
# NAVAL POSTGRADUATE SCHOOL

**MONTEREY, CALIFORNIA**

# THESIS

**TOWARDS A SCORECARD CONFIDENCE METRIC USING A TIME-LAGGED-MULTI-MODEL (TLMM) ENSEMBLE**

by

Shunika S. Johnson

June 2022

Co-Advisors: Wendell A. Nuss
Joel W. Feldmeier,
Office of Naval Research Global

THIS PAGE INTENTIONALLY LEFT BLANK

| REPORT DOCUMENTATION PAGE | | | *Form Approved OMB No. 0704-0188* |
|---|---|---|---|



| **1. AGENCY USE ONLY** *(Leave blank)* | **2. REPORT DATE** June 2022 | **3. REPORT TYPE AND DATES COVERED** Master's thesis |
|---|---|---|

**4. TITLE AND SUBTITLE**
TOWARDS A SCORECARD CONFIDENCE METRIC USING A TIME-LAGGED-MULTI-MODEL (TLMM) ENSEMBLE

**5. FUNDING NUMBERS**

**6. AUTHOR(S)** Shunika S. Johnson

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Naval Postgraduate School
Monterey, CA 93943-5000

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
N/A

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES** The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release. Distribution is unlimited.

**12b. DISTRIBUTION CODE**
A

**13. ABSTRACT (maximum 200 words)**

Consistent and quantitative methods to characterize the run-to-run performance of the Coupled Ocean Atmospheric Mesoscale Prediction System Numerical Weather Prediction (COAMPS NWP) model have not been readily available to naval forecasters. Moreover, an ensemble version of COAMPS is not typically run for uncertainty calculations. Operational forecasters rely on COAMPS output to use tactical decision aids (TDA) and ultimately provide mission-critical forecasts. COAMPS NWP forecasts are deterministic and only provide end-users with a specific value for a given area and time. An ensemble model, such as a Time-Lagged-Multi-Model (TLMM), may provide an acceptable method to combine successive deterministic forecasts into probabilistic values with associated uncertainty. This research is focused on statistical methods to investigate the uncertainty and statistical properties of COAMPS NWP deterministic model data, and ultimately develop a new statistical and analytical approach to capture model uncertainty in parameters used for electromagnetic propagation predictions. This uncertainty could then be shared with the end-user for improved mission planning and decisions. The overall goal of this research is not to improve the performance of COAMPS NWP output, but to assess the performance and uncertainty of a deterministic forecast.

| **14. SUBJECT TERMS** Bayesian statistics, time-lagged model, EM warfare, electromagnetic propagation, numerical weather prediction, mesoscale meteorology, ensemble forecasting, CASPER | **15. NUMBER OF PAGES** 119 |
|---|---|
| | **16. PRICE CODE** |

| **17. SECURITY CLASSIFICATION OF REPORT** | **18. SECURITY CLASSIFICATION OF THIS PAGE** | **19. SECURITY CLASSIFICATION OF ABSTRACT** | **20. LIMITATION OF ABSTRACT** |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UU |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release. Distribution is unlimited.**

# TOWARDS A SCORECARD CONFIDENCE METRIC USING A TIME-LAGGED-MULTI-MODEL (TLMM) ENSEMBLE

Shunika S. Johnson
Lieutenant Commander, United States Navy
BA, Washington & Jefferson College, 2010
MS, West Virginia University, 2012
MS, The University of Southern Mississippi, 2018

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN METEOROLOGY AND PHYSICAL OCEANOGRAPHY**

from the

**NAVAL POSTGRADUATE SCHOOL**
**June 2022**

Approved by: Wendell A. Nuss
Co-Advisor

Joel W. Feldmeier
Co-Advisor

Wendell A. Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

Consistent and quantitative methods to characterize the run-to-run performance of the Coupled Ocean Atmospheric Mesoscale Prediction System Numerical Weather Prediction (COAMPS NWP) model have not been readily available to naval forecasters. Moreover, an ensemble version of COAMPS is not typically run for uncertainty calculations. Operational forecasters rely on COAMPS output to use tactical decision aids (TDA) and ultimately provide mission-critical forecasts. COAMPS NWP forecasts are deterministic and only provide end-users with a specific value for a given area and time. An ensemble model, such as a Time-Lagged-Multi-Model (TLMM), may provide an acceptable method to combine successive deterministic forecasts into probabilistic values with associated uncertainty. This research is focused on statistical methods to investigate the uncertainty and statistical properties of COAMPS NWP deterministic model data, and ultimately develop a new statistical and analytical approach to capture model uncertainty in parameters used for electromagnetic propagation predictions. This uncertainty could then be shared with the end-user for improved mission planning and decisions. The overall goal of this research is not to improve the performance of COAMPS NWP output, but to assess the performance and uncertainty of a deterministic forecast.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AREPS | Advanced Refractive Effects Prediction System |
| BEMOS | Bayesian Ensemble Model Output Statistics |
| BUILDER | Interactive Scenario Builder |
| CASPER West | Coupled Air-Sea Processes and Electromagnetic ducting Research West Coast |
| COAMPS | Coupled Ocean Atmospheric Mesoscale Prediction System |
| ECMWF | European Centre for Medium-Range Weather Forecasts. |
| EM | Electromagnetic |
| EMW | Electromagnetic Maneuver Warfare |
| FNMOC | Fleet Numerical Meteorology and Oceanography Center |
| GFS | Global Forecast System |
| GLM | Generalized Linear Model |
| iid | Independent and Identically Distributed |
| MCMC | Markov Chain Monte Carlo |
| ME | Mean Error |
| METOC | Meteorology and Oceanography |
| MLE | Maximum Likelihood Estimate |
| MSE | Mean Squared Error |
| NAVGEM | Navy Global Environmental Model |
| NCEP | National Centers for Environmental Prediction |
| NOAA | National Oceanic and Atmospheric Administration |
| NWP | Numerical Weather Prediction |
| ONR | Office of Naval Research |
| PPD | Posterior Predictive Distribution |
| R/P | Research Platform |
| RF | Radio Frequency |
| RH | Relative Humidity |
| RMSE | Root Mean Squared Error |
| SL/SI | Semi-Lagrangian/Semi-Implicit |
| SLP | Sea Level Pressure |

| | |
|---|---|
| SST | Sea Surface Temperature |
| TDA | Tactical Decision Aid |
| TLMM | Time-Lagged-Multi-Model |
| USGODAE | U.S. Global Ocean Data Assimilation Experiment |

# ACKNOWLEDGMENTS

First and foremost, I would like to thank God for His protection over my family and me through this journey at Naval Postgraduate School. Without God, nothing is possible.

I would like to give special attention to my family near and far for the continuous love and support. To Anthony and Avery, you have shown me unwavering support despite spending long hours and weekends working on this research and other coursework over the last two years. To my immediate family members in Maryland, West Virginia, Pennsylvania, Georgia, and New Jersey, I would like to express my extreme gratification for having you in my life. Thank you for being a sounding board and always lending an ear. Moreso, thank you for your time and willingness to travel thousands of miles away to visit.

I am incredibly thankful for all the support from my advisors, Dr. Wendell Nuss and CDR Joel Feldmeier, who were willing to progress the thesis forward during COVID, a sabbatical, and in another country. To Mr. Andrew Krammerer, this research may not have been initiated without your proposal on the topic. So for that, I am thankful for your willingness to work with me and to allow me to push your vision forward. To CDR Travis Wendt, your report and code on Bayesian prediction made this thesis successful. Thank you for your assistance, feedback, time, and great advice during the last month of research.

Last but not least, I would like to thank the Naval Postgraduate School faculty, staff, and fellow cohort members. Throughout COVID, we have all learned to adapt and excel in a new learning style. I look forward to working with you all again!

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. MOTIVATION

Understanding the accuracy and performance of electromagnetic (EM) prediction is vital to the U.S. Navy. As EM waves propagate through the atmosphere, EM radiation of various wavelengths will refract, attenuate, reflect, and scatter differently depending on the atmospheric conditions encountered. The U.S. Navy Meteorology and Oceanography (METOC) community primarily uses the Coupled Ocean Atmospheric Mesoscale Prediction System (COAMPS) Numerical Weather Prediction (NWP) model in tactical decision aids (TDA), such as Interactive Scenario Builder (BUILDER) and its predecessor Advanced Refractive Effects Prediction System (AREPS), to predict radar and EM/Radio Frequency (RF) system performance. Consequently, commanders and end-users are interested in the reliability of EM/RF predictions produced by a TDA. Moreover, personal conversation confirms that U.S. Navy METOC personnel are also concerned with the accuracy of the COAMPS data assimilated into the TDA. These uncertainties affect electromagnetic battlespace awareness and management, such as sensor use, sensor and platform placement, weapons selection, communications, and other EM/RF sensors or systems sensitive to atmospheric conditions. From the METOC professional's perspective, error in the model affects confidence and credibility with the warfighter.

The METOC community has been actively supporting Electromagnetic Maneuver Warfare (EMW). Recently, the Commander of Naval Meteorology and Oceanography Command (NMOC) announced an EMW Strategy (Gallaudet 2016). The strategy addressed three main goals: to influence the development of the Navy's EMW capability, to improve METOC environmental sensing and prediction capabilities, and to integrate EM/RF atmospheric impacts into the decision-making process (Gallaudet 2016). This thesis supports the last main goal of improving the forecasters' and commanders' decision-making, especially for enhanced battlespace awareness; improving the forecasters' decision-making depends on understanding the accuracy and performance of EM/RF predictions produced by TDAs reliant on COAMPS model data.

Department of Defense (DOD) released an updated Electromagnetic Spectrum Superiority Strategy (2020) to align DOD forces on EMW superiority, readiness, and implementation. Subsequently, in 2021 the METOC community continued support to EMW through a Large-Scale Exercise. As stated by the Commander Officer of Naval Oceanography Operations Command (NOOC), the Large-Scale Exercise provided "the opportunity to flex our capabilities, especially in regard to the new Electromagnetic Maneuver Warfare (EMW) mission" (U.S. Naval Meteorology and Oceanography Command Public Affairs Office 2021). Consequently, this research continues the EMW support campaign while keeping tactical users in mind.

Knowing and understanding forecast uncertainty can improve probabilistic meteorological understanding and lead to better decision-making. Studies (Göber et al. 2008; Murphy 1993) show that even abundant information could never truly satisfy an inexperienced or complacent forecaster. Conversely, model scorecards are designed to help make use of that information, which enables better decision-making and understanding of forecasts (Gallo et al. 2022). Specifically, forecasters question if EM/RF prediction ranges are in the realm of possibility given the current atmospheric conditions represented by the NWP model output. Furthermore, forecasters may be unsure of the model performance, which may compromise their confidence level in derived products.

As a common practice, forecasters take current atmospheric conditions to initialize the forecast model; i.e., compare the last forecast values with recent observations and adjust forecast values based on the difference between the observations and model output. This process, although common practice, provides only a generalized comparison of model performance for a grided space and particular time. Representativeness within that grid space is not necessarily present or may not be available. Skill during the initialization process is minimal for less experienced forecasters. So it follows, Göber et al. Claim, that "even a perfect model could not nearly satisfy a naïve forecaster who just wants to pass a model value to a customer in need of a point forecast" (2008). However, what is the goal of the forecaster? The goal is to have confidence in the propagation prediction produced by METOC personnel, and for any uncertainty present, clear communication to the warfighter of the error and risk margin. No forecaster is perfect, and no forecast is perfect.

Nonetheless, knowledge enables understanding, leading to better communication and decision-making.

While previous methods to model and understand COAMPS NWP model forecast uncertainty exist, these have not been readily available to naval end-users who rely on it for immediate TDA model assimilation. COAMPS NWP forecasts are deterministic and only provide the end-users with a specific value for a given area and time. COAMPS is a limited-area atmospheric-mesoscale model whose domain is initiated by the Navy Global Environmental Model (NAVGEM), a global numerical weather prediction model. FNMOC is the Navy's Operational Numerical Weather Prediction Center, which provides "weather and ocean prediction products in support of Fleet Safety and Warfighting Effectiveness" (U.S. Naval Meteorology and Oceanography Command Public Affairs Office 2014). FNMOC produces forecast verification metrics for COAMPS and NAVGEM over a month-long period for different levels of the atmosphere and for specific model forecast times.

In general, FNMOC verification products help describe systematic model bias by comparing model performance to observations when available. Specifically, for EM propagation, verification of model parameters near the surface; i.e., FNMOC Surface Observation Bias Verification Parameters, in some ways help identify the sources of error being ingested into the TDAs. However, FNMOC's tools do not inform the users of model-to-model-run bias. Therefore, this tool cannot help with the immediate analysis of forecast consistency between model runs, which could be critical knowledge in a tactical setting with planning and decisions being made in stride. In other applications, the advantages of FNMOC verification products are to determine long-term regional trends and show model bias for fixed observation stations, which naval ships certainly are not.

Moreover, there are various statistical methods, including machine learning, used by other organizations to model and understand NWP model uncertainty. These methods could be applied to naval mesoscale and global ensemble models. Ensemble forecasts from deterministic models result in probabilistic predictions that yield a sample distribution and statistics. Previous studies through the National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Prediction (NCEP), Fleet Numerical

Meteorology and Oceanography Center (FNMOC), and the European Centre for Medium-Range Weather Forecasts (ECMWF) have explored the use of ensemble forecasts to improve the predictability and consistency of weather forecast variables (Branković et al. 1990; Lu et al. 2006; Zhou et al. 2010; Du et al. 2018). Typically, ensembles are made from multiple runs of the same deterministic model, with slight variations leading to different forecasts. However, other variations are possible.

Through a multilinear regression model, Lu et al. (2006) explored two approaches, equally and unequally weighted means, to construct a short-range ensemble model using time-lagged ensemble members; i.e., succeeding runs of the same model rather than multiple runs done at the same time. Their research found that model-derived weights yielded better results for weather prediction variables. Subsequently, other supervised learning methods can produce even better results. Previous doctoral work by Wendt (2017) on multivariate Bayesian ensemble modeling shows that "Bayesian PPDs [posterior predictive distributions] reliably characterized forecast uncertainty and outperformed the parent ensemble and a classical least-squares approach to multivariate multiple linear regression."

An ensemble model, specifically a Time-Lagged-Multi-Model (TLMM), will provide a way to transform COAMPS deterministic forecasts into a probabilistic prediction with associated uncertainty. In addition, the global NAVGEM forecast and COAMPS forecasts produced from various grid resolutions will add to the number of members within a TLMM. It is important to note that there may be some variability between model runs and grid resolutions during the initialization or assimilation of new atmospheric observations and data. As a result, this research focuses on using statistical methods to investigate the uncertainty and statistical properties of the NWP deterministic model data and to develop a new approach to capture model uncertainty in near-surface conditions. Three variations to building the TLMM are explored, and their relative performance will be evaluated. The three TLMM variations are a simple mean, weighted mean, and Bayesian model average. Figure 1 illustrates the concept of a time-lagged-multi-model (TLMM) ensemble. In this example, ensemble members are derived from three sequential COAMPS analyses and a global model analysis.

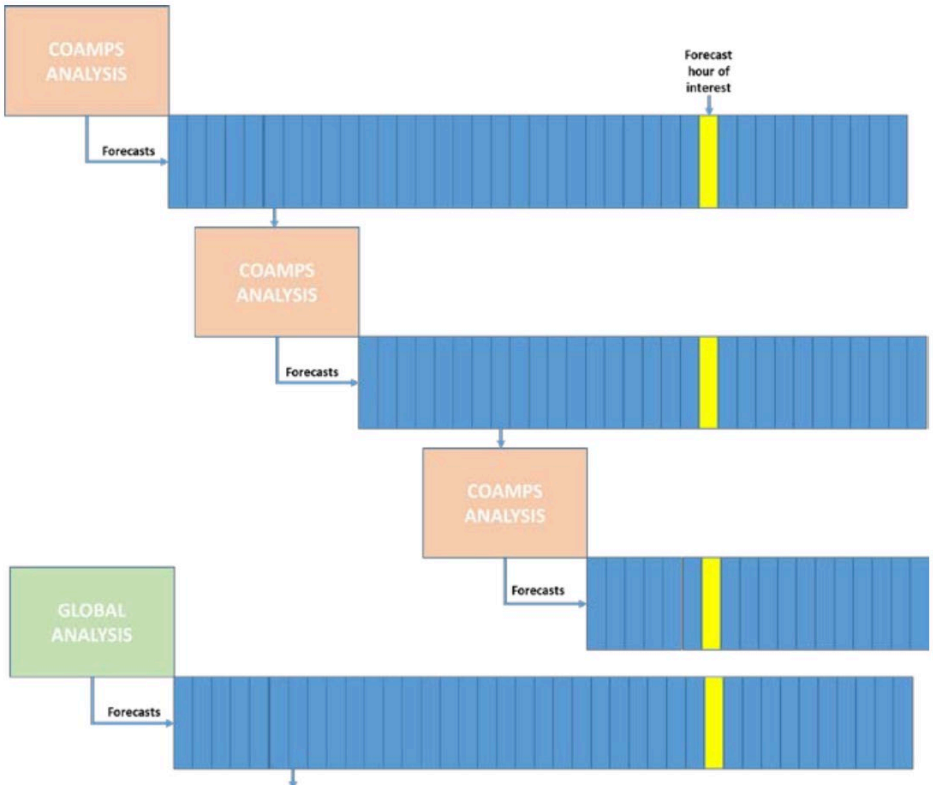


Figure 1. Example of a time-lagged-multi-model (TLMM) ensemble. Ensemble members are three sequential COAMPS forecasts and a global model analysis; i.e., NAVGEM. Source: Naval Research Laboratory Monterey (2020).

## B. OBJECTIVE

This research is conducted with METOC forecasters and users in mind. METOC operators and forecasters requested a scorecard or uncertainty confidence metric to help improve forecaster decision-making and confidence in producing operational TDA products. Therefore, communicating the atmospheric impact to EM/RF performance of radar and weapon systems requires the known variance of each key variable. Although interesting, conventional qualitative and primarily visual analysis of model performance for previous forecast periods is not sufficient to assess the accuracy and uncertainty for a particular forecast time and area of interest.

This research does not aim to improve the performance of COAMPS NWP deterministic output for key variables, but rather to convey uncertainty of the variables to

the end-user. While additional variables to those examined here may ultimately be necessary, this analysis focuses on the most tractable variables as determined in initial data exploration. An objective of this research is to employ selected variables in a TLMM; i.e., simple mean, weighted mean, and Bayesian model averaging, and evaluate the performance of various approaches to build the TLMM. Discussed in later sections of this thesis, the Bayesian regression model performs well compared to the simple mean and weighted mean models. Therefore, it is suggested that COAMPS variability in grid-to-grid differences and model forecasts lead times can be assessed by Bayesian regression statistics and that these parameter estimates can be applied to an operational scorecard.

Another objective of this thesis is to analyze and understand the uncertainty of COAMPS data for specific observational sites. Multiple nested COAMPS grids were constructed over a research area within the Southern California Bight for use in a 2017 field research campaign, the Coupled Air-Sea Processes and Electromagnetic Ducting Research West (CASPER-West). CASPER-West research was funded by the Office of Naval Research (ONR) under its Multidisciplinary University Research Initiative program (Wang et al. 2018). Observational data from three participating sites is utilized in this thesis report.

# II. BACKGROUND

## A. WORKFLOW

The thesis design follows a general workflow process: developing or initializing the model, determining proof-of-concept, and validating the time-lagged model. First, initializing the model consists of exploratory analysis, which considers descriptive statistics and characterization of all variables used in the statistical models. This section of the thesis thoroughly discusses the exploratory analysis and characterization process.

Consequently, the next step is determining proof-of-concept for the TLMM. In this step, three models are fitted to the time-lagged ensemble members. These models include a simple mean, a poor man's ensemble or weighted mean, and a multivariate Bayesian ensemble. For each model, the performance is assessed using common error metrics: mean error (ME), mean squared error (MSE), and root mean squared error (RMSE). Error estimates are calculated by the difference between the observations, the forecast ensemble, and individual ensemble members. These estimates were chosen to provide insight on the error due to the model's fit compared to the observations. Another error statistic for model bias is the variance, which is computed to evaluate the model performance without observation data available. Each model will be characterized by the distribution and dispersion of the testing dataset, biases, model limitations, etc. As such, calibration of ensemble members through variable selection methods is also considered. Between the fourth and fifth steps is model optimization. A summary of the design flow is shown in Figure 2.

The design of the model should be broken into two versions: ideal and observational models. An ideal model contains grids of COAMPS and NAVGEM time-lagged forecast members, and "uncertainty metrics will be calculated discretely for each grid point and forecast hour" (Naval Research Laboratory Monterey 2020). The ideal regression model assigns appropriate weights for each time-lag member, grid point, and forecast hour. As a result, overall uncertainty is calculated for the grid at each forecast hour of interest. Alternatively, the observational model considers COAMPS and NAVGEM time-lagged

members regressed on the observations. The model was chosen to understand and identify model variable characteristics. The NAVGEM model is included because it provides the background forcing of the limited area model, COAMPS. The observational model is the exclusive focus of this thesis.

The last step is to validate the robustness of the model. Ultimately, the aim is to show that the TLMM can be used to create an end-user product for Navy forecasters and operators.

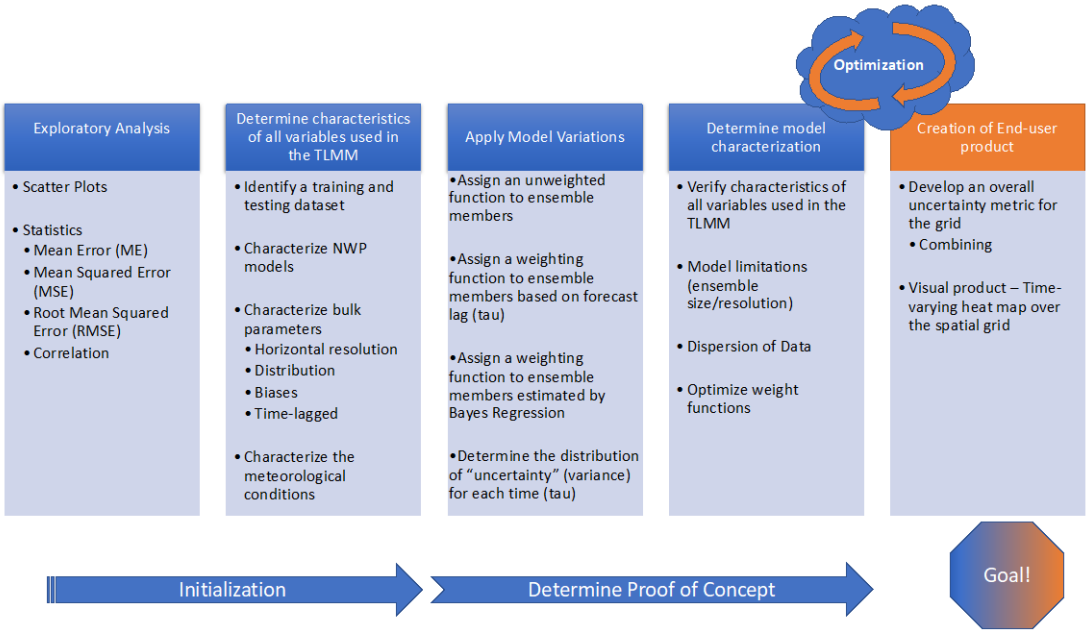


Figure 2. Design workflow chart.

## B. DESCRIPTION OF MODEL DATA

The real atmospheric environment is nonlinear and chaotic (Du et al. 2018) and is modeled by NWP models whose output is deterministic. NWP models are derived from governing dynamic and thermodynamic equations. Atmospheric dynamics consider the changes in atmospheric conditions due to forces affecting the motion of air. Thermodynamics considers the exchange of heat and energy between air masses and parcels, which changes the temperature and moisture properties. Physical properties that

are unable to be represented in the model equations are parameterized. Parameterization may also be necessary for phenomena that are small in scale, complex, and too chaotic to resolve computationally. The level of parameterization can vary based on the model type, such as mesoscale or global, and the weather variables forecasted. Examples of parameterized physical properties are convection, microphysics, surface and planetary boundary layer turbulence, and radiation (Hogan et al. 2012).

Although NWP models have performed well to capture general features in the atmosphere, there are limitations that lead to inherent error. Limitations include imperfect initial conditions from previous forecast runs, boundary conditions, the process for assimilation of observations, model parameterization schemes, computer processing speed and capacity, and model horizontal and vertical resolution (Du et al. 2018). To address NWP models limitations, ensemble models leverage statistical methods to assess the uncertainty of the deterministic forecast and to minimize the impact of model limitations. The subsections below provide a description of each deterministic NWP model, their characteristics, and limitations.

### 1. COAMPS Model Description

COAMPS is the U.S. Navy's mesoscale NWP model developed by Naval Research Laboratory (NRL) and operationally used by FNMOC. COAMPS is a coupled, high-resolution, non-hydrostatic, limited area model driven by NAVGEM (FNMOC 2018). COAMPS' basic atmospheric core solves the fully compressible, non-hydrostatic equations of motion on a system of staggered grid points (Arakawa-Lamb Scheme C) and uses the sigma-height, terrain-following vertical coordinate grid (Naval Research Laboratory Marine Meteorology Division 2003, Hsu et al. 2007). COAMPS output grid resolution is variable, ranging from coarse resolution at the synoptic level to a finer than 1.6 km grid spacing, and confined by a reduction in the scale of 1/3 that of the coarser grid. High resolution is attained through nested grids containing sub-grid physics parameterizations: vertical mixing (Bulk-Richardson number -dependent), surface flux (Louis time-implicit, surface and planetary boundary layer), cumulus (Relaxed Arakawa-Schubert), radiation (Harshvardhan), and other physics parameterizations (Naval Research

Laboratory Marine Meteorology Division 2003, Hsu et al. 2007, FNMOC 2018). For the outer COAMPS grid, boundary conditions are determined from NAVGEM output. Similarly, for the inner COAMPS nests, lateral boundary conditions are forced by the COAMPS outer nest at three times the number of grid points (FNMOC 2018).

COAMPS has been frequently updated since its debut in June 1997, when a real-time forecast capability was tested aboard USS Nimitz (FNMOC 2018). Six years later, COAMPS Version 3.0 began using the Navy Atmospheric Data Assimilation System (NAVDAS) 3-D variational (3DVAR) to assimilate more observation sources into its forecast model analysis (Naval Research Laboratory Marine Meteorology Division 2003, FNMOC 2013). COAMPS Version 5.0 came online in January 2012, and upgrades included a coupled ocean/atmosphere capability, new microphysics parameterization, and a new NOAA land surface model parametrization (FNMOC 2018). Most recently, the COAMPS upgrade from v5.2 to v5.6 "increase [d] model skill by improving aerosol forecast and tracking fields, land surface model initialization, and atmospheric physics" (FNMOC 2020). There are continuing improvements to all specifications and properties of the COAMPS NWP system.

