Two Stage Ensemble Smoothing for Intraday Load Forecasts in the Presence of Distributed Solar PV Generation

Distinguished Webinar Series in Artificial Intelligence and Cyber Security April 10th, 2023

Dr. Frank Monforte | Director, Forecasting Solutions
Frank.monforte@itron.com

Abstract

» With growing penetration of distributed (e.g., rooftop) solar PV systems what power grid operators measure as demand for power is masked by offsetting unmetered solar PV generation. The inherent volatility of cloud movement is translating into volatile demand measurements.

» Forecast models that rely on real-time measurement of demand (aka, load) are experiencing increased forecast instability due to the inherent volatility of cloud movement.

» This presentation illustrates the challenges of forecasting loads under increasing levels of distributed solar PV generation. We will discuss:

- how loads are measured,
- the impact solar PV generation has on measured loads,
- how the industry forecasts solar PV generation, and
- why traditional load forecast models lead to forecast instability.

» We then introduce a framework for developing a stable sequence of load forecasts.

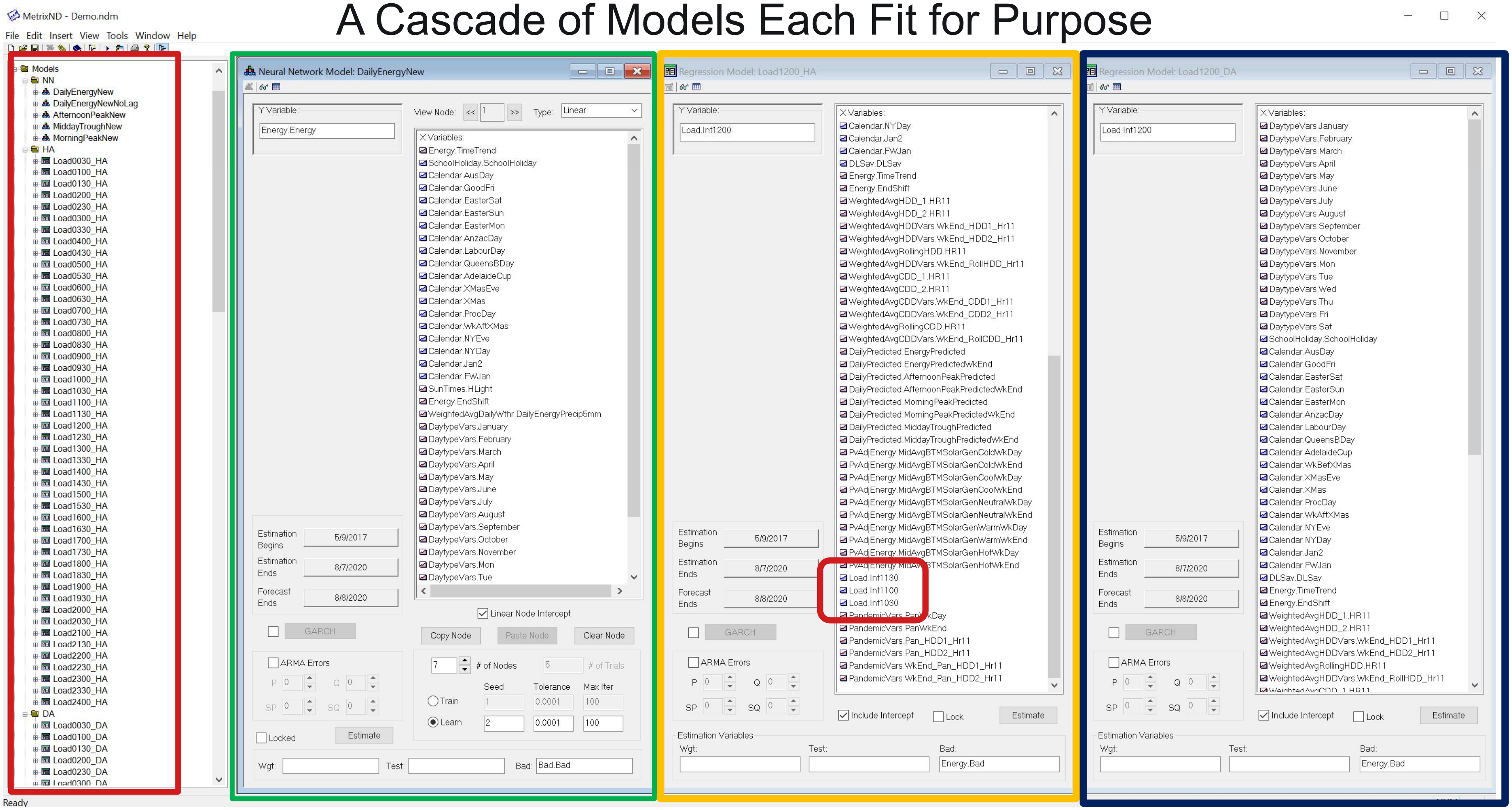
WHAT A LONG STRANGE TRIP IT'S BEEN….

EVOLUTION OF OPERATIONAL LOAD FORECASTING

1980s	1990s	2000s	2010s	2020s
Daily Temperatures Newspaper/Radio/TV	Hourly Temperatures Newspaper/Radio/TV Email/Download	Hourly Temperatures FTP Download	Hourly Temperatures Multiple Vendors BTM Solar PV FTP Download	Hourly Temperatures Multiple Vendors BTM Solar PV FTP Download
System	System	System, Transmission Zones	System, Transmission Zone	System, Transmission Zone, Substation, POD
Forecast Once a Day	Forecast Updated Hourly	Forecast Updated Sub-Hourly	Limited Forecast Ensemble, Updated Sub-Hourly	Complex Forecast Ensemble Updated Sub-Hourly
Expert Judgement Regression-Based Formulas	Expert Judgement Regression-Based Formulas Neural Nets	Expert Judgement Automated Forecasting Solutions	Expert Judgement Automated Forecasting Solutions	Expert Judgement Automated Forecasting Solutions Machine Learning
Limited Mainframe/ PC Computing	Proliferation of PC Computing	Internet FTP/SFTP	Cloud Computing Solar PV	Data Science Smart Grid

WHAT DOES AN OPERATIONAL FORECAST MODEL LOOK LIKE?

A Cascade of Models Each Fit for Purpose



KEY CHALLENGES

Our Role is to Help the Power Industry Overcome these Challenges

» COVID-19 impacts the mix of residential/non-residential HVAC equipment

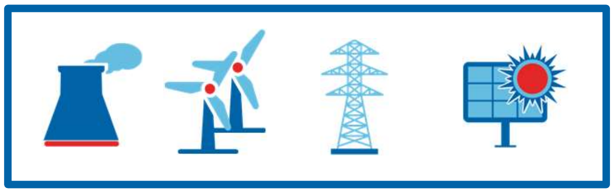
» Climate Change is manifesting as extreme weather events
- As more of these event occur utilization of HVAC equipment is evolving

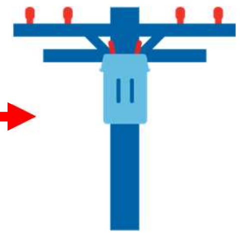
» Strategic Adoption of Grid-Connected Renewable Generation
- Recent focus and R&D is on forecasting grid-connected resources
- Limited focus on distributed generation forecasting

» Deep penetration of Distributed solar PV generation & EV charging push the technical limits of the Low Voltage Grid
- Creating a need for greater geospatial forecast detail

