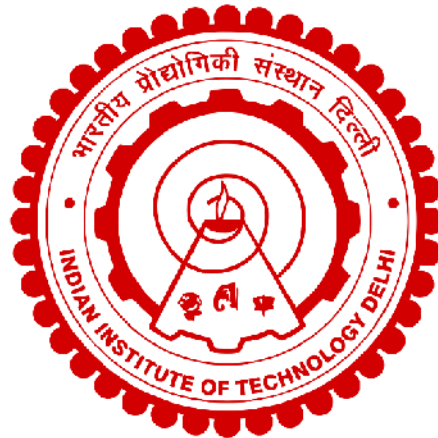


**SUBSEASONAL TO SEASONAL (S2S) SCALE  
FORECASTING OF WIND SPEEDS OVER INDIA:  
IMPLICATIONS FOR THE WIND ENERGY SECTOR**

**AHELI DAS**



**CENTRE FOR ATMOSPHERIC SCIENCES  
INDIAN INSTITUTE OF TECHNOLOGY DELHI  
JANUARY 2025**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2025

**SUBSEASONAL TO SEASONAL (S2S)  
SCALE FORECASTING OF WIND SPEEDS  
OVER INDIA: IMPLICATIONS FOR THE  
WIND ENERGY SECTOR**

*by*

**AHELI DAS**

**Centre for Atmospheric Sciences**

*Submitted*

*In fulfilment of the requirements of the degree of Doctor of Philosophy*

*to the*



**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**JANUARY 2025**

*Dedicated to the incredible determination and resilience of all  
the researchers*

## Certificate

This is to certify that the thesis entitled “**Subseasonal to Seasonal (S2S) Scale Forecasting of Wind Speeds over India: Implications for the Wind Energy Sector**” being submitted by **Ms. Aheli Das** to the Indian Institute of Technology Delhi for the award of the degree of **DOCTOR OF PHILOSOPHY** is a record of original bonafide research carried out by her. Ms. Aheli Das has worked under my guidance and supervision and has fulfilled the requirements for the submission of this thesis. The results contained in this thesis have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.

**Dr. Somnath Baidya Roy**

Professor  
Centre for Atmospheric Sciences  
Indian Institute of Technology Delhi  
New Delhi-110016, India

## **Acknowledgments**

First of all, I would like to express my heartfelt gratitude to my mother, Mrs. Ratna Das, and father, Mr. Asit Das, for their immense support and encouragement in my academic journey over the years, for putting their faith in me, and for being my pillars of strength in turbulent times. I laud my mother for always being optimistic when I felt hopeless, bearing with my regular rants and tantrums, and even pitching mini-lecture sessions at times. On the other hand, my father has always been my silent supporter and has never discouraged me from taking on challenges in life. I don't think I could have traversed this far without them.

I express my sincerest gratitude to my advisor, Dr. Somnath Baidya Roy, for considering me a suitable PhD candidate and providing me with the opportunity to work on a very significant and globally relevant topic for my doctoral thesis. He constantly strived to elevate the standard of my research by encouraging critical thinking. He dedicated countless hours to helping me improve my scientific writing and presentation skills. His unconventional supervision approach has instilled in me the confidence to address challenging research questions. I thoroughly enjoyed his courses, where being a Teaching Assistant over the past five years helped me revisit the fundamentals of atmospheric data analysis and numerical weather modelling, and additionally gave me a first-hand classroom teaching experience. His collaborative project with ReNew was my first exposure to working with wind energy industry professionals. I hugely admire his teaching skills, especially his simplified explanation of complex concepts. I consider myself very fortunate to have come across a teacher like him.

I thank my SRC members, Dr. Krishna AchutaRao, Dr. Sarvesh Kumar Dubey, and Dr. Nilanjan Senroy, for their constructive comments during my PhD progress presentations and seminars. I also thank all the CAS Office staff and the past and present technical team members Ms. Namita Sinha and Mr. Manpreet Singh, for handling all the indirect work-related issues.

Over the past few years, I was befriended by an amazing and vibrant group of people at IITD. We all embarked on a similar voyage and have thus played an indispensable part in each other's lives during the few years of our interaction. Starting with Shrestha, the only close friend I had in my 1<sup>st</sup> semester, to Indranil da, Sushovan, Subhojit, Chiranjit, Sankar, and ending with the newest addition, Ruchira, we all cheered for each other and pulled each other through tough times. It would have been an extremely monotonous journey without the interesting discussions over coffee, hanging out together, and celebrating achievements. I also thank all the students and project staff, Arunima, Pranay, Ajay, Sathish, and Rupanjan, for engaging in the exchange of scientific ideas while working on the ReNew and the Monsoon Mission projects. The list remains incomplete without crediting my seniors, Anasuya and Dr. Tanvi Gupta, for sparing their time for scientific and non-scientific discussions.

Pursuing a PhD at IITD would have only remained a dream without the inspiration of my professors back at the Department of Atmospheric Sciences, University of Calcutta, Dr. Sutapa Chaudhuri, for introducing us to meteorology and daring us to dream big, and Dr. Sarbari Ghosh, from whom I learned the fundamentals of statistics and forecasting. I also thank Dr. Mourani Sinha and Dr. Ramkrishna Datta for their support and motivation.