The TLMM analysis uses COAMPS data from 2017. It is recognized that COAMPS has been updated since then, but the overall TLMM approach should still yield interesting results. COAMPS data for the CASPER-West field campaign was collected from 06 September to 27 October 2017, offshore of Southern California. The COAMPS domain configuration was defined to produce four nested grids: 54 km, 18 km, 6 km, and 2 km (denoted as g01, g02, g03, and g04, respectively). Figure 3 depicts the COAMPS nested grids at 4 levels over the study area. Additionally, hourly forecasts spanning at most 24 hours were produced at 00, 06, 12, and 18 UTC. At analysis times of 12 UTC and 00 UTC, forecasts went out to 24 hours; however, at an analysis time of 06 UTC and 18 UTC, forecasts only extended to 6 hours. Because of the inconsistency in forecast lengths, the number of members available for the TLMM is limited. Moreover, CASPER-West field operations did not commence until 27 September 2017; hence, there were a few analysis times when the COAMPS forecast was not produced.

COAMPS has many meteorological parameters available as outputs. Specific to the TLMM analysis, available output variables are as follows:

- 10 m Wind Speed and direction ($u$, $v$)
- 2 m Air Temperature (T)
- Sea Level Pressure (SLP)
- 2 m Relative Humidity (RH)
- Sea Surface Temperature (SST)

These variables reside in the lowest layer of the COAMPS CASPER-West dataset. COAMPS has a vertical resolution of 60 user-specified sigma layers (FNMOC 2018). For the TLMM study, error in the vertical boundary conditions is not considered because observations and model data are constrained to the lowest 10 m of the atmosphere.

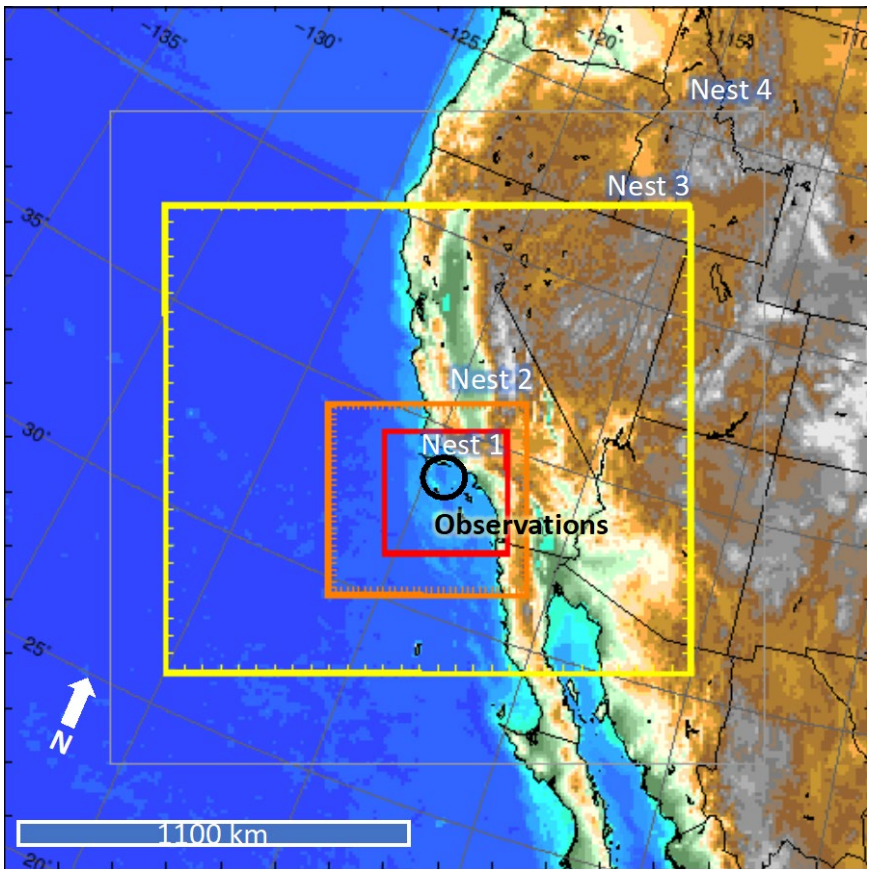


Figure 3. Visual of COAMPS nested grids at four levels with observation centered in the finest resolution grid. Adapted from Naval Research Laboratory Monterey (2017).

## 2. NAVGEM Model Description

NAVGEM is the U. S. Navy's global NWP model also developed by NRL and operationally used by FNMOC (Hogan et al. 2014; FNMOC 2018). NAVGEM is a global weather model that is designed to accurately predict synoptic or large-scale weather patterns. NAVGEM's basic atmospheric core is based on the primitive dynamic equations with hydrostatic approximation and is approximated by "a 3-time level, semi-Lagrangian/ semi-implicit (SL/SI) numerical integration of the adiabatic equations of motion and the first law of thermodynamics" (FNMOC 2018).

NAVGEM first became operational on 13 February 2013 and has seen many updates. With the 2015 update, NAVGEM 1.3 increased its spectral resolution from T359 to T425, with the Gaussian grid of 1080 x 540 grid points increased to1280 x 640 grid

points and grid spacing reduced from 37 km to about 31 km in the tropics. The NAVGEM 1.4 update in October 2016 maintained the same horizontal resolution and basic dynamical core package. It is important to note that NAVGEM 1.4 is the version of data available for this thesis, and biases associated with this version are known. In April 2020, FNMOC announced NAVGEM 2.0, "a higher horizontal resolution forecast model with improved scalability, dynamics and physics-dynamics coupling compared to NAVGEM 1.4" (FNMOC 2020). Important differences in the update were the improved horizontal resolution, improved numerical technique, and updated physics-dynamics coupling. This update acknowledged general bias in the NAVGEM 1.4 version, and a preliminary validation test showed enhanced performance in frontal placement, low-level clouds representation, and wind speed bias (FNMOC 2020). Nevertheless, all versions of NAVGEM output model parameters on half-degree latitude by half-degree longitude grids for various height or pressure levels. An update to NAVGEM 1.4.3 output model field enables quarter degree horizontal resolution permitting a higher detail of atmospheric structure (FNMOC 2018). A summary of some highlighted NAVGEM model specifications is included in Table 1.

Because the domain of COAMPS is initiated by NAVGEM, this research will use the NAVGEM model to add to the number of members in the TLMM ensemble. NAVGEM is run four times a day for an analysis time of 00, 06, 12, and 18 UTC. NAVGEM produces forecasts over an extended time period, allowing for 20 possible TLMM members over a 5-day forecast period for 6-hour times steps. The NAVGEM dataset was obtained from U.S. Global Ocean Data Assimilation Experiment (USGODAE) site of archived oceanographic and atmospheric datasets. The NAVGEM data overlapped the COAMPS data for 26 September to 26 October 2017.

All model parameters at their respective heights selected in the COAMPS dataset will be also used in the NAVGEM dataset. For the sea surface temperature (SST) variable, the data comes from the U.S. Navy Coupled Ocean Data Assimilation (NCODA) analysis and is analyzed on the NAVGEM computational grid. By design, the NAVGEM model holds SST constant through the forecast. For many applications this has proven practical,

as SST is a slowly varying parameter. Therefore, only the 00 UTC analysis is available in the NAVGEM model output.

Table 1. Summary of highlighted COAMPS and NAVGEM model specifications for 2016 to present. Sources: Naval Research Laboratory, FNMOC (2018), and FNMOC (2020).

| Version | COAMPS 5.6 | NAVGEM 1.4 | NAVGEM 2.0 |
|---|---|---|---|
| Status | Current | Replaced (Operational October 2016 to April 2020) | Current |
| Basic Equations | Primitive equations with non-hydrostatic approximation | Primitive equations with hydrostatic approximation | Primitive equations with hydrostatic approximation |
| Numerical Technique | Arakawa-Lamb Scheme C | Three-time-level semi-Lagrangian/semi-implicit (SL/SI) time differencing (Ritchie et al.1995) with spectral differencing for derivatives. | Two-time-level SL/SI time differencing. |
| Horizontal Resolution | User specified, triple nested with a factor of three | T425 (~0.28 º/31 km on the Gaussian grid) | T681 (~0.17/19 km on a Gaussian grid) |
| Vertical Resolution | Hybrid sigma-height, terrain-following staggered grid, user defined levels from surface to top of the atmosphere (~ 10 mb). | Hybrid sigma – pressure; 60 sigma levels with approximately 14 sigma levels below 850 mb, depending on terrain elevation | Same |
| Updates | Improved horizontal grid spacing, more options for vegetation types, improved physics and convective parametrization | | Improved horizontal resolution, improved scalability, updated physics-dynamics coupling |

## C. DESCRIPTION OF OBSERVATION DATA

### 1. CASPER-West Observation Platforms

An objective of this thesis is to analyze and understand NWP uncertainty for specific observational sites. Atmospheric and oceanographic observational data were collected during the CASPER-West field experiment in 2017. The CASPER-West research represented a substantial collaboration between Office of Naval Research (ONR), Naval Postgraduate School, Naval Surface Warfare Center Dahlgren Division, U.S. Naval Research Laboratory, Scripps Institution of Oceanography, and more organizations (Wang et al. 2018). Because of the availability and quality of data obtained by this research and

its application to the EMW mission, observational data from three participating sites will be utilized in this report.

Point measurements were obtained from the Research Platform Floating Instrument Platform (R/P FLIP) and two surface buoy stations, number 21 and 22, for 26 September 2017 0600 UTC to 26 October 2017 0600 UTC. Figure 4 shows the locations of the observing sites within the Southern California Bight. The platforms were largely stationary for the month-long observing period. Positioning information was recorded continuously, and its variability is considered in the analysis.

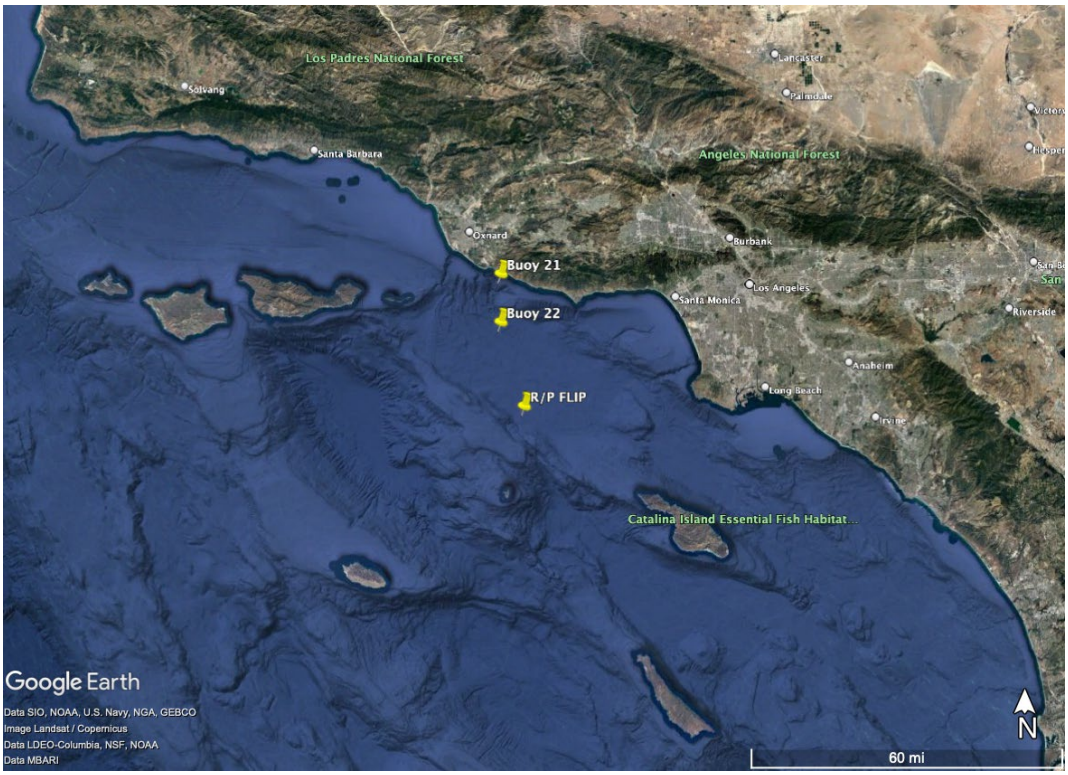


Figure 4. Location of observational sites (Buoy 21, Buoy 22, and R/P FLIP platform). Source: Google Earth (2021).

All platforms were passive continuous collection sensors. Meteorological sensors on each platform allowed for collection of in situ measurements of pressure, wind speed and direction, temperature, humidity, and sea surface temperature. There were no data coverage gaps observed for Buoy 21 and Buoy 22; however, for the R/P FLIP, observations

did not start until 29 September 2017 at 0000 UTC and concluded on 23 October 2017 at 1800. UTC. Processing and quality control of observations were done by the CASPER-West team to provide values every hour for comparison with forecast model output.

The R/P FLIP is a 108 m Floating Instrument Platform constructed in 1962 and designed to support the U.S. Navy SUBROC (SUBmarine ROCket) program (Scripps Institution of Oceanography). The original design of the platform enables rotation and submersion of its elongated tail under the water. While in operation, the platform can withstand extended motion due to waves and winds. Modification and reapplication of R/P FLIP now supports meteorology and oceanographic research efforts, such as that obtained during the CASPER campaigns. Figure 5 is an operational image of R/P FLIP during the 2017 CASPER-West experiment.


Figure 5. Image of R/P FLIP during the 2017 CASPER-West experiment with key features identified. Source: Ortiz-Suslow et al. (2019).

Identifying sources of errors and ensuring quality control of the data is a significant consideration of the analysis. Quality control of the buoy data was accomplished by a previous Naval Postgraduate School (NPS) thesis student (Yanez, 2021) who used the data in support of atmospheric boundary layer machine learning methods. During the CASPER-West experiment, atmospheric measurements from the R/P FLIP were continuously monitored for accuracy. R/P FLIP data processing and quality control of the observations considered stability of the platform with respect to wave motion, platform heading and orientation, reliability of sensors during operations affecting retention of measurements, calibration of the sensors prior to deployment, and potential sources of interference affecting the quality of data (Ortiz-Suslow et al., 2019).

## 2. Weather Summary

In general, the weather in the Southern California Bight is affected by synoptic high-pressure, transient low-pressure systems, and mesoscale phenomena, such as diurnal heating and alongshore boundary winds. Climatologically, weather observed during the observation period was typical for September to October meteorological influences on temperature, relative humidity, sea surface temperature, surface pressure, and wind speed and direction. The observation period of the study is from 26 September to 26 October 2017. A summary of the weather history was obtained from the National Weather Service and provided in Table 2. A synopsis for the day prior to the observation period is also provided to give context of the weather conditions prior to the start of the observing period.

Weather in the Southern California Bight was affected by 4 offshore winds or Santa Anna Wind events (SAWE), several Catalina Eddies, and a few occurrences of transient upper-level troughing. The onset of SAWEs was observed on 26 September, 09 October, 14 October, and 22 October 2017. Characteristically, SAWEs in the Southern California region consist of strong, adiabatic offshore winds which produce hot and dry conditions with clear skies (Naval Research Laboratory Monterey, 2022). Each SAWE varied in length, strength, and duration; and consequently, each event was captured differently in the models. Similarly, Catalina Eddies were also a significant weather phenomenon for the region. Catalina Eddies are short onshore wind events that result in cool, breezy winds and

low-level clouds and fog. Propagating low-pressure systems and troughing reaching the Southern California Bight were another significant weather phenomena. Under these conditions, northwesterly flow constrained by coastal topography increases the onshore sea breeze, low clouds remain present, temperatures are observed to be cooler than average, and gusty wind conditions are possible within the Channel Islands and along the southern coast. Figure 6 shows synoptic charts and gives an example for each significant weather phenomena.

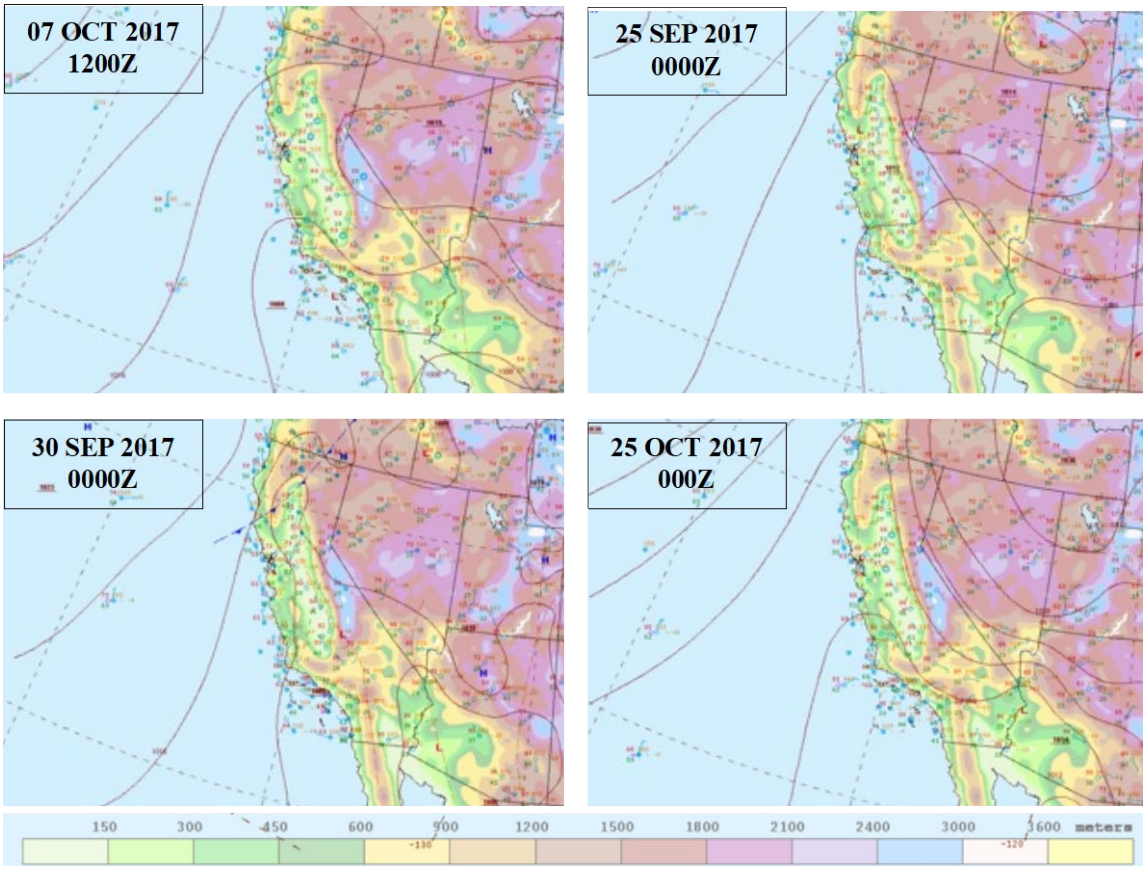


Figure 6. NOAA Weather Prediction Center (WPC) Surface Weather Map with plotted station observations. Snapshots show SAWEs (25 September and 25 October), a Catalina Eddy (07 October), and a Pacific Northwest trough (30 September). Source: NOAA (2017).

Table 2. Summary of weather conditions for the Southern California Bight for 25 September to 26 October 2017, obtained from NWS forecast area discussions. Adapted from Iowa State University – Iowa Environmental Mesonet (2001).

| Date | | Summary | SAWE | Eddy | Synoptic Pressure System |
|---|---|---|---|---|---|
| September 25, 2017 | Monday | There is a gradually strengthening offshore flow causing a warming and drying trend and clear skies. The gradient offshore or inverted pressure trough and upper-level support help to generate gusty conditions over the region. | X | | |
| September 26, 2017 | Tuesday | N/NE offshore flow is weak, but warm and dry conditions with clear skies persist over the area. Temperatures are above average. There is a weak eddy generating, bringing low-level marine layer clouds. | X | | |
| September 27, 2017 | Wednesday | There is weak offshore flow and high pressure aloft. Temperatures remain above normal. Patchy low clouds are present due to the weak eddy. | | X | |
| September 28, 2017 | Thursday | Conditions from the previous day are the same. There is a slowly approaching board upper-level trough that will flatten the upper-level ridging and provide cooling temperatures. | | X | |
| September 29, 2017 | Friday | Conditions are influenced by the Pacific Northwest Low which flattens the flow over California. Onshore or NW flow increases the sea breeze and low clouds. Temperatures are cooler, and winds are gusty at the Santa Barbara south coast. | | | X (Weak trough) |
| September 30, 2017 | Saturday | The Pacific Northwest trough extends south to the Great Basin. There is onshore flow with cooler temperatures and a deepening marine layer. | | | X |
| October 1, 2017 | Sunday | A positive tilt in the trough produces a cut off low and broad synoptic scale cyclonic flow affects the weather conditions. Temperatures remain cool, winds are gusty, and morning fog and low-level clouds persist. | | X | X |
| October 2, 2017 | Monday | Conditions from the previous day are the same. | | X | X |

| Date | | Summary | SAWE | Eddy | Synoptic Pressure System |
|---|---|---|---|---|---|
| October 3, 2017 | Tuesday | Conditions from the previous day are the same. Onshore flow keeps temperatures below normal, and there is overnight clouds and fog present. | | | |
| October 4, 2017 | Wednesday | Building high pressure begins to affect the weather in the region. | | | |
| October 5, 2017 | Thursday | Building high pressure produces an offshore flow of winds. There is a warming trend with mostly clear skies. | | | |
| October 6, 2017 | Friday | There is weak to moderate gusty offshore wind due to the weak/moderate offshore gradient and weak upper-level support. Temperatures warm into Saturday. | X | X | |
| October 7, 2017 | Saturday | With the developing eddy circulation, there is a return to onshore flow with associated low clouds and coastal cooling. | | X | |
| October 8, 2017 | Sunday | The onshore flow brings low clouds and significant cooling. Clouds are present due to the Catalina Eddy effects. | | X | |
| October 9, 2017 | Monday | Catalina Eddy effects are seen in the morning. There is a transition to a significant Santa Anna Wind event which causes damaging winds and an increase of temperatures. | X | | |
| October 10, 2017 | Tuesday | Offshore wind conditions are weak to locally moderate; however, warm and dry conditions with patchy marine layer clouds remain. | X | | |
| October 11, 2017 | Wednesday | There is a broad upper-level trough over California and will persist for a couple days. The upper-level trough brings cold air advection to the region. Temperatures are slightly cooler, onshore winds are gusty, and low clouds are present. | | | X |
| October 12, 2017 | Thursday | Conditions are like the previous day. Onshore winds help bring slightly cooler temperatures with low coastal clouds. | | | X |
| October 13, 2017 | Friday | There is a weak gradient offshore which offshore flow begins to slowly develop. Gusty and warm temperatures are expected with the offshore flow. | | | X |
| October 14, 2017 | Saturday | Ridging begins to build in the region and gusty offshore flow increases. Santa Anna wind conditions are slow to develop because of a lack of upper air support. The gradients are rapidly trending offshore as high pressure builds across the Great Basin, but winds are purely gradient driven (Not a classical SAW). | X | | |

| Date | | Summary | SAWE | Eddy | Synoptic Pressure System |
|---|---|---|---|---|---|
| October 15, 2017 | Sunday | The Santa Anna Wind event brings above normal temperatures and gusty conditions. | X | | |
| October 16, 2017 | Monday | The offshore gradient relaxes but offshore winds and warm conditions remain. | X | | |
| October 17, 2017 | Tuesday | The offshore gradient is weak, and the offshore flow is weak and light. Ridging over the area helps keep temperatures above normal. | X | | |
| October 18, 2017 | Wednesday | There is a weak short-wave trough with some mid/upper-level moisture transiting through the region. There is minimal impact to clouds, and temperatures slightly cool. | | | X |
| October 19, 2017 | Thursday | A significant trough moves into the Pacific Northwest and pushes out upper-level ridge. The flow reverses to onshore (west-northwest wind). | | X | X |
| October 20, 2017 | Friday | Weather conditions are affected by the trough through Friday night. Winds increase with the passing trough. | | | X |
| October 21, 2017 | Saturday | Similar conditions to the previous day. Over the region are clouds, gusty winds, and showery weather. Building high pressure and offshore winds develop behind the trough and persist through Wednesday. | | | X |
| October 22, 2017 | Sunday | There is a strengthening offshore gradient associated with a Santa Anna Wind Event. Gusty offshore winds become widespread and increase. Warm and dry air influences the region. | X | | |
| October 23, 2017 | Monday | Strong high pressure aloft with the gusty Santa Anna winds create above normal temperatures and dry conditions. | X | | |
| October 24, 2017 | Tuesday | Conditions from the previous day are the same. | X | | |
| October 25, 2017 | Wednesday | Conditions from the previous day are the same. There are some isolated gust events over the region. | X | X | |
| October 26, 2017 | Thursday | The offshore gradient relaxes, and offshore winds weaken. Cooler temperatures and onshore flow returns. | | X | |

THIS PAGE INTENTIONALLY LEFT BLANK

# III. DATA CHARACTERIZATION

## A. DATA REVIEW

In accordance with the design flow, this section is intended to characterize all model grids and forecast variables in the TLMM ensemble. An exploratory analysis is a simple numerical and graphical summary which helps to illustrate data characteristics and address how resolution or grid-to-grid errors translate into uncertainty in ensemble prediction. Data characterization precedes the methodology chapter (IV) because it is important to understand what statistical model assumptions can or cannot be made, which subsequently affects the type of TLMM ensembles applied to the data. Depending on the ensemble model applied to the TLMM members, error in the grids and variables may be reduced in the final uncertainty calculations.

An interest of this thesis is to analyze atmospheric variables that capture the greatest variability in EM near surface propagation. These variables are air temperature, sea surface temperature, sea level pressure, relative humidity, and wind speed and direction. The selection of these variables relates to Bean and Dalton (1966), who elegantly summarize the effect of the troposphere on EM wave propagation:

> The atmosphere causes a downward curvature of horizontally launched radio waves which is normally about one quarter of that of the earth. Under unusual meteorological conditions, however, the radio energy may be confined to thin layers near the earth's surface with resultant abnormally high field strengths being observed beyond the normal radio horizon. At other times a transition layer between differing air masses will give rise to the reflection of radio energy. In addition to these gross profile effects, the atmosphere is always more or less turbulent, with the result that radio energy is scattered out of the normal antenna pattern.

Moreover, Bean and Dalton (1966) state that only the effects of $CO_2$, dry air, water vapor, and absolute temperature influence the wave field. Therefore, apart from $CO_2$, only these principal atmospheric variables will be considered in the data characterization of the TLMM.

However, the TLMM analysis also considers wind speed and direction as important parameters because wind speed and direction determine advection and mixing of

temperature, moisture, and air molecules (Wallace and Hobbs 2006). Wind speed and direction are decomposed into their *u*- (West-East direction) and *v*-components (North-South direction) magnitudes. *U*- and *v*-component winds were available from the model output, but observations only reported wind speeds and directions. In meteorology, wind direction is read as the direction from where the wind is blowing with respect to true north. Therefore, the conversion between wind components and speed and direction is calculated by vector calculus from the following relationships:

$$WS = \sqrt{u^2 + v^2}$$

$$WD \equiv \left[180° + \frac{180°}{\pi}\tan^{-1}\left(\frac{u}{v}\right)\right] \ (mod\ 360°)$$

$$u = -WS * \sin(WD)$$

$$v = -WS * \cos(WD) \tag{1}$$

where *WS* is the wind speed in meters per second and *WD* is the wind direction in degrees (ECMWF 2021).