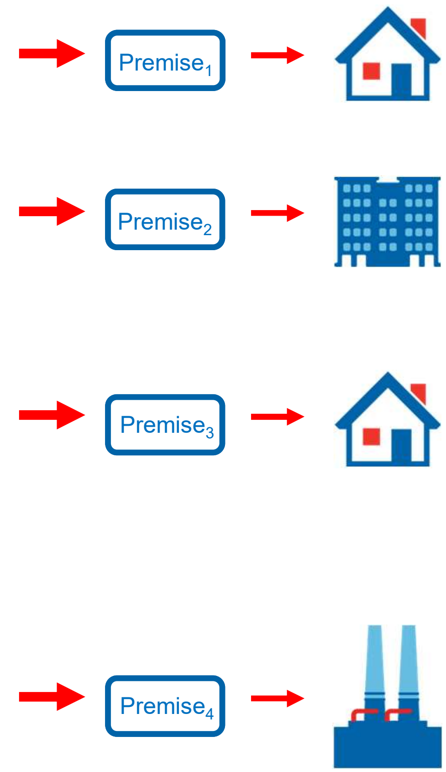
We used to Measured Power Consumption

SCADA Metered





Delivered



Before Distributed Energy Resources (DER)

- **Delivered Power = Consumption**

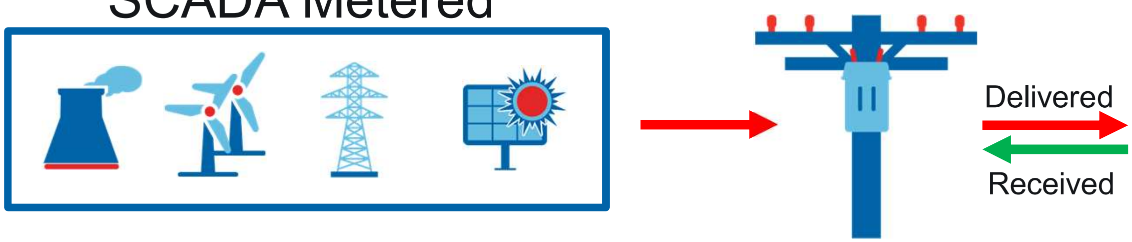
$$C_{d,i}^{SCADA} = L_{d,i}^{SCADA}$$

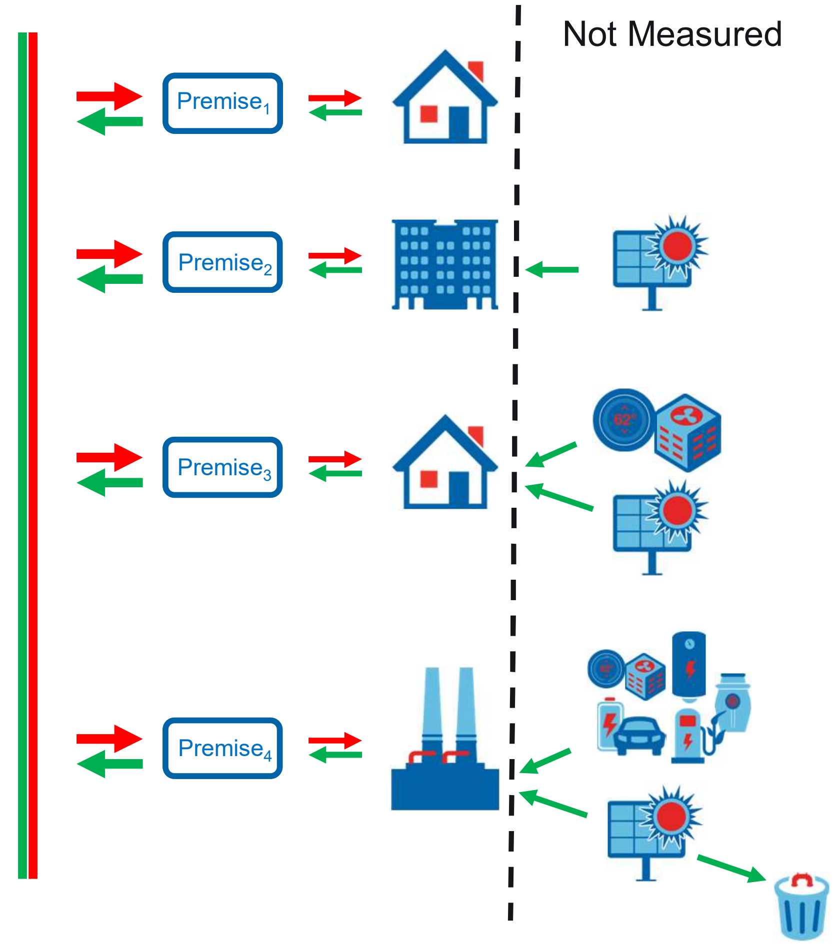
- Short-term forecasting uses SCADA measurements at the transformer / circuit etc. to describe consumption load shapes

$$L_{d,i}^{SCADA} = \left[\sum_{s=1}^{S} \mathrm{GridConnectedGeneration}_{d,i}^{s} - \sum_{n=1}^{N} \mathrm{NetTransmissionGridExchanges}_{d,i}^{n}\right] \times \left(1 - \mathrm{SystemLosses}_{d,i}\right)$$

Now we Measure Net Load Masking Consumption

SCADA Metered





With DER: Where Most Utilities are Today or in the Very Near Future

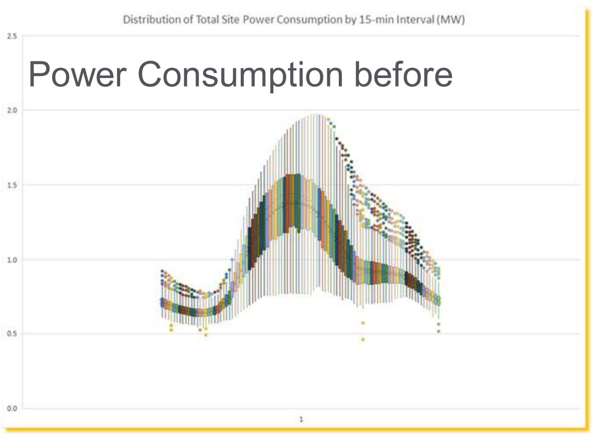
» Delivered Power ≠ Consumption

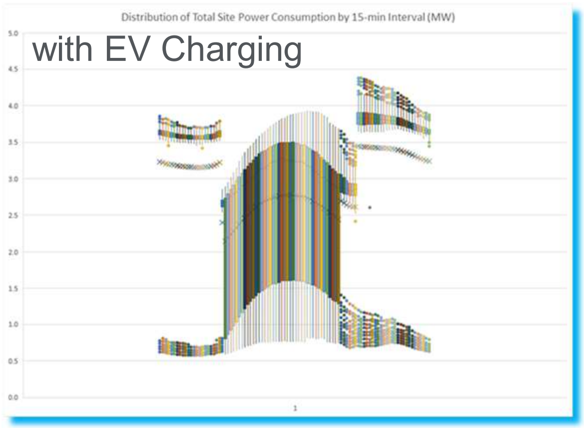
$$\boxed{C^{SCADA}_{d,i} = L^{SCADA}_{d,i} = \emptyset\left(C_{d,i}, PV_{d,i}\right)}$$

» Short-term forecasting uses SCADA measurements at the transformer / circuit etc. to describe the Behind-the-Meter Energy Imbalance

