Forecast Practitioner's Handbook: Incorporating the Impact of Embedded Solar Generation into a Short-term Load-Forecasting Model

Dr. Frank A. Monforte,
Director of Forecasting Solutions
frank.monforte@itron.com

WHITE PAPER

Contents

Introduction

More than 102 GW of installed solar generation capacity exists worldwide, according to the European Photovoltaic Industry Association (EPIA),[1] and that number is rising. The EPIA estimates that approximately 31 GW of solar generation capacity was installed in 2012 which was about 15% higher than the estimated 30.4 GW installed in 2011. In the EPIA report, Europe remains the world leader in installed solar capacity, accounting for approximately 70 GW of the 2012 total installed capacity of 102.2 GW worldwide. Together, China (8.3 GW), USA (7.8 GW), Japan (6.9 GW), Australia (2.4 GW) and India (1.2 GW) account for approximately 26% of the world solar capacity. The Solar Energy Industries Association estimates that by the end of 2013 there will be a new solar project installed in the USA every four minutes.[2]

These worldwide statistics represent both utility solar installations, where the electricity generated feeds directly to the grid, and non-utility installations (referred to elsewhere as embedded solar generation), where generation offsets on-site consumption. From the perspective of load forecasting, the non-utility installations are of critical interest since these installations directly affect measured load. Since short-term load-forecast models are based on measured load, the following examples illustrate how embedded solar generation can affect a load forecast. In these examples, assume the demand for electricity at noon, is 1,300 KW, regardless of its source.

No Embedded Solar Generation. Without embedded solar generation, metered demand, or the load that a system operator sees, equals actual demand.

$$\text{Metered Demand}_{\text{d}}^{\text{Noon}} = \text{Demand}_{\text{d}}^{\text{Noon}}$$

Now consider developing a forecasting model of demand for electricity. If we have a year's worth of measured demand, we could fit the following regression model:

$$\text{Metered Demand}_{\text{d}}^{\text{Noon}} = \beta_1 \text{ Constant}_{\text{d}}^{\text{Noon}} + \text{e}_{\text{d}}^{\text{Noon}}$$

Here, Metered Demand regresses on a variable that takes on a value of 1.0 for every observation. In this case of no embedded solar generation, the estimated coefficient on the Constant variable will be equal to the average metered demand, or 1,300 KW. As a result, the forecast from the estimated model will provide a forecast of actual demand for electricity.

With Constant Embedded Solar Generation. Now assume that 100 KW of embedded solar generation is produced every day at noon. We can rewrite Metered Demand as follows:

$$\text{Metered Demand}_{\text{d}}^{\text{Noon}} = \text{Demand}_{\text{d}}^{\text{Noon}} - \text{SolarGeneration}_{\text{d}}^{\text{Noon}}$$

Because metered demand will be 100 KW lower, the estimated coefficient from regressing the new, lower-metered demand on the Constant variable will lead to a different estimated coefficient of 1,200 KW, the new, lower-average metered demand. In this case, the resulting forecast model will under-predict actual demand for electricity by 100 KW.

From the perspective of system operations, the fact that the forecast model under-predicts actual demand for electricity is not a concern, since operators can rely on the 100 KW of solar generation's being there all the time.

With Volatile Embedded Solar Generation. In reality, solar generation is not this reliable. We can introduce uncertainty into the amount of solar generation available by assuming that, half the time, cloud cover is thick enough to drive solar generation to 0. The other days are perfectly clear, and the solar generation is 100 KW. This means that half the time the noon load is 1,200 KW, and half the time it is 1,300 KW. If the cloudy and sunny days are equal in number, the estimated coefficient will equal the average load, or 1,250 KW.

---

1 Global Market Outlook for Photovoltaics 2013-2017, European Photovoltaic Industry Association

2 www.seia.org

The variability in solar generation means that the statistical model that was fitted to metered demand will under-predict loads on cloudy days and over-predict loads on sunny days. System operations will need spinning reserves available to cover the load variability and subsequent load forecast error introduced by the volatile embedded solar generation.

Accounting for Average Solar Generation. ***Is it possible to improve the accuracy of the load forecast?*** Let's assume we can obtain a perfect forecast of cloud cover, and can accurately predict how much solar generation will be available tomorrow. It seems reasonable to adjust the baseline-load forecast with the solar-generation forecast, as follows:

$$\text{Demand}_{\text{d}}^{\text{Noon}} = \text{Predicted Metered Demand}_{\text{d}}^{\text{Noon}} + \text{PredictedSolarGeneration}_{\text{d}}^{\text{Noon}}$$

On a sunny day, the forecast of demand will equal the predicted value of 1,250 KW from the model of metered demand plus 100 KW of solar generation, or 1,350 KW. On a cloudy day, the forecast of demand will equal the predicted value of 1,250 KW from the model of metered demand plus 0 KW of solar generation. Unfortunately, both forecasts of actual demand are in error. On a sunny day, this approach over-predicts actual demand by 50 KW, which equals the average amount of solar generation that took place over the time period used to estimate the coefficient of the model of metered demand. On a cloudy day, this approach under-predicts by 50 KW, which, again, is the average amount of solar generation that took place over the time period used to estimate the coefficient of the model of metered demand.

In the current example, the average solar generation over the estimation period was 50 KW. As a result, the estimated coefficient of the metered-demand model embodies this average. Since 50 KW is already accounted for by the metered-demand model, we need to add the difference between the predicted solar generation for the day in question and the average solar generation already accounted for by the model coefficient. This results in the following calculation:

$$\text{Demand}_{\text{d}}^{\text{Noon}} = \text{Predicted Metered Demand}_{\text{d}}^{\text{Noon}} + \text{SG}_{\text{d}}^{\text{Noon}} + \left(\text{AvgSG}^{\text{Noon}} - \text{SG}_{\text{d}}^{\text{Noon}}\right)$$

In the above equation, SG represents the actual level of solar generation on day (d) at noon. AvgSG is the average solar generation over the model-estimation period at noon. The third part in the equation corrects for how much of the current day solar generation is already embedded in the model coefficients.