COAMPS data used in this thesis has been tailored and exploited for application to the TLMM ensemble. COAMPS data for the CASPER-West field experiment was run at a temporal output of one-hour forecasts and at model resolutions of 54 km, 18 km, 6 km, and 2 km grid spacings from 06 September to 28 October 2017 offshore Southern California. All nested COAMPS grids are employed in the TLMM analysis; however, with consideration of the METOC operator in mind, only the six-hour outputs are chosen. Moreover, observations made during the CASPER-West field experiment only span a month; therefore, the COAMPS data is subset for 26 September 0000 UTC to 26 October 0000 UTC.

Similarly, NAVGEM data was selected to overlap information available in the COAMPS dataset. NAVGEM is run four times a day for an analysis time of 00, 06, 12, and 18 UTC. NAVGEM sea surface temperature is only computed at the analysis time of 0000 UTC and is anticipated to remain constant for the 24-hour period. As a result, there

are few forecast variables, with five TLMM members spanning the 24-hour preferred time-lag window; i.e., there are 0-, 6-, 12-, 18-, and 24-hour forecasts.

Point observations as shown in Figure 4, were compared directly with the gridded model forecasts. The forecast model was interpolated to each observational site using a bilinear interpolation function discussed in the methodology chapter (IV, Section A). For each grid resolution of COAMPS and NAVGEM, this interpolation process consists of identifying the four grid points nearest to the observation point and performing a scaled average of the forecast variable to the observation location. However, because the COAMPS and NAVGEM datasets were incomplete for all forecast times and variables, the distribution of the data after all missing values are removed is given in Table 3. The number of observations for each forecast variable and grid resolution ranged from 666 to 1670 observations.

Table 3. Summary of COAMPS and NAVGEM dataset for all locations during the observation period.

| Forecast Variable | Grid Resolution | Data Range | Forecast Tau Available | | | | | Number of Obs. |
|---|---|---|---|---|---|---|---|---|
| | | | 00 | 06 | 12 | 18 | 24 | |
| 2 m Air Temperature | COAMPS – 01 | 287.77 – 300.61 K | X | X | X | X | X | 1177 |
| | COAMPS – 02 | 285.83 – 300.69 K | | | | | | |
| | COAMPS – 03 | 283.79 – 301.65 K | | | | | | |
| | COAMPS – 04 | 283.05 – 301.65 K | | | | | | |
| | NAVGEM | 287.56 – 308.93 K | X | X | X | | | 1017 |
| Sea Level Pressure | COAMPS – 01 | 1004.12 – 1021.16 mb | X | X | X | X | X | 1191 |
| | COAMPS – 02 | 1006.07 – 1021.14 mb | | | | | | |
| | COAMPS – 03 | 1006.06 – 1021.36 mb | | | | | | |
| | COAMPS – 04 | 1006.21 – 1021.71 mb | | | | | | |
| | NAVGEM | 1004.59 – 1020.63 mb | X | X | X | X | X | 1029 |
| Relative Humidity (pct) | COAMPS – 01 | 27.13 – 92.58 | X | X | X | X | X | 1177 |
| | COAMPS – 02 | 36.40 – 94.47 | | | | | | |
| | COAMPS – 03 | 34.80 – 98.01 | | | | | | |
| | COAMPS – 04 | 34.50 – 96.95 | | | | | | |
| | NAVGEM | 23.18 – 88.15 | X | X | X | | | 1017 |
| Sea Surface Temperature | COAMPS – 01 | 291.50 – 293.75 K | X | X | X | X | X | 1157 |
| | COAMPS – 02 | 291.57 – 295.00 K | | | | | | |
| | COAMPS – 03 | 290.27 – 295.62 K | | | | | | |
| | COAMPS – 04 | 289.40 – 295.68 K | | | | | | |
| | NAVGEM | 286.12 – 306.38 K | X | X | | | | 666 |
| U-Component Wind | COAMPS – 01 | -11.11 – 12.03 m/s | X | X | X | X | X | 1159 |
| | COAMPS – 02 | -13.27 – 12.43 m/s | | | | | | |
| | COAMPS – 03 | -13.77 – 12.51 m/s | | | | | | |
| | COAMPS – 04 | -15.75 – 13.76 m/s | | | | | | |
| | NAVGEM | -7.97 – 13.76 m/s | X | X | X | X | X | 1670 |
| V-Component Wind | COAMPS – 01 | -9.24 – 3.27 m/s | X | X | X | X | X | 1159 |
| | COAMPS – 02 | -15.34 – 3.24 m/s | | | | | | |
| | COAMPS – 03 | -12.12 – 4.82 m/s | | | | | | |
| | COAMPS – 04 | -15.72 – 3.31 m/s | | | | | | |
| | NAVGEM | -7.07 – 5.38 m/s | X | X | X | X | X | 1670 |

From Table 3, *u*- and *v*- component winds are the focus of this analysis. That is, the *u*- and *v*- component winds dataset is the most complete that allows for a thorough investigation. Moreover, other noteworthy reasons for choosing *u*- and *v*- component winds include the following:

1. The wind speed and direction are very important parameters that help to define and quantify advection of temperature, moisture, and air particulates (Wallace and Hobbs 2006).

2. Geographically, the Southern California Bight has a consistent wind regime which is confined by the coastal topography. Semi-permanent high-pressure and coastal troughing help direct winds through the Channel Islands and along the Californian coast.

3. For specific weather phenomena, such as SAWEs, winds are downslope, or katabatic, from the interior of California and funnel through passes out to the Southern California Bight. (Rolinski et al. 2019)

4. Wind direction and speed influence local weather, and winds change rapidly enough to observe the diurnal and daily variability.

Subsequently, the characteristic of wind that can influence the other forecast variables is summarized as advection. Bean and Dalton (1966) define advection as "the horizontal flow of air having different heat properties… [and] it may lead to a different rate of exchange of heat and moisture between the air and the underlying ground or ocean surface, thus affecting the physical structure of the lowest layers of the atmosphere." Consequently, winds remain a significant impact on the region of interest. It is noted that this approach to using the *u*-, *v*-component winds may not apply everywhere. This approach is most applicable to comparable coastal regions.

Furthermore, the data was then rearranged and concatenated to be in a time-lag format, in which grid resolution is ignored in the final data frame. There are 334 observations in the final data frame of *u*- and *v*- component winds (i.e., no values are missing). The data frame was significantly reduced; however, a sufficient number of observations remained, facilitating a reliable analysis.

## B. ERROR STATISTICS

Error and model performance is assessed using standard error estimates, which are calculated from the mean bias and squared error differences. Depending on how the data is analyzed, error estimates are calculated across each grid, for each forecast parameter, across platforms, and for each TLMM ensemble member. Equations used to calculate the error are as follows:

$$ME = \sum \frac{(y-x)}{N} \tag{2}$$

$$MSE = \sum \frac{(y-x)^2}{N} \tag{3}$$

$$RMSE = \sqrt{\sum \frac{(y-x)^2}{N}} \tag{4}$$

where $x$ represents the forecast value, $y$ represents the observation, *ME* is the mean error, *MSE* is the mean squared error, and *RMSE* is the root mean squared error. These estimates provide a quantitative measurement of the spread and deviations of the forecast values to the observations.

Statistically, the *ME*, *MSE*, and *RMSE* are biased estimates of the variance in the model-to-observation differences, such that the variance of the difference is modeled as

$$\sigma_d^2 = \sum \frac{(d-\mu_d)^2}{N}, \tag{5}$$

where $d = y - x$ is the model-to-observation difference, $\mu_d$ is the mean difference (should expect it to be zero for a perfect forecast), and N represents the number of observation-model pairs in the dataset.

The correlation between the forecast and the observation is also of interest. Correlation is the linear association of dependence between random variables irrespective of the variables' scale of measurement (Härdle and Simar 2003). Correlation is derived from the standard deviation and covariance of $x$ and $y$ and is given by

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{6}$$

Where $\sigma_{xy}$ represents the covariance of $x$ and $y$, $\sigma_x$ is the standard deviation of $x$, and $\sigma_y$ is the standard deviation of $y$ (Härdle and Simar 2003). Because each variables' standard deviation scales the covariance, the correlation falls between -1 and 1. A correlation close to 1 indicates a strong positive linear relationship; likewise, a correlation close to -1 indicates a strong negative linear relationship. Therefore, correlation values close to zero

imply a weak or absent linear relationship. Practically, scatterplots can help assess the correlation strength. When points align closely to a linear line, the correlation is strong.

## C. ERROR ANALYSIS

An element of determining the characteristic of all variables used in the TLMM is an error analysis: descriptive, numerical, and graphical exploratory data analysis on all variables of interest (Härdle and Simar 2003). While numerous techniques for analyzing quantitative variables can be used, scatterplots are effective for graphically comparing two quantitative variables and looking at relationships. Scatterplots of forecast parameters to observations also permit detection of groups or cluster of points. Direct comparisons in scatterplots are made in each grid level; i.e., g01, g02, g03, g04, and NAVGEM, to understand error in the forced lateral boundary conditions and initial conditions provided by the coarser resolution grids. Direct comparisons can be made because observations locations were interpolated to model grids (see Chapter IV.B) to obtain specific forecast values.

Ideally, when the forecast and observation are consistent, we would observe a perfect linear relationship in the scatterplots. Otherwise, scatterplots may indicate when the forecast model under- or overforecasts the atmospheric conditions. These bivariate relations are seen in the scatterplots provided in the subsections below. Most scatterplots have a dashed line representing the perfect linear relationship between two variables. Plots that fall above the dash line indicate the model under forecasting the observed conditions; on the other hand, values below the dashed line indicate an overforecast model value. Nevertheless, the initial summary and analysis focus is to characterize the data and provide context for the methodology section (IV).

### 1. Temperature Error Analysis

For all grid resolutions, a temperature error analysis is performed on the 2-meter air temperature, sea surface temperature, and air-sea temperature difference. True atmospheric conditions are referenced during the temperature analysis to better associate periods when the model's performance decreased. During September and October, temperature variations in the Southern California area are associated with the directional

change of onshore to offshore flow conditions. Onshore flow conditions are most persistent and are driven by thermal inland low pressure and coastal inverted troughing (NRLMRY). Intense periods of offshore flow conditions are characteristic of SAWEs: advection of warm, dry desert winds originating from the canyons and mountains of inland California (NRLMRY). At the air-sea interaction, fluxes in the lower boundary conditions result from highly variable air temperatures caused by normal flow conditions and SAWEs.

Generally, each COAMPS grid resolution has a similar error and shape of dispersion. Figure 7 (left) displays COAMPS data at the finest resolution. Note that for this plot and all COAMPS 2 m air temperature plots, data is available for all 6 hourly forecast periods up to the 24-hour forecast. In Figure 7 (left), forecasted 2 m air temperatures at about 290 K are roughly the same as what was seen in the actual atmosphere. However, for values greater than approximately 292 K, the model mostly overforecasts the temperature. Conversely, there are times when the model severely underpredicts the air temperatures by as much as 9 K.

As for the coarsest resolution grid of 2 m air temperatures, NAVGEM has the most error observed and consistently under forecasted the air temperatures. A scatterplot of forecasted versus observed 2 m air temperatures is provided in Figure 7 (right). Note that NAVGEM data is only available out to the 12-hour forecast, which is a limitation of the data available in the USGODAE repository. The forecasted 2 m air temperature range is more significant than that observed in the COAMPS grid. Additionally, there are periods where the model overforecasts the temperatures. Interestingly, these plots also show model initialization error, where there is a range interpolated forecast values for the same observation value. A data filter by observation platform shows this statement to be true among all platforms.

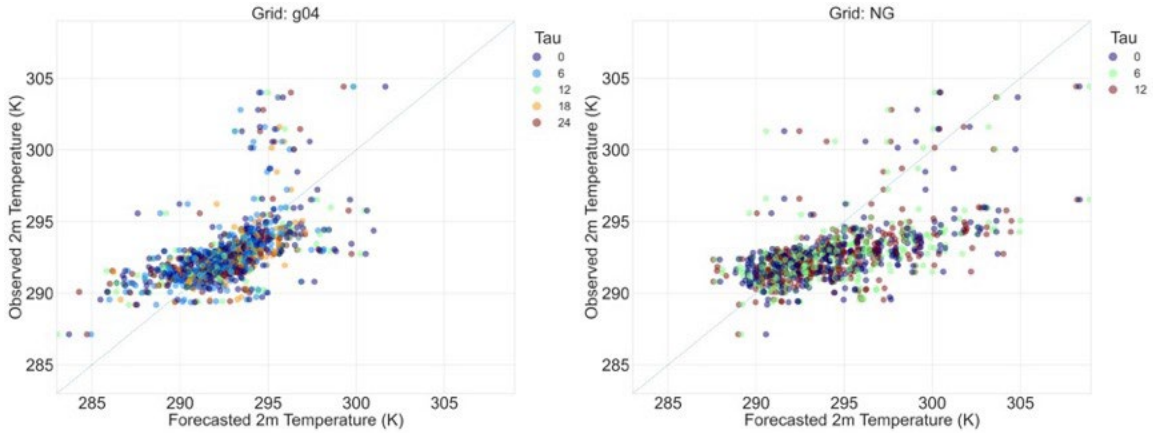


Figure 7. Scatterplots of forecasted versus observed 2 m air temperatures at all observation sites depict error at different grid resolutions. (Left) COAMPS data at the finest resolution is available for all 6hly forecast periods up to 24 hours. (Right) NAVGEM data at six-hourly updates is only available for up to 12 hours.

Scatterplots of forecasted versus observed sea surface temperature at all observation sites are also created at different grid resolutions: g01, g02, g03, g04, and NAVGEM. For SST, most of the error is attributed to a limitation of oceanographic data in the assimilation cycle and the post-processed output. SST is a slowly varying parameter, and only the earlier forecast times are available. In Figure 8, the finer resolution grid (left) has a strong linear relationship with the observed SST condition. For the global grid (right), forecast values span a small range, indicating a weak relationship between the two variables.

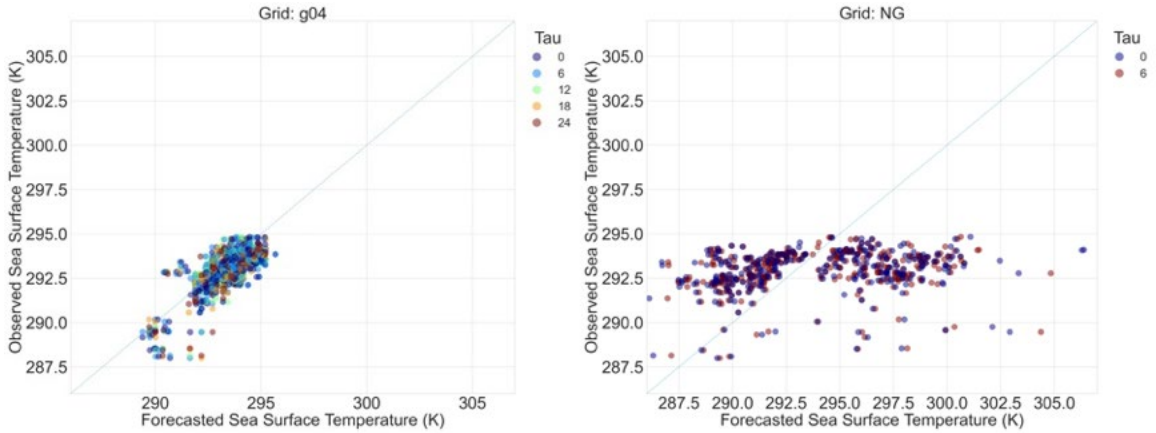


Figure 8. Scatterplots of forecasted versus observed sea surface temperatures at all observation sites depicts error at different grid resolutions. (Left) COAMPS data at the finest resolution is available for all 6hly forecast periods up to 24 hours. (Right) NAVGEM data at six hourly updates is only available for up to six hours.

Figure 7 and Figure 8 are summarized in Table 4, which also shows the 2 m air temperature and SST differences. The correlation provides a quantitative estimate of the linear relationship between the forecast variable and the observation. Values close to one indicate a strong positive relationship. In general, all COAMPS grids perform about the same. On average, NAVGEM has a higher bias and overforecasts the 2 m air temperature and SSTs. And lastly, most forecast grids are highly correlated to the observations. Other comments about Table 4 include:

- Negative ME suggests the model overforecasts the observation conditions.
- Positive ME suggests the model under forecasts the observation conditions.
- A smaller MSE is desired and indicates better grid performance.

Table 4. Summary statistics for 2 m Air Temperature, Sea Surface Temperature, and Air-Sea Temperature difference from various NWP grid resolutions to observations.

| **Forecast 2 m Temperature Model Error (K)** | | | | |
|---|---|---|---|---|
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | -1.321 | 6.072 | 2.464 | 0.529 |
| g02 | -0.11 | 3.7 | 1.924 | 0.583 |
| g03 | -0.381 | 5.006 | 2.237 | 0.533 |
| g04 | 0.048 | 4.656 | 2.158 | 0.57 |
| NG | -1.628 | 11.497 | 3.391 | 0.582 |
| **Forecast Sea Surface Temperature Model Error (K)** | | | | |
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | 0.028 | 1.141 | 1.068 | 0.422 |
| g02 | -0.498 | 1.062 | 1.03 | 0.65 |
| g03 | -0.584 | 0.917 | 0.958 | 0.766 |
| g04 | -0.511 | 0.866 | 0.931 | 0.772 |
| NG | -0.974 | 15.006 | 3.874 | 0.199 |
| **Forecast Air-Sea Temperature Difference Model Error (K)** | | | | |
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | -1.365 | 7.392 | 2.719 | 0.445 |
| g02 | 0.379 | 4.95 | 2.225 | 0.513 |
| g03 | 0.197 | 5.808 | 2.41 | 0.507 |
| g04 | 0.556 | 5.723 | 2.392 | 0.572 |
| NG | -0.653 | 12.464 | 3.53 | 0.411 |

## 2. Sea Level Pressure Error Analysis

Sea level pressure is a well-forecasted variable. There is a strong linear relationship between each forecast grid to the observations. Figure 9 displays the strong linear relationship through scatterplots of forecasted mean sea level pressure versus the observed mean sea level pressure at all observation sites. Both plots depict minimal bias; for all grid resolutions, there is a high correlation of about 90% and negative ME; i.e., on average the models overforecast the mean sea level pressure. However, under scrutiny, the COAMPS

model (left) at the finest resolution is slightly more dispersive than the NAVGEM model (right). A numerical summary of the sea level pressure variable is provided in Table 5.

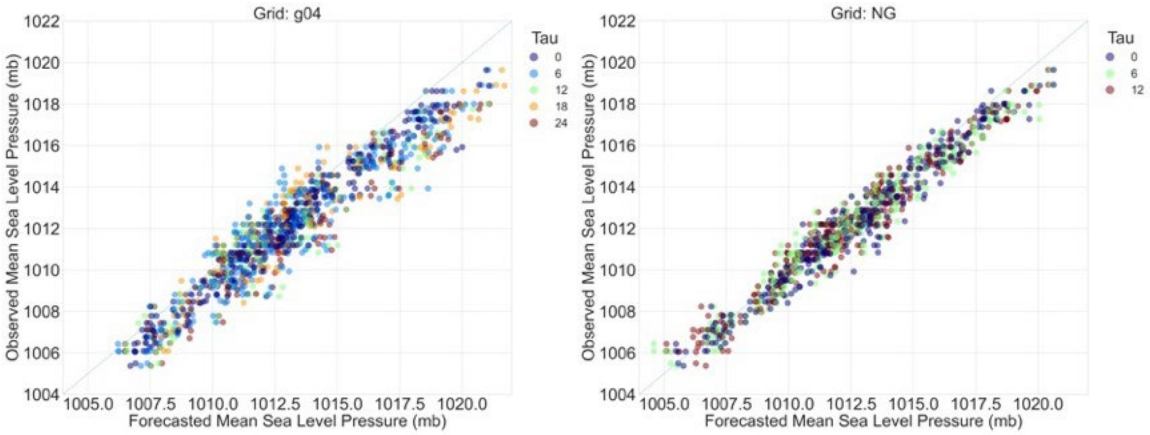


Figure 9. Scatterplots of forecasted mean sea level pressure versus the observed mean sea level pressure at all observation sites depict minimal error at different grid resolutions. (Left) COAMPS data at the finest resolution is available for all 6hly forecast periods up to 24 hours. (Right) NAVGEM data at 6 hourly updates is only available up to 12 hours.

Table 5. Summary statistics for Sea Level Pressure from various NWP grid resolutions to observations.

| **Forecast Sea Level Pressure Model Error (mb)** | | | | |
|---|---|---|---|---|
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | -0.442 | 1.705 | 1.306 | 0.94 |
| g02 | -0.751 | 1.569 | 1.253 | 0.954 |
| g03 | -0.818 | 1.639 | 1.28 | 0.954 |
| g04 | -0.98 | 1.969 | 1.403 | 0.952 |
| NG | -0.371 | 0.634 | 0.796 | 0.977 |

### 3. Relative Humidity Error Analysis

Similar to 2 m air temperature, relative humidity over the Southern California Bight is perturbed by wind flow conditions because the advection or the horizontal flow of air carries moisture. Wind flow conditions can be defined by local meteorological phenomena:

land and sea breeze diurnal circulations, Catalina eddies, migratory low-pressure systems, and SAWEs. For instance, dry conditions or low humidity are paired with SAWEs which bring dry air offshore; on the contrary, Catalina Eddies enhance the marine layer, and moisture is advected onshore.

Overall, the observed 2 m relative humidity is not well represented by the COAMPS and NAVGEM models. Figure 10 shows scatterplots of the forecasted 2 m relative humidity versus the observed 2 m relative humidity at all observation sites for the finest resolution COAMPS grid (left) and NAVGEM (right). Both plots display concentrated areas of 60% to 90% observed relative humidity, suggesting that the models have a wide forecast range for a small observational range. There are outliers present in all plots of various grid resolutions. The pair-wise correlations are moderate, indicating that the linear relationship between the forecasts and observations is weak. In fact, in Figure 10, the point clusters fans out as relative humidity decreases from 90%. In addition, Table 6 is an error summary of 2 m relative humidity from various NWP grid resolutions to observations.

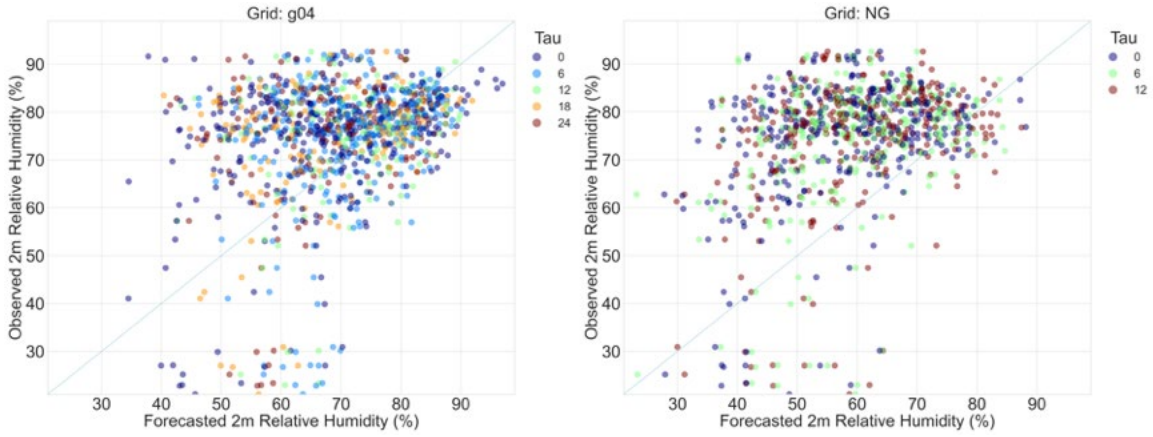


Figure 10. Scatterplots of forecasted 2 m relative humidity versus the observed 2 m relative humidity at all observation sites depicts significant error at different grid resolutions. (Left) COAMPS data at the finest resolution is available for all 6hly forecast periods up to 24 hours. (Right) NAVGEM data at 6 hourly updates is only available up to 12 hours.

Table 6. Summary statistics for 2 m Relative Humidity from various NWP grid resolutions to observations.

| Forecast 2 m Humidity Model Error (%) | | | | |
|---|---|---|---|---|
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | 10.809 | 328.001 | 18.111 | 0.316 |
| g02 | 4.767 | 221.319 | 14.877 | 0.293 |
| g03 | 7.072 | 270.704 | 16.453 | 0.242 |
| g04 | 5.444 | 235.418 | 15.343 | 0.286 |
| NG | 15.52 | 436.321 | 20.888 | 0.344 |

## 4. Wind Error Analysis

The wind error analysis is more thorough than other variables since the wind is analyzed through different configurations of the data frame. Wind speed is plotted to understand the magnitude difference of the observed winds to the forecasted winds. In Figure 11, scatterplots of the 10 m wind speeds demonstrate compounded bias as each model grid is initialized from the coarser resolution grid. Wind direction is plotted to understand the general flow conditions of each model grid. In Figure 12, the wind rose plots of the 10 m winds depict directional bias at different grid resolutions. Figure 13 is another scatterplot, similar to those shown before, of forecasted 10 m *u*- and *v*-component winds versus the observed winds at all observation sites and for each grid resolution. Also, time-series plots (Figure 14) of the observed to model grid differences are provided as an initial assessment of the wind's temporal variability. All grid-to-grid plots are summarized in Table 7.

From the grid-to-grid plots and table summary, differences between the finest resolution COAMPS grid and the coarsest global resolution grid, NAVGEM, indicate that flow conditions are more likely affected by broad synoptic conditions compared to local mesoscale effects. NAVGEM accurately models large-scale weather patterns, whereas the refined resolution grids of COAMPS simulate small-scale local mesoscale effects. For example, in Figure 11 the 10 m wind speed at the finest resolution model is more dispersive than the 10 m wind speed of the global grid. These differences imply that the COAMPS

grid occasionally over and inaccurately forecasts the observed wind speeds. Furthermore, another example is the directional bias at different grid resolutions represented in Figure 12. Observed wind conditions are predominately northwesterly associated with coastal flow, and forecasted conditions are biased to the northeast. Lastly, the observed-to-forecast grid differences in Figure 14 portray shifts from diurnal to synoptic patterns, such as migratory low-pressure systems and SAWEs.

