Finding a Representative year

Kabitri Nag   Elke Lorenz
Alexander Kies   Lüder von Bremen
Detlev Heinemann

Oldenburg University

October 13, 2014

CARL VON OSSIETZKY universität OLDENBURG

Overview

In order to find a representative year, the following analyses were performed:

1. Interannual fluctuation of capacity factors for different technologies
2. Kolmogorov-Smirnov Integral (KSI), the area between curves, for power timeseries compared to the reference ones
3. Interannual fluctuation of standard deviations for different technologies
4. Analysis of incremental power timeseries on different temporal resolutions using KSI
5. Finally use of statistical hypothesis testing for decision making at a given confidence level

Overview of the analysis

Different means are adopted for the analysis to elaborate specific characteristics:

1. analyses are made to understand inter-annual fluctuations
2. variations are analysed over power production as well as over capacity factor for better comparison among different technologies
3. fluctuations are estimated for timeseries of individual years as well as in relation to the characteristics of the reference timeseries
4. incremental timeseries at different temporal resolutions are analysed to better understand the changes from one time step to the next
5. use of cumulated distribution functions to estimate the range of values
6. frequency of occurence of values studied via analysis of histograms

CARL VON OSSIETZKY universität OLDENBURG

Analysis of interannual fluctuation of capacity factors for different technologies

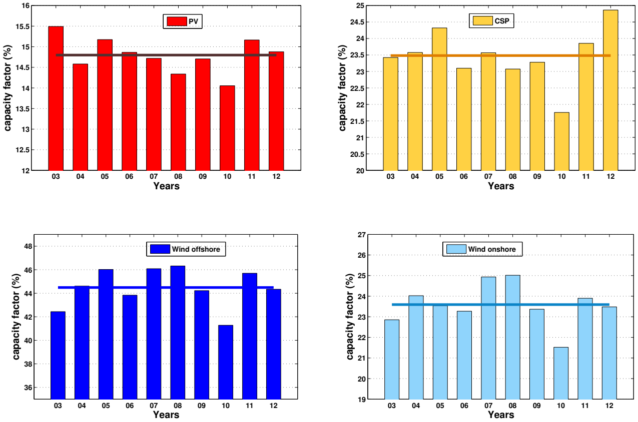
Data description of capacity factor

- Capacity factor is defined and computed as:

$$CF|_{hourly} = \frac{\left(\sum_{j=1}^{365}\sum_{i=1}^{24} P_{ij}\right)}{365 \times 24 \times P_{nom}}$$

- Capacity factors are calculated from hourly timeseries of Europe
- Values of individual years are compared against the capacity factor of the reference timeseries which is computed through the entire duration of our analysis (2003-2012)
- In general, 2004, 2006, 2009 & 2011 show close proximity to the reference while 2010 lies furthest away from the reference for all technologies

Fluctuations of capacity factors



In general, 2004, 2006, 2009 & 2011 show close proximity to the reference for all technologies.

Analysis of Kolmogorov-Smirnov integral for power timeseries compared to the reference ones

Kolmogov-Smirnov statistics

- serves as a measure of 'goodness of fit' by comparing the cumulative distribution functions of the samples
- The Kolmogorov-Smirnov statistic for a given cumulative distribution function F(x) is:

$$D_n = \sup_x |F_{n_2}(x) - F_{n_1}(x)|$$

  where $\sup_x$ is the supremum of the set of distances.

- this KSI statistics is applied to individual technologies and at different temporal resolutions
- to maintain clarity for better understanding, only a few example years will be shown

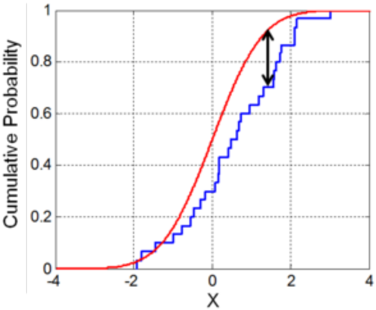


Image courtesy: Wikipedia

'A measure of 'Goodness of Fit'

