Renewable Energy 117 (2018) 474–487

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Impact of different time series aggregation methods on optimal energy system design

Leander Kotzur [a, *], Peter Markewitz [a], Martin Robinius [a], Detlef Stolten [a, b]

[a] Institute of Electrochemical Process Engineering (IEK-3), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Str., 52428 Jülich, Germany
[b] Chair for Fuel Cells, RWTH Aachen University, c/o Institute of Electrochemical Process Engineering (IEK-3), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Str., 52428 Jülich, Germany

A R T I C L E I N F O

Article history:
Received 22 May 2017
Accepted 6 October 2017

Keywords:
Energy systems
Renewable energies
Mixed integer linear programming
Typical periods
Cluster analysis
Extreme periods
Time-series aggregation

A B S T R A C T

Modeling renewable energy systems is a computationally-demanding task due to the high fluctuation of supply and demand time series. To reduce the scale of these, this paper discusses different methods for their aggregation into typical periods. Each aggregation method is applied to a different type of energy system model, making the methods fairly incomparable.

To overcome this, the different aggregation methods are first extended so that they can be applied to all types of multidimensional time series and then compared by applying them to different energy system configurations and analyzing their impact on the cost optimal design.

It was found that regardless of the method, time series aggregation allows for significantly reduced computational resources. Nevertheless, averaged values lead to underestimation of the real system cost in comparison to the use of representative periods from the original time series. The aggregation method itself e.g., k-means clustering plays a minor role. More significant is the system considered: Energy systems utilizing centralized resources require fewer typical periods for a feasible system design in comparison to systems with a higher share of renewable feed-in. Furthermore, for energy systems based on seasonal storage, currently existing models integration of typical periods is not suitable.

1. Introduction

Developing an energy system design that minimizes costs and environmental impact is a complex task due to the spatial and temporal gap between energy production and demand. In consequence, optimization algorithms are required for solving these design problems [1–11].

However, the algorithms used hitherto are computationally demanding: The size of the input data directly influences that of the related optimization problem, and with it the requirement for processing resources. For this reason, it is often necessary to simplify the design problem in advance.

Therefore, different options for complexity reduction exist and include: Spatial aggregation which reduces the number of nodes in an energy system network [12]; simplifying the technology models by reducing nonlinearities or discontinuities so as to avoid non-convexity of the program [13,14]; and temporal aggregation, which creates typical periods representing the original input time series.

The creation of recurring periods is popular because of the existing patterns in the hourly, daily and seasonal variation for the majority of design relevant time series. Therefore, it is reasonable to reduce redundant data until the minimal representative data set required for the problem is reached. Fig. 1 visualizes this redundancy by showing the result of a Fast Fourier Transformation (FFT) of different time series that are typically required for an energy system design. The frequencies with the highest amplitudes are highlighted and are, as anticipated, the daily and annual variations.

For this reason, many different methods for the selection of typical periods have been presented. Aside from custom exact optimization methods [17,18], and graphical methods [19], the majority use heuristic methods or greedy clustering algorithms for the aggregation of typical periods. Creating representative days by averaging time series, for example over a type of day defined by month or weekday, has been popular [20–23]. [24] refers to it as time-chronological averaging. Recent attempts use the k-means clustering [25–28], hierarchical clustering [29,30], or k-medoids clustering - either based on a greedy algorithm [31,32] or an exact

* Corresponding author.
E-mail address: l.kotzur@fz-juelich.de (L. Kotzur).

https://doi.org/10.1016/j.renene.2017.10.017

Nomenclature

Variables

δ	Binary variable determining the existence of a technology
μ	Representative values of a typical period
D	Scaling of a device
E	Energy flow between two devices
SOC	State of charge
y	Binary variable determining if the candidate period is a cluster center
z	Binary variable determining candidate cluster assignment

Parameters

Δt	Duration of a single time step
η	Efficiency
τ	Lifetime
C	Set of periods inside a cluster
L	Set of device connections
N	Size of an index set
x	Normalized candidate value