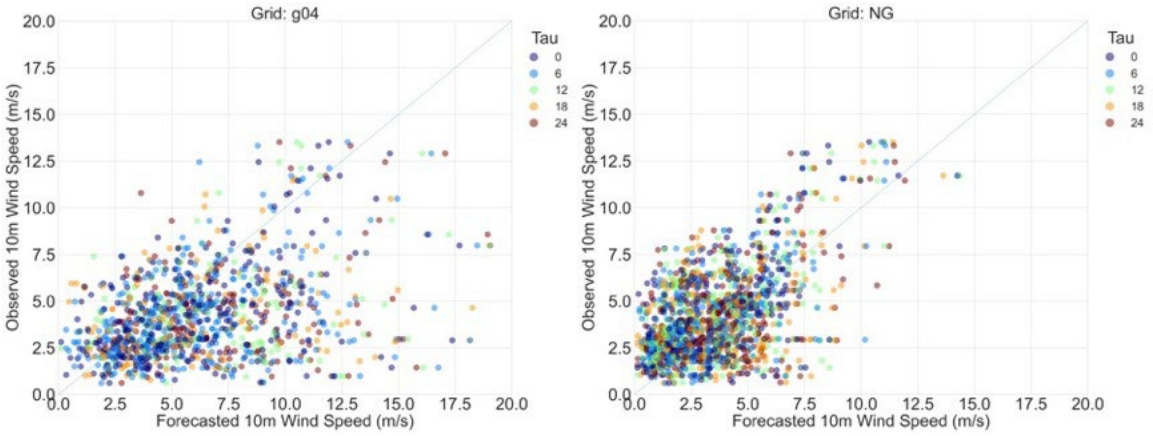


Figure 11. Scatterplots of forecasted 10 m wind speeds versus the observed 10 m wind speeds at all observation sites depict significant errors at different grid resolutions. (Left) COAMPS data at the finest resolution is available for all 6hly forecast periods up to 24 hours. (Right) NAVGEM data at 6 hourly updates are only available up to 12 hours.

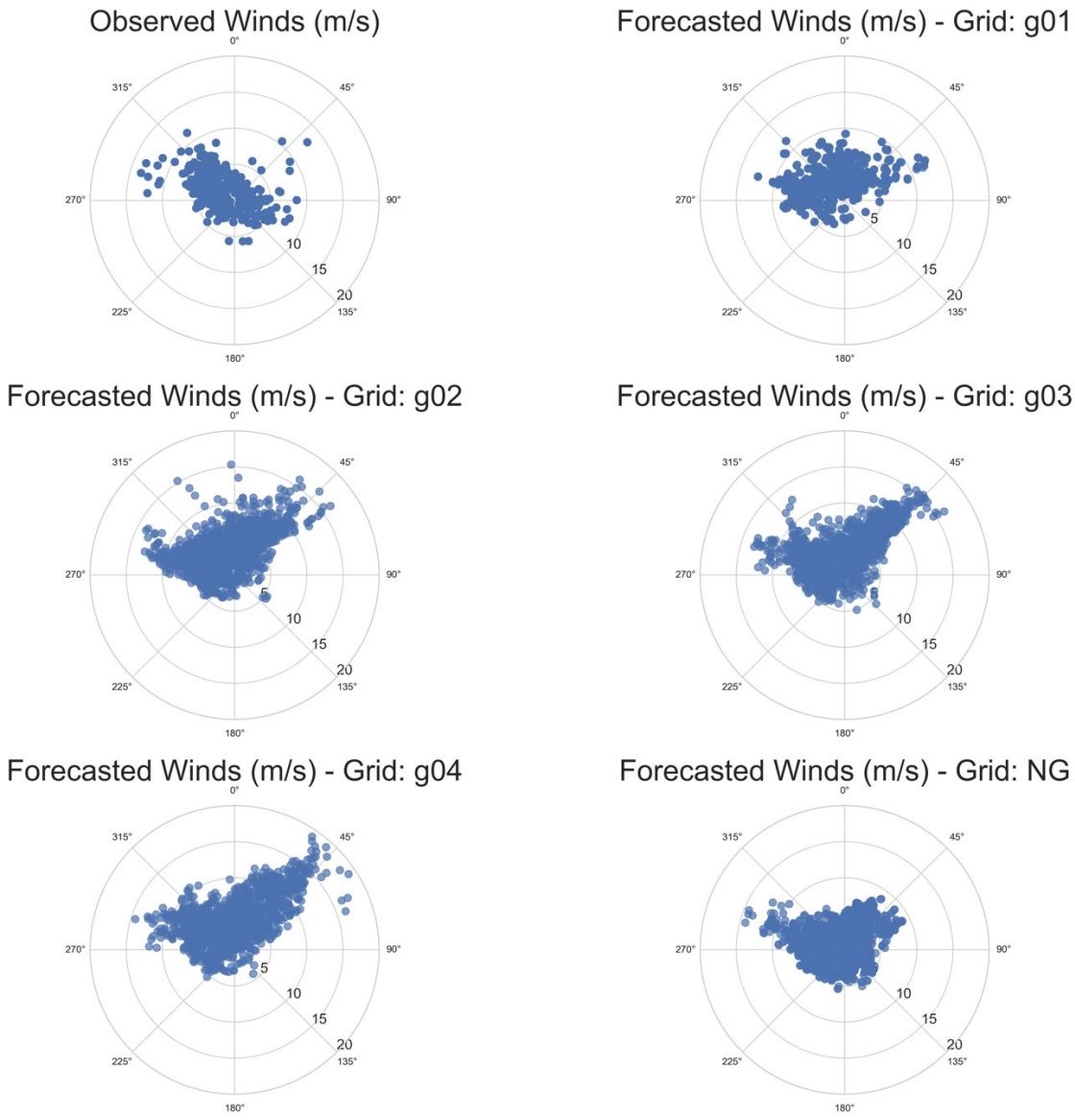


Figure 12. Scatterplots of forecasted 10 m winds versus the observed 10 m winds at all observation sites depicts directional bias at different grid resolutions.

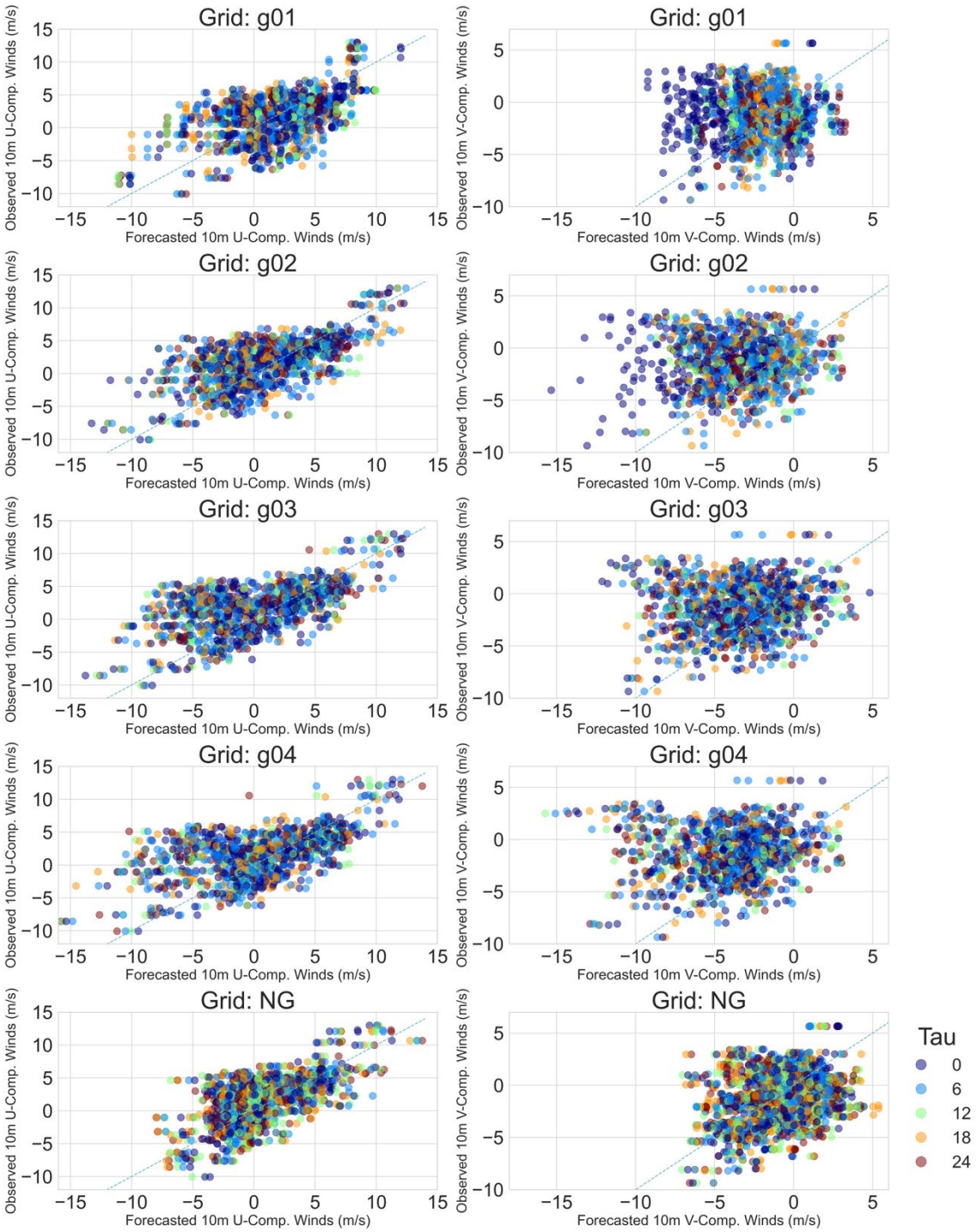


Figure 13. Scatterplots of forecasted 10 m U- and V-component winds versus the observed 10 m winds at all observation sites depicts large error at different grid resolutions.

Table 7. Summary statistics for 10 m Winds from various NWP grid resolutions to observations.

| **Forecast 10 m U-Component Wind Speed Model Error (m/s)** | | | | |
|---|---|---|---|---|
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | -0.183 | 13.179 | 3.630 | 0.536 |
| g02 | 0.804 | 12.828 | 3.582 | 0.641 |
| g03 | 1.771 | 18.172 | 4.263 | 0.592 |
| g04 | 1.328 | 18.962 | 4.355 | 0.559 |
| NG | 0.736 | 8.661 | 2.943 | 0.660 |
| **Forecast 10 m V-Component Wind Speed Model Error (m/s)** | | | | |
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | 0.994 | 12.004 | 3.465 | 0.040 |
| g02 | 2.129 | 17.996 | 4.242 | 0.073 |
| g03 | 2.173 | 18.626 | 4.316 | 0.099 |
| g04 | 2.812 | 23.067 | 4.803 | 0.088 |
| NG | -0.416 | 9.208 | 3.035 | 0.258 |
| **Forecast 10 m Wind Speed Model Error (m/s)** | | | | |
| Grid Resolution | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| g01 | -0.634 | 7.422 | 2.724 | 0.358 |
| g02 | -1.468 | 10.039 | 3.168 | 0.434 |
| g03 | -1.572 | 12.508 | 3.537 | 0.371 |
| g04 | -2.024 | 15.92 | 3.990 | 0.371 |
| NG | 0.593 | 5.096 | 2.258 | 0.561 |

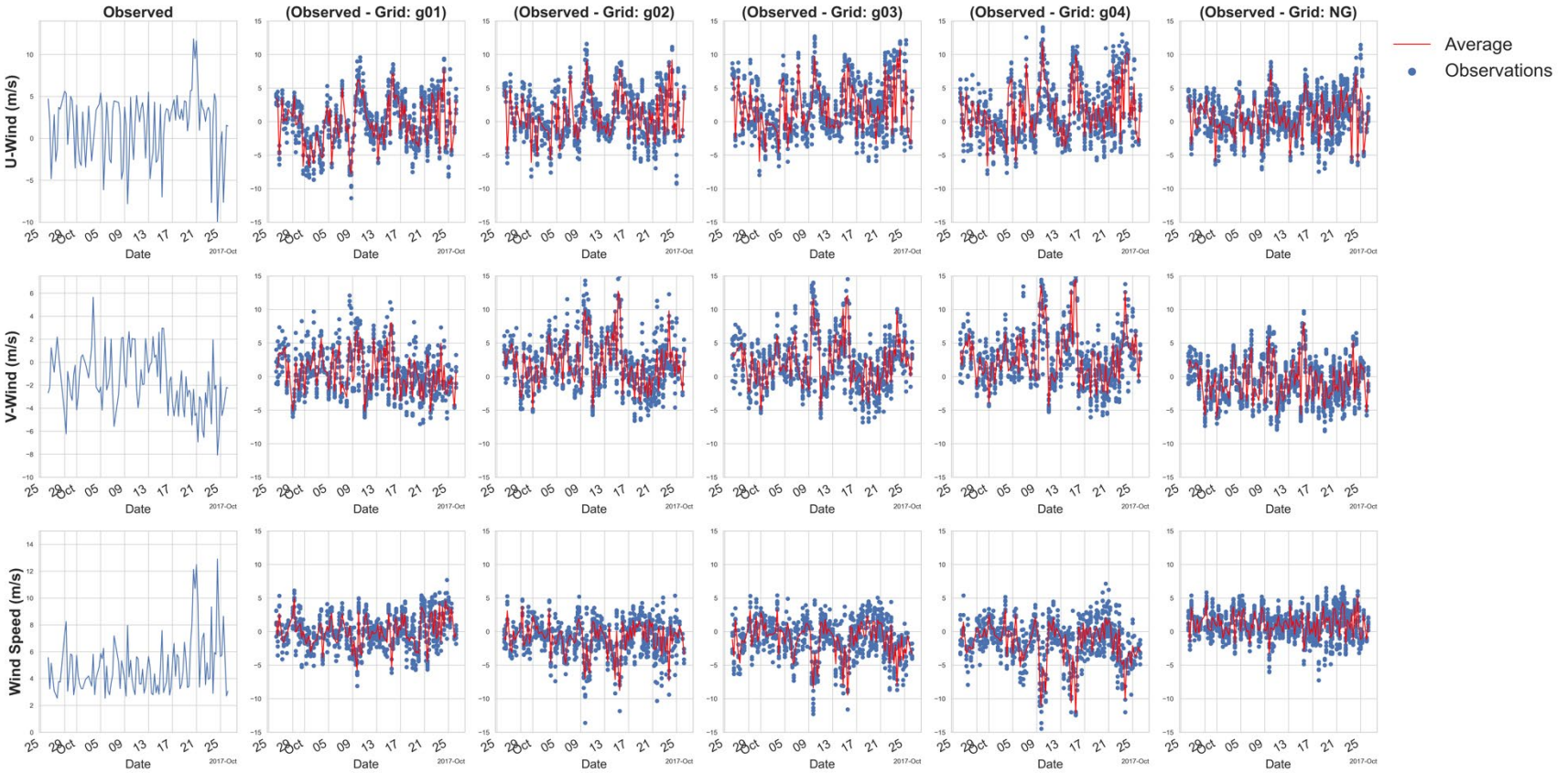


Figure 14. Time series of 10 m winds (*u*-Component winds, *v*-Component winds, and wind speed in m/s) observed to forecast differences. As indicated, plots show the observed minus forecast difference for 27 September to 27 October 2017; red line on plots represents the mean difference for a given day and time; blue dots are the differences between the observations to the forecasts; the x-axis is the date and is the same for all plots.

The wind data frame is analyzed through the TLMM prospective. As a result, plots of observed 10 m winds (*u*- and *v*- components) are compared to the models at various forecast times: 0-, 6-, 12-, 18-, and 24-hour forecasts. This comparison of the observations to the TLMM members does not differentiate the data by grid resolutions. Scatterplots of the TLMM members compared to the observations are exceptionally important because these plots show the relationship of each predictor; i.e., time-lagged forecasts, to the predictand; i.e., observations. Figure 15 and Table 8 provide graphical and numerical summaries for *u*- and *v*-component 10 m winds time-lagged members. Overall, there is a positive bias for *u*-component winds for all forecast periods; that is, the component magnitudes are under forecasted, and RMSE observed is approximately 3 m/s or 5.8 knots. For the *v*-component winds there is negative bias for all forecast periods and RMSE is around 3 m/s or 5.8 knots. Therefore, the winds would have about a 4.5 m/s or 8.7 knot speed error.

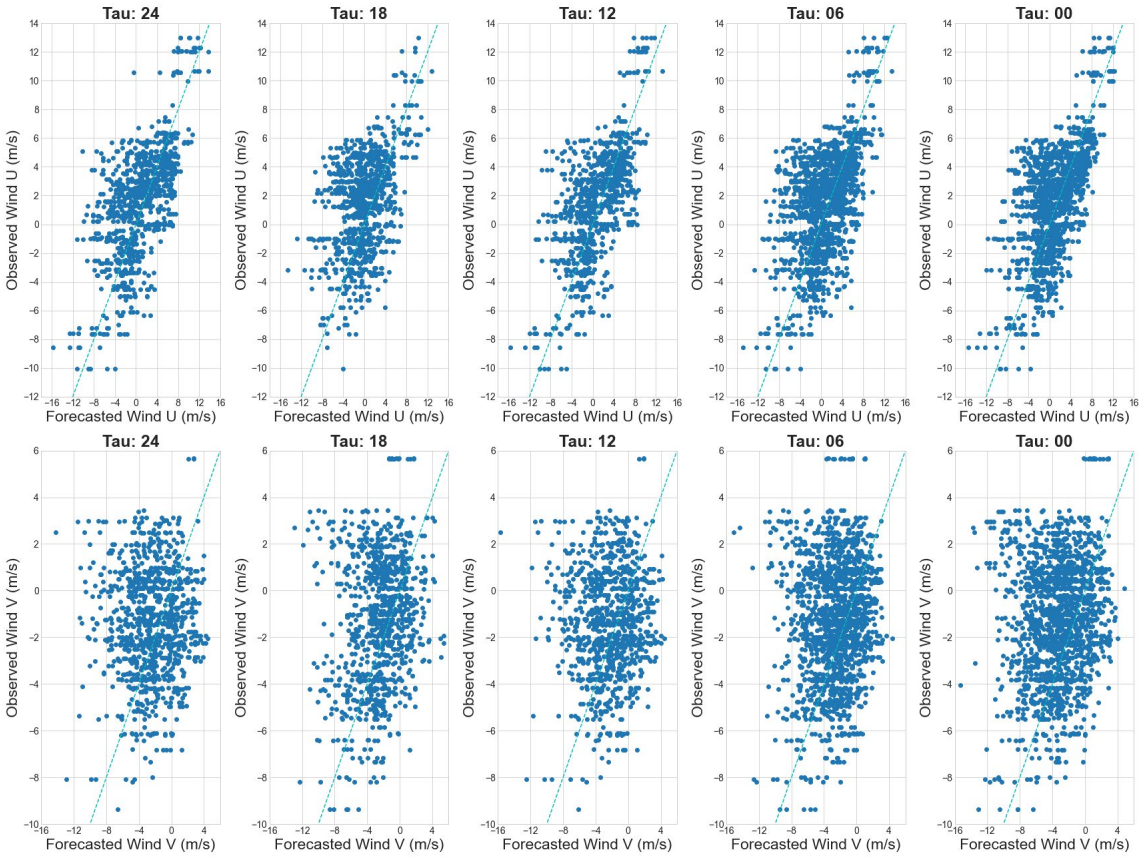


Figure 15. Scatterplots of observed 10 m Winds (*u*- and *v*- components) to various time-lag members: 0-, 6-, 12-, 18-, and 24-hour forecasts. This is a comparison of the observations to the time-lag members. This plot includes all observation before null values are removed.

Table 8. Summary statistics for 10 m winds for all time-lagged members compared to the observations. The table is broken into *u*-component and *v*-component wind estimates. Calculations are for the clean *u*- and *v*-component dataset (334 observations).

| **Forecast 10m U-Component Wind Speed Model Error (m/s)** | | | | |
|---|---|---|---|---|
| Forecast Hour | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| Tau 00 hr | 0.814 | 8.105 | 2.847 | 0.684 |
| Tau 06 hr | 0.584 | 7.889 | 2.809 | 0.679 |
| Tau 12 hr | 0.635 | 8.972 | 2.995 | 0.641 |
| Tau 18 hr | 0.771 | 8.853 | 2.975 | 0.652 |
| Tau 24 hr | 0.877 | 9.484 | 3.080 | 0.650 |
| **Forecast 10m V-Component Wind Speed Model Error (m/s)** | | | | |
| Forecast Hour | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | Correlation |
| Tau 00 hr | -0.563 | 7.852 | 2.802 | 0.358 |
| Tau 06 hr | -0.479 | 7.596 | 2.756 | 0.347 |
| Tau 12 hr | -0.402 | 9.308 | 3.051 | 0.257 |
| Tau 18 hr | -0.330 | 10.768 | 3.282 | 0.177 |
| Tau 24 hr | -0.304 | 10.517 | 3.243 | 0.178 |

## 5. Error Sources

Identification of all potential error sources is another TLMM characterization element. Error sources can be random or systematic, simple or compounding, or reducible or irreducible. Also, there are fundamental errors present in NWP models, such as observational error (instrument error, reporting and coding error, human error in recording, and representativeness), data assimilation error (selection of observations incorporated in the model updates), numerical model error (model approximations to principal atmospheric dynamic equations), physical error (a model approximation to the natural atmosphere and lateral boundaries), and forecast error (extrapolating on time-domain) (Du et al. 2018). In addition, there is an error in the structure of the analysis data set, which includes data coverage gaps, temporal and horizontal grid resolutions, initialization provided by the background field, and interpolation of the observations to the grid. Therefore, it is essential to acknowledge known sources of error and minimize these errors, if possible.

Fundamental errors are “largely unavoidable, and perhaps even unknown to us in the real world” (Du et al. 2018). Some errors in the dataset structure and analysis can be minimized by Bayesian regression to estimate weights for each TLMM parameter.

FNMOC produces monthly verification products for NAVGEM, COAMPS, and Global Forecast System (GFS). Figure 16 and Figure 17 are verification plots and show bias and RMSE at the 0000 UTC and 1200 UTC forecasts for the Southern California region in October 2017. These plots confirm the grids’ 24-hour data assimilation error and horizontal resolution error. There is a wave-like periodicity of increasing and decreasing error as the number of observations varies. Moreover, for the beginning forecast period, NAVGEM has a near-zero bias, whereas COAMPS has a consistent positive bias. NAVGEM and COAMPS model performance assessed by RMSE is variable: performance depends on the analysis time, forecast time, and the number of observations assimilated in the model. Subsequently, these sources of error observed are unavoidable when creating a TLMM from COAMPS and NAVGEM, but imperative to acknowledge.

The TLMM is an ensemble model and, when employed, will provide an ensemble forecast with pertinent statistics used for uncertainty calculations. “Ensemble forecasting is a dynamical and flow-dependent approach of quantifying this forecast uncertainty,” but ensemble forecasting reduces random error and not systematic error directly (Du et al. 2018). The TLMM approach is careful not to introduce systematic errors in the process. Varying statistical methods applied to the TLMM help to reduce the error observed in the TLMM data frame, NWP models, and observations. As a result, uncertainty calculations from the TLMM ensemble could be the basis for developing a reliable confidence metric to assess COAMPS NWP performance.

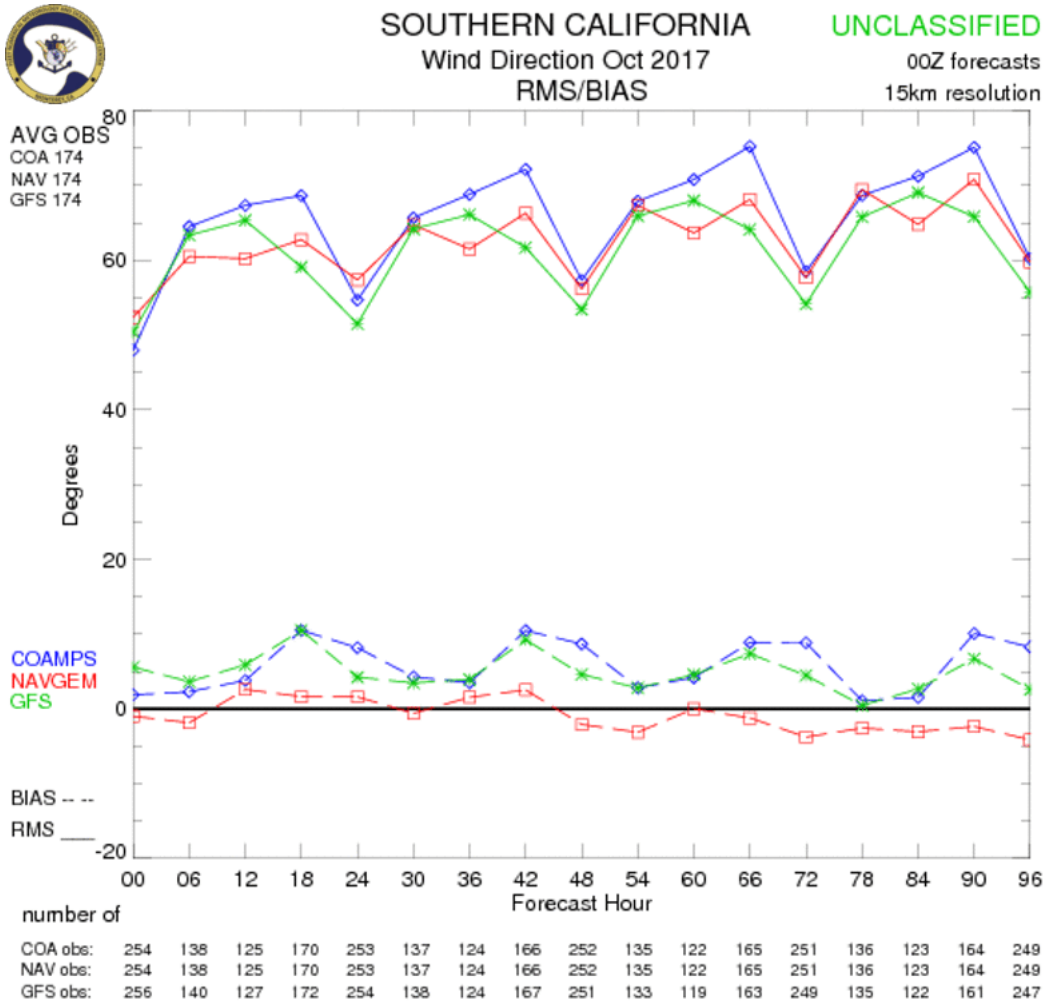


Figure 16. FNMOC monthly verification products for NAVGEM, COAMPS, and Global Forecast System (GFS). Plots show bias differences in the 00 UTC forecasts for the Southern California region in October 2017.