- To find a representative year which can potentially reproduce the characteristic features of the reference timeseries, a statistical measure is adopted: the Kolmogorov-Smirnov integral (KSI)
- A measure of 'goodness of fit' is computed using this as:

$$A = \left| \sum_{i=1}^{n} (F_{n_2}(x) - F_{n_1}(x)) \right| \times dx$$

  where F_{n_1} and F_{n_2} are the cumulated distribution functions of the reference timeseries and the specific year to analyse, respectively. dx is a constant value specific to each distribution type.
- Hence, lower value of A indicates that the corresponding distribution lies closer to the reference year and vice versa.
- For comparing different technologies, all timeseries are normalised with their respective installed capacities
- For PV and CSP, night-time values are filtered out from the hourly timeseries to remove the effects of strong deterministic course of the diurnal solar variability

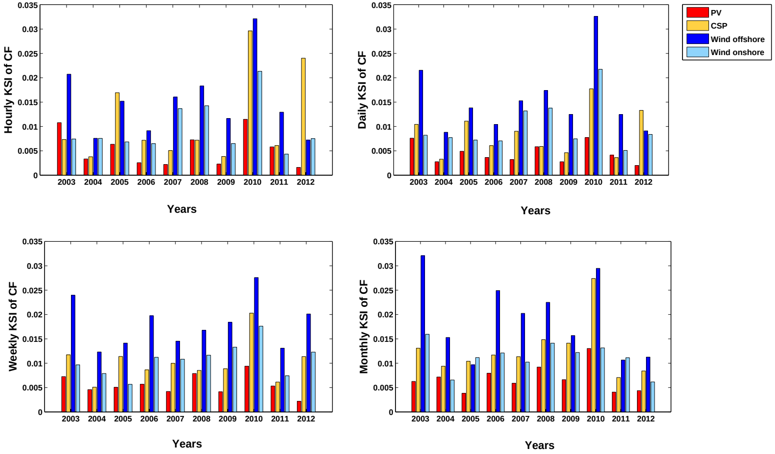
Interpretation of KS integral

The Kolmogorov-Smirnov integral computes the area between the cumulated distribution function of the reference year and of the particular year in question.

- Close proximity of the curves produces low value of the KS integral and is interpreted that the particular year closely resembles the reference timeseries and may serve good as a *'Representative year'*.
- On the other hand, if the cumulated distribution functions lie further apart from each other producing high values of the KS integral, the result is interpreted as lack of resemblance between the reference timeseries and that of the particular year and it is not quite suitable to be used as a *'Representative year'*.

The Kolmogorov Smirnov integral is interpreted as the area in between the cumulated distribution curves. It does not have a any unit as it can be represented as the product of cumulated distribution [0 1] and the capacity factor (P/P_{nom}), which are both unit-less quantities.

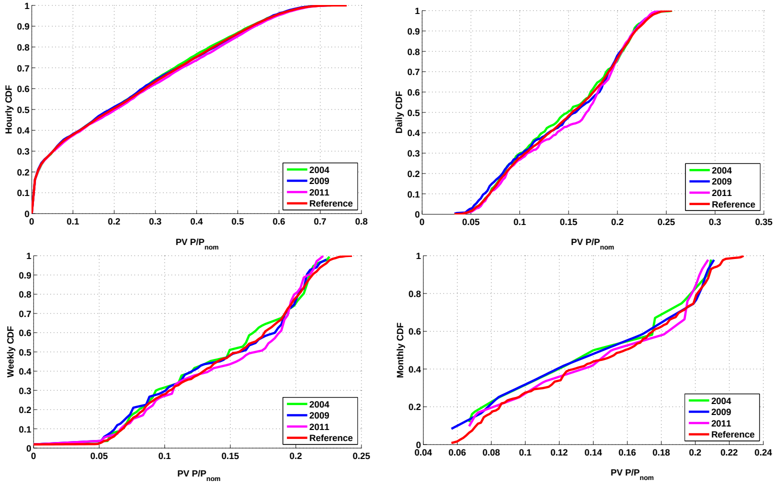
Variation of KSI values:



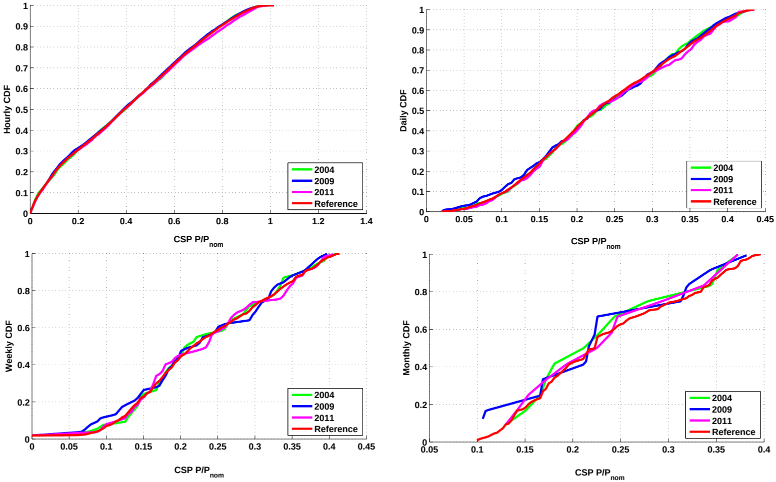
Analysis of KSI values:

- In general, 2004, 2006, 2009, 2011 show lower KSI values at all temporal resolutions compared to the other years that are analysed. However, offshore wind shows comparatively higher KSI magnitudes for 2006 (specially for weekly and monthly values) and 2009.
- 2005, 2007, 2008, 2012 show quite high values of KSI, specially for offshore and onshore winds. However, for 2005 and 2012, CSP technology shows particularly high values in hourly and daily series, where it even exceeds those of winds.
- 2010 shows very large values for each technology and at each temporal resolutions. These high values are next followed by 2003 which gives particularly high values for offshore wind regimes.

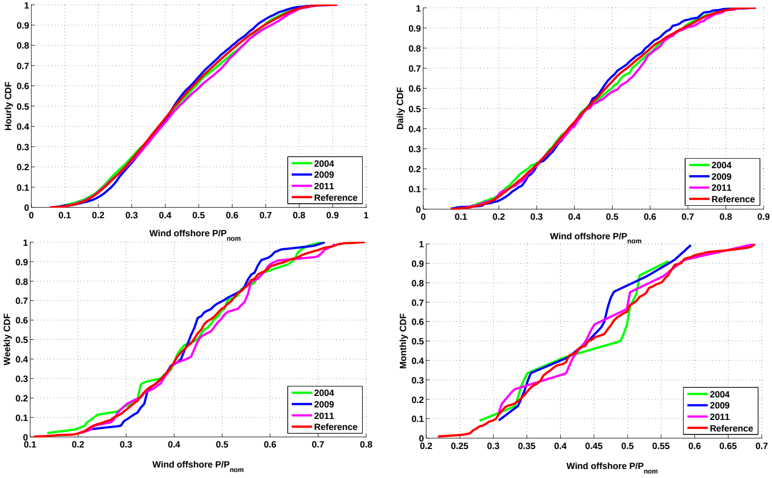
PV 2004,2009,2011:



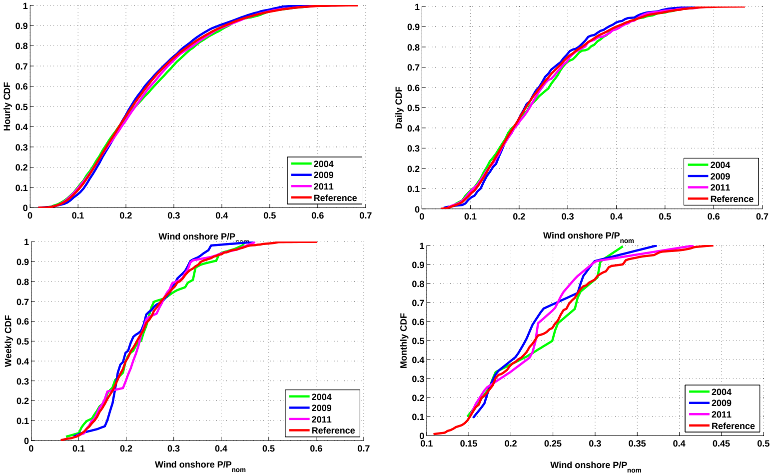
CSP 2004,2009,2011:



Offshore wind 2004,2009,2011:



Onshore wind 2004,2009,2011:



Overview of comparison

Based on the results shown before, the following table is created for comparing 2004, 2009 and 2011. The best year in each category is chosen based on the smallest KSI values obtained within these three years:

	PV	CSP	Offshore	Onshore
Hourly	2009	2004	2004	2011
Daily	2004	2004	2004	2011
Weekly	2009	2004	2004	2011
Monthly	2011	2011	2011	2004

Analysis of interannual fluctuation of standard deviations for different technologie

Assessment of fluctuations

Fluctuations are studied in terms of standard deviations with two-fold manifestations:

- they are estimated as deviations of capacity factors which are important to compare different technologies due to the normalisation made with-respect-to the nominal power of each technology

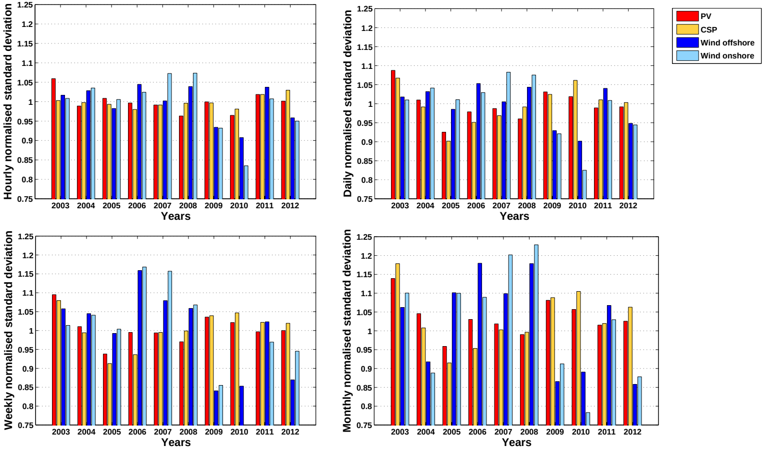
$$\boxed{result = std\left(\frac{P}{P_{nom}}\right)}$$

- they are also evaluated as normalised to the standard deviations of the reference timeseries as well as to the mean of reference timeseries:

$$\boxed{result_1 = \frac{std(P)}{std(P_{ref})}} \qquad \boxed{result_2 = \frac{std(P)}{mean(P_{ref})}}$$

It is to be noted that for PV and CSP, night-time values are filtered out from the hourly timeseries to remove the effects of strong deterministic course of the diurnal solar variability

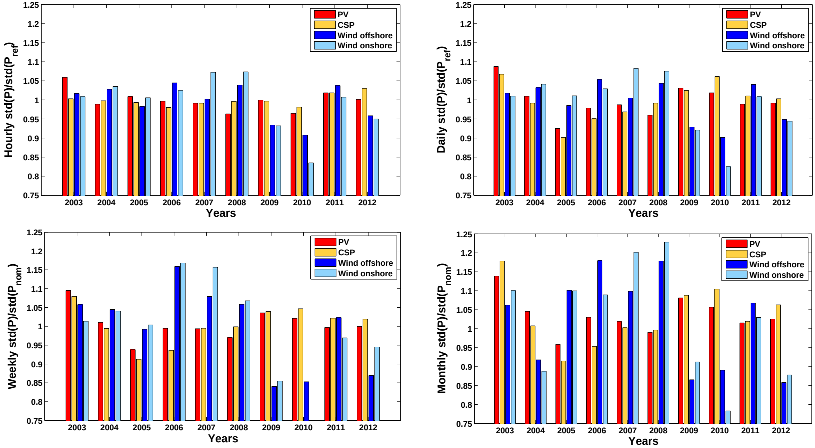
Standard deviations of CF



Analysis of fluctuations

- The standard deviations of capacity factor at all temporal resolutions considered here show higher fluctuations during 2004, 2006, 2007, 2008, 2011 and lower fluctuations during 2009 and 2010 for offshore and onshore wind
- On the other hand, solar technologies show relatively less prominent fluctuations over the 10-year time period at hourly resolution. The variability of solar PV and CSP is quite homogeneous in hourly scale. It should be noted here that the night-time values are filtered out for hourly solar fluctuation calculations.
- However, for daily, weekly and monthly resolutions, solar technologies show quite some fluctuations over the 10-year time period. This fluctuation is minimum for 2005 and maximum for 2003 at all resolutions, except for CSP where 2012 hourly fluctuation exceeds that of 2003.
- Focusing only on the specific years chosen so far (2004,2009 and 2011), one can see that for 2004 and 2011 almost all technologies were in similar range (~ 1) at all temporal resolutions. However, 2009 showed particularly low deviations for wind technologies, specially on weekly and monthly resolutions.