Using the current example, where AvgSG equals 50 KW, let's consider the two scenarios of a sunny day when SG = 100 KW and a cloudy day when SG = 0 KW. The calculations are:

Sunny Day

$$Demand_d^{Noon} = 1{,}250 + 100 + (50 - 100) = 1{,}300$$

Cloudy Day

$$Demand_d^{Noon} = 1{,}250 + 0 + (50 - 0) = 1{,}300$$

These examples illustrate the potential for additional forecast error arising from embedded solar generation. It is important to recognize that the coefficients of the short-term forecast model embody the average impact of solar generation on loads. This means that, in areas where there has been significant penetration of embedded solar generation, the short-term forecast will tend to under-forecast loads on cloudy days and over-forecast loads on sunny days. Ignoring the problem is not an option. There are two practical approaches to dealing with the impacts of embedded solar generation. In the first approach, an *ex post* adjustment is made to the short-term model forecast. To do so, determine the difference between the average solar generation over the estimation period and the current day's activity. Since the goal is to forecast metered demand, not actual demand, the *ex post* adjustment has the following general form:

$$\text{AdjustedForecast}_{\text{d}}^{\text{Noon}} = \text{PredictedMeteredDemand}_{\text{d}}^{\text{Noon}} + \left(\text{AvgSG}^{\text{Noon}} - \text{SG}_{\text{d}}^{\text{Noon}}\right)$$

Using the example data from above, the adjusted forecast of metered demand on a sunny day versus that on a cloudy day is computed as follows.

Sunny Day

$$AdjustedForecast_d^{Noon} = 1{,}250 + (50 - 100) = 1{,}200$$

Cloudy Day

$$AdjustedForecast_d^{Noon} = 1{,}250 + (50 - 0) = 1{,}300$$

Under the second approach, explanatory variable(s) that account for embedded solar generation are included directly into the short-term load forecast model. Developing a strong forecast model of metered demand that incorporates embedded solar generation is challenging. The focus of this paper is to present practical steps to incorporate embedded solar generation into an existing short-term load forecast model. The discussion begins with a high-level description of the mathematical relationships that determine the amount of solar energy striking the Earth's surface. These calculations are then combined with estimates of installed solar generation to form an *a priori* estimate of solar generation output for any location and time of year. The paper then discusses how to incorporate the *a priori* estimates into a statistical load forecasting model.

Solar Panel Basics

There is an abundance of information on how solar generation works. Summarized below is the basic information needed to develop a solar-generation-forecast model.

The most commonly used solar-generation technology for homes and businesses is solar photovoltaic (PV) panels. PV cells convert light (photons) to electricity (voltage). The first practical PV cell was developed in 1954 when scientists at Bell Telephone Laboratories noticed that silicon created an electric charge when exposed to sunlight. When light strikes a PV cell, a certain portion of the light is absorbed within the silicon material. The absorbed energy knocks electrons loose, allowing them to flow freely. The "freed" electrons form a current. The amount of electricity produced by a solar panel depends on the size of the panel, the panel's efficiency, and the amount of solar energy that reaches the panel surface.

- **Panel size** is typically measured in units of watts/m^2 of peak output. Peak output is the amount of electricity the panel produces when it receives the maximum amount of sunlight possible and at optimal ambient temperatures (13°C or 55°F). Typically, a solar panel comprises 40 or so solar cells that, in total, generate approximately 150 watts/m^2 of peak-electricity output. From the perspective of generation forecasting, panel size is treated as a known, exogenous forecast driver to the forecast framework. For example, the size of a solar plant is given by the total peak MW output of the plant. Embedded solar is measured typically in total MWs of installed rooftop panels for a given geographic area.
- **Panel efficiency** measures the percentage of solar energy hitting the panel that is converted into electricity. Solar panel efficiencies for a typical residential or commercial application range between 10% and 20%. For example, let's say the solar energy reaching the panel surface is 1,000 watts/m^2. A panel with a maximum output rating of 150 watts/m^2 would have an efficiency of 15% (computed as 150 watts-out/m^2 over 1,000 watts-in/m^2).

Efficiency and panel-size ratings are under ideal conditions. Factors affecting a solar panel's actual electricity output include (a) panel orientation and tilt, (b) panel temperature and (c) shade.

- Panel orientation and tilt are important factors affecting how much direct sunlight reaches the panel surface. For installations in the northern hemisphere, the ideal orientation is south-facing, while in the southern hemisphere, the ideal orientation is north-facing. Further, under ideal circumstances the panel will tilt throughout the day to track the path of the Sun through the sky. Because the costs of solar-tracking systems are high, they tend to be only found in large commercial installations and solar plants. This implies that most residential installations do not operate under peak conditions.

- Solar panels are most efficient at temperatures around 13°C or 55°F. The hotter the temperature, the less efficiently a panel converts sun energy into electricity. For temperatures above 25°C (77°F), the efficiency of a rooftop solar panel will degrade 0.48% per degree Celsius (0.27% per degree Fahrenheit).
- Shade lowers the output of a panel by blocking the amount of solar energy reaching the panel's surface. Trees, clouds, nearby structures, and even dirt or snow buildup can effectively block the amount of solar energy reaching the panels. Clouds reduce solar-panel output by reducing the amount of solar energy striking the panel. At 100% cloud cover, roughly 80% of solar energy reflects back out to space and 20% filters through to Earth's surface.

Unlike solar plants where the panel orientation, tilt, and shading (other than cloud cover) are known, the specifics of each rooftop installation are unknown. In this case, an average operating efficiency of 15% should account for all likely combinations of panel orientation, tilt, and shade that a collection of residential and commercial installations would have. From the perspective of forecasting embedded solar generation, the impact of temperature and cloud cover will be treated as separate factors influencing solar-panel output on an hour-by-hour basis.

How Much Solar Energy Reaches a Solar Panel? Given the basics of solar-panel technology, the big question is, how do we predict the amount of solar energy that will reach the surface of a solar panel for any location and time? *Solar Insolation* measures how much solar energy (watts/m^2) reaches Earth's surface under a cloudless day, and is the key input to solar-generation forecasting. For any given point on Earth, the amount of solar energy will vary not only throughout the day as the Sun tracks across the sky, but also throughout the year as the Sun cycles between the Tropic of Capricorn and the Tropic of Cancer.

Johannes Kepler's First, Second, and Third Laws of Motion combined with Sir Isaac Newton's explanation of the motion of planets give us everything we need to predict solar insolation for any location and time. The detailed calculations for estimating solar insolation are presented below. Three parameters are needed to track the position of the Sun for a specific location, date, and time: (a) *solar-declination angle*, (b) *solar-hour angle*, and (c) *solar-altitude angle*. In addition, the *sunrise-hour angle* and *sunset-hour angle* are used to determine the time of sunrise and sunset. The formulas for each parameter are presented below.