I duly acknowledge the Doctoral fellowship provided by the Council of Scientific and Industrial Research, Government of India, to carry out my research. I am also grateful to ReNew for giving me a flavour of what the real-life application of my PhD research looks like and for also sponsoring my visit to the Bureau of Meteorology Annual RnD Workshop 2023 in Melbourne and the AGU Fall meeting 2023 in San Francisco. I also acknowledge the IIT Delhi High Performance Computing facility PADUM for providing all the necessary computing resources.

Lastly, a big shoutout to all the cats that I came across during my PhD. Starting with the ones back at home, Meowie, Mini, the floofy cat, Mini's kittens Miu, and Pookie. They kept me

company during my isolated working hours. In addition, there are the ones roaming around our academic block, the grey tabby and her kittens, the orange tabby and her kittens, and the one with markings similar to a clouded leopard. All of them acted as stress-relievers and thus played small but vital roles during my PhD.

## **Abstract**

The global power sector is shifting from generating electricity using conventional fossil fuel-based energy sources to cleaner and sustainable renewable energy sources to tackle climate change caused by anthropogenic greenhouse gas emissions. Among the several renewable energy options available, the expansion of wind energy has been picking up pace worldwide and contributed to one-tenth of the global electricity generation in 2023. The Indian wind energy sector, which has the 4<sup>th</sup> highest installed wind capacity in the world, has produced around 5% of the electricity generated in the country in the last financial year. However, the wind energy industry faces challenges in electricity grid integration because wind is an intermittent resource. This intermittency can lead to sudden excess or shortfall of power thereby destabilizing the grid. Accurate wind speed forecasts are integral to solving the intermittency problem and accelerating the growth of wind energy. This thesis attempts to solve the challenge of providing high-quality skillful wind speed forecasts over India in the Subseasonal to seasonal (S2S) scale, a time-scale with low predictability.

The thesis focuses on the S2S scale having a forecast lead time spanning two weeks to a season. This time-scale is widely regarded as the ‘predictability desert’ because atmospheric predictability is known to be limited to two weeks. But the S2S forecasts are highly valued, especially in the wind energy sector, where they help in decision-making and financial planning. Various meteorological agencies worldwide have started providing experimental S2S forecasts by employing high-end coupled ocean-atmospheric models. These models are capable of simulating slowly varying earth system features like the sea-surface temperature, soil moisture, and snow cover, which aid in long-term predictability. This suggests that the forecasts of meteorological variables from these models may have some skill and therefore should be evaluated. This thesis evaluates the performance of the experimental S2S forecasts over India and calibrates the best performing forecast to further improve its performance.

Forecast evaluation requires good quality reference ground truth, such as station observations, before being deployed for applications by the wind energy sector. As the wind speed observation network over India is sparse, not homogeneously distributed, and discontinuous in time, the wind energy sector relies on meteorological reanalysis datasets. Being a numerical weather model derived product, the methods of estimating wind speeds in the reanalyses differ, due to which these reanalyses wind speeds need to be validated to find out which one poses the closest statistical resemblance to the observations. In this thesis, validation is done for 10 m wind speeds, also known as near-surface wind speeds, as it is the standard height recommended by the World Meteorological Organization for wind speed measurements. A comparison of 10 m wind speeds from NCEP1, NCEP2, JRA55, IMDAA, ERA5, and MERRA2 with station observations from NCEI-GSOD, spatially averaged over the homogeneous climate zones of India are carried out using Root Mean Squared Error (RMSE), Pearson's correlation, distribution, trend, and robust coefficient of variation for the period 1980-2020. Results demonstrate that JRA55 has the lowest RMSE and the highest correlation and represents the observed distribution, trends, and variability. Therefore, JRA55 best represents observed near-surface wind speeds over India. Among the others, the performance of ERA5 follows JRA55. These reanalyses are used to evaluate the S2S wind speed forecasts.

The JRA55 and ERA5 reanalyses are used in this thesis, to evaluate the monthly mean 10 m wind speed forecasts from six ocean-atmospheric coupled models SEAS5, GCFS2.0, MF6, CFSv2, GS5-GC2-LI, and SPS3 at 1, 2, 3, 4, and 5 month lead times for the period 1994-2016 over the homogeneous climate zones of India. Using two different reanalyses helps to understand how the forecast skill varies with the reference used for evaluation. The evaluation using deterministic and probabilistic metrics like standard deviation, RMSE, Normalized RMSE (NRMSE), Fair Continuous Ranked Probability Skill Score (FCRPSS) show that the raw S2S wind speed forecasts have little to no skill, and their skill is relative to the reanalysis

used as reference. On aggregation, all the models show 33.83% skillful forecasts using the 2<sup>nd</sup> best reanalysis ERA5 but 2.95% skillful forecasts using the best reanalysis JRA55. The raw forecasts also overestimate the magnitude of Intra-seasonal Oscillation (ISO), which is one of the dominant modes of intraseasonal variability and a significant driver of subseasonal and seasonal predictability. Overall, these forecasts are not fit for direct applications and may gain some skill through calibration.

