Data Scientist
Tasks for a candidate

bahdanchyk@faradaic.io
yakushenko@faradaic.io

February 11, 2025

Abstract

This document contains a set of exercises designed for Data Scientist candidates, reflecting routine daily tasks at FaradaIC. Please download them from this `repository`.

Please complete as many exercises as possible and submit your answers via email with a link to an accessible repository containing both the data and a Jupyter Notebook (or a script with comments) that can be cloned and executed within the same directory.

Based on your evaluation, you may be invited to the next stages of the interview process.

For inquiries or support, contact `bahdanchyk@faradaic.io`

Contents

1 Noise analysis

The Task 1 in `GitHub repository` contains data from noise measurements, which are part of routine data analysis performed in our laboratory to identify problematic sensors or equipment.

Noise measurements are recorded by applying a constant potential close to the rest potential and capturing current transients. These transients typically exhibit near-zero current behavior, as shown in Fig. 1.

The dataset includes measurements from eight channels - last number in the filename, where each channel corresponds to a physically separate sensor connected to a multiplexer and potentiostat. The noise measurement was repeated 15 times per each channel - identified as the second-to-last number in the file name The data consists of two columns: time (s) and recorded current (µA).

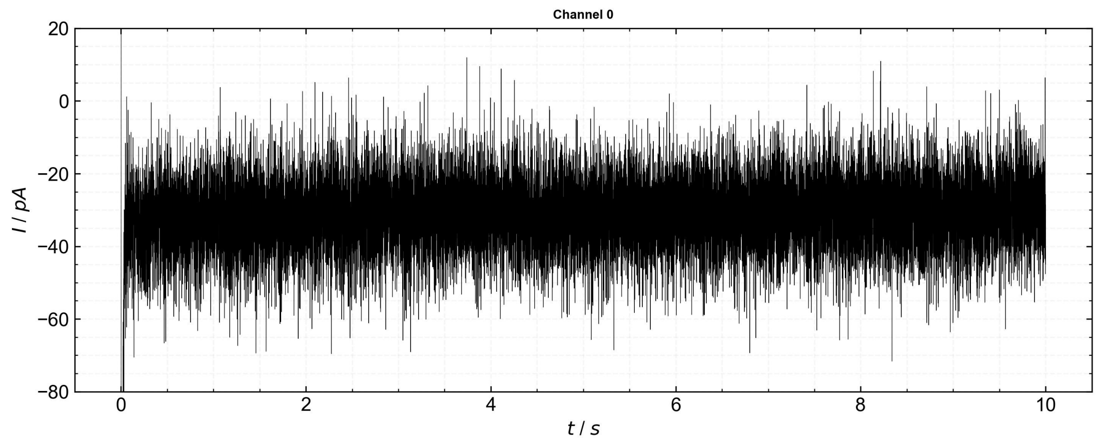

Fig. 1 – Example of a noise measurement

Tasks

1. Visualize the noise data for each channel.
2. Perform noise analysis and provide a brief description of your approach.
3. Identify, quantify and visualize any corrupted channels.

2 Sensor calibration

The Task 2 in `repository` contains data from a calibration test, which is essential for developing robust sensors. The primary objective is to establish a correlation between the measured current (or charge), humidity, temperature, and oxygen concentration.

Calibration measurements are recorded by applying a constant potential away from the rest potential and capturing current transients. These transients typically exhibit a dependence on gas concentration—in this case, oxygen.

The dataset includes measurements from four sensors (referred to as channels), recorded across a range of oxygen concentrations, relative humidities (h_{sht}), and temperatures (t_{sht}). This information is provided in the second row of the file header. The third row contains the set values, which may differ from the actual measured values.

The oxygen concentration (vol%) can be calculated from the MFC (Mass Flow Controller) data using the following equation:

$$\frac{\mathrm{MFC}_2}{\mathrm{MFC}_1 + \mathrm{MFC}_2} \times 20.9$$

The main dataset consists of four columns: time (s), voltage (V), recorded current (µA) and data type specification. You are primarily interested in rows labeled as 'Current', which contain the recorded current values. However, you may also explore 'CE potential', which represents the cell voltage as a function of time instead of current.

Please note that the technique used to record these curves consists of two polarization steps: negative and positive polarizations. We would also suggest to use either current in the end of transient or total charge as descriptive parameter.

Tasks

1. Analyze, interpret, and visualize the recorded signal as a function of oxygen concentration for each channel.
2. Determine which polarization step is most effective for sensing oxygen levels.
3. Identify and visualize any apparent dependencies in the data.
4. Detect and flag any corrupted sensors.
5. Attempt to calibrate the functional sensors and evaluate the oxygen level calibration quality using metrics such as MAE, RMSE, R^2, etc.

6. Considering the intentional addition of moisture to the supplied gas, please describe its potential impact on the absolute values of oxygen concentration, effect on calibration, and how it can be accounted for.

3 Cyclic Voltammetry

Task 3 in the `GitHub repository` contains data from a cyclic voltammetry (CV) test. This is a common and rapid test performed in our laboratory to evaluate the ability of a material or sensor to detect gases, such as oxygen. The test is conducted by linearly sweeping the electrode potential while recording the resulting current at the electrode. The potential sweep is then repeated in a triangular waveform. An example is shown in Fig. 2.

The dataset includes measurements from four channels. The description of the header fields is similar to that provided in Task 2, however, this dataset contains only three columns: potential (V) and measured current (µA). Rows labeled as 'CE Potential' can be ignored.

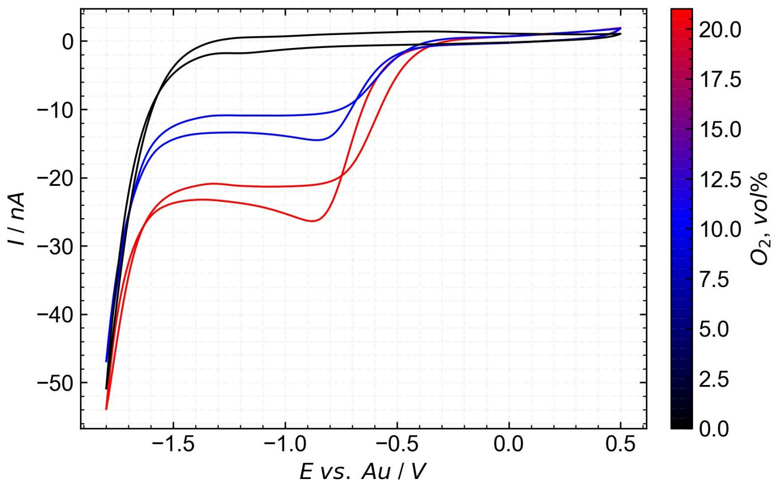

Fig. 2 – Example of cyclic voltammetry at different oxygen concentrations

Tasks

1. Visualize the CVs for each channel. Each file may contain multiple cycles. Propose a method to distinguish individual cycles, considering the full triangular potential sweep.

2. Propose a method to summarize the curves using descriptive metrics, such as the width of the plateau, the potential at which the current decays, or any other relevant characteristics of your choice.

3. Compare the channels based on the selected metrics and discuss the results.

Contact us

FaradaIC is at the forefront of innovation in gas sensing technology. Our miniature sensors are poised to revolutionize industries by providing compact, reliable, and cost-effective solutions. To learn more about our technology or explore partnership opportunities, visit our website at www.faradaic.io.

