



LUND UNIVERSITY

Using Stochastic Processes and Machine Learning to Solve Environmental Challenges Applications in renewable energy and remote sensing Zeinali, Shokoufa

2026

[Link to publication](#)

Citation for published version (APA):

Zeinali, S. (2026). *Using Stochastic Processes and Machine Learning to Solve Environmental Challenges: Applications in renewable energy and remote sensing*. Centre for Mathematical Sciences, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Using Stochastic Processes and Machine Learning to Solve Environmental Challenges

Using Stochastic Processes and Machine Learning to Solve Environmental Challenges

Applications in renewable energy and remote sensing

by Shokoufa Zeinali



LUND
UNIVERSITY

Doctoral Thesis

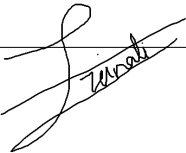
Thesis advisors: Johan Lindström, Magnus Wiktorsson, Ted Kronvall,
and Per-Ola Olsson

Faculty opponent: Professor Thordis Thorarinsdottir, University of Oslo

To be presented, with the permission of the LTH of Lund University, for public criticism in
the room MH:G at the Centre for Mathematical Sciences on Wednesday, the 28th of October
2026 at 13:00

Organization LUND UNIVERSITY Centre for Mathematical Sciences Box 118 SE-221 00 LUND Sweden		Document name Doctoral thesis	
Author(s) Shokoufa Zeinali		Date of presentation 2026-10-28	
		Sponsoring organization Swedish Energy Agency and MERGE (Modelling the Regional and Global Earth System)	
Title and subtitle Using Stochastic Processes and Machine Learning to Solve Environmental Challenges			
Abstract This thesis investigates the application of artificial intelligence (AI), machine learning, and environmental statistics to address challenges in renewable energy optimization and environmental monitoring. Environmental systems are inherently complex, dynamic, and uncertain, making data-driven methods particularly valuable for modeling and analysis. Advances in remote sensing technologies and the increasing availability of large-scale environmental data have further enabled the development of AI-based approaches capable of extracting spatial and temporal patterns from heterogeneous and incomplete datasets. The first part of this thesis focuses on the optimization of Wave Energy Converters (WECs). Ocean waves represent a promising renewable energy source due to their high energy density and predictability, but efficient energy extraction requires adaptive system design under stochastic sea conditions. Using statistical models of ocean wave behavior, Papers I and II investigate the optimization of a WEC system developed by the company Waves4Power. The second part of the thesis addresses environmental monitoring using aerial and satellite imagery. In Paper III, the focus is on the early detection of bark beetle infestations in forests using Sentinel-2 satellite time series data and static environmental features classified by XG-Boost. In Paper IV, a Conditional Variational Autoencoder (CVAE) combined with a Gaussian Mixture Model (GMM) is proposed for semi-supervised classification by integrating temporal satellite observations with static environmental variables. In Paper V, a U-Net-based deep learning model is developed to reconstruct distorted or noisy Shortwave Infrared (SWIR) spectral information in optical remote sensing imagery. Although the applications considered in this thesis differ, both are unified by a common methodological perspective: the use of data-driven models to analyze and optimize environmental systems characterized by uncertainty, temporal variability, and incomplete observations.			
Key words Machine Learning; Environmental Statistics; Optimization; Wave Energy Converters; Remote Sensing; Deep Learning; Forest Disturbance Detection			
Classification system and/or index terms (if any)			
Supplementary bibliographical information		Language English	
ISSN and key title 1404-0034, Doctoral Thesis in Mathematical Sciences, 2026:9		ISBN 978-91-6858-028-9 (print) 978-91-6858-029-6 (pdf)	
Recipient's notes	Number of pages 231	Price	
	Security classification		

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature 

Date 2026-08-31

Using Stochastic Processes and Machine Learning to Solve Environmental Challenges

Applications in renewable energy
and remote sensing

by Shokoufa Zeinali



LUND
UNIVERSITY

Funding information: This research was supported by Swedish Energy Agency grant 50197-1 and carried out in cooperation with the company Waves4Power. Financial support has also been received from the Swedish Strategic Research Area (SRA) MERGE (ModELLing the Regional and Global Earth System). The study is a contribution to the SRA BECC (Biodiversity and Ecosystem Services in a Changing Climate) and to the Lund University profile area Nature-based future solutions.

