9-2021

Wavelet Methods for Very-short Term Forecasting of functional Time Series

Jared K. Nystrom

Wavelet Methods for Very-short Term Forecasting of Functional Time Series

DISSERTATION

Jared K. Nystrom, Lieutenant Colonel, USA

AFIT-ENS-DS-21-S-050

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-DS-21-S-050

Wavelet Methods for Very-short Term Forecasting of Functional Time Series

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of PhD in Operations Research

Jared K. Nystrom, M.S., M.A., B.S., B.A.

Lieutenant Colonel, USA

June 24, 2021

AFIT-ENS-DS-21-S-050

Wavelet Methods for Very-short Term Forecasting of Functional Time Series

DISSERTATION

Jared K. Nystrom, M.S., M.A., B.S., B.A.
Lieutenant Colonel, USA

Committee Membership:

Raymond R. Hill, Ph.D
Chair

Eric Chicken, Ph.D
Member

Andrew Geyer, Ph.D
Member

Joseph J. Pignatiello Jr., Ph.D
Member

AFIT-ENS-DS-21-S-050

Abstract

Space launch operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) require near-real time determination of lightning risk. Meteorological sensor networks produce data that are often noisy, high volume, and high frequency time series for which traditional forecasting methods are often ill-suited. Current approaches result in significant residual uncertainties and consequentially may result in operational policies that are excessively conservative or inefficient. This work proposes a new methodology of wavelet-enabled semiparametric modeling to develop accurate and timely forecasts robust against chaotic functional data. Wavelet methods are first used to de-noise the weather data, which is then used to estimate a single-index model for forecasting of lightning. This semiparametric technique mitigates noise of the chaotic signal while avoiding any possible distributional misspecification. A screening experiment with augmentations is used to demonstrate how to explore the complex factor space of model parameters, guiding decisions regarding model formulation and gaining insight for follow-on research. Imputation methods are applied on the spatially-based sensor time series making use of the inherent autocorrelation within the data, resulting in improved modeling using machine learning and artificial intelligence techniques. Results indicate a promising technique for operationally relevant lightning prediction from chaotic sensor measurements.

Dedication

For my beautiful family, who have shown me that we can accomplish nearly anything together. I am so proud of you all! We made it.

Acknowledgements

This work would not have been possible without the support and mentorship of many mentors and colleagues. I would like to thank Dr. Raymond Hill for his academic guidance and counsel over the past three years. Your calm demeanor kept me focused on the research and grounded during stressful times.

I would also like to thank my committee members: Dr. Eric Chicken, Dr. Andrew Geyer, and Dr. Joseph J. Pignatiello Jr. Your support and feedback provided a solid foundation for this research. I appreciate your professional guidance, and your comments always proved instrumental in improving my work.

Next, I would like to thank my Masters' advisor Dr. Matthew Robbins. I was offered this amazing opportunity due to your technical guidance and mentorship. I would also like to thank my fellow Army cohort, most especially Dr. Brian Lunday. Your patient counsel and kind words were greatly appreciated in countless situations.

Thank you all for your support and professional advice.

Table of Contents

Page

List of Figures

List of Tables

I. Introduction

Weather operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) are complicated by unique requirements for near-real time determination of risk from lightning. KSC/CCSFS experiences one of the world's highest incidence of lightning, which impacts both the launch of space vehicles and daily support activity. Accurate lightning forecasts are essential for safe flight line operations through the prediction of lightning onset and the cessation of lightning events following a storm. The accuracy of these forecasts is complicated by sensor data that is both inherently noisy and collected in time series. KSC/CCSFS weather policy literature suggests current methodologies are far too conservative in nature, resulting in widespread operational inefficiencies. This study proposes a method to apply discrete wavelet transformations and semi-parametric single index model to time series weather sensor data to improve the timeliness and accuracy of lightning prediction for KSC/CCSFS.

Cape Canaveral possesses a dense array of weather sensors that includes both traditional sensors and tailor-made systems such as the electric field mill (EFM) network and lightning detection and ranging (LDAR) system. Weather forecasters also use traditional weather measurements, a local weather radar (WSR-88D), National Lightning Detection Network (NLDN), and daily weather balloon launches. These sensor networks inform an operational warning system that manages ten warning regions spread throughout Cape Canaveral. These warning regions consist of 5NM or 6NM circles centered on key infrastructure locations, sometimes heavily overlapping [81].

The data produced by these sensor networks is inherently noisy and inappropriate for standard modeling approaches due to the complexity of movement of atmospheric electrostatic potential [44]. EFM data collection is further perturbed by dense networks of antennas, radar arrays, and other equipment and facilities supporting space launch and communication. There have been attempts made to mitigate these disturbances, such as ceasing collection of an individual EFM sensor if maintenance crews are mowing grass in the area, but these disturbances remain.

The LDAR detects both radar and flashes emitted by lightning to produce a 3D map of all lightning events within 54NM of Cape Canaveral [87]. The system was originally designed by NASA to meet their unique operational requirements that includes the ability to detect total lightning. The system has above a 90% correct detection rate out to 54NM, increasing to over 99% within 14NM of Cape Canaveral [87]. LDAR data contains timestamps for all detected lightning events, to include a detection range and azimuth from the system's central tower. The LDAR data are used in this study as the response for model training and evaluation.

Predictive models such as linear regression are prevalent due to their ease of interpretation; however, they require certain assumptions to be made concerning underlying relationships within the data. These assumptions may cause a model to over-smooth a predicted response, resulting in a failure to capture a significant event that is of most interest to research. Furthermore, these models are not the best suited to time series data and can not mitigate statistical noise. The proposed approach employs DWT as a computationally efficient method to transform a meteorological time series for accurate modeling while simultaneously reducing observed noise. Additionally, semi-parametric models are used to capture complex phenomena observed in meteorological events without assumptions of the underlying data.

Wavelet transformations are a relatively new method that allows re-expression of

data from the time domain into a frequency domain in an very computational efficient manner. These methods facilitate accurate modeling of a complex response in time, producing analysis of both frequency and time content simultaneously. Motivated by the Fourier transform, the DWT consists of a linear transformation that reduces a complex response to a single vector of coefficients. An inverse DWT (IDWT) can then be applied to perfectly reproduce the original data. This method allows manipulation to remove noise or extraneous data, with common applications in signal analysis, data compression, and image analysis. This method is especially useful in evaluation of time series as it allows analysis without auto-correlation that would otherwise cause overestimation of a response.

Semi-parametric models, such as a single index model or generalized linear model, are methods that bridge the capabilities of parametric and non-parametric models. Non-parametric methods approximate a function strictly using the data and without any required assumptions; however, these models often fail to converge for higher dimension problems. A semi-parametric model makes some basic assumptions of linearity to accommodate high dimensional problems while maintaining some beneficial properties of non-parametric models. The semi-parametric single-index model (SIM) is a generalization of many parametric models to include Normal regression, Logit, Probit, and Tobit. Similar to these methods, the SIM models the relationship between a response and predictive variables but without any distributional assumptions.

1.1 Problem Statement

Senior leaders require forecasting models that provide the timeliness and accuracy required to effectively informs critical decisions. Modern systems produce data that is high volume, high frequency time series, and of differing data types that traditional time series forecasting methods are ill-suited to address. This research identifies,

evaluates, and applies a methodology for accurate and timely operational forecasting derived from complex and noisy time series data.

1.2 Summary

Chapter II presents a survey of wavelet methods for time series analysis and forecasting, to include an examination of novel wavelet techniques from three disparate fields developed to address unique requirements. These techniques offer powerful techniques for pre-processing time series by de-noising or feature extraction, thus facilitating greatly improved model estimation and performance in artificial intelligence and machine learning applications. Unlike other filtering methods, such as the Fourier transform and exponential smoothing, wavelets offer an efficient method to model a function in terms of time and frequency simultaneously. This facilitates improved de-noising and feature extraction techniques.

Chapter III proposes a forecasting methodology using wavelet decomposition of chaotic weather sensor time series and semiparametric single-index models to mitigate the chaotic signal and any possible distributional misspecification. Space launch operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) are complicated by unique requirements for near-real time determination of risk from lightning. Weather sensor networks for lightning forecasting produce data that are noisy, high volume, and high frequency time series for which traditional forecasting methods are often ill-suited. Current approaches result in significant residual uncertainties and consequentially may result in forecasting operational policies that are excessively conservative or inefficient. A screening experiment with augmentations is used to demonstrate how to explore the complex factor space of model parameters, guiding decisions regarding model formulation and gaining insight for follow-on research. Results indicate a promising technique for operationally relevant lightning

prediction from chaotic sensor measurements.

Chapter IV employs a spatiotemporal imputation technique that simultaneously accounts for autocorrelation between spatially correlated measurements collected as a time series. Wavelet methods are used as an additional pre-processing step, serving to de-noise the chaotic EFM measurements to allow faster convergence and estimation of spatiotemporal models. Instead of a purely time series or spatial model, spacetime approaches use all available data to infer predicted values. These methods prove highly useful in situations in which large amounts of a particular time series are missing and need to be estimated. Although complex in application, such methods are of increasing importance due to the increasing prevalence of modern sensor systems. Results indicate significant improvements upon the previous wavelet-enabled single-index model.

1.3 Contributions

The literature review, survey of wavelet methods, advances in methods and techniques developed by this body of work contribute to the general field of Operations Research, specifically to both meteorological forecasting and military and security operations research. The survey of wavelet methods (Chapter II) provides the first cross-functional analysis of current applications in three disparate fields using predictive models. The paper provides a discussion of wavelet theory, to include a concise presentation of the application of wavelets in time series applications. The paper also provides an overview of current applications of wavelets in machine learning and artificial intelligence applications to present general application methods and best practices. This includes an overview of applications in wind speed prediction, very short-term prediction of earthquake magnitude, and traffic congestion prediction.

A new method of wavelet-enabled semiparametric modeling builds upon exist-

ing literature, improving both the safety and efficiency of flight line operations at KSC/CCSFS (Chapter III). This work identifies and evaluates a wavelet-enabled semiparametric single-index modeling approach for lightning warning derived using chaotic time series data at KSC/CCSFS. This approach develops forecasts designed to meet operational requirements for timeliness and accuracy from a data source previously considered too noisy for successful use in machine learning and artificial intelligence applications. The novel application of a designed experiment is used to aide in model formulation, guiding the selection of model parameters to ensure the best possible forecast.

Building upon the initial model formulation, spatiotemporal kriging is applied as an imputation method to further improve model performance (Chapter IV). This approach accounts for both spatial and temporal autocorrelation within the EFM data to estimate values missing due to sensor maintenance, interference, or technical malfunctions. Most applications using machine learning are not robust to missing values, and poorly estimated values from competing imputation methods could perturb any modeling forecast. Spatiotemporal kriging proves to be a powerful method for completing the EFM dataset for machine learning and artificial intelligence applications. In this particular implementation, the model developed a forecast with over 95% accuracy using the EFM data with imputed estimates.

II. Wavelet Methods for Pre-processing Time Series for Forecasting using Artificial Intelligence and Machine Learning

Accurate and timely time series forecasts have become increasingly important for short-term weather forecasting. However, parameter estimation and interpretation in such models has become particularly difficult due to the high volume of data produced by modern meteorological sensors. Wavelet methods offer powerful techniques for pre-processing time series by de-noising or feature extraction, thus facilitating greatly improved model estimation and performance in artificial intelligence and machine learning applications. Unlike other filtering methods, such as the Fourier transform and exponential smoothing, wavelets offer an efficient method to model a function in terms of time and frequency simultaneously. This facilitates improved de-noising and feature extraction techniques. This paper presents a survey of wavelet methods for time series analysis and forecasting, to include an examination of novel wavelet techniques from three disparate fields developed to address unique requirements. [1]

2.1 Introduction

Accurate and timely time series forecasts are becoming increasingly important; however, estimation of such models are likewise becoming increasingly difficult due to the high volume of data produced by modern meteorological sensor networks. These complex systems collect data that can be noisy, high volume, and high frequency time series. Developing a forecast with traditional time series analysis using this type of data may be inappropriate as parametric assumptions may not hold. Additionally, parametric assumptions may over-smooth the response and lose the signal of interest within the noise. The result is a model with a high degree of residual uncertainty

[1] Paper submitted to the journal Weather and Forecasting.

that forces decision makers to implement policies that are excessively conservative or inefficient.

Extensive current literature points to the power of these methods for time series analysis; however, each modeling approach has limitations and includes specific requirements for full specification. The Box-Jenkins methodology models a time series using polynomials, which can sometimes over-smooth a particularly abrupt response. Furthermore, time series must be a stationary process with constant variance for full specification. Although this condition can be met in many industrial processes, it can be too strict an assumption for situations with a chaotic response. For instance, some sensor networks produce high frequency, high dimensional datasets often collected as noisy and non-stationary time series. The artifacts of interest within these series frequently consist of sharp and abrupt changes that traditional modeling applications may fail to accurately capture. Section 2.2 includes a concise presentation of the Fourier transform, which is a powerful filtering method but lacks the ability to model a function in terms of both frequency and time. Wavelet methods are being increasingly used in such situations due to their ability to model abrupt change in a computationally efficient manner without the requirement for a stationary time series. Early works, such as Lau and Weng [45], point to the power of wavelet techniques in meteorological time series analysis. These methods have only developed into more powerful tools, especially with the growth of machine learning and artificial intelligence. Wavelet methods are being employed as a preprocessing method, either for de-noising and smoothing a time series or serving as a feature selection method. Wavelet techniques offer new avenues of analysis overcoming some of the the limitations of traditional approaches.

This work is a cross-functional analysis of current applications in three disparate fields of wavelet methods in predictive models. These methods are common across

applications using machine learning and artificial intelligence, however the author is unaware of any attempts to survey these methods to investigate best practices for use in weather forecasting. The paper is organized as follows: Section 2.2 provides background to wavelet methods and Section 2.3 provides a brief overview of generalized applications methods. Section 2.4 provides an overview of current applications of wavelets in machine learning and artificial intelligence applications. This includes an overview of applications in wind speed prediction, very short-term prediction of earthquake magnitude, and traffic congestion prediction. Section 2.5 provides analysis into assessed best practices, assessed weaknesses of wavelet methods, and areas for additional research.

2.2 Wavelet Theory

Wavelets model a function in time and frequency simultaneously by approximating functions at increasing levels of resolution expressed as a linear combination of scaling functions $\phi_{j,k}$ combined with the difference in approximations expressed as a linear combination of wavelets $\psi_{j,k}$ [69]. This is accomplished by projecting approximations of that function into a series of nested subspaces, each of which provide a different level of resolution in time. Wavelet functions represent a family of unique functions designed to be localized in time and frequency, typically defined as a mother wavelet (ψ) and father wavelet (ϕ). Through dilation and translation operations, these wavelets produce an entire basis of wavelet functions [69]. These basis functions can be used to model a function in a Multiresolution Analysis (MRA) which consists of successively detailed approximations of the function. Wavelets provide significant advantages over competing methods, namely the discrete Fourier transform and windowed Fourier transform, to localize frequency in time by adapting the size of their window of approximation to the frequency at each resolution level [69]. The

result is a time to resolution level analysis method that optimizes the tradeoff between certainties in frequency and time across each nested and consecutive resolution level.

2.2.1 Properties of Wavelets

A wavelet is a small wave that grows and decays in a relatively limited time. Percival and Walden [73] define wavelets as real-value functions $\psi(\cdot)$ over the real axis $(-\infty, \infty)$ that satisfy the following two properties:

1. The function $\psi(\cdot)$ integrates to zero,

$$\int_{-\infty}^{\infty} \psi(u)du = 0. \tag{1}$$

2. The square of $\psi(\cdot)$ integrates to unity,

$$\int_{-\infty}^{\infty} \psi^2(u)du = 1. \tag{2}$$

Equation 1 forces $\psi(\cdot)$ into a wave shape, where any non-zero activity must be mirrored in an integral equivalent non-zero activity of opposite sign. Equation 2 forces non-zero activity and ensures that the function can be used to form an orthonormal basis within $L^2(\mathbb{R})$. The space $L^2(\mathbb{R})$ forms a Hilbert space of square integrable functions with a defined inner product where

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \middle| \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \right\}.$$

Furthermore, for a given ϵ where $0 < \epsilon < 1$ there exists an interval $[-T, T]$ of finite length such that

$$\int_{-T}^{T} \psi^2(u)du > 1 - \epsilon. \tag{3}$$

As ϵ approaches zero, $\psi(\cdot)$ can only deviate insignificantly outside of $[-T, T]$. The non-zero activity of $\psi(\cdot)$ is considered small and the interval $[-T, T]$ is insignificant compared to the real number line, resulting in the formation of a little wave [73]. The consequence of this is a function whose dilations and translations are localized in both time and frequency that are capable of serving as basis functions.

2.2.2 Wavelet Features

The following section overviews how wavelets approximate functions and is derived from [69] unless otherwise specified.

The oldest and most basic example of a wavelet was developed by [25] and is given by

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \cdot \tag{4}$$

As defined, $\psi(x)$ is known as the *mother wavelet* for the Haar system. The dilation operation compresses or stretches the wavelet, while the translation moves it back and forth in time. These operations manipulate a wavelet to best approximate a function. Letting j represent the dilation index and k represent the translation index, the mother wavelet is then defined as

$$gt\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \quad j, k \in \mathbb{Z}. \tag{5}$$

These dilation and translation operations allow any arbitrary function $f \in L^2(\mathbb{R})$ to be reasonably approximated with linear combinations of the mother wavelet, $\psi_{j,k}$. An example of the approximation of a function is shown in Figure 1, the presentation of which was motivated by Percival and Walden [73]. Piecewise constant functions are

used to approximate a function with increasing levels of dilation j. The approximation of the function improves as j increases.

The piecewise continuous nature of the Haar function results in a blocky representation of the signal, however works of Daubechies [13] and others provide much more elegant wavelets. Figure 2 provides a representation of two such wavelets compared to the Haar. Wavelet basis function selection is made by application and based upon a particular wavelet's ability to model a function.

2.2.3 Multiresolution Analysis

MRA is one of the most important consequences of basic wavelet mechanics. The basic principles of MRA state that an approximation of a function can be accomplished through an additive decomposition: an approximation f^j at resolution level j can be decomposed into a coarser approximation f^{j-1} and a detail function g^{j-1} at level $j-1$. Mallat [59] first identified the properties required for a sequence of subspaces to result in a wavelet system. Frazier [20] defines the properties of a MRA with a sequence of functional spaces $(V_j)_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ as follows:

1. *Monotonicity.* The sequence of subspaces is increasing for all $j \in \mathbb{Z}$ where

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$$

2. *Scaling function.* There exists a scaling function $\phi \in V_0$ with resolution level factor j and shift factor k defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \quad j,k \in \mathbb{Z}. \tag{6}$$

This function is commonly referred to as the father wavelet [69]. This function scales through dilation and scaling operations, providing orthonormal bases for

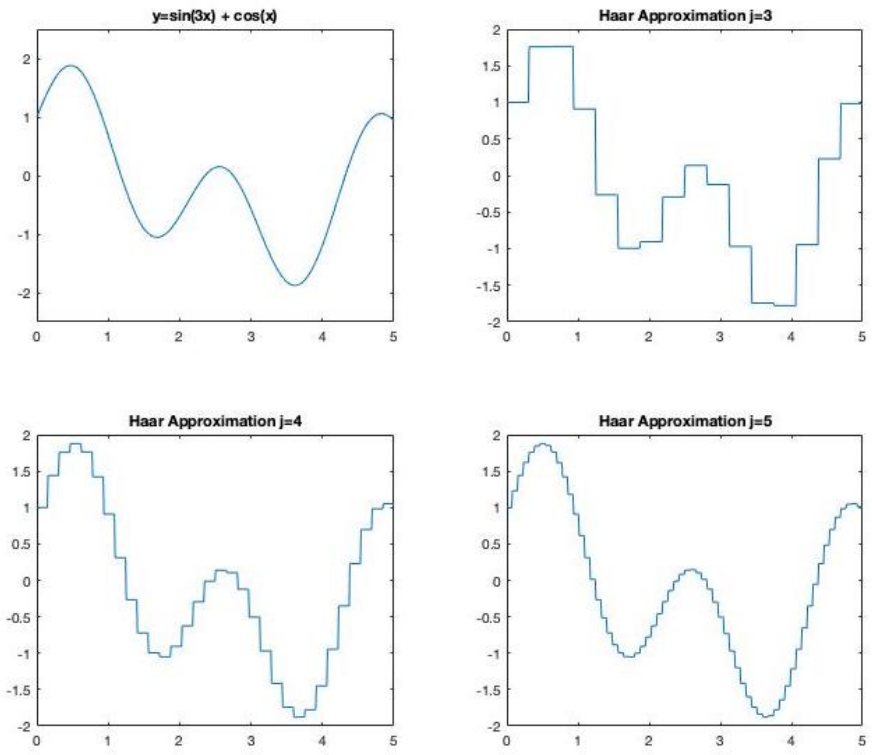


Figure 1: Piecewise constant approximations of a function (top left) with increasing levels of dilation j. The approximation by the piecewise continuous Haar function starts blocky (top right) but becomes smoother at higher levels of dilation (bottom right).

V_0 and all j resolution levels of this subspace or

$$V_j = \text{span}\{\phi_{j,k}, k \in \mathbb{Z}\} \quad j \in \mathbb{Z}.$$

3. *Dilation property.* $f \in V_j$ if and only if $f(2\cdot) \in V_{j+1}$. This implies every subspace is a scaled version of the original space V_0.

4. *Trivial intersection property.* $\cap_{j\in\mathbb{Z}} V_j = \{0\}$

5. *Density.* $\cup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$, or for any $f \in L^2(\mathbb{R})$ there exists a sequence $\{f_n\}_{n=1}^{\infty}$ such that each $f_n \in \cup_{j\in\mathbb{Z}} V_j$ and $\{f_n\}_{n=1}^{\infty}$ converges to f in $L^2(\mathbb{R})$.

With these properties we can now fully articulate the mechanics of the MRA. A

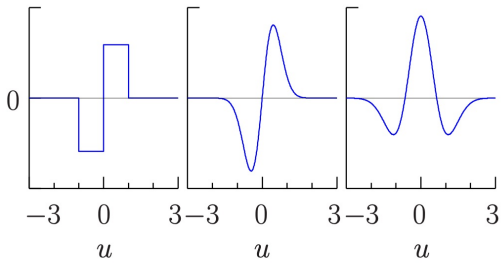


Figure 2: Three examples of mother wavelets (ψ). From left to right, the Haar wavelet; a wavelet related to the first derivative of the Gaussian probability density function (pdf); and the Mexican hat wavelet related to the second derivative of the Gaussian PDF [73].

"detail space" is defined based upon the mutually orthogonal property of wavelets, for wavelets of the same dilation index j where

$$W_j = \text{span}\{\psi_{j,k}, k \in \mathbb{Z}\}.$$

A consequence of the monotonicity of the sequence of subspaces is that for certain choices of ψ and ϕ, the scaling function $\phi_{j,k}$ is orthogonal to a wavelet $\psi_{j',k'}$ whenever $j \leq j'$. This implies that for an MRA

$$V_j = V_{j-1} \oplus W_{j-1} \tag{7}$$

where $\oplus$ denotes the direct sum of subspaces V_{j-1} and W_{j-1}, where $V_j = V_{j-1} + W_{j-1}$ and $V_{j-1} \cap W_{j-1} = \{0\}$. Extending this recursively results in

$$V_j = V_{j-2} \oplus W_{j-2} \oplus W_{j-1}$$

and thus

$$V_j = V_{j_0} \oplus \bigoplus_{\ell=j_0}^{j-1} W_\ell.$$

Wavelet Transform	Benefits	Constraints
CWT	-Beneficial for data exploration -Produces 2D image of 1D signal	-Highly Redundant -Difficult to apply in hybrid models
DWT	-Time/scale decomposition -Efficient computation ($\mathcal{O}(N)$) -Succinct representation of coefficients	-Requires sample size of dyadic length -Shift variant filter; does not align with original time series - Resolution scales with level -Assumes periodicity; boundary effects
MODWT	-Shift invariant; stationary representation -Well defined for all sample sizes -Provides high resolution at every resolution level -Does not assume periodicity in data	-Highly redundant -Slower computation($\mathcal{O}(N \log_2 N)$) -Each resolution level results in a vector length N
DWPT	-Time/frequency decomposition -Mimics DFT on intervals of time -Succinct representation of coefficients -MODWPT method available	-Assumes periodicity; boundary effects -Downsampling results in coarser approximations at high level

Table 1: Summary of wavelet decomposition methods for time series analysis

This result gives a key insight to wavelet analysis; a function can be approximated at increasing levels of resolution expressed as a linear combination of scaling functions $\phi_{j,k}$ combined with the difference in approximations expressed as a linear combination of wavelets $\psi_{j,k}$, all accomplished through the use of translation and dilation operations [69].

2.2.4 The Wavelet Transform

This section briefly introduces various wavelet transforms being used in current time series analysis applications. This presentation includes an assessment of relative strengths and weaknesses of particular approaches to complement later discussion of current literature. The wavelet transform is well documented, and particularly helpful and in-depth presentations are found in Percival and Walden [73] and Ogden [69]. Table 1 provides a concise analysis of the benefits and constraints for the wavelet transforms presented later in this section. Wavelet methods require the selection of one of these transforms based upon capability tradeoffs within each application.

2.2.5 Continuous Wavelet Transform

Equation 5 defines the translations and dilations of the mother wavelet for integer values of j and k. Relaxing the restrictions on the indices and allowing them to take

continuous values results in the Continuous Wavelet Transform (CWT). Ogden [69] defines the continuous mother wavelet for $a > 0, b \in \mathbb{R}$ as

$$\psi_{(a,b)}(x) = a^{-1/2}\psi\left(\frac{x-b}{a}\right) \tag{8}$$

and the CWT defined for any $f \in L^2(\mathbb{R})$ as

$$(\mathcal{W}_\psi f)(a,b) = a^{-1/2}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt. \tag{9}$$

If $\psi_{(a,b)}$ is assumed to be a suitable window function, then the CWT provides information about a signal in the time domain centered at b with radius $a\Delta_\psi$ [69]. This results in a window defined as

$$(b - a\Delta_\psi, b + a\Delta_\psi)$$

where the size of the window scales relative to its continuous dilation index a. Unlike the rigid windowing of the Discrete Fourier Transform (DFT), wavelet methods scale windows automatically to frequency. This unique property of wavelets optimizes the approximation of a function in frequency and time simultaneously.

2.2.6 Discrete Wavelet Transform

These results naturally lead to the introduction of the discrete wavelet transform (DWT). The DWT can be applied to the additive decomposition of a time series into constituent detailed time series ($\psi_{j,k}$) reflecting variations at resolution level j and a smoothed version of the time series ($\phi_{j,k}$) reflecting averages at resolution level j [73].

