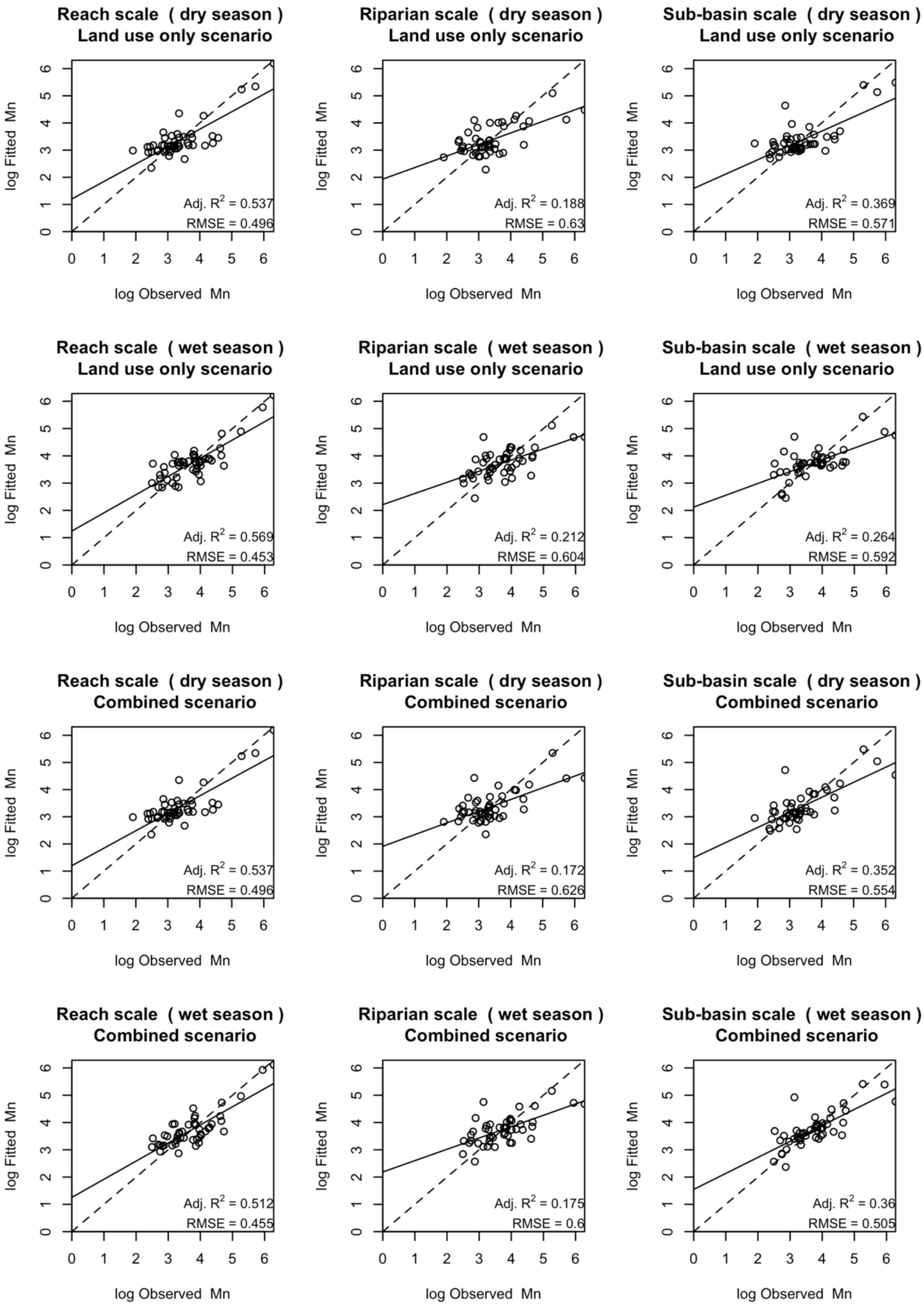


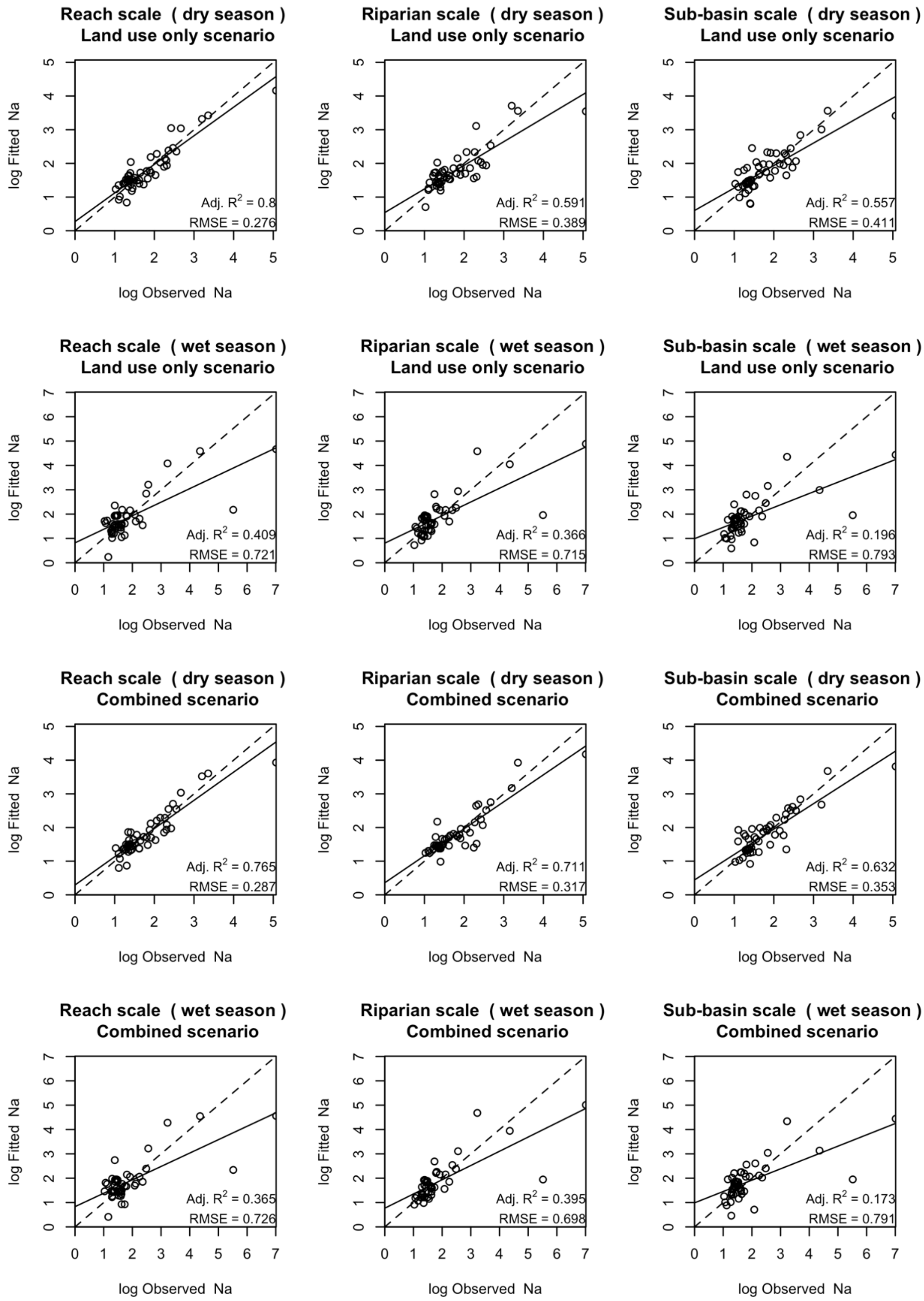
Reach scale (dry season)
Land use only scenario
log Fitted Fluoride
log Observed Fluoride
Adj. R^2 = 0.156
RMSE = 0.0518
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.398
RMSE = 0.0432
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.326
RMSE = 0.0451
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.0479
RMSE = 0.139
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.063
RMSE = 0.146
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = -0.324
RMSE = 0.147
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.177
RMSE = 0.0491
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.26
RMSE = 0.0479
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.373
RMSE = 0.0409
Reach scale (wet season)
Combined scenario
Adj. R^2 = -0.0669
RMSE = 0.136
Riparian scale (wet season)
Combined scenario
Adj. R^2 = -0.29
RMSE = 0.147
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = -0.4
RMSE = 0.146