Mathematical Statistics
Centre for Mathematical Sciences
Box 118
SE-221 00 LUND
Sweden

www.maths.lu.se

Doctoral Thesis in Mathematical Sciences 2026:9
ISSN: 1404-0034

ISBN: 978-91-6858-028-9 (print)
ISBN: 978-91-6858-029-6 (pdf)
LUTFMS-1053-2026

© Shokoufa Zeinali, 2026

Printed in Sweden by Media-Tryck, Lund University, Lund 2026



*Beyond belief and unbelief, there is an open field;
our hearts are drawn toward that boundless place.
When the mystic reaches there, he bows his head,
for there, neither belief nor unbelief has a place.*

— Rumi (Molana)

Dedicated to my Parents
Rezvan – Nooshin

Abstract

This thesis investigates the application of artificial intelligence (AI), machine learning, and environmental statistics to address challenges in renewable energy optimization and environmental monitoring. Environmental systems are inherently complex, dynamic, and uncertain, making data-driven methods particularly valuable for modeling and analysis. Advances in remote sensing technologies and the increasing availability of large-scale environmental data have further enabled the development of AI-based approaches capable of extracting spatial and temporal patterns from heterogeneous and incomplete datasets.

The first part of this thesis focuses on the optimization of Wave Energy Converters (WECs). Ocean waves represent a promising renewable energy source due to their high energy density and predictability, but efficient energy extraction requires adaptive system design under stochastic sea conditions. Using statistical models of ocean wave behavior, Papers I and II investigate the optimization of a WEC system developed by the company Waves4Power.

The second part of the thesis addresses environmental monitoring using aerial and satellite imagery. In Paper III, the focus is on the early detection of bark beetle infestations in forests using Sentinel-2 satellite time series data and static environmental features classified by XGBoost.

In Paper IV, a Conditional Variational Autoencoder (CVAE) combined with a Gaussian Mixture Model (GMM) is proposed for semi-supervised classification by integrating temporal satellite observations with static environmental variables.

In Paper V, a U-Net-based deep learning model is developed to reconstruct distorted or noisy Shortwave Infrared (SWIR) spectral information in optical remote sensing imagery.

Although the applications considered in this thesis differ, both are unified by a common methodological perspective: the use of data-driven models to analyze and optimize environmental systems characterized by uncertainty, temporal variability, and incomplete observations.

Publications

Publications included in this thesis:

- I **Optimizing the hydraulic power take-off system in a wave energy converter (Published)**
Shokoufa Zeinali, Magnus Wiktorsson, Jan Forsberg, Georg Lindgren, Johan Lindström
Ocean Engineering, Volume 310, Part 1, 2024
- II **The effects of sea-state on optimal generator parameters for a wave energy converter (Published)**
Shokoufa Zeinali, Magnus Wiktorsson, Johan Lindström, Georg Lindgren, Jan Forsberg, Xinyuan Shao, Jonas Ringsberg
Innovations in renewable energies offshore, Proceedings of the 6th International Conference on Renewable Energies Offshore, November 2024
- III **Detection of Bark Beetle Attacks Using Time-Aggregated Satellite Data With Machine Learning (Published)**
Shokoufa Zeinali, Per-Ola Olsson, Ted Kronvall, Magnus Wiktorsson, Johan Lindström
Remote Sensing, 18(13), 2026
- IV **Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series (Submitted)**
Shokoufa Zeinali, Per-Ola Olsson, Ted Kronvall, Magnus Wiktorsson, Johan Lindström
- V **Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction**
Shokoufa Zeinali, Ted Kronvall, Langning Hou, Per-Ola Olsson, Johan Lindström

Publications not included in this thesis:

- VI **A comparison of approaches integrating power take-off systems into wave energy converters simulations (Published)**
Xinyuan Shao, Jonas W Ringsberg, Hua-Dong Yao, Erland Johnson, Jan Forsberg, Shokoufa Zeinali, Johan Lindström, Magnus Wiktorsson
Innovations in renewable energies offshore, Proceedings of the 6th International Conference on Renewable Energies Offshore, November 2024
- VII **Effect of front–back wave asymmetry on a hydraulic power take-off system (Published)**
Shokoufa Zeinali, Georg Lindgren, Jan Forsberg, Johan Lindström, Magnus Wiktorsson
Ocean Engineering, Volume 366, Part 4, 2026

Acknowledgements

First, I would like to express my sincere gratitude to my main supervisor, **Johan Lindström**, for always being there for me throughout this journey. Thank you for your guidance, support, and encouragement, and for always making time when I needed help or advice. I especially appreciate the freedom and trust you gave me to explore my own ideas and find my own way as a researcher. Whenever I felt stuck or unsure about the next step, I knew I could come to you and leave with a clearer mind and a way forward. Thank you for your patience, for believing in my work, and for supporting me through both the good and difficult parts of this PhD. I have learned a lot from working with you, not only about research, but also about how to approach problems with calmness and an open mind. I feel very lucky to have had you as my supervisor.

I would also like to thank *Georg Lindgren* and my co-supervisors, **Magnus Wiktorsson**, **Ted Kronvall**, and **Per-Ola Olsson**, for their valuable insights, discussions, and support throughout my PhD.

My thanks also go to **Fredrik Olsson** and **Erik Lindström** and **Michael MacDougall**, the heads of the division and department, and IT expert, for their support and for creating a welcoming and supportive working environment.