Solar-Declination Angle is the angle between the Sun's rays and a plane passing through the equator. In the northern hemisphere, the solar-declination angle has a maximum value of 23.45° on June 21 and a minimum value of -23.45^{o} on December 21. The annual cycle for the solar-declination angle is depicted in

Figure 1. The solar-declination angle is computed as follows:

$$\mathrm{SolarDeclinationAngle_d} = 23.45^{o}\mathrm{xSIN}\left(360^{o}\mathrm{x}\left(\frac{284+\mathrm{d}}{365}\right)\right)$$

Where *d* is the number of the day in the year (e.g., *d* =1 to 365). The solar-declination angle for January 1, 2010 through December 31, 2012 is shown in

Figure 1.

Solar-Hour Angle measures the position of the Sun relative to solar noon at a given location and time. The solar-hour angle will be 0.0° at local solar noon—this is the point the Sun is highest in the sky. The solar-hour angle will be negative before local solar noon and positive after local solar noon. The solar-hour angle changes 15° each hour or 1° every four minutes. For five-minute modeling, this means the solar-hour angle changes 1.25° every five minutes. At local solar hour 08:00, the solar-hour angle will equal -60°. At local solar hour 16:00, the solar hour angle will equal 60°.

Solar-Altitude Angle is the angle between the Sun's rays and a horizontal plane as the Sun traverses the sky between sunrise and sunset. At the Sun's zenith (solar noon), this angle will vary across the year. The solar-altitude angle can be calculated for any location and time as follows:

$$\mathrm{SIN}(\alpha) = \mathrm{SIN}(\mathrm{L})\mathrm{SIN}(\delta) + \mathrm{COS}(\mathrm{L})\mathrm{COS}(\delta)\mathrm{COS}(\omega)$$

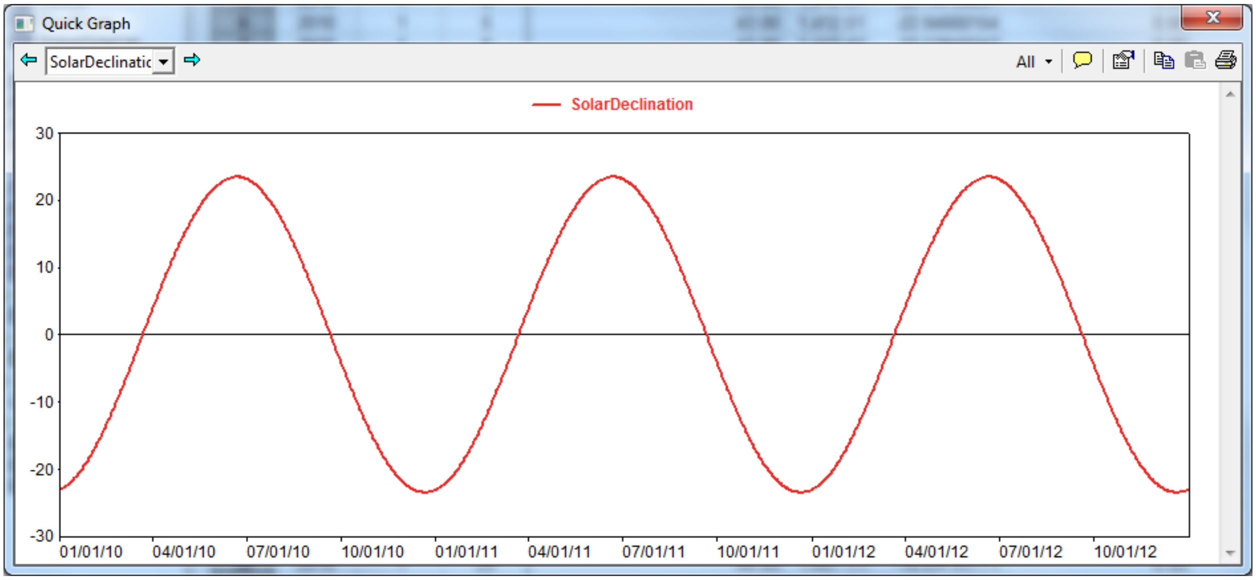


Figure 1: Solar Declination Angle

Where,

α *is the Solar Altitude Angle*
δ *is the Solar Declination Angle*
L *is the lattitude*
ω *is the Solar Hour Angle*

We can use this equation to determine the solar-altitude angle for solar noon for a specific day and location. For example, to calculate the solar-altitude angle for solar noon on February 15 in Honolulu, Hawaii (latitude 21.3069°N; longitude 157.8583°W), the computations are as follows:

First, we need to compute the solar-declination angle for February 15.

$$SolarDeclinationAngle_{Feb-15} = 23.45^{o} x SIN\left(360^{o} x \left(\frac{284+46}{365}\right)\right) = -13.3$$

Second, since it is solar noon, the solar-hour angle is $\omega = 0°$.

Given this information, we have:

$$SIN(\alpha) = SIN(21.3069°) \times SIN(-13.3°) + COS(21.3069°) \times COS(-13.3°) \times COS(0°)$$

$$SIN(\alpha) = 0.823175$$

$$\alpha = SIN^{-1}(0.823175) = 55.4°$$

In the northern hemisphere, as Earth rotates toward the summer solstice, the solar-altitude angle will steepen. If we re-do the calculations for solar noon on June 21, we have a solar-altitude angle of 87.9°. The solar-altitude angle for February 15 and June 21 are depicted in Figure 2 and Figure 3.

