

# A high-fidelity weather time series generator using the Markov Chain process on a piecewise level

**K Hersvik and O-E V Endrerud**

Shoreline AS, Stavanger, Norway

E-mail: [hersvik@shoreline.no](mailto:hersvik@shoreline.no) / [kjetil@hersvik.net](mailto:kjetil@hersvik.net)

**Abstract.** A method is developed for generating a set of unique weather time-series based on an existing weather series. The method allows statistically valid weather variations to take place within repeated simulations of offshore operations. The numerous generated time series need to share the same statistical qualities as the original time series. Statistical qualities here refer mainly to the distribution of weather windows available for work, including durations and frequencies of such weather windows, and seasonal characteristics. The method is based on the Markov chain process. The core new development lies in how the Markov Process is used, specifically by joining small pieces of random length time series together rather than joining individual weather states, each from a single time step, which is a common solution found in the literature. This new Markov model shows favorable characteristics with respect to the requirements set forth and all aspects of the validation performed.

## 1. Introduction

### 1.1. Problem description

Weather conditions are a key limiting factor when it comes to offshore operations. For a specific work task to be carried out e.g. in an offshore wind farm, a certain weather window is needed. The weather window is described by a set of operational weather criteria over a certain time span, also known as the window length. The weather criteria in this study include maximum mean wind speed and maximum significant wave height. Detailed analysis of weather window lengths has high economical value. A study of the impact of uncertainty in repair time for offshore wind farm maintenance planning [1] confirms the importance of studying weather uncertainty.

To find suitable weather windows for a certain offshore operation we can use weather forecasts, however forecasts are only applicable when doing short term planning. When doing long term planning, e.g. for wind farm installation or lifespan maintenance operations, weather forecasts are not available. Instead, historical weather data for the relevant site need to be used, and predictions need to be made based on a large set of possible weather scenarios.

### 1.2. Weather data time series

The historical weather data is stored as a time series. A weather time series consists of measurements at each time step throughout a selected period of time. The measurements at each evenly spaced time step record a predefined set of weather parameters. The measurements can contain observations such as



atmospheric pressure, ocean current speeds, visibility, etc., however the type of time series being processed in this study contains observations of the mean wind speed and the significant wave height.

### 1.3. Aims

As described above, to make the best possible predictions for the future in a statistical way, a somewhat large and appropriate variety of possible weather outcomes is needed. The historical weather data time series is an excellent starting point to use for simulation purposes, however it is desirable to have an endless set of unique, realistic weather data time series for repeated simulations. The generated weather data time series must on average represent the historical weather data well. This has in part been achieved previously with a standard Markov chain model [2]. One major challenge with the standard Markov chain model is related to the discretization of weather parameters into weather state categories, as explained more in section 3.3 “Classification of weather states” below.

The loss of details when categorizing weather parameters into a set of weather states can be made non-relevant if the weather state transition thresholds are aligned with the weather criteria limits. For example, if the significant wave height  $H_s$  for a certain operation needs to be below  $H_s = 2$  m, then as long as the weather states distinguish clearly when transitioning across  $H_s = 2$  m, then it becomes non-relevant for that specific operation what the exact  $H_s$  value is. However, it is not always easy to tailor the categorization threshold levels. For instance, when analyzing composite offshore operations, many different weather criteria are involved, such as access criteria for different types of vessels, lifting operations, etc. Thus, it can become challenging to tailor the categorization limits according to all the different weather criteria involved in a specific study. Increasing the detail level by using a very fine-grained categorization of weather states is not necessarily a good idea either. It can result in too many weather states, causing too few transition options between each of the weather states.

The generated time series is traditionally limited in detail because of the categorization of weather parameters into weather state categories. The aim of the Markov model presented in this paper is to conserve output details that are normally lost within each weather state category. Although building on the favorable computational strategy from [3], as described in section 3.2 “The proposed Markov model”, improvements were also sought with respect to the computational efficiency of the Markov chain implementation.

