Generation of Synthetic Multi-Resolution Time Series Load Data via Generative Adversarial Networks

Andrea Pinceti, Lalitha Sankar, and Oliver Kosut
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ, 85287

I. Introduction

Historically, power system research has been based around the development of simplified physical models and the study of their interactions within a system. However, in recent years, the field of machine learning has matured and improved to the point where it can provide real value to power system operations; for this reason, a large portion of the research work focuses on applying machine learning techniques to power system applications. This change in direction represents a major shift: moving away from physics-based device models to a data-centric analysis of system behaviors.

Within this new paradigm, the availability of large amounts of real data is crucial. Unfortunately, while power system models of all kinds are readily available, data is a much more scarce resource. The few researchers who have relationships with electric utilities can get access to real measurements through long processes involving non-disclosure agreements; in general though, the broader research community must rely on the very few and limited datasets that are publicly available.

The goal of our project is to develop an open source tool for the generation of synthetic time-series load data at varying sampling rates and for different time lengths. Leveraging a proprietary dataset of high resolution measurements from hundreds of phasor measurement units (PMUs) across many years of operation, we can model the behavior of real system loads and subsequently generate realistic-looking data on demand. The focus on load data is motivated by the fact that loads are one of the main drivers of power system behaviors and they represent a latent variable: loads depend on phenomena outside of the power system itself (consumer behaviors, weather, etc.). Thus, realistic load profiles can be used as an input to existing power system programs and, running dynamic simulations, electric quantities such as voltages and currents can be accurately determined.

II. Load Generation Scheme

Having access to a dataset of almost 100 TB of uncompressed PMU data allows us to observe load profiles at 30 samples per second for many consecutive years, thus capturing fast dynamic behaviors as well as long term seasonal patterns. By first decomposing and down-sampling the raw PMU load data at different levels, we can use advanced generative adversarial networks (GANs) to learn a generative model for each time horizon and resolution. GANs are a powerful machine learning algorithm in which a generator (usually a deep neural network) is trained to generate realistic data by making it "compete" against a discriminator whose job is to distinguish between real samples and those created by the generator.

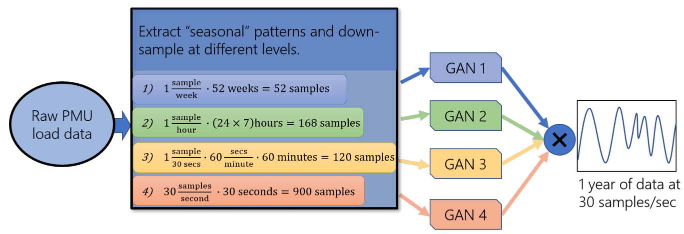

Fig. 1. Overview of the time-series load data generation scheme.

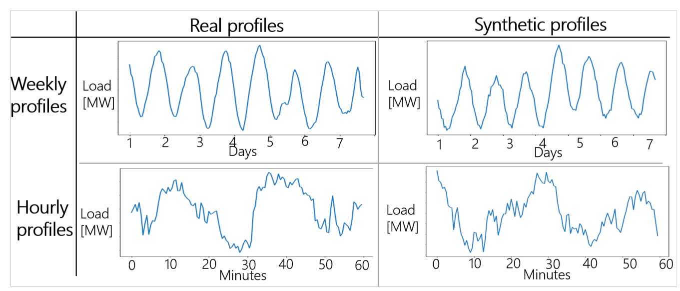

Fig. 2. Overview of the time-series load data generation scheme.

Figure 1 shows an overview of the generation scheme. The raw PMU load data is down-sampled at four different levels: 1) yearly profile, 2) weekly profile, 3) hourly profile, and 4) minute profile. A GAN model is trained individually for each level. The fully trained model can then be shared and used by researchers to generate any type of data required by their specific application. For example, for a dynamics study, two hours of data at 30 samples per second might be required; GANs 3 and 4 will be used to generate the hourly patterns and sub-second patterns, respectively. Figure 2 shows examples of real and generated load profiles at the weekly and hourly levels; the GANs learn to create completely new realistic-looking profiles.