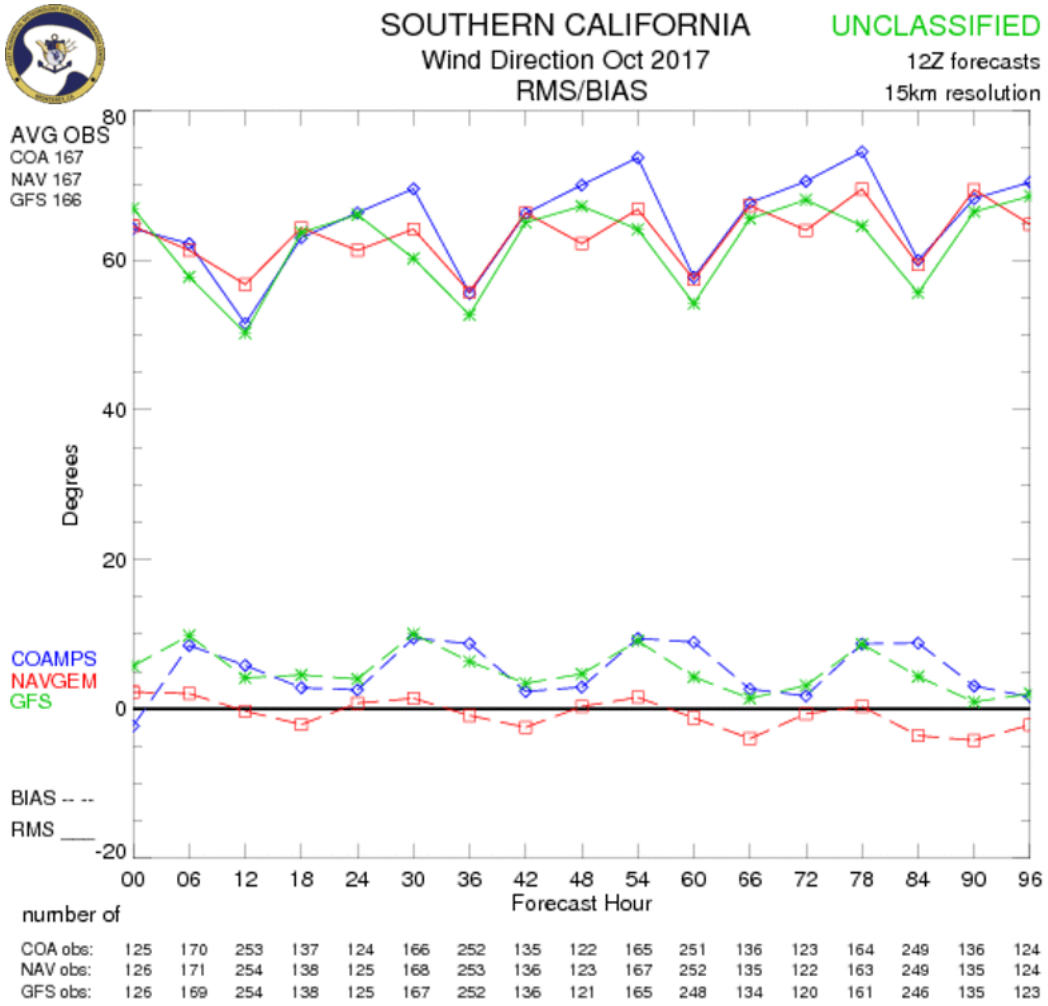


Figure 17. FNMOC monthly verification products for NAVGEM, COAMPS, and Global Forecast System (GFS). Plots show bias differences in the 12 UTC forecasts for the Southern California region in October 2017.

## D. TEMPORAL AND SPATIAL VARIABILITY

Temporal and spatial variability of the *u*- and *v*- component winds are analyzed to further identify possible sources of error in the data set. The spatial variability is determined by looking at the position of the observation sites relative to the coast, to each other, and within a grid. Temporal variability is assessed by a spectral analysis. A spectral analysis of the *u*- and *v*-component times series is applied to the data, transforming the time series from the time domain to the frequency domain. Information about performing a spectral

analysis can be found in the NPS OC3150 Time Series Analysis Course Notes (Thornton, 2007). Cowpertwait and Metcalfe (2009) also provide an excellent description of time series analysis and how to interpret data results.

The spectral analysis to determine underlying periodic signals within the time series data was performed in MATLAB. Figure 18 is a frequency plot of the time series data decomposed by observation location and wind component. From these spectral density plots for *u*- and *v*-component winds, a prominent peak at the 24-hour period corresponds to diurnal variations of the land/sea breeze. Land and sea breezes result from daily heating and cooling of the land and water at different rates. There are also peaks at lower frequency (longer time periods) observed in the *u*-component wind, which suggest quasi-periodic frontal passage or SAWEs.

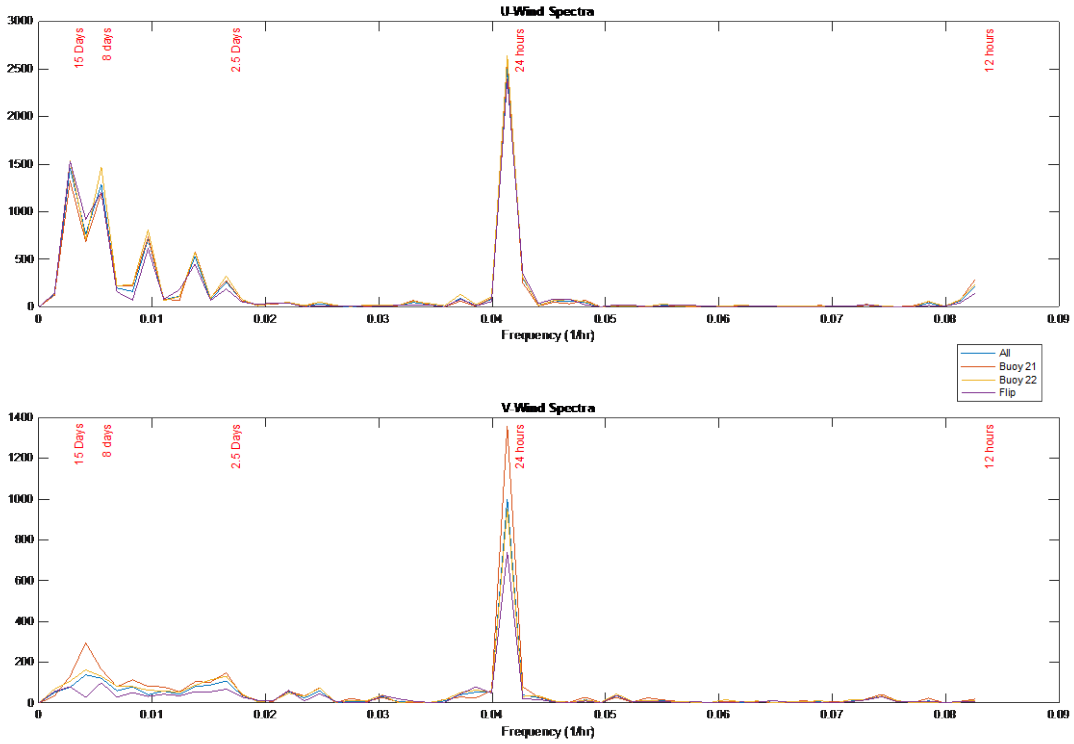


Figure 18. The spectral analysis identifies underlying periodic signals within the time series data. Spectral analysis is performed on the entire dataset and for each observation site.

A spatial summary of the observation sites and grid models is as follows:

- Buoy 21 is approximately 3.5 km from the shore.
- Buoy 22 is about 14.9 km from Buoy 21 and approximately 7.3 km from the shore.
- Flip is approximately 39.5km from the nearest shore point, 27.2 km from Buoy 22, and 41.3 km from Buoy 21.
- COAMPS grid resolutions are 54 km, 18 km, 6 km, and 2 km.
- NAVGEM grid resolution is 31 km.
- Within the coarser resolution grids, the grid has an excellent spatial representation.

After reviewing the data distribution across grids and platforms, it is concluded that the data exhibits good temporal and spatial variability; however, some bias is observed in the directionality of the winds for the observed diurnal frequency.

## E. DATA CORRECTIONS

The observations, the COAMPS dataset, and the NAVGEM dataset are thoroughly examined in the quality assurance process: identifying outliers, correcting data-entry inconsistencies, conversion of units, and removal of null and erroneous values. During the quality assurance process of the observations, corrections were made to the wind speed and direction. For example, wind directions from R/P FLIP were reported as -180° to 180°, which were translated to 0° to 360°. Similarly, observed wind speeds and directions are decomposed into their u- and v-components. Moreover, observation temperatures were reported in Celsius; and because temperature model data is available in Kelvin, a conversion from Celsius to Kelvin is performed on the observations by adding 273.15 K to each observation temperature value. Lastly, null values for Buoy 21 and Buoy 22 were reported as -999, and null values for R/P FLIP were reported as -99. All null values were converted to NaNs in the python pandas data frame.

THIS PAGE INTENTIONALLY LEFT BLANK

# IV. METHODOLOGY

## A. OVERVIEW

Continuing with the design flow from Figure 2, this chapter discusses model variations applied to the dataset. This thesis explores three approaches to fit the TLMM: a simple mean, a weighted mean, and a Bayesian model average. All models are supervised regression models applied to theobservations and interpolated forecasts. The observations are considered the predictands, whereas interpolated model data are identified as the predictors. The supervised TLMM produces a probabilistic prediction with two beneficial characteristics: mean and variance (Dowdy et al. 2004). Although prediction values are made, the focus of this paper is to leverage the statistical information calculated by the ensemble models. The term "prediction" will be used often and represents the TLMM estimate for a given forecast time.

Regression analysis is implemented to determine ensemble members' individual weights. Appropriate weights (e.g., Bayesian model averaging) are assigned to various ensemble members based on forecast time (referred to as tau) and horizontal grid resolution. More recent forecast cycles will have a higher overall weight. Ideally, higher resolution COAMPS grids have a higher overall weight than the global NAVGEM model; however, the COAMPS outer grid has a coarser resolution than NAVGEM. Therefore, in this analysis of the TLMM, grid resolutions are not assigned their own weight. In future analysis of the TLMM, model resolution should be considered. Additionally, model parameters are uniform or unweighted, weighted, or optimized by a Bayesian risk (Bayesian regression model). Model optimization is assessed using error metrics as defined in Chapter III.B.

## B. REGRESSION OF DATASET ON OBSERVATIONS

COAMPS and NAVGEM output is gridded, and grid point values are utilized to obtain an interpolated forecast value. A bilinear interpolation function is applied to the model grid to interpolate the data to the observation locations: Buoy 22, Buoy 21, and the

R/P FLIP. The *i*-th, *j*-th position in the latitude and longitude matrix that surrounds the observational point is found using four conditional statements:

$$\lambda_{mat,i\,j} \leq \lambda_{obs}$$

$$\phi_{mat,i\,j} \leq \phi_{obs}$$

$$\lambda_{mat,i+1\,j+1} \leq \lambda_{obs}$$

$$\phi_{mat,i+1\,j+1} \leq \phi_{obs}. \tag{7}$$

Then the limits of a box containing the observation becomes:

$$(i, j)$$

$$(i, j+1)$$

$$(i+1, j)$$

$$(i+1, j+1) \tag{8}$$

A COAMPS land surface file, a 2-D binary variable that indicates if the grid points are over land or water, categorizes the surrounding grid locations as either water or land, corresponding to a value of 1 and 0, respectively. The grid extents are defined as:

$$ll = \text{land}(i, j)$$

$$lr = \text{land}(i, j+1)$$

$$ul = \text{land}(i+1, j)$$

$$ur = \text{land}(i+1, j+1). \tag{9}$$

The following variables describe the differences in latitude and longitude between grid points and the observation location within the box of grid points. Subscripts *i* and *j* represent the index of the data frame. The grid difference are defined as,

$$\Delta\phi_i = \phi_{i\,j+1} - \phi_{ij}$$

$$\Delta\lambda_i = \lambda_{i\,j+1} - \lambda_{i\,j}$$

$$\Delta\phi_{obs,i} = \phi_{obs} - \phi_{ij}$$

$$\Delta\phi_{i+1} = \phi_{i+1\,j+1} - \phi_{i+1\,j}$$

$$\Delta\lambda_{i+1} = \lambda_{i+1j+1} - \lambda_{i+1\,j}$$

$$\Delta\phi_{obs,i+1} = \phi_{obs} - \phi_{i+1\,j}. \tag{10}$$

To begin the process, the data is interpolated over lines of constant longitude. The latitude along the bottom of the area surrounding the observation is given by:

$$Bottom\ Lat = \lambda_{ij} + \Delta\lambda_i \cdot \frac{\Delta\phi_{obs,i}}{\Delta\phi_i}, \tag{11}$$

and the top latitudinal value is

$$Top\ Lat = \lambda_{i+1\,j} + \Delta\lambda_{i+1} \cdot \frac{\Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}}. \tag{12}$$

These latitudinal values are variable along constant lines of longitude. Next, to determine the longitudinal value, interpolation over latitude is found by first defining the latitudinal difference of the northern latitude to the southern latitude. The result is the latitudinal difference at the top of the box minus the latitudinal difference at the bottom of the box scaled by the longitudinal difference due to the *i+1*-th point, plus the latitudinal difference at the bottom of the box scaled by the longitudinal difference due to the *i*-th point, plus the contribution due to the observation location. The latitudinal difference reduces to

$$\Delta Lat = Top\ Lat - Bottom\ Lat$$

$$= \lambda_{i+1j} - \lambda_{i\,j} - \left(\lambda_{i+1j+1} - \lambda_{i+1\,j}\right) \cdot \frac{\phi_{i+1j}}{\phi_{i+1\,j+1} - \phi_{i+1\,j}} + (\lambda_{ij+1} - \lambda_{ij}) \cdot \frac{\phi_{ij}}{\phi_{i\,j+1} - \phi_{i\,j}}$$

$$+ (\lambda_{i+1j+1} - \lambda_{i+1\,j}) \cdot \frac{\phi_{obs}}{\phi_{i+1\,j+1} - \phi_{i+1\,j}} - (\lambda_{i\,j+1} - \lambda_{i\,j}) \cdot \frac{\phi_{obs}}{\phi_{i\,j+1} - \phi_{i\,j}}. \quad (13)$$

Similarly, the observation latitudinal difference to the southern latitude limit is

$$\Delta Obs\ Lat = Observation\ Lat - Bottom\ Lat$$

$$= \lambda_{obs} - \lambda_{i\,j} - (\lambda_{i\,j+1} - \lambda_{i\,j}) \cdot \frac{\phi_{obs}}{\phi_{i\,j+1} - \phi_{i\,j}} + (\lambda_{ij+1} - \lambda_{ij}) \cdot \frac{\phi_{ij}}{\phi_{i\,j+1} - \phi_{i\,j}}. \quad (14)$$

The distances to determine the weights for interpolation are summarized by Figure 19, where *y* refers to the latitudinal grid and *x* refers to the longitudinal grid. The one-dimensional scaled distance from $(i, j)$ to $obs$ at the bottom longitudinal limit is

$$s_{xb1} = \frac{\Delta\phi_{obs,i}}{\Delta\phi_i} = \frac{\phi_{obs} - \phi_{ij}}{\phi_{i\,j+1} - \phi_{ij}}; \quad (15)$$

the one-dimensional scaled distance from $obs$ to $(i, j+1)$ at the bottom longitudinal limit is

$$s_{xb2} = \frac{\Delta\phi_i - \Delta\phi_{obs,i}}{\Delta\phi_i} = \frac{(\phi_{i\,j+1} - \phi_{ij}) - (\phi_{obs} - \phi_{ij})}{\phi_{i\,j+1} - \phi_{ij}} = \frac{\phi_{i\,j+1} - \phi_{obs}}{\phi_{i\,j+1} - \phi_{ij}}; \quad (16)$$

the one-dimensional scaled distance from $(i+1, j)$ to $obs$ at the top longitudinal limit is

$$s_{xt1} = \frac{\Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}} = \frac{\phi_{obs} - \phi_{i+1j}}{\phi_{i+1\,j+1} - \phi_{i+1\,j}}; \quad (17)$$

and the one-dimensional scaled distance from $(i+1, j+1)$ to $obs$ at the top longitudinal limit is

$$s_{xt2} = \frac{\Delta\phi_{i+1} - \Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}} = \frac{(\phi_{i+1\,j+1} - \phi_{i+1\,j}) - (\phi_{obs} - \phi_{i+1j})}{\phi_{i+1\,j+1} - \phi_{i+1\,j}} = \frac{\phi_{i+1\,j+1} - \phi_{obs}}{\phi_{i+1\,j+1} - \phi_{i+1\,j}} \quad (18)$$

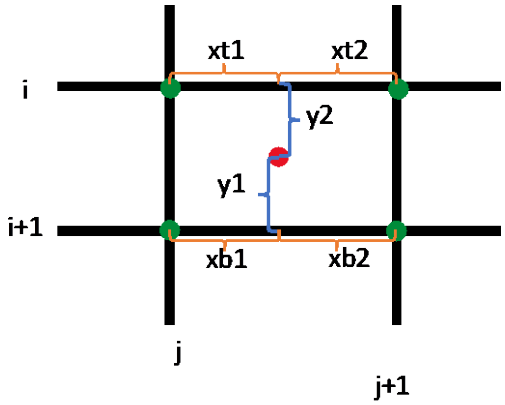


Figure 19. Latitude and longitude matrix surrounding the observation.

Similarly, for the latitudinal grid, the one-dimensional scaled distance from the observation to the southern (bottom) latitude is

$$s_{y1} = \frac{\Delta Obs\ Lat}{\Delta Lat} = \frac{\lambda_{obs} - \lambda_{ij} - \Delta\lambda_i \cdot \frac{\Delta\phi_{obs,i}}{\Delta\phi_i}}{\left(\lambda_{i+1\,j} + \Delta\lambda_{i+1} \cdot \frac{\Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}}\right) - \left(\lambda_{ij} + \Delta\lambda_i \cdot \frac{\Delta\phi_{obs,i}}{\Delta\phi_i}\right)} \tag{19}$$

and the one-dimensional scaled distance from northern (top) latitude to the observation is

$$s_{y2} = \frac{\Delta Lat - \Delta Obs\ Lat}{\Delta Lat} = \frac{\lambda_{i+1j} - \lambda_{obs} + \Delta\lambda_{i+1} \cdot \frac{\Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}}}{\left(\lambda_{i+1\,j} + \Delta\lambda_{i+1} \cdot \frac{\Delta\phi_{obs,i+1}}{\Delta\phi_{i+1}}\right) - \left(\lambda_{ij} + \Delta\lambda_i \cdot \frac{\Delta\phi_{obs,i}}{\Delta\phi_i}\right)} \tag{20}$$

The two-dimensional scaled distances are normalized to account for the sea-land identifier. Calculations ignore the values when the model grid point is flagged as a land point.

The weights for the distance from the grid points to the observations are

$$w_{i+1j} = \sqrt{s_{xb1}^2 + s_{y1}^2},$$

$$w_{i+1j+1} = \sqrt{s_{xb2}^2 + s_{y1}^2},$$

$$w_{ij} = \sqrt{s_{xt1}^2 + s_{y2}^2},$$

$$w_{ij+1} = \sqrt{s_{xt2}^2 + s_{y2}^2}. \tag{21}$$

Then, the sum for the normalized weights is defined by

$$w_{sum} = \frac{(1-ul)}{w_{i+1\,j}^2} + \frac{(1-ur)}{w_{i+1\,j+1}^2} + \frac{(1-ll)}{w_{ij}^2} + \frac{(1-lr)}{w_{ij+1}^2}. \tag{22}$$

The bilinear interpolated value for each forecast variable is computed as

$$x_{obs} = \frac{\frac{(1-ul)}{w_{i+1\,j}^2} \cdot x_{i+1\,j} + \frac{(1-ur)}{w_{i+1\,j+1}^2} \cdot x_{i+1\,j+1} + \frac{(1-ll)}{w_{ij}^2} \cdot x_{ij} + \frac{(1-lr)}{w_{ij+1}^2} \cdot x_{ij+1}}{w_{sum}}. \tag{23}$$

It follows that a direct comparison between the observation and $x_{obs}$ can be made in the TLMM analysis.

## C. SIMPLE AND WEIGHTED MEAN MODELS

The simple mean is the most elementary construction of the TLMM ensemble. A simple mean does not give any preference to recent forecast cycles or grid resolutions; a simple mean equally weights each TLMM ensemble. Another way to describe a simple mean is an unweighted poor-man's ensemble: an assembly of many deterministic forecasts from various sources (Du et al. 2018). Derivations for the TLMM are specific to the *u*- and *v*-components. A data matrix of *p* predictors for the *u*-component wind, *q* predictors for the *v*-component wind, and *n* forecasts interpolated to the observations is denoted as $\boldsymbol{X}_{p+q}$ and is defined by

$$\boldsymbol{X}_{p+q} = \left(\begin{array}{ccc|ccc} x_{11} & \cdots & x_{1p} & x_{1p+1} & \cdots & x_{1p+q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} & x_{np+1} & \cdots & x_{np+q} \end{array}\right), \tag{24}$$

where p = 5 and q = 5 for forecast members of 0, 6, 12, 18, and 24 tau. For $x_{ij}$ in the data matrix $X_{p+q}$, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p+q$, $x_{ij}$ is the same as $x_{obs}$ as defined in Section IV.B.

The predictand or observation matrix is defined as

$$\boldsymbol{y} = \begin{pmatrix} y_{11} & y_{12} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{pmatrix}, \tag{25}$$

where $y_{i1}$ represents the *u*-component winds and $y_{i2}$ represents the *v*-component winds. It follows that the simple mean for the *i*-th observation is calculated component-wise by

$$\hat{y}_{i1} = \frac{1}{p}\sum_{j=1}^{p} x_{ij} \tag{26}$$

and

$$\hat{y}_{i2} = \frac{1}{q}\sum_{j=p+1}^{q} x_{ij}, \tag{27}$$

such that

$$\hat{\boldsymbol{y}} = \begin{pmatrix} \hat{y}_{11} & \hat{y}_{12} \\ \vdots & \vdots \\ \hat{y}_{n1} & \hat{y}_{n2} \end{pmatrix} \tag{28}$$

is the forecast matrix for *n* observations (Lu et al. 2006).

The weighted mean is chosen to make the analysis time of Tau 0 the most significant, and a forecast time of Tau 24 the least important. Weights are ordinal and normalized, such that

$$\boldsymbol{w} = (w_1 \;\; w_2 \;\; w_3 \;\; w_4 \;\; w_5) = \left(\frac{1}{3} \;\; \frac{4}{15} \;\; \frac{1}{5} \;\; \frac{2}{15} \;\; \frac{1}{15}\right) \tag{29}$$

represents the weight vector. Then, the weighted mean estimate for the *i*-th observation is

$$\hat{y}_{i1} = \sum_{j=1}^{p} w_j x_{ij} \tag{30}$$

and

$$\hat{y}_{i2} = \sum_{j=p+1}^{q} w_{j-p} x_{ij}. \tag{31}$$

Weights shown in Equation (29) assume that the analysis time is the best performer and skill decreases as forecast time increases. Weights could be optimized based on errors

attributable to a given forecast. In Chapter III.C.4 of the wind error analysis, it is shown some forecast hours have minimal ME. For example, in Table 8 of the *u*- component summary statistics for all time-lagged members, the 24-hour, 12-hour, and 0-hour forecasts have the lowest ME of 0.769 m/s, 0.696 m/s, and 0.656 m/s, respectively. Therefore, appropriate weights for the TLMM ensemble could be selected by error attributes. As a result, the 6-hour and 12-hour forecast would have less importance.

## D. BAYESIAN REGRESSION MODEL

The Bayesian Regression Model is a supervised regression model that trains a probability model with *a priori* assumptions on the parameters before seeing the data, and then updates the parameterization after observing the data (Hastie et al. 2009, Gelman et al. 2021). It has been said by Wendt (2017), Hastie et al. (2009), and Gelman et al. (2021) that the Bayesian approach to data analysis is different than the traditional frequentist approach, where model assumptions on the data and model errors are required. The difference is the "use of a prior distribution to express the uncertainty present before seeing the data, and to allow the uncertainty remaining after seeing the data to be expressed in the form of a posterior distribution" (Hastie et al. 2009). The uncertainty in the posterior distribution is of interest to the TLMM prediction.

The TLMM approach leverages the statistical information from the posterior distribution to predict a value or calculate a Bayesian model average for each tau of the forecast. Python code to determine the Bayesian model average is obtained from Wendt (2017), who explains his method of Bayesian forecasts the following way:

> NPS BEMOS [Bayesian ensemble model output statistics] forecasts were produced …by drawing 10,000 random posterior parameter samples from the thinning region of their respective Markov chains—that is, the portion of the Markov chain beyond burn-in [of 1 million samples]. In this way, NPS BEMOS PPDs [posterior predictive distribution] …are explicitly probabilistic and compose joint probability distributions in the 3-dimensional state space of the predictands. Nevertheless, the NPS BEMOS estimates…were derived from descriptive statistics extracted from the full Bayesian PPDs.

Without providing considerable detail of the theory of the full Bayesian PPD and BEMOS estimates, the Bayesian model is trained on the observations with the predictors identified as the forecast time-lagged members. The posterior parameter estimates, denoted as B, are used to create the Bayesian TLMM predictions. Wendt's (2017) dissertation can be referenced for more detail on model structure. The TLMM parameter estimate matrix of regression coefficients is defined as

$$\mathrm{B} = \begin{array}{cc} u\text{-obs} & v\text{-obs} \\ \left(\begin{array}{cc} \beta_0 & \beta_1 \\ \beta_2 & \beta_3 \\ \beta_4 & \beta_5 \\ \beta_6 & \beta_7 \\ \beta_8 & \beta_9 \\ \beta_{10} & \beta_{11} \\ \beta_{12} & \beta_{13} \\ \beta_{14} & \beta_{15} \\ \beta_{16} & \beta_{17} \\ \beta_{18} & \beta_{19} \\ \beta_{20} & \beta_{21} \end{array}\right) & \begin{array}{l} \text{Intercept} \\ u\text{-00} \\ v\text{-00} \\ u\text{-06} \\ v\text{-06} \\ u\text{-12} \\ v\text{-12} \\ u\text{-18} \\ v\text{-18} \\ u\text{-24} \\ v\text{-24} \end{array} \end{array} \tag{32}$$

where $\beta_0$ represents the sensitivity of the intercept on the observed *u*-component winds, $\beta_1$ represents the sensitivity of the intercept on the observed *v*-component winds, $\beta_2$ is the sensitivity of the *u*-component winds at Tau 0 on the observed *u*-component winds, $\beta_3$ is the sensitivity of the *u*-component winds at Tau 0 on the observed *v*-component winds, etc. Understanding the meaning of each posterior parameter estimate helps determine which interactions are significant.

Note in subsequent sections, the "Bayes Regression Model," "Bayes Model Averaging," and "Bayes Model" will all be used interchangeably and indicate that the Bayesian Regression model presented by Wendt (2017) is applied to the TLMM to obtain prediction or a Bayesian Model Average.

## E. ASSUMPTIONS

Sections C and D of this chapter discuss three approaches to fit the TLMM, all of which require assumptions on the data. These models are a simple mean, a weighted mean, and a Bayesian model average. Most statistical models require assumptions, such as

normalized errors, about the dataset, $\boldsymbol{X}_{p+q}$, prior to use. In the TLMM analysis, there is no specific assumption made on $\boldsymbol{X}_{p+q}$ except that the mean and variance vectors exist, and for the Bayesian model, a training dataset exists.

Bayesian modeling requires *a priori* information about the data distribution before fitting the model (Wendt 2017). While knowing prior Bayesian information would prevent model misspecification, this information may not be available to the analyst or forecaster. Undoubtedly, information about $X_{p+q}$ permits the correct model choice with satisfied assumption; thus, accurate inference is made. Faraway (2005) summarizes model assumptions and model misspecification succinctly:

> Most statistical theory rests on the assumption that the model is correct. In practice, the best one can hope for is that the model is a fair representation of reality. A model can be no more than a good portrait. As George Box said,
>
> All models are wrong, but some are useful.