I am very grateful to my colleagues who made these years more enjoyable. **Linda Hartman**, thank you for all the nice experiences we shared while teaching together. **Niloofar**, thank you for all your help and support along the way. **Stas**, **Satyaki**, **David**, **Maria**, and **Tadas**, thank you for all the warm conversations, good moments, and friendly atmosphere. You made everyday life at work much more enjoyable.

Most importantly, I would like to thank my family, without whom I could never have made it this far. To my **parents**, thank you for your unconditional love, support, and belief in me throughout every stage of my life. No matter how far away I have been, I have always felt that you were there for me. I am sorry for all the moments I have missed and for being far away when I wished I could have been there with you. Thank you for understanding the choices that took me far from home and for always giving me the strength and freedom to follow my own path. Everything I have achieved carries a part of your love and support with it.

To my siblings, **Shakiba**, **Sadra**, and **Mehrdad**, thank you for always being part of my life, for supporting me, making me laugh, and reminding me that there is a whole world outside the PhD. Thank you for your patience, for hosting

me in Canada, for your love and support, and for sharing both the difficult and happy moments of this journey with me. Having you beside me made the difficult days easier and the good days much better. I could not have done this without you.

A very special thanks goes to my Iranian crew, who brought so much friendship, laughter, and warmth into these years. **Rezvan**, thank you for all the genuine support and all the laughs we have shared, and for teaching us the most subtle revenges and reposting whatever you see on Instagram. You have such a big heart, and you are one of the kindest and most supportive people I know. **Elham**, my friend firmly on the left, I wish you all the best, and still hope that one day you will find your way to the other side of politics. **Asie**, my wind-driven friend, whose side seems to change depending on which way the wind blows, I hope you finally settle on one side one day. And **Zahra**, the best chef and boss, thank you for all the amazing food and cakes. To all of you, thank you for all the support, the girls' nights, the endless conversations, and what often felt like our own group therapy sessions. I am very grateful to have had you by my side throughout this journey.

And of course, the boys: **Ata**, my right-wing true monarchist friend, thank you for the friendship and all the conversations and debates along the way. **Mehran**, **Mohammad**, and **Hadi**, thank you for all the good times, laughs, and discussions, even though, unfortunately, you somehow ended up on the wrong side! You all made life in Sweden much more fun and memorable, although I barely laughed at your jokes; sorry about that! I have bad taste in jokes.

Finally, I am grateful to everyone who has been part of this journey, in one way or another. This PhD has been much more than research, papers, and deadlines. It has also been about the people I met, the friendships I made, and the memories I will carry with me long after this chapter is finished.

Funding

Funding information: This research was supported by Swedish Energy Agency grant 50197-1 and carried out in cooperation with the company Waves4Power. Financial support has also been received from the Swedish Strategic Research Area (SRA) MERGE (Modelling the Regional and Global Earth System). The study is a contribution to the SRA BECC (Biodiversity and Ecosystem Services in a Changing Climate) and to the Lund University profile area Nature-based future solutions.

Popular Summary in English

The Earth we live on is like a mother to us. From the oceans to the skies, protecting our environment is essential for future generations. But how can mathematics and statistics help us take better care of nature? Mathematical and statistical models provide powerful tools for understanding complex environmental systems and solving real-world challenges.

This thesis explores how mathematics, statistics, and artificial intelligence can contribute to renewable energy and environmental monitoring. The work addresses challenges ranging from harvesting renewable energy from ocean waves to using satellite imagery for monitoring forests and ecosystems.

The first part of the thesis focuses on the optimization of Wave Energy Converters (WECs). Ocean waves are a promising renewable energy source because they contain large amounts of energy and are more predictable and continuous than many other renewable sources. However, efficient energy extraction requires systems that can adapt to changing sea conditions.

In Papers I and II, the goal is to optimize the parameters of a WEC to maximize the generated energy. To do so, first, waves with different characteristics are simulated and used as input to the WEC model.

In Paper I, we investigate how the wave energy system can be adjusted to produce more electricity under different wave conditions. The results show that adapting the system to changing waves can significantly increase the amount of energy captured from the ocean. Building on these results, Paper II uses a more advanced and realistic model of the wave energy system. Here, we develop a control strategy that automatically determines how the system should operate under different wave conditions to extract as much energy as possible.

The second part of the thesis focuses on environmental monitoring using aerial and satellite imagery. This part of the thesis explores how remote sensing can be used to better understand forest health and improve large-scale environmental monitoring systems.

In Papers III and IV, we investigate how satellite observations can improve the detection of bark beetle attacks in forests by different machine learning methods. Using satellite image time series, together with environmental and forest features, we evaluate how early in the growing season bark beetle infestations can be detected. Early detection is important since it allows forest owners better opportunities to fight the outbreaks. Paper III also identifies the most important features and time periods for detecting bark beetle infestations. In Paper IV, we

introduce a framework that integrates prior knowledge about tree susceptibility, derived from environmental characteristics. The results indicate that integrating these sources of information can support monitoring forest health and, with timely action, help prevent the spread of bark beetles.