Among the six S2S models evaluated, the raw wind speed forecasts from the SEAS5 model are chosen for calibration. Unlike the others, SEAS5 raw wind speed forecasts show some skill over climatology and have long hindcast data that provides ample data points for training, validating, and testing the calibration algorithms. The calibration uses statistical and Machine Learning (ML) based methods with JRA55 as the reference. The statistical methods are Bias Adjustment, Quantile Mapping, Ratio of Predictable Components, and the ML-based methods are Random Forest (RF), Light Gradient Boosting Machine (LGBM), RF\_lags, and LGBM\_lags. The latter two methods use past observations as features in their data. The data from 1982-2018 are used for training and validation, whereas 2019-2021 are used for testing. The raw and calibrated forecasts for the testing period are compared using metrics similar to those in the earlier section. The results show that calibration significantly improves the skill of raw S2S wind speed forecasts; the statistical methods enhance the raw forecast skill by 125-197% whereas the ML-based methods enhance the skill by 143-279%. The ML-based methods significantly outperform the statistical methods and can be considered as the best method for forecast calibration.

Apart from the individual monthly mean calibrated S2S forecasts, the wind energy sector also wants to know if the upcoming monsoon wind speeds will be stronger or weaker than climatology. This information aids the industry in chalking out their plans for the forthcoming financial year because more than 50% of the total annual power is generated during the

monsoon months. To address this question, the SEAS5 wind speed forecasts of the peak monsoon months of June, July, and August (JJA), initialized on the 1<sup>st</sup> of March, are calibrated using ML-based methods. The results show that the ML-based calibrated JJA forecasts have wind speed anomalies of -0.13 to -0.64 m/s, which is close to the observed JRA55 anomalies of -0.17 to -0.82 m/s. Therefore, the ML-based calibration algorithms can reduce the systematic errors of the raw S2S wind speed forecasts, elevate their skill, and predict anomalies that are important for the wind energy sector.

This is a novel study because it addresses a very challenging scientific problem on S2S wind speed forecasting. It is one of the first studies to implement ML-based calibration algorithms to improve S2S wind speed forecasts. Introducing past observations or lags is also a new addition to the ML-based calibration algorithms, which have not been applied before. Overall, this study provides a statistically robust workflow towards obtaining skillful wind speed forecasts that the wind energy sector can adopt for their operations.

## सार

विश्वीय ऊर्जा क्षेत्र पारंपरिक जीवाश्म ईंधन आधारित ऊर्जा स्रोतों का उपयोग करके बिजली पैदा करने से हटकर स्वच्छ और संधारणीय अक्षय ऊर्जा स्रोतों की ओर जा रहा है, ताकि मानवजनित ग्रीनहाउस गैस उत्सर्जन के कारण होने वाले जलवायु परिवर्तन से निपटा जा सके। उपलब्ध कई अक्षय ऊर्जा विकल्पों में से, पवन ऊर्जा का विस्तार दुनिया भर में गति पकड़ रहा है और 2023 में विश्वीय बिजली उत्पादन में दसवां हिस्सा योगदान देगा। भारतीय पवन ऊर्जा क्षेत्र, जिसकी दुनिया में चौथी सबसे अधिक स्थापित पवन क्षमता है, ने पिछले वित्तीय वर्ष में देश में उत्पादित बिजली का लगभग 5% उत्पादन किया है। हालाँकि, पवन ऊर्जा उद्योग को बिजली ग्रिड एकीकरण में चुनौतियों का सामना करना पड़ता है क्योंकि पवन एक अस्थायी संसाधन है। इस अस्थायीता के कारण अचानक बिजली की अधिकता या कमी हो सकती है जिससे ग्रिड अस्थिर हो सकता है। सही पवन गति पूर्वानुमान अस्थायीता की समस्या को हल करने और पवन ऊर्जा के विकास में तेजी लाने के लिए अभिन्न अंग हैं। यह थीसिस भारत में उप-मौसमी से मौसमी (S2S) पैमाने पर उच्च-गुणवत्ता वाले कुशल पवन गति पूर्वानुमान प्रदान करने की चुनौती को हल करने का प्रयास करती है, जो कम पूर्वानुमान के साथ एक समय-पैमाना है।

यह थीसिस S2S पैमाने पर केंद्रित है, जिसमें दो सप्ताह से एक मौसम के लिए तक का पूर्वानुमान समय होता है। इस समय-पैमाने को व्यापक रूप से 'पूर्वानुमान रेगिस्तान' के रूप में माना जाता है क्योंकि वायुमंडलीय पूर्वानुमान दो सप्ताह तक सीमित माना जाता है। लेकिन S2S पूर्वानुमानों को बहुत महत्व दिया जाता है, खासकर पवन ऊर्जा क्षेत्र में, जहाँ वे निर्णय लेने और वित्तीय नियोजन में मदद करते हैं। दुनिया भर में विभिन्न मौसम विज्ञान एजेंसियों ने उच्च-स्तरीय युग्मित महासागर-वायुमंडलीय मॉडल का उपयोग करके प्रयोगात्मक S2S पूर्वानुमान प्रदान करना शुरू कर दिया है। ये मॉडल धीरे-धीरे बदलती पृथ्वी प्रणाली विशेषताओं जैसे समुद्र की सतह के तापमान, मिट्टी की नमी और बर्फ के आवरण का अनुकरण करने में सक्षम हैं, जो दीर्घकालिक पूर्वानुमान में सहायता करते हैं। यह थीसिस भारत पर

प्रयोगात्मक S2S पूर्वानुमानों के प्रदर्शन का मूल्यांकन करता है और इसके प्रदर्शन को और बेहतर बनाने के लिए सबसे अच्छा प्रदर्शन करने वाले पूर्वानुमान को कैलिब्रेट करता है।