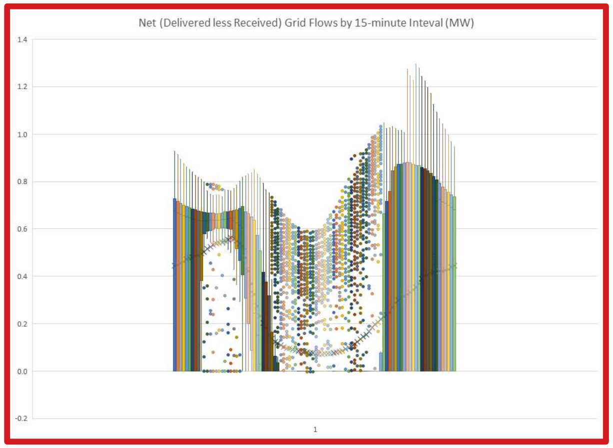
$$L^{SCADA}_{d,i} = \left[\sum_{s=1}^{S} \mathrm{GridConnectedGeneration}^{s}_{d,i} - \sum_{n=1}^{N} \mathrm{NetTransmissionGridExchanges}^{n}_{d,i}\right] \times \left(1 - \mathrm{SystemLosses}_{d,i}\right)$$

The Operational Forecasting Problem is Evolving

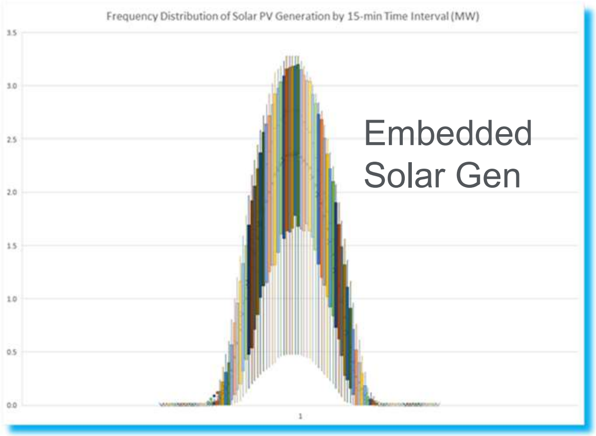


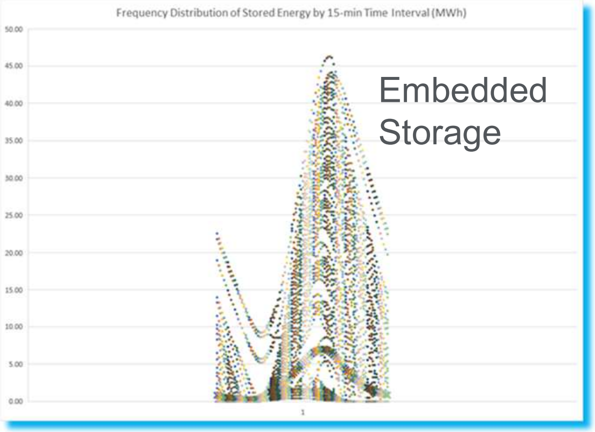


- Before we forecast Consumption that was well understood. Accuracy depends on weather forecast performance.
- Spinning reserves at minimum levels given certainty of demand.

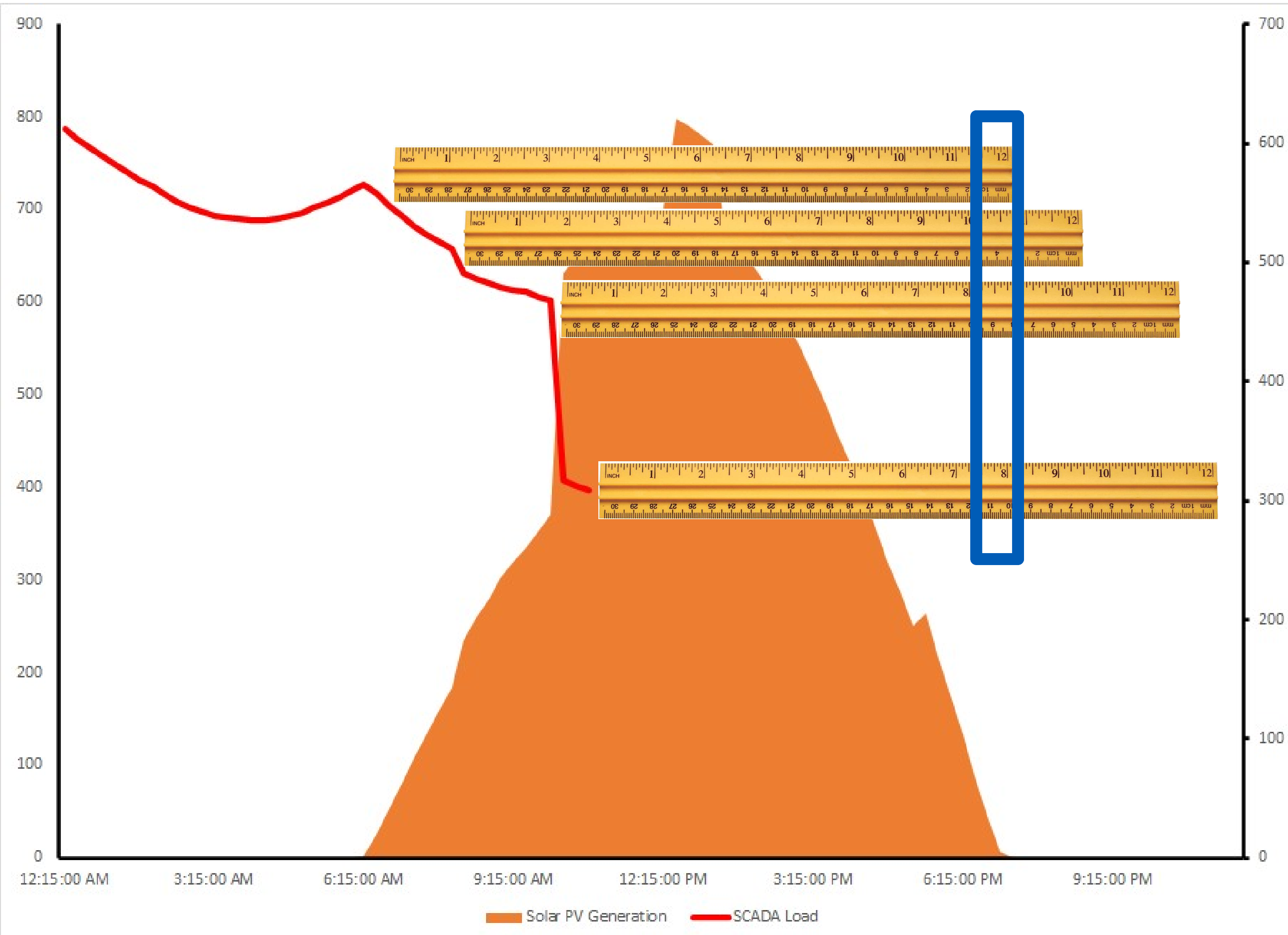


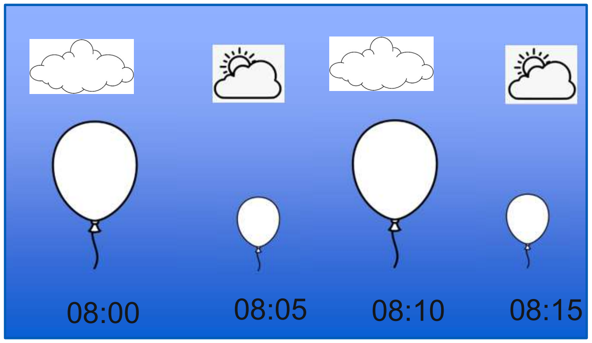
- Now forecast energy imbalances leading to forecast instability
- Higher spinning reserves to cover the uncertainty resulting in higher system operating costs in order of magnitude of millions dollars.