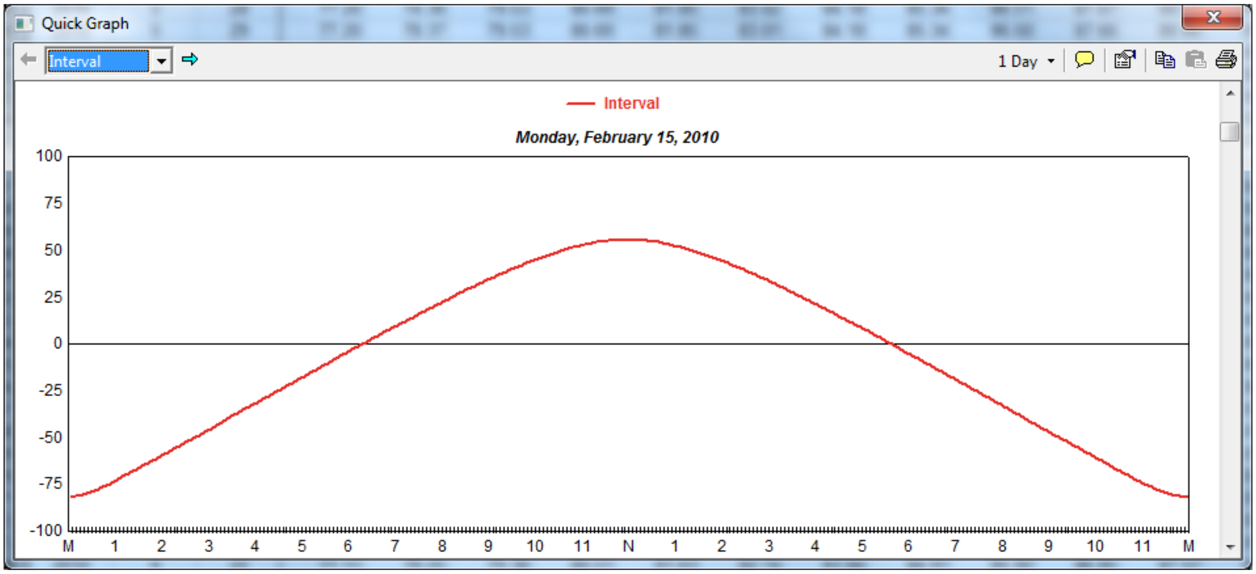


Figure 2: Solar-Altitude Angle (February 15)

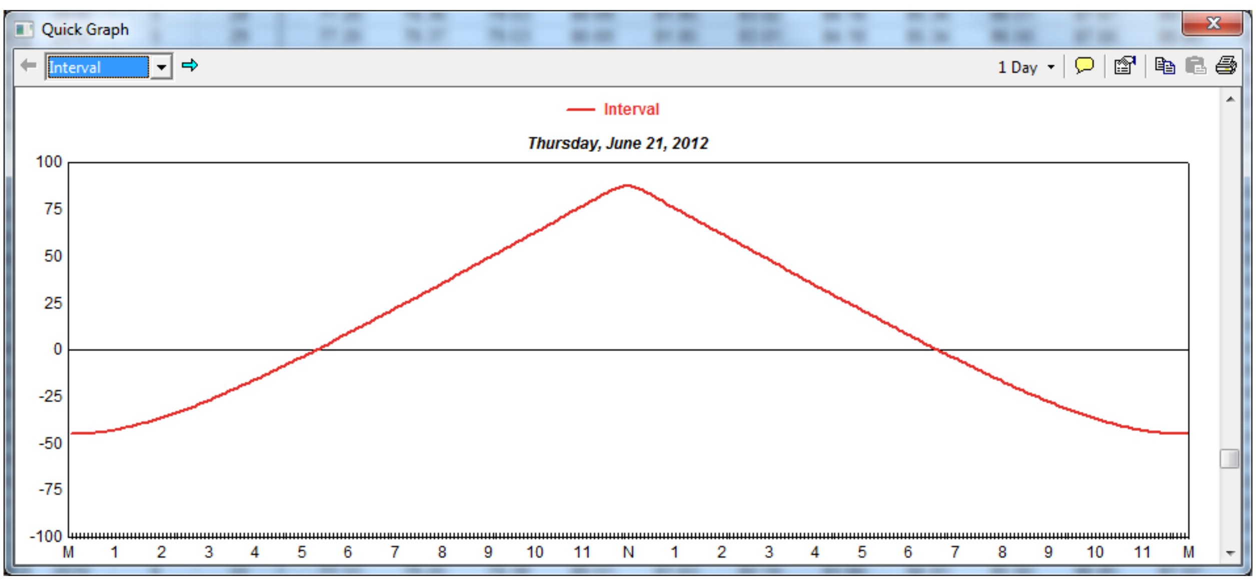


Figure 3: Solar-Altitude Angle (June 21)

To determine the time difference between solar noon and sunrise and sunset, it is useful to compute the sunrise and sunset solar-hour angles. To compute the time difference, all you need to do is multiply solar-hour angle at sunrise (or sunset) by four minutes per degree. The sunrise solar-hour angle will be that point at which the solar-altitude angle equals

0.0. Using the information above, we can compute the solar-hour angle at the time of sunrise and sunset for February 15 as follows.

In this case, we know the solar- altitude angle at the time of sunrise and sunset is 0°. This gives:

$$SIN(0°) = SIN(21.3069°)\times SIN(-13.3°) + COS(21.3069°)\times COS(-13.3°)\times COS(\omega)$$

Solving for ω gives:

Sunrise-Hour Angle = -84.7

Sunset-Hour Angle = 84.7

This implies the Sun rises approximately 5 hours and 39 minutes (computed as 84.7° times 4 minutes per degree) before solar noon and sets approximately 5 hours and 39 minutes after solar noon. If we repeat the calculations for June 21, we have the Sun rising approximately 6 hours and 39 minutes before solar noon and setting 6 hours and 39 minutes after solar noon.

It is important to note that local solar time is not the same as local standard time. For example, solar noon as measured by a sundial will not always occur at the same time every day. The difference between the time of solar noon and noon standard time can be up to +/-16 minutes. This also accounts for the asymmetry in the times of sunrise and sunset. The difference between local standard time and local solar time is referred to as the **Equation of Time**.

The Equation of Time has two causes:

- *Angle of Obliquity*. The plane of Earth's equator is inclined to the plane of Earth's orbit around the Sun
- *Elliptical Orbit*. The orbit of Earth around the Sun is an ellipse and not a circle.

Because of the angle of obliquity, solar time changes throughout the year as the Sun moves above and below the equator. The elliptical orbit means the distance between Earth and the Sun is at minimum near December 31 and is at maximum near July 1.

The Equation of Time can be written as follows:

$$\mathrm{Equation\_of\_Time} = 9.87\times \mathrm{SIN}(2B) - 7.53\times \mathrm{COS}(B) - 1.5\mathrm{SIN}(B)$$

$$\text{Where,} \quad B = \frac{360}{365}\times(d - 81)$$

Here, the Equation of Time is measured in minutes and *d* is the number of days since the start of the year.