## 2. Markov Chain theory

The Markov chain concept is fully relevant for the Markov model in this paper. Markov chains are stochastic chains which fulfill the Markov property. The term stochastic refers to the probabilistic or random nature of a process. The Markov property defines that the probability of a certain state appearing as the next element in the chain is dependent only on the current state in the chain [3]. In other words, when selecting the next state in the chain, only the current state is needed to determine the probability distribution for selecting the next state, and the elements before the most recent element in the chain are ignored. Note that the Markov property does not imply that the coming element in the chain is uncorrelated with respect to the elements that are further away than the nearest neighbor. The probability of a certain state appearing two steps down the road can be calculated using the probability distribution for the first next state combined with the probability distributions for the second next state which are conditional on the first next state.

How does the Markov property fit with the physical real world weather system? The real-world weather system is known as a chaotic system. Chaotic systems are deterministically defined given an exact starting condition for the system, despite their appearance as random. However, the exact condition of the global weather system is far out of reach of modelling, and the available data for a certain analysis will certainly be a far greater limitation. Given the limited data available for each time step in the historical data, it can be tempting to make that next weather state dependent on more than just the current weather state when defining the model to be used. In the end, nonetheless, it is the statistical properties of the output which matter, and using the principles from a Markov chain model proves efficient and suitable.

### 3. Method

#### 3.1. Input data – Historical weather series

As described in the introduction, the historical weather data serves as input and consists of weather measurements for each time step over several years. The time step  $\Delta t$  is typically 1 or 3 hours, and is the average measured quantity over that time step. The weather measurements always include significant wave height  $H_s$  and mean wind speed  $U$ , and more weather parameters can be included in the algorithm.

#### 3.2. The proposed Markov model

The Markov model algorithm developed consists of two main parts. The first half parses and analyzes the historical weather data. The last part combines appropriate weather sequences stochastically into a whole new time series, which means using weather transition probabilities determined in the first part of the algorithm.

Again, the first part of the algorithm parses through the historical weather file. It categorizes the weather state at every time step. This is done by classifying the wave height into a category defined by a maximum and minimum value of significant wave height. Equivalently, the mean wind speed is classified into a category of wind speed magnitude. The two categories together define the weather state. More details are described in the “Classification of weather states” section below.

When using the traditional state-by-state Markov process to generate new time series as described in [3], it is common to describe the process by transition probabilities. These transition probabilities can be calculated while parsing the historical time series file, and they can be stored in a large matrix. The matrix will be  $n$  by  $n$  large, where  $n$  is the number of unique weather states available in the mentioned weather state classification. For any reasonable weather state resolution, that matrix will easily be too large to store in computer memory. Adding a few gigabytes of memory will not help much since the memory requirement is proportional to  $n^2$ . It is a highly sparse matrix, meaning that the matrix contains for the most part elements with value zero. On the other hand, storing the entire historical time series in memory is far more manageable. It simply means storing a copy of the entire weather data file in memory.

So even though describing the process with transition probabilities is correct and useful for understanding the process, the entire calculation of the transition probabilities can be omitted. Instead, the historical time series can be searched in a strictly random fashion with the right constraints, and the probability of the next weather state will be correctly represented. The process takes place in the last part of the algorithm. This computational strategy was preferred in [3] and is even more suitable with the new piecewise level of using the Markov process which is described in the last part of the algorithm.

The last part of the algorithm starts out making a new time series by selecting a random year from the historical time series and a random start date which is close to the desired start date of the generated time series. A small piece of the historical time series is then copied from the selected starting point and used as the first part of the generated time series. The small piece copied from the historical time series is set to have a random length, uniformly distributed with minimum and maximum length defined to be 6 - 48 time steps, which is conversely 6 - 48 hours in most cases.

The exact choice of 6 – 48 time steps as the building block length is a design choice, which was tried out and found to balance characteristics of the output in a good way. The balance is mainly considering variability in the output vs. keeping imperfections of joining blocks together from affecting the natural weather window distribution. The end validation is needed to confirm if both these considerations are properly achieved. Smaller and larger blocks are assumed to also work well and produce similar quality output, and keeping the block sizes a random length reduces the possibility for systematic impurities. The block sizes do affect the performance of the algorithm slightly since larger block sizes result in fewer search and join operations, however performance does not influence the choice of block sizes for the current application.