Load Masking Leads to Measurement & Forecast Instability



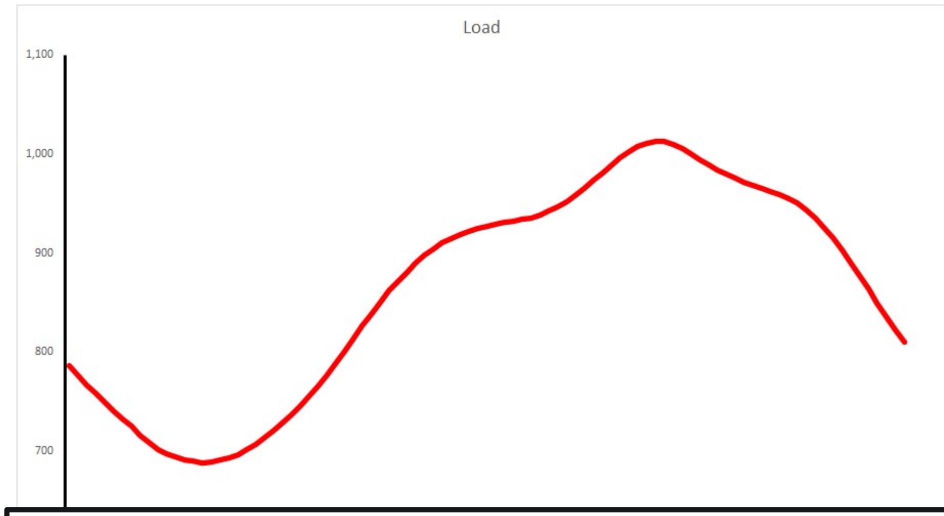


As solar PV generation cuts in and out, measurements of Net Load bounce around.

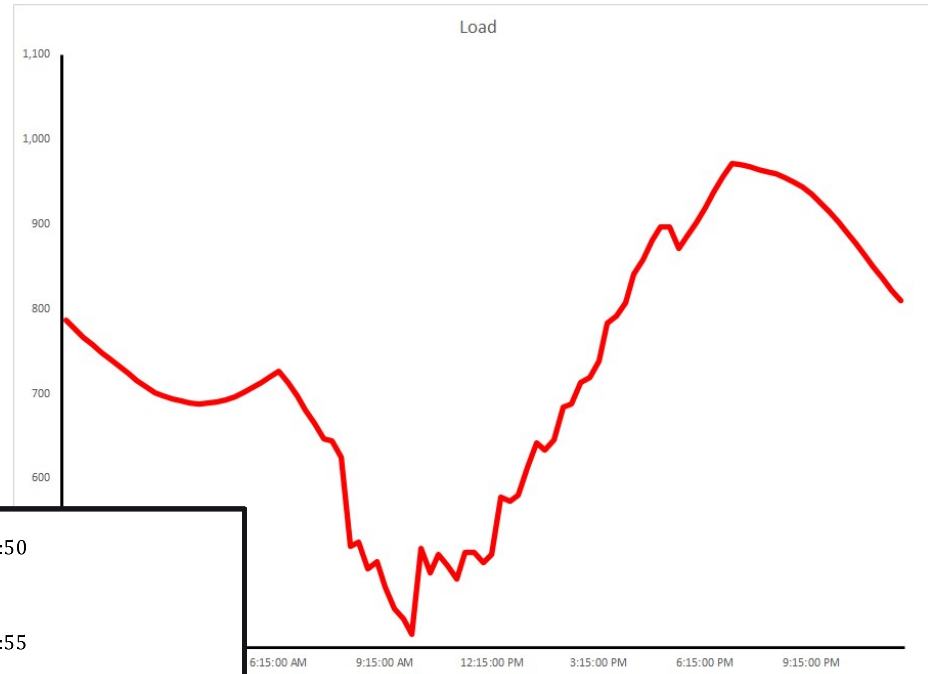
Autoregressive models that leverage lagged Net Loads risk casting that load volatility into the forecast period.

Autoregressive Terms and Forecast Instability

Autoregressive Terms Work Well



Autoregressive Terms Break Down



$$L_{10:50} = \beta_0^{10:50} + \beta_1^{10:50} L_{10:45} + \beta_2^{10:50} L_{10:40} + \beta_3^{10:50} L_{10:35} + e_{10:50}$$

$$L_{10:55} = \beta_0^{10:55} + \beta_1^{10:55} L_{10:50} + \beta_2^{10:55} L_{10:45} + \beta_3^{10:55} L_{10:40} + e_{10:55}$$

$$L_{11:00} = \beta_0^{11:00} + \beta_1^{11:00} L_{10:55} + \beta_2^{11:00} L_{10:50} + \beta_3^{11:00} L_{10:45} + e_{11:00}$$

What to do with Solar PV Generation?

$$L_{d,i}^{SCADA} = F(X_{d,i}\beta_i\ ) + G(SolaPVGen_{d,i}\alpha_i\ ) + L(L_{d,i-j}^{SCADA}\delta_{i-j}) + e_{d,i}$$

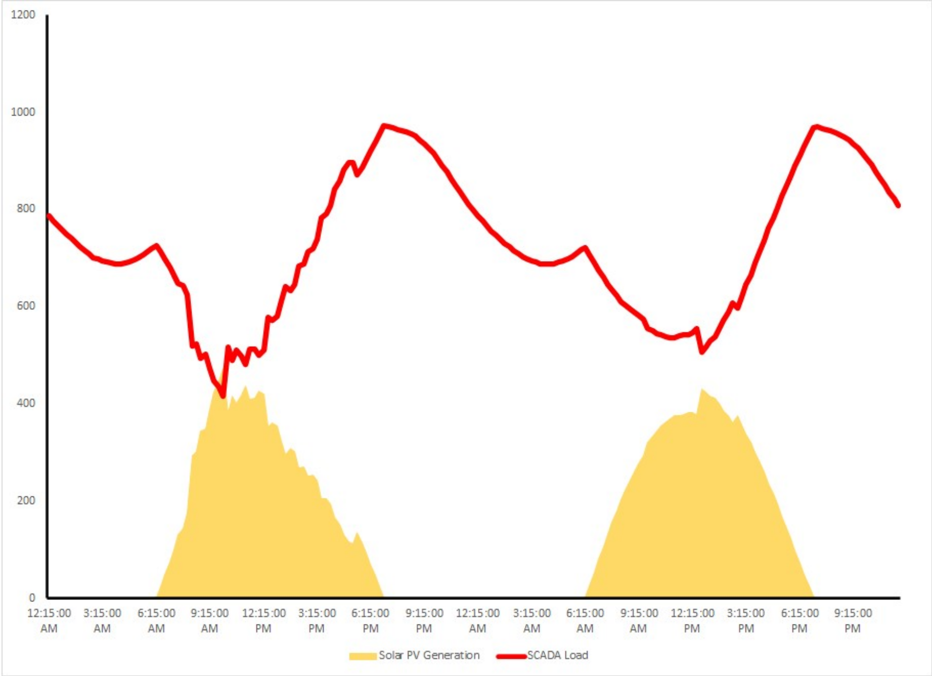
Direct Modelling Approach

» Distributed solar PV generation is not metered
  - Engineering-based estimates driven by GHI

