STATISTICS OF TIME-DEPENDENT PHYSICAL QUANTITIES

Some comments on statistical measures

Wendelin Sprenger

Fraunhofer Institute for Solar Energy Systems ISE, Solar façades group

BIPV workshop, September 14th, 2013

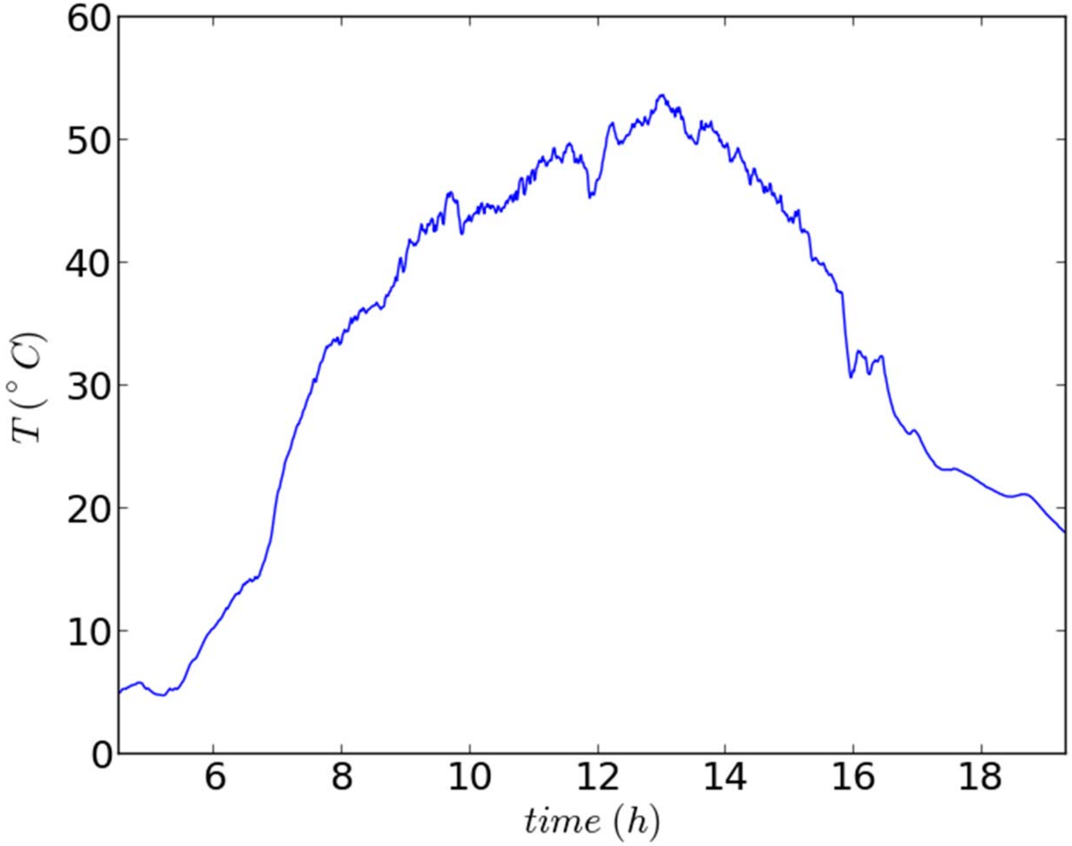
www.ise.fraunhofer.de

Fraunhofer ISE

AGENDA

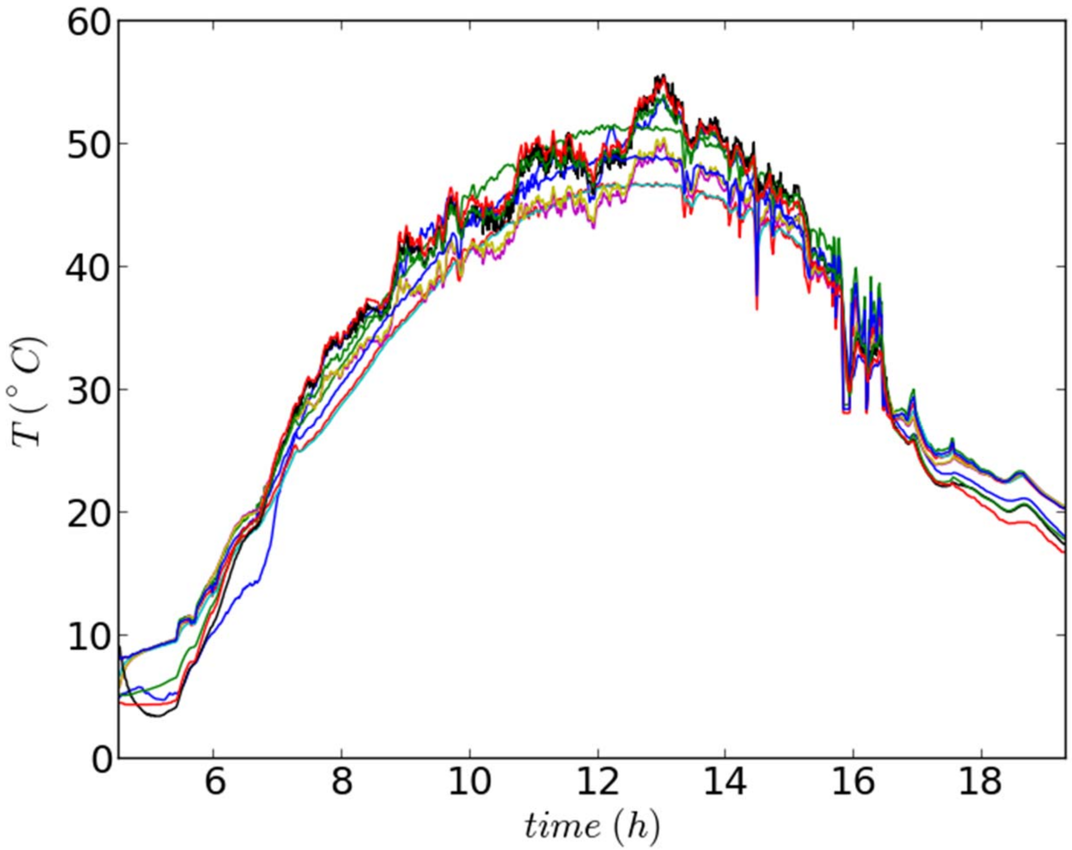
- Motivation
- The “uncertainty” of a time-dependent physical quantity
- The frequency distribution of deviations
- The statistical measures RMSE, MBE, MAE
- Dealing with the night time
- Percentual statistical measures
- Conclusions

Project SOPHIA: Simulation of the PV modules' back-surface temperature of a BIPV system



measured back-surface temperature of one PV module

Project SOPHIA: Simulation of the PV modules' back-surface temperature of a BIPV system



measured back-surface temperature of one PV module

and the results of nine different simulation approaches

Fraunhofer ISE

Statistics of time-dependent physical quantities

How can the „accuracy“ (or „uncertainty“) of a simulation method be quantified?

→ For the DC power:
Important for checking/proving the functionality of a building-integrated photovoltaic system!

→ Important for the validation of simulation approaches that are applied to optimize physical parameters of BIPV systems!

1) Analysis of the uncertainty with Gaussian error propagation method

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + \cdots}$$

Common misunderstanding:

The approach is only correct if the uncertainty of the relation **f itself** can be neglected (physical laws, relations based on physical laws etc.).

There is also the uncertainty of the approach itself
that the Gaussian law of error propagation does not cover.

Fraunhofer ISE

2) Comparison of simulated and measured time series

Many people would like to hear a percentage.

→ Common approach:
Compare the average values [or: totals] of the simulated and the measured time series. BUT:

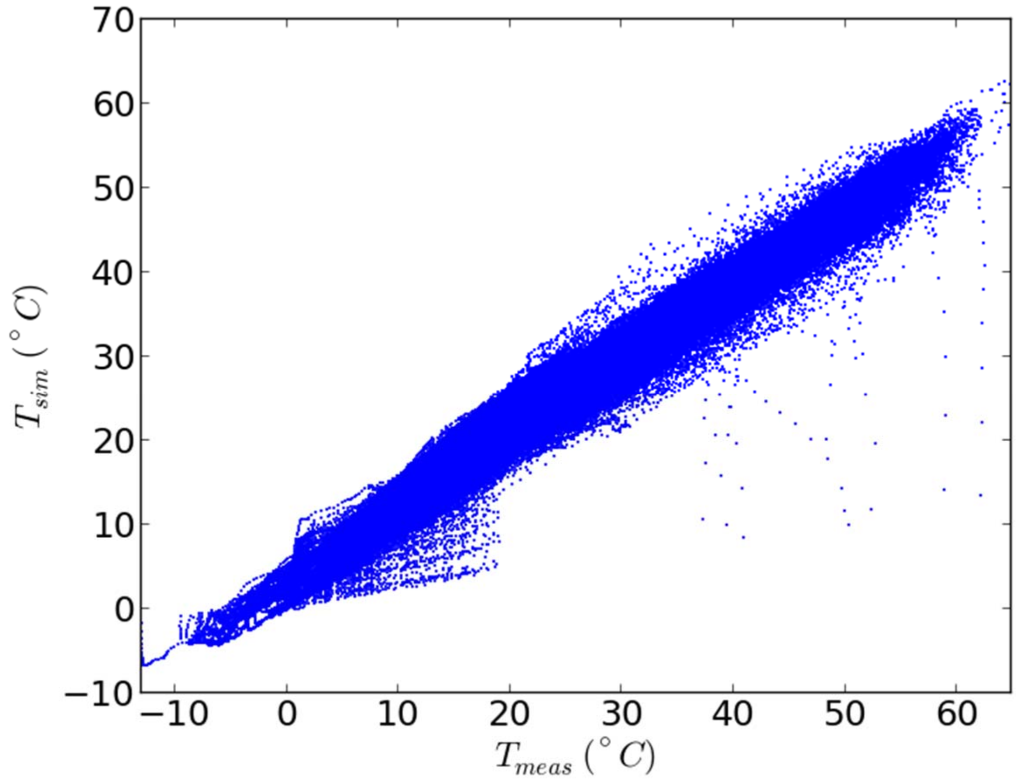
- No information about the dependence on time of the physical quantity!
- The relative deviation between the simulated and measured mean value varies from year to year!

Better approach: Analyze the differences for the simulated and the measured value for every time step and analyze the frequency distribution.

The frequency distribution of deviations

Common approach:

Plot measured values vs. simulated values.



Advantages:

- Fast method to check if the simulation has worked
- Outliers can be detected and interpreted

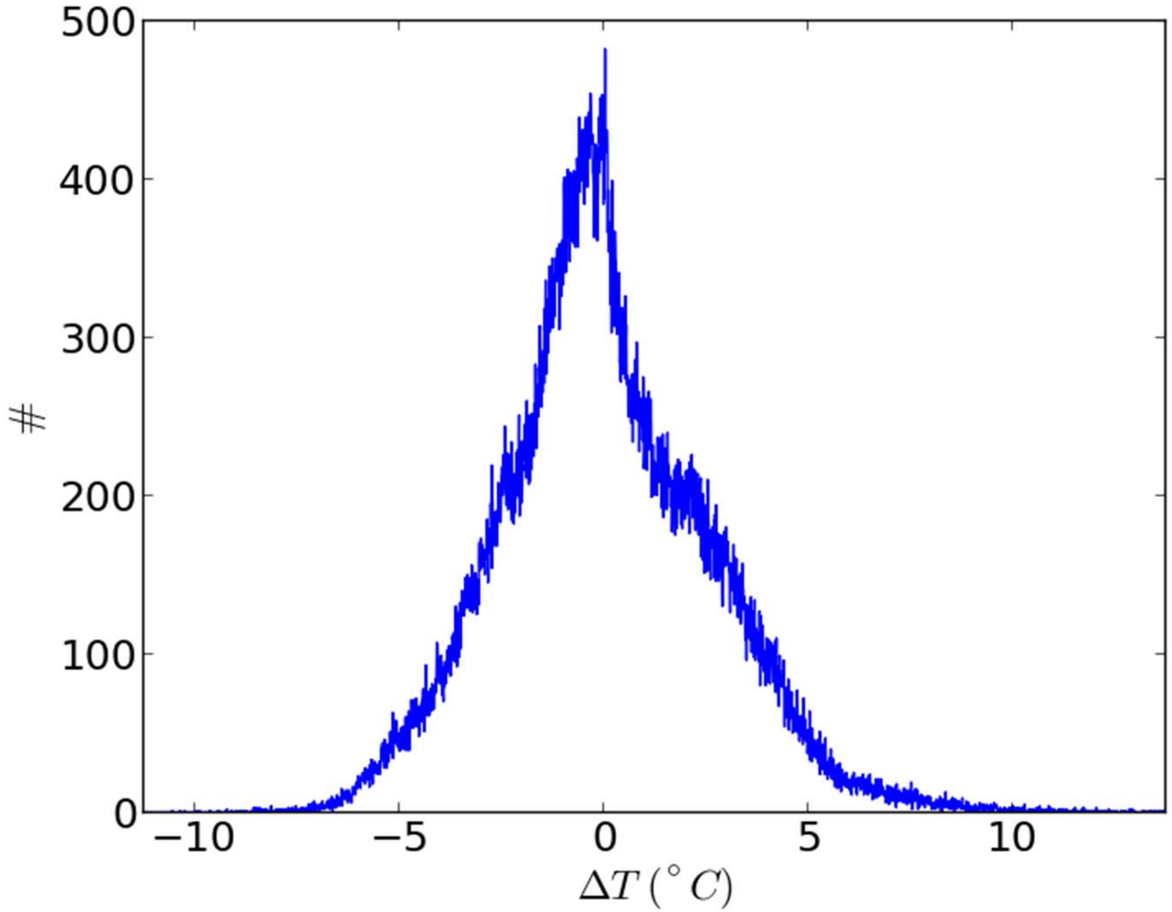
Disadvantages:

- The „standard deviation“ in the curve may depend on the number of time steps (!) as well as on the size of the dots (!)