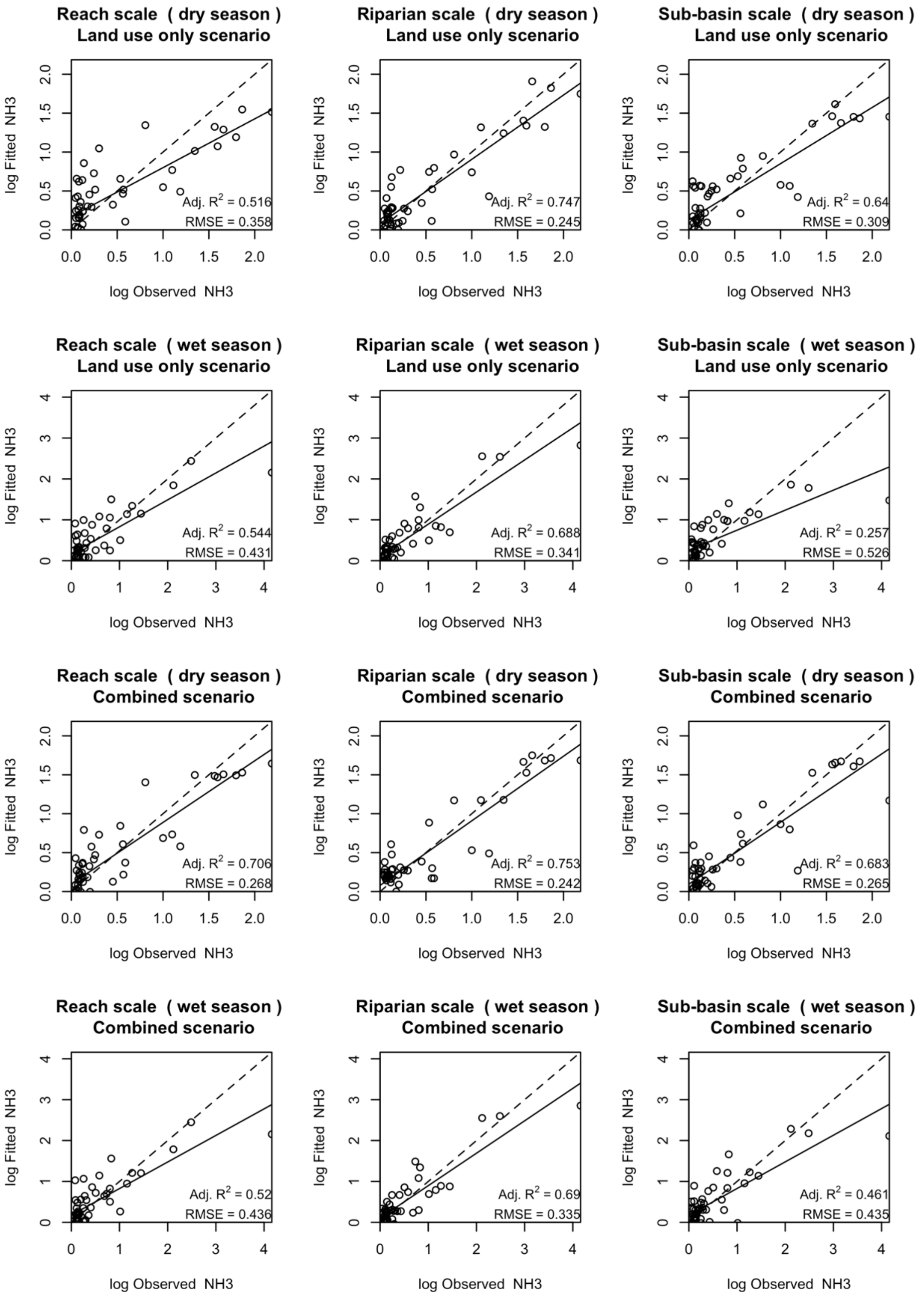
Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.318
RMSE = 0.419
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.462
RMSE = 0.348
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.493
RMSE = 0.333
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.38
RMSE = 0.469
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.472
RMSE = 0.421
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.394
RMSE = 0.445
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.336
RMSE = 0.375
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.481
RMSE = 0.332
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.471
RMSE = 0.33
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.476
RMSE = 0.401
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.509
RMSE = 0.394
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.46
RMSE = 0.407
log Fitted K
log Observed K

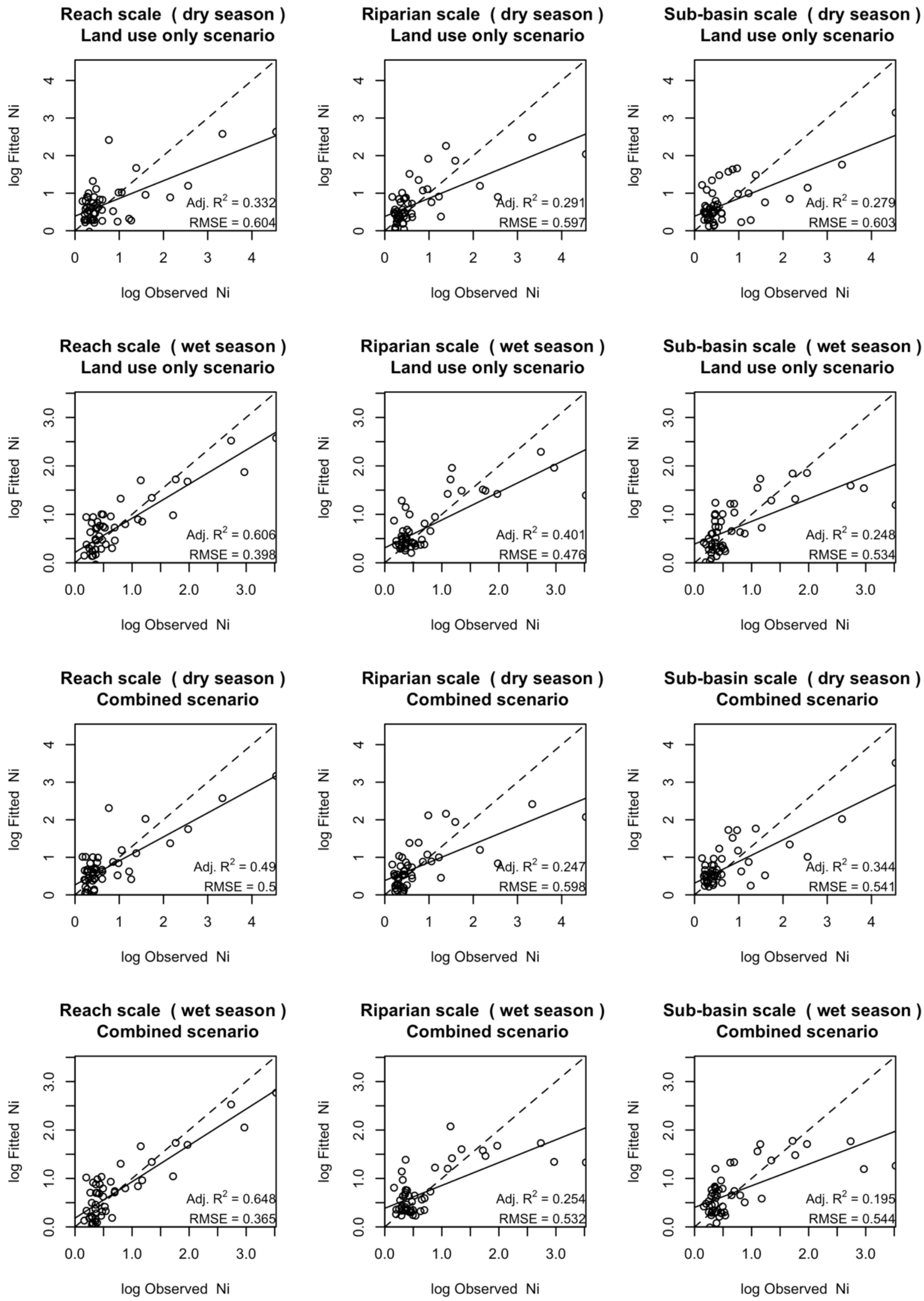

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.33
RMSE = 0.372
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.352
RMSE = 0.356
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.223
RMSE = 0.379
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.238
RMSE = 0.593
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.193
RMSE = 0.576
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.0413
RMSE = 0.618
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.275
RMSE = 0.36
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.28
RMSE = 0.37
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.193
RMSE = 0.375
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.136
RMSE = 0.623
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.27
RMSE = 0.54
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.0664
RMSE = 0.601
log Observed Mg
log Fitted Mg

