# New Technologies homework assignment

September, 2018

Welcome to ClimaCell!

We enjoyed the phone conversation with you, and are excited to move forward to the next step: the homework assignment. The purpose of this assignment is twofold: not only will it help us assess your skill-set, but it will also help you understand the type of challenges we currently work on and ask yourself if these challenges are of the kind you would be interested in.

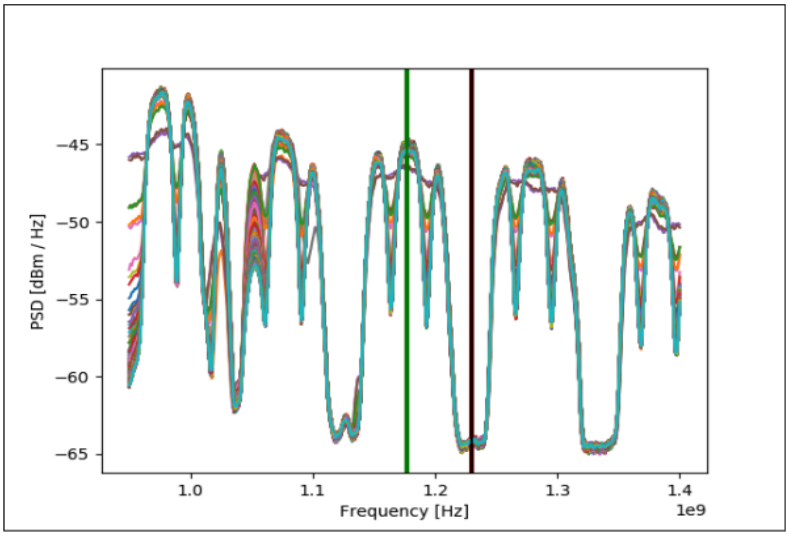
The challenge we deal with in ClimaCell are diverse in nature: on one day we could be dealing with a signal processing challenge and on another a machine learning data classification challenge. Another day we could be writing front-end software or other production code. We understand you can't do everything - no one here can - but we do expect you to be eager to learn. With a positive can-do approach and mentality (and a good chunk of Stack Overflow and Google) - no challenge will be left unsolved. **We are looking for team players.**

**Please do not spend more than a few hours on this assignment.**

**Part 1: A little bit of signal processing**. These are open-ended questions, so please write them in a neat word-processing doc or as the body of the email. Feel free to attach images and diagrams as necessary.

Questions 1 and 2 are warm-up questions. Question 3 is meant to explore how we can explain technical terms in concise ways to business people, or people who may be technical but aren't necessarily experienced with signal processing. Question 4 should provide a good sense some of our challenges.

1) What is the gain of a signal if the output voltage is 500V and input voltages is 50mV?
2) What is the gain in decibels if an amplifier has an input of 3mV and output of 5V?
3) Please briefly describe (in a few sentences and 1-2 diagram) what a QPSK modulation scheme is.
4) Consider the below graph:




The graph is taken from a spectrum analyzer that sits between a satellite dish LNB downlink and its connection to the set-top box receiver. It shows the power spectral density as a function of the frequency, and the colored lines represent measurements in different times. **Tell us the story of this graph, as detailed as possible:** What else can we tell about this graph? What do you think are the peaks in the different spectra? How can we measure the Signal-to-Noise ratio from this graph? Where in the spectrum can we do so?

**Part 2: Data Science**

**Data acquisition**: Here is the ID of a US "ASOS" weather station: CID (Cedar Rapids, Iowa). Retrieve all historical temperature data from that station (search around for how to access it) for the last 10 years.

**Forecasting Assignment**:

- Design a simple machine learning/statistics-based forecasting model using this data. (No physical/meteorological modeling.) The goal is to predict the temperature every hour for the day of 2018-06-01 (i.e. midnight to midnight in local time), trained only on data prior to that day.

A. Please use Python 3. (If you have a strong preference for another language, let us know and we can try to accommodate you.)
B. Feel free to use any packages available on PyPI/conda.
C. Please write good, organized code, with appropriate documentation.

- Some particular intermediate questions to answer: 2. Are there large data gaps? How big? How are you handling this? 3. Is the data on a regular time grid? If not, how are you handling this?
- Make a plot showing the results of your model, compared to the true temperature data for 2018-06-01.
- Produce an aggregate error metric to assess your model's performance -- how's it do?
- Please write a short blurb explaining your methodology/assumptions/etc. Also please discuss other methodologies that could be investigated, and how they might improve upon your results.

**Desired output:**

- A well-organized Jupyter notebook with all responses/plots requested in the previous section.
- All analysis should be fully reproducible -- we should be able to execute all cells in order to regenerate all your outputs. (All plots should be pregenerated, though.)
- Any data files you download should also be included. (Shouldn't be larger than a few MiB.)