The Equation of Time accounts for the physical reasons why solar noon is not the same as noon clock time. The time-correction factor (in minutes) accounts for the variation of local solar time within a given time zone due to the longitudinal variations within the time zone and incorporates the Equation of Time.

$$\mathrm{TimeCorrectionFactor} = 4\times(\mathrm{LocalStandardMeridian} - \mathrm{Longitude}) + \mathrm{Equation\_of\_Time}$$

Here, longitude is set by the location of the solar panel. The factor of four minutes reflects the fact that Earth rotates 1° every four minutes. In cases where the local longitude equals the local standard meridian, the time-correction factor is simply equal to the Equation of Time.

Given these equations, we can then compute local solar time by adjusting the local standard time as follows:

$$\text{LocalSolarTime} = \text{LocalClockTime} + \frac{\text{TimeCorrectionFactor}}{60}$$

Using the example above, we can compute the time of local solar noon for February 15 as follows. Given that February 15 is the 46th day of the year, we have:

$$B = \frac{360}{365} \times (46 - 81) = -34.5205$$

$$Equationoftime = 9.87 \times SIN(2 \times -34.5205) - 7.53 \times COS(-34.5205) - 1.5 SIN(-34.5205) = -14.57\ Minutes$$

Given that the local prime meridian for Hawaii is 160°, we can compute the time correction factors as:

$$TimeCorrectionFactor = 4 \times (160° - (157.858°)) - 14.57 = -6.00\ Minutes$$

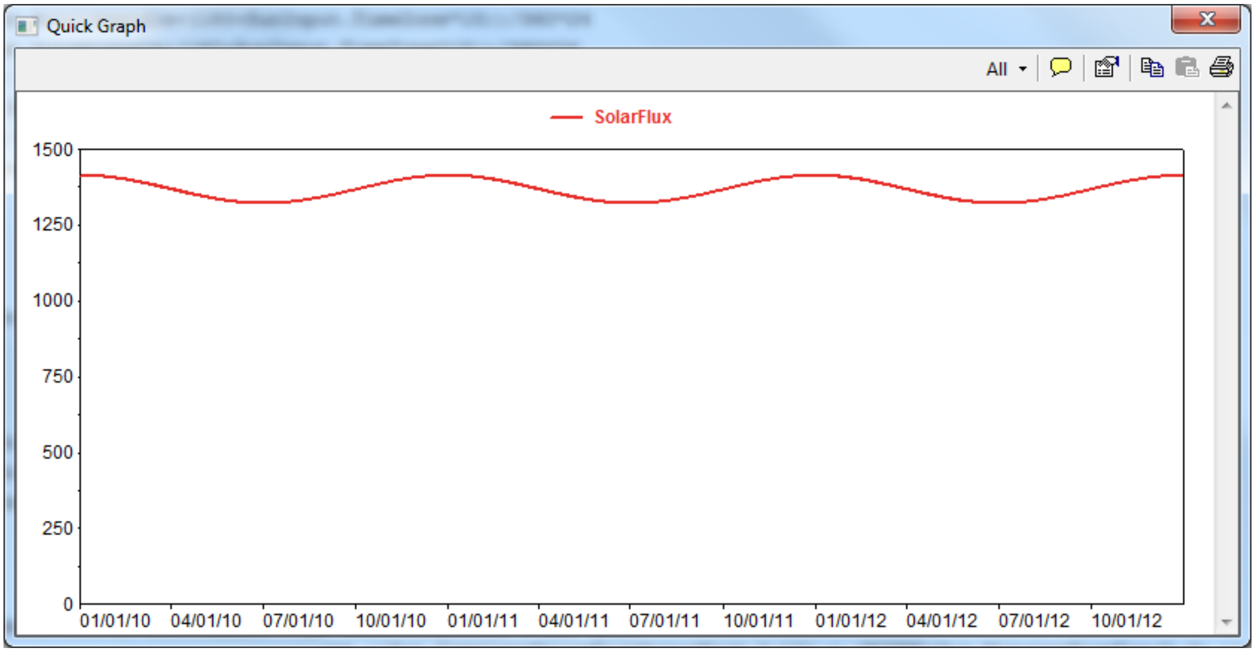
This implies that solar noon will take place at roughly 11:54 a.m. local clock time, not adjusted for daylight savings. From the perspective of forecasting solar generation, it is important to recognize this distinction in time. Specifically, if the generation forecast is to be in local clock time, then adjustments for daylight savings need to be made to the engineering estimates of solar generation in order to keep the engineering estimates in line with metered generation.

Solar Flux. Now that we know where the Sun is in the sky for any location and time, we can determine how much solar energy will strike a horizontal surface (i.e., solar panel). Scientists know that solar radiation strikes Earth's outer atmosphere on average at a rate of 1367 watts/m^2. This is commonly referred to as the *solar constant*. To account for seasonal variation due to the annual cycle in the distance between Earth and the Sun, the actual solar radiation (*solar flux*) hitting Earth's atmosphere on any day of the year can be calculated as follows:

$$\text{SolarFlux}_{\text{d}} = \text{SolarConstant}\left[1 + 0.034\text{COS}\left(\frac{360\text{d}}{365.25}\right)\right]$$

Here, *d* indexes the day of the year. A depiction of the annual cycle of solar flux is shown in Figure 4.

Figure 4: Solar Flux



Solar Insolation. The amount of solar energy hitting a horizontal plane on Earth's surface for any location and time of day can then be computed as follows:

$$\mathrm{SolarInsolation}_d^i = \mathrm{SolarFlux}_d \times \mathrm{COS}(\emptyset_d^i)$$

Here, the time interval of the day (d) is indexed by (i) and ∅ is the solar-zenith angle. The solar-zenith angle is computed as 90° – Solar-Altitude Angle. Given this, we can rewrite the equation for the amount of solar energy that hits a horizontal plane (i.e., a solar panel) on Earth's surface for any location and time as follows:

$$\mathrm{SolarInsolation}_d^i = \mathrm{SolarFlux}_d \times \mathrm{COS}(90^o - \alpha_d^i)$$

Using the example data from above, the amount of solar insolation at solar noon on February 15 is 1,152 watts/m^2 and 1,321 watts/m^2 on June 21. The patterns of solar insolation for the weeks of February 14 and June 20 are depicted in Figure 5 and Figure 6.