पूर्वानुमान मूल्यांकन के लिए पवन ऊर्जा क्षेत्र द्वारा अनुप्रयोगों के लिए तैनात किए जाने से पहले स्टेशन अवलोकन जैसे उच्च गुणवत्ता वाले संदर्भ ग्राउंड सत्य की आवश्यकता होती है। क्योंकि भारत में पवन गति अवलोकन नेटवर्क विरल है, समरूप रूप से विस्तारित नहीं है, और समय में असंतत है, इसलिए पवन ऊर्जा क्षेत्र मौसम संबंधी पुनर्विश्लेषण डेटासेट पर निर्भर करता है। एक संख्यात्मक मौसम मॉडल व्युत्पन्न उत्पाद होने के नाते, पुनर्विश्लेषण में पवन गति का अनुमान लगाने के तरीके अलग-अलग होते हैं, जिसके कारण इन पुनर्विश्लेषण पवन गति को यह पता लगाने के लिए मान्य करने की आवश्यकता होती है कि कौन सी विधि अवलोकनों के लिए सबसे निकटतम सांख्यिकीय समानता प्रस्तुत करती है। इस थीसिस में, 10 मीटर पवन गति के लिए सत्यापन किया जाता है, जिसे निकट-सतह पवन गति के रूप में भी जाना जाता है, क्योंकि यह पवन गति माप के लिए विश्व मौसम विज्ञान संगठन द्वारा अनुशंसित मानक ऊंचाई है। NCEP1, NCEP2, JRA55, IMDAA, ERA5 और MERRA2 से 10 मीटर की हवा की गति की तुलना NCEI-GSOD के स्टेशन अवलोकनों के साथ की गई है, जो भारत के समरूप जलवायु क्षेत्रों में स्थानिक रूप से औसत है, 1980-2020 की अवधि के लिए आरएमएसई, सहसंबंध, वितरण, प्रवृत्ति और भिन्नता के मजबूत गुणांक का उपयोग करके किया जाता है। परिणाम दर्शाते हैं कि JRA55 में सबसे कम आरएमएसई और सबसे अधिक सहसंबंध है और यह देखे गए वितरण, प्रवृत्तियों और परिवर्तनशीलता का प्रतिनिधित्व करता है। इसलिए, JRA55 भारत में देखी गई सतह के निकट हवा की गति का सबसे अच्छा प्रतिनिधित्व करता है। अन्य के बीच, ERA5 का प्रदर्शन JRA55 का अनुसरण करता है। इन पुनर्विश्लेषणों का उपयोग S2S पवन गति पूर्वानुमानों का मूल्यांकन करने के लिए किया जाता है। इस थीसिस में JRA55 और ERA5 पुनर्विश्लेषण का उपयोग किया गया है, ताकि भारत के समरूप जलवायु क्षेत्रों में 1994-2016 की अवधि के लिए 1, 2, 3, 4 और 5 महीने के लीड टाइम पर छह महासागर-

वायुमंडलीय युग्मित मॉडल SEAS5, GCFS2.0, MF6, CFSv2, GS5-GC2-LI और SPS3 से मासिक औसत 10 मीटर वायु गति पूर्वानुमान का मूल्यांकन किया जा सके। दो अलग-अलग पुनर्विश्लेषण का उपयोग यह समझने में मदद करता है कि मूल्यांकन के लिए उपयोग किए जाने वाले संदर्भ के साथ पूर्वानुमान कौशल कैसे भिन्न होता है। मानक विचलन, RMSE, NRMSE, FCRPSS जैसे नियतात्मक और संभाव्य मीट्रिक का उपयोग करके मूल्यांकन से पता चलता है कि मूल S2S वायु गति पूर्वानुमानों में बहुत कम या कोई कौशल नहीं है, और उनका कौशल संदर्भ के रूप में उपयोग किए जाने वाले पुनर्विश्लेषण के सापेक्ष है। एकत्रीकरण पर, सभी मॉडल 2nd सर्वश्रेष्ठ पुनर्विश्लेषण ERA5 का उपयोग करके 33.83% कुशल पूर्वानुमान दिखाते हैं, लेकिन सर्वश्रेष्ठ पुनर्विश्लेषण JRA55 का उपयोग करके 2.95% कुशल पूर्वानुमान दिखाते हैं। मूल पूर्वानुमान भी ISO के परिमाण को बढ़ा-चढ़ाकर बताते हैं, जिसे उप-मौसमी और मौसमी पूर्वानुमान का एक महत्वपूर्ण चालक माना जाता है। कुल मिलाकर, ये पूर्वानुमान प्रत्यक्ष अनुप्रयोगों के लिए उपयुक्त नहीं हैं और अंशांकन के माध्यम से कुछ कौशल प्राप्त कर सकते हैं।