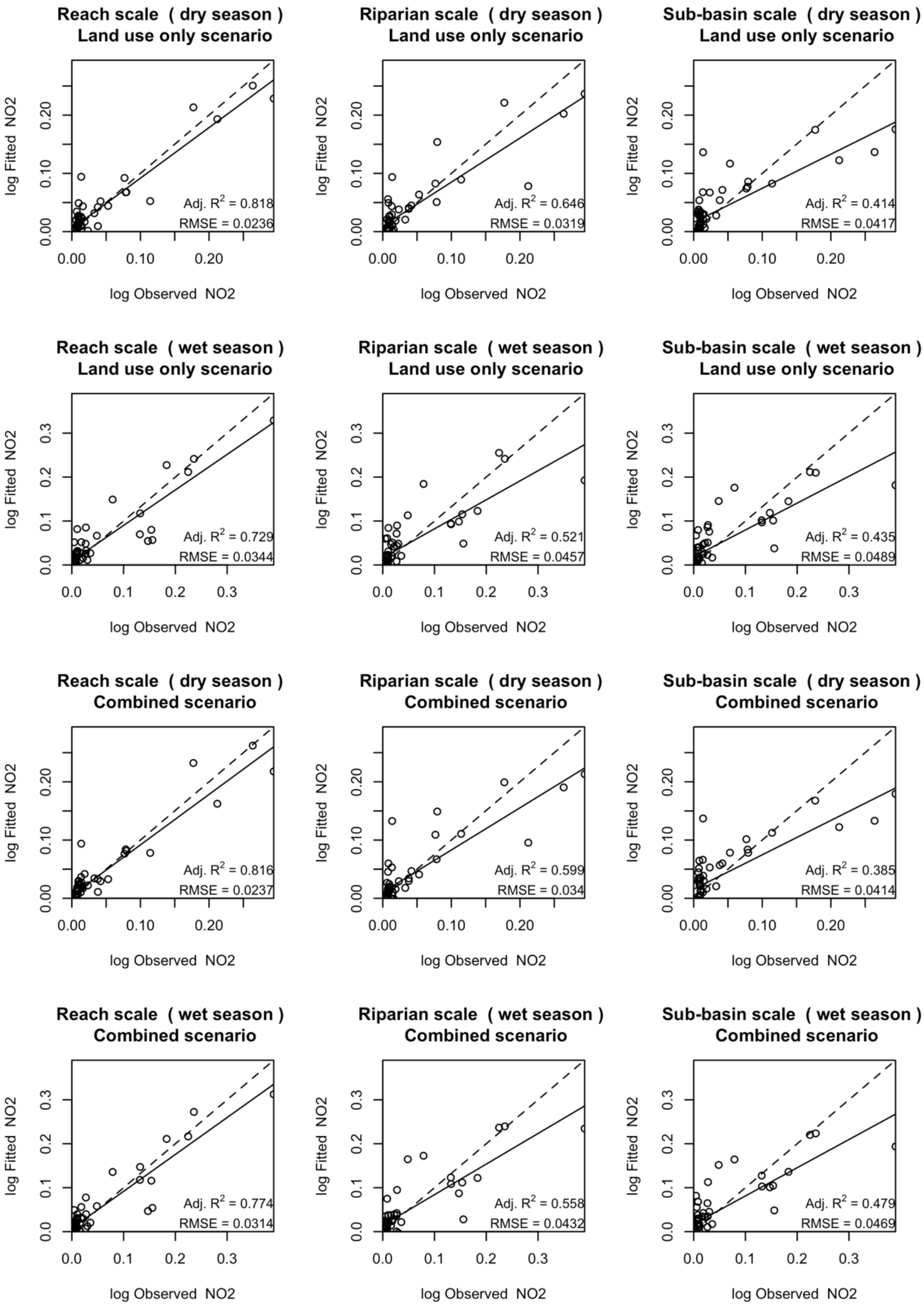

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.537
RMSE = 0.496
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.188
RMSE = 0.63
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.369
RMSE = 0.571
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.569
RMSE = 0.453
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.212
RMSE = 0.604
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.264
RMSE = 0.592
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.537
RMSE = 0.496
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.172
RMSE = 0.626
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.352
RMSE = 0.554
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.512
RMSE = 0.455
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.175
RMSE = 0.6
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.36
RMSE = 0.505
log Fitted Mn
log Observed Mn