Next, a new sample from the historical time series is drawn at random with two restrictions. The first restriction ensures that the new sample starts out with a weather state which is selected from one of the possible state transitions, according to the historical data set. This is ensured by selecting one of the existing weather transitions from the historical file directly. It is a critical criterion for maintaining realistic weather development in the generated time series. It is equivalent to the Markov Chain process, in which the next state is drawn by transition probabilities which depend on the current state. Specifically, the historical weather file is traversed in a random order, looking for an historical point in time where the weather state is identical to the last weather state of the time series being generated. The weather state immediately after the first random matching historical state is selected as the next random weather state. However, for the piecewise process level in this algorithm, the next 6 or up to 48 data points are selected together to become the next piece of the new time series.

As the first search restriction ensured that the new piece for the generated time series matches the previous piece, the second restriction ensures seasonal correctness in the new time series. The second restriction ensures that the new piece originates from the same calendar month as the calendar month it is representing in the generated series, although it can originate from any year.

If no matching weather state is found within the historical data for the specific calendar month, a transition is selected without restriction on calendar month. This can be relevant in transitions from one month to another, in combination with a rare weather state. In that case, there will be at least one matching transition in the historical time development, unless of course the sampling reached the very end of the entire historical file. If no transition is found at all, i.e. at the very end of the imported data, then a completely new state is drawn corresponding to the current season.

A new sample from the historical time series is repeatedly drawn in the same way, based on the preceding weather state, and added to the output time series.

### 3.3. Classification of weather states

The needed classification of weather states represents a discretization of weather parameters. The representation of each weather parameter becomes constricted to the set of available categories. An important reason for not using a too fine-grained classification of weather states is to avoid significant quantities of deterministic transitions. Deterministic transitions occur when a certain state only has one possible state transition to continue from, namely the next state in the historical data.

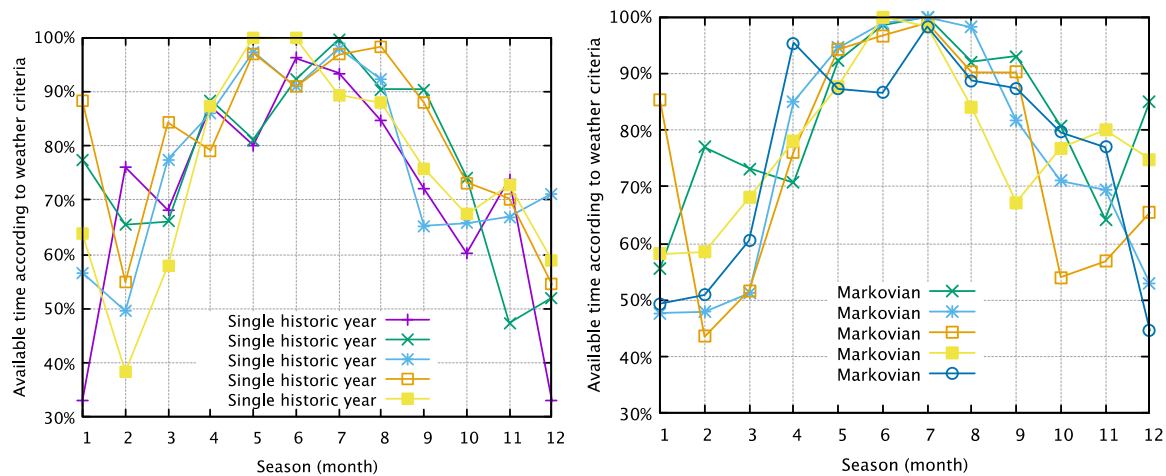
For the current Markov model implementation, the weather state categories are defined by the following. The step size for wave height categories  $\Delta H_s$  is  $\Delta H_s = 0.1 \text{ m}$  while the wind speed category step size  $\Delta U$  is  $\Delta U = 1.0 \frac{\text{m}}{\text{s}}$ .