मूल्यांकन किए गए छह S2S मॉडलों में से, SEAS5 मॉडल से मूल वायु वेग पूर्वानुमानों को अंशांकन के लिए चुना गया है। दूसरों के विपरीत, SEAS5 मूल वायु वेग पूर्वानुमान जलवायु विज्ञान पर कुछ कौशल दिखाते हैं और उनके पास लंबे समय तक पूर्वानुमानित डेटा होते हैं जो अंशांकन एल्गोरिदम के प्रशिक्षण, सत्यापन और परीक्षण के लिए पर्याप्त डेटा बिंदु प्रदान करते हैं। अंशांकन संदर्भ के रूप में JRA55 के साथ सांख्यिकीय और मशीन लर्निंग (ML) आधारित विधियों का उपयोग करता है। सांख्यिकीय विधियाँ पूर्वाग्रह समायोजन, क्वांटाइल मैपिंग, पूर्वानुमानित घटकों का अनुपात हैं, और ML-आधारित विधियाँ रैंडम फ़ॉरेस्ट (RF), लाइट ग्रेडिएंट बूस्टिंग मशीन (LGBM), RF\_lags और LGBM\_lags हैं। बाद की दो विधियाँ अपने डेटा में सुविधाओं के रूप में पिछले अवलोकनों का उपयोग करती हैं। 1982-2018 के डेटा का उपयोग प्रशिक्षण और सत्यापन के लिए किया जाता है, जबकि 2019-2021 का उपयोग परीक्षण के लिए किया जाता है। परीक्षण अवधि के लिए मूल और अंशांकित पूर्वानुमानों की तुलना पहले अनुभाग में दिए गए समान मीट्रिक का उपयोग करके की जाती है। परिणाम दर्शाते हैं कि अंशांकन मूल S2S वायु

गति पूर्वानुमान के कौशल में महत्वपूर्ण रूप से सुधार करता है; सांख्यिकीय विधियाँ मूल पूर्वानुमान कौशल को 125-197% तक बढ़ाती हैं जबकि ML-आधारित विधियाँ कौशल को 143-279% तक बढ़ाती हैं। ML-आधारित विधियाँ सांख्यिकीय विधियों से काफी बेहतर प्रदर्शन करती हैं और इन्हें पूर्वानुमान अंशांकन के लिए सबसे अच्छी विधि माना जा सकता है।

व्यक्तिगत मासिक औसत कैलिब्रेटेड S2S पूर्वानुमानों के अलावा, पवन ऊर्जा क्षेत्र यह भी जानना चाहता है कि आगामी मानसून की हवा की गति जलवायु विज्ञान की तुलना में अधिक मजबूत होगी या कमजोर। यह जानकारी उद्योग को आगामी वित्तीय वर्ष के लिए अपनी योजनाएँ बनाने में सहायता करती है क्योंकि कुल वार्षिक बिजली का 50% से अधिक उत्पादन मानसून के महीनों के दौरान होता है। इस प्रश्न का समाधान करने के लिए, 1 मार्च को आरंभ किए गए जून, जुलाई और अगस्त के चरम मानसून महीनों के SEAS5 पवन गति पूर्वानुमानों को ML-आधारित विधियों का उपयोग करके कैलिब्रेट किया जाता है। परिणाम बताते हैं कि ML-आधारित कैलिब्रेटेड JJA पूर्वानुमानों में -0.13 से -0.64 मीटर/सेकंड की पवन गति विसंगतियाँ हैं, जो कि -0.17 से -0.82 मीटर/सेकंड की देखी गई JRA55 विसंगतियों के करीब है। इसलिए, ML-आधारित कैलिब्रेशन एल्गोरिदम मूल S2S पवन गति पूर्वानुमानों की व्यवस्थित त्रुटियों को कम कर सकते हैं, उनके कौशल को बढ़ा सकते हैं और उन विसंगतियों की भविष्यवाणी कर सकते हैं जो पवन ऊर्जा क्षेत्र के लिए महत्वपूर्ण हैं। यह एक नया अध्ययन है क्योंकि यह S2S पवन गति पूर्वानुमान पर एक बहुत ही चुनौतीपूर्ण वैज्ञानिक समस्या को संबोधित करता है। यह S2S पवन गति पूर्वानुमानों को बेहतर बनाने के लिए ML-आधारित अंशांकन एल्गोरिदम को लागू करने वाले पहले अध्ययनों में से एक है। पिछले अवलोकनों या अंतरालों को पेश करना भी ML-आधारित अंशांकन एल्गोरिदम में एक नया जोड़ है, जिसे पहले लागू नहीं किया गया है। कुल मिलाकर, यह अध्ययन कुशल पवन गति पूर्वानुमान प्राप्त करने की दिशा में एक सांख्यिकीय रूप से मजबूत वर्कफ़्लो प्रदान करता है जिसे पवन ऊर्जा क्षेत्र अपने संचालन के लिए अपना सकता है।

# Contents

Certificate.....	i
Acknowledgments.....	ii
Abstract.....	v
संक्षेप.....	ix
List of Figures.....	xvi
List of Tables.....	xx
List of Abbreviations.....	xxii
CHAPTER 1 .....	1
1 Introduction .....	2
1.1 Wind as a renewable energy resource .....	2
1.2 Wind speed forecasting .....	3
1.2.1 Applications of wind speed forecasts at different time-scales.....	4
1.2.2 Forecasting wind speeds in the S2S scale: Challenges and ways to overcome them .....	5
1.3 Literature review .....	8
1.3.1 Necessity of an accurate reanalysis representation of observed near-surface wind speeds.....	8
1.3.2 Evaluation of raw S2S forecasts .....	12
1.3.3 Calibration of raw S2S forecasts using statistical methods .....	16
1.3.4 Calibration of raw S2S forecasts using Machine Learning (ML) methods .....	17
1.3.5 Harnessing the wind speeds during periods of high supply.....	20
1.3.6 Limitations of the above-mentioned research.....	21
1.4 Broad goal and specific objectives of the thesis .....	22
CHAPTER 2 .....	23
2 Data and methods .....	24
2.1 Identifying the best reanalysis representing observed 10 m wind speeds over India ....	24
2.1.1 Reanalysis data.....	24
2.1.2 Station observations .....	26
2.1.3 Study area and time-period .....	27
2.1.4 Processing wind speed reanalyses and observations .....	28
2.1.5 Validation metrics .....	33
2.2 Evaluation of raw S2S 10 m wind speed forecasts .....	35