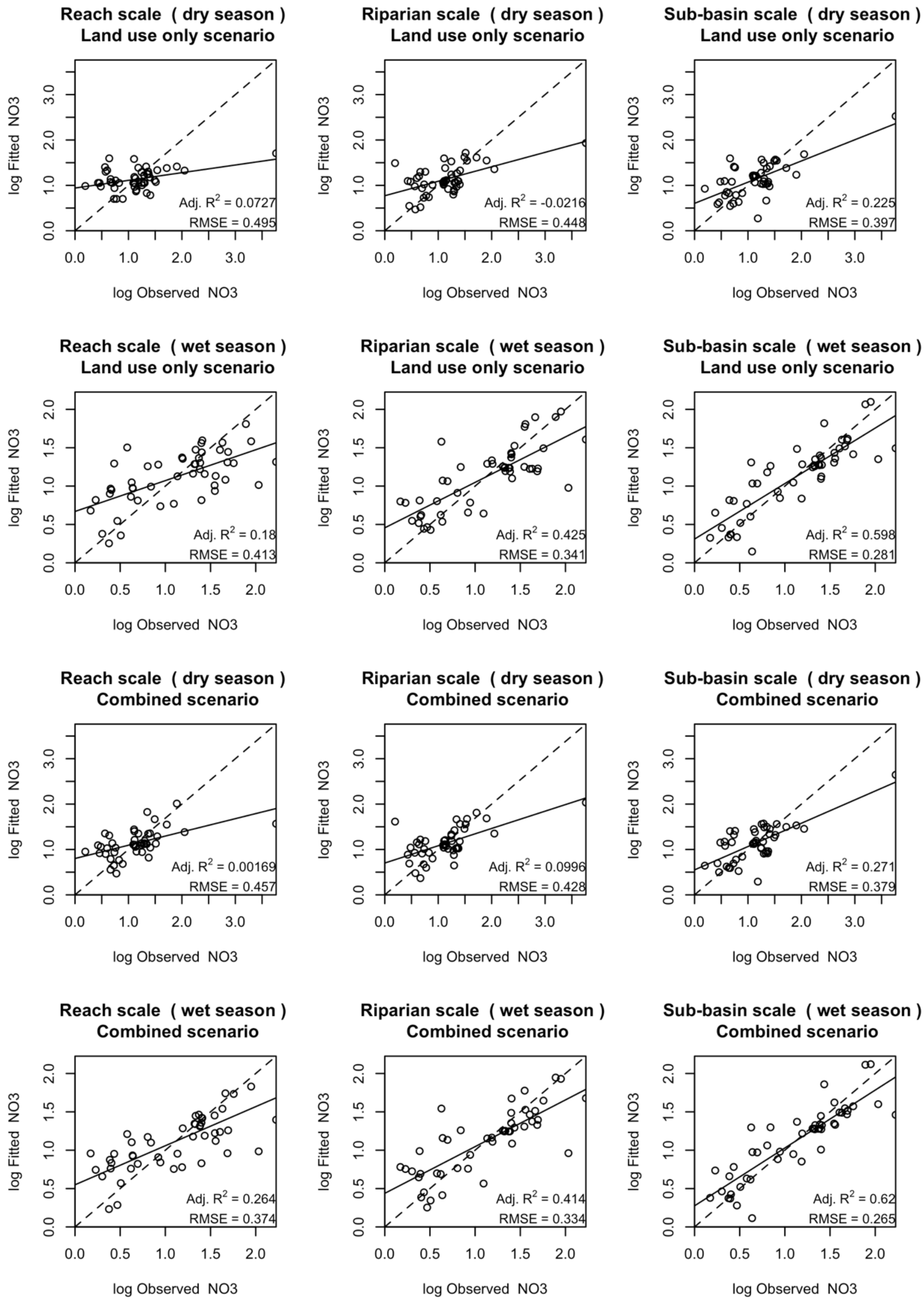


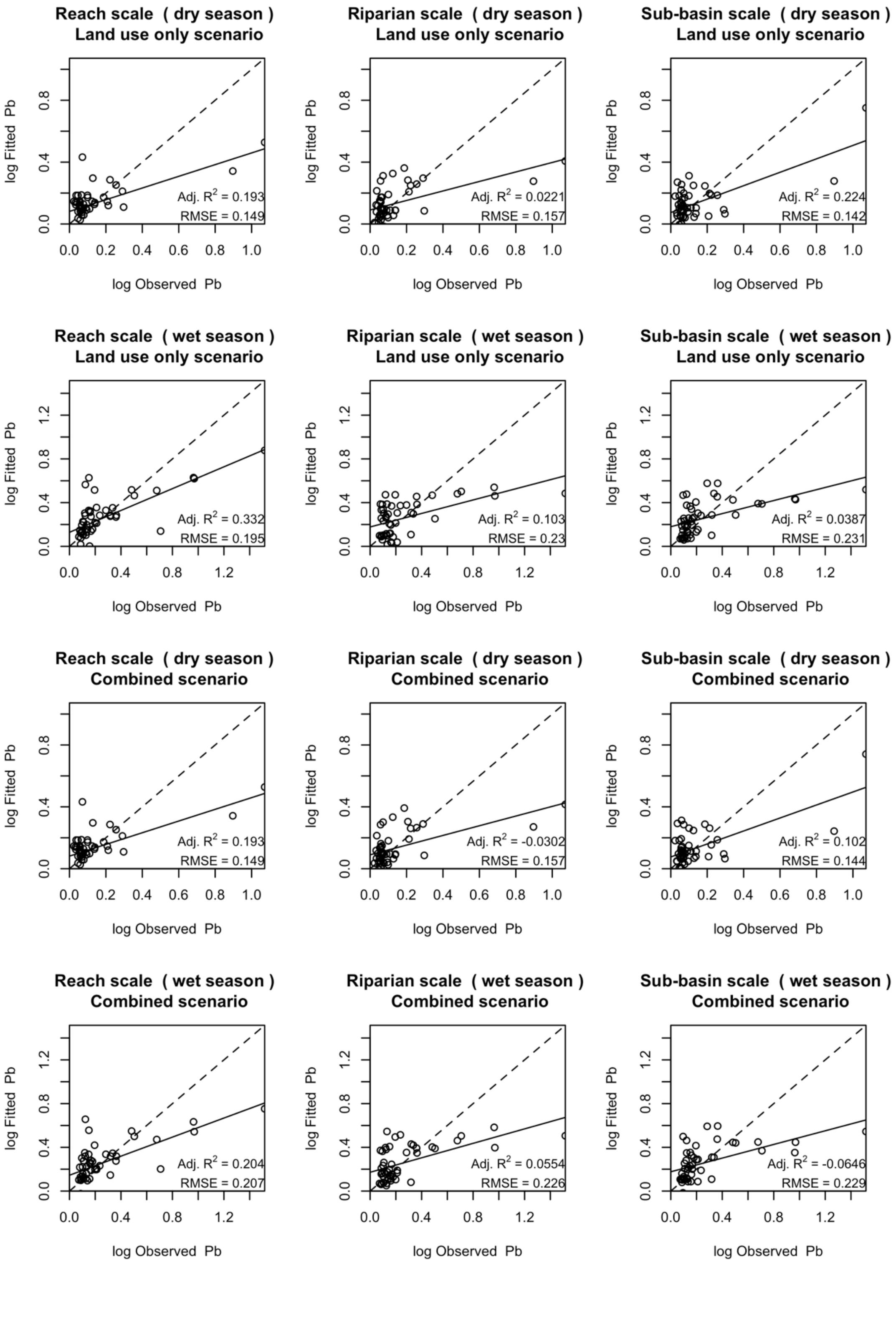
Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.8
RMSE = 0.276
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.591
RMSE = 0.389
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.557
RMSE = 0.411
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.409
RMSE = 0.721
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.366
RMSE = 0.715
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.196
RMSE = 0.793
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.765
RMSE = 0.287
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.711
RMSE = 0.317
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.632
RMSE = 0.353
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.365
RMSE = 0.726
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.395
RMSE = 0.698
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.173
RMSE = 0.791
log Fitted Na
log Observed Na


Reach scale (dry season)
Land use only scenario
Adj. R² = 0.516
RMSE = 0.358
Riparian scale (dry season)
Land use only scenario
Adj. R² = 0.747
RMSE = 0.245
Sub-basin scale (dry season)
Land use only scenario
Adj. R² = 0.64
RMSE = 0.309
Reach scale (wet season)
Land use only scenario
Adj. R² = 0.544
RMSE = 0.431
Riparian scale (wet season)
Land use only scenario
Adj. R² = 0.688
RMSE = 0.341
Sub-basin scale (wet season)
Land use only scenario
Adj. R² = 0.257
RMSE = 0.526
Reach scale (dry season)
Combined scenario
Adj. R² = 0.706
RMSE = 0.268
Riparian scale (dry season)
Combined scenario
Adj. R² = 0.753
RMSE = 0.242
Sub-basin scale (dry season)
Combined scenario
Adj. R² = 0.683
RMSE = 0.265
Reach scale (wet season)
Combined scenario
Adj. R² = 0.52
RMSE = 0.436
Riparian scale (wet season)
Combined scenario
Adj. R² = 0.69
RMSE = 0.335
Sub-basin scale (wet season)
Combined scenario
Adj. R² = 0.461
RMSE = 0.435
log Fitted NH3
log Observed NH3

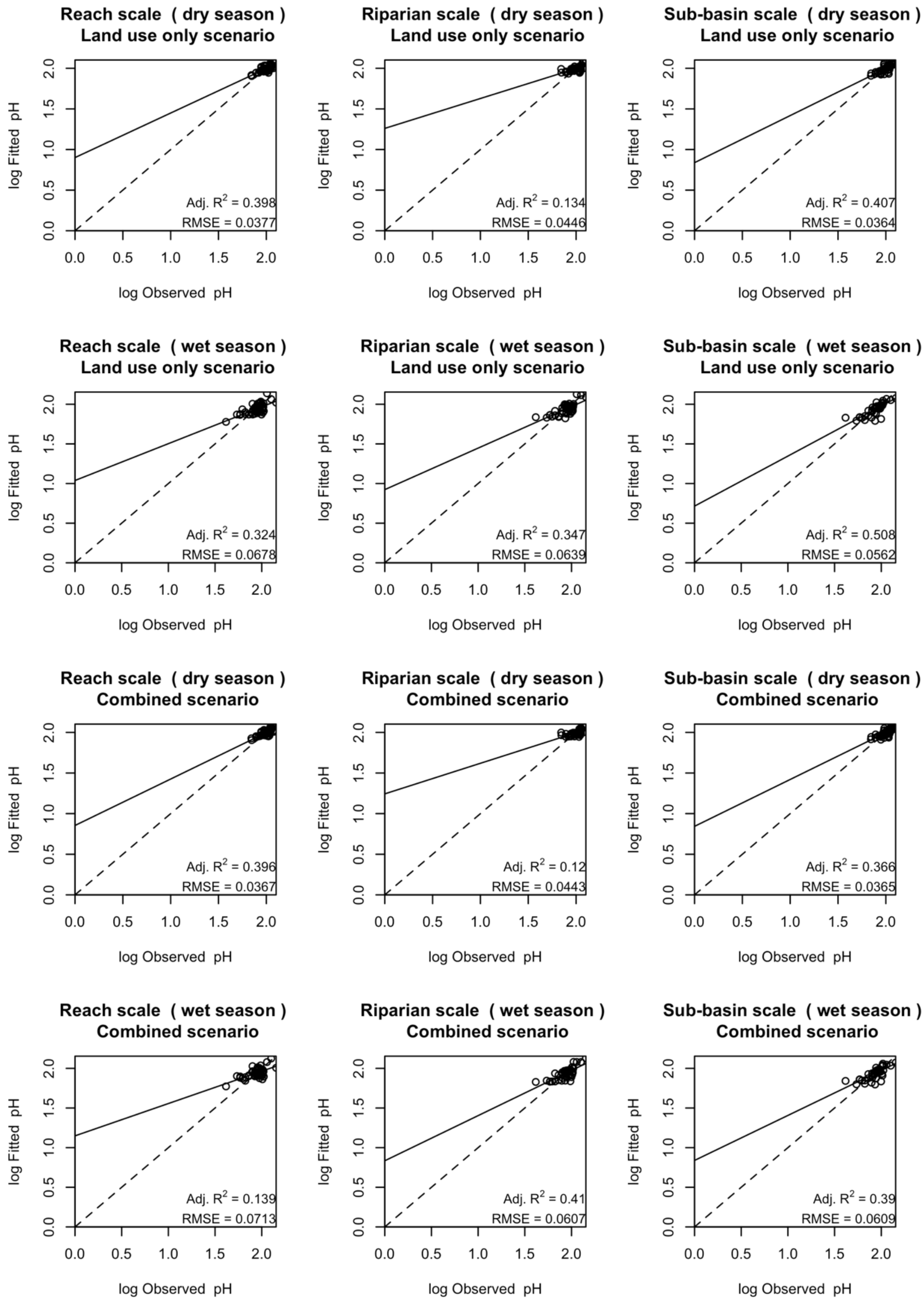

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.332
RMSE = 0.604
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.291
RMSE = 0.597
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.279
RMSE = 0.603
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.606
RMSE = 0.398
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.401
RMSE = 0.476
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.248
RMSE = 0.534
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.49
RMSE = 0.5
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.247
RMSE = 0.598
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.344
RMSE = 0.541
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.648
RMSE = 0.365
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.254
RMSE = 0.532
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.195
RMSE = 0.544
log Fitted Ni
log Observed Ni