Figure 5: Solar Insolation Week of February 14

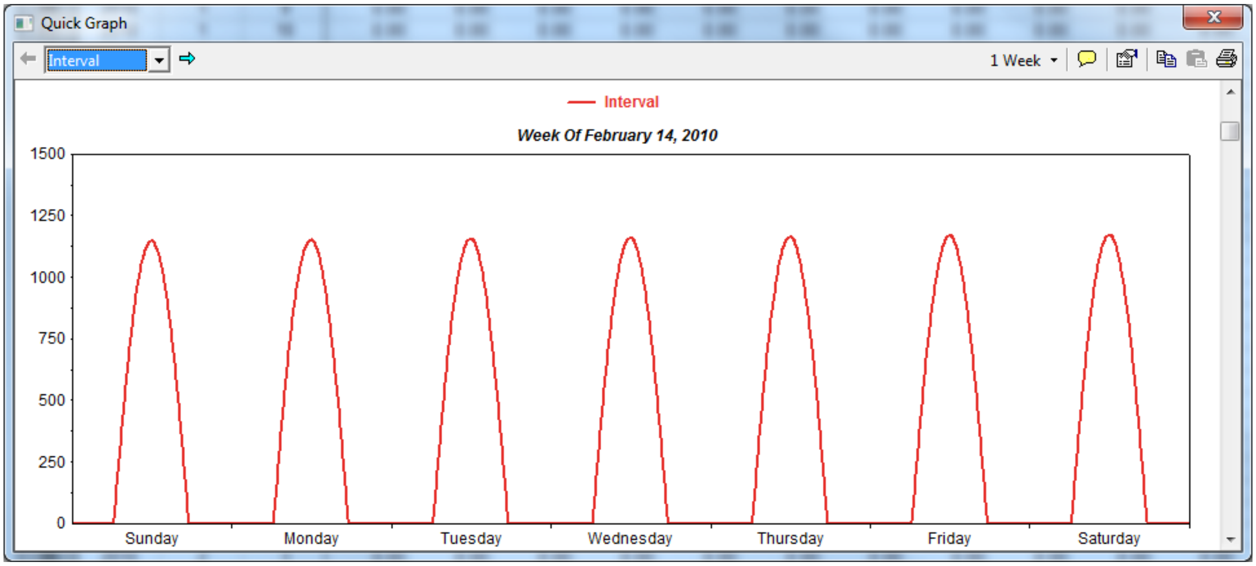
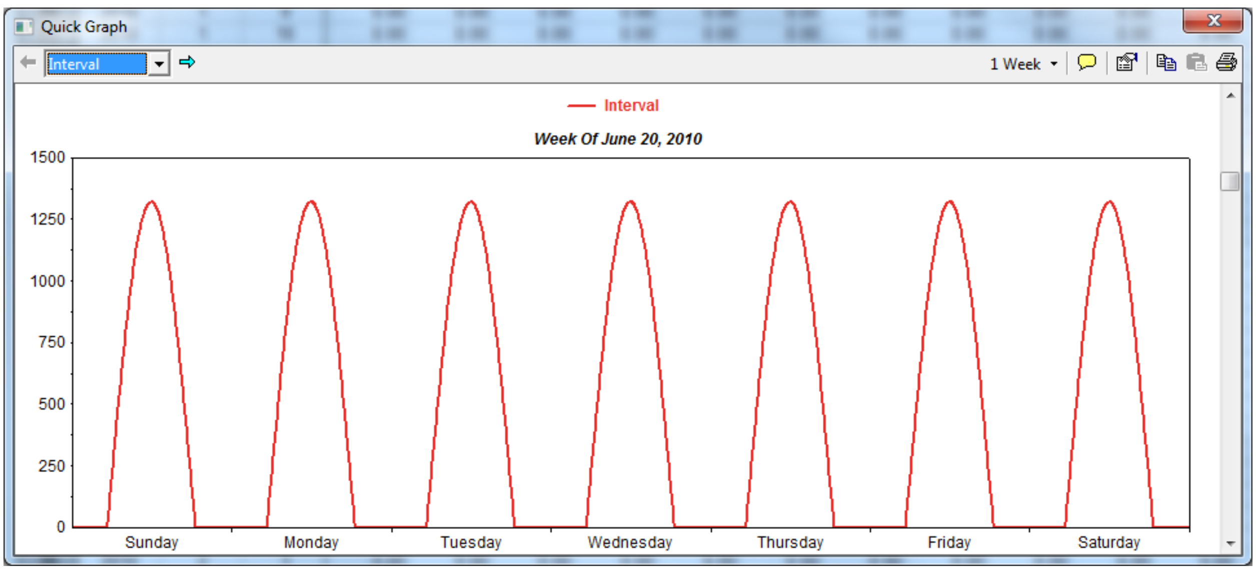


Figure 6: Solar Insolation Week of June 20



Engineering Model of Solar Generation

Given an estimate of how much solar energy is delivered to Earth's surface for any location and time, we can construct an engineering estimate of how much electricity is generated using the following relationship.

$$\mathrm{SolarGeneration}_{\mathrm{d}}^{\mathrm{i}} = \mathrm{SolarInsolation}_{\mathrm{d}}^{\mathrm{i}} \times \mathrm{SolarPanelCapacity}_{\mathrm{d}} \times \mathrm{SolarPanelEfficiency}_{\mathrm{d}}^{\mathrm{i}}$$

Here,

$SolarGeneration_d^i$ *is the electricity generated on day* (d) *time interval* (i) *in* $Watts\ Out$

$SolarInsolation_d^i$ *is the solar energy delivered to the panel in* ${}^{Watts\ In}/_{m^2}$

$SolarPanelCapacity_d$ *is the installed capacity in* m^2

$SolarPanelEfficiency_d^i$ *is the solar panel efficiency in* ${}^{Watts\ Out}/_{Watts\ In}$

To help fix ideas, assume solar insolation at noon of June 12 is 1,000 Watts/m^2, installed capacity is 2.5 kW, and the solar panel efficiency is 15%. If we assume 150 Watts/m^2 for the average panel size, we can say the installed capacity is approximately 16.66 m^2 (computed as 2500 Watts over 150 Watts/m^2). With these numbers we have:

$$SolarGeneration = 2500\ Watts = 1000\ {}^{Watts}/_{m^2} \times 16.667 m^2 \times 0.15$$

Alternatively, this can be written as:

$$SolarGeneration = 2500\ Watts = 1000\ {}^{Watts}/_{m^2} \times 2.5\ kW$$

This can be expressed as:

$$\mathrm{SolarGeneration}_{\mathrm{d}}^{\mathrm{i}} = \mathrm{SolarInsolation}_{\mathrm{d}}^{\mathrm{i}} \times \mathrm{SolarCapacity}_{\mathrm{d}}$$

Where, $SolarCapacity_d$ is the installed capacity on day (d) in kW.