Therefore with wavelets defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{10}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{11}$$

a time series can be represented as

$$f(t) = \sum_k c_{j_0,k}\phi_{j_o,k} + \sum_j \sum_k d_{j,k}\psi_{j,k} \tag{12}$$

where $c_{j,k} = \langle f, \phi_{j,k} \rangle$, $d_{j,k} = \langle f, \psi_{j,k} \rangle$, and $j, k \in \mathbb{Z}$. The time series is thus represented as a linear combination of the shifted and scaled versions of the wavelet functions as estimated using the wavelet coefficients $c_{j,k}$ and $d_{j,k}$. An important consequence of equation 34 is the separation of the approximation and detailed representations of a signal.

Figure 15a, motivated by and adapted from presentations in the MATLAB Wavelet Toolbox [61], provides a rudimentary representation of a three-level, $j = 3$, DWT of a signal X, where $X \in \mathbb{R}^N$. The levels D_1, D_2, and D_3 represent the detailed resolution levels whereas S_3 is representative of the smoothed approximation of the function. The decomposition results in a concatenation of these resolution levels into a single vector of wavelet coefficients $W \in \mathbb{R}^N$ the length of the original sample.

In practice, execution of this transform is accomplished through a filter bank approach. This approach processes a signal using decimation or downsampling by two, where every other value of the signal is removed. This reduces the size of the signal by half at every level of decomposition. This results in a quick and highly efficient algorithm as every iteration requires half the number of calculations. The inverse implementation requires a similar filter bank approach governed by upsampling, or doubling the size of the sample by inserting zeros between every value. However, this

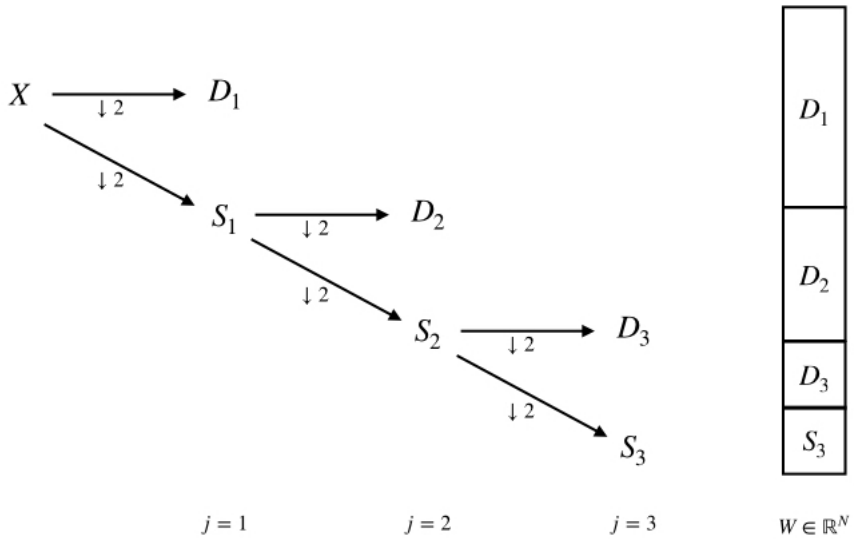


Figure 3: Depiction of three-level wavelet decomposition of signal X to wavelet coefficients W with decimation

approach suffers from some limitations and weaknesses.

1. The filter bank method of the DWT requires a signal sample size of dyadic length, or an integer multiple of 2^j.

2. The DWT is not shift invariant, meaning the values of the details and smooths do not shift with the values of the original signal. The result is that the inverse DWT can give a different reconstruction compared to the original time series even when accounting for the shifts.

3. The DWT requires a periodicity assumption in the signal. For non-stationary time series, this means that the DWT transform is highly dependent upon when the time series is sampled. Significant changes in the time series across the sample will result in significant boundary effects.

2.2.7 Maximal Overlap Discrete Wavelet Transform

The maximal overlap discrete wavelet transform (MODWT) is a modified version of the DWT better suited for certain applications, such as time series analysis. This particular transform is found throughout wavelet literature under different names, such as undecimated DWT, shift invariant DWT, wavelet frames, translation invariant DWT, stationary DWT, time invariant DWT, and non-decimated DWT [73]. This research adopts the use of MODWT as in Percival and Walden [73] due to their thorough and foundational work in applying wavelets to time series. Essentially, the MODWT does not include downsampling as in the DWT and thus uses all values of the original signal at every level of decomposition.

The use of the MODWT provides the following key advantages over the DWT.

1. The MODWT is well defined for all sample sizes, unlike the decimated DWT that requires a sample of dyadic length.

2. The MODWT is shift invariant, meaning each level of decomposed coefficients aligns with the original time series. The MODWT also avoids boundary effects found in the decimated wavelet transforms.

3. The MODWT does not down sample at each level, meaning each resolution level contains the same number of coefficients as the original sample. This produces a redundant but higher resolution at coarser levels compared to the decimated wavelet transforms.

These advantages are not without costs. A notable cost is that the transform is highly redundant and loses orthogonality. This results in dependencies between the empirical coefficients of the scaling function and wavelets. The details and smooth resolution level of the MODWT each contain the same number of samples as the original signal. Although this gives a finer resolution at each level, it results in the

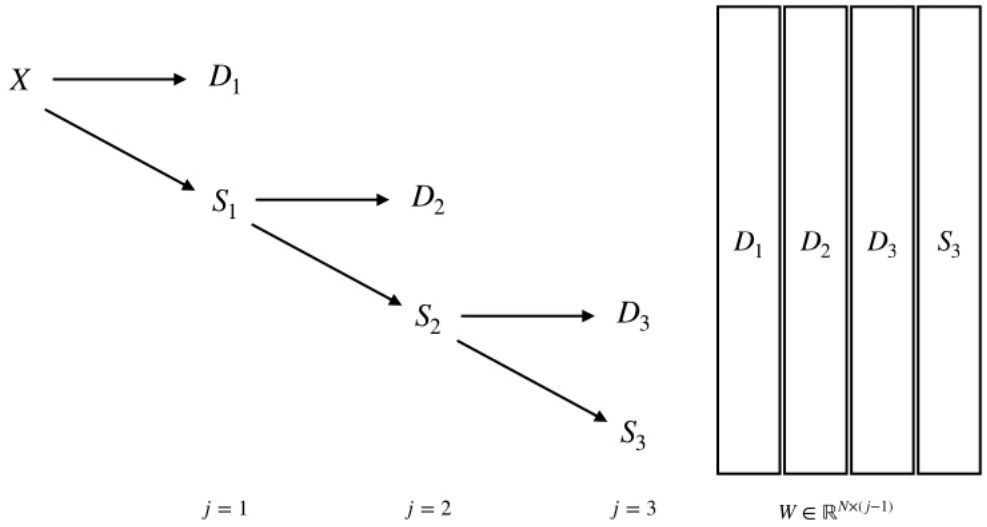


Figure 4: Depiction of three-level MOWDT decomposition of signal X to wavelet coefficients W.

number of required computations $\mathcal{O}N \log_2 N$ or a cost of $\mathcal{O} \log_2 N$ when compared to the DWT.

The MODWT can be analyzed using a MRA just as in the DWT. Figures 5 and 6 provide examples of this process using four-level wavelet MRAs, utilizing the "la8" Daubechies wavelet filter, of an ECG signal as presented in Percival and Walden [73]. The ECG sample (top panels) is obtained nasally from a patient who occasionally experiences arrhythmia. The transform is clearly shift invariant, as the spikes in the details $(D_1, \ldots, D_4)$ align perfectly with the sharp and abrupt changes in the original data (top panel). Unlike the progressively coarse levels of the DWT, the resolution in the higher levels of detail remain the same as the sample size for each level is identical.

2.2.8 Discrete Wavelet Packet Transform

The discrete wavelet packet transform (DWPT) decomposes every resolution level of coefficients, resulting in a tree-like decomposition of the original signal. The DWPT does mimic the DFT somewhat and results in a time/frequency decompo-

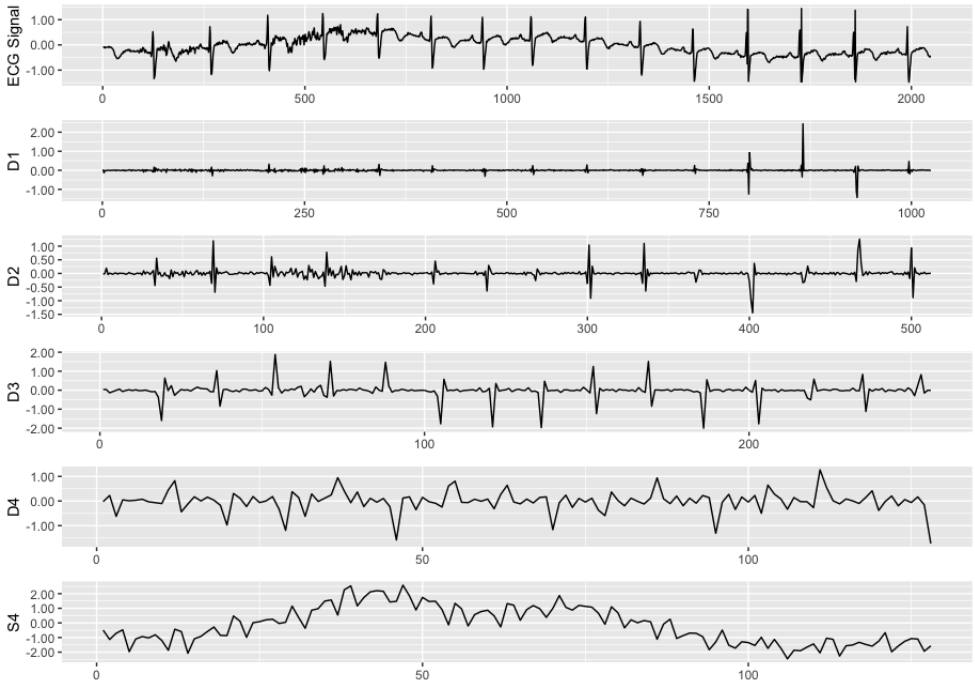


Figure 5: The shift variant DWT where movement of coefficients not necessarily aligned across resolution levels using four-level wavelet MRAs, utilizing the "la8" Daubechies wavelet filter, of an ECG signal as presented in Percival and Walden [73].

sition, whereas earlier methods resulted in a time/scale representation. Unlike the DWT, where each detail resolution level is preserved and not reanalyzed, each level is subsequently decomposed. As such, a representation such as those seen in Figures 5 and 6 are not directly applicable. Similar to the DWT, this method uses downsampling and results in a non-redundant representation. DWPT is used as it can provide a more detailed analysis of a signal compared to DWT.

Figure 7 depicts the DWPT of a signal X, resulting in a tree-like decomposition. The resulting vector of wavelet coefficients W is the same length of the original sample, and represents the original sample at scale and at different frequency sub-bands.

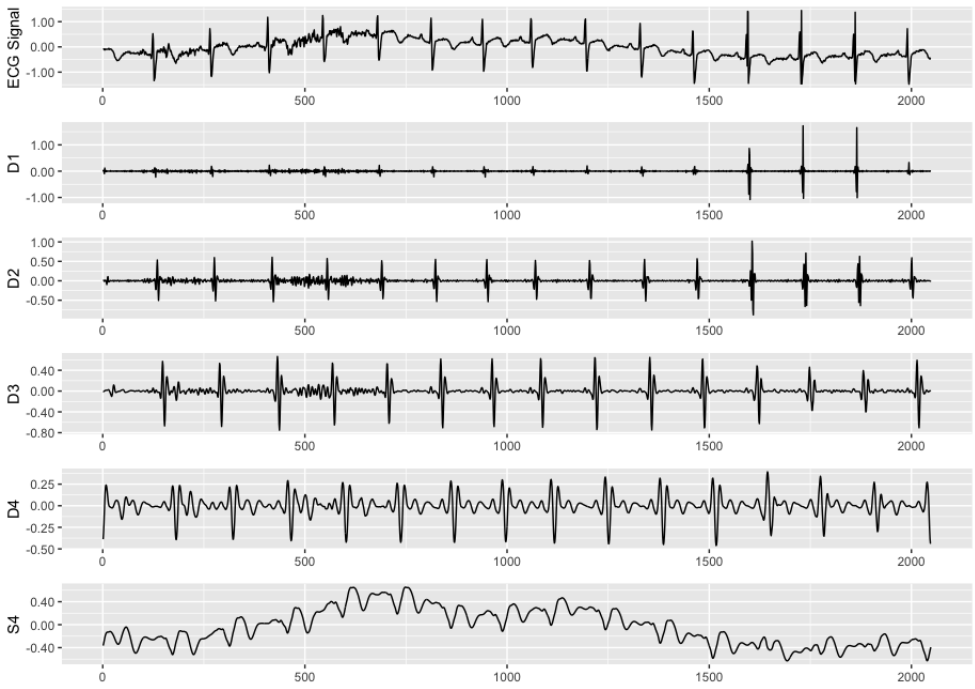


Figure 6: The shift invariant MODWT. Movement of coefficients align across resolution levels using four-level wavelet MRAs, utilizing the "la8" Daubechies wavelet filter, of an ECG signal as presented in Percival and Walden [73].

2.2.9 Wavelet Thresholding

Wavelet thresholding is a dimension reduction and de-noising method that manipulates the transformed wavelet coefficients. This section introduces thresholding using a brief discussion on the sparsity of the wavelet representation, followed by both universal and adaptive thresholding techniques.

2.2.10 Sparsity of Effects

The wavelet transformation results in a sparsity of effects, where most of the key features of a signal are captured and represented by only a few coefficients. Figure 8 depicts the sorted values of the first four levels of details for both a DWT and MODWT from Figure 6. It is readily apparent that most of the coefficient values are near zero for both of the transforms. However, the redundancies of the MODWT

result in a less sparse representation of the signal, effectively increasing the number of significant coefficients that describe the power of the signal. This denser representation becomes more pronounced at higher levels of detail.

The computational efficiency of the DWT is now apparent due to the sparse representation. The MODWT provides a finer resolution at each level of decomposition at the cost of a much denser, redundant representation.

When an observed signal is contaminated with stochastic noise, then an additional consequence of the sparsity seen in Figure 8 where the noise in the signal is concentrated in smaller valued nonzero coefficients. These coefficients can be manipulated to reduce or remove stochastic noise while the power of the true signal is retained in only a few significant coefficients. Therefore, the ability of wavelet methods to model a signal in frequency and time simultaneously grants a powerful ability to capture and isolate signals of random noise. Manipulation of the coefficients to reduce or remove random noise is known as thresholding, which can be applied globally to the entire set of coefficients or adaptively applied using localized rules. Unless otherwise stated, thresholding methods require the assumption of normally distributed observational errors.

2.2.11 Global Thresholding

Global thresholding uses a single threshold value λ applied uniformly to all or nearly all coefficients of the wavelet transform. Consider for a given threshold value λ, then

$$\hat{f}_\lambda(t) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} I_{\{|d_{j,k}^{(n)}|>\lambda\}} d_{j,k}^{(n)} \psi_{j,k}(t) \tag{13}$$

where I represents the indicator function [69]. This representation of "keep or kill" is known as hard thresholding, where any value less than or equal to the given value of λ is set to zero. This enforces sparsity in the wavelet coefficients, resulting in

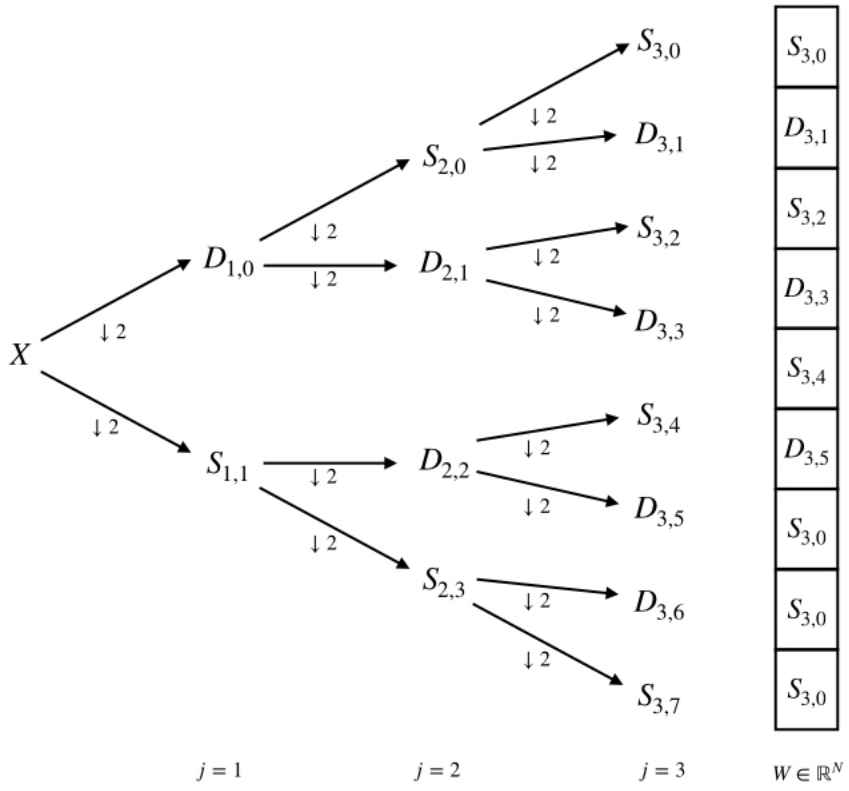


Figure 7: Depiction of three-level DWPT decomposition of signal X to wavelet coefficients W with decimation.

maintaining only those coefficients significant for representing the original signal. An inverse wavelet transform can then be applied to recreate the original signal with random noise removed. Then, defining the thresholded coefficients as

$$\hat{\theta}_{j,k} = \delta_\lambda(\tilde{\theta}_{j,k}) \tag{14}$$

allows for reexpression of the hard thresholding rules as

$$\delta_\lambda^H = \begin{cases} x & \text{if } |x| > \lambda \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

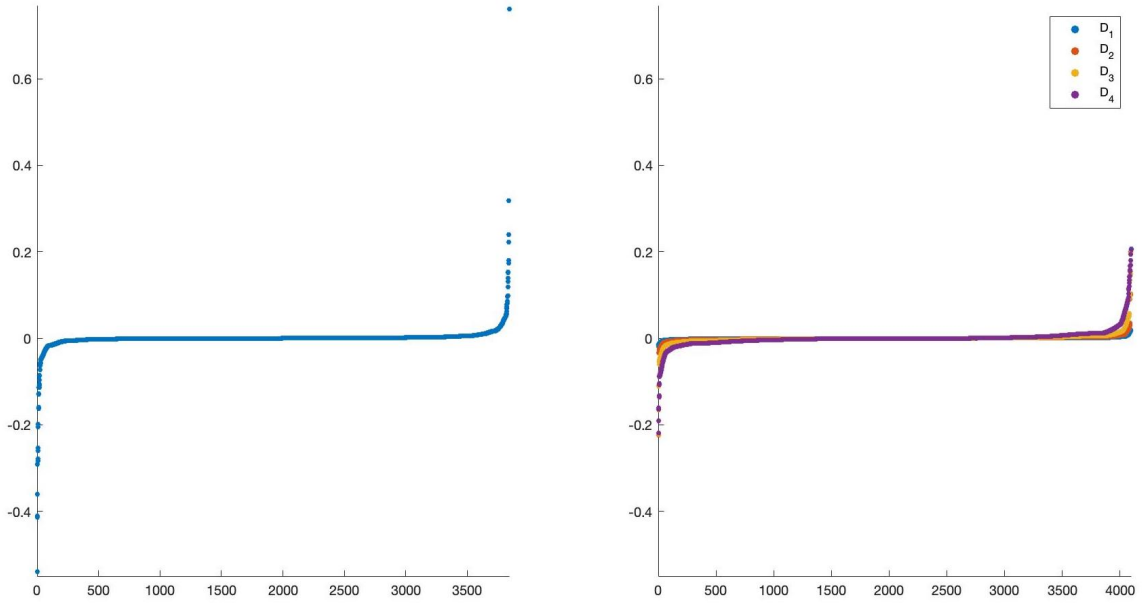


Figure 8: Wavelet coefficients of four detail resolution levels, combined and sorted by value, of DWT (left) and four detail resolution levels MODWT (right) sorted individually by value using Haar wavelet from the ill-behaved time series in Figure 6.

Donoho and Johnstone [16] propose an alternative method of soft thresholding defined as

$$\delta_\lambda^S = \begin{cases} x - \lambda & \text{if } x > \lambda \\ 0 & \text{if } |x| \leq \lambda \\ x + \lambda & \text{if } x < -\lambda \end{cases} . \tag{16}$$

Similar to hard thresholding, only wavelet coefficients greater than a threshold are kept, however their value is shrunk closer towards zero by an amount equal to the threshold λ [69].

These two methods are widely applied in current applications as a dimension reduction method. However proper choice of the threshold value remains subjective, based upon an assessed tradeoff between over- and under-smoothing the function. Furthermore, these universal methods may underperform adaptive techniques in large sample sizes. Donoho and Johnstone [16] propose two universal thresholds, the first

of which is

$$\delta = \sqrt{2\sigma^2 \log(N)} \tag{17}$$

to be used when the variance of the original signal (σ^2) is known. This method, commonly referred to as VisuShrink, is a computationally efficient method that can be applied through either soft or hard thresholding techniques. In application, when σ^2 is frequently unknown, Percival and Walden [73] recommend the use of the median absolute deviation standard

$$\hat{\sigma}_{(mad)} = \frac{\text{median}\{|W_{1,0}|, |W_{1,1}|, \ldots, |W_{1,\frac{N}{2}-1}|\}}{0.6745} \tag{18}$$

using the $N/2$ values of the first details level of decomposition. Donoho and Johnstone [16] also introduce minimax thresholding, where the threshold value is numerically calculated based upon sample size N.

As an example of thresholding, Figure 17 displays annual Nile River minima measured from 622-1284 A.D. [73]. The raw time series is in blue, and a reconstructed time series following a MODWT and soft thresholding is in red. Thresholding the function has effectively smoothed the response, an action that may allow easier interpretation and implementation into a hybrid model that requires convergence. However, oversmoothing this function could eliminate some sharp and abrupt changes that may be the signal of interest. Careful application of thresholding is required dependent upon application.

2.2.12 Data Adaptive Thresholding

Data adaptive techniques attempt to improve upon global techniques by varying the threshold within the decomposition. Donoho and Johnstone [15] present SureShrink as an extension of VisuShrink, combining a level-dependent thresholding

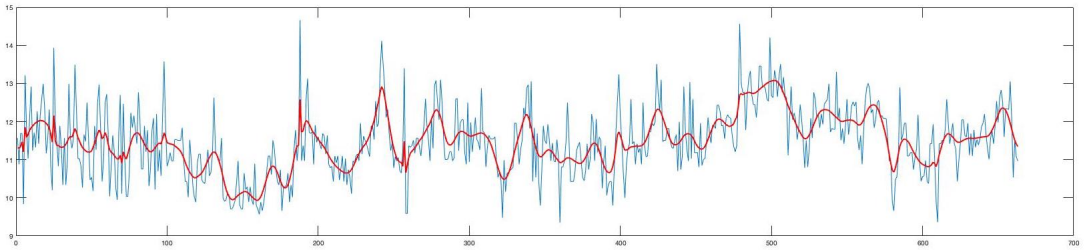


Figure 9: Annual Nile River minima 622-1284 A.D. (blue) [73] and values of wavelet approximated smoothed function (red).

technique with Stein's unbiased risk estimator (SURE) [88]. This method is more computationally demanding compared to VisuShrink, but is shown to reduce the mean squared error in estimation.

Cai [9] introduces block thresholding, where a thresholding δ is determined using groups of coefficients within a resolution level. An important note is that both VisuShrink and SureShrink assume normality of the errors. McGinnity et al. [63] propose a nonparametric method of block thresholding that does not require this normality assumption.

2.3 General Approaches for Wavelet Methods in Forecasting

Wavelet methods offer a powerful approach to decompose a time series; however, these techniques must be implemented in conjunction with other methods to estimate a predicted response. This section provides an overview of such methods, generalized into three distinct approaches: data-preprocessing, forecasting in the wavelet domain, and hybrid models. A more detailed examination of these methods follows in Section 2.4 through a review of current applications across three disparate fields.

2.3.1 Data Preprocessing

The multiresolution analysis method provided by wavelet techniques offers a computationally efficient approach for data preprocessing of a noisy time series. Resolu-

tion levels of wavelet coefficients associated with random noise can be either removed or manipulated through thresholding techniques, providing a smoothed approximation of the true signal of interest. An inverse wavelet transform returns the smoothed approximation to the original factor space, allowing application of traditional time series techniques. Figure 17 provides an example of wavelet methods to smooth a chaotic time series.

2.3.2 Forecasting Wavelet Resolution Levels

The MODWT is both shift invariant and defined for any sample size, facilitating a unique method of forecasting within the wavelet domain. The resolution levels resulting from a MODWT MRA can be viewed as individual time series, each more well-behaved then the original time series. Forecasting models, such as autoregressive integrated moving average (ARIMA), can thus be applied to produce forecasted values of the wavelet coefficients at each resolution level. The inverse MODWT is applied to these extended resolution levels to create a reconstruction of the original signal that includes forecasted values.

2.3.3 Hybrid Models

The third general application of wavelet methods consists of using the wavelet coefficients as predictive variables in a hybrid modeling approach. This is the most complicated approach as the transform adds to the dimensionality of the formulation by the number of wavelet resolution levels, requiring care in implementation and possible data reduction techniques. Furthermore, each of these steps requires the manipulation of tunable parameters within the transform as well as possible manipulation of the coefficients into the predictive model. Wavelet resolution levels can be highly collinear, requiring special considerations in certain hybrid modeling approaches.

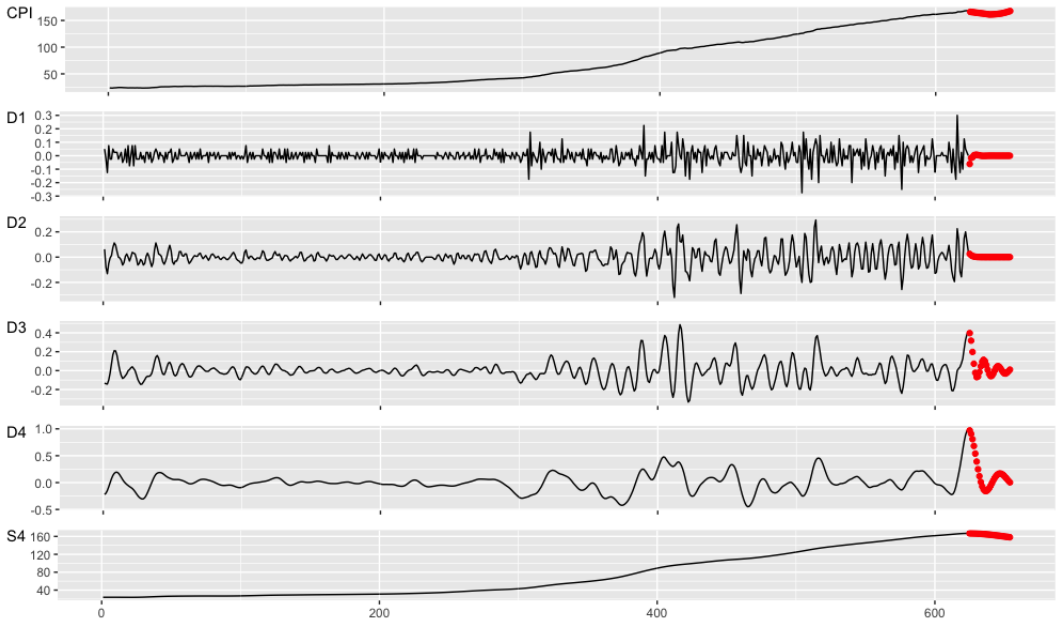


Figure 10: Forecast developed in the wavelet domain using the monthly U.S. consumer price index (CPI) from 1948 to 1999 dataset from the *waveslim* R package [95]. The original data is provided in the top subplot, with descending resolution levels of a four-level MODWT beneath. ARIMA models are fit each resolution level to produce thirty forecasted values (red), extending each individual resolution level. An inverse MODWT is applied to the extended resolution levels to a reconstructed CPI to include thirty forecasted values (red).