Normalisation with reference

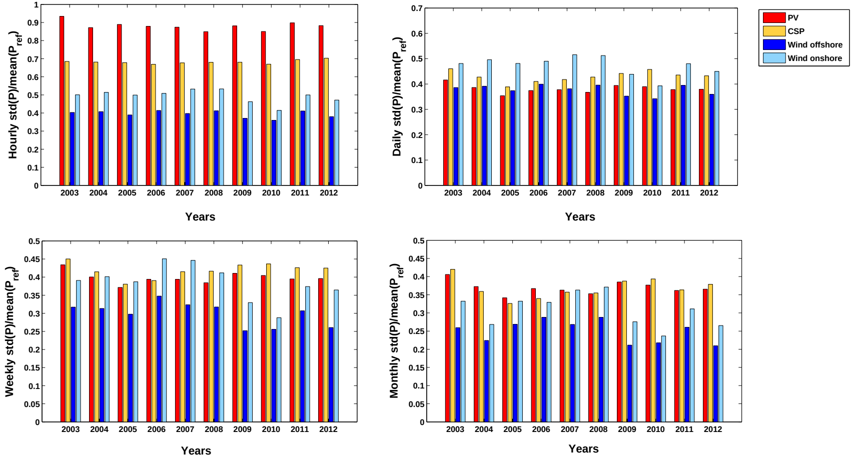


Analysis of fluctuations

The variation of the ratio of standard deviation of power to the standard deviation from the reference timeseries is discussed in this section.

- 2009 and 2010 show values lower than 1 for wind technologies at each temporal resolutions considered. This follows directly from the previous discussion where standard deviations of wind capacity factors are shown to have quite low values.
- On hourly and daily scales, offshore and onshore wind show 2004, 2006, 2007, 2008 have higher fluctuations compared to the reference timeseries.
- However, wind 2011 shows nearly equivalent fluctuations on all time scales as compared to the reference timeseries. Wind 2004, on the other hand show very low fluctuations on monthly scale.
- The fluctuations for solar technologies are primarily visible over daily, weekly and monthly time scales. For most cases, 2003, 2009 and 2010 show higher solar fluctuations compared to the reference.
- 2005 shows minimum fluctuations for both PV and CSP on all temporal resolutions except for the hourly timeseries, where it almost equals that of the reference.
- since 2004, 2009 and 2011 are already on top of the list chosen for finding a representative year, analysis of standard deviation of capacity factor and the ratio of standard deviations of the power to that of the reference timeseries suggest further filtration of 2009

Normalisation with 10 years mean

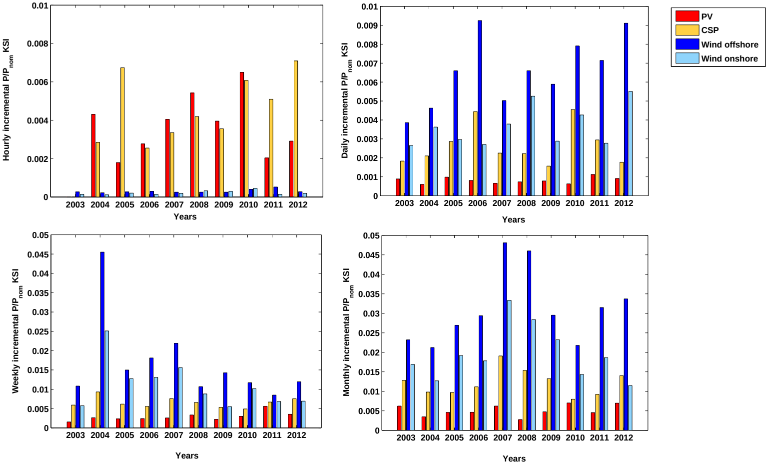


Analysis of incremental power timeseries at different temporal resolutions via calculation of their KSI value