In Paper V, we develop a machine learning model to reconstruct the noisy parts of the shortwave infrared band with the help of visible and near-infrared bands. Based on the findings of Paper III, we identify the shortwave infrared band as one of the most important spectral bands for detecting forest attacks. However, due to sensor limitations, shortwave infrared data are often affected by noise. The results show that visible and near-infrared can be used as a spatial guide for reconstruction of the shortwave infrared band.

Contents

Abstract	ix
Publications	xi
Acknowledgements	xiii
Popular Summary in English	xv
List of Acronyms	xix
1 Introduction	1
1.1 Statistical Modeling of Renewable Energy	2
1.2 Machine Learning for Environmental Monitoring	2
1.3 Thesis Perspective	4
2 Mathematics	5
2.1 Statistical Methods for Wave Modeling	5
2.2 Interpolation and Optimization	11
2.3 Extreme Gradient Boosting (XGBoost)	15
2.4 Deep Generative Models	17
2.5 Attention	23
3 Waves and Wave Energy	27
3.1 Ocean Waves	28
3.2 Wave Energy	30
3.3 Wave Modelling	34
3.4 Analysis of the WEC	39
3.5 Paper I	41
3.6 Paper II	42
4 Remote Sensing	43
4.1 Fundamentals of Remote Sensing	43
4.2 Sensor Platforms	47
4.3 Bark Beetles and Their Spectral Response	50
4.4 Multispectral Image Fusion	52
4.5 Paper III	53
4.6 Paper IV	54
4.7 Paper V	55

References	56
Author contributions	67
Paper I: Optimizing the hydraulic power take-off system in a wave energy converter (Published)	69
Paper II: The effects of sea-state on optimal generator parameters for a wave energy converter (Published)	85
Paper III: Detection of Bark Beetle Attacks Using Time-Aggregated Satellite Data With Machine Learning (Published)	95
Paper IV: Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series (Submitted)	119
Paper V: Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction	163

List of Acronyms

AE autoencoder.

AI artificial intelligence.

CVAE conditional variational autoencoder.

DRS distance red and shortwave infrared.

ELBO evidence lower bound.

ESA European Space Agency.

GPM generalized Pierson–Moskowitz.

GSD ground sampling distance.

JONSWAP Joint North Sea Wave Project.

KL Kullback–Leibler.

LCoE leveled cost of energy.

ML machine learning.

MSI Multispectral Instrument.

NBR normalized burn ratio.

NDRE normalized difference red-edge index.

NDRS normalized distance red and shortwave infrared.

NDVI normalized difference vegetation index.

NDWI normalized difference water index.

NIR near-infrared.

OWC oscillating water column.

PM Pierson–Moskowitz.

PSD power spectral density.

PTO power take-off.

RE red edge.

RGB red, green, and blue.

SAR synthetic aperture radar.

SWIR shortwave infrared.

UAV unmanned aerial vehicle.

VAE variational autoencoder.

VIS visible spectrum.

VNIR visible and near-infrared.

WAFO Wave Analysis for Fatigue and Oceanography.

WEC wave energy converter.

XGBoost extreme gradient boosting.

Chapter 1

Introduction

In recent years, the need to address complex environmental challenges has become increasingly urgent. Human activities since the Industrial Revolution have led to significant increases in greenhouse gas emissions, resulting in global warming and long-term shifts in climate patterns [43]. These changes are associated with a wide range of impacts, including more frequent extreme weather events, ecosystem degradation, and increased pressure on infrastructure and public health systems. Addressing these challenges requires not only reducing emissions through sustainable technologies, but also improving our ability to monitor, understand, and respond to environmental changes.

To support these efforts, a range of statistical and computational modelling approaches can be used to analyse environmental systems and extract information from increasingly large and complex datasets. Environmental systems such as oceans, forests, and the atmosphere are inherently dynamic, nonlinear, and influenced by a wide range of interacting processes. Statistical methods and established modelling approaches provide important tools for characterising these systems and quantifying relationships and uncertainties. At the same time, the increasing availability of large-scale, high-dimensional, noisy, and incomplete datasets has created opportunities for artificial intelligence (AI) and machine learning (ML) methods to complement these approaches by identifying patterns, relationships, and dynamics directly from data [11].

Building on these challenges, this thesis focuses on two areas of environmental research: renewable energy and remote sensing for environmental monitoring. The first addresses the development and optimisation of renewable energy technologies, while the second focuses on the use of Earth observation data to moni-

tor and understand environmental changes. Together, these areas contribute to the broader goal of supporting sustainable energy systems and improving the monitoring and management of changing environmental conditions.

1.1 Statistical Modeling of Renewable Energy

Renewable energy systems are key to reducing reliance on fossil fuels and mitigating climate change. However, their performance depends strongly on environmental conditions, which are often variable and uncertain. Wave energy, derived from ocean surface waves, is a promising renewable source due to its high energy density and predictability [55, 59]. Nevertheless, efficiently capturing this energy requires careful system design and adaptation to changing sea states.