The category steps are widened at the more extreme and rare ends of the spectra. The classification of weather states is both a necessity to solve the task and, at the same time, it cannot be very fine-grained in order to keep the random nature of the process, as already discussed. The main novelty of the developed piecewise approach to using the Markov chain process is to produce time series which display weather details beyond the weather state classification. This can be explained by giving an example from the traditional Markov chain model. In the traditional model, the resolution of the weather measurements is limited by the category step size in the weather state classification. For the current setup, that would result in wind speed values stepping 1 m/s at a time, i.e. no intermediate values.

The core idea in the new development is to process small pieces of the time series as one unit, i.e. batches of a few time steps, with all uncategorized details preserved. This contrasts with traditionally processing individual data points from single time steps which are limited to representing only a general weather state category.

## 4. Validation analysis

The aim of the validation analysis is to determine both the degree of diversity among the generated weather series and how closely the group averages of key time series properties resemble the historical weather.



**Figure 1.** Seasonal available time for work for 5 single-year time series, historic data in left view and Markov model output in right view, given weather criteria:  $H_s < 1.5$  (m)  $\cap$   $U < 12$  (m/s).

The key properties relate to weather windows which are time slots available for work. These weather windows are all dependent on the weather criteria, which dictate at what times the certain work tasks can be performed or not, offshore. The criteria consist of maximum allowed significant wave height  $H_{s,MAX}$  and maximum allowed mean wind speed  $U_{MAX}$  for any specific work task to be possible.

The weather time series used for validation and presentation of results is an anonymous 21-year length weather data time series with time step  $\Delta t = 1$  hour, from a relevant site. The characteristics of the data series are illustrated graphically in the figures, where sets of generated time series are compared to the original data series. The algorithm has not in any way been adjusted according to the result of the specific validation performed here. Thus, any weather data time series can be used to reproduce the validation of the algorithm. The key to validating the algorithm is the comparison between generated and original time series, which is the purpose of all the figures.

A typical weather criteria ( $U_{MAX}$  and  $H_{s,MAX}$ ) is selected for the validation analysis presented here, and is displayed in the caption of each graph. The most common weather requirements are related to transferring personnel onto the wind turbines using for instance crew transfer vessels. The weather criteria chosen in this validation analysis is a typical criterion, but not related to a specific vessel or task. The significant wave height is typically the limiting factor for traveling offshore to work on wind farms, and most working limits are set between 1 and 2 m significant wave height ( $H_s$ ).

