Enter your title here

(A title should be specific, informative, and brief. Use abbreviations only if they are defined in the abstract. Titles that start with general terms then specific results are optimized in searches—delete these notes when done)

Enter authors here: A. B. Author1, B. C. Author2, and D. E. Author3

(List authors by first name or initial followed by last name and separated by commas. Use superscript numbers to link affiliations, and symbols \*†‡ for author notes. For example, X. Jones1\*, P. Smith1,2 Authors are individuals who have significantly contributed to the research and preparation of the article. Group authors are allowed, if each author in the group is separately identified in an appendix.)

1Affiliation for author 1.

2Affiliation for author 2.

3Affiliation for author 3.

(Affiliations should be preceded by superscript numbers corresponding to the author list. Each affiliation should be run in so that the full affiliation list is a single paragraph.)

Corresponding author: first and last name ([email@address.edu)](mailto:email@address.edu))

†Additional author notes should be indicated with symbols (for example, for current addresses).

(include name and email addresses of the corresponding author. More than one corresponding author is allowed in this Word file and for publication; but only one corresponding author is allowed in our editorial system.)

Key Points:

* List up to three key points (at least one is required)
* Key Points summarize the main points and conclusions of the article
* Each must be 140 characters or less with no special characters or acronyms.

(The above elements should be on a title page)

Abstract

Principal Component analysis following by auto-regressive model gives us a reasonable regional model, and captures most of the variance in the first several important PCs. PC1 represents the mean regional behavior, and PC2 represents a regional pattern. Prediction with PC gives us a good result up to 10 days.

MS-AR model has a good performance in modeling the 30 sites with reasonably memory and can present trustworthy simulation for wind speed and win energy. Also we are interested in site by site risk of persistent outages below some threshold from a risk analysis point of view which MS-AR model can do the job pretty good while HMM model does not.

The abstract should be a single-paragraph of less than 250 words, or for Geophysical Research Letters, less than 150 words. A good abstract sets the general question or topic that you are studying for the general reader, provides background on the specific question or problem, briefly describes key data or analyses, and describes the key results and uncertainties. Please avoid acronyms or if used, define them.

**Plain Language Summary**

This is optional but will help expand the reach of your paper. Information on writing a good plain language summary is available [here](http://sharingscience.agu.org/creating-plain-language-summary/).

1 Introduction

The main text should start with an introduction. Except for short manuscripts (such as comments and replies), the text should be divided into sections, each with its own heading. Sections are numbered (1, 2, 3, etc.). A maximum of four levels of heads may be used, with subsections numbered 1.1., 1.2.; 1.1.1., 1.2.1; 1.1.1.1., and so on. Headings should be sentence fragments. Examples of headings are:

2 Materials and Methods

In this paper, we use two models, MS-AR model (Markov-switching autoregressive models) and HMM model (Hidden Markov model)

From the

The MS-AR models we use in this paper is detailed described in the paper ‘Markov-switching autoregressive models for wind time series’(2012) by Pierre Ailliot , Valérie Monbet.

MS-AR model is a combination of Hidden Markov model and auto-regressive model. It has St∈{1,...,S} represents the state at time t. yt denotes the wind speed at time t.

The conditional distribution of state St only depends on the value of pervious state St-1, and is independent of other pervious state as we are using a Markov first order chain in our application.

And the conditional distribution of the observation yt depends on the value of current state St, and depends on previous observation yt-1 and previous observation yt-2 as we are using a auto-regressive second order in our application.

In our application of the MS-AR model, we assume it to be homogenous. Meaning that the transition probabilities P(St=s’|St-1=s) are constant in time. And p(k) = pk, so that iterated algorithm has the same result as a direct algorithm when we perform a prediction using the MS-AR model.

Using the model, we do 2 things. One, we can do a simulation of the wind-speed data for the past 71 years. Second, we can do a prediction of wind-speed given the pervious observations.

Before using the model, we do box-cox transformation on all of our data. In order to make our variables to be more in a normal shape, because when we are using the MS-AR model, we assume a Gaussian distribution.