In the first part of this work (Papers I and II), we focus on the optimization of a wave energy converter (WEC) using statistics and data-driven methods. The stochastic nature of ocean waves is modeled using statistical approaches, allowing the simulation of a wide range of realistic sea states (see section 3.4). A WEC system developed by the company Waves4Power is studied, with particular emphasis on the power take-off (PTO) mechanism by which the WEC converts mechanical motion into electrical energy. In Paper I, the pressure of the accumulator tank is optimized using a simplified model, and a relationship between sea state and optimal operating conditions is established by concave regression. In Paper II, using a more detailed model, a control strategy is designed to optimize the generator speed to improve the WEC’s performance further. These studies demonstrate that statistical representations of environmental variability can effectively guide the design of adaptive control strategies, improving both efficiency and robustness.

1.2 Machine Learning for Environmental Monitoring

Environmental monitoring is an important area where AI plays a transformative role. Advances in satellite and aerial imaging technologies now allow continuous observation of the Earth’s surface at multiple spatial and temporal scales [85], providing valuable information about vegetation, water resources, land use, and environmental disturbances. However, the resulting datasets are often large, heterogeneous, and affected by uncertainties such as noise, missing values, and irregular sampling. ML models, particularly deep learning approaches, often

related to computer vision, are well-suited to handle such data, as they can learn complex spatial and temporal structures [53].

Forests play a crucial role in carbon sequestration but are highly vulnerable to stressors such as drought, fire, and insect outbreaks. Among these, bark beetle infestations have become a major concern in many regions [42]. Early-stage detection of infestations is challenging because the initial physiological changes in trees generate only subtle spectral signatures, requiring high-quality, noise-free imagery. However, a further challenge in optical remote sensing is the presence of missing or corrupted data caused by image-capture noise or atmospheric effects [9]. This is particularly problematic for spectral bands such as the shortwave infrared (SWIR), which is critical for detecting vegetation water stress but is often degraded by cloud cover and atmospheric interference. In Papers III, IV, and V, we address these challenges using ML methods applied to aerial and satellite imagery for environmental monitoring.

Paper III uses Sentinel-2 satellite imagery in combination with static environmental features. Using these data, we investigate how early bark beetle attacks can be detected through a supervised learning framework based on extreme gradient boosting (XGBoost) [10], and identify which spectral regions contribute most to distinguishing between healthy and infested forest stands. This analysis provides insight into the physical drivers of detection and informs feature selection for more advanced models.

In Paper IV, we extend this approach by modeling the temporal dynamics of vegetation using a conditional variational autoencoder (CVAE) [73]. First, each tree’s susceptibility is modelled by an XGBoost model with static features. Then the CVAE learns latent representations of healthy and attacked vegetation dynamics from time-series satellite data, with the trees’ susceptibilities acting as priors for the latent representation. These latent representations are then classified into healthy and infested classes in a semi-supervised setting that does not need temporally aligned labels, allowing the model to capture complex temporal patterns and enable early detection of disturbances even under limited or uncertain labeling conditions.

In Paper V, we employ a U-Net architecture [69] for cross-sensor band reconstruction, specifically to recover missing or distorted spectral information. We focus on reconstructing the SWIR band from the available multispectral bands acquired from a different sensor with spatial and spectral mismatch. The SWIR band is sensitive to vegetation water content and therefore critical for detecting early-stage stress signals. By learning spatial patterns from available bands, the model restores the distorted band and improves input quality for analysis.

1.3 Thesis Perspective

Although the two parts of this thesis address different applications, renewable energy optimization and environmental monitoring, they are unified by a common idea of using mathematical modelling to solve environmental challenges. Both rely on data-driven approaches to model complex environmental systems characterized by uncertainty, temporal variability, and incomplete observations. In the wave energy case, the goal is to optimize system performance under stochastic environmental forcing. In the remote sensing case, the goal is to detect subtle deviations from normal behavior in high-dimensional time series data.

By integrating environmental statistics with modern AI techniques, this thesis contributes to the development of methods that support both sustainable energy production and effective environmental monitoring. The presented works illustrate how data-driven models can be used to analyze environmental systems and adaptive strategies in the face of a changing climate.

Chapter 2

Mathematics

In this chapter, we review the mathematical, statistical, and ML methodologies employed in this thesis. We first introduce the classical statistical techniques applied in the wave optimization in Papers I and II, after which we analyze the ML approaches utilized in the remote sensing Papers (III, IV, and V).

2.1 Statistical Methods for Wave Modeling

Ocean waves are inherently variable in time and space, and individual wave realizations cannot be predicted deterministically over long periods. This natural variability motivates representing sea states as stochastic processes, where the wave elevation is treated as a random function of time. Rather than representing measurement noise or uncertainty in the estimated spectral density, the stochastic formulation captures the intrinsic variability of the underlying wave field.