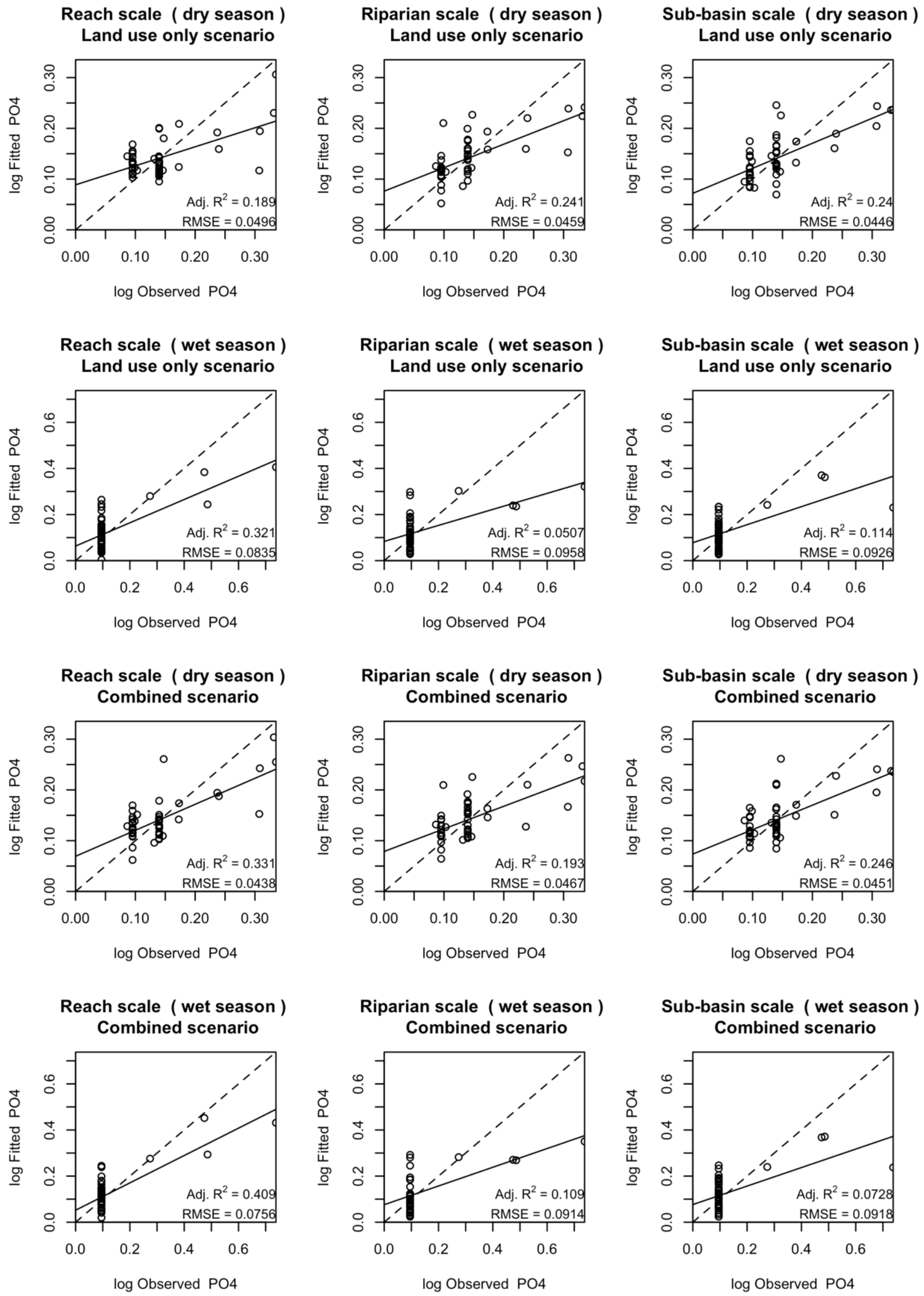

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.818
RMSE = 0.0236
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.646
RMSE = 0.0319
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.414
RMSE = 0.0417
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.729
RMSE = 0.0344
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.521
RMSE = 0.0457
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.435
RMSE = 0.0489
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.816
RMSE = 0.0237
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.599
RMSE = 0.034
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.385
RMSE = 0.0414
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.774
RMSE = 0.0314
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.558
RMSE = 0.0432
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.479
RMSE = 0.0469
log Fitted NO2
log Observed NO2

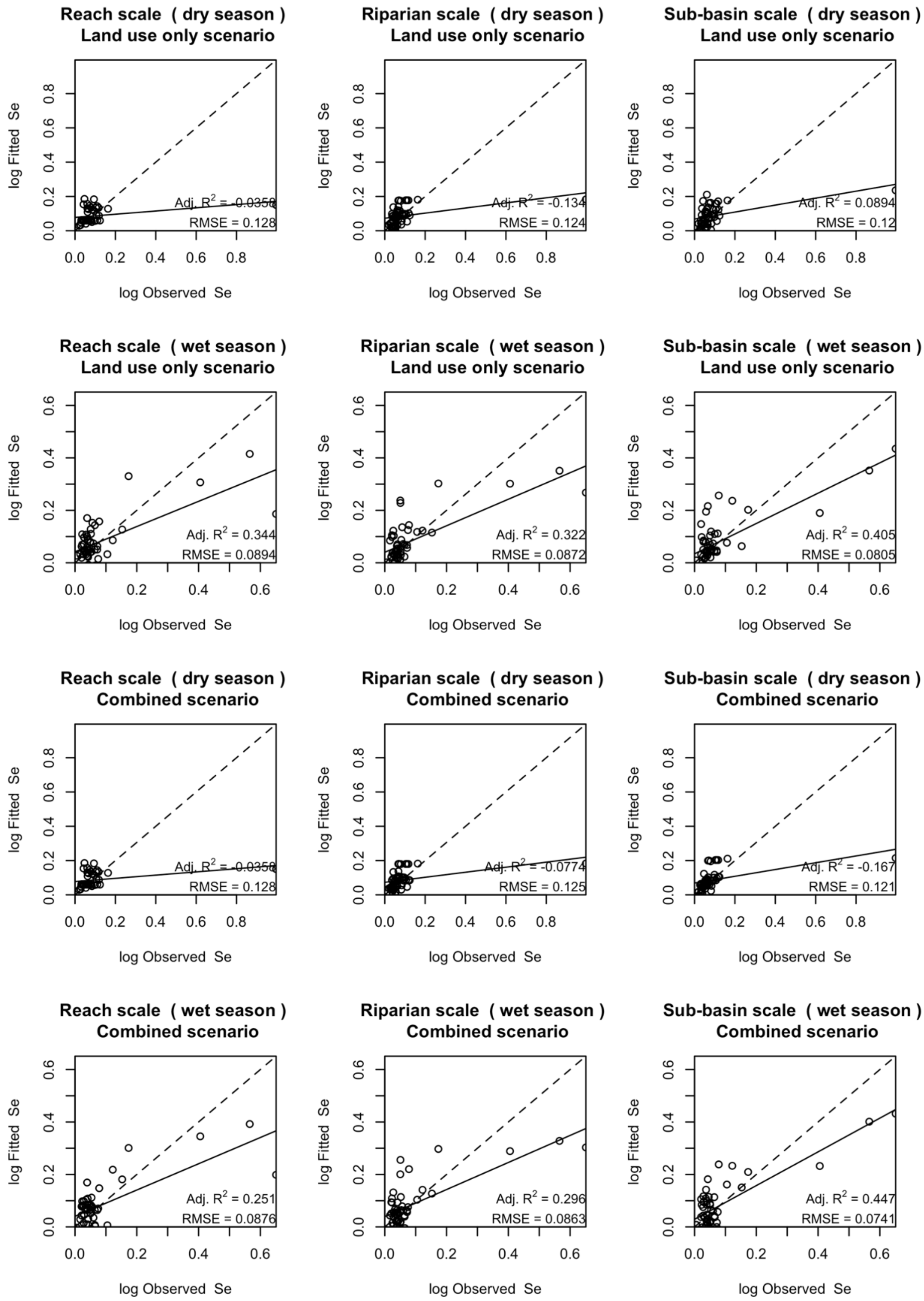

Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.0727
RMSE = 0.495
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = -0.0216
RMSE = 0.448
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.225
RMSE = 0.397
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.18
RMSE = 0.413
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.425
RMSE = 0.341
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.598
RMSE = 0.281
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.00169
RMSE = 0.457
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.0996
RMSE = 0.428
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.271
RMSE = 0.379
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.264
RMSE = 0.374
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.414
RMSE = 0.334
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.62
RMSE = 0.265
log Fitted NO3
log Observed NO3