This thesis shows that the simple mean, weighted mean, and Bayesian model average are all useful, but Bayesian model averaging of the TLMM performs most effectively to capture uncertainty of the deterministic forecasts.

Wendt (2017) also discusses employing the Bayesian model without *a priori* specification. The author states, "This is often implemented with early versions of the Bayesian data analysis process, when confidence in a priori beliefs is either low or incomplete, through so-called noninformative or weakly informative priors." For this thesis, *a priori* information may not be available to train the Bayesian model, such as when a forecast is for a remote area with limited observations, or when there is an urgent operational requirement to initialize the limited area model. Nevertheless, the Bayesian approach to the TLMM will be trained and fit to the data by "letting the data speak for itself" (Wendt 2017). There is an advantage to taking this approach where *a priori* beliefs are not predefined and are incorrectly specified; i.e., "permits an equivalent [Maximum Likelihood Estimate] MLE analysis while simultaneously delivering the associated Bayesian uncertainty estimates—with no separate or additional calculations required"

(Wendt 2017). The TLMM approach leverages Bayesian Model averaging to optimize the ensemble member weighting scheme.

Multiple linear regression and a generalized linear model (GLM) are intermediate between a mean model and Bayesian model averaging. The multiple linear regression model is not applied to the dataset because the linear model assumes a linear relationship of the predictors; i.e., the ensemble members, to the predictands; i.e., the observations, and errors are minimized through a squared error loss function. This relationship is not directly seen in the TLMM data, and there are compounding factors that cause variability in the observation to forecast error. Limitations of a linear regression model are as follows:

- The linear model assumes errors are independent and identically distributed (*iid*) and Gaussian with a mean of zero and variance of $y_i$ independent of $x_i$ (Faraway 2005).
- A multiple linear model does not fit a joint distribution to $y_1$ and $y_2$. The preference is to not fit the data to a multivariate multiple regression model because wind speed and direction are dependent on both the *u*- and *v*-component winds. However, a similar assumption applies to the errors.
- All TLMM ensemble members arise from a common statistical distribution.

The GLM is not applied to the dataset because the GLM also requires specification on the errors: a known "correlation and relative variance between the errors" (Faraway 2005). In this case, *iid* error and linearity of $\boldsymbol{X}_{p+q}$ on the expected value of $\boldsymbol{y}$ are not required; however, the variance matrix structure is needed or must be estimated by the GLM (Faraway 2006). Moreover, the GLM optimizes the maximum likelihood estimate (MLE) of the model parameters, providing best probabilities to the observed data. By iteration of the GLM and MLE optimization, the variance matrix can be estimated.

A similarity between the GLM and the Bayesian model is malleability to data with curvature; a difference is the requirement of MLE parameter estimation by GLM and

specification of a prior distribution for the parameters before analyzing the data by Bayesian methods (Hastie et al. 2009). Moreover, differences between the two approaches translate to inferences about parameter estimates. As stated by Hastie et al. (2009), "the Bayesian approach differs from the standard ('frequentist') method for inference in its use of a prior distribution to express the uncertainty present before seeing the data, and to allow the uncertainty remaining after seeing the data to be expressed in the form of a posterior distribution." Notwithstanding, this thesis predominantly explores the Bayesian model approach presented by Wendt (2017), who employs Monte Carlo Markov Chain sampling methods in a Bayesian model framework with "*a priori* model assumptions conditioned on available training data."

Although the analysis will not make specific assumptions about the prior probability distribution, further exploration of $\boldsymbol{X}_{p+q}$, such as how the predictors correlate and the covariance matrix structure, provides understanding of model parameter estimates. The covariance measures the amount of dependence between two random variables, while the correlation measures similarity and is calculated from the covariance (Härdle and Simar 2003). The equation for the correlation is provided in Equation (6), and covariance can be deduced by Equation (5). Figure 20 is a correlation heatmap of *u*- and *v*-component winds for each time-lagged ensemble member. Correlation within $\boldsymbol{X}_{p+q}$ assesses collinearity and independence of individual ensemble members.

To test which correlations are significant, a standard *F*-test is applied. The *F*-test determines if the two random variables are independent estimates of the same variance or if the linear relationship is significant (Dowdy et al. 2004). The null hypothesis of the *F* test is $\rho_{xy} = 0$; i.e., there is no correlation. The test statistic is defined by

$$\mathrm{F} = \frac{r^2}{(1-r^2)/(n-2)} \tag{33}$$

where $r = \hat{\rho}_{xy}$ represents the correlation estimate for *x* and *y*. An equivalent form to the F-test is the t-test with n-2 degrees of freedom (Dowdy et al. 2004). The t-test statistic is defined by

$$T = \sqrt{\mathrm{F}} = \frac{r}{\sqrt{(1-r^2)/(n-2)}}, \tag{34}$$

and the critical value for a 95% confidence level is $\mathrm{t}_{1-\alpha/2,n-2} = 1.969$ for $n$=334. The null hypothesis is rejected if $|T| \geq \mathrm{t}_{1-\alpha/2,n-2}$. In reference to Figure 20, all values less than 0.107 are not statistically significant at the $\alpha = 0.05$ significance level. Note that a rejection of the null hypothesis does not mean complete independence. Prior assumptions acknowledge that all variables are derived from the same underlying distribution and atmospheric model.

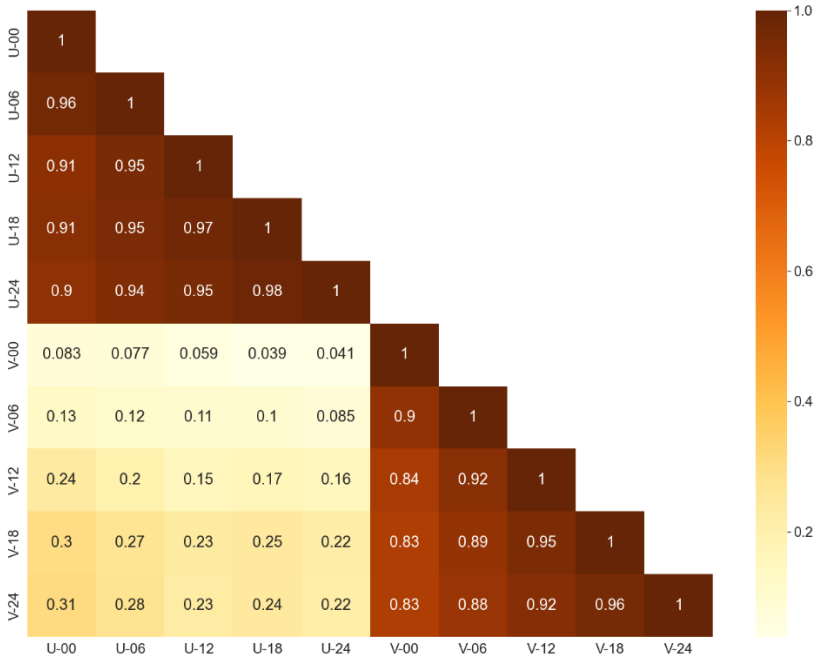


Figure 20. Correlation heatmap of $u$- and $v$-component winds for each time-lagged ensemble member.

## F. MEASUREMENT OF PERFORMANCE

Error estimates were examined to review model performance. Error for the three TLMMs and individual ensemble members are calculated from the mean bias and squared error estimate. These performance measurements are provided in Chapter III.B. First, in the review, there is interest in the variability in $\hat{\boldsymbol{y}}$ and the residual error. For the mean and

weighted mean models, error arises from the dispersion of the TLMM ensemble members and the collective accuracy to the observations. Error in the Bayesian model averaging is associated with the influence of the prior distribution for the parameters and posterior distribution of the estimate. Second, there is interest in the bias of individual ensemble members. Large bias in the member parameter estimates will help to determine which members are most or least valuable.

## G. VARIABLE SELECTION

The TLMM is linear in all predictors, and it does not include any interaction or higher-order terms. Therefore, the ensemble size is defined by the number of forecast taus for each atmospheric predictor. For example, the Bayesian TLMM has 22 parameters corresponding to the intercept and all time-lagged ensemble members for the *u*- and *v*-component winds. Variable selection of the ensemble size is a balance between data representation and temporal resolution. Significance intervals or credible intervals of parameter estimates, e.g., Bayesian model parameters, provide a method for variable selection.

The python code for this thesis is adopted from Wendt (2017) and is fundamentally unmodified. Therefore, the Bayesian model considers all variables, and none are removed. Although this thesis does not illustrate variable selection methods, it does raise questions as to how the TLMM can be optimized. Some questions about the variable selection of the TLMM ensemble are as follows:

- How many members are needed for the ensemble and is there a minimum number of members?
- Is there a maximum number of members based on contribution significance?
- How does forecast lead time affect the required ensemble size?
- How does vertical resolution affect model performance and ensemble size requirements?

- Do ensemble size requirements vary over TLMM averaging methods?

## H. DISTRIBUTION OF TRAINING, TESTING, AND VALIDATION DATASETS

Given the characteristics of the dataset explained in Chapter III, careful consideration is taken to distribute the TLMM dataset into training and testing parts. The TLMM dataset is broken into two parts: a random selection of 75% for training and 25% for testing. A random sample of the data was made because the length of the complete dataset spans only 31 days and higher frequencies are seen in the data set that random block sampling would not be able to resolve. Schultz et al. (2021) argues that random block sampling strategy is a preferred method to test machine learning methods because it does not "over-estimate the true generalization capability of the [model and] …because the test set contains information already used for training." Figure 21 shows different options to divide meteorological data with periodic features into training and testing datasets. A requirement for random block sampling is choosing a sufficient block length to observe the longest periodicity. However, in the TLMM case and with the data available from the CASPER-West experiment, the testing dataset would be insufficient, and there would be limits to repeat the random block sampling strategy over the 31-day time series. Besides, as stated by Hastie et al. (2009), it is "too difficult to give a general rule on how much training data is enough; among other things, this depends on the signal-to-noise ratio of the underlying function, and the complexity of the models being fit to the data." Therefore, the dataset selection method employed here provides the best possible assessment of the TLMM ensemble performance.

For validation of the model, a 10-fold cross validation was applied twice, yielding 20 iterations. A 10-fold cross validation method is advantageous when there is not enough data to split the dataset into three parts (training, testing, validation), and the quality of the model prediction is degraded (Faraway 2005). The data are evaluated for each iteration by the ME, MSE, and RMSE statistics. An average ME, MSE, and RMSE statistic is used to assess the overall model performance and selection.

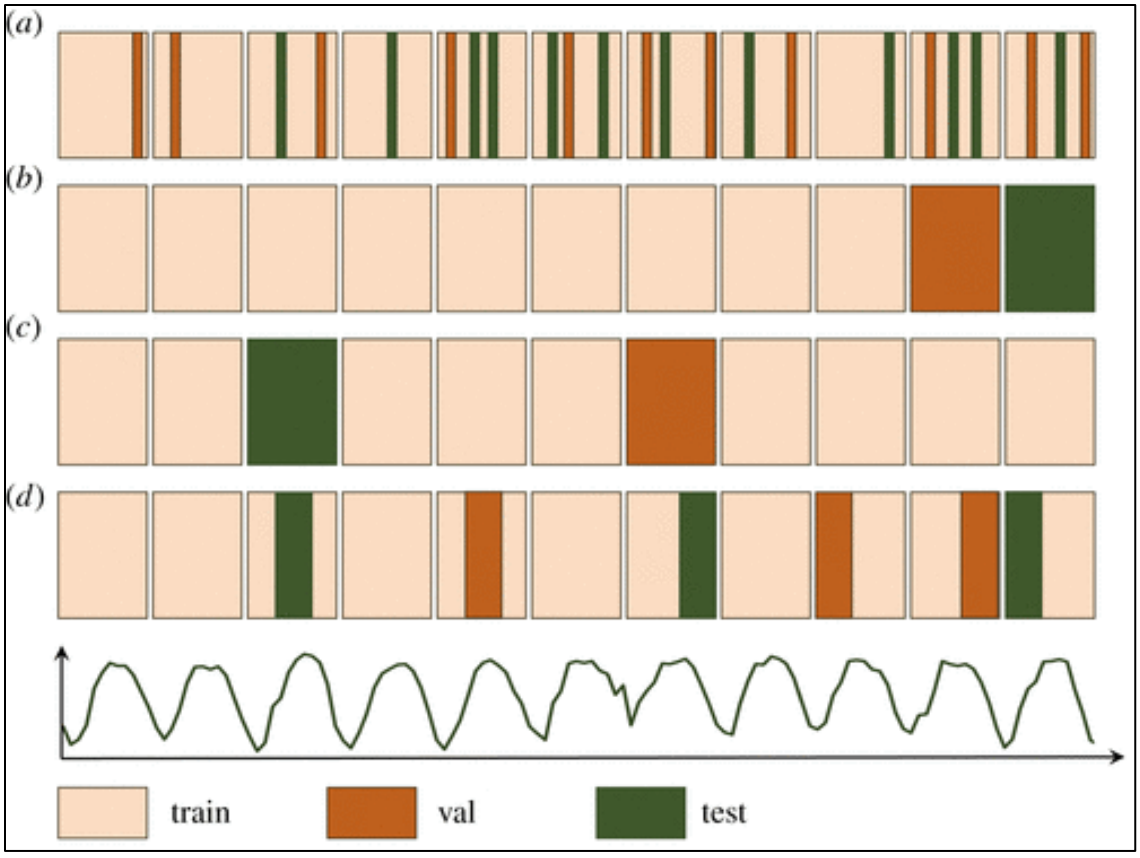


Figure 21. "Different train-dev-test splitting strategies for meteorological data with periodic features as indicated in the conceptual time series at the bottom of the figure. Every sand coloured block stands for 1 year of data. Case (a) depicts random sampling as is commonly applied in many DL applications. Cases (b–d) show different variants of random block sampling, which avoid spurious correlations between the train, val and test sets, if the block length chosen is long enough" (Schultz et al. 2021).

# V. ANALYSIS

## A. RESULTS

This chapter examines overall model performance and is the culmination of the TLMM proof-of-concept. For each model variation, model performance and characterization are discussed in detail. Models are evaluated on ME, MSE, and RMSE statistics obtained from the testing dataset against the observations. The testing dataset is a random selection of the larger modified dataset, as explained in Chapters III and IV.H. It is shown in following sections of this chapter that construction of the TLMM is useful to smooth forecast prediction for large member dispersion, can provide a better estimate of the forecast, provides probability information for a deterministic forecast, and thus probability forecasts offer a tool for a confidence metric.

### 1. Simple Mean Model

The simple mean model is an unweighted arithmetic mean of the individual TLMM ensemble members. The number of ensemble members for each ensemble mean varies between 5 and 15, corresponding to 1 to 3 observation sites and a forecast period of Tau 0 to 24. Figure 22 is a time series plot of the *u*-and *v*-component winds for each platform's time-lagged mean and shaded two standard deviations around the mean. Black dots are the interpolated forecasts to the grids for the 0- to 24-hour forecasts. The four SAWE are shaded in gray according to their duration. The purpose of Figure 22 is to see error or dispersion of the simple mean model, highlighting time periods of high bias.

For the month period of 26 September 0600 UTC to 26 October 0600 UTC, Buoys 21 and 22 have similar means and standard deviations. However, the mean and standard deviation of the R/P FLIP is slightly different from the other two locations, most likely due to the distance from the coast and interpolation to the grids. During SAWE, there is large spread around the ensemble mean prediction. Changes in the forecast values between model updates lead to large variance in the mean prediction. In fact, in Figure 23 of the overall mean, the blue shaded region shows large dispersion in the *v*-component winds around 10 October 2017, noted as a significant SAWE.

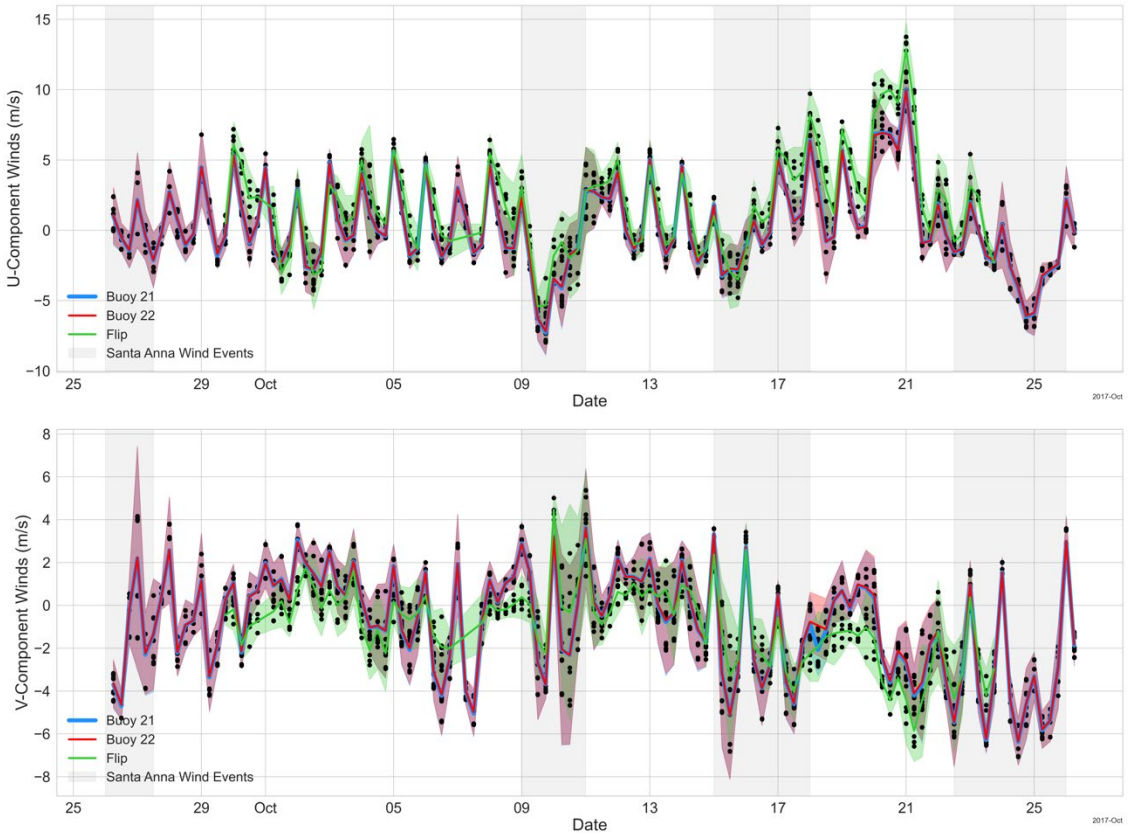


Figure 22. A time series plot of the simple mean TLMM average and two standard deviations for each platform. Values are averaged over all platforms; i.e., calculations do not differentiate observation locations.

Although limitations arise from the simple and unsophisticated construction of the unweighted mean model, it also provides insight into the variability of the ensemble mean. The central tendency of the mean is more accurate when TLMM are in close alignment; conversely, the mean bias is high when TLMM are dispersive. For example, diurnal variability in winds relate to periods of high and low mean bias. Also, there is more bias observed at the end of or after SAWE. For example, there is large spread around 27 September 2017. In both scenarios, the simple mean model captures variability of the ensemble mean. A limitation of averaging over all locations and time-lagged members is that the mean estimate is smoothed, and variability can be minimized such as that seen in Figure 23 and Figure 22 for 10 October 2017. In this case, the bias is high compared to neighboring time periods, but the bias is reduced by averaging over all locations.

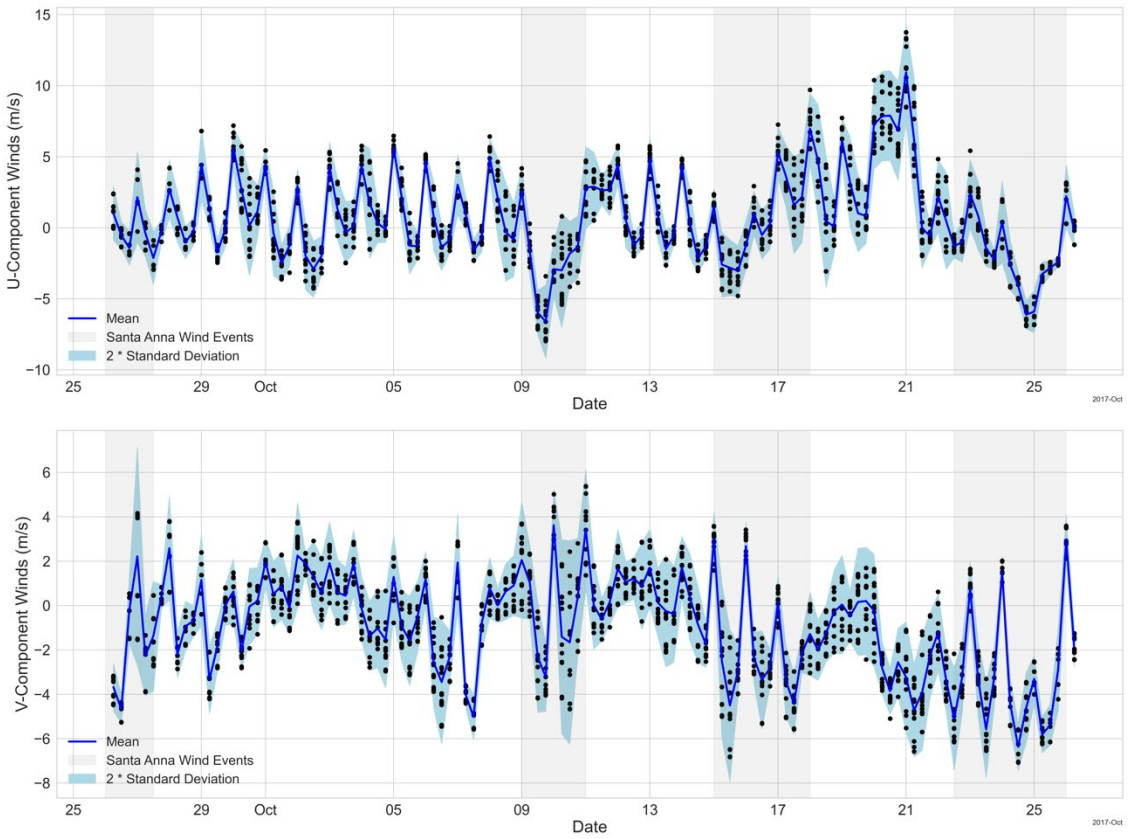


Figure 23. This is a time series plot of the simple mean TLMM average and two standard deviations. Values are averaged over all platforms; i.e., calculations do not differentiate observation locations.

Simultaneously comparing the expected value and variance is another way to analyze probabilistic information of the simple mean model. Figure 24 is a time series plot of the time-lagged mean and variance averaged over all platforms. The thin blue line represents the ensemble mean, and its vertical axis is provided on the left given in meters per second. The thick light blue line represents the variance, and its vertical axis is provided on the right given is meters squared per second squared. Light gray shading corresponds to the four SAWE with their duration. This plot is better at identifying periods of high bias in the TLMM prediction. In Figure 24, the significant variance in the *v*-component plot correlates well with SAWE. The *u*-component mean shows high variance around those time periods, but it also shows high variance about 30 September. High variance around 30 September relates to a transient trough.


Figure 24. Time series plot of the simple mean TLMM average and variance. Values are averaged over all platforms; i.e., calculations do not differentiate observation locations.

Error between the TLMM prediction and the observation measures model performance. Again, the purpose of this thesis is not the accuracy of the prediction; however, error in the ensemble prediction discloses uncertainty in the TLMM. The RMSE calculations are a direct assessment of model performance and is defined in Equation (4). High RMSE values indicate poor performance of the prediction, whereas low RMSE values indicate good performance.

Figure 22, Figure 23, and Figure 24 do not consider the observations in development of the TLMM ensemble; nonetheless, a time series plot (Figure 25) of the TLMM RMSE for each platform is constructed to assess how well the simple mean ensemble model predicts the wind conditions. Depending on the observation location, RMSE values range from near 0.5 m/s to as high as 10 m/s. Similar to the mean and

variance plot of *v*-component wind (Figure 24, bottom), large RMSE values correlate well with SAWE. There is also a diurnal pattern of RMSE for both *u*- and *v*-component wind. In addition, R/P FLIP (green) seems to have a different RMSE pattern than the buoys, suggesting a coastal influence. Also, Buoy 22 (red), which is half the distance between Buoy 21 and R/P FLIP, has *v*-component RMSE values that are closely aligned with the average RMSE, suggesting spatial variability in the mean estimates. In summary, overall large and variable RMSE suggest that there is high uncertainty and low confidence in the TLMM ensemble.

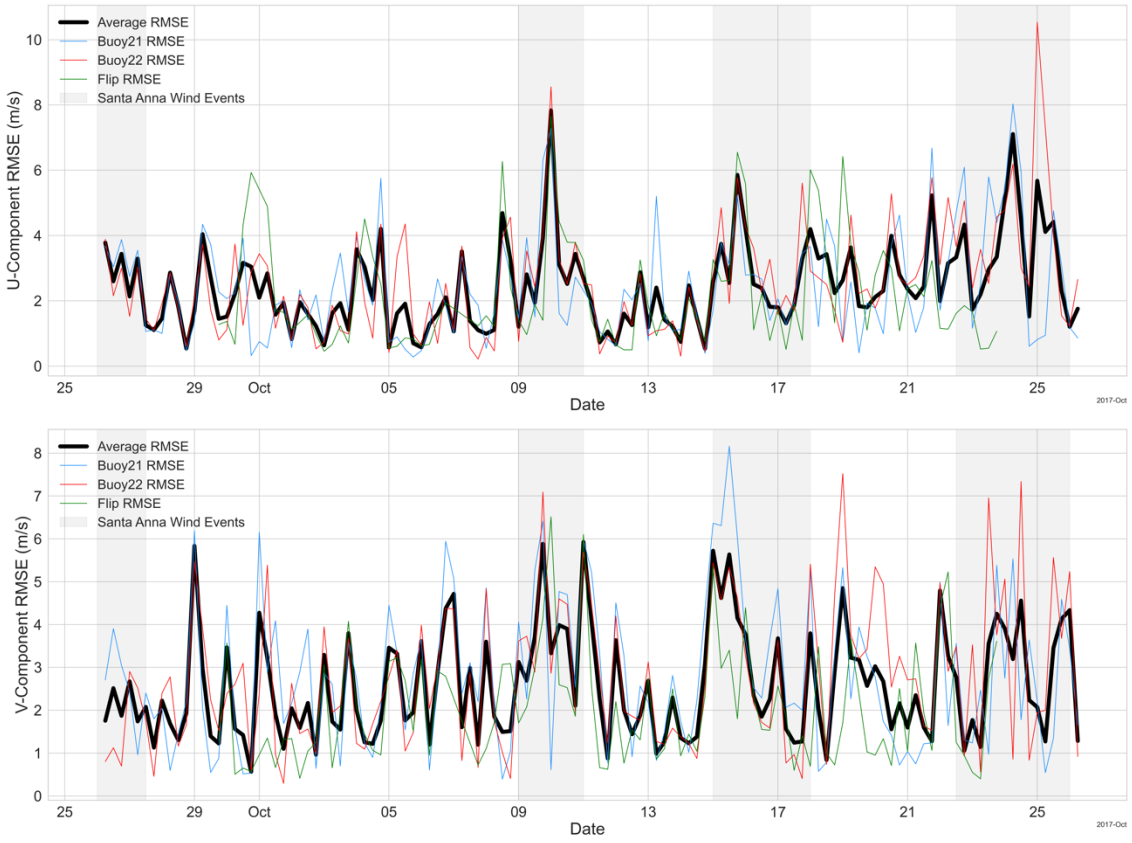


Figure 25. A time series plot of the simple mean TLMM RMSE for each platform and wind component. The black line is the mean RMSE for all platforms. Platform-specific RMSE values are given by the blue, red, and green lines for Buoy 21, Buoy 22, and R/P Flip, respectively.