Despite these difficulties, estimating a time series model within wavelet space offers several highly desirable qualities in terms of predictive power, flexibility in implementation, and interpretation of the wavelet parameters for insight. Primarily, this approach allows full exploitation of the power of the wavelet decomposition as predictive variables. Use of the coefficients of each resolution level facilitates the identification and prioritization of individual resolution levels that are most predictive, while less weight can be applied upon those resolution levels that are comprised primarily of noise. This approach can many times accommodate additional exogenous variables separate from the wavelet transform, allowing for a wide variety of application. Proper selection of wavelet transform allows for a wide selection of prediction models, with neural networks being the most common application in current

literature.

Although approaches vary considerably, a generalization of this approach begins with a wavelet decomposition of the original time series. The wavelet coefficients of each resolution level are manipulated into vectors of predictive variables, and observations divided into both training and testing datasets. The training dataset and respective observed response are used to estimate model parameters, as used in general for estimation of models such as support vector machines and artificial neural networks. The resulting model is the used to produce forecasted results using the wavelet coefficients in the testing dataset. The model's forecasted values are finally compared to the known observed response to develop metrics to measure the model's effectiveness.

2.4 Review of Current Applications

The variety of implementation methods and parameters in wavelet methods allows for tailored application across a wide array of disciplines. As such, the following section presents a survey of current wavelet methods for forecasting to highlight best practices and potential research gaps across three disparate fields. These fields are selected for review due to unique requirements for timeliness and accuracy, driving unique approaches in the application of wavelet methods. This diversity is examined deliberately to show the wide variety of possible applications, any of which can be applied towards meteorological data.

The following section is organized by discipline: wind speed, earthquake, and traffic. Wind speed prediction is analyzed due to its predominantly short-term prediction requirement to inform resource and operational decisions in a relatively simple, easily understood system. Conversely, earthquake prediction is reviewed due to its requirement for a very short-term prediction within seconds. This has driven the

Time Horizon	Range	Applications
Very short-term	Few seconds to 30 minutes	- Electricity Market Clearing - Regulation Actions
Short-term	30 minutes to 6 hours	- Economic Lead Dispatch Planning - Load Increment/Decrement Decisions
Medium Term	6 hours to 1 day ahead	- Generator Online/Offline Decisions - Operational Security in Day-Ahead Electricity Market
Long Term	1 day to 1 week or more ahead	- Unit Commitment Decisions - Reserve Requirements Decisions - Maintenance Scheduling to Obtain Optimal Operating Cost

Table 2: Time scale classifications for wind speed prediction [85]

development of novel approaches for earthquake warning optimized for streamlined computational efficiency. Finally, a survey of traffic prediction methods is presented due to the wide variety of requirements that has driven a very diverse set of approaches and wavelet method solutions.

2.4.1 Wind Speed Prediction

Research into accurate wind speed prediction models is becoming increasingly crucial as power systems become more reliant upon wind-driven systems. Intermittency of the wind is the biggest challenge for integration of these systems for managers of both the electric power grid and electricity markets [85]. Efficient prediction of wind speed allows improvement efforts in process and prediction yielding optimized output from wind-driven power generators, reducing consumption of fossil fuels and increasing technical advantage in a rapidly growing market [85]. Comprehensive reviews of all wind speed prediction methods, to include some wavelet methods, can be found in Soman et al. [85], Tascikaraoglu and Uzunoglu [90], and Wang et al. [94].

Soman et al. [85] define a series of time horizons for interest in wind speed prediction, shown in Table 2. These provide a good categorization tool as each time

horizon implies unique requirements in terms of data complexity and computational efficiency. Wavelet methods for wind speed prediction are commonly aligned with a short time horizon due to the method's ability to efficiently analyze high dimension, high frequency data.

Current studies employ wavelets in a variety of methods, predominantly through either de-noising the signal, extending the series of wavelet coefficients, or using the coefficients in a hybrid model. Liu et al. [50] and Singh and Mohapatra [83] both produce a forecast by extending the series of decomposed wavelet coefficients. These resolution levels of coefficients are treated as multiple time series that are better behaved than the original time series. An Auto-Regressive, Integrated, Moving-Average (ARIMA) model is fit to extend these decomposed resolution levels, and then the inverse wavelet decomposition is applied to reassemble the original time series to now include extended forecast values.

The most prevalent wavelet method is to use the decomposed coefficients to train a hybrid model, most commonly a neural network. One notable exception is the use of support vector machines (SVM) in Zeng and Qiao [98]. Table 3 summarizes the literature in wind speed prediction to include the method of wavelet application. This ranges from an application in wavelets and thresholding to de-noise the original signal, extending the wavelet coefficients with reconstruction of an extended signal, or use of the wavelet features to train a hybrid model.

Applications of wavelet methods in wind forecasting vary considerably, but common trends consist of the time period of interest and requirement for data reduction. With few exception, these models produced predictions of short-term interest to inform decisions in economic planning and network management. This is most likely due to the computational efficiency enabled by wavelet methods, facilitating accurate predictions must faster compared to competing methods. Data reduction require-

ments can be seen across methods as redundant wavelet transformations create large data structures that are both difficult to manage and prevent timely convergence in hybrid model applications. None of these works point to existing literature to guide their decision over which data reduction technique to employ, and application of these methods appears to be ad hoc without existing comparisons.

2.4.2 Earthquake Prediction

Some of the earliest applications of wavelets, such as Goupillaud et al. [23], were used by the geoscience and seismology communities for the exploration of oil and gas deposits. Therefore, it should come as no surprise that geoscience researchers continue to apply these methods in areas such as the prediction of earthquake magnitude and location. Earthquake seismology is characterized by a series of waves, most notably the P and S waves [82]. Ruptures emit both low amplitude, high velocity P waves and high amplitude but slower S waves that can arrive at a monitoring station several seconds later. The time interval between the arrival of the P and S waves increases as monitoring stations are located further from the site of the rupture. The P waves represent low-level and imperceptible motion whereas S waves consist of the destructive movements typically associated with earthquakes. Characterizations of these waves are exceedingly complex and depend upon numerous factors such as ground composition and relative position of rupture and sensor. Through evaluating these factors, Allen and Kanamori [3] first show that exploiting this time differential between the initial P waves and damaging S waves could be used to produce an earthquake warning system.

Earthquake early warning systems attempt to determine the location and magnitude of an earthquake in sufficient time to issue a timely alert. Although only seconds separate the arrival of P and destructive waves, timely alerts allow authorities to mit-

Study	Wavelet	Method	Hybrid Model	Time Horizon
De Aquino et al. [14]	DWT, Db3, j=3	Hybrid	ANN	Short/Medium
Bhaskar and Singh [8]	MODWT, Mexican hat	Hybrid	AWNN	Long
Catalão et al. [10]	DWT, Db4 j=3	De-noising	PSO-ANFIS	Short
Catalão et al. [11]	DWT, Db4 j=3	Hybrid	Levenberg-Marquardt NN	Short
Chitsaz et al. [12]	Morlet	Hybrid	WNN using CSO	Short
Doucoure et al. [17]	DWT, Mexican hat	Hybrid	AWNN, Hurst predictability	Short
Faria et al. [19]	MODWT	De-noising	ARIMA	Medium
Hunt and Nason [32]	DWPT	Hybrid	PCA and linear model	Short
Khan and Shahidehpour [42]	DWT, Db1-4, j=3	Hybrid	Spline smoothing/linear model	Medium
Lei and Ran [46]	DWT, j=6	Extension	ARIMA	Short
Liu et al. [49]	DWT, j=1	Hybrid	Support Vector Machine, Genetic Algorithm	Short
Liu et al. [50]	DWT, Db4, j=3	Hybrid	ARIMA	Short
Liu et al. [51]	DWT, DWPT	Hybrid	Neuro-fuzzy ANFIS, RBF NN	Short
Liu et al. [52]	DWT, DWPT	Hybrid	ANFIS, MLP	Short
Liu et al. [54]	EWT	Hybrid	Elman NN	Short
Meng et al. [65]	DWPT	Hybrid	Crisscross optimization NN, PSO, ANN	Short
Osório et al. [71]	DWT, Db4, j=3	Hybrid	ANFIS and mutual information	Short
Singh and Mohapatra [83]	MODWT	Extension	ARIMA	Short
Zeng and Qiao [98]	DWT, Mexican hat	Hybrid	Support Vector Machine	Short
Zhang et al. [100]	DWT	De-noising	RBFNN and seasonal adjustment	Short

Table 3: Summary of current wavelet methods in prediction of wind speed

Study	Wavelet	Method	Hybrid Model	Time Horizon
Hloupis and Vallianatos [29]	MODWT	Hybrid	Linear Model	Very-short
McGuire et al. [64]	DWT, CDF(2,4)	Hybrid	Linear Model	Very-short
Reddy and Nair [74]	DWT, CDF(2,4)	Hybrid	SVM	Very-short
Simons et al. [82]	DWT, CDF(2,4)	Hybrid	Linear Model	Very-short

Table 4: Summary of current wavelet methods in earthquake prediction

igate damage through actions such as stopping trains and alerting the populace [39]. These systems rely upon the findings of previous studies that identified the radiated seismic energy from the first few seconds of a rupture through the P wave scale with the final magnitude [3] [96] [70]. Olson and Allen [70] conclude that earthquake ruptures are deterministic in nature, allowing early warning systems to calculate a great deal of information concerning a rupture from only the first few seconds of readings. These results established requirements in earthquake early warning systems for robustness to noise and computational speed. Systems must be capable of capturing the arrival of a P wave inside an inherently noisy seismological time series and then compute the projected magnitude within the wave arrival time differential to allow for an operationally relevant alert.

2.4.3 Analysis Using Wavelet Coefficients

Simons et al. [82] were the first to apply wavelet methods to seismological time series for the estimation of earthquake magnitude using the P wave. This research focused on providing a fully-automated algorithm with the speed and simplicity to be deployed in real-world sensor networks. Previous studies conclude direct calculation of the time-domain expression of recorded waveforms is notoriously difficult to compute and competing spectral methods produced limited predictive capability. Simons et al. [82] present the predominant period estimator (PDE) for computing τ_c^2 the predominant period of the P wave using

$$\tau_c^2 = 4\pi^2 \frac{\int_0^{\tau_0} u^2(t)dt}{\int_0^{\tau_0} \dot{u}^2(t)dt} = \frac{\int_0^{\infty} |\hat{u}(f)|^2 df}{\int_0^{\infty} f^2 |\hat{u}(f)|^2 df}$$

where $u(t)$ and $\dot{u}(t)$ are the ground motion displacement and velocity as a function of time t, and τ_0 is the duration of the P waveform. The value for τ_0 is usually assumed to be 3 or 4, and τ_c is determined using an iterative algorithm in real-time. However,

this method suffers with convergence failure resulting in significant scattering due to the iteration and recursive calculation. Wavelet methods are then used as they are complimentary to seismic waveforms, providing the requisite computational speed and stability while including accepted methods for simultaneously de-noising the data.

Simons et al. [82] analyze 2,272 seismograms recorded by 142 monitoring stations in California that record 53 seismic events at 34 distinct magnitudes. They employ the discrete wavelet transform (DWT) using a wavelet basis of biorthogonal construction with two and four vanishing moments for the primal and dual wavelets termed Cohen-Daubechies-Feauveau (CDF(2,4)). A DWT is calculated for each seismogram over five resolution levels using the fast lifting algorithm of Sweldens [89] due to computational speed and applicability to real-world systems that may have limited on-board computational power. A threshold is T_j is defined at resolution level j in terms of the number of coefficients at that resolution level N_j and $\hat{\sigma}_j$, the median absolute deviation from the median of the coefficients, as

$$T_j = \hat{\sigma}_j \sqrt{2 \ln N_j}.$$

Soft thresholding of the wavelet coefficients is then applied by replacing original coefficients by their signed distance from the threshold. This effectively removes all random noise, leaving only significant coefficients related to P wave detection.

Simons et al. [82] isolate a wavelet coefficient in a particular resolution level which provides the greatest predictive power using their methodology. This coefficient is averaged across all detecting stations and used in a linear model to produce an estimate of the resulting earthquake magnitude. Results show this method predicts the magnitude to within approximately one unit.

The methodology of McGuire et al. [64] build upon Simons et al. [82] to provide a key insight linking frequency to magnitude in the prediction of devastating

earthquakes. The study uses the same method of wavelet transformation on sea floor seismograms of the 8.1 magnitude 2003 Tokachi-Oki earthquake. Notably, McGuire et al. [64] omit soft thresholding of wavelet coefficients as the arrival of the P wave was known to be within the window of provided data. Results indicate that smaller earthquakes result in significant wavelet coefficients typically located very near to the initial arrival of the P wave. This trend is not found within large magnitude earthquakes, with the largest-scale coefficients increasing in amplitude as earthquake magnitude increases. Earthquakes with exceptional large final magnitudes build in strength during the initial rupture, as shown in the behavior of the P wave. Use of this method on the 2003 Tokachi-Oki earthquake confirm the findings, providing a method to predict extremely high magnitude events.

Hloupis and Vallianatos [29] and Hloupis and Vallianatos [30] further build upon Simons et al. [82], first with an improved magnitude estimator (WME) and later with a wavelet-based epicenter estimator (WEpE). Hloupis and Vallianatos [29] evaluate seismograms of 325 earthquakes collected between 2008 and 2011 from the South Aegean Sea, focusing exclusively upon the Island of Crete. This region contains two seismological networks resulting in average distance of coverage of 60 km. This coverage means the network can not be characterized as a dense sensor network used in previous studies of Simons et al. [82] and McGuire et al. [64].

The use of the MODWT further differentiates Hloupis and Vallianatos [29] work from previous studies. Previous use of the CDF (2,4) wavelet basis was justified due to computational speed when paired with the Sweldens [89] lifting algorithm. This method is incredibly fast, requiring $\mathcal{O}(N)$ operations compared to $\mathcal{O}(N \log_2 N)$ for MODWT. However, this is the same computational price for the popular Fast Fourier Transformation (FFT) and is therefore deemed acceptable.

Noise is removed using methods presented in Vallianatos and Hloupis [92], where

MODWT are applied and certain nuisance resolution levels are removed. This application focuses on automation, but specific criteria or automation methods are not provided. Like previous studies, correlation is shown between maximal values of wavelet coefficients at certain resolution levels. Therefore, a linear model is fit using the maximum coefficient of the seventh resolution level. The results of Hloupis and Vallianatos [29] show that the WME outperforms PDE estimators, but a comparison with previous wavelet-based methods is not included. This implies wavelet-based methods may be superior to competing methods on non-dense sensor networks, allowing deployment of early warning systems to networks that do not meet the system's strict requirements.

Hloupis and Vallianatos [30] propose the WEpE using a wavelet azimuth estimation (WAE) and two stations' sub array method. The WAE relies upon the polarization of the P wave for a regional earthquake. This polarization implies that a de-noised P wave signal will have zero, or minimal, variance except in the line of travel. The WAE is automated using the methods of Galiana-Merino et al. [21] of wavelet de-noising, P wave detection, and azimuth estimation. The WAE is implemented in real-world application in 20 shallow earthquakes against Hypoinverse software, the current industry standard for epicenter estimation. The results show WAE provides reduced error compared to Hypoinverse and at significantly greater computational speed, implying the method would be acceptable for use in an early warning system.

The WEpE combines results of the WAE with an existing method, the two stations' sub array, to greatly improve epicenter estimation. The two stations' sub array combines detections from two monitoring sites to form an ellipse area of interest. The inclusion of the direction azimuth found by WAE greatly reduces the size of the ellipse and significantly improves the predictive power of the method compared to existing

methods. This is especially apparent in sparse sensor networks, where competing methods provided unstable estimates.

Reddy and Nair [74] extend the work of Simons et al. [82] by applying a support vector machine (SVM) statistical learning machine to the decomposed wavelet coefficients. Reddy and Nair [74] utilize 1,689 seismograms associated with 108 earthquakes from KiK-net, an earthquake detecting network in Japan. The magnitude and epicenter of each earthquake is provided by the National Research Institute for Earth Science and Disaster Prevention (NIED) and the Japanese Meteorological Agency (JMA). The wavelet decomposition method closely mirrors that of Simons et al. [82] with a seven resolution level decomposition using a biorthogonal CDF(2,4) basis. Soft thresholding from Simons et al. [82] is applied and a SVM is fit using a Matlab toolbox. Results indicate improvements over Simons et al. [82] from one unit of earthquake magnitude to 0.4 units.

2.4.4 Traffic Congestion Prediction

Short-term traffic forecasting can be applied to traffic incident detection using factors such as traffic volume, density, speed, or travel times. These forecasts can be evaluated to decrease emergency resources' response time to incidents, routing of traffic around the incident, and civil planning resources to improve roadway design. Traditionally performed by human analysis, these methods rely upon recorded values for lane occupancy using spatial and temporal measures for incident detection. Automated traffic systems have been developed since the 1980's to accommodate increasing requirements for traffic modeling. The timeliness of these forecasts depend upon application, ranging from several hours to only seconds. Traffic patterns are noisy and difficult to fully characterize as they derive from human action. Influences such as the presence of traffic accidents or changing weather patterns can result in

ill-behaved time series from traffic sensors, making traditional modeling techniques difficult. Vlahogianni et al. [93] provide a thorough survey of short-term forecasting methodologies.

2.4.5 Traffic Incident Detection

Wavelet methods were first applied to the traffic detection problem through a series of companion papers: Samant and Adeli [79], Adeli and Samant [2], Adeli and Karim [1], and some extensions to these works. Samant and Adeli [79] present a wavelet-based two-stage feature extraction algorithm as a preprocessing tool for training a neural network. This process utilizes a DWT and Linear Discriminant Analysis (LDA) sequentially to both de-noise and reduce the dimensionality of raw traffic pattern data. The DWT de-noising is accomplished using Daubechies wavelets by removing complete detail resolution levels believed to be comprised exclusively of noise. Human logging errors in observed data require the use of simulated traffic incident datasets for feature extraction and modeling.

Adeli and Samant [2] apply this pre-processing algorithm to train a neural network for traffic incident detection. The study aims to improve upon the false alarm rate in contemporary real-world systems based upon a moving average analysis. Researchers use an adaptive conjugate gradient neural network learning model, where weights are chosen in the direction of the greatest improvement to system error. The neural network displays significantly improved time for converge using pre-processed data. Some experimentation determines this improvement is primarily due to the effects from wavelet-based de-noising. The final model results in a faster traffic incident algorithm with improved accuracy of approximately a 98% detection rate with less than 1% false alarm rate. These results are shown again in Samant and Adeli [80] when the pre-processed traffic data is applied to a fuzzy-based neural network.

Adeli and Karim [1] provide an alternative application of the pre-processed dataset of a single-station sensor using the methods of Samant and Adeli [79]. Unlike earlier applications, this research applies a soft thresholding technique to the wavelet coefficients prior to applying an inverse DWT, and then feeding the de-noised signal into a fuzzy clustering algorithm. The clustering algorithm is applied to reduce the dimensionality of the dataset, since the DWT is applied only for smoothing the data and not feature extraction. The smoothed, reduced dataset is then used to train a radial basis function neural network (RBFNN). Karim and Adeli [41] evaluate and compare this method against the California algorithm, a contemporary real-world traffic incident detection method. Both real-world and simulated datasets to assess the methods' performance in detection rate, false alarm rate, and detection time. The wavelet-based method outperformed the California algorithm consistently through improvements in detection and false alarm rates. However, both algorithms shared near identical practical detection time rates. One additional strength of the wavelet-based method is the lack of tunable parameters. The existing California model requires selection of certain threshold parameters based upon localized traffic patterns. However, the wavelet-based neural network lacks any tunable parameters as it is derived exclusively from a nonparametric approach and requires only a training period for the neural network. Ghosh-Dastidar and Adeli [22] further expand this work with exploration of differing wavelet and clustering approaches, finding potential improvements in the use of a Coifman wavelet and Mahalanobis distance data clustering technique applied to a Levenberg-Marquardt backpropagation neural network. Xie and Zhang [97] replicate this work, comparing a more basic implementation of wavelets and the Levenberg-Marquardt backpropagation neural network to show increased performance over existing neural network methods.

Teng and Qi [91] propose an alternative approach, utilizing the wavelet transform as a feature extraction of occupancy data and directly applying the resulting coefficients in a neural network for categorization. Only occupancy data is used to provide a fair comparison of wavelet methods with legacy models based solely on this information, such as the California algorithm. Soft thresholding is applied to DWT coefficients and any significant values in the detailed resolution levels are used for a neural network classification of changing traffic patterns. Results of this method are compared agains a multi-layer feed-forward (MLF) neural network, a probabilistic neural network, the fuzzy-wavelet RBFNN algorithm of Adeli and Karim [1], a low-pass filtering algorithm, and the California algorithm. Results indicate significant improvements are derived from direct application of the wavelet coefficients into training the neural network. Furthermore, using wavelets for both de-noising and feature extraction eliminates the requirement for a clustering algorithm which may or may not be optimal for use with a neural network.

2.4.6 Traffic Flow Detection

Increased availability of GPS and autonomous detection systems transformed short-term traffic forecasting from an issue of single incident detection to overall traffic flow management. These intelligence systems now provide a greater breadth of situational awareness and control to traffic managers to optimize regional congestion.

Jiang and Adeli [36] present a wavelet-based process to detect atypical changes, or singularities, in traffic flow. Wavelet methods are use to identify perturbation in traffic flow outside daily and weekly traffic patterns to allow routing systems to divert traffic to alternate routes. This research employs the discrete wavelet packet transform (DWPT) to provide a richer decomposition of the signal. Although computationally expensive and redundant, the DWPT provides finer detail in the frequency

information that facilitates an improved de-noising capability compared to the DWT. Jiang and Adeli [36] apply the DWPT to the traffic sensor data to obtain a MRA of the de-noised signal, to which statistical autocorrelation function (ACF) is applied to analyze the correlation between MRA decomposition level and the characteristics of the original time series. This method is presented to apply additional rigor and objective processes to the selection of wavelet decomposition level, which is usually accomplished using trial and error [73]. The hybrid modeling approach is shown to have promising applicability for traffic forecasting models.

Jiang and Adeli [37] present the first integration of wavelet-methods with a dynamic neural network model for both short and long-term traffic forecasting. This novel application utilizes a nonparametric dynamic time-delay recurrent wavelet neural network model that relies upon data preprocessed using a MODWT and modified Gram-Schmidt algorithm. The non-decimated wavelet transform is redundant, however provides some excellent properties for multidimensional decomposition of a time series. The MODWT produces far too many vectors of wavelet coefficients to be computationally feasible in actual implementation. The modified Gram-Schmidt algorithm, first proposed in Zhang [99], is used to select only those wavelet coefficients required to produce an accurate result and discarding the rest. The resulting data facilitates timely convergence of the dynamic neural network capable of producing adequate results for forecasting.

2.5 Analysis of Wavelet Applications

Wavelet methods are an advanced application in signal analysis; however, their prevalence in software and literature enables easy interpretation and implementation. Despite this ease of use, several tunable parameters are required that strongly impact the predictive accuracy of the model to include choice of level of decomposition

and choice of wavelet function. Tascikaraoglu and Uzunoglu [90] suggest that the complexity of wind speed analysis require the development of site-specific predictive models. Research in both earthquake and traffic pattern prediction suggest similar approaches. The tunable parameters of wavelet methods makes them highly adaptable to individual sites and allow rapid development of a site-specific model. These methods are also highly adaptive to varied datatypes, and lack any requirement for a stationary time series.

Applications suggest that choice of wavelet transform depends upon a tradeoff between computational efficiency and predictive accuracy. The DWPT and MODWT and consistently preferred for accuracy; however, a DWT is preferred when computational efficiency is paramount such as in earthquake prediction. None of the application methods used the MODWPT that combines the resolution in frequency of the DWPT with the desirable properties of the MODWT in time series analyses.

Thresholding wavelet coefficients is commonly used as a method to reduce noise in the data, with soft thresholding techniques such as VisuShrink being the most popular due prevalence in software packages. Very few of the studies explicitly state the assumption for random Gaussian errors, a critical assumption to most thresholding techniques. None of the studies under evaluation employed block thresholding and few used any data adaptive thresholding techniques.

2.5.1 Weakness and Limitations of Wavelet Methods for Forecasting

Current literature indicates wavelet methods are a powerful tool for time series analysis and forecast estimation; however, these techniques have some inherent weaknesses and limitations that must be considered. Primarily, wavelet methods can be complex to implement and report. Wavelet techniques can require some investment of time to delve into and understand how to best pair a wavelet method for a particular

application. Application of wavelet methods is complicated by the wide variety of parameters for the wavelet transform such as the type of wavelet transform, number of decomposition layers, and selection of the mother wavelet. As shown above, wavelet methods are further complicated by their reliance upon other modeling disciplines. Application of these techniques requires and understanding of both wavelets as well as traditional time series models or hybrid models, depending upon the approach. Even if the individual time series analyst understands how to apply wavelet methods, it can be very difficult to justify conclusions if a thorough understanding of this complex underlying methodology is required.

The complexity of wavelet methods makes them best suited for very large, complex datasets. These techniques are not meant to completely replace traditional time series approaches, which remain highly relevant across many applications. Wavelet methods do fill critical needs for modeling approaches where computational efficiency is required or particular assumptions do not hold.

2.5.2 Prospects for Future Research

Many studies apply the MODWT due to its preferential qualities in time series analysis. However, the redundancy of this transform results in a series of vectors, equal to the level of decomposition plus one, each of which are the same length as the original data. This creates issues from both simple data management as well as in application for convergence in hybrid modeling. Several studies seek to address this issue through data reduction techniques as seen in Table 5. Most of these studies apply traditional time series techniques to assess which resolution levels of the wavelet transform contribute to the predictability in the response, allowing removal of non-predictable resolution levels and reducing the overall dimensionality. Development of such measures would facilitate more efficient application of the MODWT, as well

as enable more analysis through the scarcely used MODWPT. Additional research is required to assess these competing approaches through comparison and generalization for use in wavelet methods for time series analysis.