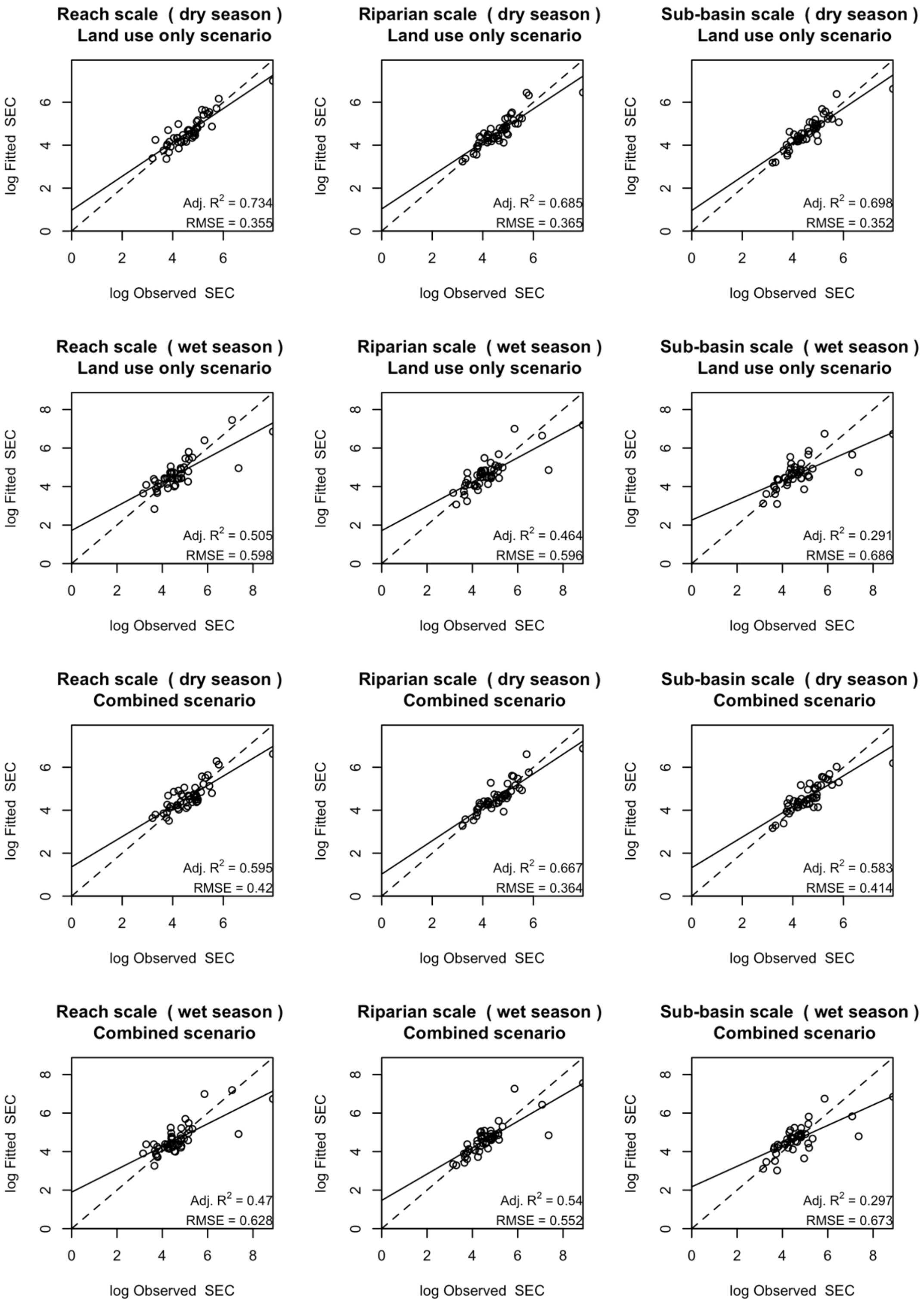


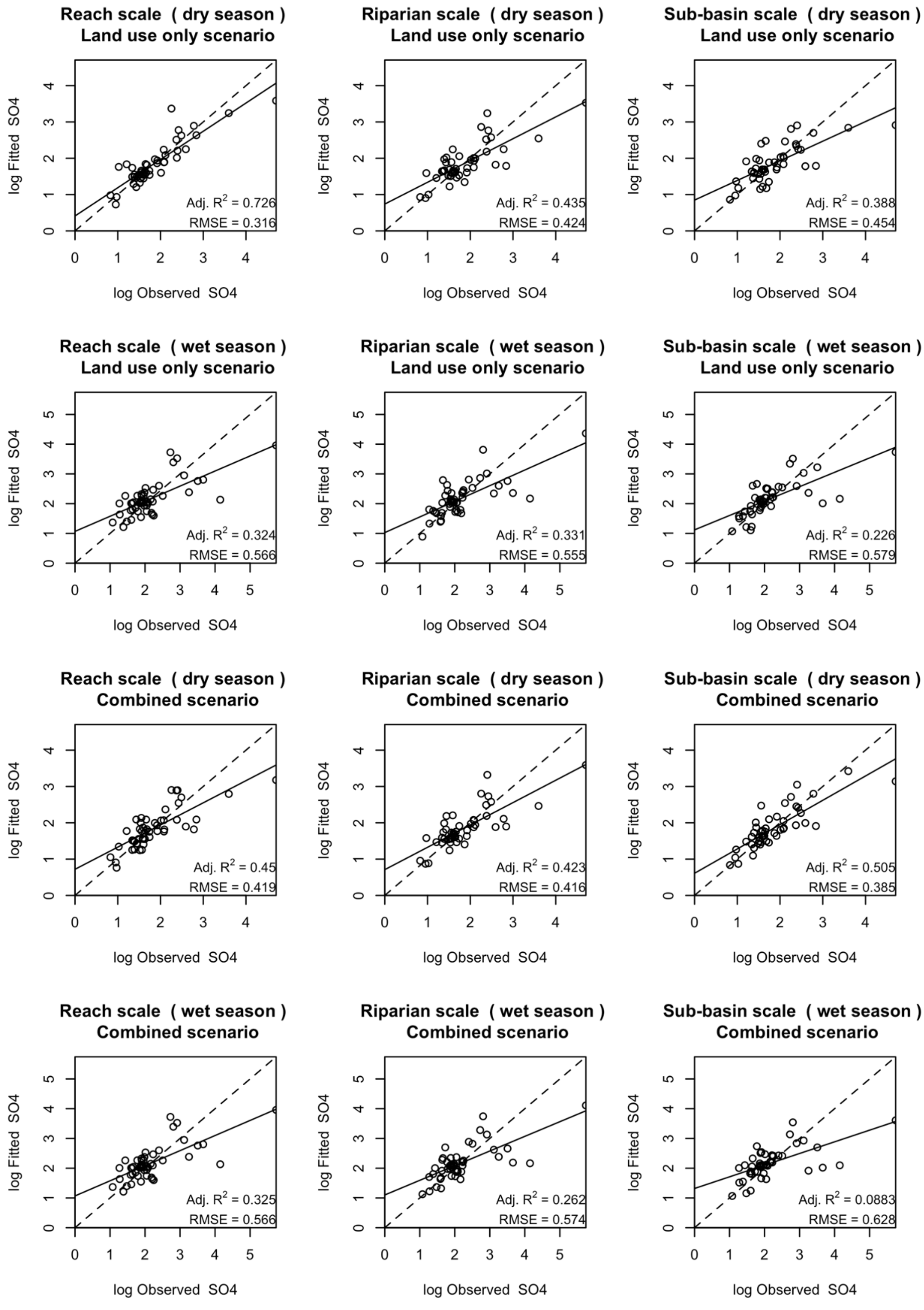
Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.193
RMSE = 0.149
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.0221
RMSE = 0.157
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.224
RMSE = 0.142
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.332
RMSE = 0.195
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.103
RMSE = 0.23
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.0387
RMSE = 0.231
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.193
RMSE = 0.149
Riparian scale (dry season)
Combined scenario
Adj. R^2 = -0.0302
RMSE = 0.157
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.102
RMSE = 0.144
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.204
RMSE = 0.207
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.0554
RMSE = 0.226
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = -0.0646
RMSE = 0.229
log Fitted Pb
log Observed Pb