Factoring in Temperature Impacts. The hotter a solar panel becomes, the less efficiently it converts sun energy into useful electricity. This leads to the following adjustment to the solar-panel-efficiency calculation.

$$\mathrm{SolarPanelEfficiency}_{\mathrm{d}}^{\mathrm{i}} = \mathrm{RatedEfficiency} \times \left(1 - \left[\mathrm{MAX}\left(\mathrm{Temp}_{\mathrm{d}}^{\mathrm{i}} - \mathrm{ThresholdTemp}, 0\right) \times \nabla\right]\right)$$

Here,

$SolarPanelEfficiency_d^i$ *is the solar panel operating efficiency for day* (d) *time interval* (i)

$RatedEfficiency$ *is the peak output efficiency*

$Temp_d^i$ is the temperature of the panel

$ThresholdTemp$ is the temperature above which the efficiency of the panel degrades

∇ is the rate of efficiency degradation per degree (0.48%℃ or 0.27%℉).

Factoring in Cloud Cover. Cloud cover lowers the output of a solar panel by reducing the amount of solar energy reaching the panel. While the exact impact of cloud cover on a particular location is difficult to measure, we can assume that at 100% cloud cover, only about 20% of the solar flux reaches Earth's surface. That is, the cloud albedo is 80% at 100% cloud cover. We can use this information to adjust the engineering estimate of solar insolation by incorporating the following relationship:

$$\boldsymbol{CloudAlbedo_d^i = CloudCoverPercentage_d^i \times 80\%}$$

$$\mathrm{SolarInsolation_d^i = SolarFlux_d \times COS(90^o - \alpha_d^i) \times (1 - CloudAlbedo_d^i)}$$

The final engineering model of solar generation can then be written as follows:

$$\mathrm{SolarGeneration_d^i = SolarInsolation_d^i \times (1 - CloudAlbedo_d^i) \times SolarCapacity_d \times (1 - [MAX(Temp_d^i - ThresholdTemp, 0) \times \nabla])}$$

Incorporating Embedded Solar Generation into a Load Forecast Model

In practice, what the control room sees is metered demand, not actual demand. Following the example presented above, on a sunny day the control room will see a metered demand of 1,200 KW, and on a cloudy day metered demand will be 1,300 KW. Since we do not measure actual demand, we can ask: *How does embedded solar generation affect forecasts of metered demand?* To answer this question, we first re-arrange the equation of demand to provide the following equation of metered demand:

$$\mathrm{MeteredDemand_d^i = Demand_d^i - SG_d^i = PredictedMeteredDemand_d^i + (AvgSG_d - SG_d^i)}$$

The above equation provides the basis for a regression of metered demand on a set of explanatory variables that drive the demand for electricity, and the difference between average solar generation over the estimation period and the current-period solar generation.

$$\mathrm{MeteredDemand_d^i = G(X_d^i) + \partial(AvgSG_d - SG_d^i) + \varepsilon_d^i}$$

Here, G() is the set of explanatory variables driving the demand for electricity. This set of explanatory variables would include calendar conditions (e.g. day of the week, season, and holidays), weather conditions (e.g. temperature, humidity), and economic conditions (e.g. growth trends). It is expected that the predicted value from the estimated function G() will be the metered demand expected when embedded solar generation equals the average solar generation over the estimation period. The expected value sign on ∂ is 1.0. This implies that on sunny days, metered demand will be lower than average due to higher-than-average embedded solar generation. On cloudy days, metered demand will be higher than average due to lower-than-average embedded solar generation.

Here are practical steps for estimating a model of Metered Demand:

Constructing an Historical Time Series of Embedded Solar Generation. First, an historical time series of embedded solar generation by day and time-of-day is needed. To construct this time series, we need three pieces of information: (1) estimated solar insolation for the geographic region being studied, (2) installations of embedded solar generation capacity over the estimation period, and (3) cloud-cover observations over the estimation period.

- **Solar Insolation by Location**: The above section on solar-panel basics presents the framework for computing solar insolation. Rather than keying up these calculations by hand, let us use the National Oceanic & Atmospheric Administration's (NOAA) useful solar calculation spreadsheet that will compute a year of daily solar insolation values for any location on Earth.[3] This spreadsheet can help us develop estimates of solar insolation by location and day of year.
- **Solar Insolation by Time-of-Day**. The solar insolation value computed on the NOAA spreadsheet is for solar noon. To compute a value of solar insolation for a specific time of the day, we need to know the solar-altitude angle for that time point. Again, we can use the NOAA spreadsheet, which provides an estimate of the time of solar noon as corresponding to a solar-altitude angle of 90°. We also have the estimated sunrise and sunset times. Since the solar-altitude angle at the time of sunrise and sunset is 0°, we can back into the average decay per minute in the solar-altitude angle.

$$\text{Angle Lost Per Minute}_d = 90^\circ / (\text{Time of Solar Noon}_d - \text{Time of Sunrise}_d)$$

Typically, the value for Angle Lost Per Minute will range between 0.2° per minute and .31° per minute, with the average value of approximately 0.25° per minute; or 4° per minute.

Given this value, we can then compute the solar-altitude angle for any time period as:

$$\text{SolarAltitudeAngle}_d^t = 90^\circ - \left(\text{Angle Lost Per Minute}_d \times \left|\text{Time of Solar Noon}_d - \text{Time of Interval of the day}_d\right|\right)$$

Here, the absolute-value function returns the number of minutes between the time of solar noon and the time of the time interval (t) under study.

Given the solar insolation at solar noon, we can then compute the solar insolation for time interval (i) as follows:

$$\text{SolarInsolation}_d^i = \text{SolarInsolation}_d^{\text{SolarNoon}} \times \text{COS}\left(\text{SolarAltitudeAngle}_d^i - 90^\circ\right)$$

- **Embedded Solar Generation Capacity**. In many regions, solar-generation installations are driven by government incentives. This means that, in many regions, good estimates of installed-solar capacity are available.
- **Cloud Cover & Temperatures**. Much of the focus in solar-generation forecasting is on developing accurate forecasts of cloud cover. The techniques range from vector decomposition of satellite imagery to vector decomposition of location-specific cloud-cover observations. Most of this analysis is geared for forecasting generation at utility solar installations and/or solar generation over a small geographic footprint. This micro-focus is most useful when the exact locations of the solar installations are known. Unfortunately, this is not the case for most non-utility solar installations. In these cases, a market operator may have good estimates of the total

[3] See http://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html for the NOAA Solar Calculation spreadsheets.

installed capacity by transmission zone and/or possibly by postal code. Further, the load forecasted typically spans multiple postal codes. For purposes of non-utility solar generation forecasting, having cloud-cover data for a handful of weather stations spanning the geographic footprint of the metered demand is sufficient to capture the overall impact of embedded solar generation on loads. As a result, the weather-station-level cloud-cover and temperature observations and forecasts used by the load-forecast model can be used to drive the embedded solar-generation data and forecasts.