Thresholding techniques for time series are commonly used but rarely explored. Most applications used soft thresholding techniques, primarily VisuShrink. There was rarely an assumption or analysis for normality in error terms required by this approach. Furthermore, application of data adaptive thresholding techniques are very rare. Further research is required to assess these techniques in time series applications, particularly to develop guidance for preferred method by application. Soft thresholding through VisuShrink would most likely continue to be the best in most applications, yet chaotic time series may benefit from hard or data adaptive thresholding techniques. Furthermore, these same abrupt changes may violate the required assumption for normality in the errors, requiring a non-parametric thresholding approach such as the method of McGinnity and Chicken [62].

Selection of wavelet transform varies throughout the literature based upon the application requirements. DWT is preferred when computational efficiency is required, whereas the MODWT and DWPT are both preferred for time series when the situation allows. The growing prevalence of both the MODWT and DWPT in more recent literature may be indicative of increasing capabilities of computing power. Liu et al. [53] offer direct comparison of the performance of competing transforms in application, with the DWPT offering the best results. None of the studies employed the

Study	**Method**
Doucoure et al. [17]	Hurst Predictability
Hunt and Nason [32]	Principle Component Analysis (PCA)
Jiang and Adeli [36]	Statistical Autocorrelation Function (ACF)
Jiang and Adeli [37]	Modified Gram-Schmidt
Zhang et al. [100]	Seasonal Adjustment

Table 5: Examples of data reduction techniques found in current literature using wavelet methods.

MODWPT transform or matching pursuit methods outlined in Percival and Walden [73].

Several studies, such as Tascikaraoglu and Uzunoglu [90], state the complexity of variables for individual sites requires unique models for every forecast location. Conversely, several earthquake studies note that models appear to be overfit to either specific devastating events or niche implementations within regional seismograms of low grade events. None of the literature examines the impact to forecast, or some measure of uncertainty, created by generalizing such a model. Wavelet methods have been shown to be robust across implementation, such as the approach of Hloupis and Vallianatos [29] on a sparse network array. These methods may be able to provide a universal model that sacrifices a small amount of predictive accuracy for ease in implementation.

An extension of Hloupis and Vallianatos [29] could consist of further analysis into robustness of wavelet methods to sparsity of network, especially focused upon optimization of network design. Procurement and maintenance of a sensor array is often a relatively high cost initiative for any organization. The sparse sensor networks are problematic as they produce fewer readings and the increased distances in sensors result in greater levels of noise. Wavelet methods are perfectly suited to such an implementation due to the ability to utilize reduced datasets to identify underlying features. A possible contribution would be an assessment of network sparsity on predictive power using various quantities of sensor returns on the same dense network. Varying sensor quantity would evaluate a model's ability to predict with limited inputs. The ability of wavelet methods to provide comparable forecasts using less sensor would allow organizations to optimize network management.

Wavelet methods rarely include a rigorous evaluation of proposed model parameters such as wavelet basis, level of wavelet decomposition, correlation of coefficient to

response, or analysis of assumptions and fit of linear models. Parameters are chosen arbitrarily or due to some assessed quality; however, no attempt is made to quantify or defend impacts of these selections. A designed experiment using each of these parameters as an experimental factor would reveal how robust each model is towards change in a parameter. This in turn would allow researchers to select a model that is the most accurate and robust to these tunable parameters.

None of the literature reviewed uses wavelets to predict volatility in the response. Use of wavelet-enabled autoregressive conditional heteroskedasticity (ARCH) models is common to econometric literature to predict periods of volatility or relative calm in financial time series. Introducing these methods to the physical sciences may enable better prediction of periods of uncertainty.

III. Experimental design in complex model formulation for lightning prediction

Space launch operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) are complicated by unique requirements for near-real time determination of risk from lightning. Weather sensor networks for lightning forecasting produce data that are noisy, high volume, and high frequency time series for which traditional forecasting methods are often ill-suited. Current approaches result in significant residual uncertainties and consequentially may result in forecasting operational policies that are excessively conservative or inefficient. This work first proposes a forecasting methodology using wavelet decomposition of chaotic weather sensor time series and semiparametric single-index models to mitigate the chaotic signal and any possible distributional misspecification. Then, a screening experiment with augmentations is used to demonstrate how to explore the complex factor space of model parameters, guiding decisions regarding model formulation and gaining insight for follow-on research. Results indicate a promising technique for operationally relevant lightning prediction from chaotic sensor measurements.[1]

3.1 Introduction

Advances in sensor production and scalability have driven the development of sensor networks that are both relatively cheap to produce and easily deployable. The Department of Defense has become increasingly reliant upon such networks to perform tasks such as battlefield surveillance of remote areas through seismic and acoustic monitoring [4], space-borne missile defense [40], and monitoring of lightning risk at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) [87]. These sensors collect data that are noisy, high volume, and high frequency time

[1]Paper to appear in the International Journal of Experimental Design and Process Optimisation.

series which can be problematic when developing operationally relevant forecasts using traditional time series analysis. Parametric assumptions may not hold or models may over-smooth the response, losing the signal of interest in the smoothing process. The result is a model with a high degree of residual uncertainty that forces operational commanders to employ policies that are possibly excessively conservative or inefficient. This research identifies and evaluates a wavelet-enabled semiparametric single-index modeling approach for lightning warning derived using chaotic time series data at KSC/CCSFS to meet operational requirements for timeliness and accuracy.

Accurate prediction modeling provides vital insight into complex systems for risk assessment and management of resources; yet, the implementation and use of these predictive models is complicated by information availability. Modern sensor networks meant to feed such models produce high frequency, high dimensional datasets often collected as noisy and non-stationary time series. The artifacts of interest within these series frequently consist of sharp and abrupt changes that traditional modeling applications may fail to accurately capture. Wavelet methods are being used in these situations due to their ability to tackle these artifacts in a computationally efficient manner. These wavelet methods are being employed as a preprocessing method, either for de-noising and smoothing a time series or serving as a feature selection method. While a hybrid wavelet approach provides ample flexibility for tailored application, there are several challenges presented during model formulation. Wavelet techniques can require some investment of time to delve into, especially to comprehend how to best pair a wavelet method to a particular application. Application of wavelet methods is complicated by the wide variety of parameters for the wavelet transform such as the type of wavelet transform, number of decomposition levels, and selection of the mother wavelet. Application of these techniques requires an understanding of both wavelets as well as traditional time series models or hybrid models, depending

upon the approach. Furthermore, the single-index model requires significant computational resources for model estimation from a multivariate dataset. Therefore, any exploration of model parameters must be efficient and judiciously use available computational resources.

This study proposes a new modeling framework for lightning prediction and employs a design of experiments (DOE) approach using a screening experiment with augmentation to guide and inform the complex model formulation. Section 3.2 provides an overview of techniques used in this formulation, to include both wavelet methods and the semiparametric single index model. The intent of this first model formulation is to apply the wavelet methodology of Section 3.3 using only chaotic Electric Field Mill (EFM) time series in an attempt to evaluate if the sensors are indeed predictive of lightning activity, and then evaluate the potential limits of this particular approach. The use of designed experiments provide a clear structure to efficiently examine model parameters and their possible interactions. Section 3.4.1 presents the series of experiments and their impact in guiding parameter selection that produced the results discussed in Section 3.4.2.

3.1.1 Wavelets in Forecasting

Some of the earliest applications of wavelets, such as those of Goupillaud et al. [23], were used by the geoscience and seismology communities for the exploration of oil and gas deposits. Therefore, not surprisingly geoscience researchers continue to apply these methods in areas such as the prediction of earthquake magnitude and location. Earthquake early warning systems attempt to determine the location and magnitude of an earthquake in sufficient time to issue a timely alert. Although only seconds may separate the warning to the arrival of destructive waves, timely alerts allow authorities to mitigate damage through actions such as stopping trains and alerting the populace

[39]. These systems rely upon exploiting the findings of previous studies that identified the radiated seismic energy from the first few seconds of a rupture scale with the final magnitude [3] [96] [70]. Olson and Allen [70] conclude that earthquake ruptures are deterministic in nature, allowing early warning systems to calculate a great deal of information concerning a rupture from only the first few seconds of readings. These results established requirements in earthquake early warning systems for robustness to noise and computational speed. Systems must be capable of capturing the arrival of initial waves inside an inherently noisy seismological time series and then compute the projected magnitude within the wave arrival time differential to allow for an operationally relevant alert.

Research into developing accurate wind speed prediction models is becoming increasingly crucial as power systems become more reliant upon wind-driven systems. Intermittency of the wind is the biggest challenge for integration of these systems for managers of both the electric power grid and electricity markets [85]. Efficient prediction of wind speed allows optimized output from wind-driven power generators, reducing consumption of fossil fuels and increasing technical advantage in a rapidly growing market [85]. Comprehensive reviews of all wind speed prediction methods, to include some wavelet methods, can be found in Soman et al. [85], Tascikaraoglu and Uzunoglu [90], and Wang et al. [94].

Wavelet methods were first applied to the traffic detection problem through a series of companion papers: Samant and Adeli [79], Adeli and Samant [2], Adeli and Karim [1], and some extensions to these works. Samant and Adeli [79] present a wavelet-based, two-stage feature extraction algorithm as a preprocessing tool for training a neural network. This process utilizes a Discrete Wavelet Transform (DWT) and Linear Discriminant Analysis (LDA) sequentially to both de-noise and reduce the dimensionality of raw traffic pattern data. The DWT de-noising is accomplished

using Daubechies wavelets by removing complete detail resolution levels believed to be comprised exclusively of noise.

3.1.2 The Lightning Prediction Problem

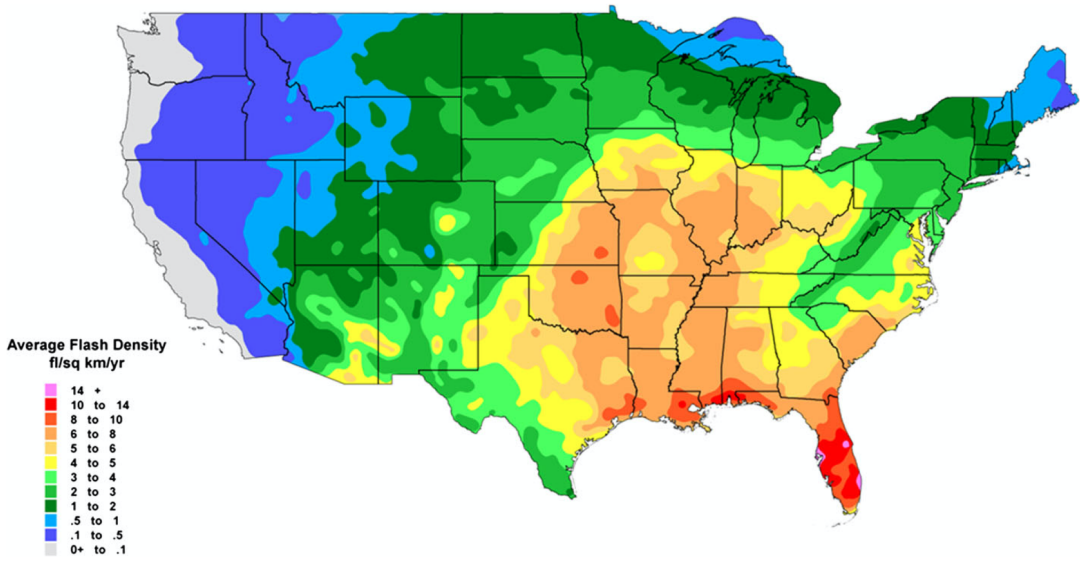


Figure 11: Cloud-to-ground lightning flash density (1997-2010) for the USA from the National Lightning Detection Network [77]

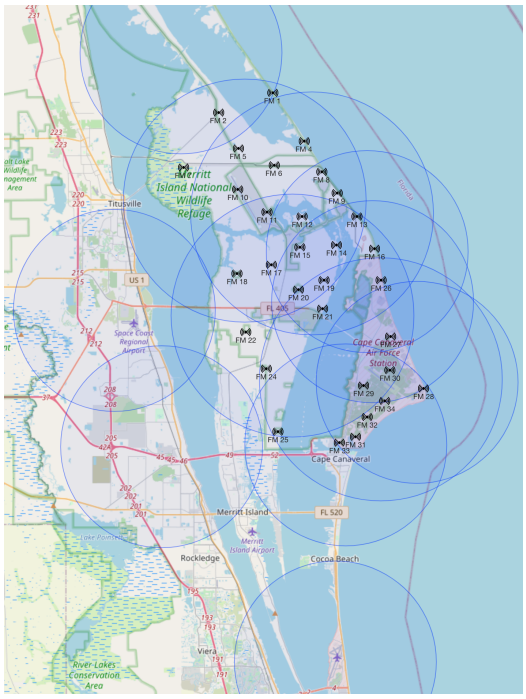
KSC/CCSFS experiences one of the world's highest incidence of lightning, impacting both the launch of space vehicles and daily support activity. Figure 26 provides a heatmap of cloud-to-ground lightning for the United States, where a high density of activity can be seen in Florida's central peninsula. Accurate lightning forecasts are essential for safe flight line operations to protect both personnel and high-value equipment, requiring both the prediction of lightning onset and the cessation of lightning events following a storm. A wide array of sensors are employed at KSC/CCSFS to inform a lightning warning system comprised of ten 5 nautical miles (NM) or 6NM circular warning regions as seen in Figure 28. The Lightning Detection and Ranging (LDAR) system is a sensor network developed by NASA that detects and records total lightning (both cloud-to-cloud and cloud-to-ground) within 100NM of KSC/CCSFS.

A network of Electric Field Mills (EFM) measures the ground-level electric potential of the atmosphere, detecting when electrified clouds move into the area. The EFM sensors are spread throughout the KSC/CCSFS region, as seen in Figure 28. These networks collect measurements at a 50 hertz rate, resulting in very large data sets available for modeling. Current literature suggests lightning risk can be predicted by a sudden change of polarity and increase of magnitude of the atmospheric electric potential as recorded by EFM networks [6] [55]. However, the electric potential is constantly altered within the clouds which results in a chaotic and nonstationary EFM signal [44]. The EFM signals have also been shown to experience a strong diurnal cycle and spatial variability specific to KSC/CCSFS [56]. These issues of high chaotic noise and high frequency/volume data have confounded recent attempts for more accurate predictive modeling of lightning onset or cessation.

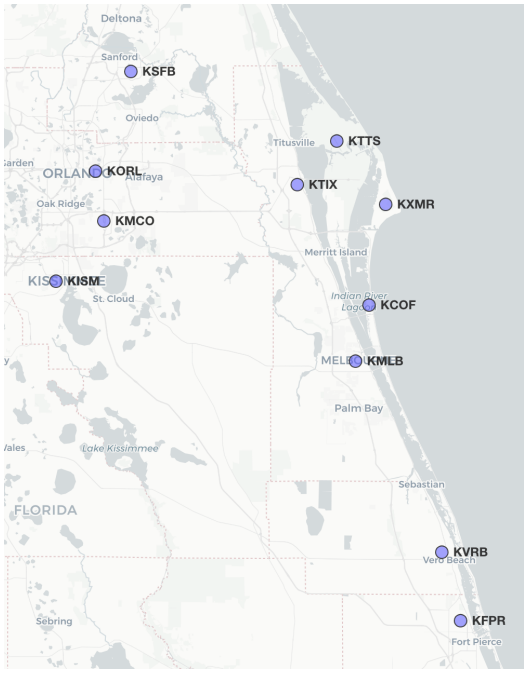
Figure 12b depicts the eleven collection sites for Meteorological Terminal Aviation Routine (METAR) data. These sites collect hourly measures of common meteorological factors to include cloud cover and height, pressure, temperature, dewpoint, visibility, wind direction, wind speed, wind gust, and altimeter.

Lightning is a relatively rare event that poses significant risk towards personnel and property. The study of lightning patterns within the USA is well documented, with the map in Figure 26 providing one example of a visualization of detected cloud-to-ground lightning in the contiguous United States. Notably, the concentration of lightning activity in Florida receives more cloud-to-ground lightning than any other state. KSC/CCSFS receives approximately 4-10 lightning flashes per kilometer every year.

The prevalence of lightning in this region drives the need for near real-time determination of lightning risk to support operations. Launch activities at Cape Canaveral require the forecasting of lightning activity. Personnel and high value equipment are



(a) KSC/CCSFS map with locations of EFM sensors and lightning warning circles



(b) Regional map of Eastern central Florida with eleven METARs collection locations

Figure 12: On the left, location of 12 lightning warning circles (blue) and 31 active EFM sensors throughout the region containing both KSC/CCSFS and Patrick Air Force Base (southernmost warning circle). On the right, a regional map of the same area providing the location of the 11 locations for METARs data collection.

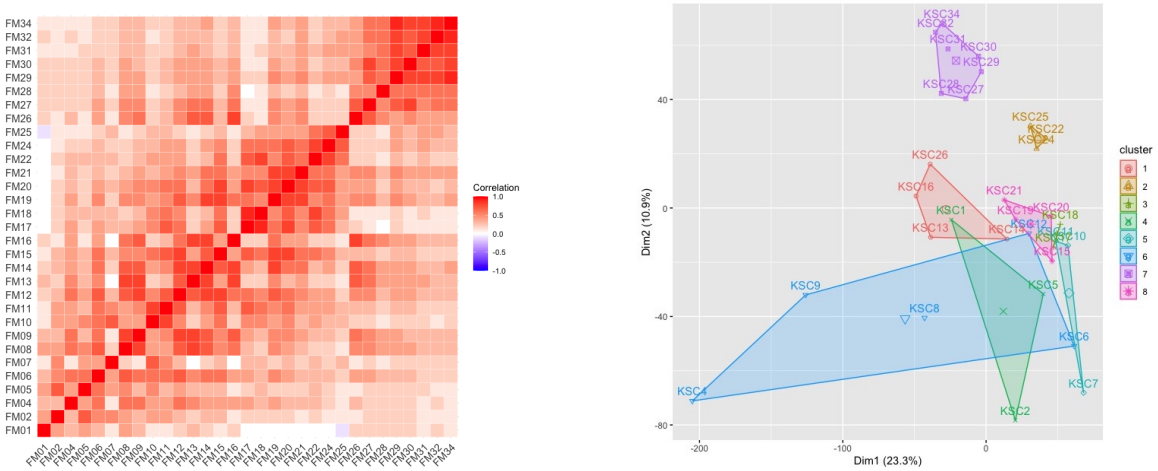
frequently moved about the launch complex, which geographically consists primarily of flat, open coastal land with few spots for refuge from lightning. For instance, rockets loaded with propellant may be moved from a staging area to a launch facility. A lightning strike to this piece of equipment would result in catastrophic loss of expensive equipment, not to mention the threat to flight line personnel. Although cloud-to-ground lightning is the main concern for most operations, flight line operations also require accurate forecasting of cloud-to-cloud lightning. This forecast of both cloud-to-cloud and cloud-to-ground lightning, or "total lightning", is required for both the onset of lightning activity for stop work safety concerns, as well as the cessation of lightning activity to allow personnel to safely return to work. Current policy consists of an "all clear" signal after a thirty minute period without lightning within a 10NM radius. Available literature suggests the aforementioned process is far too conservative and results in operational inefficiencies [76]. These inefficiencies are becoming increasingly problematic due to the growth of private space industry such as SpaceX, Blue Origin, and numerous start-ups such as Firefly Aerospace. This private industry growth has dramatically increased the utilization rates of Cape Canaveral launch facilities, making improved operational efficiencies increasingly important.

Cape Canaveral possesses a dense array of weather sensors that includes both traditional sensors and tailor-made systems such as the EFM network and LDAR system. Weather forecasters also use traditional weather measurements, a local weather radar (WSR-88D), National Lightning Detection Network (NLDN), and daily weather balloon launches. These sensor networks inform an operational warning system that manages ten warning regions spread throughout Cape Canaveral. These warning regions consist of 5NM or 6NM circles centered on key infrastructure locations, sometimes heavily overlapping [81]. Of note, three of the lightning warning circles do not geographically contain a EFM sensor to provide direct coverage, as seen in Figure 28.

The data produced by these sensor networks is inherently noisy and inappropriate for standard modeling approaches due to the complexity of movement of atmospheric electrostatic potential [44]. EFM data collection is further perturbed by dense networks of antennas, radar arrays, and other equipment and facilities supporting space launch and communication. There have been attempts made to mitigate these disturbances, such as ceasing collection of an individual EFM sensor if maintenance crews are mowing grass in the area, but these disturbances remain.

The LDAR detects both radar and flashes emitted by lightning to produce a 3D map of all lightning events within 54NM of Cape Canaveral [87]. The system was originally designed by NASA to meet their unique operational requirements that includes the ability to detect total lightning. The system has above a 90% correct detection rate out to 54NM, increasing to over 99% within 14NM of Cape Canaveral [87]. LDAR data contains timestamps for all detected lightning events, to include a detection range and azimuth from the system's central tower. The LDAR data are used in this study as the response for model training and evaluation.

EFM sensors measure the vertical electric potential of the atmosphere at ground level [44]. Each EFM site contains a series of vertically-oriented sensors that are covered and uncovered by a grounded rotor turning at 1800 rpm yielding a recorded measurement every 0.1 seconds [35]. The intent is to detect when an electrified cloud moves into the area that could signify an increased chance for lightning activity, characterized by a sudden change of polarity and increase of magnitude of the electric potential [48] [57] [58]. The original Cape Canaveral array used 34 EFM sensors. However, currently only 31 sensors remain in active service. The network of EFM sensors are currently used to inform decisions for both daily launch operations and incorporate into the Launch Pad Lightning Warning System (LPLWS), a legacy system used to inform operational decisions for space vehicle launch. A threshold value



(a) Correlation heatmap of EFM data

(b) K-means clustering of EFM sensors

Figure 13: A correlation heatmap of EFM data for 1-14 June 2013 shows predominantly positive relationships between all sensors roughly aligned with geographic location. Similarly, k-means clustering identifies groups of sensors primarily based upon geographic location.

approach is employed due to the chaotic nature of the measurements, where a lightning warning or launch delay is issued if a sensor reports a measurement that exceeds that predetermined threshold value.

Figure 13a displays a correlation heatmap between all sensors' data, showing a high degree of correlation amongst most of the EFM network. There are highly complex relationships between EFM sensors, not easily explained by geographic location. Nearly all correlations are positive, which makes sense as weather patterns should move mostly uniformly despite the large associated geographic area. Of note, sensor 'FM01' seems uniquely uncorrelated with a large number of the sensors. Figure 13b shows clustering analysis using k-means that results in eight identified clusters of EFM sensors, which seem to closely align geographically as seen in Figure 28. Normal clustering analysis procedures may be converging towards too many clusters due to a high degree of chaotic noise within the EFM data. Further research, such as applications in wavelet-enabled discriminant analysis, may better define relationships between sensors.

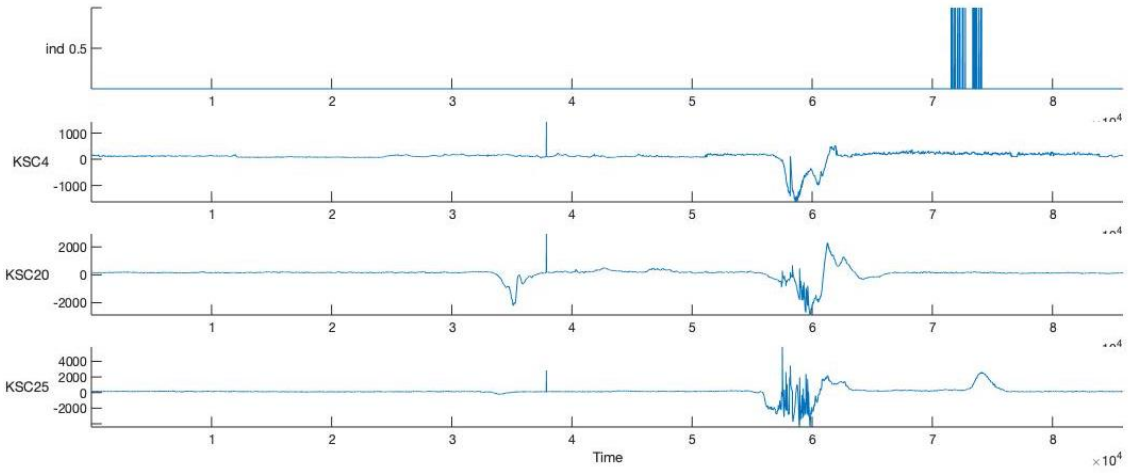


Figure 14: Detected lightning flashes as binary variable (top) for Central Cape warning circle and raw EFM data from three sensors showing predictive yet chaotic response, over time (seconds) for 22 May 2013.

There are several examples available of field mill data being used for lightning prediction both for KSC/CCSFS and international locations. Speranza [86] employed a neural network approach to attain a maximum 84% accuracy for KSC/CCSFS lightning prediction using both EFM data and surface weather measurements. Skrovan [84] also looked at the lightning prediction problem using regression models based upon EFM threshold values, concluding that EFM measurements too noisy to effectively fit a predictive model. Lucas et al. [56] examined the EFM data for time series components and characterized both a strong diurnal cycle as well as spatial variability between sensors. A series of studies looked at lightning warning systems based on EFM data in Spain and Medellín, Columbia [5] [6] [55]. These studies employed a reverse in field mill polarity and threshold values to successfully indicate an approaching thunderstorm.

The noisy and chaotic nature of EFM data complicates its direct application through traditional modeling [44]. Figure 14 provides an example of the behavior of three randomly selected EFM sensors seven hours prior to a detected lightning event. A signal is a reverse in electrostatic potential building within the atmosphere prior

to a storm, and the overall behavior of this building energy is quite erratic. The relationship between the electrostatic charge of the atmosphere and lightning is well documented; however, incorporating EFM readings into a model has found limited success. Krider [44] summarizes the results of a series of early studies that attempt to model lightning using linear regression and EFM measurements as predictors. Results indicate that any attempt to apply such a model produces an over-simplification as the electric charge within a thunderstorm is in constant change due to churning within the clouds and lightning discharge.

3.2 Background

This section provides a concise overview of the methods and techniques used in model formulation, to include wavelet methods, the semiparametric single index mode, and principal component analysis.

3.2.1 Wavelet Transforms

The wavelet transform is used for feature extraction and noise reduction. Wavelets model a function in time and frequency simultaneously by approximating functions at increasing levels of resolution expressed as a linear combination of scaling functions $\phi_{j,k}$ combined with the difference in approximations expressed as a linear combination of wavelets $\psi_{j,k}$ [69]. This is accomplished by projecting approximations of that function into a series of nested subspaces, each providing a different level of resolution in time. Wavelet functions represent a family of unique functions designed to be localized in time and frequency, typically defined as both a mother wavelet (ψ) and father wavelet (ϕ). Through dilation and translation operations, these wavelets produce an entire basis of wavelet functions [69]. These basis functions can be used to model a function in a Multiresolution Analysis (MRA) which consists of successively de-

tailed approximations of the function. Wavelets provide significant advantages over competing methods, such as the discrete Fourier transform and windowed Fourier transform. Wavelets localize frequency in time by adapting the size of their window of approximation to the frequency at each resolution level [69]. The result is a time to resolution level analysis method that optimizes the tradeoff between certainties in frequency and time across each nested and consecutive resolution level.

