




Article

# Dataset Reduction Techniques to Enable SVD Analysis

Laurens Bogaardt <sup>1</sup>, Romulo Goncalves <sup>1</sup>, Raul Zurita-Milla <sup>2,\*</sup> and Emma Izquierdo-Verdiguier <sup>3</sup>

<sup>1</sup> Netherlands eScience Center; l.bogaardt@esciencecenter.nl, r.goncalves@esciencecenter.nl

<sup>2</sup> Faculty ITC, University of Twente; r.zurita-milla@utwente.nl

<sup>3</sup> Faculty IPL, Universitat de Valencia; emma.izquierdo@uv.es

\* Correspondence: r.zurita-milla@utwente.nl

Academic Editor: name

Version July 30, 2018 submitted to ISPRS Int. J. Geo-Inf.

**Abstract:** Performing an SVD analysis on large datasets can be time consuming and costly. Often, techniques exist to arrive at the same output, or at a close approximation, which require far less effort. This article looks at several such techniques and at the inherent scale of the structure within the data. When the values of a dataset vary slowly, e.g. in a spatial field of temperature over a country, there is a high level of autocorrelation and the structure of the field has a large scale. Datasets need not have a high resolution to describe such fields. Using generated *Gaussian Random Fields* with various levels of autocorrelation, we examine rank decomposition, coarsening and approximate SVD procedures. This article outlines when certain techniques can be useful and makes predictions about the error incurred in the approximate techniques based on the level of autocorrelation of the input datasets. Finally, these techniques and predictions are verified using real-world geospatial datasets.

**Keywords:** Singular value decomposition, autocorrelation, data reduction, coarsening, approximate SVD, Gaussian Random Fields

## 1. Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance. The current state of the research field should be reviewed carefully and key publications cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the principal conclusions. As far as possible, please keep the introduction comprehensible to scientists outside your particular field of research. Citing a journal paper [1]. And now citing a book reference []. Please use the command[] for the following MDPI journals, which use author-date citation: Administrative Sciences, Arts, Econometrics, Economies, Genealogy, Humanities, IJFS, JRFM, Languages, Laws, Religions, Risks, Social Sciences.

## 2. Materials and Methods

Materials and Methods should be described with sufficient details to allow others to replicate and build on published results. Please note that publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.

Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.

Interventionary studies involving animals or humans, and other studies require ethical approval must list the authority that provided approval and the corresponding ethical approval code.

### 3. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

#### 3.1. Subsection

##### 3.1.1. Subsubsection

Bulleted lists look like this:

- First bullet
- Second bullet
- Third bullet

Numbered lists can be added as follows:

1. First item
2. Second item
3. Third item

The text continues here.

#### 3.2. Figures, Tables and Schemes

All figures and tables should be cited in the main text as Figure 1, Table 1, etc.



**Figure 1.** This is a figure, Schemes follow the same formatting. If there are multiple panels, they should be listed as: (a) Description of what is contained in the first panel. (b) Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered.

**Table 1.** This is a table caption. Tables should be placed in the main text near to the first time they are cited.

Title 1	Title 2	Title 3
entry 1	data	data
entry 2	data	data

### 3.3. Formatting of Mathematical Components

This is an example of an equation:

$$a + b = c \quad (1)$$

Please punctuate equations as regular text. Theorem-type environments (including propositions, lemmas, corollaries etc.) can be formatted as follows:

**Theorem 1.** *Example text of a theorem.*

The text continues here. Proofs must be formatted as follows:

**Proof of Theorem 1.** Text of the proof. Note that the phrase ‘of Theorem 1’ is optional if it is clear which theorem is being referred to. □

The text continues here.

## 4. Discussion

Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

**Author Contributions:** Conceptualization, R.Z.; Methodology, L.B.; Software, L.B.; Validation, R.G., R.Z. and E.I.; Formal Analysis, L.B.; Investigation, L.B.; Resources, R.G., R.Z. and E.I.; Data Curation, R.G., R.Z. and E.I.; Writing—Original Draft Preparation, L.B.; Writing—Review & Editing, R.G., R.Z. and E.I.; Visualization, L.B.; Supervision, R.G. and R.Z.; Project Administration, R.G.; Funding Acquisition, R.Z.

**Funding:** Please add: “This research received no external funding” or “This research was funded by [name of funder] grant number [xxx].” Check carefully that the details given are accurate and use the standard spelling of funding agency names at <https://search.crossref.org/funding>, any errors may affect your future funding.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

SVD	Singular value decomposition
SI-x	Extended spring indices
AVHRR	Advanced very-high-resolution radiometer

1. Bogaardt, L. Dataset Reduction Depending On Structure Scale. <https://github.com/phenology/>, 2018.

© 2018 by the authors. Submitted to *ISPRS Int. J. Geo-Inf.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).