

Performance Comparison of Spatial Outlier Detection Algorithms

Zetty Alwani Zulki Alwani¹, Rossita Mohamad Yunus^{2,*}, and Adriana Irawati Nur Ibrahim³

Rossita M. Yunus

^{1,2,3} Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya

4 - 5 AUGUST 2021



CONTENTS

- 1 INTRODUCTION
- 2 OBJECTIVES
- 3 LITERATURE REVIEW/JUSTIFICATIONS
- 4 METHODOLOGY
- 5 RESULTS AND DISCUSSIONS
- 6 CONCLUSIONS



Introduction

A spatial outlier is a spatially referenced object whose non-spatial attribute values are significantly different from those of other spatially referenced objects in its spatial neighborhoods (Haslett et al., 1991; Cressie, 1993; Shekhar & Chawla, 2002).

Several procedures outlined in the literature can only detect outliers in a non-spatial manner. Thus, the results obtained will have no protection against spatial outliers.



Objectives

In this study, we propose two neighborhood-based approach algorithms based on robust median.

The objective of this study is to assess the detection performance of four algorithms for detecting spatial outliers in the PM10 concentration data.

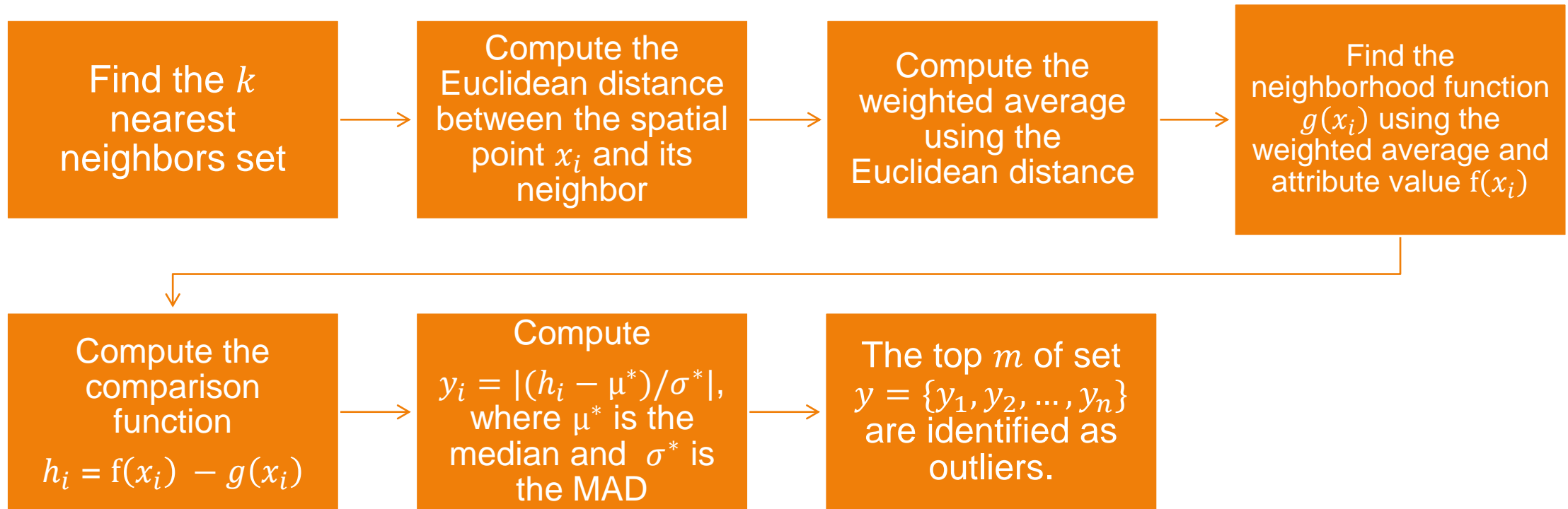


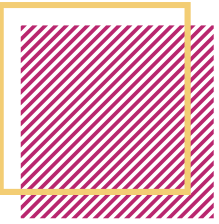
Literature Review

Common procedures in detecting spatial outliers such as the Moran's scatter plot and Z algorithm (Anselin, 1995; Haining, 1993; Shekhar et al., 2001) have a drawback where they may choose some good points as spatial outliers (cf. Chen et al., 2008; Lu et al., 2003).

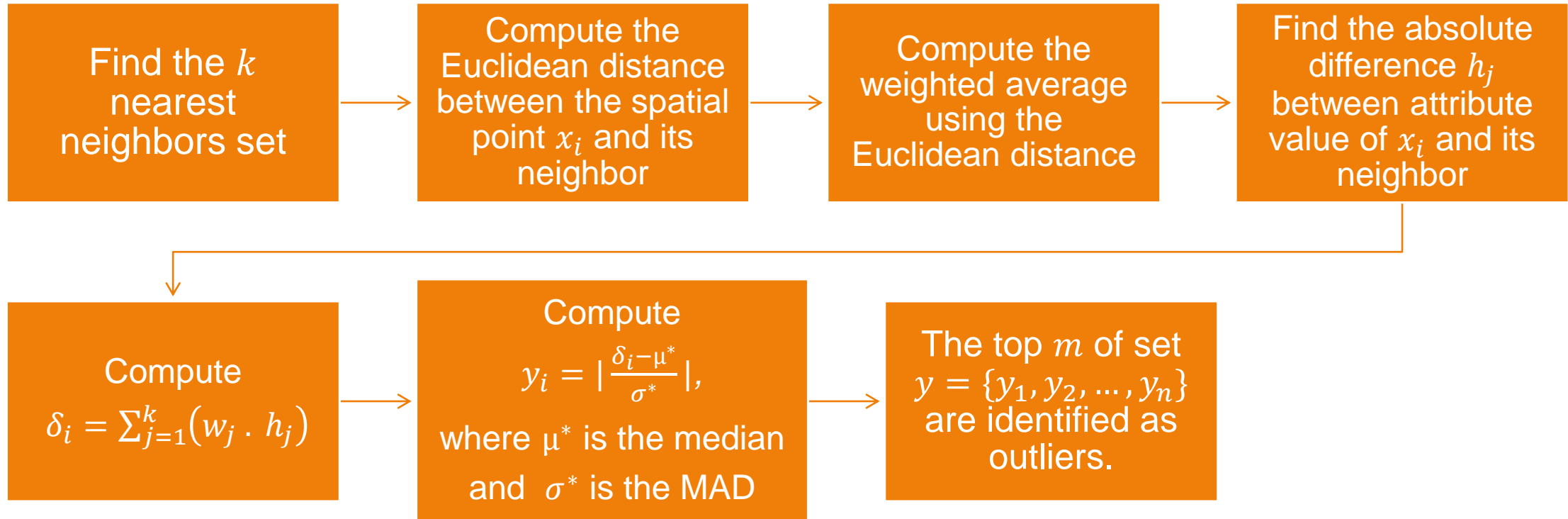
A few neighborhood-based approach algorithms such as the Weighted Z algorithm and the Average Difference algorithm (Kou et al, 2006) proposed in the literature have resolved this issue.