3.2.1.1 Discrete Wavelet Transform(DWT)

The DWT can be applied to discrete time series, resulting in an additive decomposition having constituent detailed time series ($\psi_{j,k}$) reflecting variations at resolution level j and a smoothed version of the time series ($\phi_{j,k}$) reflecting averages at resolution level j [73]. With wavelets defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{19}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{20}$$

a function of time can be represented as

$$f(t) = \sum_j \sum_k d_{j,k}\psi_{j,k} + \sum_k s_{j_0,k}\phi_{j_o,k} \tag{21}$$

where $s_{j,k} = \langle f, \phi_{j,k}\rangle$, $d_{j,k} = \langle f, \psi_{j,k}\rangle$, and $j, k \in \mathbb{Z}$. The time series is thus represented as a linear combination of the shifted and scaled versions of the wavelet functions as estimated using the wavelet coefficients $c_{j,k}$ and $d_{j,k}$. An important consequence of Equation 34 is the separation of the approximation and detailed representations of a signal.

Figure 15a, motivated by and adapted from presentations in the MATLAB Wavelet Toolbox [61], provides a rudimentary representation of a three-level, $j = 3$, DWT of a

signal X, where $X \in \mathbb{R}^N$. The levels D_1, D_2, and D_3 represent the detailed resolution levels whereas S_3 is representative of the smoothed approximation of the function. The decomposition results in a concatenation of these resolution levels into a single vector of wavelet coefficients $W \in \mathbb{R}^N$ the length of the original sample. The wavelet decomposition of a time series $X_t, t = 1, 2, \ldots, T$ is therefore

$$X_t = \sum_{k=1}^{j} D_k + S_j \tag{22}$$

where D_j is the wavelet detail coefficients at scale j and S_j are the smoothed coefficients.

In practice, execution of this transform employs a filter bank approach. This approach processes a signal using decimation, or downsampling by two, where every other value of the signal is removed. This reduces the size of the signal by half at every level of decomposition, resulting in a quick and highly efficient algorithm as every iteration requires half the number of calculations. The inverse implementation requires a similar filter bank approach governed by upsampling, or doubling the size of the sample by inserting zeros between every value.

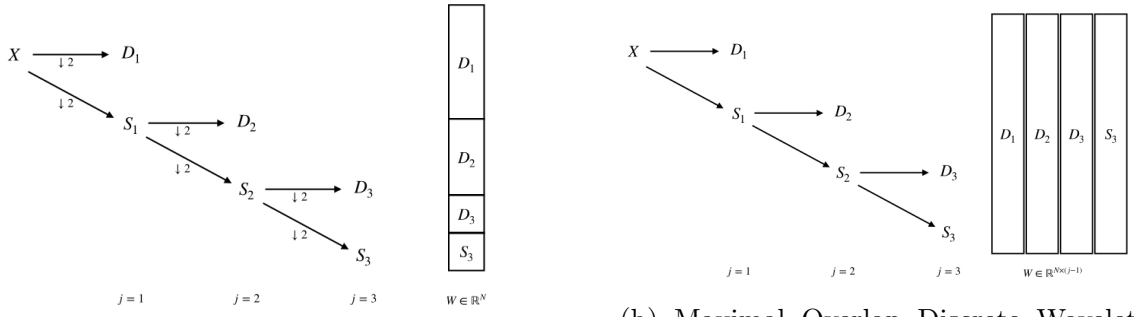
Although the DWT possesses many desirable attributes, it suffers from some limitations in time series applications. Primary issues of note are that the filter bank estimation method of the DWT requires a signal sample size of dyadic length, or an integer multiple of 2^j and the DWT is not shift invariant. As such, the values of the details and smooths do not shift with the values of the original signal. The inverse DWT can accordingly give a different reconstruction compared to the original time series even when accounting for the shifts. Finally, the DWT requires a periodicity assumption in the signal. For non-stationary time series, this means that the DWT transform is highly dependent upon when the time series is sampled. Significant changes in the time series across the sample will result in significant boundary

effects.

3.2.1.2 Maximal Overlap Discrete Wavelet Transform (MODWT)

The maximal overlap discrete wavelet transform (MODWT) is a modified version of the DWT better suited for applications like time series analysis. This particular transform is found throughout the wavelet literature under different names, such as undecimated DWT, shift invariant DWT, wavelet frames, translation invariant DWT, stationary DWT, time invariant DWT, and non-decimated DWT [73]. This research adopts the use of MODWT as in Percival and Walden [73] due to their thorough and foundational work in applying wavelets to time series. Essentially, the MODWT does not include downsampling as in the DWT and thus uses all values of the original signal at every level of decomposition.

MODWT provides some key advantages over the DWT. It is well defined for all sample sizes. The decimated DWT requires a sample of dyadic length, complicating its use in time series applications. The MODWT is shift invariant, meaning each level of decomposed coefficients aligns with the original time series. The MODWT also avoids boundary effects found in the decimated wavelet transforms. The MODWT does not downsample at each level, meaning each resolution level contains the same



(a) Discrete Wavelet Transform (DWT)

(b) Maximal Overlap Discrete Wavelet Transform (MODWT)

Figure 15: Depiction of three-level DWT and MOWDT decompositions of signal X to wavelet coefficients W

number of coefficients as the original sample. This produces a redundant but higher resolution at coarser levels compared to the decimated wavelet transforms.

These advantages are not without costs. A notable cost is that the transform is highly redundant and loses orthogonality. This results in dependencies between the empirical coefficients of the scaling function and wavelets. The details and smooth resolution level of the MODWT each contain the same number of samples as the original signal. Although this gives a finer resolution at each level, it results in the number of required computations $\mathcal{O}(N \log_2 N)$ or a cost of $\mathcal{O}(\log_2 N)$ when compared to the DWT.

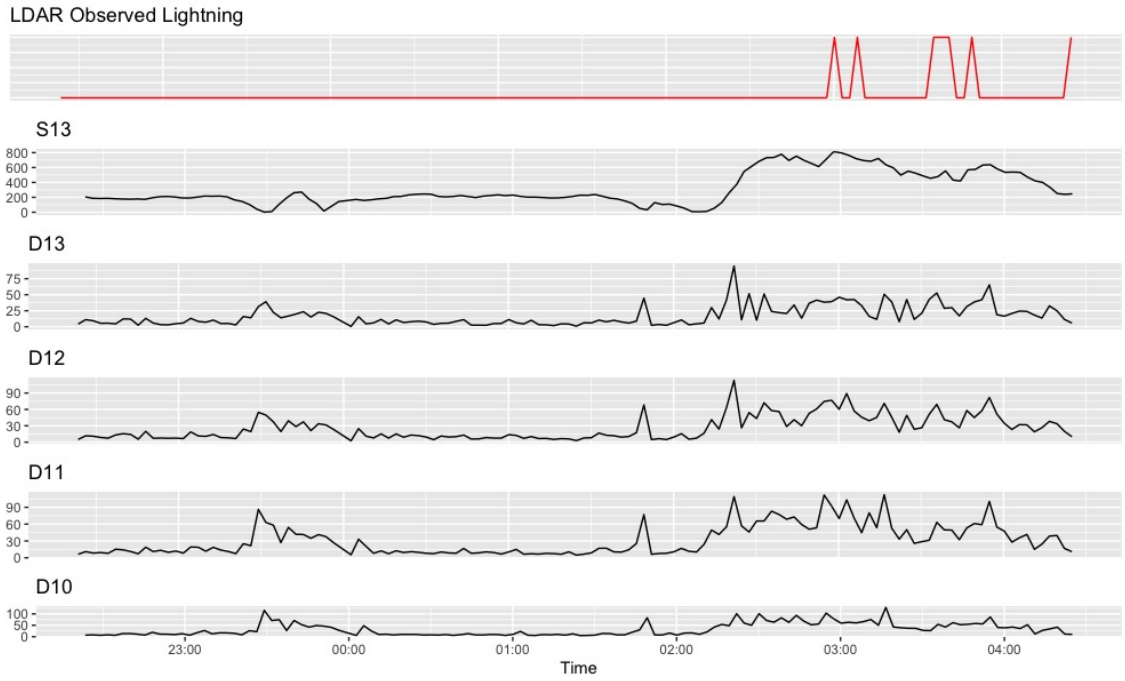


Figure 16: LDAR observed lightning (red) and wavelet coefficients of a 13 level MODWT of EFM sensor FM7 for 1 June 2013

The MODWT can be analyzed using a MRA just as in the DWT. Figure 16 provides the top levels of a 13-level MODWT of some EFM data from 1 June 2013. The transform is clearly shift invariant, as the perturbations in the details align perfectly with the sharp and abrupt changes in the LDAR data. This includes some movement

in wavelet coefficients several hours prior to the lightning activity, indicating that these high level detail levels may be predictive of lightning activity. Unlike the progressively coarse levels of the DWT, due to downsampling, the density of coefficients in the higher levels of MODWT detail remain identical to the original sample.

3.2.1.3 Wavelet Thresholding

Wavelet thresholding is a dimension reduction and de-noising method that manipulates the transformed wavelet coefficients. This section introduces thresholding using a brief discussion on the sparsity of the wavelet representation, followed by both universal and adaptive thresholding techniques.

The wavelet transformation results in a sparsity of effects, where most of the key features of a signal are captured and represented by only a few coefficients. These coefficients can be manipulated to reduce or remove stochastic noise while the power of the true signal is retained in only a few significant coefficients. Therefore, the ability of wavelet methods to model a signal in frequency and time simultaneously grants a powerful ability to capture and isolate signals of random noise. Manipulation of the coefficients to reduce or remove random noise is known as thresholding, which can be applied globally to the entire set of coefficients or adaptively applied using localized rules. Unless otherwise stated, thresholding methods require the assumption of normally distributed observational errors.

Global thresholding uses a single threshold value λ applied uniformly to all or nearly all coefficients of the wavelet transform. Consider for a given threshold value λ and set

$$\hat{f}_\lambda(t) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} I_{\{|d_{j,k}|>\lambda\}} d_{j,k} \psi_{j,k}(t) \tag{23}$$

where I represents the indicator function [69]. This representation of "keep or kill" is known as hard (H) thresholding, where any value less than or equal to the given

value of λ is set to zero. This enforces sparsity in the wavelet coefficients, resulting in maintaining only those coefficients significant for representing the original signal. An inverse wavelet transform is then applied to recreate the original signal with random noise removed. Then, defining the thresholded coefficients as

$$\hat{d}_{j,k} = \delta_\lambda(d_{j,k}) \tag{24}$$

allows for reexpression of the hard (H) thresholding rules as

$$\delta_\lambda^H(x) = \begin{cases} x & \text{if } |x| > \lambda \\ 0 & \text{otherwise} \end{cases}. \tag{25}$$

Donoho and Johnstone [16] propose an alternative method of soft (S) thresholding defined as

$$\delta_\lambda^S(x) = \begin{cases} x - \lambda & \text{if } x > \lambda \\ 0 & \text{if } |x| \leq \lambda \\ x + \lambda & \text{if } x < -\lambda \end{cases}. \tag{26}$$

Similar to hard thresholding, only wavelet coefficients greater than a threshold are kept, however their value is shrunk closer towards zero by an amount equal to the threshold λ [69].

These two methods are widely applied in current applications as a dimension reduction method. However, proper choice of the threshold value remains subjective, based upon an assessed tradeoff between over- and under-smoothing the function. Furthermore, these universal methods may underperform adaptive techniques in large sample sizes. Donoho and Johnstone [16] propose two universal thresholds, the first of which is

$$\lambda = \sqrt{2\sigma^2 \log(N)} \tag{27}$$

to be used when the variance of the original signal (σ^2) is known. This method, commonly referred to as VisuShrink, is a computationally efficient method that can be applied through either soft or hard thresholding techniques.

As an example of thresholding, Figure 17 displays annual Nile River minima measured from 622-1284 A.D. [73]. The raw time series is in blue, and a reconstructed time series following a MODWT and soft thresholding is in red. Thresholding the function has effectively smoothed the response, an action that may allow easier interpretation and implementation into a hybrid model that requires convergence. However, oversmoothing this function could eliminate some sharp and abrupt changes that may be the signal of interest. Careful application of thresholding is required dependent upon application.

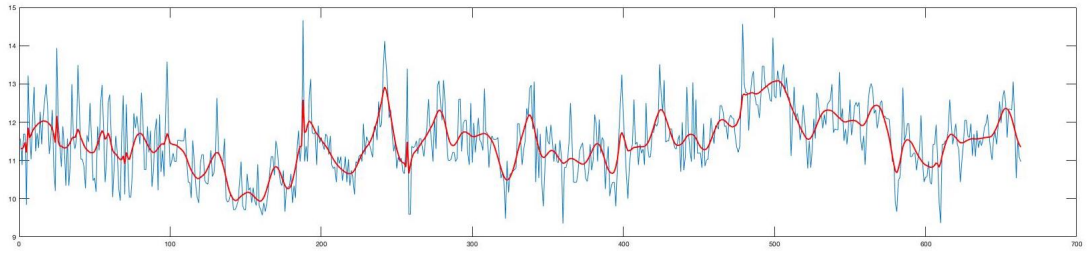


Figure 17: Annual Nile River minima 622-1284 A.D. (blue) [73] and values of wavelet approximated smoothed function (red)

Data adaptive techniques attempt to improve upon global techniques by varying the threshold within the decomposition. Donoho and Johnstone [15] present SureShrink as an extension of VisuShrink, combining a level-dependent thresholding technique with Stein's unbiased risk estimator (SURE) [88]. This method is more computationally demanding compared to VisuShrink, but is shown to reduce the mean squared error in estimation.

Cai [9] introduces block thresholding, where a thresholding δ is determined using groups of coefficients within a resolution level. An important note is that both VisuShrink and SureShrink assume normality of the errors. McGinnity et al. [63] propose

a nonparametric method of block thresholding that does not require this normality assumption.

3.2.2 Principal Component Analysis

The dataset of wavelet coefficients and LDAR observed lightning events are divided into training and testing datasets. Stratified samples are developed from the training data, where only those observations three hours prior to each lightning event are selected for model estimation. This helps improve precision of the random sampling and facilitate improved convergence of estimated model parameters. The resulting data consist of highly redundant MODWT resolution levels that exhibit a high degree of multicollinearity. A Principal Component Analysis (PCA) is applied to these resolution levels to produce orthogonal principal components. Only the principal components that describe 99% of the variance are retained, resulting in significant dimension reduction. This is particularly helpful for convergence of the single-index model.

The MODWT results in an additive decomposition consisting of redundant resolution levels. In certain applications, such as found with EFM sensors, this further results in a high degree of multicollinearity amongst the highly correlated sensors. This multicollinearity can cause erratic changes in model estimates and result in significant issues with numerical estimation methods. PCA re-express these multicollinear vectors as orthogonal index vectors, while simultaneously providing the opportunity for overall dimension reduction.

PCA finds combinations of the p EFM sensors across all j sensors to produce uncorrelated indices $Z_1, Z_2, \ldots, Z_p$ known as principal components [60]. The procedure results in a ranked order of indices by relative importance in contribution towards explaining overall variance in the data $\text{Var}(Z_1) \geq \text{Var}(Z_2) \geq \cdots \geq \text{Var}(Z_p)$ [60].

High levels of multicollinearity within the original data result in much of the variance being represented by relatively few principal components. This may not be the case in datasets that are not highly correlated.

3.2.3 Semiparametric Single-Index Models

The single-index model (SIM) combines the lack of distributional assumptions of nonparametric methods with the dimension reduction capabilities of parametric methods. The first attribute avoids any issue of distribution misspecification resulting in model inconsistency. Some basic assumptions of linearity in the index, resulting in a semiparametric method, significantly reduces the negative impacts of dimensionality in nonparametric methods. However, these benefits do not come without a cost as the SIM requires estimation of both a parameter vector and link function. The SIM is a generalization of many popular parametric models such as normal regression, logit, probit, and Tobit [28]. The following provides a formal definition of the SIM to include identification requirements, dominant methods for estimation, and how this approach is implemented in the forecasting of lightning onset.

3.2.3.1 Defining the Single-Index Model

The SIM is well documented and a general presentation can be found in Li and Racine [47], Härdle et al. [26], and Henderson and Parmeter [28]. The most general form of the SIM is

$$y_i = g(\varphi(\mathbf{x}_i, \beta)) + u_i \quad i = 1, 2, \ldots, n \tag{28}$$

where $g(\cdot)$ is an unknown smooth function, $\varphi(\cdot, \cdot)$ is a known parametric function with regressors $\mathbf{x}$ and parameter vector β, and the additive error term u is uncorrelated and independent [28]. The dependent variable y_i can be either continuous or discrete, although some applications restrict y_i to be a binary variable. The function $\varphi(\cdot, \cdot)$ is

not required to be linear, however Henderson and Parmeter [28] state that linearity is commonly assumed. This results in $\varphi(\cdot,\cdot)$ being equivalent to

$$\varphi(\mathbf{x}_i, \beta) = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_q x_{qi}$$

where the model contains an equal number of regressors p and parameters q. The result is a semiparametric model where the linearity of $\varphi(\cdot,\cdot)$ is specified while the form of $g(\cdot)$ is unspecified. The most common representation of the SIM is in matrix form

$$Y = g(\mathbf{X}'\beta_0) + u \tag{29}$$

where Y is the dependent variable, $\mathbf{X} \in \mathbb{R}^q$ is the vector of explanatory variables, β_0 is the $q \times 1$ vector of unknown parameters, and u is the additive error term uncorrelated with the index where $E[u|X] = 0$ [47]. The model derives its name from the scalar value for $\mathbf{X}'\beta_0$ that provides a "single index" even though $\mathbf{X}$ is a vector.

3.2.3.2 Identification Conditions

Certain restrictions are imposed on the index vector to estimate the SIM. As in linear regression, the $\mathbf{X}$ matrix can not be singular [47]. The unknown function $g(\cdot)$ must be differentiable and can not be the constant function, otherwise β_0 can never be identified [47]. Furthermore, $g(\cdot)$ is assumed monotonically increasing to identify bounds on β_0 [47].

The vector of explanatory variables must have at least one continuous component, and varying the discrete components will not divide the support of $\mathbf{X}'\beta_0$ into disjoint subsets [47]. The use of discrete variables is especially important when determining the estimation method to be employed. For instance, average derivative estimation (ADE) is a non-iterative method that solves many of the computational

hurdles of competing methods such as Ichimura [33] and Klein and Spady [43]. The iterative methods require significant computational cost in terms of n nonparametric regressions with every evaluation of the objective function while also being susceptible to nonlinear local minima and saddle points [28]. The non-iterative ADE shows improved performance in this regard, however requires continuous variables in calculation of the gradients. Horowitz [31] details some methods of evaluating discrete and continuous variables separately, however the use of discrete variables should be closely examined in any implementation of SIM.

As specified above, the parameter β_0 is only identifiable up to a scale [47]. To demonstrate this, given two constants α_1 and α_2, any $g(\cdot)$, and a fixed β, another function $g_2(\cdot)$ can always be identified where $g_2(\alpha_1+\alpha_2\mathbf{X}'\beta) = g(\mathbf{X}'\beta)$ [47]. Therefore, β_0 is not identifiable without some kind of restriction to the index vector, otherwise known as normalization. Normalizations of the index vector ensure that β_0 can be identified in location and scale. A common location normalization is to restrict the vector of explanatory variables $\mathbf{X}$ to not contain a constant, meaning the parameter vector β_0 does not contain an intercept or location parameter [47]. Popular scale normalizations include normalizing the vector β to unit length, $\|\beta\| = 1$, or assuming the first component of $\mathbf{X}$ is both continuous and a unit coefficient [47]. Note that coefficients for two SIMs can only be compared if the same normalization is applied in both models [26].

3.2.3.3 Estimation Procedures

Estimating the SIM is complicated by both β_0 and the link function $g(\cdot)$ being unknown, making direct estimation impossible. If β_0 were known, then the model would simply become a univariate regression problem. If $g(\cdot)$ were known it would become a standard nonlinear regression problem to estimate β_0 [47]. Härdle et al.

[26] present the following general algorithm for estimating a SIM:

1. Estimate β_0 by $\hat{\beta}$
2. Compute index values $\hat{\eta} = \mathbf{X}'\hat{\beta}$
3. Estimate the link function $g(\cdot)$ by using a univariate nonparametric method for the regression of Y on $\hat{\eta}$

Estimating the link function $g(\cdot)$ is relevant for most common estimation procedures, to include all procedures mentioned here. Therefore, estimation of $\hat{\beta}$ is of primary concern and can be accomplished using iterative methods such as semiparametric least squares (SLS) and psuedo maximum likelihood estimation (PMLE). Both SLS and PMLE focus on estimating β_0 through use of an objective function that achieves convergence at the $\sqrt{n}$ parametric rate. Nonparametric estimates of $\hat{\beta}$ or $\hat{g}(\cdot)$ are used in the objective function, resulting in a complicated and non-trivial estimation procedure. The objective function is not guaranteed to converge nor is it guaranteed to converge to a unique global optimum. As a result, most methods employ numerous random starts.

Ichimura [33] introduced methods using both SLS and a weighted version (WSLS) which propose estimating $g(\mathbf{X}'\beta_0)$ by the leave-one-out nonparametric kernel estimator

$$\hat{G}_{-i}(X_i'\beta) \equiv \hat{E}_{-i}(Y_i|X_i'\beta) = \frac{(nh)^{-i}\sum_{j=1,j\neq n}^{n} Y_j K\left(\frac{X_j'\beta - X_i'\beta}{h}\right)}{\hat{p}_{-i}(X_i'\beta)} \tag{30}$$

where h denotes the bandwidth. This method is effective for both a continuous and binary response variable. The denominator is defined as

$$\hat{p}_{-i}(X_i'\beta) = (nh)^{-1}\sum_{j=1,j\neq i}^{n} K\left(\frac{X_j'\beta - X_i'\beta}{h}\right)$$

but is problematic as it is random. Ichimura [33] compensates for this through the use of a trimming function to make the denominator positive and relatively large with high probability to aid in uniform convergence. In the WSLS, the result is to estimate β_0 by minimizing the objective function

$$S_n(\beta_0) = \sum_{i=1}^{n} \left[Y_i - \hat{G}_{-i}(X_i'\beta) \right]^2 w(X_i)\mathbf{1}(X_i \in A_n) \tag{31}$$

where $\hat{G}_{-i}(X_i'\beta)$ is estimated from Equation 30, $w(X_i)$ is a non-negative weight function, and $\mathbf{1}(\cdot)$ is an indicator function. The result is an unbiased estimator $\hat{\beta}$ that converges at the parametric $\sqrt{n}$ rate.

3.3 Methodology

While there has been some very good work in the area of lightning prediction, the current suite of methods are still somewhat lacking. A new methodology for lightning prediction is outlined in Figure 18 and is comprised of three phases: data preparation, model development, and evaluation. The following subsections describe the methodology in detail.

3.3.1 Data Preparation

The raw EFM data consists of very large data frames of EFM measurements in time series collected at a 50Hz rate. Sensors periodically miss large sections of collection, either due to routine maintenance or due to local disturbances to the individual sensor such as mowing activities. The METARs data frame contains hourly weather measurements from regional airports and weather stations, as seen in Figure 12b. Both EFM and METARs data frames are summarized to the minute and mean imputation is used to complete the datasets, where the missing values are assigned

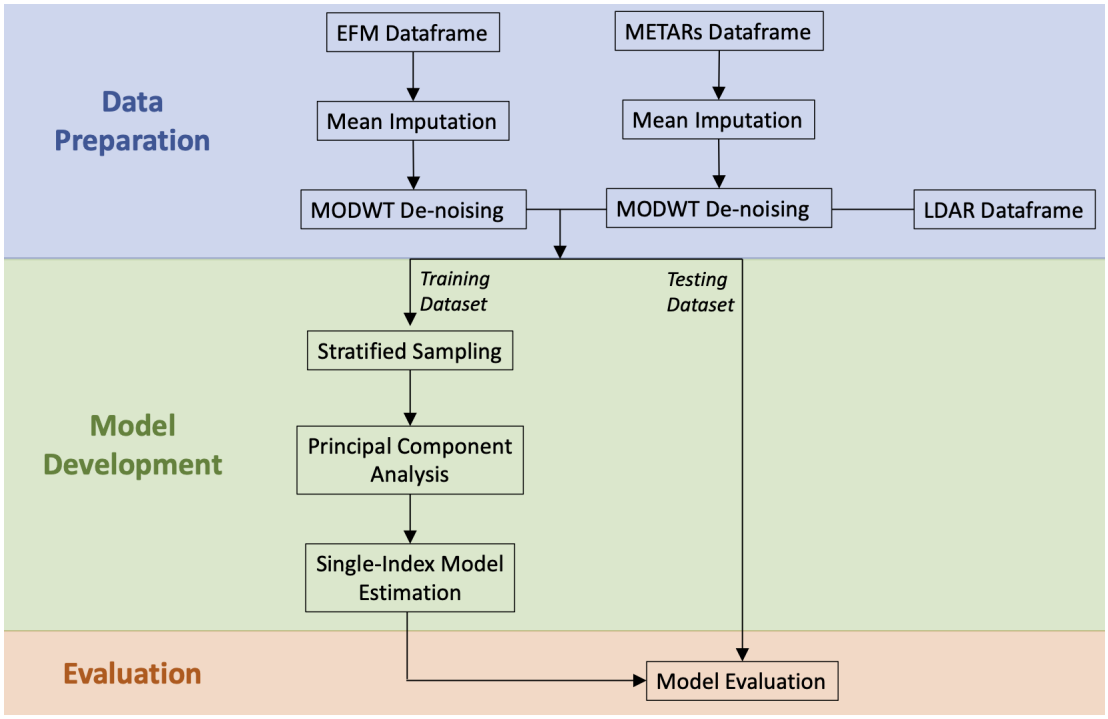


Figure 18: Outline of methodology. The original dataset is partitioned so the first third becomes a training set, with the rest of the data used as a testing dataset.

the average value of the series. A MODWT is applied to the modified and completed dataset using the R package "waveslim" [95]. The decomposed wavelet coefficients are modified by a wavelet thresholding technique, and an inverse MODWT is used to reconstruct the de-noised data. The resulting time series of EFM measurements contains far less chaotic noise that facilitates convergence in a hybrid model.

The LDAR dataset consists of timestamped observations of each detected lightning event, to include three-dimensional coordinates and distance of the lightning relative to the main LDAR tower in central KSC/CCSFS. This system detects lightning up to 54NM, and so detected events range from the Tampa region in the west to the Atlantic in the east. Geodesic distances assuming a spherical earth are calculated for each LDAR detected lightning event to the central point of the desired lightning warning circle and observations are filtered to those that fall within the desired lightning warning circle. The LDAR data frame is then joined with the EFM and METARs

data by time stamp to produce a final dataset for use in model development and evaluation.