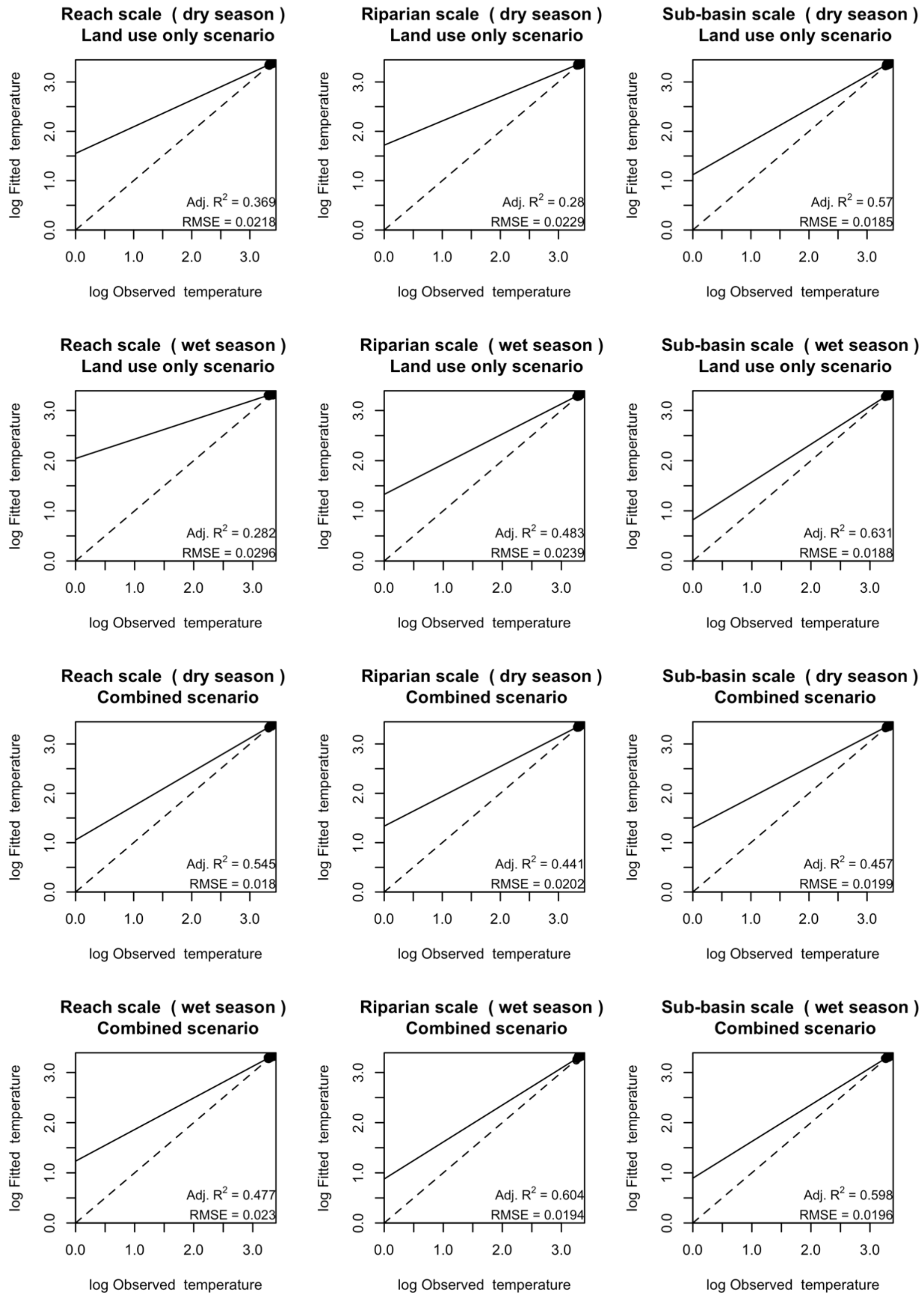

Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.398
RMSE = 0.0377
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.134
RMSE = 0.0446
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.407
RMSE = 0.0364
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.324
RMSE = 0.0678
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.347
RMSE = 0.0639
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.508
RMSE = 0.0562
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.396
RMSE = 0.0367
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.12
RMSE = 0.0443
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.366
RMSE = 0.0365
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.139
RMSE = 0.0713
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.41
RMSE = 0.0607
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.39
RMSE = 0.0609
log Fitted pH
log Observed pH

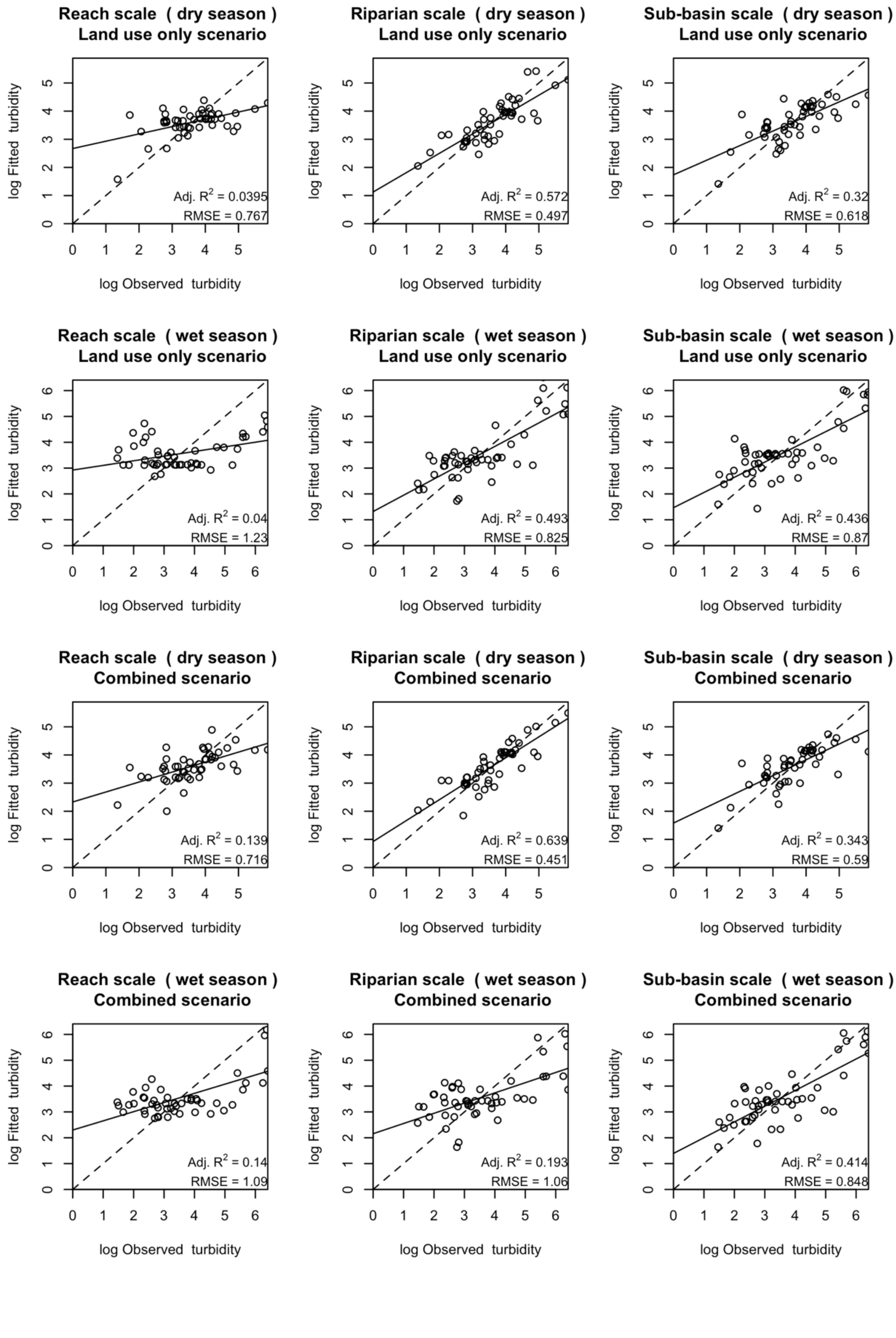

Reach scale (dry season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.189
RMSE = 0.0496
Riparian scale (dry season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.241
RMSE = 0.0459
Sub-basin scale (dry season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.24
RMSE = 0.0446
Reach scale (wet season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.321
RMSE = 0.0835
Riparian scale (wet season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.0507
RMSE = 0.0958
Sub-basin scale (wet season)
Land use only scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.114
RMSE = 0.0926
Reach scale (dry season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.331
RMSE = 0.0438
Riparian scale (dry season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.193
RMSE = 0.0467
Sub-basin scale (dry season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.246
RMSE = 0.0451
Reach scale (wet season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.409
RMSE = 0.0756
Riparian scale (wet season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.109
RMSE = 0.0914
Sub-basin scale (wet season)
Combined scenario
log Fitted PO4
log Observed PO4
Adj. R^2 = 0.0728
RMSE = 0.0918

