Introduction

The goal of this data rescue project is to design and implement a program that facilitates interpretation of Arctic data files that were deposited at the National Oceanographic Data Center (NODC; now NCEI) in the 1970's and 1980's.

These legacy data were collected by numerous scientists in the first large field effort to describe the oceanography, chemistry and biology of the American Arctic under the Outer Continental Shelf Environmental Assessment Program (OCSEAP).

The files were stored in a hierarchical database format which required extensive programming to create useful data records. Without strong support for the necessary programming, the majority of the data submitted by participants in the OCSEAP program disappeared into the NODC database and effectively became inaccessible.

Rescue of National Oceanographic Data Center (NODC) Files: The Benthic Infauna of the Beaufort Sea Coast and Shelf, 1975-1980

Susan Schonberg[1], Tim Whiteaker[2], Ken Dunton[1]
Marine Science Institute[1] & Center for Water and the Environment[2]
The University of Texas at Austin

ALASKA MARINE SCIENCE SYMPOSIUM

Methods

This project developed a protocol and supporting Python scripts to turn the NODC files into usable records that can be opened in Excel. The parsed NODC files enable the post-processing steps required to create useful data for an end user. This basic framework could be applied to multiple NODC data types posted on the NCEI web site.

This study uses infaunal invertebrate data (F030) collected along the nearshore Beaufort Sea as a test case to develop and check the project protocol and manipulations.

F030 Parsing Instructions and Python Module

Input File

Output Files

Keywords

NODC (now NCEI)
OCSEAP data rescue
Python programming
9 track reel data tapes
Beaufort Sea
Infauna

Results

Working with the NODC F030 infaunal data available on the NCEI web site was **challenging**.

1. Original station names were lost.
2. Format descriptions of record layouts and NODC taxonomic code evolved over time.
3. We discovered erroneous format descriptions, invalid code values, and numerous records with a taxon name but no data.
4. In one case, repeated data permanently overwrote another data type.
5. The parsed data showed numerous types of errors that required data cleanup decisions based on common sense, mathematics, and experience.