Then, we tried different number of regime and AR iteratively to find the optimum number of regime and the optimum degree of AR for our model by choosing the combination that give us the smallest BIC.

Using our data, we found out the optimum combination is with 4 regime and AR equals to 2.

Transition matrix and BIC of different combination of regime and AR can be found in the supplementary materials.

3 Data, or a descriptive heading about data

The data we use is from National Oceanic and Atmospheric Administration (<https://www.noaa.gov/>).

The data is 71 years of 10m daily wind-speed data from 1947-2018, covering the area of mainly California and Nevada. It consists of 30 locations; each location is at a distance of 1.875 degree in longitude or latitude with nearby locations.

Detailed characteristic of the data include spatial correlation can be found in the supplementary materials.

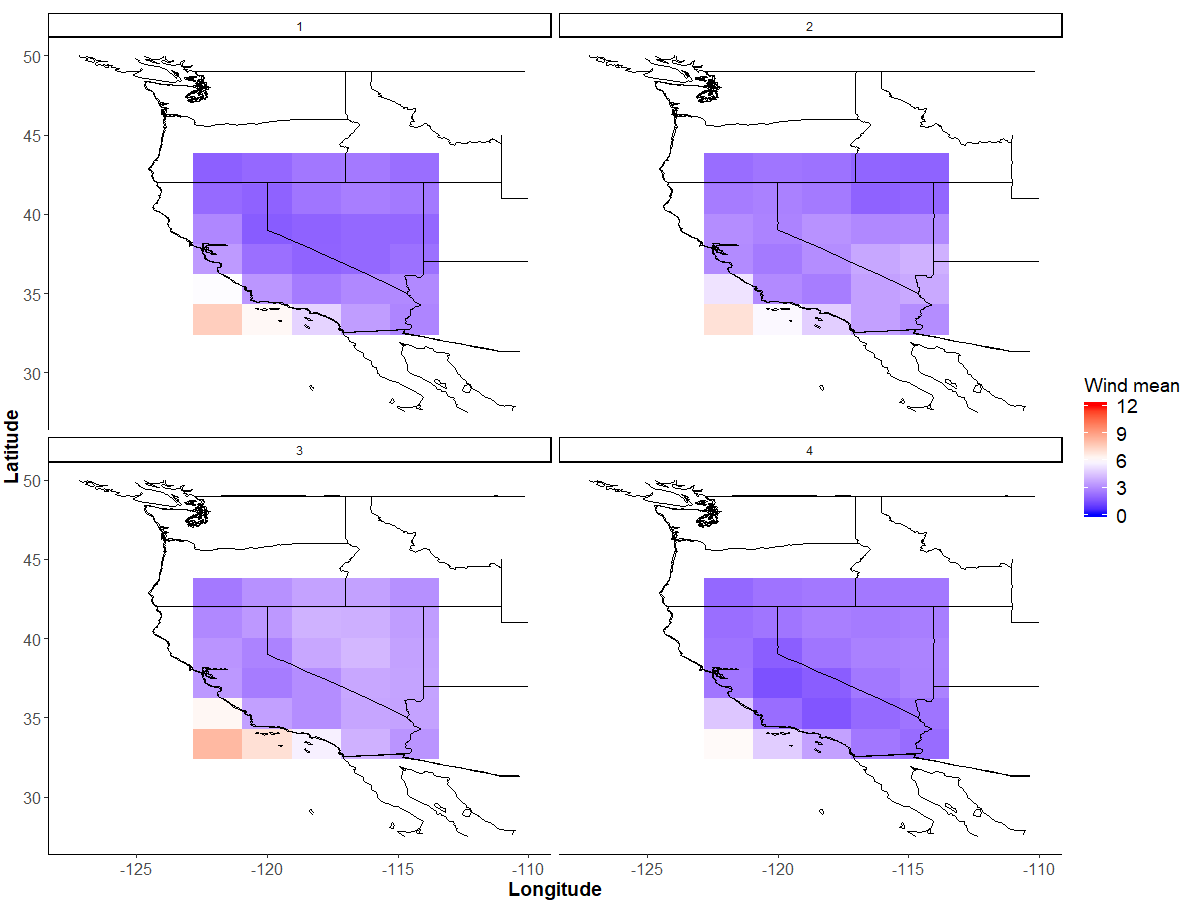
4 Results, or a descriptive heading about the results

Below is the average wind speed for different location for the 4 regimes.