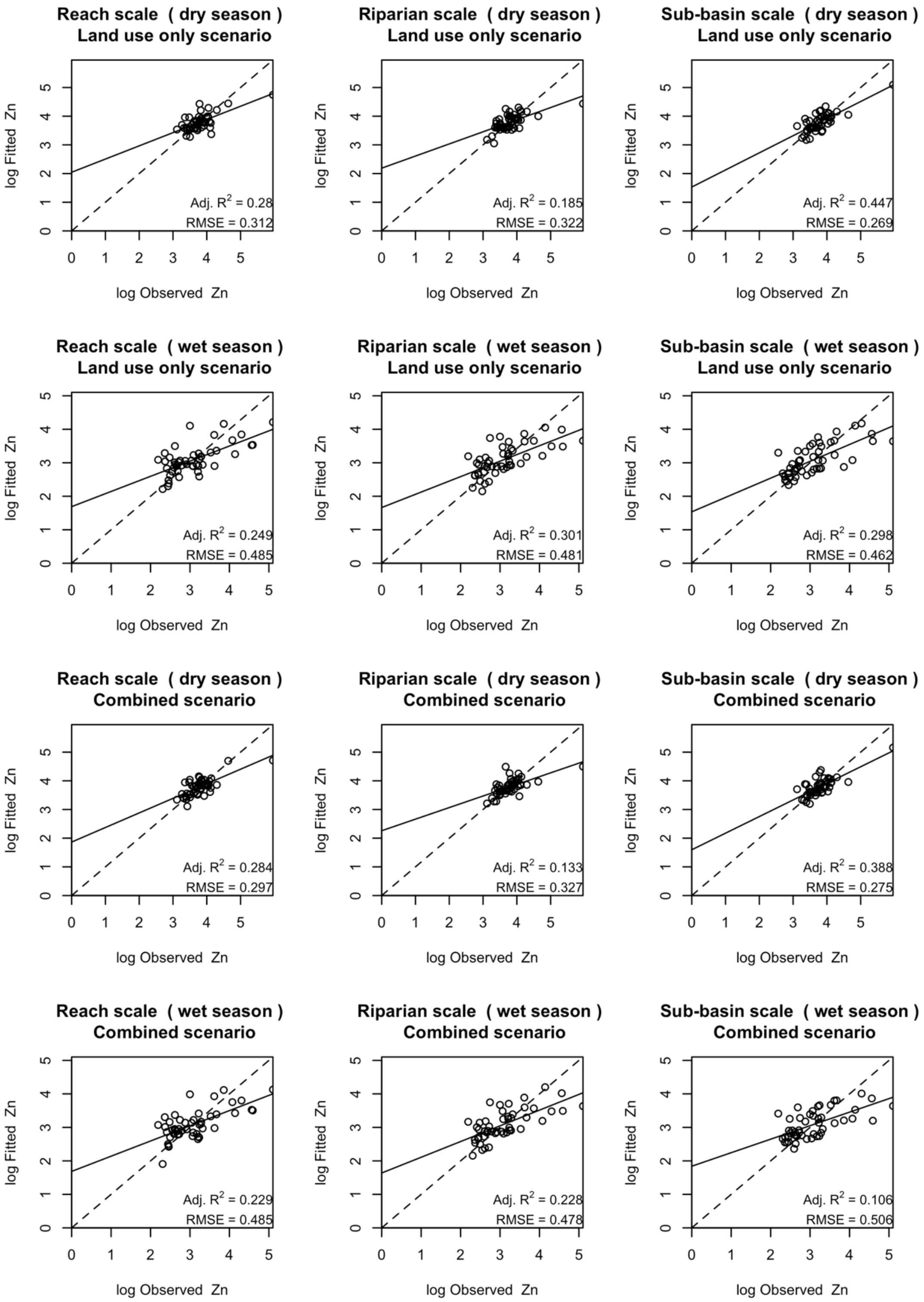

Reach scale (dry season)
Land use only scenario
Adj. R^2 = -0.0358
RMSE = 0.128
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = -0.134
RMSE = 0.124
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.0894
RMSE = 0.12
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.344
RMSE = 0.0894
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.322
RMSE = 0.0872
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.405
RMSE = 0.0805
Reach scale (dry season)
Combined scenario
Adj. R^2 = -0.0358
RMSE = 0.128
Riparian scale (dry season)
Combined scenario
Adj. R^2 = -0.0774
RMSE = 0.125
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = -0.167
RMSE = 0.121
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.251
RMSE = 0.0876
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.296
RMSE = 0.0863
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.447
RMSE = 0.0741
log Fitted Se
log Observed Se


Reach scale (dry season)
Land use only scenario
Adj. R2 = 0.734
RMSE = 0.355
Riparian scale (dry season)
Land use only scenario
Adj. R2 = 0.685
RMSE = 0.365
Sub-basin scale (dry season)
Land use only scenario
Adj. R2 = 0.698
RMSE = 0.352
Reach scale (wet season)
Land use only scenario
Adj. R2 = 0.505
RMSE = 0.598
Riparian scale (wet season)
Land use only scenario
Adj. R2 = 0.464
RMSE = 0.596
Sub-basin scale (wet season)
Land use only scenario
Adj. R2 = 0.291
RMSE = 0.686
Reach scale (dry season)
Combined scenario
Adj. R2 = 0.595
RMSE = 0.42
Riparian scale (dry season)
Combined scenario
Adj. R2 = 0.667
RMSE = 0.364
Sub-basin scale (dry season)
Combined scenario
Adj. R2 = 0.583
RMSE = 0.414
Reach scale (wet season)
Combined scenario
Adj. R2 = 0.47
RMSE = 0.628
Riparian scale (wet season)
Combined scenario
Adj. R2 = 0.54
RMSE = 0.552
Sub-basin scale (wet season)
Combined scenario
Adj. R2 = 0.297
RMSE = 0.673
log Fitted SEC
log Observed SEC


Reach scale (dry season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.726
RMSE = 0.316
Riparian scale (dry season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.435
RMSE = 0.424
Sub-basin scale (dry season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.388
RMSE = 0.454
Reach scale (wet season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.324
RMSE = 0.566
Riparian scale (wet season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.331
RMSE = 0.555
Sub-basin scale (wet season)
Land use only scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.226
RMSE = 0.579
Reach scale (dry season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.45
RMSE = 0.419
Riparian scale (dry season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.423
RMSE = 0.416
Sub-basin scale (dry season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.505
RMSE = 0.385
Reach scale (wet season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.325
RMSE = 0.566
Riparian scale (wet season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.262
RMSE = 0.574
Sub-basin scale (wet season)
Combined scenario
log Fitted SO4
log Observed SO4
Adj. R^2 = 0.0883
RMSE = 0.628


Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.369
RMSE = 0.0218
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.28
RMSE = 0.0229
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.57
RMSE = 0.0185
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.282
RMSE = 0.0296
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.483
RMSE = 0.0239
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.631
RMSE = 0.0188
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.545
RMSE = 0.018
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.441
RMSE = 0.0202
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.457
RMSE = 0.0199
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.477
RMSE = 0.023
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.604
RMSE = 0.0194
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.598
RMSE = 0.0196
log Fitted temperature
log Observed temperature
0.0
1.0
2.0
3.0


Reach scale (dry season)
Land use only scenario
Adj. R^2 = 0.0395
RMSE = 0.767
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.572
RMSE = 0.497
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.32
RMSE = 0.618
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.04
RMSE = 1.23
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.493
RMSE = 0.825
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.436
RMSE = 0.87
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.139
RMSE = 0.716
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.639
RMSE = 0.451
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.343
RMSE = 0.59
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.14
RMSE = 1.09
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.193
RMSE = 1.06
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.414
RMSE = 0.848
log Fitted turbidity
log Observed turbidity


Reach scale (dry season)
Land use only scenario
log Fitted Zn
log Observed Zn
Adj. R^2 = 0.28
RMSE = 0.312
Riparian scale (dry season)
Land use only scenario
Adj. R^2 = 0.185
RMSE = 0.322
Sub-basin scale (dry season)
Land use only scenario
Adj. R^2 = 0.447
RMSE = 0.269
Reach scale (wet season)
Land use only scenario
Adj. R^2 = 0.249
RMSE = 0.485
Riparian scale (wet season)
Land use only scenario
Adj. R^2 = 0.301
RMSE = 0.481
Sub-basin scale (wet season)
Land use only scenario
Adj. R^2 = 0.298
RMSE = 0.462
Reach scale (dry season)
Combined scenario
Adj. R^2 = 0.284
RMSE = 0.297
Riparian scale (dry season)
Combined scenario
Adj. R^2 = 0.133
RMSE = 0.327
Sub-basin scale (dry season)
Combined scenario
Adj. R^2 = 0.388
RMSE = 0.275
Reach scale (wet season)
Combined scenario
Adj. R^2 = 0.229
RMSE = 0.485
Riparian scale (wet season)
Combined scenario
Adj. R^2 = 0.228
RMSE = 0.478
Sub-basin scale (wet season)
Combined scenario
Adj. R^2 = 0.106
RMSE = 0.506