We created a table of infaunal records, by taxa, that includes collection location, abundance and biomass along with associated physical data. Managing the errors and data handling decisions assuredly introduced some variation to the originally submitted data.
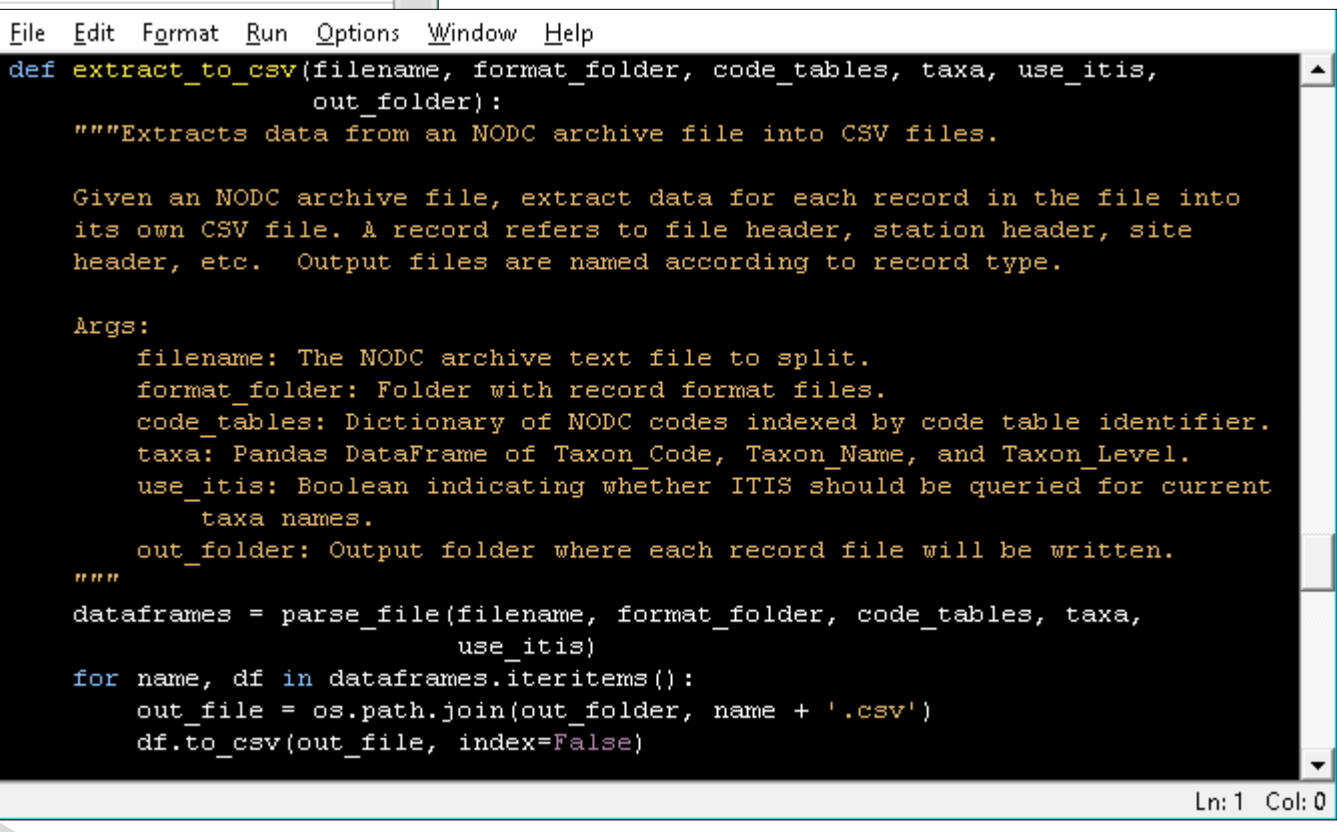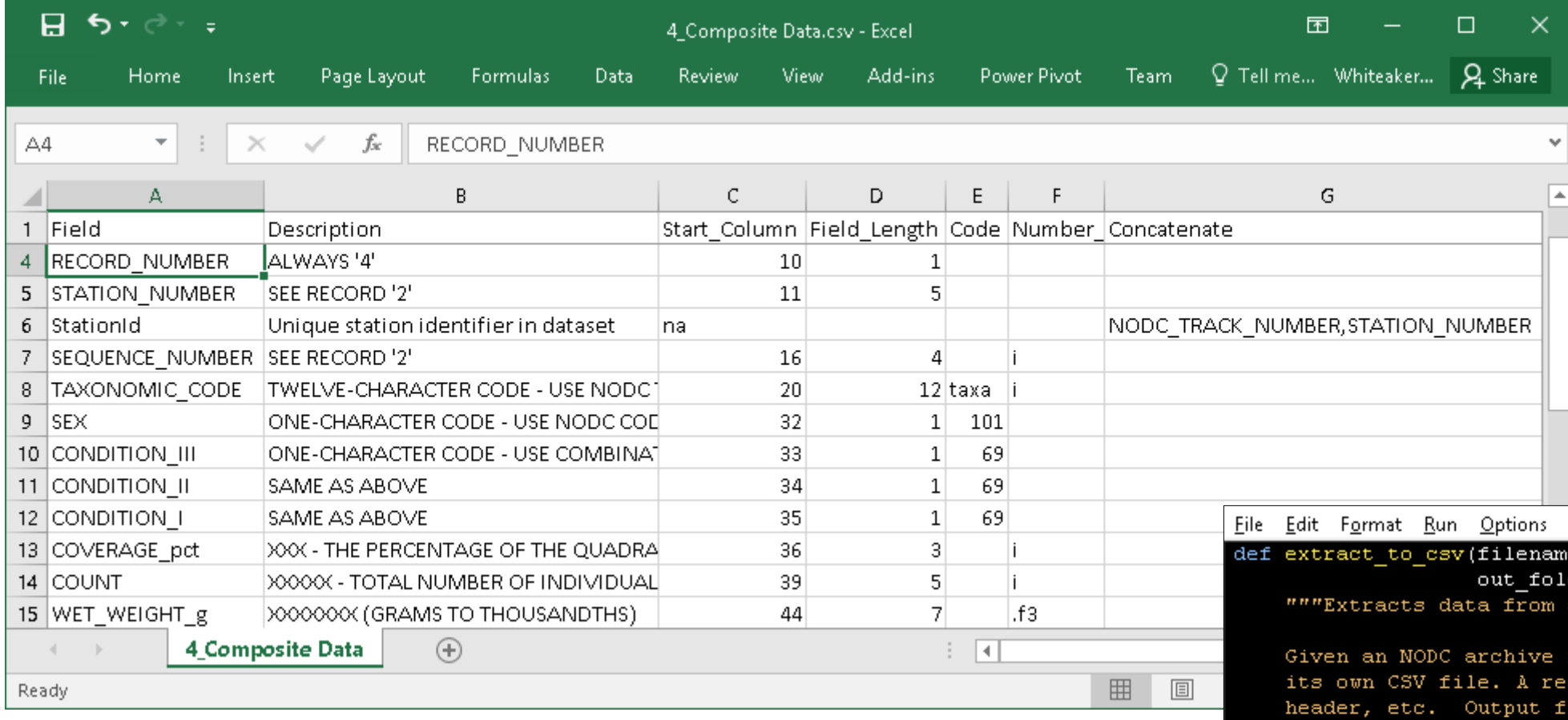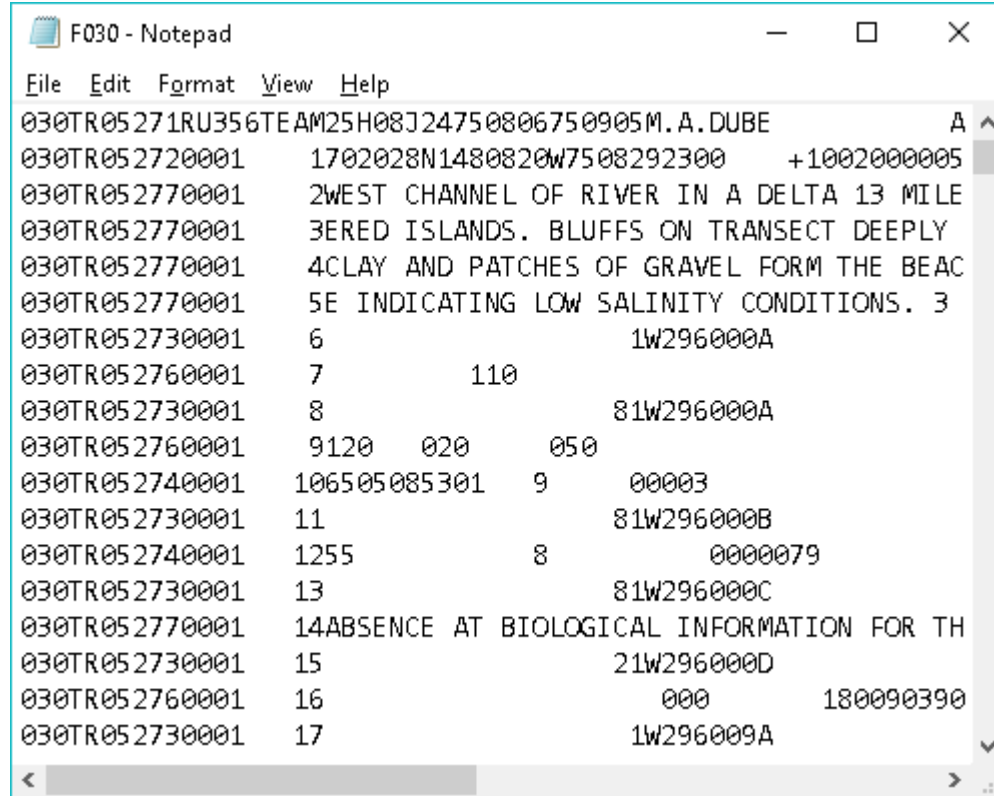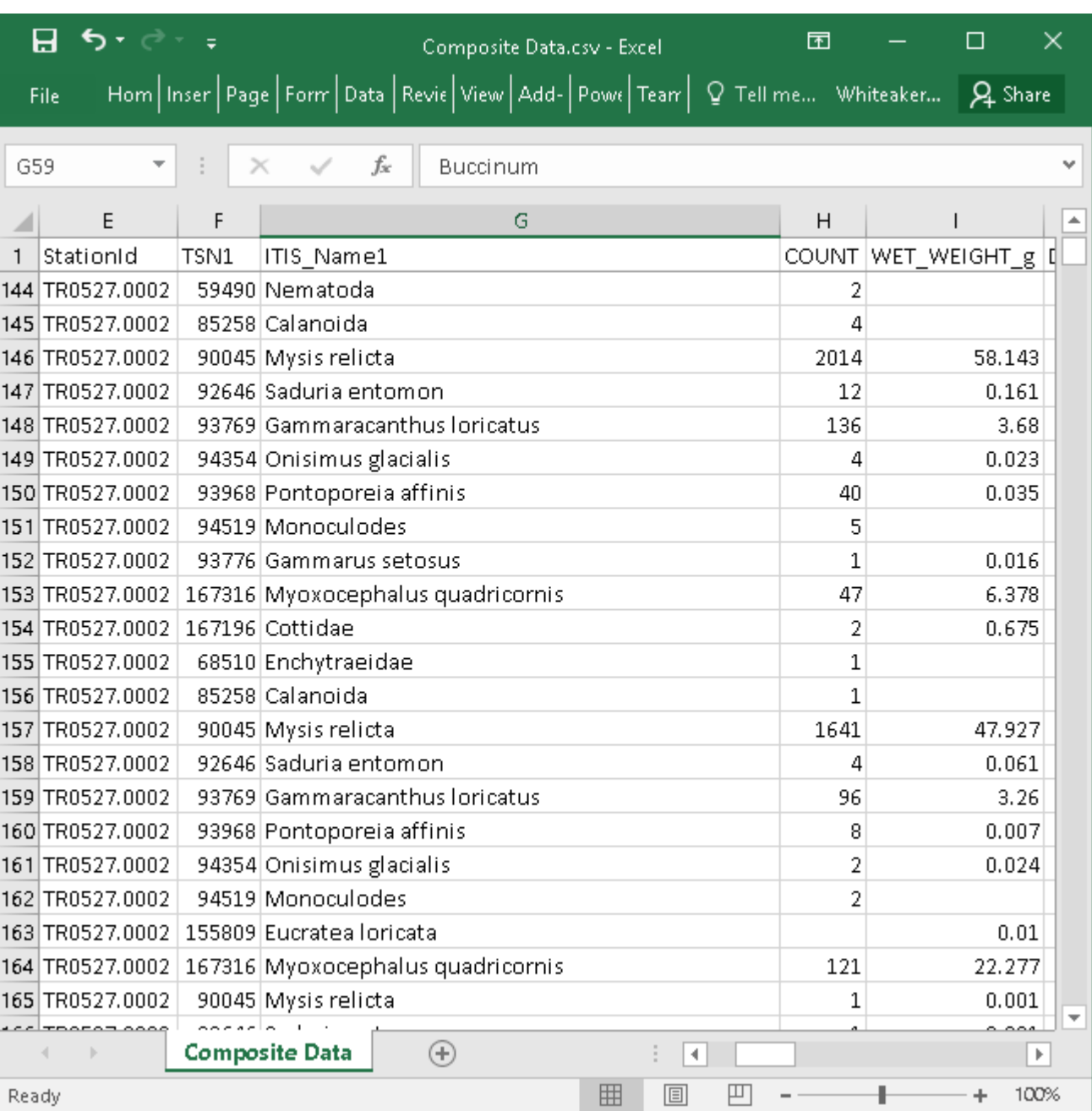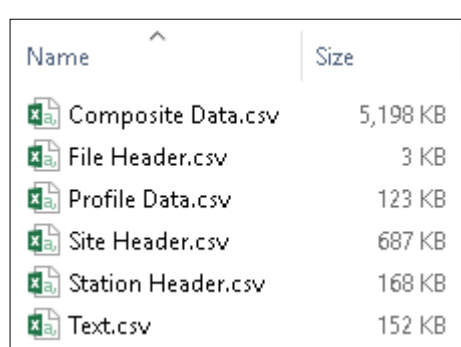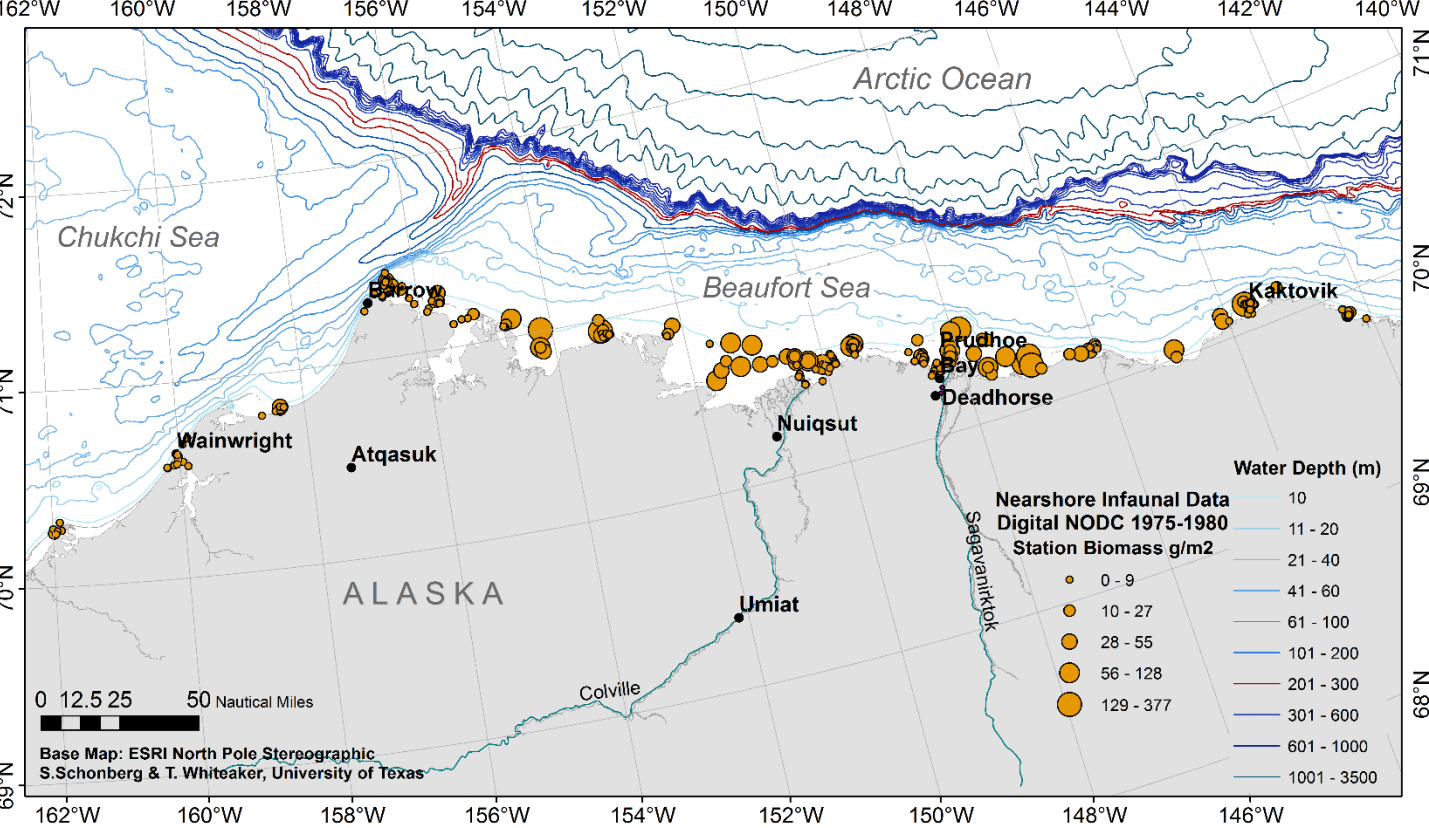