» Direct Modelling provides statistically-adjusted solar PV generation values

» To make it work the autoregressive terms need solar PV generation interactions to free up the slopes

» Getting the specification right is THE challenge

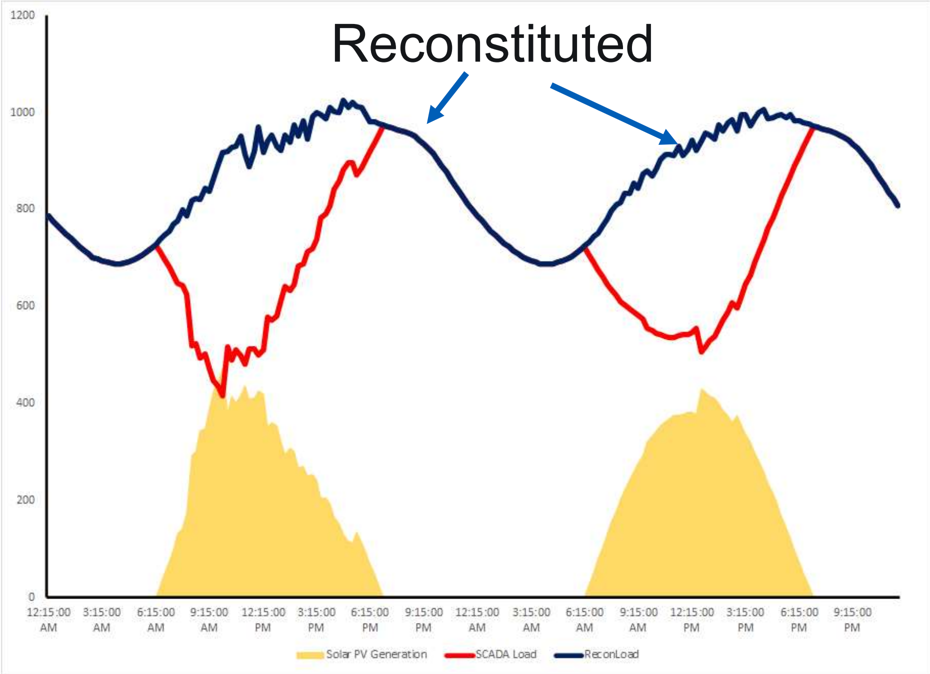


What to do with Solar PV Generation?

$$L_{d,i}^{SCADA} + SolarPVGen_{d,i} = F(X_{d,i}\beta_i) + L([L_{d,i}^{SCADA} + SolarPVGen_{d,i}]\alpha_i) + e_{d,i}$$

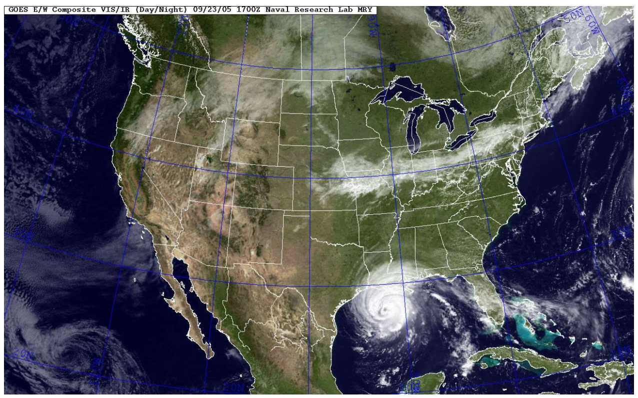
Reconstituted Load Approach

- Assumes solar PV generation estimates are correct & the impact is 1.0 KW of solar PV generation lowers loads by 1.0 KW
- The autoregressive process is relatively stable with Reconstituted loads
- Getting the specification right is THE challenge

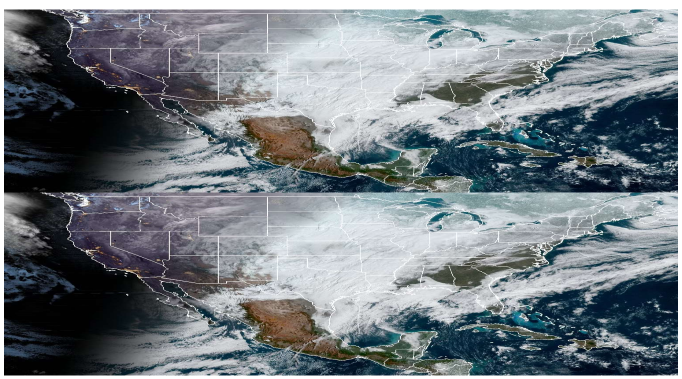


Evolving from Direct Modelling to Reconstituted Loads

Dealing with Uncertain Solar PV generation estimates



Easy Days To Predict GHI



Hard Days To Predict GHI

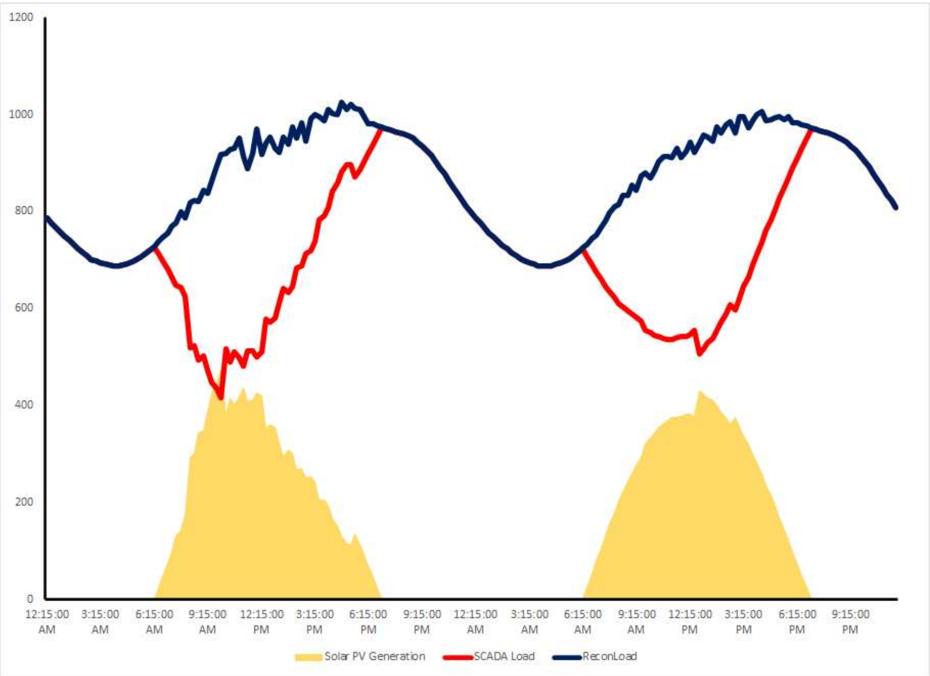
» There are two major approaches to GHI estimation and forecasting

- **Numerical Weather Prediction Models** use mathematical models to predict cloud cover movement. These forecasts are then translated to forecasts of GHI which drive solar PV generation estimates.
  - BEST for Forecast Horizons of 4 Hours +
- **Satellite Image Decomposition** provide estimates of cloud cover over 1km x 1km squares. Mathematical models then infer the GHI values.
  - BEST for Forecast Horizons up to 4 Hours

Evolving from Direct Modelling to Reconstituted Loads

Dealing with Uncertain Solar PV generation estimates

$\mathrm{VAR}\left(\mathrm{L}_{d,i}^{\mathrm{SCADA}} + \mathrm{Solar}\widehat{\mathrm{PV}}\mathrm{Gen}_{d,i}\right)$ is Greater on Cloudy Days than Clear Sky or Dark Sky Days



- » Few utilities collect in real-time the population of solar PV generation. As a result, we reconstitute with an estimate of rooftop solar PV generation.
  - On cloudy to partially cloudy days solar PV generation estimates are at their highest levels adding volatility to the reconstituted loads.
  - E.g., Net Load measurement goes down but estimated Solar PV generation goes down compounding the swing in Reconstituted loads
  - Defeats the purpose of using Reconstituted Loads

Two Stage Ensemble Smoothing

- How can we save the Reconstituted Approach?