3.3.2 Model Development

The dataset is divided so that the first third is used for model training and the remainder for testing. Stratified sampling is applied to the training dataset, reducing the overall size to only those time periods around the relatively rare occurrence of lightning. The stratified sampling greatly reduces the overall computational requirements for convergence in SIM estimation. A PCA is applied to both the EFM and METARs data individually, and resulting index vectors are combined into a single data frame. The SIM is estimated using the "np" package in R [27]. This estimation package automatically selects the method of Ichimura [33] for a continuous response or the method of Klein and Spady [43] for a binary response. Both of the aforementioned methods normalize the first element to one and jointly estimate model parameters and bandwidth. The method also requires at least one continuous variable, which is satisfied in application with EFM and METARs data.

Despite steps such as dimension reduction, estimation of a SIM remains computationally demanding within multivariate applications. Computational methods such as parallel processing can speed up estimation routines. However, there is no method to reduce the overall requirement for computational resources for parameter estimation. Furthermore, estimation methods use nonlinear optimization routines that can result in convergence to localized minima. Multiple estimation routines, or multistarts, are required to overcome convergence issues. Reduced multistarts and relaxed relative convergence tolerance can be used for faster convergence during data exploration [27].

3.3.3 Model Evaluation

The resulting SIM is evaluated by using the predicted response generated from the training dataset. The training PCA indices are used to estimate a complimentary set of testing PCA indices. A single PCA of the overall dataset is not used as it would contaminate the testing dataset with information from the training. Likewise, a PCA of the testing dataset alone is not used as the orthogonal indices will not align with those used to train the model. The predicted response from the single-index model is then evaluated against real-world LDAR observed lightning to produce a confusion matrix describing model accuracy.

3.4 Analysis and Results

3.4.1 Designed Experiment

The flexible nature of wavelet methods allows for a wide variety of application. However, this same flexibility results in a large number of tunable parameters which complicate implementation. As such, wavelet techniques can require some investment of time to delve into and understand how to best pair a wavelet method for a particular application. Without existing literature on how to apply wavelet methods to EFM data, a series of designed experiments are conducted to guide and inform model formulation. This approach efficiently explores individual factors and interactions to conserve required time and preserve resources required due to the computational requirements of multivariate single-index model estimation. The intent of these screening experiments is to explore the complex factor space and gain an understanding of potential model performance using this methodology derived exclusively from EFM data. The results of these experiments guide the development of research extensions to improve upon a model not limited only to EFM inputs. This section presents

concise summaries of these experiments followed by a brief discussion of some of the challenges and lessons learned.

3.4.1.1 Experimental Design

Development of the first screening experiment is complicated by a lack of existing knowledge of the range of values for factors and their effect upon a lightning forecast. The results of some initial "one factor at a time" experimentation develop factors and levels, shown in Table 6. Factor A is the size of window used to produce stratified sampling in the training data, where larger window size correspond to longer periods of EFM data used prior to each observed lightning event. Values are selected based upon observation of perturbations in EFM readings one to four hours prior to a lightning event, as seen in Figure 14. Factor B is the total percent of variance retained following the PCA, where higher degree of variance results in more PCA indices used for model formulation. Factor C is a two-level categorical denoting use of the entire EFM sensor network or restriction of the data to only EFM sensors within the particular lightning warning circle under evaluation. Factors D and E denote the thresholding method employed to manipulate the wavelet coefficients, resulting in smoothed EFM data.

The experimental design produced is a fractional factorial design using the design of experiments (DOE) tool in JMP. The design chosen is a 16 run 2_V^{5-1} fractional factorial design where no main effects or two-factor interactions are aliased with any

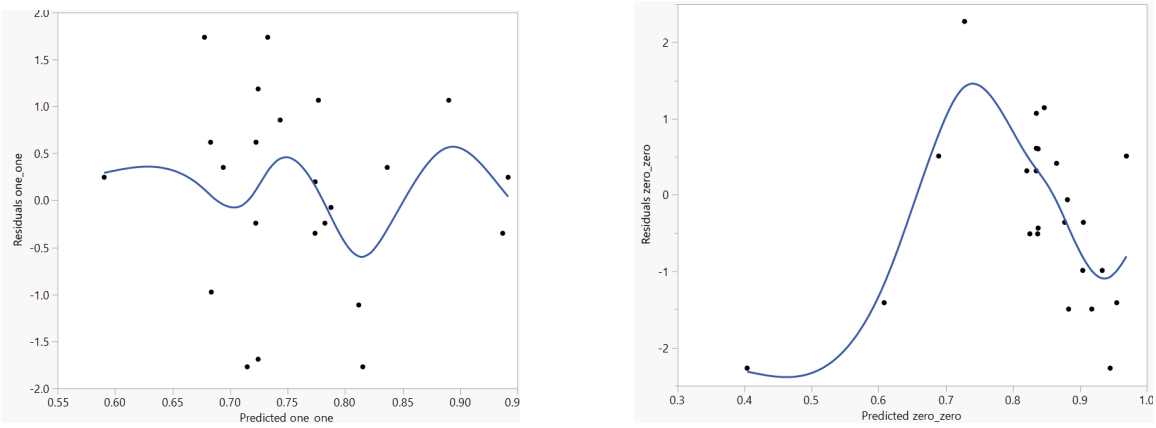
Factor	**Name**	**Low Level**	**Center**	**High Level**
A	Sample Stratification Window Size	45 minutes	2 hours	3.25 hours
B	Percent of PCA variance	85%	92%	99%
C	Localization of Sensors	Off		On
D	Thresholding approach	Universal		SURE
E	Thresholding method	Hard		Soft

Table 6: Factors and levels for screening experiment

other main effects or two-factor interactions [66]. Ten center point runs are included to assess for curvature in the response functions.

Several models are fit using the results of the first experiment, producing Figures 19a and 19b that plot predicted responses against residuals. Figure 19b displays a slight curvature in shape that could be indicative of curvature in the response. A pair of t-tests for curvature indicates a lack of evidence for curvature in the true positive response, but shows sufficient evidence to conclude curvature in the true negative response to the .05 confidence level. The original design is augmented with eight additional experimental runs, using D optimality, to include polynomial effects as necessary. Figure 20 compares the fraction of design space for each design, showing very reasonable behavior in the prediction variance even in the original design.

Figure 21 provides the color maps of correlations between both the original and augmented design. The inclusion of eight additional runs in the augmentation allows for estimation of polynomial effects for continuous Factors A and B. The augmented design includes some increased aliasing between factors and interactions, however the impact is acceptable and the design remains near-orthogonal, reducing the standard



(a) True positive (b) True Negative

Figure 19: Plot of residuals against predicted values for both true positive and true negative rates in the original design. The plots show a generally curved pattern indicative of possible curvature within the factors.

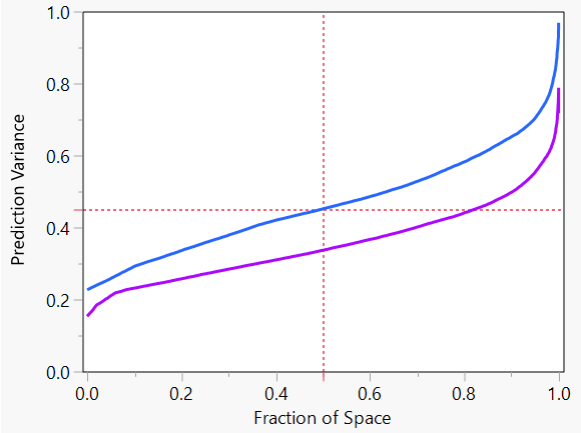


Figure 20: Fraction of design space plots for the original design (blue) and augmented design (purple). Both designs indicate a very reasonable behavior in the prediction variance across the design space.

error of estimates in resulting models.

Conducting a series of screening experiments proved to be a highly effective approach to inform decisions regarding the complexity of the wavelet model. The structure of the experimental design allowed for an efficient exploration of a complex factor space and provided valuable insight into the model. The approach provided an easy and efficient framework to guide experimentation and explore factor space, especially when compared to "one factor at a time" approaches. Initial model runs indicated several factors may be significant, but were quickly shown to have little or no impact to model effectiveness. The designed experiment approach proved effective in development of an effective training response, possibly one of the biggest challenges in the formulation of a EFM-only approach to lightning modeling. One improvement to this approach would be to use a custom design versus a classical experimental design. Although these modern custom designs can include more complicated alias structures, such techniques allow for a more targeted approach to focus experimentation on significant factors using fewer runs.

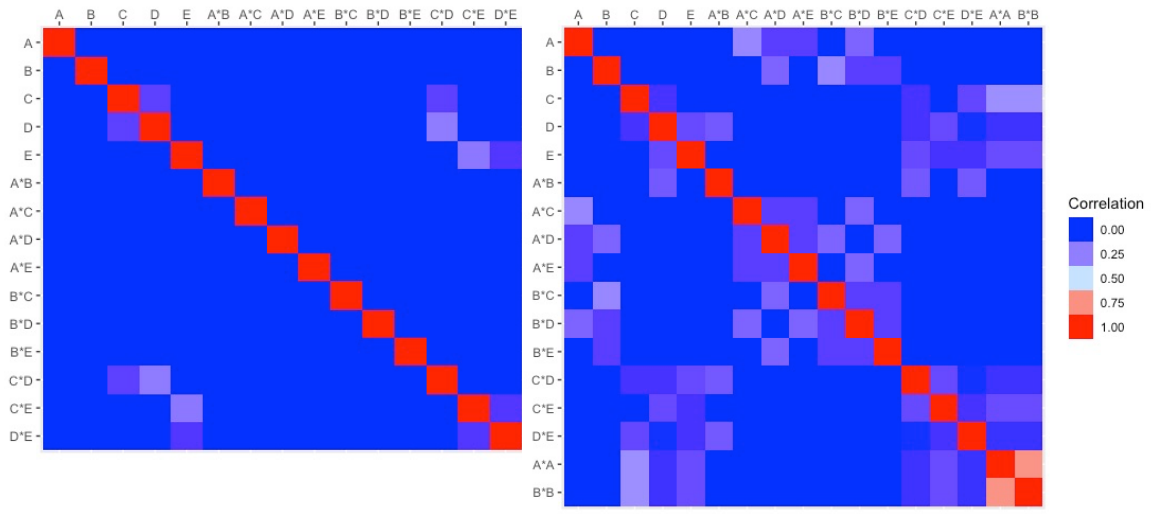


Figure 21: Color maps of the absolute value of correlations derived from the design matrices of the initial screening design (left) and augmentation (right). The inclusion of eight additional runs for estimation of polynomial effects in the augmented design results in some partial aliasing within the design. However, the impacts are acceptable and result in near-orthogonality between main effects, two-factor interactions, and polynomial factors. The near-orthogonality of these designs reduces the standard error of estimates in the resulting models.

3.4.2 Results

Results from the experimental design are analyzed using desirability functions within the JMP prediction profiler. Each factor is considered in producing factor levels for the best anticipated performance, as seen in Figure 22. Optimal parameters include the highest levels of PCA variance and stratified sample size. The model performed best through use of the entire sensor network, versus those geographically close to the Central Cape warning circle. The parameters for wavelet smoothing of the EFM data are selected as hard thresholding using the SURE technique. A final lightning prediction model is formulated using these factor levels to evaluate the overall performance of this modeling approach.

Figure 23 provides an overall picture of model predictive performance for 20 to 30 June 2013. This time period is a ten-day selection of time within the test dataset

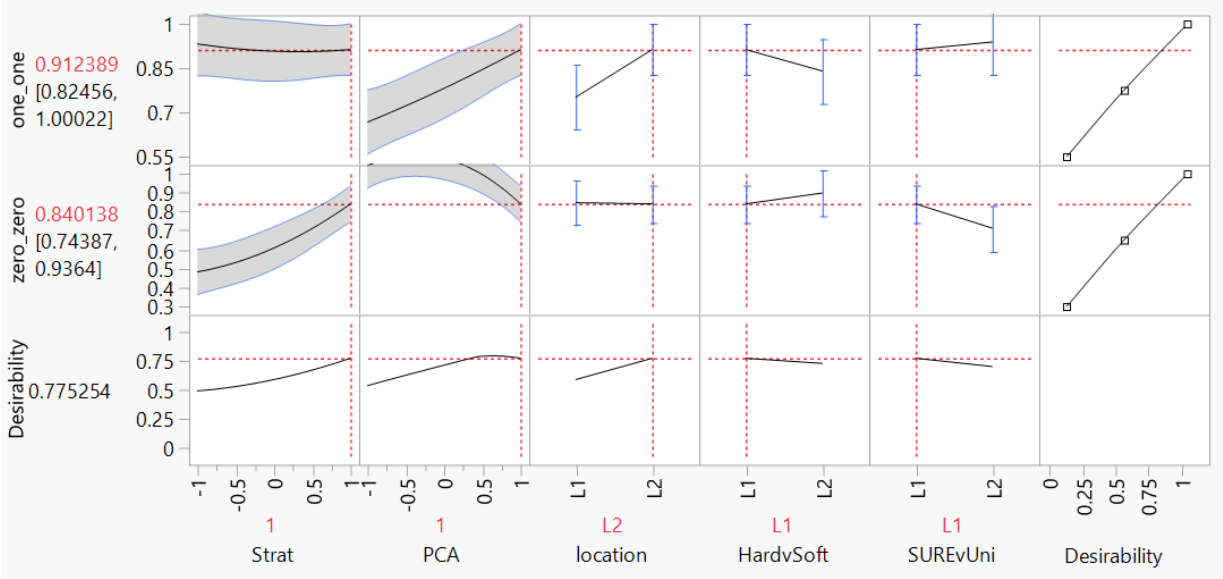


Figure 22: Prediction profiler in JMP for selection of factor levels to model performance based upon the responses "one-one", positive lightning identification, and "zero-zero", positive identification of lightning absence.

for which the EFM network remains fully operational, providing the best indicators of model performance under ideal circumstances. At this high resolution, the model produces a predictive response for all but one lightning event. Furthermore, there only appears to be one significant false alarm event.

Overall model performance is summarized by the confusion matrix in Table 7 corresponding to a 60 minute lightning warning period for the Central Cape lightning warning circle. The model's predictive response, $y_{pred} \in [0, 1]$, is assessed based upon performance in the testing dataset, resulting in a threshold value of 0.98 is chosen as a triggering event for lightning prediction. This means the model predicts lightning within the next 60 minutes if $y_{pred} \geq 0.98$. Although competing models from the experiment out-performed in certain measures, this model represents the best tradeoff between correct lightning identification against an acceptable false positive rate. The correct identification rate not only informs of approaching lightning threat, but also lowers the prevalence of false alarms. In the case of space launch, false alarms can

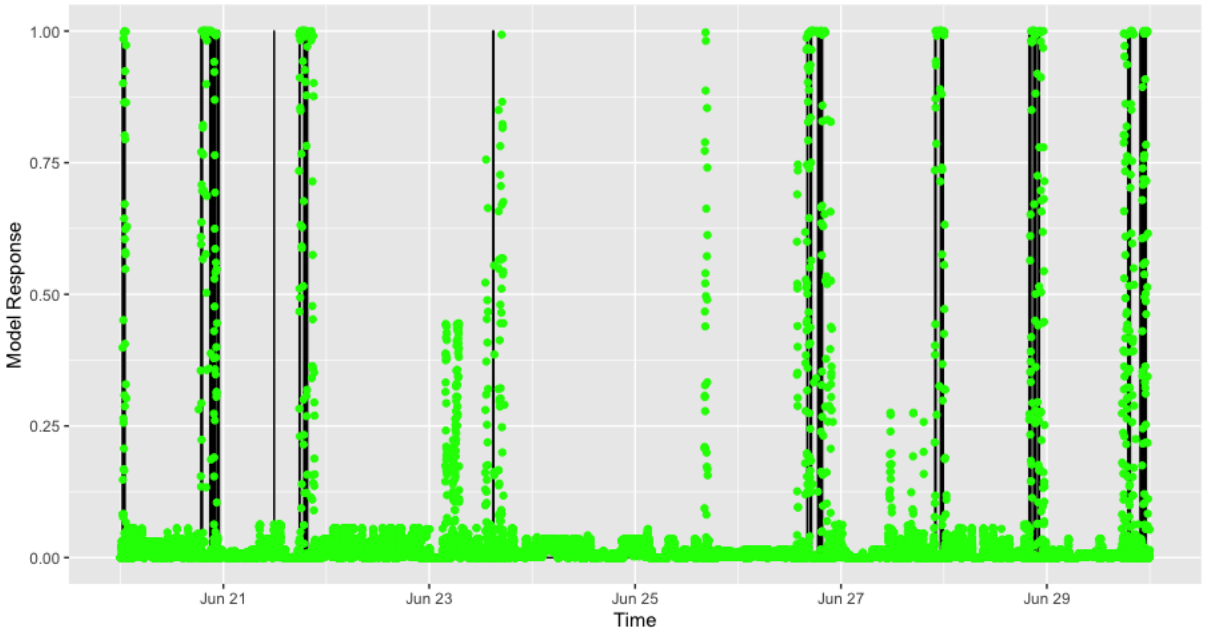


Figure 23: Plot of overall model fit of predicted response (green) to LDAR observed lightning within the Central Cape lightning warning circle (black) for 20-30 June 2013.

result in significant money waste.

		Observed	
		0	1
Predicted	0	12,624/13,815 91.38%	1,191/13,815 8.62%
	1	26/586 4.44%	560/586 95.56%

Table 7: Confusion matrix for model predictions 60 minutes prior to any observed lightning within the Central Cape lightning warning circle, 20-30 June 2013. A prediction or observed value of "0" corresponds to no lightning, whereas a "1" denotes LDAR observed lightning within the lightning warning circle. Results indicate significant improvements to existing models, with 95% accuracy in correctly identifying lightning within the Central Cape warning circle in the next hour and 91% accuracy in identifying the absence of lightning.

Figure 24 plots the performance of individual experimental runs in regards to

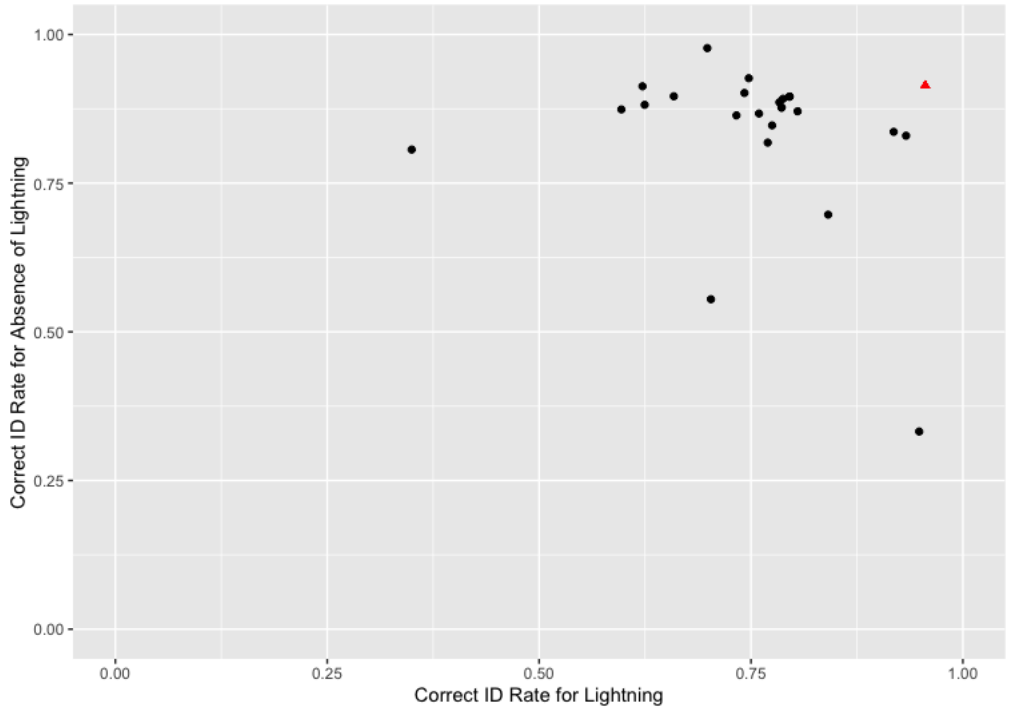


Figure 24: Plot of results for each experimental run in the designed experiment (black circles) and the results of the optimal formulation (red triangle). The results indicate a wide variation in model performance given varying experimental treatments.

overall accuracy in predicting lightning and predicting the absence of lightning. The results of the optimal formulation identified through the designed experiment is included for reference. The results indicate a wide range of trade-offs in model performance given varying experimental treatments. Although the model provides decent performance in many formulations, there can be large discrepancies in performance that would be very difficult and costly to explore using simple trial and error. This underlines the efficiency and effectiveness of the DOE approach in efficiently identifying the optimal formulation.

Table 9 provides results from competing naïve models using the same time window as above, 20-30 June 2013. First, model accuracy is assessed 24 hours prior to any observed lightning to assess if the model is reacting solely to diurnal variation. A 24 hour prediction offset results in a naïve model, akin to guessing lightning perfor-

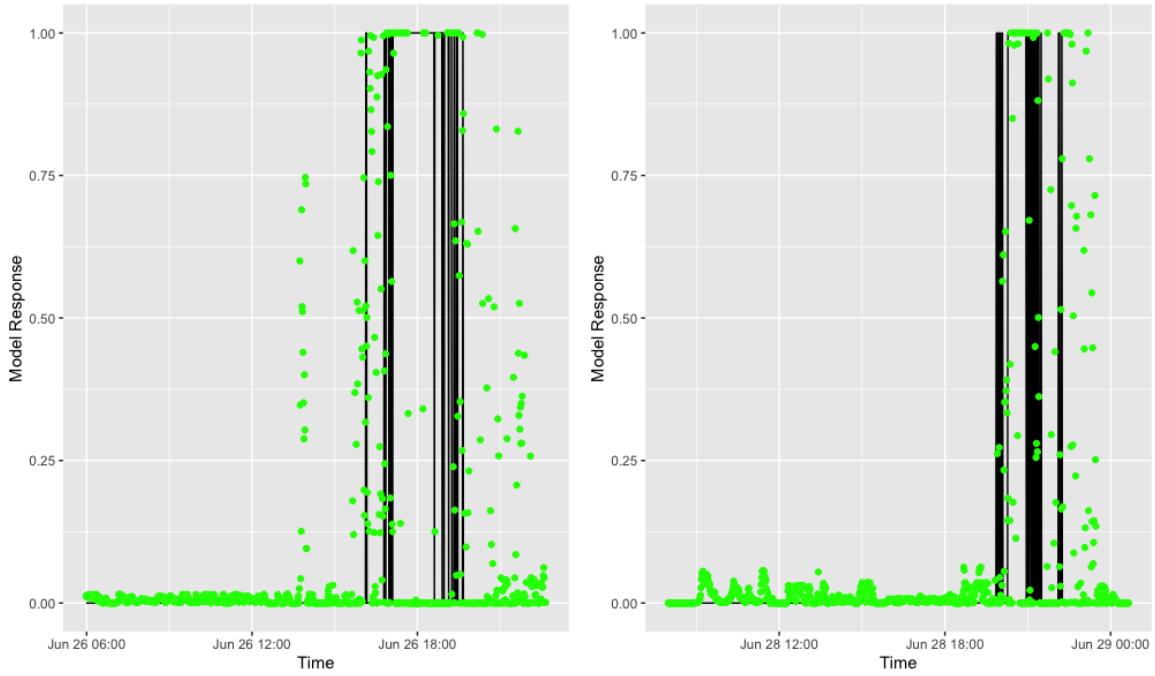
mance based upon time of day or persistence of current conditions. Results indicate a significant loss in model accuracy due to a 24 hour offset, showing that prediction is not simply indicative of diurnal variation. Next, the results of a persistence model is given where a model simply predicts lightning conditions based upon the previous timestamp's lightning condition. Results indicate a drop in accuracy for correct lightning identification, but an increase in the correct identification of "no lightning". This increase in performance is due to the rarity of lightning across the entire time period, making a naïve guess a decent predictor for the absence of lightning while always missing the onset of lightning activity. Extensions to this research will focus on improving these identification rates.

		Observed	
		0	1
Predicted	0	1,440/13,815 10.42%	12,375/13,815 89.58%
	1	365/586 62.29%	221/586 37.71%

		Observed	
		0	1
Predicted	0	13,698/13,814 99.16%	116/13,814 0.84%
	1	116/586 19.8%	470/586 80.2%

Table 8: Confusion matrices built to compare model performance against naïve models. The results of a simple naïve model (left) measures predictions 24 hours prior to observed lightning to demonstrate the model is reacting to EFM conditions and not simply a time cycle. A basic persistence model (right) develops a forecast using only the lightning state of the previous timestamp. These results indicate that the model is not just predicting diurnal variation or based upon conditions in the previous timestamp.

Figures 25a and 25b provide closeups of model behavior against real-world LDAR detected lightning. The model successfully captures lightning, particularly in periods of sustained lightning activity. Of particular concern is the amount of false positives on 26 June 2013 (left) and the lack of a strong model response to lightning onset on 28 June 2013 (right). Continued extensions of this modeling approach will be designed to increase predictive qualities for such events.



(a) (b)

Figure 25: Close-up of model predictive response (green) against binary LDAR data for detected lightning for two time periods of sustained storms.

3.5 Conclusions

The proposed methodology indicates promising improvements to lightning prediction at KSC/CCSFS. The model demonstrates significant improvement over a persistence model for correct positive identification of lightning within the next 60 minutes. The results could provide a timely predictive metric to support decision making in space vehicle launch operations. The formulation methodology demonstrates how the use of experimental design greatly informed and guided possible extensions to the research. Furthermore, the results of the experimental design offer insight into potential extensions of this research designed to improve the model's accuracy in correct identification of the absence of lightning.

The adaptability of the single-index model to data types is a significant advantage of this approach, and time series of alternative weather measurements can be quickly

incorporated. Initial studies using the wavelet coefficients of EFM data moves the prediction window to the left several hours, but chaotic noise currently prevents the fit of accurate models. The addition of other weather datasets may allow the model to not only capture time periods of high lightning risk, but also the triggering events for each lightning event to provide timely warning.

This work notes a high degree of correlation amongst EFM sensors, and clustering analysis indicates possible groupings of EFM sensors. Further analysis combining these methods with wavelet methods could provide an improved analysis of relationships between sensors within the EFM network. A more sparse EFM network may be able to provide equivalent lightning prediction, possibly providing savings in operation and maintenance costs. Furthermore, this could enable the development of small, deployable networks to remote lightning prone areas.