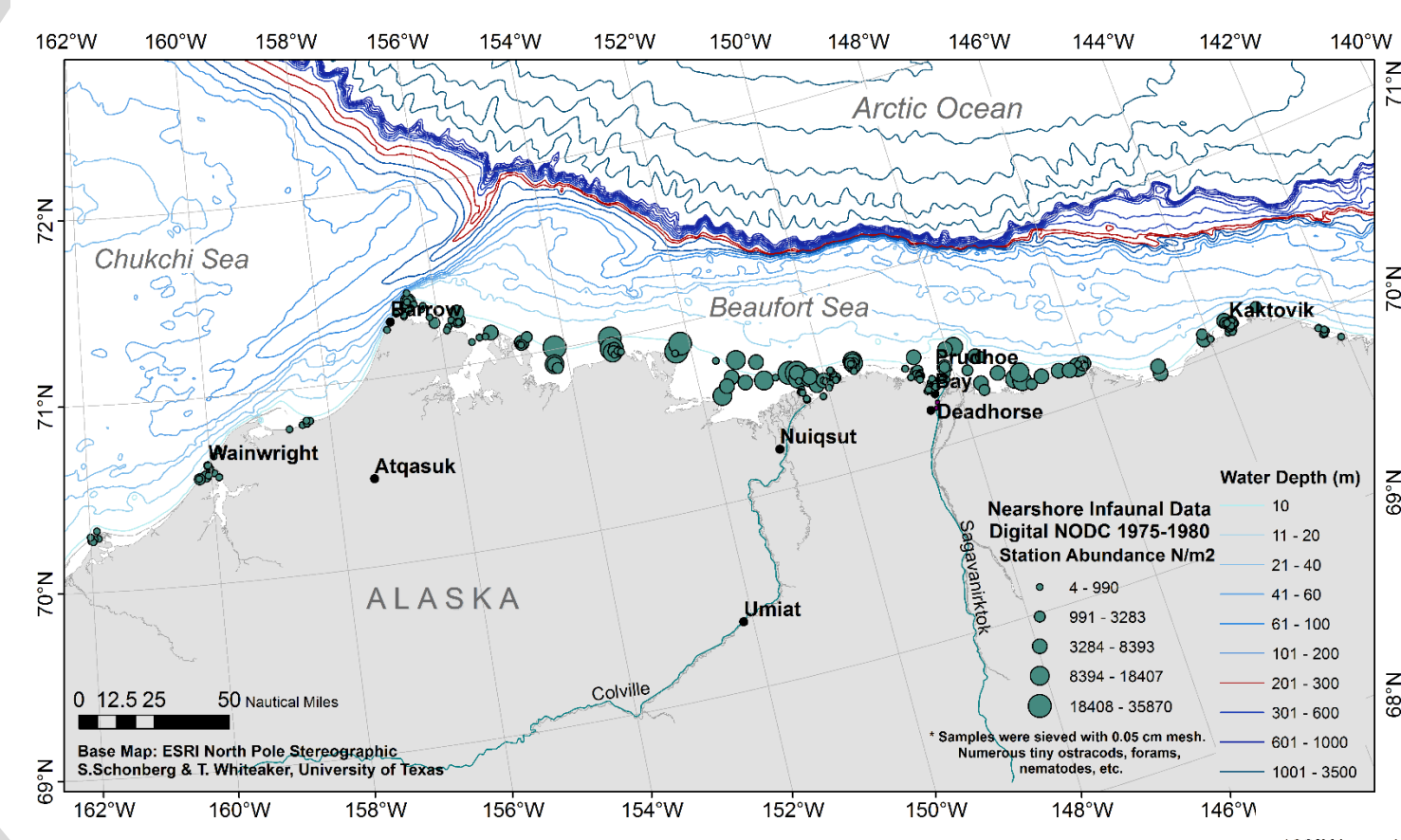
NODC Station Data
Station Abundance & Biomass

Conclusions

We have confidence that the available files were analyzed as correctly and carefully as possible using available information. We digitized station abundance and biomass data from original printed data sheets for accuracy assessment. The NODC station means fell within similar distribution ranges.

Next Steps

We do not have printed data by taxa for an accuracy assessment, but we do have the original **9 track magnetic reel data tapes** for the F030 infauna that contain these data. We are **searching for help to read these tapes** which would allow us to interpret the NODC files with greater confidence.