- Observation: Aggregate changes in Consumption oscillate at a slower frequency than Solar PV Generation
  - The field of Signal Processing suggests a range of smoothing algorithms that will filter out unwanted high frequency oscillations of “noisy solar PV generation estimates” leaving a relative smooth reconstituted load series
- But …
  - Wide smoothing windows while cutting through the noise of solar PV generation risk *smoothing* through key turning points in underlying consumption of power
  - Narrow smoothing windows maintain key changes in consumption, but also the volatility of solar PV generation

- How do we balance removing the noise from the solar PV while maintaining key features of consumption?

Savitzky-Golay Smoothing Filters

Centered Moving Average using Polynomial Weights

$$\widehat{\text{ReconLoad}}^{J}_{d,i} = \left(\sum_{j=0}^{(J-1)/2} \omega^{J}_{-j}\,\text{ReconLoad}_{d,i-j} + \sum_{j=1}^{(J-2)/2} \omega^{J}_{j}\,\text{ReconLoad}_{d,i+j}\right) \Bigg/ \left(\sum_{j=0}^{(J-1)/2} \omega^{J}_{-j} + \sum_{j=1}^{J/2} \omega^{J}_{j}\right)$$

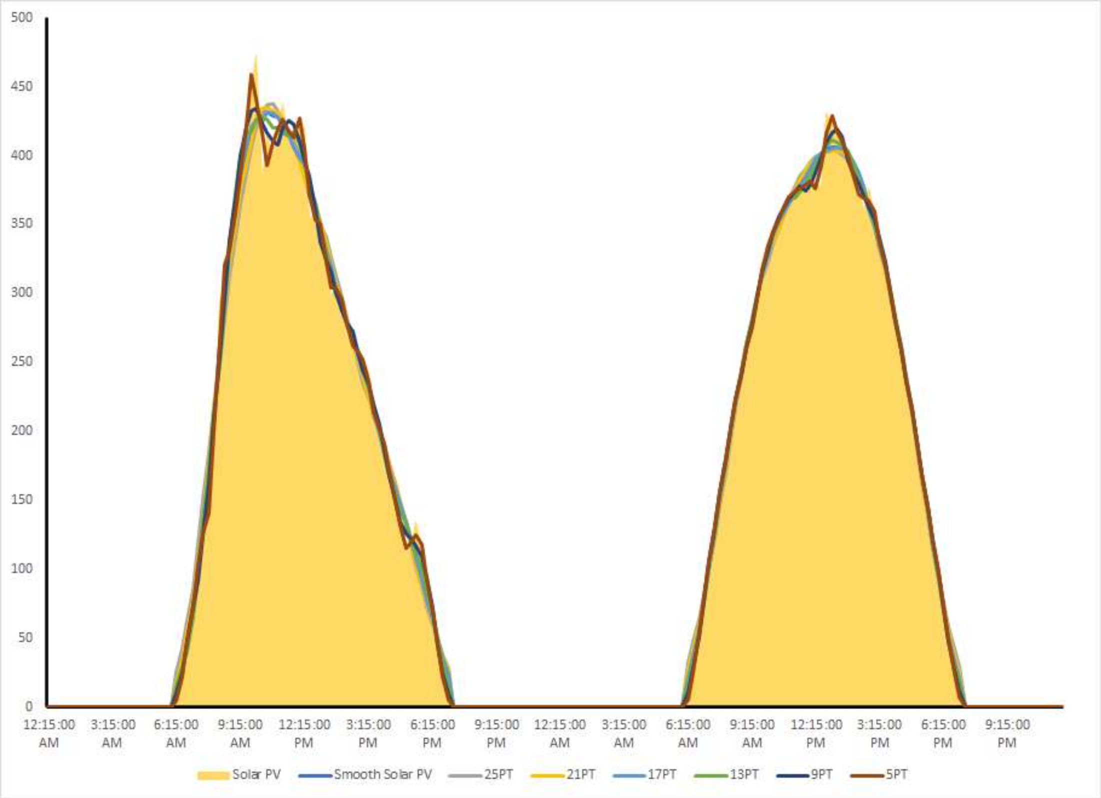
- » SG Smoothing is useful for load forecasting because the polynomial weights preserve the curvature of the load data.
  - A straight centered-moving average would produce a relatively flat result.
- » But which Smoothing Window Should be Used to Smooth Reconstituted Load?

Normalized Savistky-Golay Smoothing Weights

Smoothing Window Size	25	21	17	13	9	5
Lag 12	-253					
Lag 11	-138					
Lag 10	-33	-171				
Lag 9	62	-76				
Lag 8	147	9	-21			
Lag 7	222	84	-6			
Lag 6	287	149	7	-11		
Lag 5	322	204	18	0		
Lag 4	387	249	27	9	-21	
Lag 3	422	284	34	16	14	
Lag 2	447	309	39	21	39	-3
Lag 1	462	324	42	24	54	12
Center	467	329	43	25	59	17
Lead 1	462	324	42	24	54	12
Lead 2	447	309	39	21	39	-3
Lead 3	422	284	34	16	14	
Lead 4	387	249	27	9	-21	
Lead 5	322	204	18	0		
Lead 6	287	149	7	-11		
Lead 7	222	84	-6			
Lead 8	147	9	-21			
Lead 9	62	-76				
Lead 10	-33	-171				
Lead 11	-138					
Lead 12	-253					
Normalization Constant	5135	3059	323	143	231	35