2.2.1	S2S forecast models .....	35
2.2.2	Reanalyses.....	38
2.2.3	Study area and time period .....	38
2.2.4	Forecast verification metrics.....	39
2.2.5	Investigating sources of S2S 10 m wind speed forecast errors.....	41
2.3	Calibration of raw S2S 10 m wind speed forecasts.....	43
2.3.1	S2S model and reanalysis .....	43
2.3.2	Study area and time period .....	43
2.3.3	Pre-processing of raw forecasts .....	43
2.3.4	Calibration methods .....	44
2.3.5	Forecast verification metrics.....	51
2.4	S2S forecasts of percentage anomaly of monsoon 10 m wind speeds for the wind energy sector .....	51
2.4.1	S2S forecast model and reanalysis.....	51
2.4.2	Study area and time period .....	51
2.4.3	Pre-processing of raw forecasts .....	52
2.4.4	Calibration methods .....	52
2.4.5	Forecast verification metric .....	52
CHAPTER 3	.....	53
3	Identifying the best reanalysis representing observed 10 m wind speeds over India.....	54
3.1	Monthly mean wind speed statistics.....	54
3.1.1	RMSE.....	56
3.1.2	Correlation (R).....	57
3.1.3	Distribution .....	58
3.2	Annual mean wind speed trends.....	60
3.3	Inter-annual variability .....	66
3.4	Conclusions and discussions .....	67
CHAPTER 4	.....	70
4	Evaluating raw S2S 10 m wind speed forecasts over India.....	71
4.1	Evaluation using deterministic metrics .....	71
4.1.1	Standard deviation .....	72
4.1.2	RMSE.....	75
4.1.3	NRMSE.....	80

4.2	Evaluation using probabilistic metric.....	84
4.2.1	FCRPSS .....	84
4.3	Investigating S2S 10 m wind speed forecast errors using Singular Spectrum Analysis .....	90
4.4	Conclusions and discussions .....	95
CHAPTER 5 .....		98
5	Calibrating raw S2S 10 m wind speed forecasts over India .....	99
5.1	Evaluation using deterministic metrics .....	99
5.1.1	Standard deviation .....	99
5.1.2	RMSE.....	101
5.1.3	NRMSE.....	107
5.2	Evaluation using probabilistic metric.....	110
5.2.1	CRPSS.....	110
5.3	Conclusions and discussions .....	118
CHAPTER 6 .....		120
6	S2S forecasts of percentage anomaly of monsoon 10 m wind speeds for the wind energy sector .....	121
6.1	Evaluation of raw and calibrated monsoon mean anomaly forecasts of 10 m wind speeds .....	123
6.2	Conclusion and discussions.....	125
CHAPTER 7 .....		127
7	Conclusions, limitations, and future scope.....	128
7.1	Conclusions .....	128
7.2	Limitations .....	129
7.3	Future scope .....	131
References.....		133
Appendix.....		152
Curriculum Vitae.....		157

# List of Figures

**Figure 1.1:** Wind power generated all over India during the year 2023. Data obtained from CEA Renewable Generation Reports.....3

**Figure 1.2:** Predictability and applications of wind speed forecasts at different time-scales. The red rectangle highlights the S2S scale. ....5

**Figure 1.3:** Typical S2S model grid cell showing various roughness elements.....6

**Figure 2.1:** Map of India showing the station locations in each homogenous region with the topography shown in background. ....28

**Figure 2.2:** Station locations overlaid with the grid points from each reanalysis.....30

**Figure 2.3:** Number of stations used to calculate the mean wind speed for each day from 1980 to 2020 for each homogenous region.....31

**Figure 2.4:** Number of missing observation data points in each decade from 1980 to 2020 over the seven homogenous regions. ....31

**Figure 2.5:** (a) Burst mode initialization of SEAS5 for January where all members are initialized together on the 1<sup>st</sup> day of the month. (b) Lagged mode initialization of GS5-GC2 for January where 7 members are initialized on different dates. ....36

**Figure 2.6:** Calibration workflows of (a) statistical and (b) ML-based methods shown as an example of February L1 forecasts. The grey boxes show the input data that goes into training. Feb Y<sup>th</sup> year is the target variable or the predictand. ....46

**Figure 2.7:** Random Forest architecture. ....49

**Figure 2.8:** Light Gradient Boosting Machine architecture.....49

**Figure 3.1:** Annual cycle of 10 m wind speeds using monthly mean climatology calculated using data from 1980-2020 over the homogeneous climate regions. ....55