Data description

- this section analyses how increments on each time step of each resolution for a particular year resembles the same from the reference timeseries
- the entire calculation is carried over power normalised to the respective installed capacity (P/P_{nom}) for better comparison between different technologies
- since changes in wind speed are usually not very drastic from one hour to the next, all wind technologies show lower magnitude of variation in this domain
- on the other hand, the strong diurnal variation associated with solar technologies, shows high fluctuations on hourly level, even after filtering out the night-time values
- however, simple removal of the nightime solar values may result in loss of information during dawn and dask, so the filter is applied on appropriate conditions

KSI of incremental P/P_{nom}

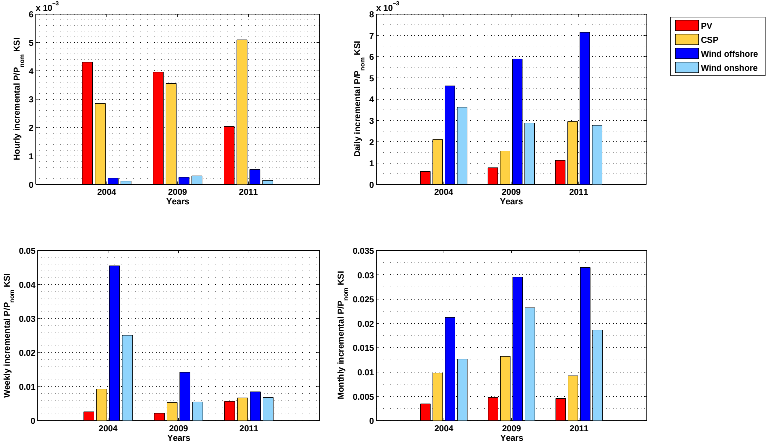


Based on the results shown before, the following table is created for comparing 2004, 2009 and 2011. The best year in each category is chosen based on the smallest KSI values obtained within these three years:

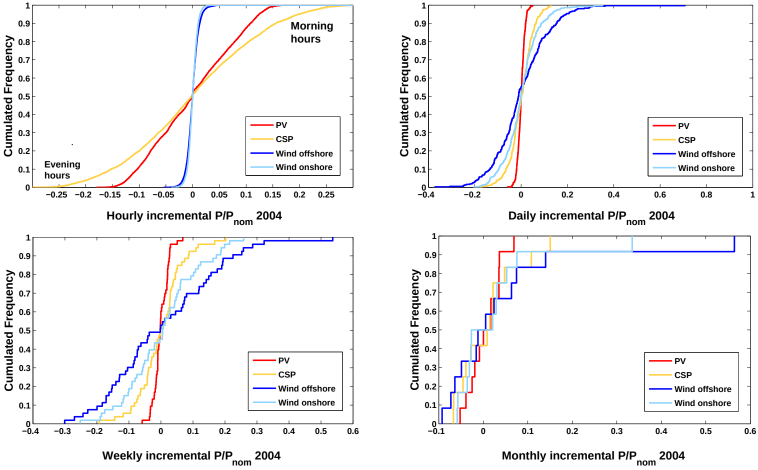
	PV	CSP	Offshore	Onshore
Hourly	2011	2004	2004	2004
Daily	2004	2009	2004	2011
Weekly	2009	2009	2011	2009
Monthly	2004	2011	2004	2004

It is to be noted that although this table depicts that 2004, 2009, 2011 share a ratio of 2:1:1, in most of the cases the differences in magnitude of KSI between 2004 and 2009 or 2004 and 2011 are very small. Hence, at this point, choosing 2004 over the other years is not quite recommended.

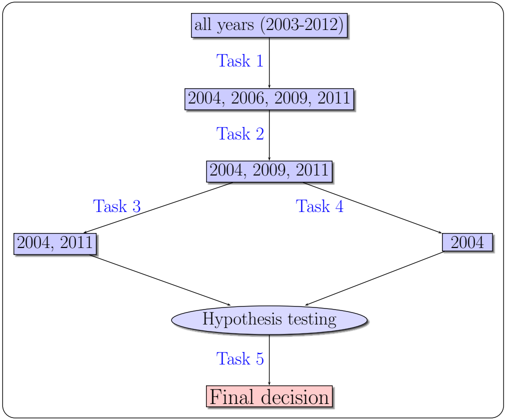
KSI of incremental P/P_{nom} for selected years



Incremental timeseries: 2004



A summary of our analysis until now yields the following results and suggests the need of further statistical testings to obtain the final result:



Kolmogorov-Smirnov hypothesis testing at a specific confidence level

The statistical measure of Kolmogorov-Smirnov integral can be extended to find the confidence limits on the cumulated distribution itself.

