California Solar Initiative 15-Minute Interval Data

As part of the program's measurement and evaluation effort, the California Solar Initiative (CSI) authorized the program contractor (Itron) to install production meters on 504 CSI systems in 2010. Fifteen minute interval data are being collected from these systems and delivered to the CSI administrators on a quarterly basis through 2016. The data sets are being made available for download on the California Solar Statistics website. The September 30, 2015 (third quarter 2015) data set (which covers performance beginning in 2010) is the first to be made publicly available.

The data consist of three files – one for each Investor Owned Utility (PG&E, SCE, and SDG&E). Each file is a compressed (zipped) comma-delimited data set. Microsoft Excel will likely not be able to fully open the files due to the very large file sizes. Instead, it is recommended that a tool capable of handling millions of rows of data be used, such as SAS or R.

It is reasonable to expect that performance characteristics will change as technology, manufacturing methods, and ownership and maintenance protocols change over time. Users of this data set should note that the data are from systems installed predominantly on single family houses, in the 2007-2010 timeframe, and manufactured in that timeframe or earlier. Systems of different vintages, sectors, or ownership types may differ from those whose performance is shown here.

Below is a key to the data:

15-Minute Interval Data - Data Set Key

Column	Field	Definition
A	ID	CSI application number
B	Inctv_type	Incentive type; note: all applications are EPBB (which represents Expected Performance Based Buydown)
C	Pa	Program administrator for the site
D	LocalDateTime	Beginning time for the 15-minute period in Pacific Standard Time (Daylight savings are ignored)
E	kWh	The energy produced by the system during that 15 minute interval
F	Flag	M, Mx, Rg, Rl, or I (see below for descriptions of these codes)

Two of the columns are described at greater length below.

"ID" column linking interval data with application data

The "ID" column of each data set contains CSI Application Numbers associated with the systems. These application numbers can be used to look up attributes of the systems using the CSI Working Data Set.

"Flag" column in data set
The "Flag" column in each data set provides a coded value for each interval row. The value in this column can be interpreted as follows:

- 'M' = Metered directly (including raw data with interval <15 minutes)
- 'Mx' = Metered directly (including raw data with interval <15 minutes), but multiplied by a constant to account for a change in system capacity.
- 'Rg' = Metered, but data is "blocky" because it is recording in 1 kWh intervals.
  - Some sites that report in 1 kWh may not be accurately flagged
  - Smaller sites reporting in 1 kWh increments may show abnormally high or low 15 minute readings in relation to expected.
- 'Rl' = Ratio using metered data down from longer interval than 15 minutes up to 2 hours
  - Some 'Rl' sites may also be 'Rg'
  - Consider excluding these sites when doing 15 minute analysis since they will tend to artificially smooth data at that level
- 'Rh' = Metered data collected either on an hourly basis or summed to an hourly basis then averaged to the 15 minute intervals over that hour.
  - Some 'Rh' sites may also be 'Rg'
  - Consider excluding these sites when doing 15 minute analysis since they will tend to artificially smooth data at that level
- 'I' = Invalidated data due to data errors etc.

Technical Support
Please direct any technical questions to: csisupport@energy-solution.com.