<b>Figure 3.2:</b> Heatmaps of RMSE for reanalyses with station observations for individual months over homogenous regions. ....	57
<b>Figure 3.3:</b> Heatmaps of correlation (R) for reanalyses with station observations for individual months over homogenous regions. The cells marked with 'x' indicate that the bias and correlation values are statistically significant at 95% confidence level. ....	58
<b>Figure 3.4:</b> Violin plots of monthly mean wind speed from station observations (OBS) and reanalyses for 1980-2020 over homogenous regions.....	60
<b>Figure 3.5:</b> Annual mean wind speeds in m s <sup>-1</sup> decade <sup>-1</sup> from 1980 to 2020 overlaid with the best fit trends for observations and reanalyses over (a) Western Himalayas (WH), (b) North West (NW), (c) North Central (NC), (d) North East (NE), (e) Interior Peninsula (IP), (f) West Coast (WC), and (g) East Coast (EC). 'OBS' stands for observed wind speeds. ....	64
<b>Figure 3.6:</b> Inter-annual variability of wind speeds calculated using Rcov for observations and reanalyses over the homogenous regions. 'OBS' stands for observed wind speeds. The error bars indicate 95% bootstrapped confidence intervals.....	67
<b>Figure 4.1:</b> Standard deviation of monthly mean 10 m wind speed forecasts from 6 models at 1, 2, 3, 4, and 5 month lead times for 1994-2016 over 6 homogeneous climate zones of India, with bootstrapped 95% confidence intervals shown by whiskers. JRA55, and ERA5 standard deviations are shown by green and blue dashed lines respectively and the respective shaded regions show their bootstrapped 95% confidence intervals.....	73
<b>Figure 4.2:</b> Standard deviation heatmaps of 10 m wind speed forecasts, JRA55, and ERA5 for 12 months from 6 models at 1 month lead time over 6 homogeneous climate zones for 1994-2016.....	75

**Figure 4.3:** RMSE of monthly mean 10 m wind speed forecasts from 6 models using JRA55 and ERA5 at 1, 2, 3, 4, and 5 month lead times for 1994-2016 over 6 homogeneous climate zones of India, with bootstrapped 95% confidence intervals shown by whiskers. .... 77

**Figure 4.4:** RMSE heatmaps of 10 m wind speed forecasts for 12 months from 6 models at 1 month lead time using JRA55 and ERA5 over 6 homogeneous climate zones for 1994-2016. .... 79

**Figure 4.5:** NRMSE of monthly mean 10 m wind speed forecasts from 6 models using JRA55 and ERA5 at 1, 2, 3, 4, and 5 month lead times for 1994-2016 over 6 homogeneous climate zones of India, with bootstrapped 95% confidence intervals shown by whiskers. .... 81

**Figure 4.6:** NRMSE heatmaps of 10 m wind speed forecasts for 12 months from 6 models at 1 month lead time using JRA55 and ERA5 over 6 homogeneous climate zones for 1994-2016. .... 83

**Figure 5.1:** Standard deviation of monthly mean raw and calibrated 10 m wind speed forecasts at 1, 2, 3, 4, and 5 month lead times for 2019-2021 over 6 homogeneous climate zones of India, with bootstrapped 95% confidence intervals shown by whiskers. JRA55 standard deviation is shown by green dashed line and the shaded regions show its bootstrapped 95% confidence intervals. .... 101

**Figure 5.2:** RMSE of monthly mean 10 m wind speed forecasts of raw and calibrated forecasts at 1-5 month lead times over 6 homogeneous climate zones for 2019-2021 with bootstrapped 95% confidence intervals shown by whiskers. .... 103

**Figure 5.3:** RMSE heatmaps of 10 m wind speed raw and calibrated forecasts for 12 months at 1 month lead times over 6 homogeneous climate zones of India for 2019-2021. .... 104

<b>Figure 5.4:</b> NRMSE of monthly mean 10 m wind speed forecasts of raw and calibrated forecasts at 1-5 month lead times over 6 homogeneous climate zones for 2019-2021 with bootstrapped 95% confidence intervals shown by whiskers.....	108
<b>Figure 5.5:</b> NRMSE heatmaps of 10 m wind speed raw and calibrated forecasts for 12 months at 1 month lead times over 6 homogeneous climate zones of India for 2019-2021. ....	109
<b>Figure 5.6:</b> CRPSS of monthly mean 10 m wind speed forecasts of raw and calibrated forecasts at 1-5 month lead times over 6 homogeneous climate zones for 2019-2021. The whiskers show the 5th and 95th percentiles of mean bootstrapped CRPSS.....	111
<b>Figure 5.7:</b> Distribution of CRPSS values of 360 forecasts across 12 months, 6 regions, 5 lead times for raw and calibrated forecasts during the testing period 2019-2021. Dashed line at 0 indicates skill of climatology. ....	112
<b>Figure 5.8:</b> Plots showing the range of CRPSS averaged over 2019-2021 of Raw, BA, QM, RPC, RF, LGBM, RF_lags, LGBM_lags forecasts of 12 months, 6 regions, and 5 lead times. Dashed line at 0 indicates skill of climatology. ....	114
<b>Figure 5.9:</b> Wind farm locations over India. ....	117
<b>Figure 6.1:</b> Climatology of monthly mean 10 m wind speeds from JRA55 and station observations (OBS) calculated with data from 1980-2020 over 6 homogeneous regions. Shading shows the wind speed range for each month from both observations and JRA55. .	123
<b>Figure 6.2:</b> Anomalies and percentage anomalies of mean JJA 10 m wind speed from JRA55, raw, and calibrated SEAS5 forecasts initialized on the 1st of March for the years 2019, 2020, 2021 over the 6 homogeneous climate zones of India. The error bars show the range of anomalies and percentage anomalies of the ensemble members.....	125