IV. Imputation by Spatiotemporal Kriging and Wavelet De-Noising of Chaotic Electromagnetic Field Sensors at Cape Canaveral for Forecasting of Lightning Risk

Space launch operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) require near-real time determination of lightning risk. Lightning forecasts are developed from large sensor networks that produce very large, chaotic time series. These time series are frequently missing data due to sensor maintenance or local perturbations in the signal. Spatiotemporal kriging estimates data that is autocorrelation both spatially and temporally. Using this method to impute missing data values for lightning prediction results in marked improvements to forecasting accuracy. [1]

Forecasters develop a risk assessment of lightning activity at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) using a dense array of Electric Field Mill (EFM) sensors. These sensors measure the ground-level electric potential within the atmosphere directly overhead each sensor, indicating changes in electromagnetic energy. These changes are a phenomena shown to be predictive of future lightning activity [6] [55]. The EFM network records data at 50Hz, resulting in very large data structures that are high frequency and high volume. Furthermore, these datasets are autocorrelated in regards to both temporal timestamps and spatial distancing of the fixed EFM sensor sites. As is common across many types of sensors, the EFM sites periodically experience periods of time missing measurements. This can be due to routine site maintenance, sensor malfunction, or a purposeful shut-down due to local disturbances that would perturb the sensor readings. These gaps in collection prove problematic in some machine learning and artificial intelligence applications as some methods are not robust to periods of missing data. Imputation

[1] Paper submitted to the journal Weather and Climate Extremes.

methods are required to fill these missing gaps of information using inference from the available data. This study applies imputation methods on the spatially-based EFM time series making use of the inherit autocorrelation in the data, resulting in improved modeling using machine learning and artificial intelligence techniques.

Imputation is a data pre-processing method which substitutes missing entries with estimated values. There are many imputation methods available based upon data type and application. The simplest imputation methods use a representative value for all missing entries, such as the mean, median, or mode of available data. Time series imputation is a sub-discipline which takes into account the autocorrelation between timestamped values. For instance, use of time stamped observations of air pollutants to produce an estimate for missing values [38]. Autocorrelation in time series is the dependence of values between time stamped observations. This results in a great deal of redundancy of the information within time series data, and if not accounted for can result in a model that overstates fit [18]. Time series imputation approaches include use of moving averages, extension of nearest observation, Kalman smoothing, and linear or spline interpolation [67]. Likewise, spatial imputation methods are a sub-discipline that estimates missing data values while accounting for autocorrelation present between spatially correlated measurements. For instance, the estimate of tree density measurements from nearby measurement sites within an especially dense forest [75].

This paper employs a spatiotemporal imputation technique that simultaneously accounts for autocorrelation between spatially correlated measurements collected as a time series. Wavelet methods are used as an additional pre-processing step, serving to de-noise the chaotic EFM measurements to allow faster convergence and estimation of spatiotemporal models. Instead of a purely time series or spatial model, spacetime approaches use all available data to infer predicted values. These methods

prove highly useful in situations in which large amounts of a particular time series are missing and need to be estimated. Although complex in application, such methods are of increasing importance due to the increasing prevalence of modern sensor systems. Section 4.1 provides an overview of the EFM dataset, wavelet methods for de-noising a time series, and spatiotemporal modeling techniques. Section 4.2 presents the methodology and results of wavelet techniques and spatiotemporal modeling as an imputation method. Section 4.3 applies the EFM dataset, to include values estimated by spatiotemporal kriging, using an existing methodology and compared to a baseline imputation method. Conclusions and applications for future research are provided in Section 4.4.

4.1 Methodology

4.1.1 EFM Sensor Network

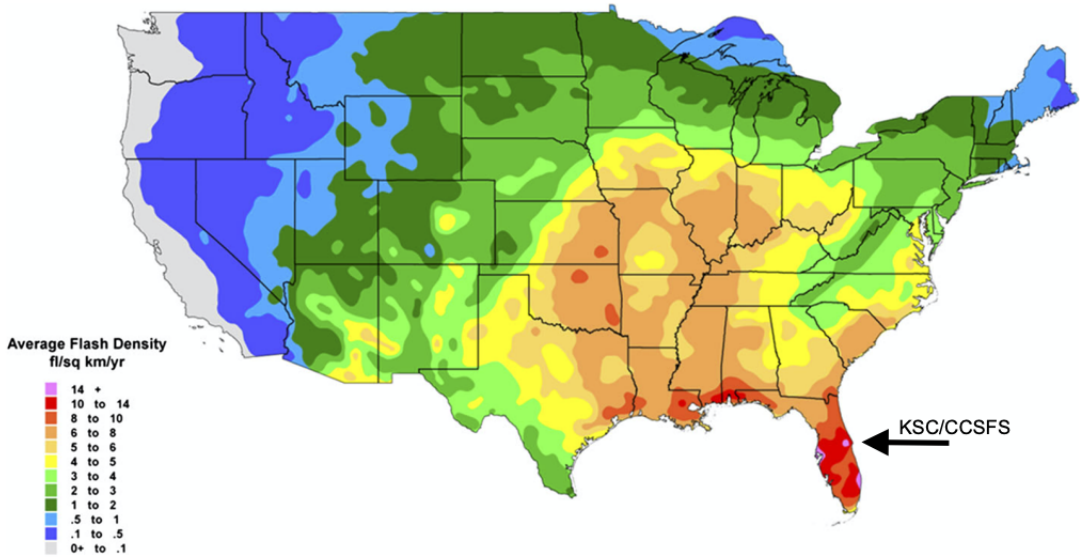


Figure 26: Cloud-to-ground lightning flash density (1997-2010) for the USA from the National Lightning Detection Network [77]

Lightning activity is particularly concentrated in the KSC/CCSFS region of cen-

tral Florida, as can be seen in the heat-map of Figure 26. Accurate and timely forecasts of lightning activity is essential to inform operational risk assessments that guide both flight line and space launch activities. Current studies indicate EFM networks can be predictive of lightning activity through either a relatively sudden change of polarity or an increase in magnitude of the atmospheric electric potential [6] [55]. However, constant movement and churning actions within the atmosphere result in a chaotic response of electrostatic potential by the EFM network [44]. Figure 27 provides three examples of typical and chaotic EFM measurements prior to observed lightning within KSC/CCSFS. Current literature also indicates a diurnal cycle to the EFM network at KSC/CCSFS [56]. The highly chaotic EFM response stored in very large datasets has confounded many attempts to create models to estimate lightning prediction.

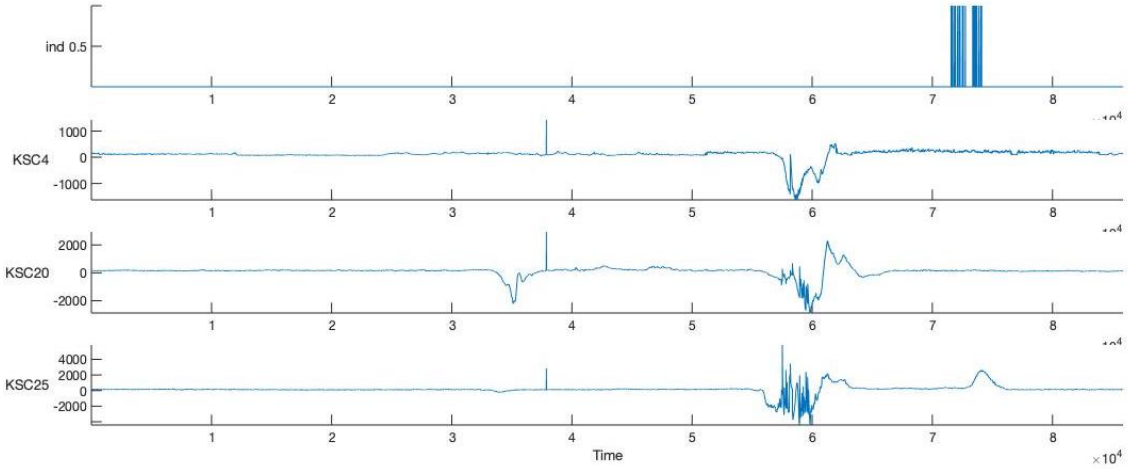


Figure 27: Top subplot is binary response of observed lightning, followed by three typical EFM measurements chosen randomly across the entire KSC/CCSFS region over time in seconds. The EFM measurements indicate a natural steady state in the absence of lightning, becoming increasingly chaotic as electromagnetic potential builds within the atmosphere.

Figure 28 provides a map of the KSC/CCSFS region with the location of all thirty-one EFM sensors. No significant shift in EFM measurements are noted at the 50Hz

rate, so the data is reduced by summarizing by the per minute mean of the 50Hz signal to reduce overall data size.

For evaluation of the imputation method, data for field mill 25 is extracted from the main data frame. The data for field mill 25 is estimated using spatiotemporal imputation methods, and then compared against the actual observed response.

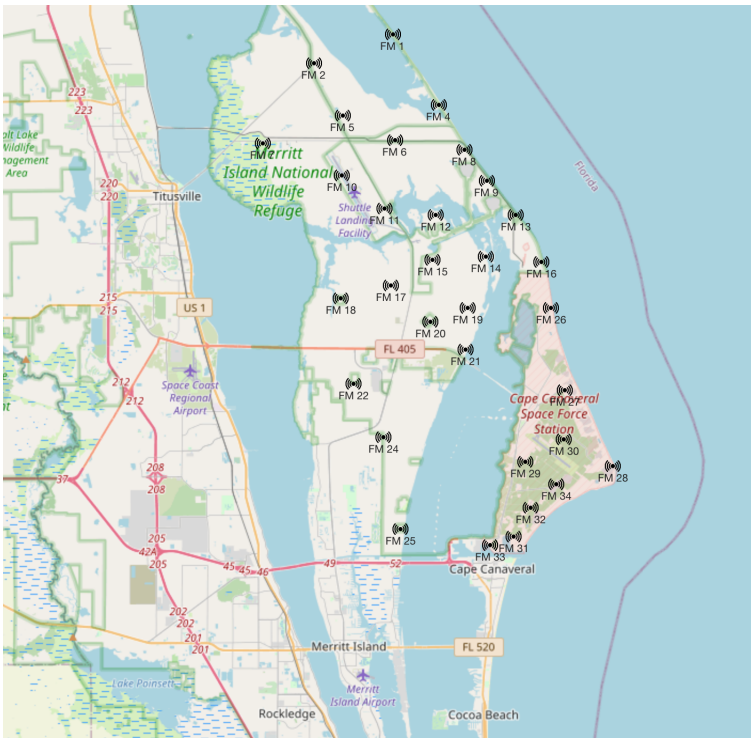


Figure 28: KSC/CCSFS map with locations of EFM sensors.

4.1.2 Wavelet De-noising

Wavelet techniques are used as part of data preprocessing to reduce chaotic noise within the EFM response. Similar to the Fourier transform, wavelet transforms model a function in terms of its constituent frequencies. However, wavelet methods employ a family of unique functions that localize this approximation in time. This allows for the simultaneous approximation of a function in terms of frequency and time. Wavelet methods accomplish this by projecting approximations of a function into a

series of nested subspaces, each providing a different resolution in time.

A Discrete Wavelet Transform (DWT) can be applied to a discrete time series to produce an additive decomposition having constituent detailed time series ($\psi_{j,k}$) reflecting variations at resolution level j and a smoothed version of the time series ($\phi_{j,k}$) reflecting averages at resolution level j [73]. Let ϕ represent the father wavelet function and ψ represent the mother wavelet. Daubechies [13] provides a wide variety of choices for this functions which generate an orthonormal basis. With wavelets defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{32}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{33}$$

a function of time can be represented as

$$f(t) = \sum_j \sum_k d_{j,k}\psi_{j,k}(t) + \sum_k s_{j_0,k}\phi_{j_o,k}(t) \tag{34}$$

where $s_{j,k} = \langle f, \phi_{j,k} \rangle$, $d_{j,k} = \langle f, \psi_{j,k} \rangle$, and $j, k \in \mathbb{Z}$. The time series is thus represented as a linear combination of the shifted and scaled versions of the wavelet functions as estimated using the wavelet coefficients $c_{j,k}$ and $d_{j,k}$. An important consequence of Equation 34 is the separation of the approximation and detailed representations of a signal.

This study employs a Maximal Overlap Discrete Wavelet Transform (MODWT), a variant of wavelet transform well-suited for applications in time series analysis. Unlike the standard DWT which requires a dyadic sample size, the MODWT is well defined for any sized sample. Also unlike the DWT, the MODWT is shift invariant. This means that the wavelet coefficients remain aligned in time with the original time series. A Haar wavelet basis is used in this implementation due to its ability

to model jumps in the response signal. Figure 29 provides a visual representation of the MODWT decomposition for three detail coefficient levels and a smooth level. These properties allow the wavelet coefficients to remain aligned with regards to the temporal position of the original time series. However, the MODWT is a redundant transform that results in $\mathcal{O}(N \log_2 N)$ required computations or a cost of $\mathcal{O}(\log_2 N)$ when compared to the DWT.

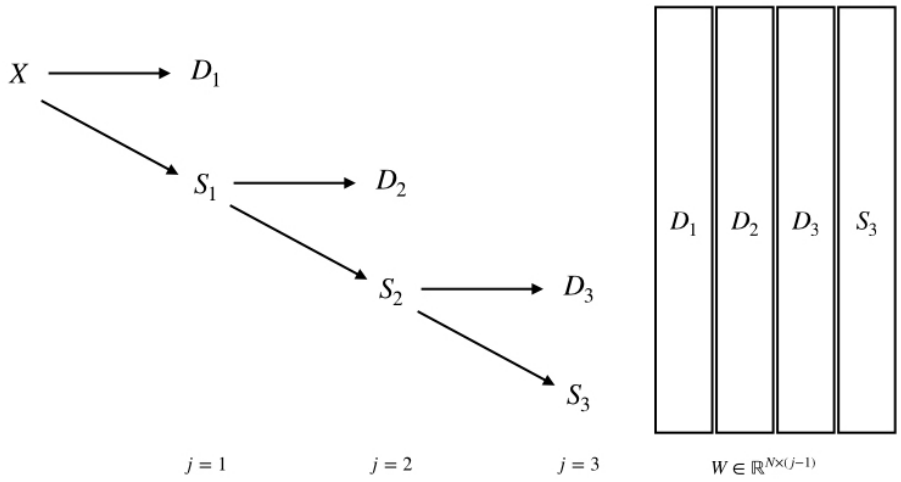


Figure 29: Depiction of a three-level MODWT decomposition of signal X to wavelet coefficients W.

4.1.2.1 Wavelet Thresholding

A DWT or MODWT results in a sparse representation of the decomposed signal in the form of detail and smooth wavelet coefficient levels. This sparse approximation contains all the power of the original signal within relatively few wavelet coefficients. The remainder of the coefficients are either zero or of relatively low magnitude, and predominantly represent stochastic noise in the original time series. Thresholding manipulates these coefficients to reduce how stochastic noise represented in the wavelet model.

This paper uses global thresholding, where a single threshold value λ is applied

uniformly to all or nearly all coefficients. Consider a given threshold value λ and set

$$\hat{f}_\lambda(t) = \sum_j \sum_k I_{\{|d_{j,k}|>\lambda\}} d_{j,k}\psi_{j,k}(t) \tag{35}$$

where I represents the indicator function [69]. This method is known as hard (H) thresholding, where the policy is to set coefficients to zero if less than or equal to the given value of λ. The result that only those high magnitude coefficients are kept that represent the original signal. Then, defining the thresholded coefficients as

$$\hat{d}_{j,k} = \delta_\lambda(d_{j,k}) \tag{36}$$

allows for representation of the hard (H) thresholding rules as

$$\delta_\lambda^H(x) = \begin{cases} x & \text{if } |x| > \lambda \\ 0 & \text{otherwise} \end{cases}. \tag{37}$$

Donoho and Johnstone [16] propose an alternative method of soft (S) thresholding defined as

$$\delta_\lambda^S(x) = \begin{cases} x - \lambda & \text{if } x > \lambda \\ 0 & \text{if } |x| \leq \lambda \\ x + \lambda & \text{if } x < -\lambda \end{cases}. \tag{38}$$

Soft thresholding is similar to hard methods, but values are shrunk towards zero by an amount equal to the threshold λ [69].

4.1.3 Spatiotemporal Modeling

Spatiotemporal modeling assumes a Gaussian spatiotemporal random field $\mathbb{Z}$ defined over a spatial domain $\mathcal{S}$ and a temporal domain $\mathcal{T}$ [24]. A vector of samples

$\mathbf{z} = (z(s_1, t_1), \ldots, (z(s_n, t_n))$ is then a collection of n measurements at distinct locations and times $(s_1, t_1), \ldots, (s_n, t_n) \in \mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{R}$ [24]. Measurements may include repeated values over time for the same location, or multiple values for various locations at the exact same time. Estimated values for unmeasured points (s_0, t_0) can be made since z can be assumed to be the realization of a spatiotemporal random function.

Spatiotemporal kriging is a modeling approach that produces estimated values for unmeasured locations and time using the values from the surrounding area. The method is named after Danie Krige who developed the technique to improve the accuracy of predicting the location of underground ore reserves [7]. Kriging requires the assumption that the response is a continuous random variable over the region of interest $\mathcal{S} \times \mathcal{T}$ [75]. Furthermore, this modeling approach requires an assumption of stationary and spatially isotropic values across the domain of interest [24]. This means independence between the univariate probability, equal probability of occurrence regardless of location, and the bivariate probability law, where the value of the underlying random function between two points depends only upon their relative distance [34].

The field $\mathbb{Z}$ can then be characterized with a covariance function C_{st} where covariance depends only upon distance $h \in \mathbb{R}$ and time $u \in \mathbb{R}$ [24]. The general spatiotemporal covariance function can thus be given as

$$C_{st}(h, u) = \mathrm{Cov}(Z(s, t), Z\left(\tilde{s}, \tilde{t}\right)) \tag{39}$$

for any pair of points $(s, t), (\tilde{s}, \tilde{t}) \in \mathcal{S} \times \mathcal{T}$ where $\|s = \tilde{s}\| = h$ and $|t - \tilde{t}| = u$ [24].

Kriging modeling parameters retain the original nomenclature from geostatistics as seen in Figure 30. The nugget effect is the point at which the semivariogram intersects the y-axis representing semivariance. Although ideally a semivariogram

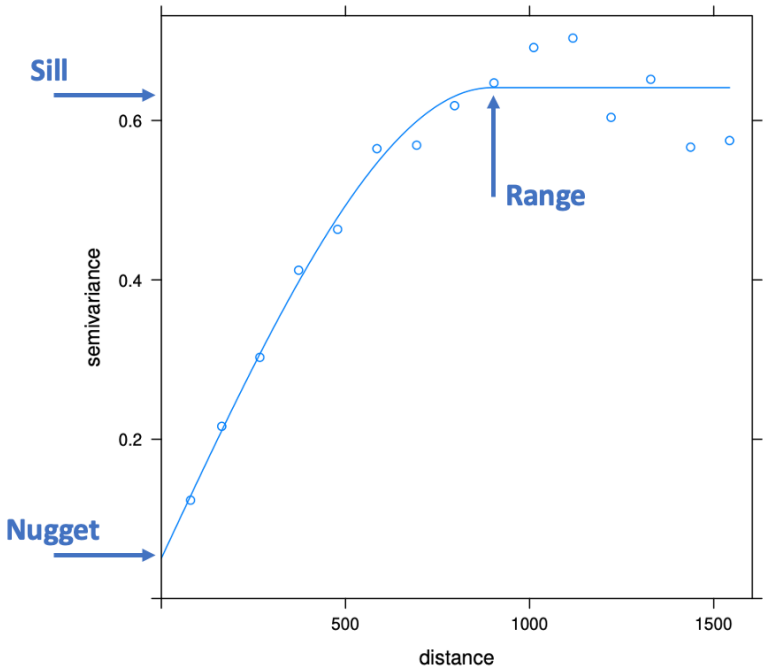


Figure 30: Example spatial semivariogram plot from gstat package [24] [72] annotated to include location of key kriging parameters nugget, sill, and range.

would intersect at the origin, in application measurement error may result in variance amongst spatially similar measurements. The nugget effect could also be due to variations at distances smaller than the sampling distances. The range is the distance at which the semivariogram function levels off, representing the distance at which measurements are no longer autocorrelated. The sill is the value of semivariance for the range.

In practice, the covariance is modeled using a series of variograms. Model estimation is performed using the gstat package for R [72] [24]. First, the observed data are used to derive an empirical variogram that depicts the spatial and temporal autocorrelation of the sample points. This empirical variogram is then used as an input to a fitting routine for a generalized variogram model capable of describing covariance at varying spatial distances and times.

There are classes of generalized covariance models such as the separable covariance model, product-sum model, metric covariance model, sum-metric covariance model, and simplified sum-metric covariance model [24]. Each class includes a trade-off between required assumptions and computational complexity. For instance, the separable covariance model assumes spatiotemporal covariance can be represented as

$$C_{sep}(h,u) = C_s(h)C_t(u)$$

or the product of the spatial and temporal term [24]. This results in the variogram represented as

$$\gamma_{sep}(h,u) = \text{sill} \cdot (\bar{\gamma}_s(h) + \bar{\gamma}_t(u) - \bar{\gamma}(h)\bar{\gamma}_t(u))$$

with standardized spatial and temporal variograms, $\bar{\gamma}_s$ and $\bar{\gamma}_t$, with separate nugget effects and joint sill of 1 [24]. This study employs the Simple Sum-Metric model as it provides the best prediction values. This modeling approach assumes identical spatial and temporal covariance functions only with spatio-temporal anisotropy [24]. Space and time are then matched using an anisotropy correction κ. The Simple Sum-Metric model is calculated by

$$\gamma_{ssm}(h,u) = \text{nug} \cdot \mathbf{1}_{h>0 \vee u>0} + \gamma_s(h) + \gamma_t(u) + \gamma_{joint}\left(\sqrt{h^2 + (\kappa \cdot u)^2}\right)$$

which uses a single nugget effect for the spatial, temporal, and joint variograms [24].

The stationary assumption of ordinary kriging further implies an assumption for an unknown and constant mean over a search neighborhood about the estimation point. This differs from simple kriging which assume a known mean over the entire domain of interest. Ordinary kriging is a best linear unbiased estimator of an

estimated point $\hat{z}(s_0, t_0)$ as

$$\hat{z}(s_0, t_0) = \sum_{i=1}^{n} w_i * z(s_i, t_i)$$

where w_i are the spatiotemporal kriging weights, which are allowed to change across time and location [34]. The optimal kriging weights are then found via a search neighborhood of n points about the estimation point by solving the system of equations

$$\begin{cases} \sum_{j=1}^{n} w_j C_{st}(s_i - s_j, t_i - t_j) + \mu = C_{st}(s_i - s_0, t_i - t_0),\ \forall i = 1, \dots n \\ \sum_{i=1}^{n} w_i = 1 \end{cases}$$

where μ is the Lagrange parameter [34] [78]. Representing the ordinary kriging system of equations in matrix form results in

$$\underbrace{\begin{bmatrix} \tilde{C}_{11} & \dots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \dots & \tilde{C}_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}}_{(n+1)\times(n+1)} \cdot \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ u \end{bmatrix}}_{(n+1)\times 1} = \underbrace{\begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1)\times 1}$$

$$\mathbf{C} \cdot \mathbf{w} = \mathbf{D}$$

whose solution, in the form $\mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D}$, yields the kriging weights [34].

4.2 Imputation Results and Discussion

This new methodology is evaluated by applying it to the EFM dataset. First, the raw EFM data are summarized to the minute to reduce the overall size of the EFM data structure. Time series data for field mill 25 is removed and stored for later comparison against the estimates produced by spatiotemporal kriging.

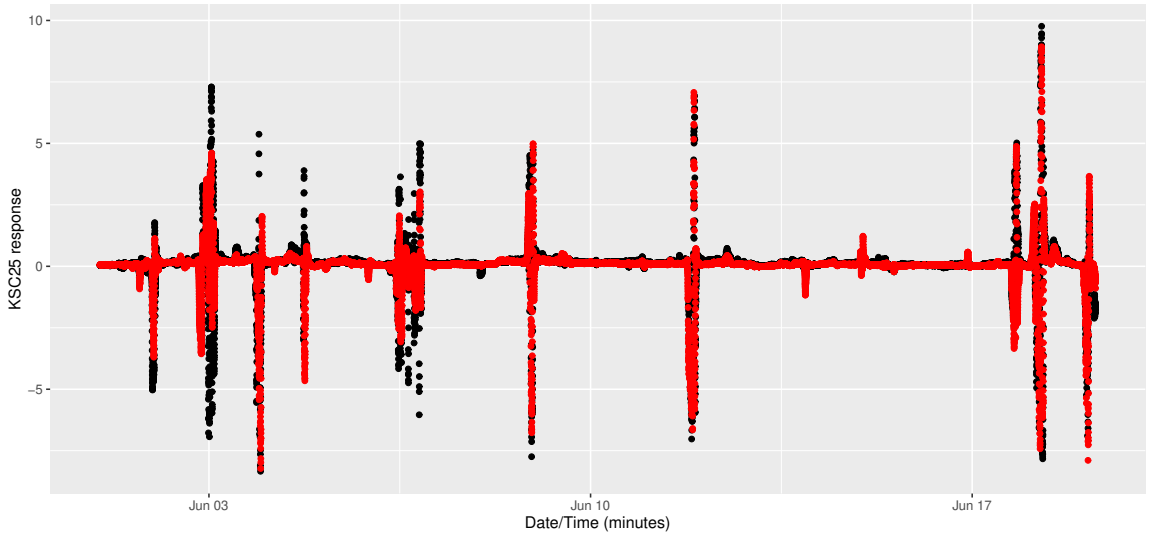


Figure 31: Observed data for field mill 25 (black) and estimated values (red) using a Simple Sum-Metric model and spatiotemporal kriging for 1-19 June 2013, MSE =0.474 and RMSE = 0.688.

A MODWT transform is applied to each individual EFM time series, hard thresholding applied, and an inverse MODWT is conducted to reproduce the de-noised time series. This pre-processing step reduces chaotic noise within the time series, facilitating more accurate and efficient convergence in later machine learning and artificial intelligence applications.

Spatiotemporal modeling is accomplish using the gstat package. An empirical spatiotemporal variogram is estimated from the EFM dataset. All available variogram models in the gstat package are fit and assessed. The Simple Sum-Metric model results in the best fit by RMSE, and is thus chosen for application. Spatiotemporal kriging is then applied to interpolate values for the geodesic position of the missing field mill site.

Figure 31 provides a visual example of the estimated response (red) against the actual observed response (black). Despite the chaotic nature of EFM data, the spatiotemporal modeling technique reconstructs much of the signal for field mill 25, with an observed Mean Squared Error (MSE) of 0.474 and Root Mean Squared Error

(RMSE) of 0.688. Many of the perturbations in the response are captured and modeled correctly, if not always to the full magnitude of the original observed response. This is possibly due to either the chaotic nature of the EFM data or wavelet thresholding. However, this may be a desirable property as the interpolated signal is relatively smooth and well-behaved in comparison to the chaotic raw signal. The benefit of this smoothing would depend entirely on the impact on any further application using machine learning or artificial intelligence.

Some modeling formulations using EFM for lightning forecasting employ mean imputation to fill for periods of lost sensor data. Mean imputation applies the mean of the existing time series to missing timestamped data points. Although this method appears to provide MSE of 0.6651 and RMSE of 0.8155, the constant response fails to provide any of the signal perturbations indicative of impending lightning activity. Furthermore, the relatively high assessed levels of MSE and RMSE are simply due to the EFM signal predominantly existing at a steady state measurement. The spikes out of steady state are the artifacts of interest in EFM applications, and are the indicators required in forecasting using machine learning or artificial intelligence.