### 2. Weighted Mean Model

The weighted mean model is modified arithmetic mean specified to make the analysis time; i.e., Tau 0, the most significant and the 24-hour forecast the least significant. In the TLMM ensemble, weights are defined in Equation (29) and predictions are provided by Equations (30) and (31). Similar to the simple mean model, plots and error statistics are analyzed to evaluate the performance of this TLMM ensemble variation. Figure 26 is a time series plot of the *u*- and *v*-component wind for each platform's weighted mean and shaded two standard deviations around the mean; black dots are the interpolated forecasts to the grids, and the four SAWE are shaded in gray. Like the simple mean ensemble, there are periods of high bias in and around SAWE; however, the weight mean model is better at showing periods of large dispersion between the time-lagged members, especially for the *v*-component winds.

In Lu et al. (2006) analysis of unequally weighted means to construct a short-range ensemble model using time-lagged ensemble members, they found that model-derived weights generated better results for weather prediction variables. Accordingly, in this TLMM analysis the weighted mean model does perform slightly better than the simple mean model; however, weights could be optimized based on errors attributable to a given forecast. The careful construction of weights could improve prediction but is not fully explored in this analysis.


Figure 26. Times series plot of the weighted time-lagged mean and two standard deviations for each platform.

### 3. Bayesian Regression Model

For Bayesian model averaging, the Bayes model first requires training data to fit and optimize the model. The data was partitioned into 75% training and 25% testing. Bayes model parameters were applied to the testing data set for ensemble prediction. Please refer to Chapter IV.D or Wendt (2017) for an explanation of how predictions are made. There are 84 observations in the testing dataset. All plots presented in this section are of the testing dataset for one iteration of the Bayesian model. Plots in this section also include the simple mean and weight mean estimates.

Figure 27 is a bias analysis of the testing dataset. The x-axis represents the index of the testing dataset vector. The index is not a reference of the date, but dates increase as the index increases. The error bars of member spread represent the range in ME of the time-

lag members. Note by design, the simple mean and weighted mean fall within the member dispersion bars. The Bayesian model average may or may not fall within this range. In some cases, such as that observed for index 6 to 9, the Bayesian model has less bias than the simple mean or weighted mean estimates.

Figure 28 presents the same information as Figure 27, except this plot shows bias by date. Notice that some dates have more than one prediction. That is because the original data set considers interpolated forecast at 3 locations. Here it is seen that there is large member dispersion in and around SAWE and around 29 September 2017. Moreover, for those same time periods the Bayesian estimate has smaller bias than the simple mean, weighted mean, and any individual TLMM member.

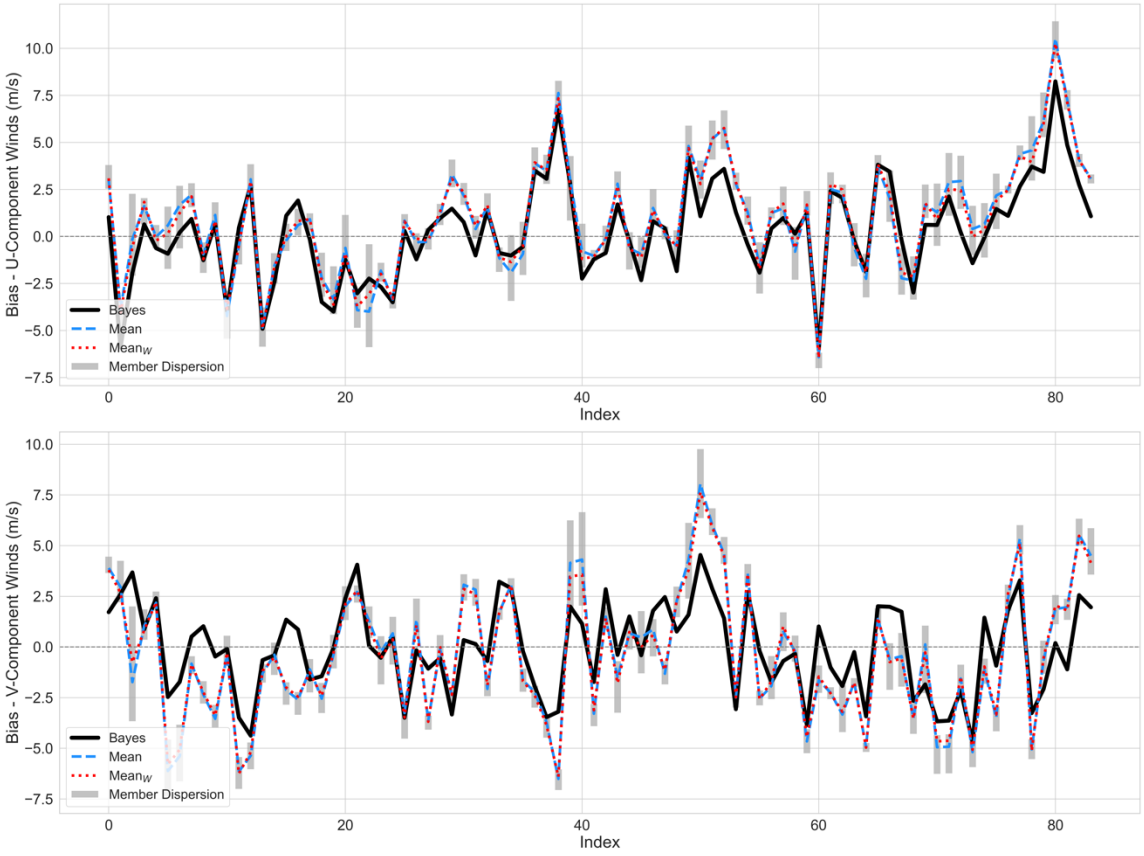


Figure 27. Bias analysis of the testing dataset. The x-axis represents the index of the test dataset vector. The index does not consider the date, but dates do increase as the index increases. The error bars represent the range in the time-lag members.

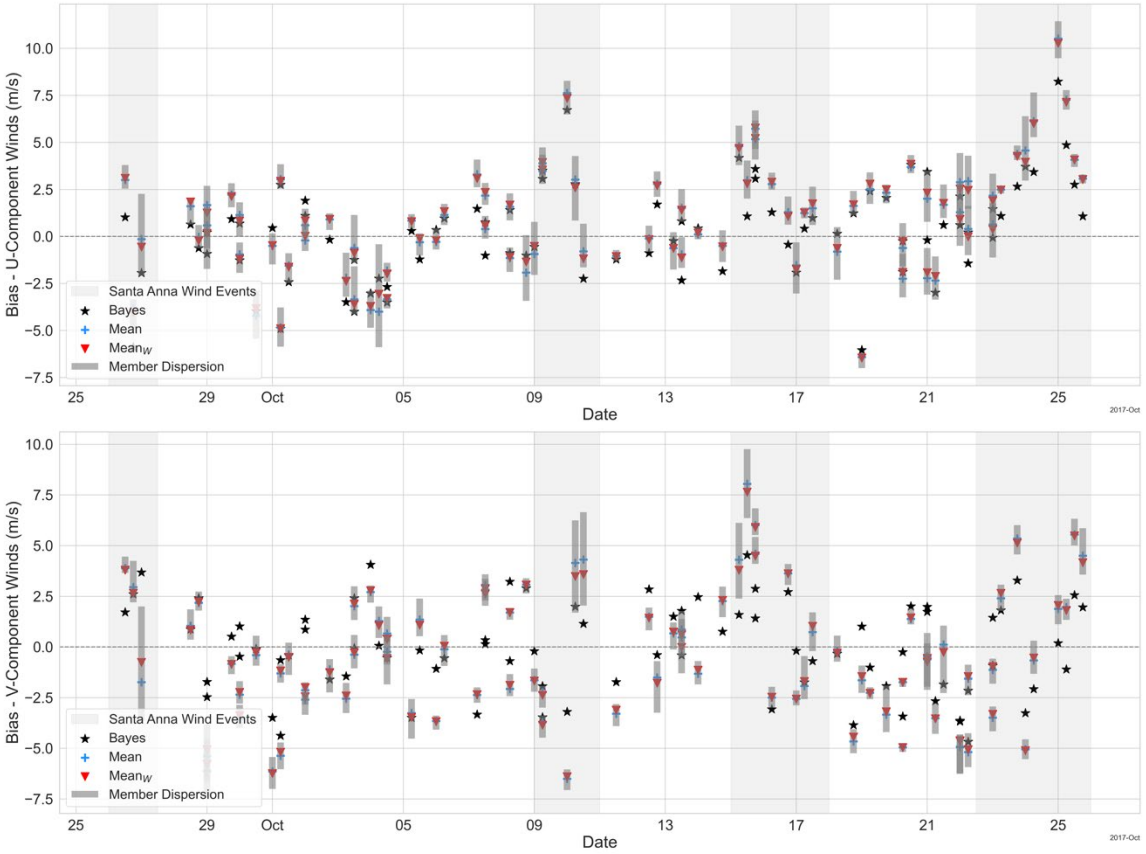


Figure 28. The plot shows the same information as Figure 27, but plots the ME or bias estimates by date. The x-axis represents the forecast valid time. The error bars represent the range of ME in the TLMM members.

Figure 29 and Figure 30 show the MSE. MSE amplifies the bias already seen in the ME plots. High tendency during SAWEs is observed in both the *u*- and *v*-component winds, with very few periods of high MSE occurring outside SAWEs. There is also some smoothing or correction happening within the Bayes model and is easily seen in the index plot (Figure 29) of the *v*-component winds. Bayesian MSE is generally less than that of the simple mean or weighted mean models. A summary of the ME, MSE, and RMSE of each TLMM variation and individual ensemble member is provided in Table 9. The table shows for the *u*- and *v*-component winds the Bayes model is a good performer and bias is reduced when compared to the observations. Moreover, the individual time members perform about the same; and even the analysis time does not outperform the Bayesian model averaging. When looking at RMSE, all models and TLMM ensemble members perform about the

same and yield an approximate 3 m/s margin of error. The *v*-component Bayesian model average does extremely well compared to the other model variations and ensemble members. The lower RMSE, ME, and MSE suggest that the model helps to correct some of the directional bias noted in the exploratory analysis.

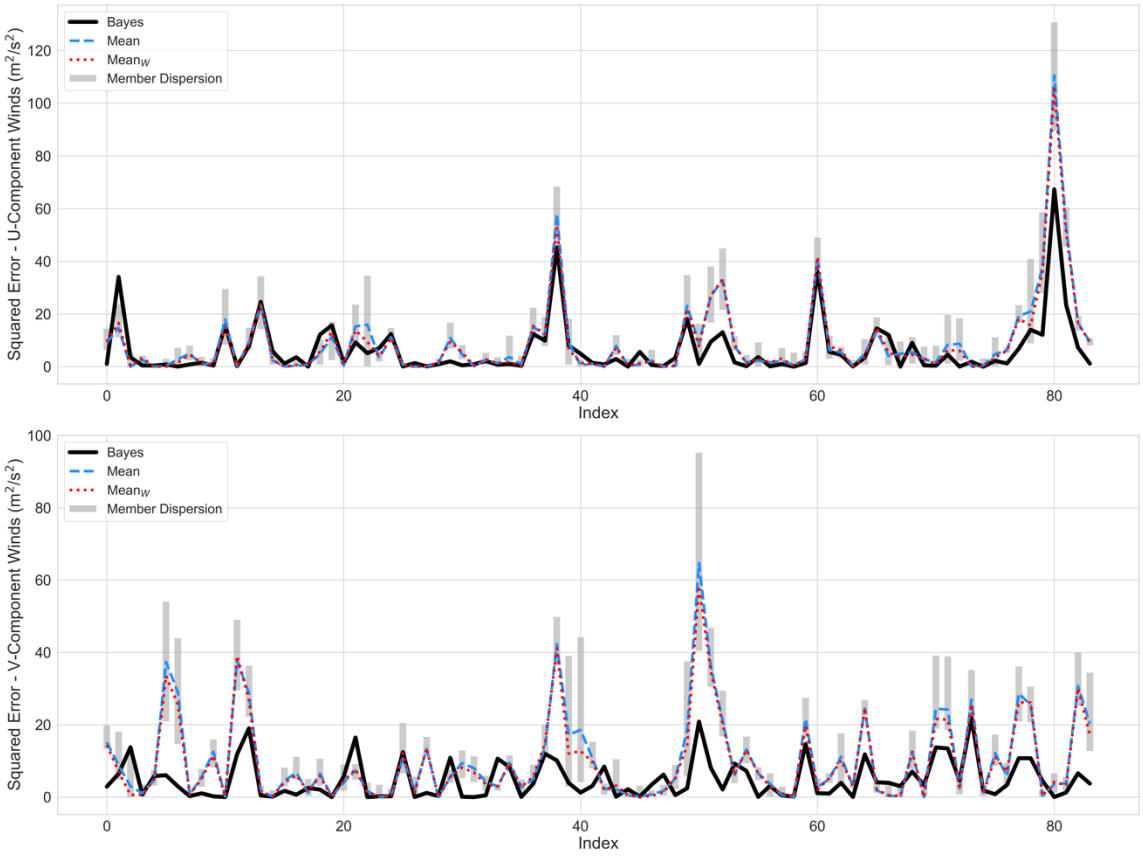


Figure 29. Bias analysis of the testing dataset. The x-axis represents the index of the test dataset vector. The index does not consider the date, but dates do increase as the index increases. The error bars represent the range in the time-lagged member squared error estimates.

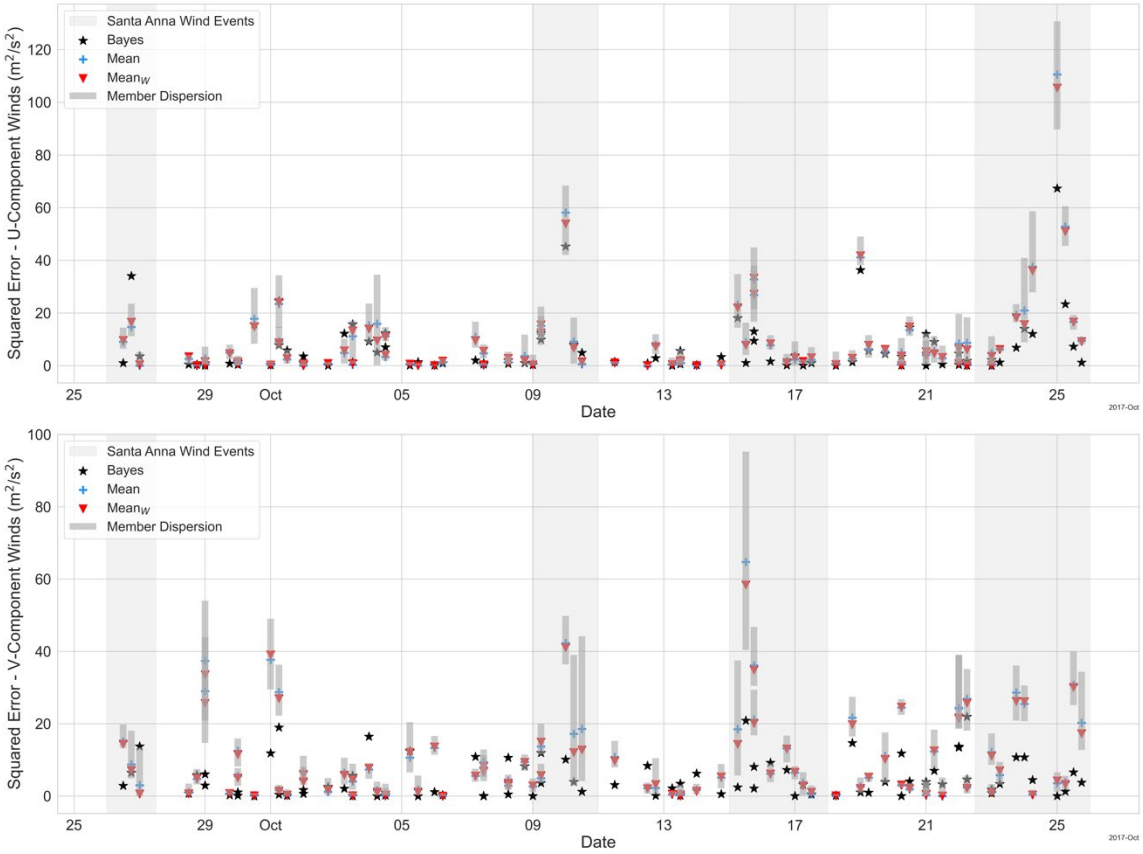


Figure 30. The plot shows the same information as Figure 29, but plots the MSE estimates by date. The x-axis represents the valid forecast time. The error bars represent the range of MSE in the TLMM members.

Table 9. Summary of the ME, MSE, and RMSE of each model and individual ensemble member.

| **10 m U-Component Wind Speeds (m/s)** | | | |
|---|---|---|---|
| | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) |
| Bayes Average | 0.287 | 6.541 | 2.557 |
| Simple Mean | 0.937 | 9.343 | 3.057 |
| Weighted Mean | 0.905 | 8.930 | 2.988 |
| Tau 0 | 0.985 | 8.828 | 2.971 |
| Tau 6 | 0.763 | 8.430 | 2.903 |
| Tau 12 | 0.845 | 11.131 | 3.336 |
| Tau 18 | 0.966 | 10.313 | 3.211 |
| Tau 24 | 1.130 | 10.900 | 3.301 |
| **10 m V-Component Wind Speeds (m/s)** | | | |
| Ensemble/ Member | Mean Error (ME) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) |
| Bayes Average | -0.099 | 4.837 | 2.199 |
| Simple Mean | -0.451 | 10.376 | 3.221 |
| Weighted Mean | -0.475 | 9.680 | 3.111 |
| Tau 0 | -0.543 | 9.472 | 3.078 |
| Tau 6 | -0.503 | 8.787 | 2.964 |
| Tau 12 | -0.403 | 10.813 | 3.288 |
| Tau 18 | -0.382 | 12.711 | 3.565 |
| Tau 24 | -0.426 | 12.661 | 3.558 |

## B. MODEL LIMITATIONS

There are some limitations to the TLMM ensemble such as the construction number of ensemble members and initialization of the Bayesian model. As mentioned in the Assumption section (Chapter IV.E), *a priori* model specifications are required and are determined from available training data. Operationally, this adds a computational requirement when trying to estimate the Bayesian parameters, but once the estimates are computed Bayesian model averaging can be applied for prediction. In addition, the model used here does not consider interactions between predictors or higher-order model terms. Considering that there are some significant correlations between *u*- and *v*-component winds

(Figure 20), adding these terms to the TLMM ensemble would help to improve model fitting and performance, but adding more terms also adds to the computational cost. Conversely, other considerations should be made when increasing the ensemble member and parameter size. There is a balance between the number of observations required and the number of parameters estimated. Generally, this should not be a problem if a large COAMPS grid is considered and not the observational TLMM ensemble.

Moreover, there are also limitations to the data employed in the TLMM. There is an anticipated initialization error with each model run as new observations and initial conditions are applied to the NWP models. There is error in the boundary conditions with each grid initialization, and there is an assimilation error of the observations as verified from the FNMOC/NRL graphics (Figure 16 and Figure 17). In an analysis of a time-lagged ensemble from successive United Model 4km mesoscale resolution model runs, Mittermaier (2007) claims that higher resolution models are more sensitive to initial boundary conditions than coarser resolution grids. Errors in the initialization can lead to bias in the predictions and time-lagged ensemble, but this is also dependent on the relative skill of the model to predict in benign or difficult atmospheric conditions (Mittermaier 2007). Generally, errors, whether systematic, random, reducible, or irreducible, must be minimized and optimized. Like many statistical models, the goal is to minimize error and capture the true variation of the dataset. However, NWP model characteristics may change over time after the Bayesian TLMM has been trained. For example, NAVGEM and COAMPS have experienced many upgrades since 2017, and their accuracy in predicting the weather has improved. Nevertheless, the TLMM Bayesian model has limitations to the NWP data, and a careful approach to optimize the TLMM may be considered.

## C. PHYSICAL LIMITATIONS

Meteorological conditions add complexity to building the TLMM. In this thesis, we focus on weather for the Southern California Bight for September to October 2017, which over the month period weather phenomena with short timescales can be captured by the TLMM ensemble from input by the NWP models (Wallace and Hobbs 2006). However, if envisioned to train the Bayesian TLMM with longer period of data, then atmospheric

variability on longer time scales needs consideration. Also, other physical concerns include the following:

- The model may be limited to specific forecast regions where observational data is available to train the model. With the consideration of satellite observations and reanalysis data, this limitation may be void.
- The model may be limited temporally. There are seasonal and intraseasonal atmospheric variability, so more than one TLMM ensemble could be applied over different time periods.
- Depending on the NWP model resolution, model winds are portrayed as (coastal) synoptically driven. There is diurnal variability in the CASPER-West observations, but COAMPS and NAVGEM models are directionally biased. COAMPS and NAVGEM capture major synoptic events but degrade for SAWE and significant pressure system.
- Wind speeds less than 2.6 m/s or 5 knots are considered light and variable; consequently, persistent wind speeds are hard to discern. Representativeness of forecasted southernly wind is an issue for the CASPER-West dataset.

## D. VALIDATION

Validation of the TLMM ensemble is crucial because it asserts if results are credible and reproducible. Similar for the case of one iteration, for the model validation the dataset was randomly partitioned into 2 parts: 75% of the data used for training and 25% for testing. This process was iterated 20 times. Table 10 is a summary of validation performance illustrating the improved performance of the Bayesian model average to the simple mean, weight mean, and individual ensemble members. For the *u*-component winds, the Bayesian RMSE is less than the simple mean, weighted mean, and individual ensemble members at least 75% of the time. More interestingly, the *v*-component RMSE is consistently the best performer. Validation results suggest that for any uncertainty

information extracted from the probability prediction, there is high confidence in the prediction produced in this TLMM analysis.

The total error of the validation model is the mean RMSE generated from each iteration. Table 11 provides the mean RMSE for the *u*- and *v*-component winds. The Bayesian TLMM ensemble has the least mean RMSE for both *u*- and *v*-component winds. The analysis time is the next lead performer for the zonal wind; however, the 6-hour forecast does better for the meridional wind.

Table 10. Summary of validation performance. Shows a comparison of the Bayes model to the simple mean, weight mean, and individual ensemble members.

| ***u*-Component RMSE (m/s)** | **COUNT** | **Probability (pct)** |
|---|---|---|
| Bayes < Simple Mean | 15 | 0.75 |
| Bayes < Weighted Mean | 15 | 0.75 |
| Bayes < Tau 0 | 16 | 0.80 |
| Bayes < Tau 6 | 15 | 0.75 |
| Bayes < Tau 12 | 19 | 0.95 |
| Bayes < Tau 18 | 20 | 1.00 |
| Bayes < Tau 24 | 20 | 1.00 |
| ***v*-Component RMSE (m/s)** | **COUNT** | **Probability (pct)** |
| Bayes < Simple Mean | 20 | 1.00 |
| Bayes < Weighted Mean | 20 | 1.00 |
| Bayes < Tau 0 | 20 | 1.00 |
| Bayes < Tau 6 | 20 | 1.00 |
| Bayes < Tau 12 | 20 | 1.00 |
| Bayes < Tau 18 | 20 | 1.00 |
| Bayes < Tau 24 | 20 | 1.00 |

Table 11. Summary of verification performance. Table show the RMSE for *u*- and *v*-component winds for each ensemble and individual ensemble members.

| | **Bayes Average** | **Simple Mean** | **Weighted Mean** | **Tau 0** | **Tau 6** | **Tau 12** | **Tau 18** | **Tau 24** |
|---|---|---|---|---|---|---|---|---|
| ***u*-Component Mean RMSE (m/s)** | 2.593 | 2.772 | 2.731 | 2.770 | 2.728 | 2.894 | 2.903 | 3.003 |
| ***v*-Component Mean RMSE (m/s)** | 2.116 | 2.982 | 2.885 | 2.830 | 2.783 | 3.081 | 3.312 | 3.281 |

An approach to optimize the Bayesian TLMM is the selection of significant parameters through credible intervals. The Bayesian TLMM produced 22 parameter estimates defined by the matrix, Equation (32). These parameters are estimated by Markov Chain Monte Carlo (MCMC) sampling methods and presented as B_0, B_1, B_2, … B_21 in Figure 31 of the original fit of the Bayesian model. Figure 31 is a box and whisker plot of the credible intervals estimated from the one million MCMC sample. Bayesian credible intervals "characterize the probability that the true parameter value lies in the indicated interval given the fixed data sample" (Wendt 2017) and can be used as a method to validate which parameters are statistically significant in the prediction. From the 20-fold validation of the TLMM, the Interquartile Range (IQR) and min-max range can determine if the posterior parameter estimates are significantly different from zero at the 75% and 99% level.

Table 12 is a posterior parameter summary for the test for significance. There is high confidence that the intercepts B_0 and B_1 are not significant in the model given that the IQR and min-max range includes zero for all 20 iterations. Interesting, B_11 and B_12 relating to the 12-hour forecasts have low probability of being deemed significant; this is also true for B_14 representing the sensitivity of the *u*-component observed winds to 18-hour forecasted *u*-component wind. Other posterior parameters such as B_15, B_16, B_19, B_20, and B_21 are statistically different from zero. B_15 corresponds to *u*-component 18-hour forecast's influence on the *v*-component observed winds, and B_16 corresponds to the sensitivity of the *u*-component observed winds to the 18-hour forecast *v*-component

wind. B_19, B_20, and B_21 relate to the 24-hour forecasts on the *u*- and *v*-component observed winds. In summary, validation results of the posterior parameter estimates suggest that the 24-hour and earlier forecast hours have a greater influence on the TLMM prediction.