Fraunhofer ISE

The frequency distribution of deviations

Better approach (in my opinion):
Plot histogram of deviations.



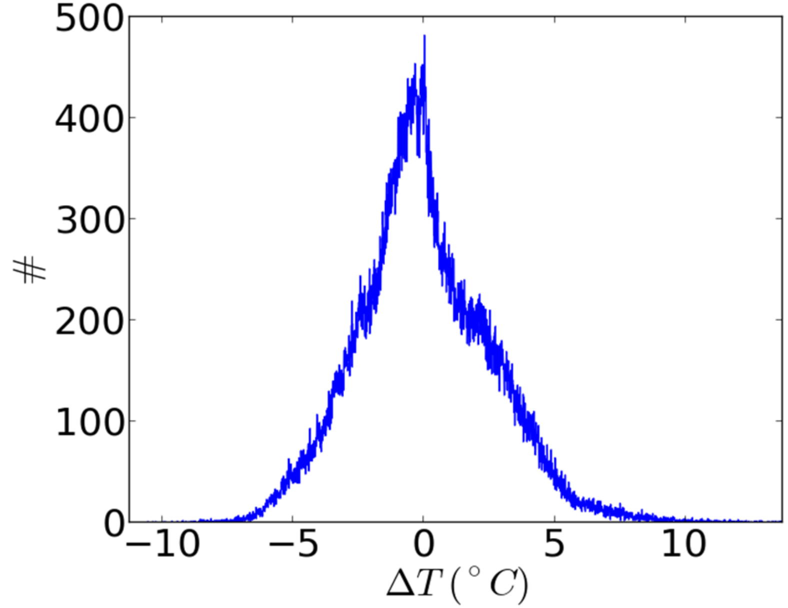
Histogram of differences between simulated and measured values.

A „standard deviation" can be defined, as well as a „shift from zero"!

The frequency distribution of deviations

$RMSE$ is similar to a standard deviation of the histogram curve:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}$$



MBE describes the shift from the zero-point in the histogram curve:

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})$$

The very similar MAE is the average absolute deviation.

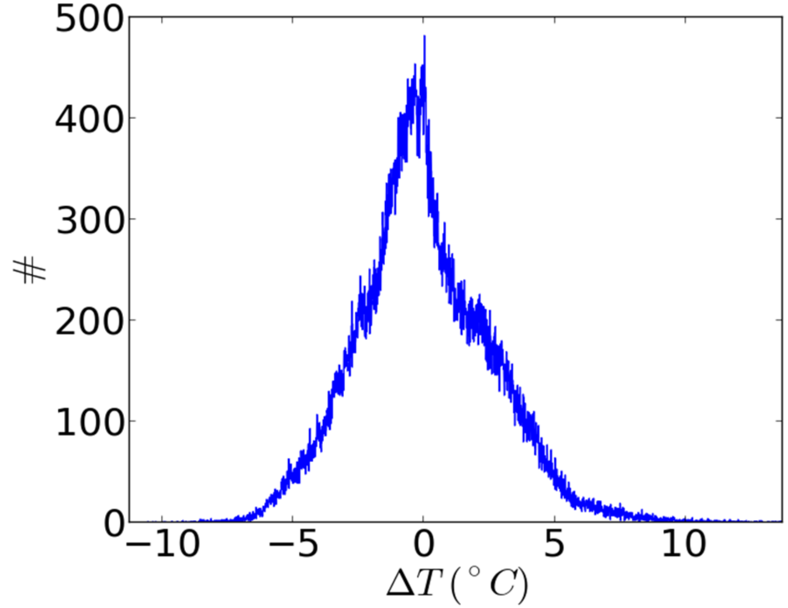
$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_{sim}^{i} - x_{meas}^{i}\right|$$

Fraunhofer ISE

The frequency distribution of deviations

R^2 is commonly applied as well (and may be useful as well), but does not belong here.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_{sim}^i - x_{meas}^i)^2}{\sum_{i=1}^{N}(x_{meas}^i - \bar{x}_{meas}^i)^2}$$



R^2 belongs to the least square method. The expression x_{sim}^i is the value of the function that should be fitted.

