

## Plan Overview

---

*A Data Management Plan created using DMPTool*

**DMP ID:** <https://doi.org/10.48321/D1W94N>

**Title:** CDSE: Quantization-Based Methods for Optimal Nonparametric Inference

**Creator:** Zuofeng Shang - **ORCID:** [0000-0003-1125-2302](https://orcid.org/0000-0003-1125-2302)

**Affiliation:** New Jersey Institute of Technology

**Funder:** National Science Foundation (nsf.gov)

**Funding opportunity number:** PD 23-8084

**Grant:** <https://new.nsf.gov/funding/opportunities/computational-data-enabled-science-engineering-3>

**Template:** Arctic Data Center: NSF Polar Programs

### Project abstract:

With rapid development in science and technology, massive data are ubiquitous. The intrinsic high-dimensional/functional structure in massive data often requires more sophisticated statistical modeling and data processing tools such as deep neural networks. Meanwhile, due to limited data storage capacity, a large portion of the massive data was either ephemeral or temporarily cached and subsequently overwritten with newer data. Limited data storage capacity, as well as the growing degree of complexity in statistical models, severely challenge standard nonparametric inferential theory and applications. A long-term goal of the PI's research is to promote advanced nonparametric methods to overcome modern massive data challenges. The specific aim of this project is to explore nonparametric inferential procedures when data are quantized, and models are high-dimensional/functional involving complicated interaction effects. Statistical optimality of the procedures, in the presence of data and modeling challenges, forms the core of this proposal. Theoretical insights gained from analyzing the proposed algorithms are beneficial for real-world problems ranging from large scale data to functional data.

**Start date:** 06-01-2024

**End date:** 05-31-2027

**Last modified:** 09-01-2023

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

## CDSE: Quantization-Based Methods for Optimal Nonparametric Inference

### Roles and responsibilities

---

1. What parties and individuals will be involved with data management in this project?
2. What will be the roles and responsibilities of each party and or individual with respect to management of the data?
3. Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

**Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.**

The project's principal investigator, Zuofeng Shang, will ultimately be responsible for all of the data management.

Cristo Leon's role as compliance assisted with policies.

### Types of data produced

---

**What types of data, samples, collections, software, materials, etc. will be produced during your project?**

The project will not generate data, samples and physical collections. Computing software such as R packages for implementing the proposed algorithms shall be made available to the public through Github at the end of the project period. Certain components of the research outputs shall be used as curriculum materials.

**What will be the approximate number and size of data files that will be produced during your project?**

The project doesn't produce data.

**What type of metadata (information others might need to use your data) will be collected during your project?**

**Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.**

The project doesn't produce metadata.

## Data and metadata formats

---

**What format(s) will data and metadata be collected, processed, and stored in?**

**Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.**

The project doesn't produce metadata. The articles may have Dublin core metadata as per APA guidelines.

## Policies for access and sharing

---

**How will data be accessed and shared during the course of the project?**

The project doesn't produce data.

**Will any of the data and/or related materials produced need provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements? If so describe them and detail any requested exceptions from the archiving requirements set for Arctic Sciences research.**

The project doesn't produce data.

**When is the approximate release date of the data products?**

**Note: Arctic Observing Network (AON) data must be deposited in a long-lived and publicly accessible archive within 6 months of collection, and Arctic Social Science Program (ASSP) research data must be deposited in a long-lived and publicly accessible archive within 5 years of the award date assuming no exceptions to the archiving requirements are requested.**

The project doesn't produce data.

## Policies for re-use and re-distribution

---

**How do you anticipate the data for this project will be used? Consider the following:**

1. Which bodies/groups are likely to be interested in the data?

## 2. What and who are the intended or foreseeable uses/users of the data?

The project doesn't produce data.

Will any permission restrictions need to be placed on the data? Consider the following:

1. Who will be allowed to use the data?
2. How will others be allowed to use the data?
3. Will others be allowed to disseminate the data.

**Note:** If you are planning on restricting access, use, or dissemination of the data, you must explain in this section how you will codify and communicate these restrictions.

The project doesn't produce data.

## Plans for archiving and preservation

---

What is the long-term strategy for maintaining, curating, and archiving the data?

**Note:** The Office of Polar Programs policy requires that metadata files, full data sets, and derived data products be deposited in a long-lived and publicly accessible archive.

The project doesn't produce data.

---

## Planned Research Outputs

Data paper - "Optimal quantization methods for nonparametric inference"

---

### Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Optimal quantization methods for nonparametric inf ...	Data paper	Unspecified	Open	None specified		None specified	None specified	No	No