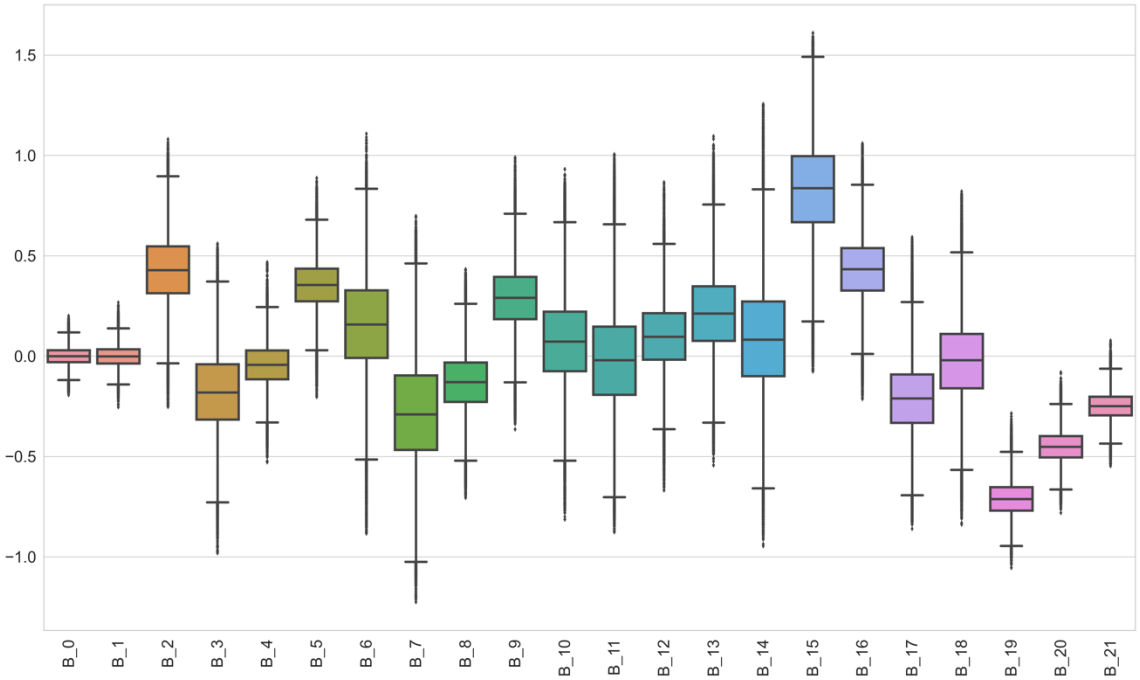


Figure 31. Example of Bayesian Parameters, estimated by MCMC sampling methods. These estimates are from the original fit of the Bayesian model

Table 12. Summary of the significance of Bayesian Parameters, estimated by MCMC sampling methods, in the validation process.

| Bayesian Parameters | Interquartile Range | | Min-Max Range | |
|---|---|---|---|---|
| | Count | Probability (pct) | Count | Probability (pct) |
| B_0 | 0 | 0% | 0 | 0% |
| B_1 | 0 | 0% | 0 | 0% |
| B_2 | 20 | 100% | 3 | 15% |
| B_3 | 13 | 65% | 0 | 0% |
| B_4 | 15 | 75% | 0 | 0% |
| B_5 | 20 | 100% | 13 | 65% |
| B_6 | 15 | 75% | 0 | 0% |
| B_7 | 16 | 80% | 0 | 0% |
| B_8 | 11 | 55% | 0 | 0% |
| B_9 | 20 | 100% | 0 | 0% |
| B_10 | 9 | 45% | 0 | 0% |
| B_11 | 4 | 20% | 0 | 0% |
| B_12 | 5 | 25% | 0 | 0% |
| B_13 | 20 | 100% | 1 | 5% |
| B_14 | 2 | 10% | 0 | 0% |
| B_15 | 20 | 100% | 16 | 80% |
| B_16 | 20 | 100% | 18 | 90% |
| B_17 | 20 | 100% | 8 | 40% |
| B_18 | 11 | 55% | 0 | 0% |
| B_19 | 20 | 100% | 20 | 100% |
| B_20 | 20 | 100% | 20 | 100% |
| B_21 | 20 | 100% | 16 | 80% |

## E. PRACTICAL APPLICATION TO OPERATIONS

Development of the TLMM ensemble is motivated by the goal of creating a COAMPS confidence metric for the forecaster. Recall in Chapter I.A, the practical objective for the forecaster is confidence in the forecast produced; and for any uncertainty present, there is clear communication to the warfighter of the error and risk margin. A single COAMPS forecast consists of one high-resolution, high-precision depiction of the current meteorological conditions. A single forecast does not provide a measure of

confidence and accuracy. Therefore, it is worth exploring probabilistic information achieved by the TLMM. In particular, "two useful characteristics of a probability distribution are its expected value and its variance. Expected value is a measure of the location of the distribution, while variance is a measure of its spread" (Dowdy et al. 2004). Hence, we may want to consider the variance of the TLMM ensembles, Bayesian estimate, simple mean, or weighted mean, to determine the "confidence level" or severity of error.

The difference between the Bayesian models and the mean models is the reference to which variance is calculated. The simple mean and weighted mean models do not consider variance relative to the observed values. In the Bayesian model, the model is trained with the observations, and variance is relative to the observed values. Bayesian model variance is partitioned into model predictive variance and posterior uncertainty in the parameter estimates (Gelman et al. 2021). Moreover, Bayesian variance also includes error in the parameter estimates of the multivariate multiple linear regression log-likelihood framework for the Bayesian *a prior* assumption (Wendt 2007).

Figure 32 is a variance scatterplot of the Bayesian estimate, simple mean, and weighted mean models. The mean variance for the Bayesian model is much higher than for the mean models, but this is because of how Bayesian model variance is estimated. A variance signal is observed in all models; however, the signal is more defined in the Bayesian model. Note that during all times of identified high uncertainty events, i.e. SAWE, there is a peak in variance.

A natural next step is to determine a threshold to identify ranges of high, medium, or low variance. For one approach, variance threshold ranges would be applied to the entire COAMPS grid for an overall uncertainty metric for the grid. Then, uncertainty metrics will be calculated for each tau of the forecast and each grid point. Calculation over each grid point enables a display of "COAMPS uncertainty as a time varying [red/yellow/green] heat map over the spatial grid, and/or as a single uncertainty metric [for the grid]" (Naval Research Laboratory Monterey 2020).

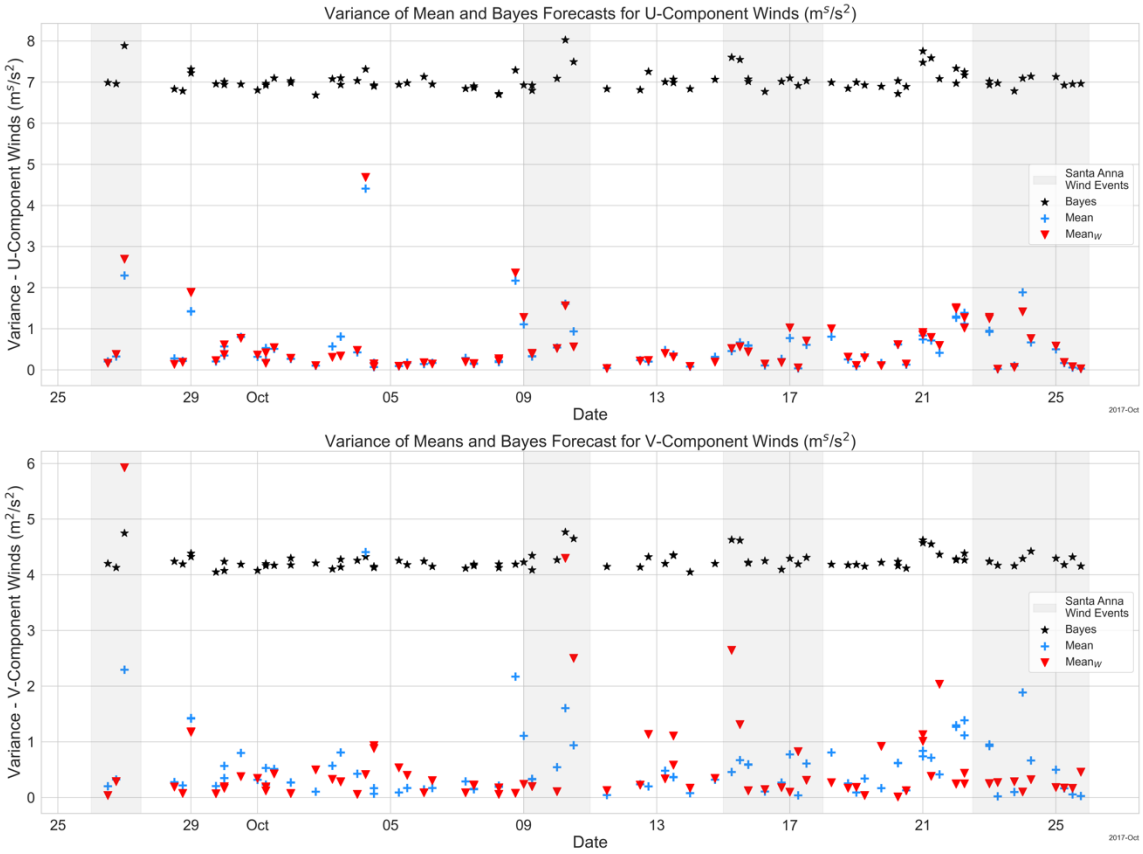


Figure 32. Variance of Bayesian estimate, simple mean, and weighted mean.

# VI. SUMMARY AND CONCLUSIONS

This statistical learning approach examined the development of a confidence metric using a TLMM ensemble: simple mean, weighted mean, and Bayesian model averaging. Keeping in mind the goal of this research, the aim was not to improve the prediction capability of COAMPS NWP deterministic output, but rather to convey uncertainty of select variables to the end-user. Toward developing this uncertainty metric, an observational model was applied to COAMPS nest grids and NAVGEM for interpolated observation locations as provided by the CASPER-West campaign. The TLMM design helps to make inference about the sources of error and relationships between variables. The uncertainty information inferred by the TLMM models provides the forecasters with knowledge of model variability over the last 24 hours and across an area of interest.

The TLMM analysis showed the Bayesian regression model improves prediction compared to the simple mean, weighted mean, and individual ensemble members. The error analysis suggests that in situations where there is large dispersion between TLMM members and in times of high model error, the Bayesian TLMM minimizes error in the prediction, especially for the *v*-component winds. A statistically significant increase in performance for the *u*-component winds is not observed; however, improvement in the prediction was observed. Consequently, as a forecaster having COAMPS model uncertainty available can help provide understanding to reliability and probability of atmospheric predictions.

The wind components were selected for the TLMM analysis due to a limitation of the CASPER-West COAMPS and NAVGEM dataset. Wind stress over coast and near shore environments has a large impact on other state variables significant to EM propagation. The Southern California Bight is dynamically and topographically complex affected by coastally constrained and synoptic-level atmospheric conditions. Therefore, the thesis design determined proof-of-concept for the TLMM ensemble applied to component winds, and an analysis for other variables such as those mentioned in Chapter III.C should be considered.

## A. RECOMMENDATIONS

The thesis is motivated to provide the METOC forecasters with a COAMPS uncertainty metric to have access to in EM TDA applications. Although progress was made to develop an uncertainty metric, time constraints did not permit the construction of a model scorecard. The analysis presented here is still useful for the operator. At a minimum, the uncertainty observed in the mean, weighted mean, and Bayesian model averaging tell the operators about the forecast consistency between model initializations and grid resolutions. The TLMM ensemble uncertainty calculations are a tool linking error in the atmospheric dynamics and NWP modeling to translations of error in the propagation analysis.

Data and time permitting, this thesis would look at other considerations to build the TLMM ensemble, such as the following:

- There are other significant variables affecting near-surface EM propagation. Many essential variables were presented in the error analysis report (Chapter III.C). For example, SST is a required input for evaporation duct calculations and is a primary factor affecting near-surface EM propagation (Wang et al. 2018). In Figure 8, grid differences in forecasted SST performance suggest model resolution is a critical consideration for a SST TLMM ensemble. Moreover, it is worth exploring the interactions and relationships between the complete set of significant variables.

- This analysis assumes forecasters may only have the 6-hourly forecast available to them. If hourly observations and predictions are made, the TLMM ensemble can be altered to include more ensemble members. An important consideration is the convergence of the Bayesian model. With the 5-member TLMM ensemble for both u- and v-component, the joint parameter matrix estimates 22 parameters to include the intercepts. Additional TLMM members will increase the computational runtime for the model. It took approximately 6 minutes each time the linear Bayesian

TLMM was run. Another approach is to include forecast periods beyond 24 hours. COAMPS and NAVGEM models can extend out to 5-days. However, careful consideration of the model run should still be made.

- The proof-of-concept model applied the Bayesian TLMM to an observational dataset. The next step in the TLMM design is to apply the TLMM model to the entire grid, mostly likely constrained by the finer resolution grid boundaries. Conversely, the TLMM could be explored over different observation locations. Depending on the selected area, a specific resolution may be needed to resolve scale-specific atmospheric features, such as eddies within the near-shore region versus transient synoptic systems over the open ocean.

- Weights in the weighted mean TLMM ensemble were selected by assuming the analysis time is the best performer, and the model degrades steadily as forecast time increases. Choosing weights appropriately for TLMM members and grid resolutions could produce a comparable prediction to the Bayesian model. A recommendation is to optimize the weights using error metrics as a function of time.

- The method of testing versus training datasets and validation of the model is specific to this analysis, given the limitations of the data. Considering what was analyzed and to see if the results are consistent, a more robust method to train, test, and validate the TLMM with new data may be of interest. Additionally, increasing the length of the dataset would allow for random block sampling; see Figure 21 for reference. For a more extended dataset, the Bayesian TLMM would be able to capture longer periodicities. Also, as Wendt (2017) explored, the Bayesian model was tested on the latter half of the time series dataset, and consideration is "a full evaluation of training data updates that properly explores the relationship between training period length and predictive performance."

- Code for the Bayesian TLMM was fundamentally unmodified such that only linear terms were estimated. A recommendation is to consider variable interactions and higher-order terms in the parameter matrix. Increasing the number of parameters also decreases the degrees of freedom in the model. It is recommended to reference the ensemble modeling text (Du et al. 2018) that mentions a balance between a number of minimum and maximum ensemble members based on required degrees of freedom. In addition, there is also a requirement to have enough data to train and test the model.

- Grid-to-grid differences were observed in the data characterization step. The pertinent question is, what is the importance of considering the mesoscale and synoptic-scale models in the TLMM ensemble? There are differences in parameterization schemes, initialization of the models, and spatial scale between NAVGEM and COAMPS. Furthermore, this analysis shows the best resolution model is not always the best for a particular area, time of the season, and weather phenomena. It is worth investigating a TLMM ensemble constructed for each grid resolution to isolate grid-to-grid effects on the inclusive TLMM ensemble.

- There were no observations excluded from the analysis based on wind speed. Wind speeds less than 5 knots are considered light winds, and directionality is variable. Therefore, when wind speed and direction are decomposed into u- and v-components, there may be some perceived directional bias when the actual wind direction cannot be ascertained. However, it is essential to understand that removing wind speeds less than 5 knots from the analysis may also affect representativeness but is location dependent.

- Lastly, as noted in the model data section (Chapter II.B), there have been improvements to COAMPS and NAVGEM NWP models. For example, NAVGEM 2.0 is currently in operation (FNMOC 2020), and ship

following COAMPS has been made available to deployed ships. This analysis could easily be reapplied to an updated version of COAMPS and NAVGEM data to compare to the study area of interest.

## B. FUTURE WORK

As presented in this report, the performance of the Bayesian TLMM model was proven to be an outperformer and has applications beyond the TLMM error metric. It would be worth exploring predictive applications to the TLMM ensemble. For example, METOC forecasts may or may not have the latest forecast or climatology available to ingest in the EM/RF TDA. The Bayesian TLMM ensemble can be applied over the last 24-hour forecasts for a Bayesian model average prediction. This data-limited scenario is probable as ships intermittently and occasionally lose internet connectivity. However, a METOC forecaster not having the latest model available should not be a limiting factor to producing accurate forecasts or to the forecaster's confidence in their prediction. Mittermaier (2007) showed that time-lagged ensemble modeling has the grid-equivalent skill to create a forecast that is "at least equivalent to, and possibly superior to, spatial upscaling of an individual forecast." Given that the finer resolution grid may not be available, a TLMM ensemble would be a better application than the synoptic model and perhaps climatology. Furthermore, this TLMM forecast ensemble applies to the Bayesian Model developed by Wendt (2017). Similarly, as shown in this analysis and Wendt (2017), the Bayesian Model is a superior contestant compared to mean models and generalized regression.

In conclusion, the primary focus of this thesis is to provide the forecaster with an uncertainty metric to reference while creating COAMPS-driven EM/RF propagation predictions. A future application to the Bayesian TLMM ensemble is to correlate high Bayesian variance with times of COAMPS-driven propagation loss. This task requires careful experiment planning because propagation loss observations are not readily available to verify errors in the propagation model. Nevertheless, the TLMM ensemble will show actual value when it can be tested and validated operationally.

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF REFERENCES


Bean, B. R., and E. J. Dutton, 1966: The radio refractive index of air. *Radio meteorology*, 92, Superintendent of Documents, U.S. Government Print Office, 3–4.

Bean, B. R., and E. J. Dutton, 1966: Meteorological conditions associated with radio refractive index profiles. *Radio meteorology*, 92, Superintendent of Documents, U.S. Government Print Office, 132–133.

Branković, C., T. Palmer, F. Molteni, S. Tibaldi, U. Cubasch, 1990: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.*, **116, 494**, 867–912, https://doi.org/10.1002/qj.49711649405.

Department of Defense, 2020: Electromagnetic Spectrum Superiority Strategy. Accessed 01 February 2022, https://media.defense.gov/2020/Oct/29/2002525927/-1/-1/0/electromagnetic_spectrum_superiority_strategy.pdf.

Du, J., and J. Berner, R. Buizza, M. Charron, P. Houtekamer, D. Hou, J. Isidora, M. Mu, X. Wang, M. Wei, and H. Yuan, 2018: Ensemble methods for meteorological predictions. *NCEP Office Notes 493*, **66**, https://doi.org/10.7289/V5/ON-NCEP-493.

Faraway, J. J., 2006: Generalized linear models. *Extending the Linear Models with R*. Chapman & Hall/CRC, 115–133.

Faraway, J. J., 2005: *Linear Models with R*. Chapman & Hall/CRC, 1–106.

Dowdy, S., S. Wearden, D. Chilko, 2004: Distributions of two variables. *Statistics for Research*, John Wiley & Son, 211–250.

FNMOC: Model Verification Main Help Page. Accessed 18 January 2022, https://www.fnmoc.navy.mil/verify_cgi/html/verf_areas_help.html

FNMOC, 2015: Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) version 5.6.5 Operational Test (OPTEST) Report. Accessed 25 April 2022, https://portal.fnmoc.navy.mil/prodnews/html/docs/COAMPS_5.6_OPTEST_Report.pdf

FNMOC, 2018: Model Information. Accessed 18 January 2022, https://portal.fnmoc.navy.mil/modelinfo/ModelInfoDoc.html

FNMOC, 2018: NAVGEM 1⁄4 degree EXECSUM. Accessed 23 May 2022, https://portal.fnmoc.navy.mil/prodnews/html/docs/NAVGEMQrtdeg_EXECSUM_12OCT2018.pdf

FNMOC, 2020: Executive Summary - NAVGEM upgrade from v1.4.3 to v2.0. Accessed 25 April 2022, https://portal.fnmoc.navy.mil/prodnews/html/docs/NAVGEM2.0_ExecSum.pdf

FNMOC, 2020: Executive Summary COAMPS upgrade from v5.2 to v5.6. Accessed 18 January 2022, https://portal.fnmoc.navy.mil/prodnews/html/docs/COAMPS_5.6_ExecSum_24AUG20.pdf

Gallaudet, T., 2016: Naval Oceanography Electromagnetic Maneuver Warfare Strategy. Naval Meteorology and Oceanography Command, 1–6, https://www.navy.mil/Press-Office/News-Stories/Article/2260634/naval-oceanography-releases-electromagnetic-maneuver-warfare-strategy/

Gallo, B. T., C. P. Kalb, J. H. Gotway, H. H. Fisher, B. Roberts, I. L. Jirak, A. J. Clark, C. Alexander, and T. L. Jensen, 2019: Initial Development and Testing of a Convection-Allowing Model Scorecard. *Bull. Amer. Meteor. Soc.*, **100**, **12,** https://doi.org/10.1175/BAMS-D-18-0218.1

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, 2021: Single-parameter models. *Bayesian Data Analysis*, Chapman and Hall/CRC, 40–41, http://www.stat.columbia.edu/~gelman/book/BDA3.pdf.

Göber, M., E. Zsótér, and D. Richardson, 2008: Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. *Meteor. Appl.*, **15**, 359–365, https://doi.org/10.1002/met.78.

Härdle, W. K., and L. Simar, 2003: Comparison of batches. *Applied Multivariate Statistical Analysis.* Springer Nature, 13–31.

Härdle, W. K., and L. Simar, 2003: Linear model for two variables. *Applied Multivariate Statistical Analysis*. Springer Nature, 95–103.

Hastie, T., R. Tibshirani, and J. Friedman, 2009: Model Assessment and Selection. *The Elements of Statistical Learning*. Springer, 219–232.

Hogan, T.F., M. Liu, J.A. Ridout, M.S. Peng, T.R. Whitcomb, B.C. Ruston, C.A. Reynolds, S.D. Eckermann, J.R. Moskaitis, N.L. Baker, J.P. McCormack, K.C. Viner, J.G. McLay, M.K. Flatau, L. Xu, C. Chen, and S.W. Chang, 2014: The Navy Global Environmental Model, *Oceanography*, **27, 3**, 116–125, http://dx.doi.org/10.5670/oceanog.2014.73.

Hsu, H., L. Oey, W. Johnson, C. Dorman, and R. Hodur, 2007: Model wind over the central and southern California coastal ocean, *Mon. Wea. Rev.*, **135.5**, 1931–1944, https://doi.org/10.1175/MWR3389.1.

Iowa State University Iowa Environmental Mesonet, 2001: NWS text products by issuing center by date. Accessed 05 April 2022, https://mesonet.agron.iastate.edu/wx/afos/list.phtml.

Lu, C., Huiling, Y., Schwartz, B., Benjamin, S. G., 2007: Short-range numerical weather prediction using time-lagged ensembles, *Wea. Forecasting,* **22**, 580–595, https://doi.org/10.1175/WAF999.1.

Metcalfe, A.V., and P. S. P. Cowpertwait, 2009: Spectral analysis. *Introductory Time Series with R*, Springer-Verlag, 171–198.

Mittermaier, M. P., 2007: Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles, *Quart. J. Roy. Meteor. Soc*, **133, 627**, 1487–1500, https://doi.org/10.1002/qj.135.

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8.2,** 281–293, https://doi.org/10.1175/1520-0434(1993)008%3C0281:WIAGFA%3E2.0.CO;2.

Naval Research Laboratory, n.d.: COAMPS Grids, Projections, and Output Files in COAMPS-OS. Accessed 24 December 2020.

Naval Research Laboratory Monterey, 2020: Time-lagged-multi-model COAMPS uncertainty metric for EM/RF TDAs, a research planning letter submitted to: Office of Naval Research. Naval Research Laboratory Monterey, 4 pp. Accessed 02 September 2020.

Naval Research Laboratory Monterey, 2022: Southern California Climatology. Accessed 30 March 2022, http://www.nrlmry.navy.mil/nrlonly/local_area_forecaster_handbooks/frames3.htm.

Naval Research Laboratory Marine Meteorology Division, 2003: *COAMPS Version 3 Model Description*, NRL Publication, 1- 66.

Ortiz-Suslow, D. G., J. Kalogiros, R. Yamaguchi, D. Alappattu, and K. Franklin, B. Wauer, and Q. Wang, 2019: The data processing and quality control of the marine atmospheric boundary layer measurement systems deployed by the naval postgraduate school during the CASPER-West field campaign (Technical Report), Naval Postgraduate School, https://calhoun.nps.edu/handle/10945/61638.

Rolinski, T., S. B. Capps, and W. Zhuang, 2019: Santa ana winds: A descriptive climatology, *Wea. Forecasting,* **34.2**, 257–275, https://doi.org/10.1175/WAF-D-18-0160.1.

Schultz, M. G., C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L.H. Leufen, A. Mozaffari, and S. Stadtler, 2021: Can deep learning beat numerical weather prediction?, *Phil. Trans. R. Soc. A.*, **379**, 20200097, https://doi.org/10.1098/rsta.2020.0097.

Scripps Institution of Oceanography, n.d.: Research Platform FLIP (FLoating Instrument Platform) History. Accessed 21 April 2022, https://scripps.ucsd.edu/ships/flip/history.

Thornton, 2007: OC3150 Analysis of air-ocean time series, Department of Oceanography Naval Postgraduate School.

U.S. Naval Meteorology and Oceanography Command Public Affairs Office, 2021: U.S. Naval Oceanography participates in LSE 2021. Accessed 01 February 2022, https://www.cnmoc.usff.navy.mil/Press-Room/News-Stories/Article/2812666/us-naval-oceanography-participates-in-lse-2021/.

U.S. Naval Meteorology and Oceanography Command Public Affairs Office, 2014: Fleet Numerical Meteorology and Oceanography Center. Accessed 25 January 2022, https://www.fnmoc.navy.mil/home/FNMOC_FACTSHEET_03-14.pdf

U.S. Naval Meteorology and Oceanography Command Public Affairs Office, 2020: Naval oceanography improves the fleet's asymmetric advantage. Accessed 02 February 2022, https://www.cnmoc.usff.navy.mil/Press-Room/News-Stories/Article/2297063/naval-oceanography-improves-the-fleets-asymmetric-advantage/.

U.S. Naval Office of Information, 2016: Naval Oceanography releases Electromagnetic Maneuver Warfare Strategy. Accessed 01 February 2022, https://www.navy.mil/Press-Office/News-Stories/Article/2260634/naval-oceanography-releases-electromagnetic-maneuver-warfare-stategy/.

Wallace, J. M., and P. V. Hobbs, 2006: Introduction and overview. *Atmospheric Science: an Introductory Survey*, 92, Elsevier, 2–5.

Wang, Q., D. P. Alappattu, S. Billingsley, B. Blomquist, R. J. Burkholder, A. J. Christman, E. D. Creegan, T. De Paolo, D.P. Eleuterio, H. J. S. Fernando, and K. B. Franklin, 2018: CASPER: Coupled air–sea processes and electromagnetic ducting research. *Bull. Amer. Meteor. Soc.*, **99, 7**, 1449–1471, https://doi.org/10.1175/BAMS-D-16-0046.1

Wendt, T., 2007: A hierarchical multivariate Bayesian approach to ensemble model output statistics in atmospheric prediction, Naval Postgraduate School, http://hdl.handle.net/10945/56188

Zhou, B., J. Du, and G. DiMego, 2010: Introduction to the NCEP very short range ensemble forecast system (VSREF). *14th Conf. on Aviation, Range, and Aerospace Meteor*., 90th Ameri. Meteor. Soc. Annu. Meet., Atlanta, GA, 17–21. Accessed 23 October 2020.

ECMWF, 2021: ERA5: How to calculate wind speed and wind direction from u and v components of the wind?, Accessed 12 May 2022, https://confluence.ecmwf.int/pages/viewpage.action?pageId=133262398.

THIS PAGE INTENTIONALLY LEFT BLANK

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California