Step One. Smooth the Solar PV Estimates/Forecasts

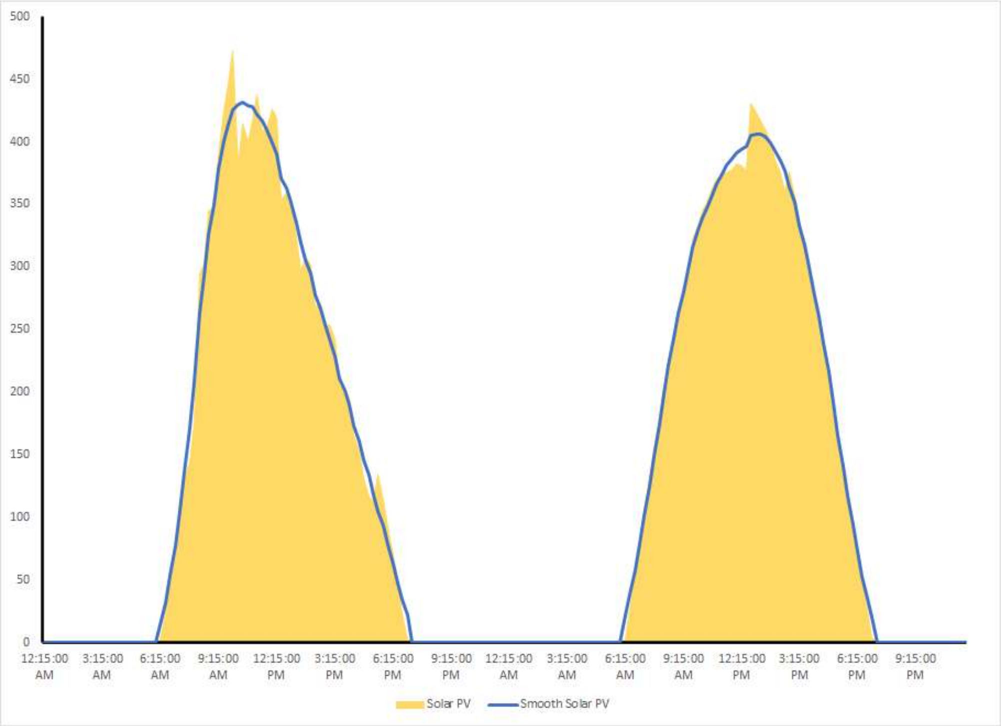
- Apply an ensemble of smoothers (e.g., different SG Smoothing Windows) to the raw solar PV estimates
- Create a weighted average solar PV estimate by weighting the alternative smoothed solar PV estimates
- Each smoother is assigned a weight which depends on the volatility of the raw solar PV generation data



Step One. Smooth the Solar PV Estimates/Forecasts

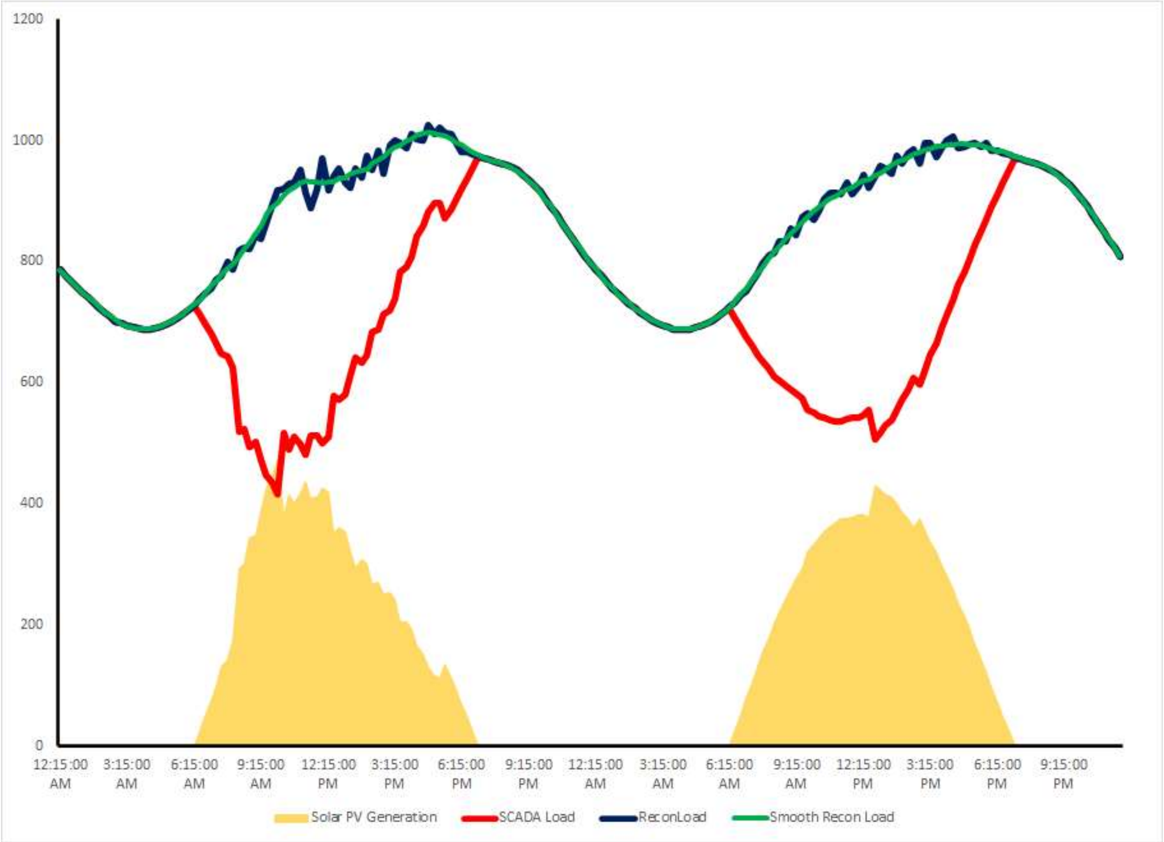
$$\varphi_d^j = \sum_{i=1}^{I} \nabla^2\left(\widehat{PV}_{d,i}^j\right) \Big/ \sum_{i=1}^{I} \nabla^2\left(ClearSky_{d,i}\right)$$

- The second order derivatives of the smoothed solar PV generation are used to form the weights
- The second order derivatives of a clear sky day are used as a normalization factor
- Narrow windows are preferred on Clear Sky days
- Wider windows are preferred on Partially Cloudy days