The above information is then combined using the engineering model of solar generation and re-written as follows:

$$\mathrm{SolarGeneration}_d^i = \mathrm{SolarInsolation}_d^i \times \left(1 - \mathrm{CloudAlbedo}_d^i\right) \times \mathrm{SolarCapacity}_d \times \left(1 - \left[\mathrm{MAX}\left(\mathrm{Temp}_d^i - \mathrm{ThresholdTemp}, 0\right) \times \nabla\right]\right)$$

It is useful to develop solar-generation estimates at the five-minute level of detail. The detailed data can then be used to support 5-, 15-, 30- and 60-minute modeling of loads and generation.

Adjusting for Observance of Daylight Savings. If the meter loads have been adjusted for Daylight Savings Time, then a similar adjustment will need to be made to solar-generation-time series data.

A Proxy for Average Solar Generation. When solar generation is introduced into the load-forecasting model, it is added as the difference between the average solar generation for the time period and the solar generation for the current day. In principle, the average solar generation should be computed as the average solar generation over the model-estimation time period. While this is a plausible value to compute by hand, in real-time forecasting systems it is not practical to re-compute the average every time the load model is re-estimated. A practical work-around is to replace the average solar generation with a proxy value based on the solar generation that can be expected under full sunshine conditions.

$$\mathrm{AvgSG}_d = \rho \mathrm{FullSunSG}_d^i$$

Here, FullSunSG is the solar generation output on day (d) and time interval (i) under cloudless skies. The parameter ρ is bounded between 0 and 1. The value for this parameter can be estimated directly from the data using the following regression:

$$\mathrm{SG}_d^i = \rho \mathrm{FullSunSG}_d^i + \varepsilon_d^i$$

In practice, the estimated coefficient on Full Sunshine Solar Generation will be higher for time intervals between, say, 9 a.m. and 3 p.m., and lower for the shoulder hours. A rule-of-thumb value would be 0.8.

Why not just include SG_d^i into the model by itself? The above approach requires knowledge, or at least a good proxy, of average solar generation over the estimation period. Since this is an extra step, it is reasonable to ask why we don't simply include SG_d^i into the model by itself and not incur the overhead of computing the delta term. If this were the approach, the *a priori* value for the estimated coefficient on SG_d^i would be -1.0. That is, when current solar-generation levels are high, metered demand, and hence the forecast of metered demand, should be low. Unfortunately, timing challenges this approach. The recent growth of embedded solar generation coincides with recovery from the global economic crisis of 2008/2009. As a result, the pace of embedded solar-generation installations correlates positively with economic recovery — specifically, the rebound of metered demand. These longer-term trends will offset the very-short-term load variation due to embedded solar generation, resulting in a positive estimated coefficient on the SG_d^i variable. This will lead to compounding of forecast errors due to the incorrect sign on the solar generation variable. This spurious positive correlation, however, works in our favor when we add $(AvgSG_d - SG_d^i)$ to the model. The *a priori* value for the delta is 1.0, which will align with the positive correlation between metered demand growth and the growth of solar installations.

Shoulder Hours. During the prime sunshine period of the day (9 a.m. to 3 p.m.), solar panels operate close to full capacity. As a result, cloud-cover variation will lead to wide swings in solar-generation output. These swings make it easier for a statistical model to isolate a strong coefficient on the difference between average solar generation and the current solar-generation value. In contrast, the variance in solar generation during the swing hours tends to be swamped by the overall load variation. As result, it is difficult to get a reasonable sign on the solar-generation variable. This suggests a hybrid approach. For the prime solar hours, estimating the load model with the solar-generation variables directly is viable. For the shoulder or swing hours, it is recommended that a coefficient value be imposed rather than estimated directly.

Autoregressive Terms. If autoregressive terms are included as explanatory variables, then a revision is needed to the solar-generation variable included in the model. To illustrate the problem, consider forecasting loads for 11:45 a.m. given load data through 11:30 a.m. Further, assume the load-forecast model for 11:45 a.m. includes the load at 11:30 a.m. as an autoregressive term in the model. In this case, the load at 11:30 a.m. embodies the amount of solar generation that took place at 11:30 a.m. Since the solar generation at 11:30 a.m. is already flowing into the forecast of loads at 11:45 a.m., we must include the change between solar generation at 11:30 a.m. and solar generation at 11:45 a.m. The model would have the following generation specification:

$$\mathrm{MeteredDemand}_d^i = G\left(X_d^i, \mathrm{MeteredDemand}_d^{i-j}\right) + \partial\left(SG_d^{i-j} - SG_d^i\right) + \varepsilon_d^i$$

Here, the prior-time interval is indexed by (i-j). In this model we have included prior-period loads in the set of explanatory variables. Further, we have included the delta between prior-period solar generation and current-period generation. The *a priori* value for ∂ is 1.0. That is, holding all things else the same, if the solar generation at 11:45 a.m. is higher than the solar generation at 11:30 a.m., we expect that the metered demand at 11:45 a.m. will be lower than the metered demand at 11:30 a.m.

Conclusions

Installed solar generation capacity is growing worldwide. The effect of this movement on load forecasts is significant. As a result, there is a growing demand to capture the impact of embedded solar generation on day-of and day-ahead load forecasts. Forecasting both utility solar installations and non-utility installations pose different challenges. This paper presents practical steps to incorporating embedded solar generation into an existing short-term load forecast model.

Itron is a global technology company. We build solutions that help utilities measure, monitor and manage energy and water. Our broad product portfolio includes electricity, gas, water and thermal energy measurement and control technology; communications systems; software; and professional services. With thousands of employees supporting nearly 8,000 utilities in more than 100 countries, Itron empowers utilities to responsibly and efficiently manage energy and water resources.

Join us in creating a more resourceful world; start here: **www.itron.com**

CORPORATE HEADQUARTERS

2111 N Molter Road
Liberty Lake, WA 99019
USA

Phone: 1.800.635.5461
Fax: 1.509.891.3355