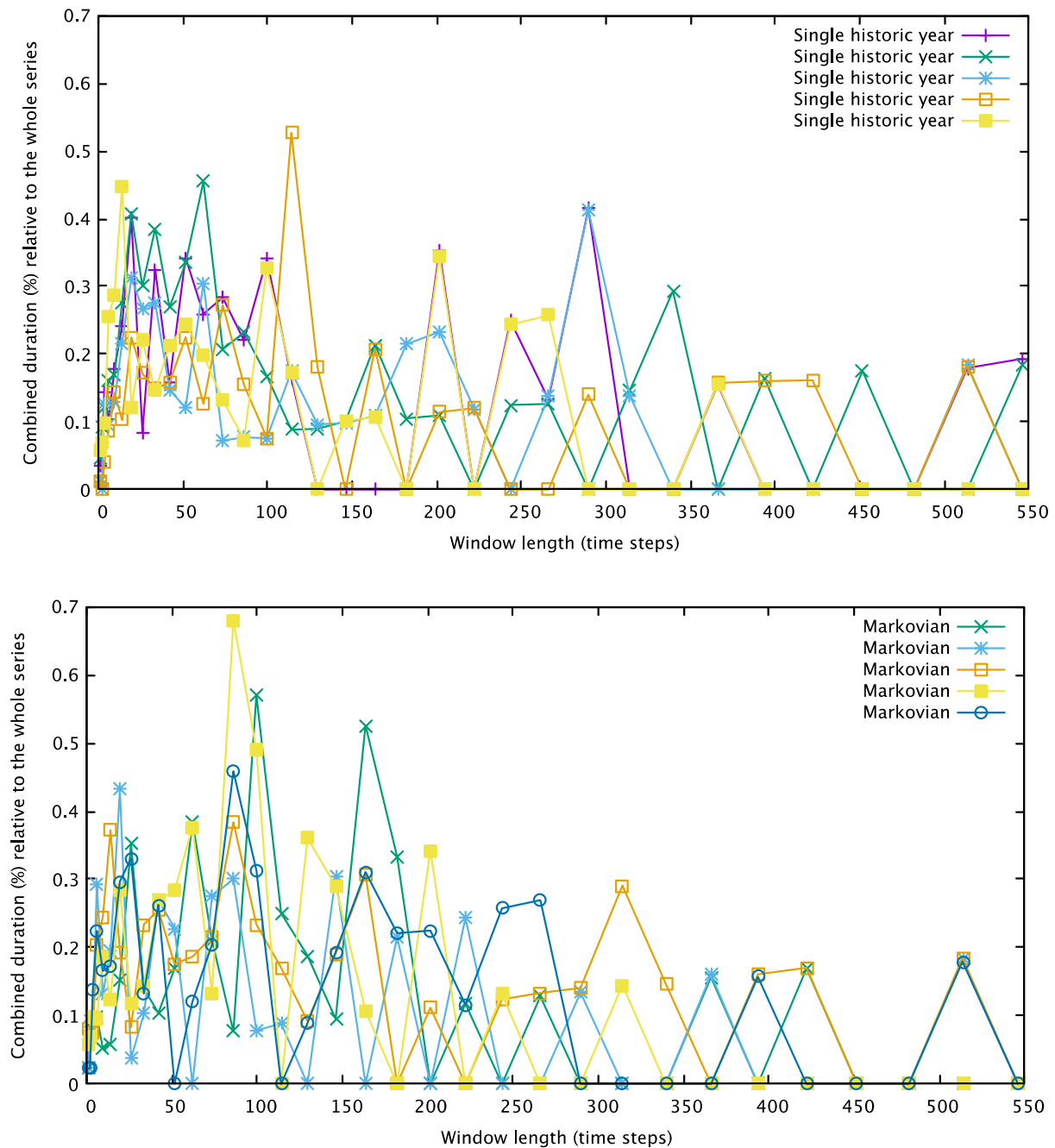
Regardless of these details, the success criterion of the algorithm is how well all the generated time series compare to the actual historical time series. That is why each of the following analysis graphs compare the generated time series directly to the original historical time series. The comparison is done for a chosen set of key properties.

Key properties of the time series include

- the proportion of time available for work,
- the distribution of weather windows according to window length metric,
- and the averages of these values for each season of the year as measured per calendar month (Figure 3).

Together, these properties give a well-defined picture of the time series on a group level.





**Figure 2.** Weather window distribution for individual 1-year series, displaying combined (sum of) weather window durations (as percentage of entire weather series) grouped and arranged as a function of individual weather window lengths, showing historical data in the top view and Markov model data in the bottom view, given weather criteria:  $H_s < 1.5$  (m)  $\cap U < 12$  (m/s).

Single-year time series are presented in Figure 1 and Figure 2. Single-year time series offer easy comparison between generated and historical time series because the multi-year historical time series can be broken down into several single-year time series. That way the variation between different historical time series can be compared to the variation within generated time series. Each of Figure 1

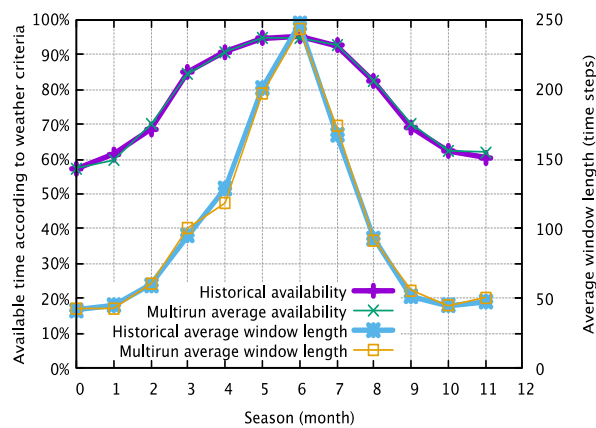
and Figure 2 consists of two separate graph displays. They contain results for historical time series and equivalent analysis results for the generated time series, respectively.