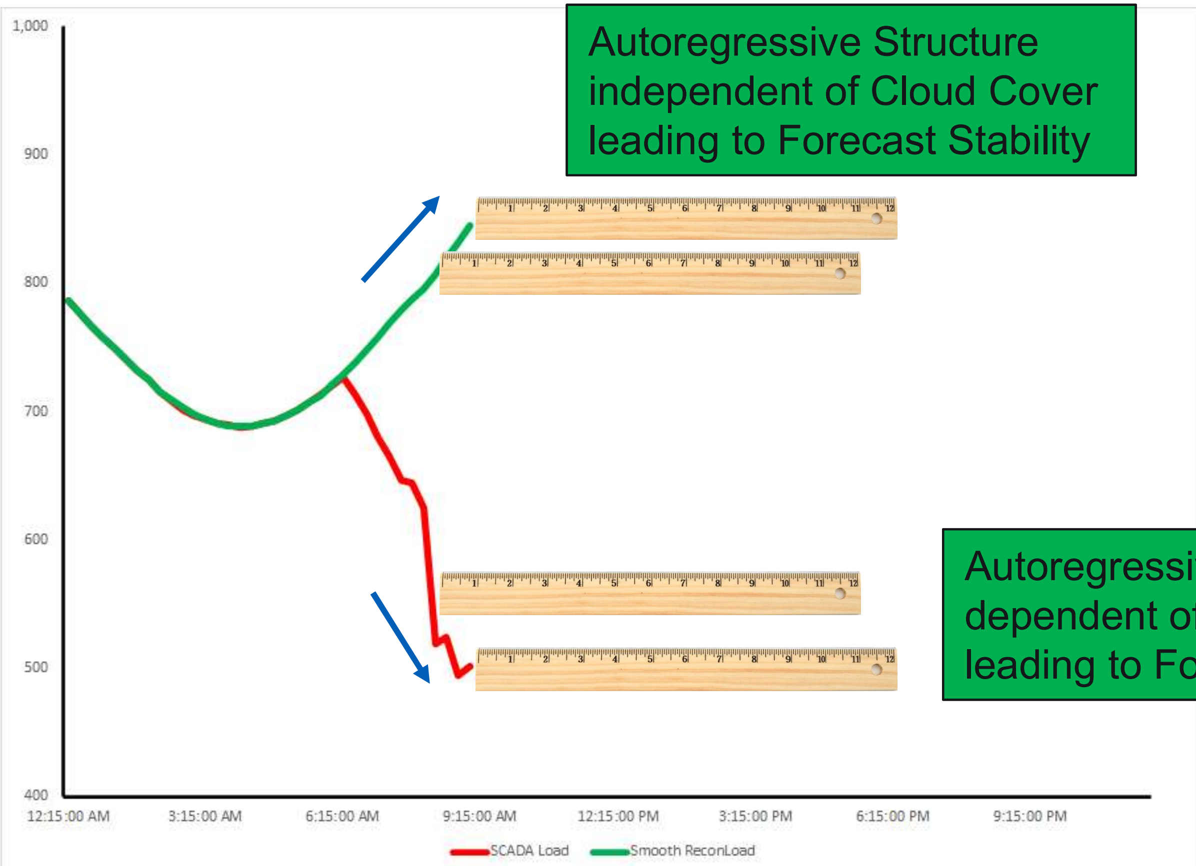


Step Two. Smooth the Reconstituted Load Data

- The same ensemble of smoothers are applied to the raw reconstituted load
- In this step, the smoothing weights from Step 1 are applied to create a weighted average reconstituted load series
- In effect, we use the volatility of the solar PV data to drive the size of the smoothing window for the reconstituted load time series

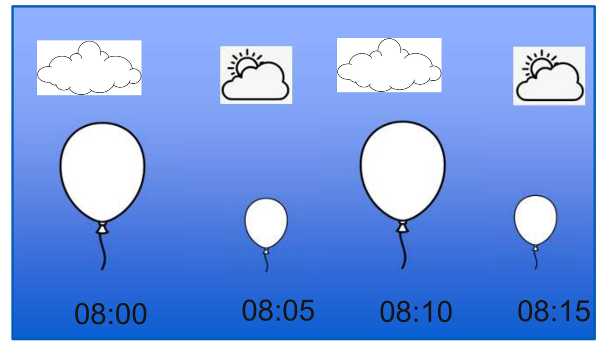


A Step Toward Forecast Stability



Autoregressive Structure independent of Cloud Cover leading to Forecast Stability

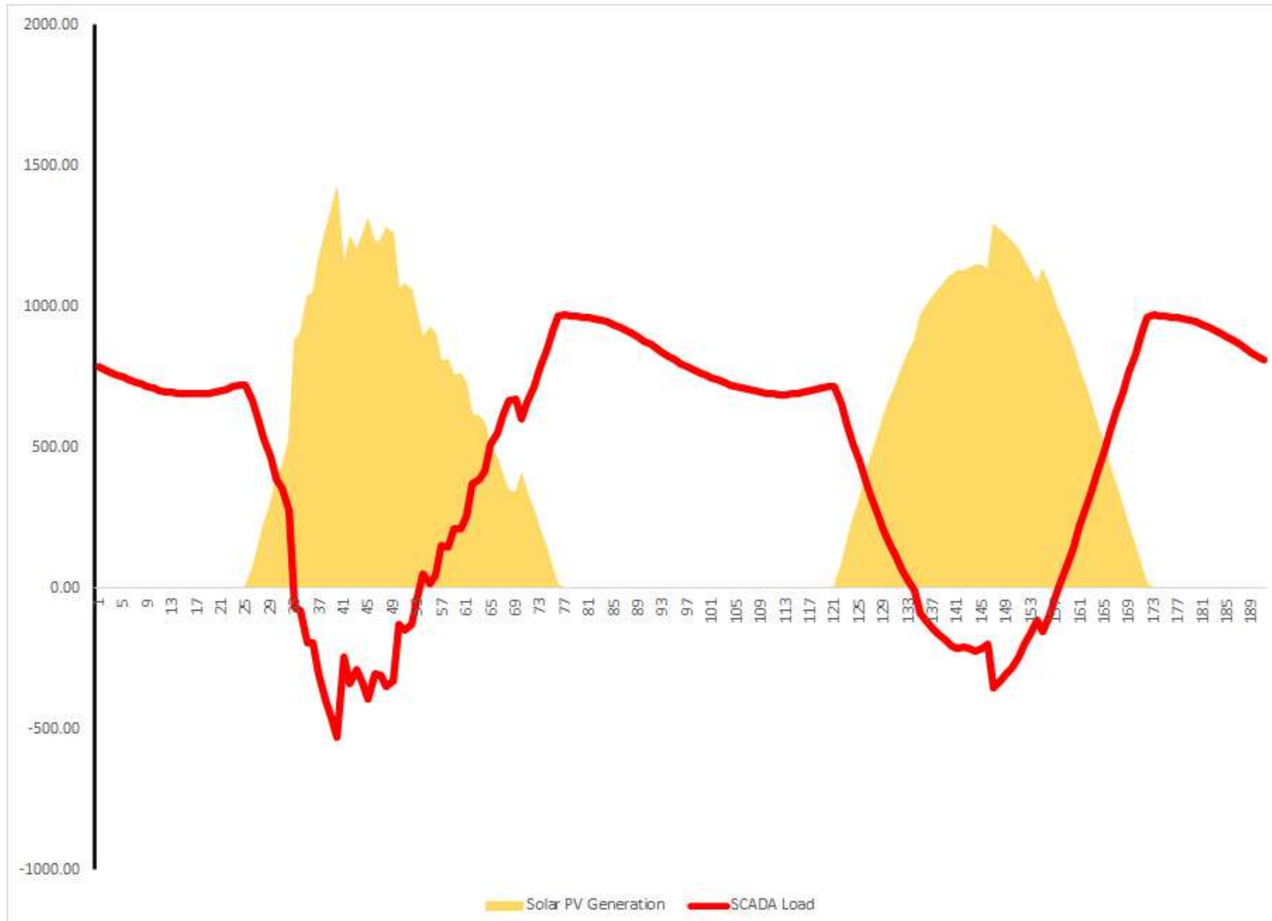
Autoregressive Structure dependent of Cloud Cover leading to Forecast Instability



A SPRING DAY IN SOUTH AUSTRALIA – A REALITY COMING SOON TO YOUR NEIGHBORHOOD

» Operational forecasting is growing in complexity as we transition to 100% renewable generation

» Solving the operational forecasting problem is a small, but critical piece to ensuring a quick & successful transition



SOME THINGS TO CONSIDER

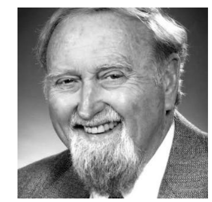
» I used to work on  and now on 

You can only imagine the evolution of tools you will experience in your careers

» Despite massive computing power & the promised miracle of machine learning the most important tools in my arsenal are EXCEL, the Internet, and a thirst to learn new mathematical approaches to solve problems.
- Excel to code up an algorithm or technique to visualize & understand how it works
- The Internet because most of what we do is based on the brilliance of those who came before us
- Learn from other disciplines
  - Constrained and Unconstrained Optimization, Signal Processing, Econometrics, Decision Trees, etc.
  - The more tools you have the better carpenter you will be

» Finally, what makes an accurate forecast is not the machine learning technique, but rather the hard part of correctly specifying the set of explanatory variables (features) included in the model.
- Like John Henry, I have yet to find a “machine” learning algorithm that beats a well specified model built with 99% hard thinking and 1% inspiration.
- It is easy to let computers do stuff. It is hard to know if the stuff was the right stuff.
  - Beware - Correlation is not Causation

Prof. George Dantzig



Prof. Clive Granger

John Henry

THANK YOU
I HOPE YOU ENJOYED THE TRIP….