Higher average wind speed has a lighter blue color, while deeper blue means lower average wind speed. Only in the sea we can find red color indicates high wind speed.

Four regimes are different in average wind speed in different locations. But we can not simply say which regime has the highest average wind speed from the graph. In each regime, 30 locations are seen as a whole, the result of calculation shows that regime 1 has the lowest overall average wind speed while regime 4 has the highest. But when we see a specific location, some will have a higher average wind speed in regime 1 than regime 4.

The 4 regimes’ average wind speed can be used for us to decide wind-power in each regime, and when we use also the distribution of 4 regimes, we can have an idea of the anticipation for building wind-turbine in the area.

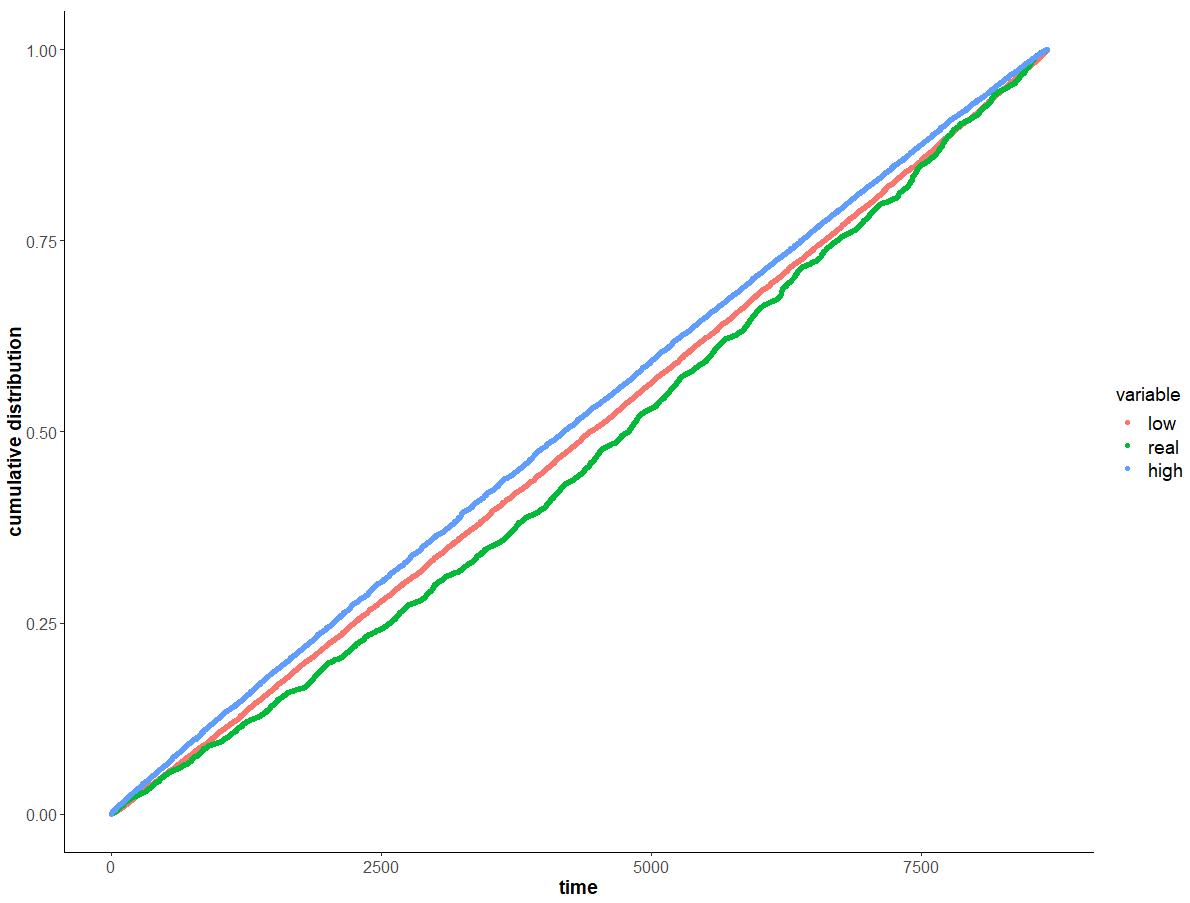


Using multiple simulation on the data to generate a same length data with 71 years. We get the empirical cumulative distribution as below.

Lower bound and upper bound is the 5% and 95% percentile of all the simulations. Real is the observed data.

We can see that

(??)



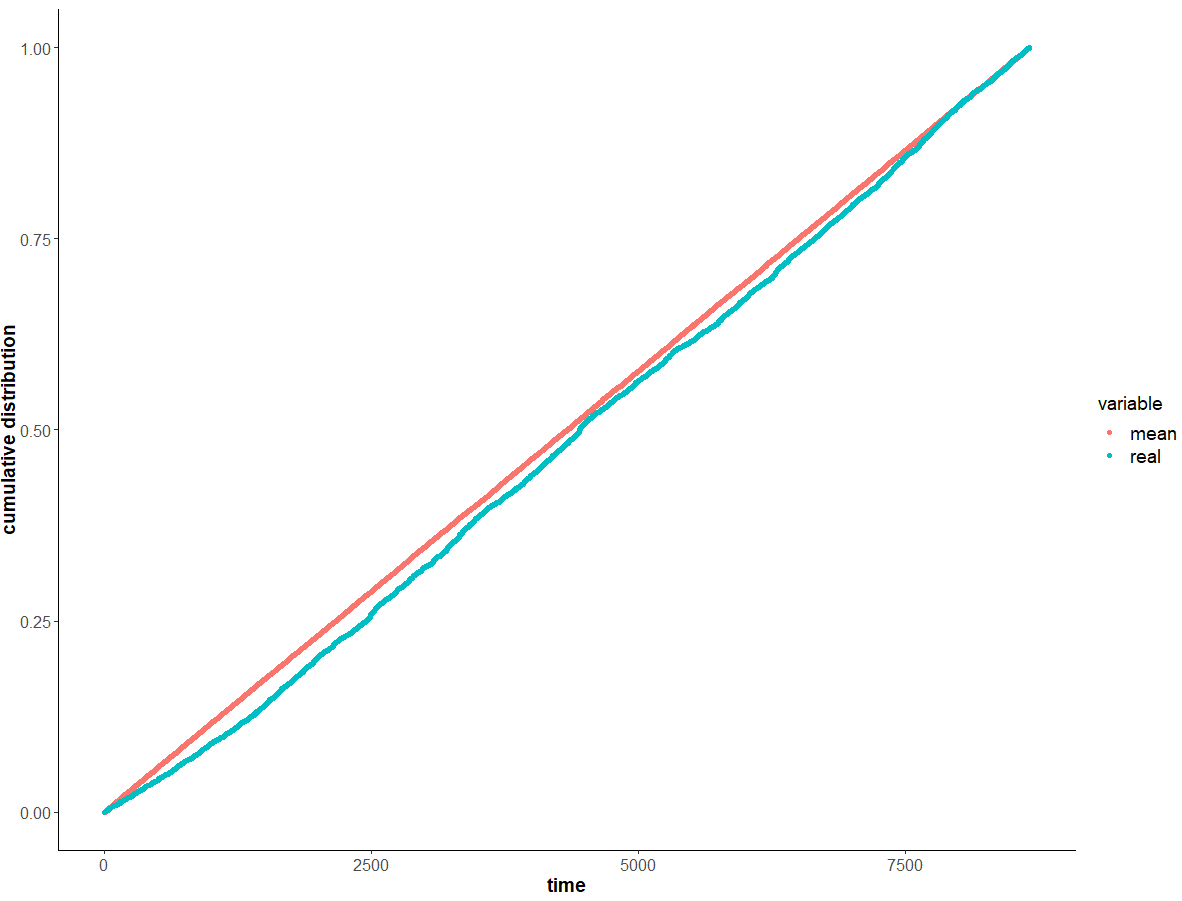


Figure 3 is the difference when doing prediction using two methods, HMM and MS-AR.