Figure 1 displays time available for work in sum for each season of the year. The left display contains results of 5 single-year historical time series, whereas the right display is based on an equivalent set of generated time series.

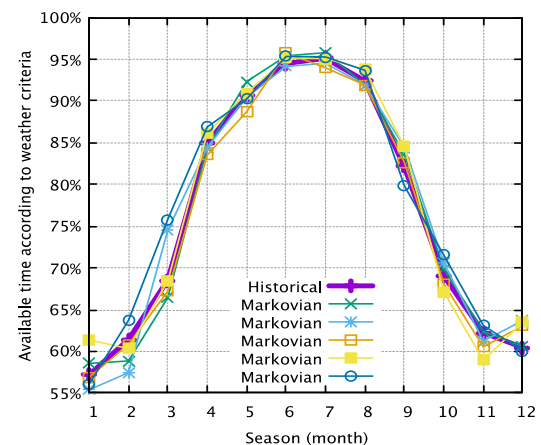
Figure 2 is based on the same set of single-year time series as in Figure 1. However, Figure 2 shows the combined window duration (sum of window lengths available for work) of all weather windows grouped and arranged by duration of the individual weather windows. Each time series is displayed with a separate graph line and legend indicator. The horizontal axis marks the size of the individual weather windows within the group, at each data point.

From the graphical comparison in Figure 1 and Figure 2, variations within historical and within modelled time series appear similar, which is a highly desired result.

Figure 3 compares average data from 100 generated time series to the historical data. The historical data and each of the generated time series consist of 21 years of weather data. The data has an hourly time resolution, which means a time step value  $\Delta t = 1$  h.



**Figure 3.** Seasonal average available time for work and average window length – historical time series compared to average of 100 runs, given weather criteria:  $H_s < 1.5$  (m)  $\cap$   $U < 12$  (m/s).



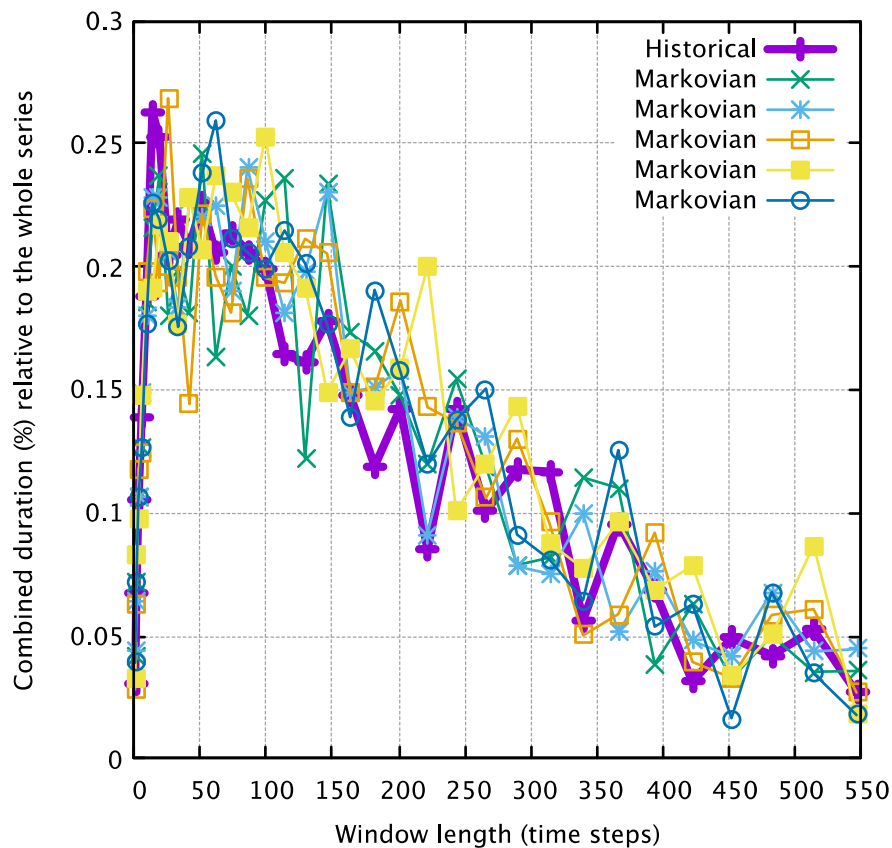
**Figure 4.** Seasonal average available time for work, comparing five 21 year runs to the equally long historical time series, given weather criteria:  $H_s < 1.5$  (m)  $\cap$   $U < 12$  (m/s).