2.1.1 Basics of Stochastic Processes

To effectively model the varying ocean waves that a WEC is exposed to, it is necessary to utilize stochastic processes. Stationary processes are characterized by statistical properties that are invariant over time. If $x(t)$ is a stochastic

process, its mean and autocovariance function are defined as:

$$\begin{aligned} m(t) &= \mathbb{E}(x(t)), \\ r(t + \tau, t) &= \mathbb{E}[(x(t + \tau) - m(t + \tau))(x(t) - m(t))]. \end{aligned}$$

For a stationary process, the mean is constant, and the autocovariance depends only on the time lag between two observations rather than on their absolute positions in time. Hence,

$$r(\tau) = r(t + \tau, t).$$

The corresponding normalized quantity is the autocorrelation function, $\rho(\tau) = \frac{r(\tau)}{r(0)}$, where $r(0)$ is the variance of the process [54].

The autocovariance describes the temporal dependence of a stationary stochastic process in the time domain. The same process can also be characterized in the frequency domain through its power spectral density (PSD). The two descriptions are related through the Fourier transform (explained below),

$$S(\omega) = \mathcal{F}\{r(\tau)\},$$

where $S(\omega)$ describes how the variance, or equivalently the power, of the stochastic process is distributed over frequency. Thus, the autocovariance and the spectral density provide complementary descriptions of the same stationary process.

To illustrate this relationship, figure 2.1 shows a simulation of a simple stationary process, an AR(1) process $x(t) = ax(t - 1) + \varepsilon(t)$ with $\varepsilon(t)$ are i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$ and $|a| < 1$. Panel (a) shows one realisation of the process over time. Panel (b) shows the autocorrelation function $\rho(\tau) = a^{|\tau|}$, which decays as the lag τ increases, meaning observations far apart in time are less similar. The empirical autocorrelation computed from the simulated signal closely follows the analytical curve, confirming stationarity. Panel (c) shows the corresponding PSD, which concentrates power at low frequencies, reflecting the smooth, slowly-varying nature of the process.

For wave modelling, the frequency-domain representation is particularly useful because sea states are commonly characterized by wave spectra. To establish how a random time-domain process can be constructed from such a spectrum, it is first useful to introduce its Fourier representation.

Fourier Theory

Fourier theory provides a representation of a time-dependent function in terms of sinusoidal components at different frequencies. For a periodic function $\zeta(t)$

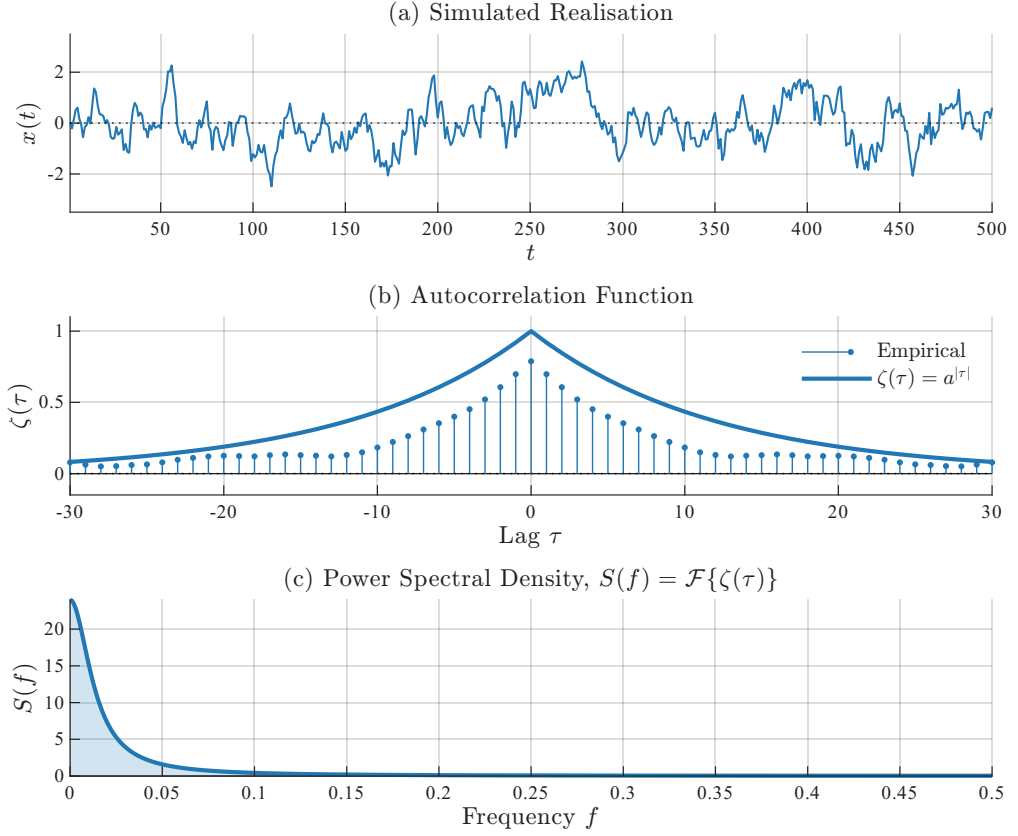


Figure 2.1: Illustration of the spectral representation of a stationary AR(1) process. (a) A simulated realisation $x(t)$. (b) Autocorrelation function $\rho(\tau) = a^{|\tau|}$: empirical (stems) and analytical (line). (c) Power Spectral Density $S(\omega) = \mathcal{F}\{\rho(\tau)\}$, showing how signal power concentrates at low frequencies.