Fraunhofer ISE

Important properties of statistical measures

What are „good“ statistical measures (in our context)?

- They should be intuitive.
- Divisions by zero or very small denominators have to be avoided.
- They should be quite insensitive to (small) statistical outliers in the measurements.
- Their values should not change systematically when the time interval (5-minute values or hourly values) changes
- They should be rather independent on the analyzed physical quantity.

Dealing with the night time

What dataset (N) should be chosen for the statistical analysis? Should it include the night time?

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_{sim}^{i} - x_{meas}^{i}\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}{\sum_{i=1}^{N}(x_{meas}^{i} - \bar{x}_{meas}^{i})^2}$$

Dealing with the night time

- Excluding the night time changes the values of all presented statistical measures.

- Cutting off values below a certain irradiance value (e.g. $< 5\, W/m^2$) makes the values dependent on the chosen border value.

- There are physical quantities that may have relevance during the night time (e.g. temperature).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_{sim}^{i} - x_{meas}^{i}\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}{\sum_{i=1}^{N}(x_{meas}^{i} - \bar{x}_{meas}^{i})^2}$$

→ two common approaches

eliminating or keeping the night time values.

Fraunhofer ISE

Percentual quantifications

Time series of some physical quantities (like the DC power) are better quantified by relative numbers.

They are then divided by the average of the measured values.

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})$$

$$MBE(\%) = \frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i}) \bigg/ \frac{1}{N}\sum_{i=1}^{N} x_{meas}^{i}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}$$

$$RMSE(\%) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2} \bigg/ \frac{1}{N}\sum_{i=1}^{N} x_{meas}^{i}$$

Percentual quantifications

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2}$$

$$RMSE(\%) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{sim}^{i} - x_{meas}^{i})^2} \Big/ \frac{1}{N}\sum_{i=1}^{N} x_{meas}^{i}$$

Sometimes, also this (dangerous and not recommendable) definition is chosen:

$$RMSE(\%) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_{sim}^{i} - x_{meas}^{i}}{x_{meas}^{i}}\right)^2}$$

For the case of values that can be very small, the latter RMSE may be influenced most by <u>small</u> values.

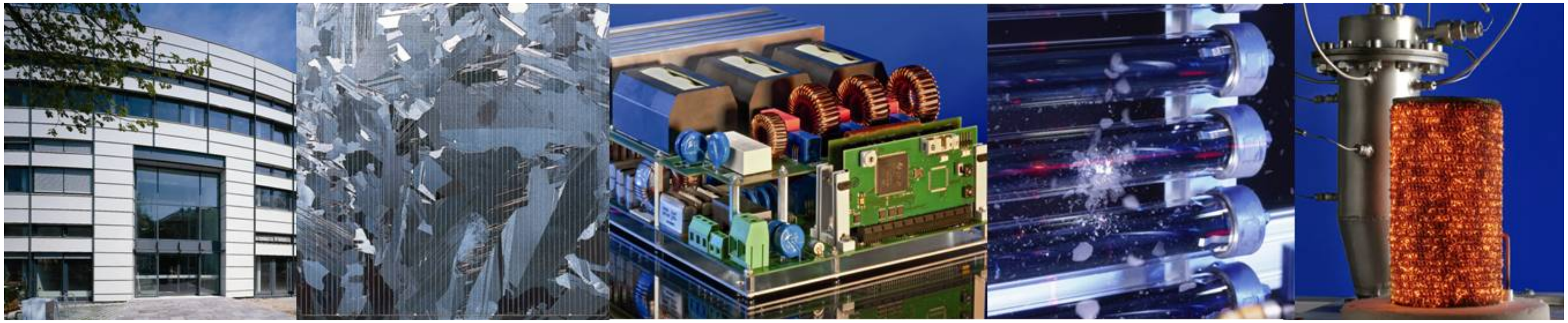
Conclusions

- No standardized way of statistical quantification has been established yet.
- In literature, many different definitions exist, also of the most common statistical measures.

- For this reason, it is always important to exactly specify the chosen statistical method.

- Caution with relative numbers per time step in statistical measures! – Any nominator can lead to outliers.

Thank you for your attention!



Fraunhofer Institute for Solar Energy Systems ISE

Wendelin Sprenger

www.ise.fraunhofer.de

wendelin.sprenger@ise.fraunhofer.de

Fraunhofer
ISE