## List of Tables

<b>Table 2.1:</b> Summary of reanalyses data.....	25
<b>Table 2.2:</b> Number of stations in each region and percentage of missing data in daily mean wind speed time series in each region.....	32
<b>Table 2.3:</b> Summary of forecast models.....	37
<b>Table 2.4:</b> Initialization times for target months May and June shown as examples.....	37
<b>Table 2.5:</b> Selection of observation lags for a target month e.g. February and lead time based on the forecast initialization.....	45
<b>Table 2.6:</b> Hyperparameter search spaces for ML-based methods. ....	51
<b>Table 3.1:</b> Annual mean wind speed trends in $\text{ms}^{-1}\text{decade}^{-1}$ for observations and reanalyses over seven homogenous regions from 1980 to 2020. Values in bold indicate significant trends ( $p < 0.05$ ).....	61
<b>Table 3.2:</b> Annual mean wind speed trends of difference time series of observations and each reanalysis ( $\text{m s}^{-1}\text{decade}^{-1}$ ) over seven homogenous regions from 1980-2020. Values in bold indicate trends that are not statistically significant ( $p > 0.05$ ). ....	62
<b>Table 3.3:</b> Total and piecewise annual mean wind speed trends in $\text{m s}^{-1}\text{decade}^{-1}$ for observations and reanalyses over seven homogenous regions during the period 1980 to 2020. Values in bold indicate significant trends ( $p < 0.05$ ).....	65
<b>Table 4.1:</b> Number of skillful forecasts ( $\text{FCRPSS} > 0$ ) out of total 2160 forecasts for each region using JRA55 and ERA5 as references.....	89
<b>Table 4.2:</b> Percentage of explained variances of each of the RCs of 10 m wind speed from JRA55 and 6 S2S models over 6 homogeneous regions for 1994-2016.....	93

<b>Table 5.1:</b> Percentage improvements in RMSE among raw and calibrated forecasts across 5 lead times over 6 homogeneous regions during 2019-2021. ....	106
<b>Table 5.2:</b> Percentage improvements in CRPSS among raw and calibrated forecasts across 5 lead times over 6 homogeneous regions during 2019-2021. ....	116

## List of Abbreviations

AMO	Atlantic Meridional Oscillation
ANN	Artificial Neural Network
ARPEGE	Action de Recherche Petite Echelle Grande Echelle
ASOS	Automated Surface Observing System
AWOS	Automated Weather Observing System
BA	Bias Adjustment
CAM	Community Atmosphere Model
CDF	Cumulative Distribution Function
CFS	Climate Forecast System
CMCC	Centro Euro-Mediterraneo sui Cambiamenti Climatici
CNN	Convolutional Neural Network
CRA-40	China Meteorological Administration 40 year Reanalysis
CRPSS	Continuous Ranked Probability Skill Score
CV	Cross-validation
DL	Deep learning
DWD	Deutscher Wetterdienst
EC	East coast
ECCC	Environment and Climate Change Canada
ECHAM	ECMWF Hamburg
ECMWF	European Centre for Medium-range Weather Forecasting
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis version 5
FCRPSS	Fair Continuous Ranked Probability Skill Score

GCFS2.0	German Climate Forecast System version 2
GCM	General Circulation Model
GEOS5	Goddard Earth Observing System-5
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecasting System
GPR	Gaussian Process Regression
GS5-GC2	Global Seasonal forecast system Global Coupled 2
GSOD	Global Summary of the Day
GWEC	Global Wind Energy Council
HP	Hyperparameter
IAV	Interannual Variability
IEA	International Energy Agency
IFS	Integrated Forecasting System
IMD	India Meteorological Department
IMDAA	Indian Monsoon Data Assimilation and Analysis
IP	Interior peninsula
IRENA	International Renewable Energy Agency
IOD	Indian Ocean Dipole
ISMN	Indian Summer Monsoon Rainfall
ISO	Intraseasonal Oscillation
JJA	June July August
JJAS	June July August September
JRA55	Japanese 55 year Atmospheric Reanalysis
LGBM	Light Gradient Boosting Machine
LSTM	Long Short Term Memory

MEI	Multivariate ENSO Index
MERRA2	Modern-Era Retrospective Analysis for Research and Applications version 2
METAR	Meteorological Aerodrome Report
MF6	Météo-France's System 6
MJO	Madden-Julian Oscillation
ML	Machine Learning
MMCFS	Monsoon Mission Climate Forecast System
MME	Multi-model Ensemble
MNRE	Ministry of New and Renewable Energy
MOM	Modular Ocean Model
MPIOM	Max Planck Institute Ocean Model
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NC	North central
NCEI	National Centers for Environmental Information
NCEP	National Centers for Environmental Prediction
NE	North east
NEMO	Nucleus for European Modelling of the Ocean
NMME	North American Multi-model Ensemble
NRMSE	Normalised Root Mean Squared Error
NW	North west
PC	Principal Component
QM	Quantile Mapping
RC	Reconstructed Component
Rcov	Robust coefficient of variation

RF	Random Forest
RMSE	Root Mean Squared Error
RPC	Ratio of Predictable Components
S2S	Subseasonal to Seasonal
SAM	Southern Annular Mode
SEAS5	Seasonal forecasting System version 5
SPS3	Seasonal Prediction System version 3
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
IITM EPS	Indian Institute of Tropical Meteorology Ensemble Prediction System
UKMO	United Kingdom Met Office
WC	West coast
WCRP	World Climate Research Programme
WH	Western Himalayas
WMO	World Meteorological Organisation
WWEA	World Wind Energy Association
WWRP	World Weather Research Programme