with period T , the Fourier series [5] is:

$$\zeta(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right) \right],$$

where a_k and b_k are the Fourier coefficients defined as:

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} \zeta(t) \cos\left(\frac{2\pi kt}{T}\right) dt, \quad k = 0, 1, 2, \dots$$

$$b_k = \frac{2}{T} \int_{-T/2}^{T/2} \zeta(t) \sin\left(\frac{2\pi kt}{T}\right) dt, \quad k = 1, 2, \dots$$

Although $\zeta(t)$ is defined continuously in time, the Fourier series represents it using a discrete set of frequencies,

$$f_k = \frac{k}{T},$$

or, in terms of angular frequency,

$$\omega_k = \frac{2\pi k}{T}.$$

The Fourier representation therefore transforms the description of the signal from the time variable t to a representation indexed by frequency.

Fourier Representation of a Stochastic Process

For a deterministic function, the Fourier coefficients a_k and b_k are fixed quantities. A zero-mean, $\mathbb{E}[Z_t] = 0$, stochastic process can be constructed by setting $a_0 = 0$ and treating the remaining coefficients as random variables. Denoting these random coefficients by α_k and β_k , a stochastic Fourier representation may be written in the form

$$Z(t) = \sum_k [\alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t)]. \quad (2.1)$$

For a Gaussian process α_k and β_k are independent zero-mean Gaussian random variables with equal variance σ_k^2 , i.e. $\alpha_k, \beta_k \sim \mathcal{N}(0, \sigma_k^2)$, see Brodtkorb et al. [7].

To determine the variances, σ_k^2 , we use the link between the autocovariance $r(\tau)$ and the spectral density $S(\omega)$. This link is given by the Wiener–Khinchin theorem [83]:

$$r(\tau) = \int_{-\infty}^{\infty} S(\omega) e^{i\omega\tau} d\omega, \quad (2.2)$$

so that $S(\omega)$ is the Fourier transform of $r(\tau)$. Equivalently, $S(\omega) d\omega$ gives the contribution to the variance from frequencies in the interval $[\omega, \omega + d\omega]$, so that the total variance of the process is

$$\text{Var}(Z) = r(0) = \int_{-\infty}^{\infty} S(\omega) d\omega.$$

The spectral density therefore determines how much variance is associated with each frequency interval. Since direct simulation of a continuous spectrum is not feasible, the frequency axis is discretised into N bins of uniform width

$\Delta\omega = 2\pi/T$, where T is the simulation duration. When the continuous spectrum is represented by discrete frequencies separated by $\Delta\omega$, the variance assigned to the k th frequency component is $S(\omega_k)\Delta\omega$.

Since α_k and β_k are independent zero-mean Gaussian random variables with equal variance σ_k^2 , the variance contributed by the k th Fourier component is

$$\text{Var}(\alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t)) = \sigma_k^2 \cos^2(\omega_k t) + \sigma_k^2 \sin^2(\omega_k t) = \sigma_k^2.$$

Matching this to the spectral contribution $S(\omega_k)\Delta\omega$ gives the identification

$$\sigma_k^2 = S(\omega_k) \Delta\omega,$$

so that the Fourier coefficients are sampled as

$$\alpha_k, \beta_k \sim \mathcal{N}(0, S(\omega_k) \Delta\omega),$$

independently for each k .

Because the representation originates from independent Gaussian sine and cosine coefficients, the resulting process is Gaussian. For the zero-mean case considered here, its covariance is determined by the prescribed spectral density. For an individual spectral component, we have,

$$r_k(\tau) = \Delta\omega S(\omega_k) \cos(\omega_k \tau).$$

Summing across all spectral components, this ensures consistency with the Wiener–Khinchin theorem: the covariance of the simulated process,

$$r(\tau) = \sum_k S(\omega_k) \Delta\omega \cos(\omega_k \tau) \xrightarrow{\Delta\omega \rightarrow 0} \int_{-\infty}^{\infty} S(\omega) \cos(\omega \tau) d\omega,$$

recovers the prescribed spectral density in the continuous limit. Consequently, the stochastic time-domain process generated from equation (2.1) retains the covariance structure associated with the prescribed spectrum.