The modelled time series do not distinguish daytime and nighttime. As mentioned in [3], some wind and sea state series have daily components, however these potential aspects are not considered relevant for the current applications of this Markov weather model, namely applications which are simulation of offshore wind farm operations.

Validation analysis of the algorithm shows that in each generated time series, approximately 99 % of the stochastic time series segments (6-48 entries in duration) were joined differently compared to in the historical time series. This applies to the 21-year data and for the granularity of state classification specified above, which is a 0.1 m resolution of wave height and 1 m/s resolution of wind speed. It proves a very good random variability in the output.

Figure 4 shows available time each season according to the weather criteria for a selection of five full 21-year output series. The historical reference values included are identical to the reference data in Figure 3.

Figure 5 shows a distribution of the available time according to individual window length. The durations of similar length weather windows are combined and distributed on the x-axis in the same way as Figure 2.



**Figure 5.** Combined duration (sum) of weather windows from six 21-year series distributed according to individual window length, given weather criteria:  $H_s < 1.5$  (m)  $\cap$   $U < 12$  (m/s).

The area under the curve gives a measure of time of the whole series covered by each interval of window lengths. Again, five full 21-year modelled time series are plotted together with the direct representation of the historical data. The combined view of six time 21-year series allows an impression of the variability between runs.

## 5. Conclusion

The improved Markov Chain model is developed for generating unique weather time-series based on an existing weather series. Despite the undesirable but necessary discretization of weather data into weather states, the method successfully recreates high detail levels in the generated time series. This is achieved by applying the Markov process using very small segments of the time series as building blocks instead of using single data points. That way the natural development within a single weather state category is recreated in the generated time series. The validation performed on the developed Markov model also shows excellent results.

In summary, all aspects of the generated time series are quite indistinguishable from original ones, both in terms of degree of variability between individual time series, as well as group averages. Even when graphing out actual parts of a time series, which is omitted here, it is quite non-trivial to distinguish a modelled time series from an original time series. This is in strong contrast to the traditional Markov chain output, which coarsely outputs average of each weather state. For example, the traditional Markov chain output would in an equivalent setup produce wind speed values with steps of 1 m/s and no intermediate wind speed values.



For the purpose of weather windows available for work, historical and generated time series are seen as interchangeable, based on any of the validation findings known. That satisfies the primary aim of the algorithm.

The developed Markov model has been used with many historical weather files and in countless simulations of offshore wind farm operations during the last year. The computational performance is also improved by more than tenfold, as well as the complexity of the algorithm being simplified to a good extent, especially since no post-processing to smooth the output is needed. From all the use cases known, no issues have been identified so far which can question the validity of the algorithm.

## References

- [1] Seyr H and Muskulus M 2016 *J. Phys.: Conf. Ser.* **753** 09200.
- [2] Scheu M, Matha D and Muskulus M 2012 *Proceedings of the 22th International Offshore and Polar Engineering Conference*, pp. 463-8.
- [3] Hagen B, Simonsen I, Hofmann M and Muskulus M 2013 *Energy Procedia*, **35**, pp. 137-147.
- [4] Monbet V, Alliot P and Prevosto M 2007 *Probab Eng Mech* **22** (2), pp. 113-126.