Subscripts and sets

ε	Energy type
f	Index of the Transformer class
n	Index of the Collector class
q	Index of the Source/Sink class
s	Index of the Storage class
a	Attribute represented by a time series
d	Considered device or technology
g	Time step index inside a period
i	Candidate period index
k	Typical period index
t	Time step index of the full series

Abbreviations

CRF	Capital Recovery Factor
LB	Lower bound
UB	Upper bound
CAPEX	Specific capital expenditure
CHP	Combined Heat and Power plant
GHI	Global Horizontal Irradiance
OPEX	Specific fix operational expenditure
RMSE	Root Mean Squared Error
WACC	Weighted Average Cost of Capital

solution of a MILP [33,34] - for the selection of typical periods. Nevertheless, each method is applied to a different system and it is difficult to identify which is the most suitable. While Schuetz et al. [35] compare methods for building energy systems, they note that future research should focus on the appropriateness of clustering algorithms for different design applications. Moreover, the period

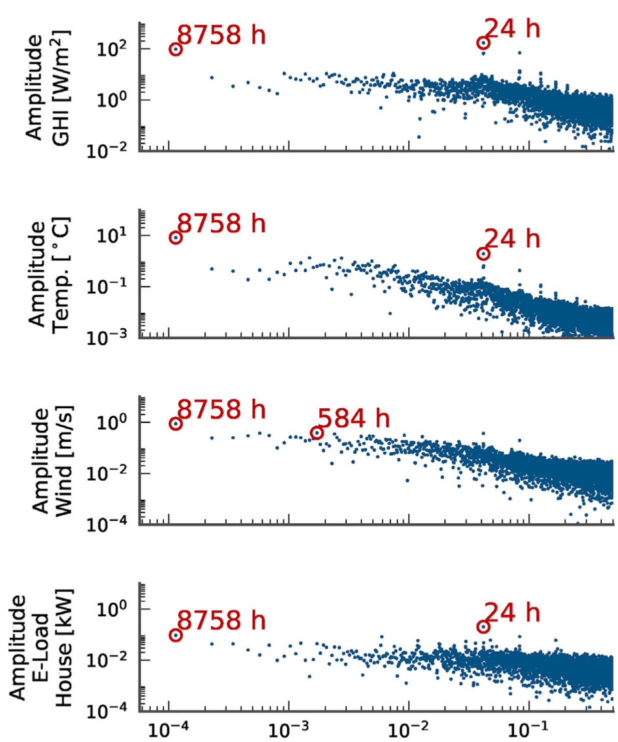

Fig. 1. Fast Fourier Transformation of the Global Horizontal Irradiance (GHI), the temperature and the wind speed of a test reference year (Location: Bad Marienberg, Germany) [15] and a representative electrical load profile of a residential building (Profile 1) [16].

length should be varied so as to assess the impact of storage effects.

A further difficulty is that the system considered determines the minimal required dataset. For renewable energy systems, a higher resolution of the input times-series is required than with their fossil counterparts [36]. For conventional system design, it could be sufficient to reduce the dataset to a few time steps [27], while for a storage based system design different typical weeks are required [30,37].

In summary, the following open research questions present themselves:

- Which time-series aggregation method is best suited for which energy system design application?
- What is the minimum number of aggregated time steps to model such a system?
- What is an appropriate period length - typical days or typical weeks?

To answer these, this paper is structured as follows: First different deterministic methods including *k-mean* clustering, *k-medoids* clustering and *hierarchical* clustering as aggregation methods are presented in section 2, where the possibilities of adding extreme periods are also discussed. In section 3, the aggregation methods are used to select four typical days of different time series that could be relevant for an energy system design. The aggregated profiles are then graphically analyzed and through accuracy indicators. In the following, the different methods are applied in section 4 to three design optimization problems of a heat and electricity supply system:

1. A cogeneration unit with a heat storage as benchmark system
2. A residential system based, amongst other elements, on photovoltaics and a heatpump
3. An island system with a high share of renewables with the support of different storage technologies

To validate the methods, the results for different numbers of