Methodology: The proposed method: Algorithm 1





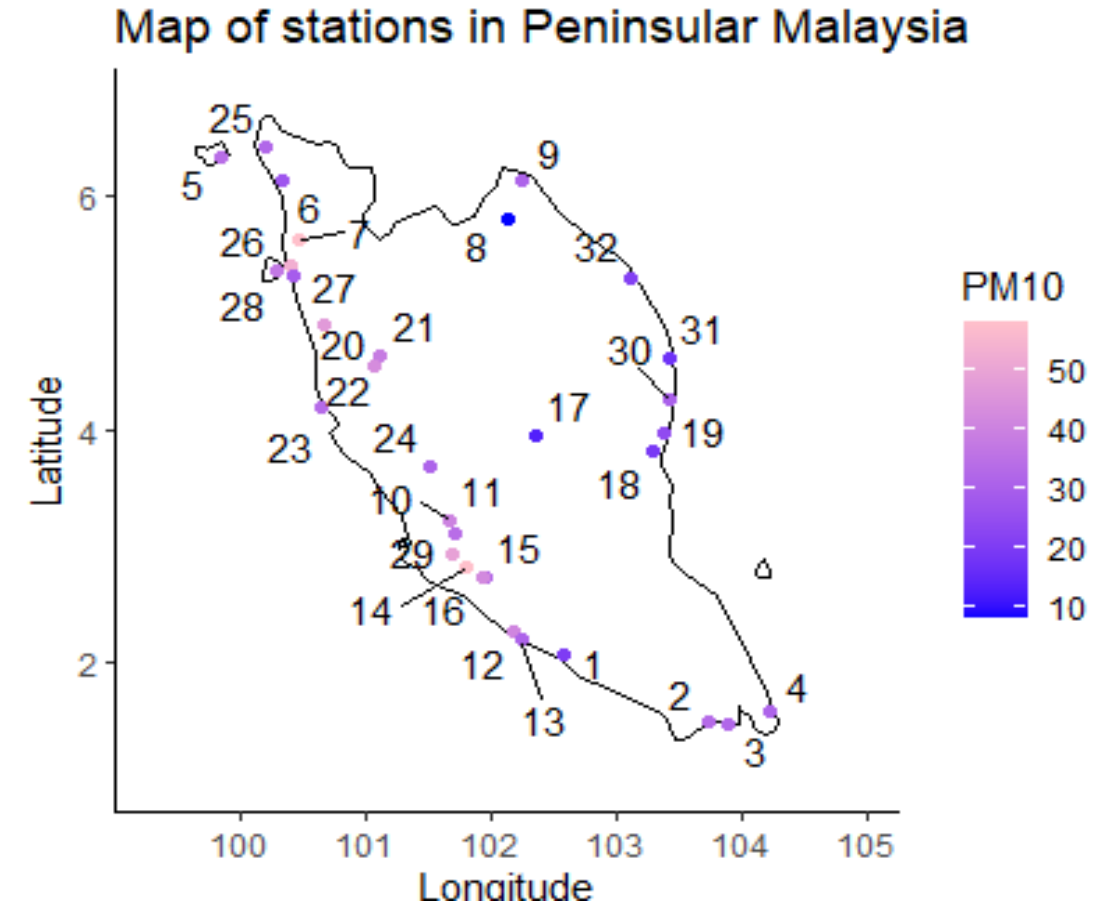
Methodology – The proposed method: Algorithm 2



RESULTS AND DISCUSSIONS

| Methods | Top four outlier candidates |
|-------------------------------------|--|
| Weighted Z (Kou et al., 2006) | 1. Station 8 2. Station 7 3. Station 14 4. Station 27 |
| Median Weighted Z (Kou et al. 2006) | 1. Station 8 2. Station 7 3. Station 27 4. Station 14 |

| Methods | Top four outlier candidates |
|-------------|---|
| Algorithm 1 | 1. Station 8 2. Station 27 3. Station 9 4. Station 7 |
| Algorithm 2 | 1. Station 8 2. Station 27 3. Station 9 4. Station 7 |





Conclusions

The methods with robust measures have shown comparable results like the other two methods when using the PM10 data.

we conclude that Stations 8, 7, and 27 are spatial outliers because they are selected in the top 4 by all the algorithms.

However, the rank of the spatial outliers in the top three is different for both existing and proposed methods. The difference in the order of the candidate shows that the performance of the algorithms dissimilar.