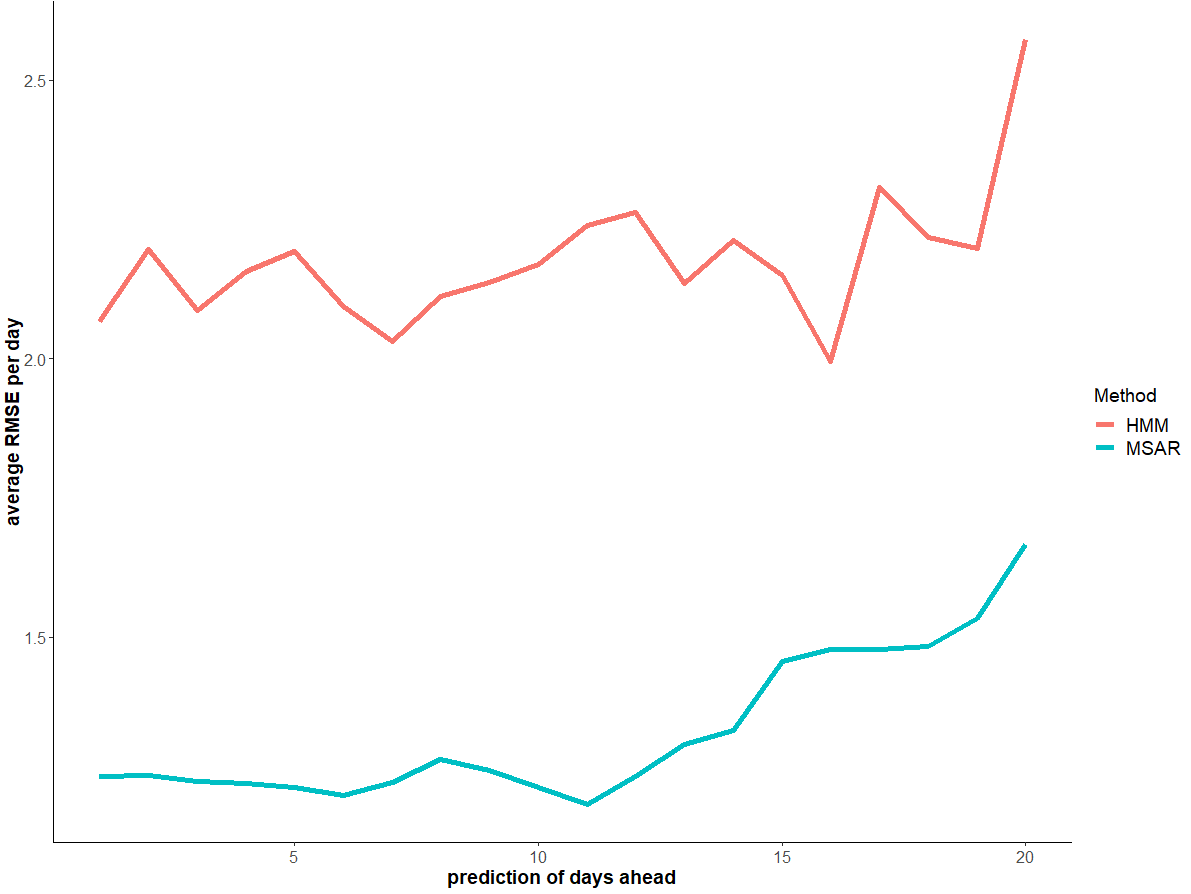
Y-axis is the average RMSE per day, while X-axis is the number of days that we are predicting.