4.3 Application of Imputed Data

The fully estimated datasets are applied to the methodology of Nystrom et al. [68] to evaluate the impact of using a fully imputed dataset. This methodology uses the same EFM data but with greatly reduced range of the time series to only those periods with a high proportion of EFM sensors active. Large blocks of data estimated by mean imputation caused the model to behave erratically. The application in this study seeks to apply the methodology using spatiotemporal imputation and without regard for any periods of EFM inactivity. Figure 32 provides the count of missing data points by minute for the EFM network in June 2013 as used in Nystrom et al. [68].

		Observed	
		0	1
Predicted	0	27,597/27,708 99.5%	111/27,708 0.5%
	1	87/1,164 7.5%	1,077/1,164 92.5%

Mean Imputation

		Observed	
		0	1
Predicted	0	27,578/27,708 99.5%	130/27,708 0.5%
	1	51/1,164 4.4%	1,113/1,164 95.6%

Spatiotemporal Imputation

Table 9: Confusion matrices for model predictions using EFM data 60 minutes prior to any observed lightning within the Central Cape lightning warning circle for 28,872 observations during 10-30 June 2013. A prediction or observed value of "0" corresponds to no lightning, whereas a "1" denotes observed or predicted lightning within the lightning warning circle. Results indicate sizable improvements in the positive identification of lightning when spatiotemporal imputation is used to complete the EFM dataset.

A majority of the sensors are missing data from short periods of less than 30 minutes when the entire network is inoperable. Linear interpolation is used to complete these time series, as there is no data available for interpolation. The spatiotemporal kriging methodology is then applied to the remaining time series to interpolate missing values.

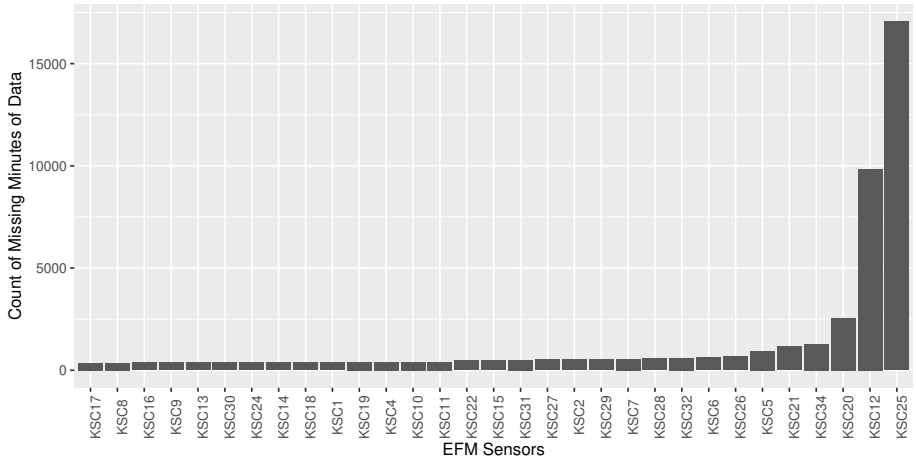


Figure 32: Count of missing data by minute for all 31 EFM sensors in June 2013, sorted by count. Sensor KSC25 is missing the most with 17,061 minutes of missing data, or about 39% of all data for the month.

Table 9 provides the results of lightning prediction using both mean imputation and spatiotemporal imputation on the original EFM dataset. Results are presented

in a confusion matrix, where the predicted state of no lightning "0" or lightning "1" is paired against actual lightning conditions observed for the same period at KSC/CCSFS. Model results predicting a lack of lightning are comparable between the two datasets. Spatiotemporal imputation results in a marked increase in the prediction accuracy for the presence of lightning (1,1) from 92.5% to 95.6%. Furthermore, this lowers the false alarm rate (1,0) that could reduce the operational impact of unnecessary lightning warnings. These improvements in model performance both increase safety for launch conditions and increase operational efficiency of launch and space flight line activities. This increase in accuracy is most likely due to the preservation of perturbations within the EFM dataset using spatiotemporal kriging, providing the semi-parametric model the key indicators for impending lightning activity.

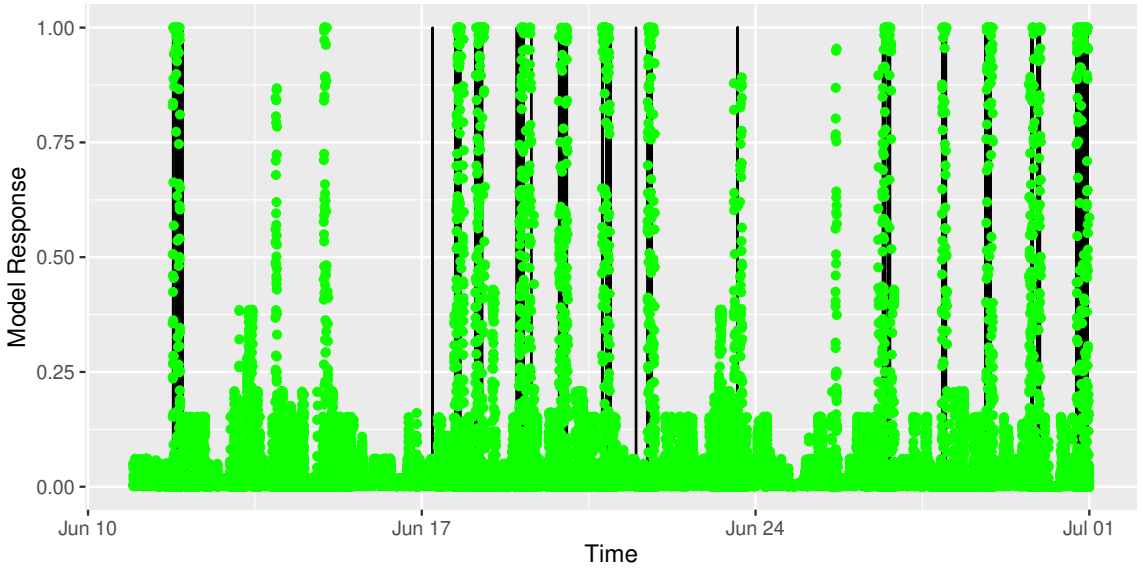


Figure 33: Predicted model response (green) using imputed EMF dataset against actual observed lightning (black) on Cape Canaveral, June 2013.

Figure 33 provides a visual representation of the model's prediction response against the actual observed lightning at KSC/CCSFS for 10-30 June 2013. This predicted response is estimated using the spatiotemporal imputed EFM dataset. The model provides a predictive response to nearly all the observed lightning, with three

apparent false alarms during the period. Some further analysis indicates the false alarm predictions align with lightning storms within the KSC/CCSFS region that did not produce lightning within the lightning warning circle under consideration for the model. Future extensions of this work will focus on reducing the impact of regional lightning storms.

		Observed	
		0	1
Predicted	0	13,698/13,814 99.16%	116/13,814 0.84%
	1	116/586 19.8%	470/586 80.2%

Table 10: Confusion matrix for performance of the naïve persistence model. This model develops a forecast using only the lightning state of the previous timestamp. For instance, if there is no lightning at time t, then the model predicts no lightning at $t + 1$. The wavelet enabled semi-parametric modeling approach outperforms a naïve model in this implementation and indicates this new methodology has explanatory power in the prediction of lightning phenomena.

Table 10 provides the results of a naïve model based upon persistence, where the model predicts the state of lightning for time $t + 1$ based exclusively on the state of lightning at time t. This manner of comparison is common in the meteorological literature, and shows whether the model under evaluation is providing explanatory insights to weather phenomena. The wavelet-enabled semi-parametric modeling approach outperforms the persistence model, most notably in the identification of the presence of lightning.

4.4 Conclusion

Spatiotemporal kriging provides an excellent method to recreate a missing time series that includes spatial autocorrelation. The technique proved robust, despite the chaotic nature of EFM measures of atmospheric electrostatic potential. Furthermore, the interpolated time series displays evidence of some smoothing while also preserving

the signal of interest for lightning prediction. Both of these qualities may aid in convergence in additional machine learning or artificial intelligence applications while still facilitation accurate and timely predictions.

Acknowledgements

This research is partially funded by the Omar Nelson Bradley Fellowships.

V. Conclusion

This research develops and evaluates methods to inform critical decisions using data that is both chaotic and incomplete. The review of current literature and survey of wavelet methods in forecasting time series are provided to examine current techniques and best applications. A wavelet-enabled forecasting methodology for lightning is proposed using the chaotic EFM data from KSC/CCSFS. Wavelet de-noising is applied to the chaotic EFM time series during pre-processing. A semiparametric single-index model is then estimated from a training dataset, and then evaluated in a testing dataset. A designed experiment is used to efficiently explore the factor space of model parameters. Results indicate a promising method for lightning forecasting against the baseline persistence model. Furthermore, once a model is estimated a forecast can be quickly and efficiently calculated using updated EFM measurements. A further research extension is the inclusion of spatiotemporal kriging as an imputation for the spatially and temporally autocorrelation EFM dataset. Results indicate the improved imputation method produces a forecasting accuracy of over 95% using a process that is robust to large missing pieces of the EFM time series. Comparing the results of the proposed methodology against naïve models clearly indicates clear improvements to lightning prediction.

Bibliography

1. Hojjat Adeli and Asim Karim. Fuzzy-wavelet RBFNN model for freeway incident detection. *Journal of Transportation Engineering*, 126(6):464–471, 2000.

2. Hojjat Adeli and Abhish Samant. An adaptive conjugate gradient neural network–wavelet model for traffic incident detection. *Computer-Aided Civil and Infrastructure Engineering*, 15(4):251–260, 2000.

3. Richard M. Allen and Hiroo Kanamori. The potential for earthquake early warning in southern California. *Science*, 300(5620):786–789, 2003.

4. Th. Arampatzis, John Lygeros, and Stamatis Manesis. A survey of applications of wireless sensors and wireless sensor networks. In *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, 2005.*, pages 719–724. IEEE, 2005.

5. Daniel Aranguren, Joan Montanya, Gloria Sola, Victor March, David Romero, and Horacio Torres. On the lightning hazard warning using electrostatic field: Analysis of summer thunderstorms in Spain. *Journal of Electrostatics*, 67(2-3): 507–512, 2009.

6. Daniel Aranguren, J. Inampués, Horacio Torres, J. López, and Ernesto Pérez. Operational analysis of electric field mills as lightning warning systems in Colombia. In *2012 International Conference on Lightning Protection (ICLP)*, pages 1–6. IEEE, 2012.

7. Margaret Armstrong. *Basic linear geostatistics*. Springer-Verlag Berlin Heidelberg, New York, NY, 1998.

8. K. Bhaskar and Sri Niwas Singh. Wind speed forecasting using MRA based adaptive wavelet neural network. In *Proc. 16th National Power Systems Conf*, 2010.

9. T. Tony Cai. Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of statistics*, 27(3):898–924, 1999.

10. João Paulo da Silva Catalão, Hugo Miguel Inácio Pousinho, and Vıctor Manuel Fernandes Mendes. Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. *IEEE Transactions on sustainable energy*, 2(1):50–59, 2010.

11. João Paulo da Silva Catalão, Hugo Miguel Inácio Pousinho, and Vıctor Manuel Fernandes Mendes. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renewable energy*, 36(4):1245–1251, 2011.

12. Hamed Chitsaz, Nima Amjady, and Hamidreza Zareipour. Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm. *Energy conversion and Management*, 89:588–598, 2015.

13. Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61. Society for Industrial and Applied Mathematics (SIAM), 1992.

14. Ronaldo R.B. De Aquino, Milde M.S. Lira, Josinaldo B. de Oliveira, Manoel A. Carvalho, Otoni N. Neto, and Givanildo J. de Almeida. Application of wavelet and neural network models for wind speed and power generation forecasting in a Brazilian experimental wind park. In *2009 international joint conference on neural networks*, pages 172–178. IEEE, 2009.

15. David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432): 1200–1224, 1995.

16. David L. Donoho and Jain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

17. Boubacar Doucoure, Kodjo Agbossou, and Alben Cardenas. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renewable Energy*, 92:202–211, 2016.

18. Gidon Eshel. *Spatiotemporal Data Analysis*. Princeton University Press, 41 William Street, Princeton, NJ, 2012.

19. Diogo L. Faria, Rui Castro, Claudia Philippart, and Alexandre Gusmao. Wavelets pre-filtering in wind speed prediction. In *2009 International Conference on Power Engineering, Energy and Electrical Drives*, pages 168–173. IEEE, 2009.

20. Michael W. Frazier. *An Introduction to Wavelets Through Linear Algebra*. Springer Science & Business Media, 2006.

21. Juan J. Galiana-Merino, Julio Rosa-Herranz, P. Jauregui, Sergio Molina, and J. Giner. Wavelet transform methods for azimuth estimation in local three-component seismograms. *Bulletin of the Seismological Society of America*, 97 (3):793–803, 2007.

22. Samanwoy Ghosh-Dastidar and Hojjat Adeli. Wavelet-clustering-neural network model for freeway incident detection. *Computer-Aided Civil and Infrastructure Engineering*, 18(5):325–338, 2003.

23. Pierre Goupillaud, Alex Grossmann, and Jean Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1):85–102, 1984.

24. Benedikt Gräler, Edzer Pebesma, and Gerard Heuvelink. Spatio-Temporal Interpolation using gstat. *The R Journal*, 1(8):204–218, 2016.

25. Alfréd Haar and Theodore von Kármán. Zur theorie der spannungszustände in plastischen und sandartigen medien. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1909:204–218, 1909.

26. Wolfgang Karl Härdle, Marlene Müller, Stefan Sperlich, and Axel Werwatz. *Nonparametric and Semiparametric Models*. Springer Science & Business Media, 2012.

27. Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 2008. URL `http://www.jstatsoft.org/v27/i05/`.

28. Daniel J. Henderson and Christopher F. Parmeter. *Applied Nonparametric Econometrics*. Cambridge University Press, 2015.

29. George Hloupis and Filippos Vallianatos. Wavelet-based rapid estimation of earthquake magnitude oriented to early warning. *IEEE Geoscience and Remote Sensing Letters*, 10(1):43–47, 2012.

30. George Hloupis and Filippos Vallianatos. Wavelet-based methods for rapid calculations of magnitude and epicentral distance: An application to earthquake early warning system. *Pure and Applied Geophysics*, 172(9):2371–2386, 2015.

31. Joel L. Horowitz. *Semiparametric Methods in Econometrics*. Springer-Verlag, New York, 1998.

32. Katherine Hunt and Guy P Nason. Wind speed modelling and short-term prediction using wavelets. *Wind Engineering*, 25(1):55–61, 2001.

33. Hidehiko Ichimura. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58(1-2):71–120, 1993.

34. Edward H. Isaaks and R. Mohan Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, 200 Madison Avenue, New York, New York, 10016, 1989.

35. Elizabeth A Jacobson and E Philip Krider. Electrostatic field changes produced by Florida lightning. *Journal of the Atmospheric Sciences*, 33(1):103–117, 1976.

36. Xiaomo Jiang and Hojjat Adeli. Wavelet packet-autocorrelation function method for traffic flow pattern analysis. *Computer-Aided Civil and Infrastructure Engineering*, 19(5):324–337, 2004.

37. Xiaomo Jiang and Hojjat Adeli. Dynamic wavelet neural network model for traffic flow forecasting. *Journal of transportation engineering*, 131(10):771–779, 2005.

38. W.L. Junger and A. Ponce De Leon. Imputation of missing data in time series for air pollutants. *Atmospheric Environment*, 102:96–104, 2015.

39. Hiroo Kanamori. Real-time seismology and earthquake damage mitigation. *Annual Review of Earth and Planetary Sciences*, 33:195–214, 2005.

40. Thomas Karako. Missile defense: Time to go big. *Center for Strategic and International Studies*, 4, 2016.

41. Asim Karim and Hojjat Adeli. Comparison of fuzzy-wavelet radial basis func-

tion neural network freeway incident detection model with california algorithm. *Journal of Transportation Engineering*, 128(1):21–30, 2002.

42. Ahmad Aarshan Khan and Mohammad Shahidehpour. One day ahead wind speed forecasting using wavelets. In *2009 IEEE/PES Power Systems Conference and Exposition*, pages 1–5. IEEE, 2009.

43. Roger W. Klein and Richard H. Spady. An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 61(2):387–421, 1993.

44. E. Philip Krider. Electric field changes and cloud electrical structure. *Journal of Geophysical Research: Atmospheres*, 94(D11):13145–13149, 1989.

45. K-M Lau and Hengyi Weng. Climate signal detection using wavelet transform: How to make a time series sing. *Bulletin of the American meteorological society*, 76(12):2391–2402, 1995.

46. Cao Lei and Li Ran. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. In *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 2525–2529. IEEE, 2008.

47. Qi Li and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

48. Chuntao Liu, Earle R Williams, Edward J Zipser, and Gary Burns. Diurnal variations of global thunderstorms and electrified shower clouds and their contribution to the global electrical circuit. *Journal of the atmospheric sciences*, 67 (2):309–323, 2010.

49. Da Liu, Dongxiao Niu, Hui Wang, and Leilei Fan. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, 62:592–597, 2014.

50. Hui Liu, Hong-Qi Tian, Chao Chen, and Yan-fei Li. A hybrid statistical method to predict wind speed and wind power. *Renewable energy*, 35(8):1857–1861, 2010.

51. Hui Liu, Hong-qi Tian, Di-fu Pan, and Yan-fei Li. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. *Applied Energy*, 107:191–208, 2013.

52. Hui Liu, Hong-qi Tian, and Yan-fei Li. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. *Energy Conversion and Management*, 89:1–11, 2015.

53. Hui Liu, Hong-qi Tian, and Yan-fei Li. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Conversion and Management*, 100:16–22, 2015.

54. Hui Liu, Xi-wei Mi, and Yan-fei Li. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network. *Energy conversion and management*, 156:498–514, 2018.

55. J. Lopez, E. Perez, J. Herrera, D. Aranguren, and L. Porras. Thunderstorm warning alarms methodology using electric field mills and lightning location networks in mountainous regions. In *2012 International Conference on Lightning Protection (ICLP)*, pages 1–6. IEEE, 2012.

56. Greg M. Lucas, Jeffrey P. Thayer, and Wiebke Deierling. Statistical analysis of spatial and temporal variations in atmospheric electric fields from a regional array of field mills. *Journal of Geophysical Research: Atmospheres*, 122(2):1158–1174, 2017.

57. Douglas M Mach, Richard J Blakeslee, Monte G Bateman, and Jeffrey C Bailey. Comparisons of total currents based on storm location, polarity, and flash rates derived from high-altitude aircraft overflights. *Journal of Geophysical Research: Atmospheres*, 115(D3), 2010.

58. Douglas M Mach, Richard J Blakeslee, and Monte G Bateman. Global electric circuit implications of combined aircraft storm electric current measurements and satellite-based diurnal lightning statistics. *Journal of Geophysical Research: Atmospheres*, 116(D5), 2011.

59. Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11(7):674–693, 1989.

60. Bryan F.J. Manly and Jorge A. Navarro Alberto. *Multivariate Statistical Methods: A Primer*. CRC press, 2017.

61. MATLAB Wavelet Toolbox. Matlab wavelet toolbox, 2020a.

62. Kelly McGinnity and Eric Chicken. Wavelet block thresholding for non-gaussian errors. Technical report, Technical Report, 2012.

63. Kelly McGinnity, Roumen Varbanov, and Eric Chicken. Cross-validated wavelet block thresholding for non-gaussian errors. *Computational Statistics & Data Analysis*, 106:127–137, 2017.

64. Jeffrey J. McGuire, Frederik J. Simons, and John A. Collins. Analysis of seafloor seismograms of the 2003 Tokachi-Oki earthquake sequence for earthquake early warning. *Geophysical Research Letters*, 35(14), 2008.

65. Anbo Meng, Jiafei Ge, Hao Yin, and Sizhe Chen. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Conversion and Management*, 114:75–88, 2016.

66. Douglas C Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, New York, NY, 10th edition, 2019.

67. Steffen Moritz and Thomas Bartz-Beielstein. imputeTS: Time series missing value imputation in R. *The R Journal*, 9(1):207, 2017.

68. Jared Nystrom, Raymond Hill, Andrew Geyer, Joseph Pignatiello Jr., and Eric Chicken. (in press). Experimental Design in Complex Model Formulation for Lightning Prediction. *International Journal of Experimental Design and Process Optimisation*, 2021.

69. R. Todd Ogden. *Essential Wavelets for Statistical Applications and Data Analysis*. Springer Science & Business Media, New York, NY, 1997.

70. Erik L. Olson and Richard M. Allen. The deterministic nature of earthquake rupture. *Nature*, 438(7065):212, 2005.

71. G.J. Osório, J.C.O. Matias, and J.P.S. Catalão. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. *Renewable Energy*, 75:301–307, 2015.

72. Edzer J Pebesma. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7):683–691, 2004.

73. Donald B. Percival and Andrew T. Walden. *Wavelet Methods for Time Series Analysis*, volume 4. Cambridge University Press, 40 West 20th Street, New York, NY, 2006.

74. Ramakrushna Reddy and Rajesh R. Nair. The efficacy of support vector machines (SVM) in robust determination of earthquake early warning magnitudes in central Japan. *Journal of Earth System Science*, 122(5):1423–1434, 2013.

75. Andrew P. Robinson and Jeff D. Hamann. *Forest analytics with R: an introduction*. Springer Science & Business Media, 233 Spring Street, New York, NY, 2011.

76. William P. Roeder and Jim E. Glover. Preliminary results from phase-1 of the statistical forecasting of lightning cessation project, paper presented at Conference on Meteorological Applications of Lightning Data. *American Meteorological Society, San Diego, CA*, 2005.

77. William P. Roeder, Benjamin H. Cummins, Kenneth L. Cummins, Ronald L. Holle, and Walker S. Ashley. Lightning fatality risk map of the contiguous united states. *Natural Hazards*, 79(3):1681–1692, 2015.

78. Christopher J Ruybal, Terri S Hogue, and John E McCray. Evaluation of groundwater levels in the arapahoe aquifer using spatiotemporal regression kriging. *Water Resources Research*, 55(4):2820–2837, 2019.

79. Abhish Samant and Hojjat Adeli. Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis. *Computer-Aided Civil and Infrastructure Engineering*, 15(4):241–250, 2000.

80. Abhish Samant and Hojjat Adeli. Enhancing neural network traffic incident-detection algorithms using wavelets. *Computer-Aided Civil and Infrastructure Engineering*, 16(4):239–245, 2001.

81. Dawn L Sanderson. Modeling the distribution of lightning strike distances outside a preexisting lightning area. Master's thesis, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 2019.

82. Frederik J. Simons, Ben D.E. Dando, and Richard M. Allen. Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival. *Earth and Planetary Science Letters*, 250(1-2): 214–223, 2006.

83. Sri Niwas Singh and Abheejeet Mohapatra. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renewable energy*, 136:758–768, 2019.

84. Charles A Skrovan. An analysis of a lighting prediction threshold for 45th weather squadron electric field mill data. Master's thesis, Air Force Institute of Technology, 2020.

85. Saurabh S. Soman, Hamidreza Zareipour, Om Malik, and Paras Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010*, pages 1–8. IEEE, 2010.

86. Dominick V. Speranza. Lightning prediction using recurrent neural networks. Master's thesis, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 2019.

87. Geoffrey T. Stano, Henry E. Fuelberg, and William P. Roeder. Developing empirical lightning cessation forecast guidance for the Cape Canaveral Air Force

Station and Kennedy Space Center. *Journal of Geophysical Research: Atmospheres*, 115(D9), 2010.

88. Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

89. Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and computational harmonic analysis*, 3(2):186–200, 1996.

90. Akin Tascikaraoglu and Mehmet Uzunoglu. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews*, 34:243–254, 2014.

91. Hualiang Teng and Yi Qi. Application of wavelet technique to freeway incident detection. *Transportation Research Part C: Emerging Technologies*, 11(3-4): 289–308, 2003.

92. Filippos Vallianatos and George Hloupis. HVSR technique improvement using redundant wavelet transform. In *Increasing seismic safety by combining engineering technologies and seismological data*, pages 117–137. Springer, 2009.

93. Eleni I. Vlahogianni, Matthew G. Karlaftis, and John C. Golias. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.

94. Xiaochen Wang, Peng Guo, and Xiaobin Huang. A review of wind power forecasting models. *Energy procedia*, 12:770–778, 2011.

95. Brandon Whitcher. *waveslim: Basic Wavelet Routines for One-, Two-, and Three-Dimensional Signal Processing*, 2020. URL `http://waveslim.blogspot.com`. R package version 1.8.2.

96. Yih-Min Wu, Hsin-Yi Yen, Li Zhao, Bor-Shouh Huang, and Wen-Tzong Liang. Magnitude determination using initial P waves: A single-station approach. *Geophysical Research Letters*, 33(5), 2006.

97. Yuanchang Xie and Yunlong Zhang. A wavelet network model for short-term traffic volume forecasting. *Journal of Intelligent Transportation Systems*, 10(3): 141–150, 2006.

98. Jianwu Zeng and Wei Qiao. Short-term wind power prediction using a wavelet support vector machine. *IEEE transactions on sustainable energy*, 3(2):255–264, 2012.

99. Qinghua Zhang. Using wavelet network in nonparametric estimation. *IEEE Transactions on Neural networks*, 8(2):227–236, 1997.

100. Wenyu Zhang, Jujie Wang, Jianzhou Wang, Zengbao Zhao, and Meng Tian. Short-term wind speed forecasting based on a hybrid model. *Applied Soft Computing*, 13(7):3225–3233, 2013.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
16-09–2021	Dissertation	Sept 2018 — Aug 2021

4. TITLE AND SUBTITLE

Wavelet Methods for Very-short Term Forecasting of Functional Time Series

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nystrom, Jared K., LTC, U.S. Army

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-DS-21-S-037

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

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Space launch operations at Kennedy Space Center and Cape Canaveral Space Force Station (KSC/CCSFS) are complicated by unique requirements for near-real time determination of risk from lightning. Lightning forecast weather sensor networks produce data that are noisy, high volume, and high frequency time series for which traditional forecasting methods are often ill-suited. Current approaches result in significant residual uncertainties and consequentially may result in forecasting operational policies that are excessively conservative or inefficient. This work proposes a new methodology of wavelet-enabled semiparametric modeling to develop accurate and timely forecasts robust against chaotic functional data. Wavelets methods are first used to de-noise the weather data, which is then used to estimate a single-index model for forecasting. This semiparametric technique mitigates noise of the chaotic signal while avoiding any possible distributional misspecification.

15. SUBJECT TERMS

wavelets, semiparametric model, design of experiments, spatiotemporal kriging, imputation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Raymond R. Hill, AFIT/ENS
U	U	U	UU	134	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x7469; Raymond.Hill@afit.edu