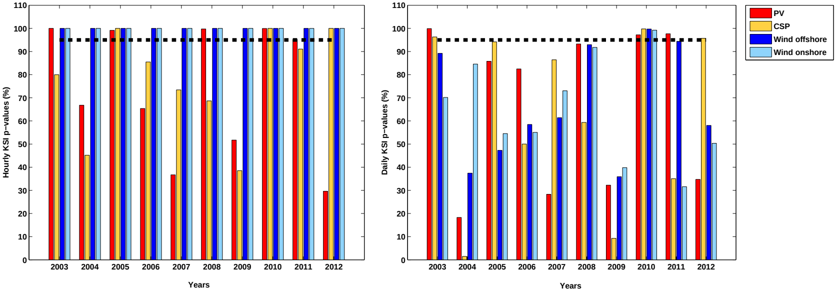
- a critical value of the test statistic D_a can be chosen such that $p = P(D_n > D_a)$
- a band of width $\pm D_a$ around $F_n(x)$ which entirely contain the distribution will have probability $1 - p$
- then using a predefined confidence interval α (in this case 95%) one can check if the samples belong from the same distribution or not
- this leads us to a statistical hypothesis testing with the following:

$$decision = \begin{cases} \textit{retain null hypothesis} & \text{if } p < \alpha \\ \textit{reject null hypothesis} & \text{otherwise} \end{cases}$$

  here the null hypothesis is that both samples are drawn from the same distribution
- the test has been performed on normalised power timeseries for each technology, at each temporal resolution, along with the use of a filter on hourly data sets of solar renewables to remove night-time values

KSI test result

KSI test results for hourly and daily normalised power timeseries at 95% confidence level:



At 95% confidence level, both weekly and monthly data sets show retaining of null hypothesis, hence not shown here explicitly.

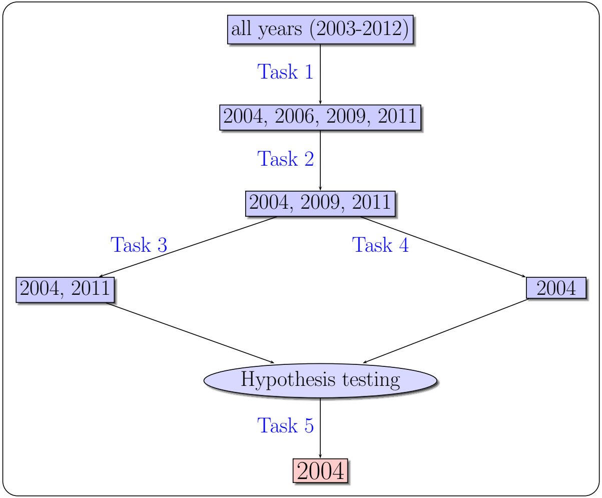
Overview of KSI test result

The following table presents an overview of the results of KSI tests for each technology at each temporal resolutions. Only 2004, 2009 and 2011 are analysed here. Only years for which the null hypothesis is retained, are shown here.

	PV	CSP	Offshore	Onshore
Hourly	2004, 2009	2004, 2009, 2011		
Daily	2004, 2009	2004, 2009, 2011		
Weekly	2004, 2009, 2011	2004, 2009, 2011	2004, 2009, 2011	2004, 2009, 2011
Monthly	2004, 2009, 2011	2004, 2009, 2011	2004, 2009, 2011	2004, 2009, 2011

This suggests a choice between 2004 and 2009, and rejection of 2011. Since 2004 was also selected from the analysis of task-4, we propose 2004 as the 'Reference year'; However, it should be noted that 2011 also has quite the potential to reproduce most features of the reference timeseries, and in certain cases, even better than 2004 (for example, weekly KSI values of incremental P/P_{nom}). So, although a suggestion has been made here, we request the readers to carefully go through the examples cited above and to adapt the year that is most reasonable for their particular analysis and desired resolution.

Final result through analysis



Thank You!!!

For further questions, please contact
kabitri.nag@uni-oldenburg.de