Both methods RMSE increase with the increase of number of days because error propagation.

While we can see that MS-AR has a lower RMSE than using simple HMM model.

The curve is not monotonically raising up because some uncertainty in the prediction, but the general trend we can easily figure out.

The error propagation is more significant when number of days of prediction increases. After day 12, it starts to increase much faster than before. It could be a significant deviation happens around that time and it start to propagate.



5 Conclusions

(All figures and tables should be cited in order. For initial submission, please embed figures, tables, and their captions within the main text near where they are cited. At revision, figures should be uploaded separately, as we need separate files for production. Tables and all captions should be moved to the end of the file.)

References should use a name-date format, not numbers. Enclose citations in brackets with authors in italics as in: [*Smith et al.,* 2009] or *Smith et al.* [2009].

Acknowledgments, Samples, and Data

The text ends with an acknowledgment section and statement that includes:

* Any real or perceived financial conflicts of interests for any author
* Other affiliations for any author that may be perceived as having a conflict of interest with respect to the results of this paper.
* A statement that indicates to the reader where the data supporting the conclusions can be obtained (for example, in the references, tables, and other databases).
* Funding information related to the work for all authors should be entered in the form in GEMS as part of your submission. This form in GEMS uses the official Fundref list, which provides a link after publication that is available to funders. Any other funding information not listed in the GEMS form should be included in the acknowledgments and/or cover letter.

It is also the appropriate place to thank colleagues and other contributors. AGU does not normally allow dedications.

AGU supports following [best practices](https://copdess.org/enabling-fair-data-project/author-guidelines/) and depositing data into domain repositories, or other public repositories. Citations to archived data should be included in your reference list. Data sets should not be listed as “available from authors.” Data sets should be deposited in a domain repository specializing in that data type or a general repository if a specific domain repository is not available. See guidelines [here](https://copdess.org/enabling-fair-data-project/author-guidelines/) and FAQ’s [here](https://copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/). AGU requires that all data necessary to reproduce the reported finding be available at the time of publication.

AGU recommends use of IGSNs ([International Geo Sample Numbers](http://www.geosamples.org/igsnabout)) for all samples reported in AGU Journals. IGSNs provide a unique identifier allowing samples to be linked across publications and searched through a central repository. We strongly encourage authors to [register samples](http://www.igsn.org/register-your-samples) and obtain IGSNs and use them throughout their manuscript and tables. We recognize IGSNs during our production process and will provide links in manuscript tables to the registered sample descriptions.

References

All sources cited in text, tables, figures, and Supporting Information must appear in the main reference list, and all entries in the reference list must be cited in main text. References that are cited in supporting information should also be included in the reference list of the paper and worked into the text, so that they will be indexed and included in citation records and given credit. References are not included in word counts for excess length fees. Data sets that are not newly reported as part of this research should also be cited in the references as a reference. New data sets or software that are deposited elsewhere with a permanent identifier should be cited. AGU follows the [Joint Declaration of Data Citations Principles.](https://www.force11.org/group/joint-declaration-data-citation-principles-final)

All references must be available to readers at the time of publication; there should be no “unpublished” or “in press” references.

An examples of reference:

Deng, A., and D. R. Stauffer (2006), On improving 4-km mesoscale model simulations, J. Appl. Meteorol. Climatol., 45(3), 361–381, doi:10.1175/JAM2341.1.

**Figure 1**. The figure caption should begin with an overall descriptive statement of the figure followed by additional text. They should be immediately after each figure. Figure parts are indicated with lower-case letters (**a, b, c…**). For initial submission, please place both the figures and captions in the text near where they are cited rather than at the end of the file (not both). At revision, captions should be placed at the end of the file, and figures should be uploaded separately. Each figure should be one complete file (please do not upload parts separately).

**Table 1.** Start this caption with a short description of your table. Format tables using the Word Table commands and structures. Do not create tables using spaces or tabs characters. Large tables especially presenting rich data should be presented as separate excel or .cvs files, not as part of the main text.