Alternatively, the sine and cosine terms at each frequency in (2.1) can also be expressed in amplitude–phase form. Defining

$$A_k = \sqrt{\alpha_k^2 + \beta_k^2}, \tag{2.3}$$

and

$$\theta_k = \text{atan2}(\beta_k, \alpha_k), \tag{2.4}$$

The corresponding frequency component can be represented by a sinusoid with random amplitude A_k and random phase θ_k ,

$$\alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t) = A_k \sin(\omega_k t + \theta_k).$$

If α_k and β_k are independent zero-mean Gaussian random variables with equal variance, $\sigma_k^2 = S(\omega_k) \Delta\omega$, then their resulting magnitude, A_k , defined in equation (2.3), follows a scaled Rayleigh distribution, while the corresponding phase is uniformly distributed over $[0, 2\pi)$. The random amplitude can now be written as

$$A_k = \sqrt{S(\omega_k) \Delta\omega} \cdot R_k$$

where R_k are independent Rayleigh-distributed random variables with density function $f_R(r) = r \exp(-r^2/2)$, $r > 0$.

This relationship is particularly useful for stochastic wave simulation, since random Fourier coefficients can therefore be represented equivalently by random amplitudes and phases.

2.1.2 Simulation

One common approach to characterizing waves is through wave spectrum analysis (discussed in more detail in section 3.3.2). First, a suitable family of wave spectra is selected. These spectra are often parameterized by significant wave height H_s and a characteristic period. Significant wave height is the average height of the highest one-third of waves in a sea state. The characteristic period is typically taken as the peak period T_p , the period corresponding to the maximum of the spectral density, or the energy period T_e . These concepts are fully discussed in chapter 3.

After selecting an appropriate family of wave spectra, the parameters (height and period) have to be determined. These can either be estimated from data [7, 54] or based on previously estimated and tabulated values from marine research centers. Here we have used data from the European Marine Energy Centre at Billia Croo, Orkney [24] that provides parameters and frequency of occurrence for different sea states (see Table 1 in Paper I). The frequency of occurrences provided in the data allows us to account for the variability in sea states during a year.

Given a prescribed spectral density $S(\omega)$, different stochastic realizations of the wave elevation can be generated by sampling the Fourier coefficients associated

with each discrete frequency. Following the Fourier representation of stochastic processes in (2.1) we have two options for sampling

$$Z(t) = \begin{cases} \sum_{k=0}^{N-1} \left[\alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t) \right], & \text{or} \\ \sum_{k=0}^{N-1} \underbrace{\sqrt{S(\omega_k) \Delta \omega} R_k}_{=A_k} \cos(\omega_k t + \theta_k), \end{cases} \quad (2.5)$$

where the random numbers are independent and drawn from either

$$\begin{cases} \alpha_k, \beta_k \sim \mathcal{N}(0, S(\omega_k) \Delta \omega), & \text{or} \\ R_k \sim \text{Rayleigh}, \theta_k \sim \mathcal{U}(0, 2\pi). \end{cases} \quad (2.6)$$

This simulation framework is the direct operational link between the abstract statistical theory above and the physical wave modelling described in chapter 3. In that context, the prescribed spectral density $S(\omega)$ is not arbitrary but is chosen to match observed sea states, for example through the Pierson–Moskowitz (PM) or Joint North Sea Wave Project (JONSWAP) spectral families described in section 3.3.2. The parameters of these spectra (significant wave height H_s and peak period T_p) determine the autocovariance structure of the simulated wave via equation (2.2), and different draws of the random amplitudes A_k and phases θ_k produce statistically equivalent but distinct wave-elevation time series, each a physically plausible realization of the same sea state. It is these simulated time series that are fed as input to the Simulink model of the WEC in Papers I and II.

2.2 Interpolation and Optimization

Optimization plays a crucial role in improving the performance and efficiency of engineering systems. Various optimization methods can be employed depending on the nature of the problem, such as the availability of derivatives, the dimensionality of the search space, and the complexity of the objective function.

In this thesis, optimization is applied to wave energy applications where multiple variables influence the outcome. The complex and often unpredictable nature of marine environments makes robust and reliable optimization methods essential for WEC parameter tuning.

2.2.1 Robust modeling

As mentioned in chapter 1, we optimize the accumulator pressure of a WEC in Paper I. The Simulink model describing the WEC, detailed in section 3.4.2, will be run with many different simulated waves, creating a set of possible outputs that we need to optimize over. To account for variability and uncertainties in these outputs, a robust method of fitting the response surface that can describe the model's general behavior is needed.

Piecewise Polynomial Fitting [8] is a powerful technique in numerical analysis and data modeling that divides a dataset into segments and fits a polynomial to each segment. This method allows for greater flexibility and accuracy in capturing the underlying trends and patterns within the data, particularly when dealing with complex or non-linear relationships. By joining these polynomial segments smoothly, Piecewise Polynomial Fitting provides a robust approach to interpolation, smoothing, and function approximation, making it widely applicable in fields such as computer graphics, signal processing, and statistical analysis.

Depending on the application and required smoothness, different types of splines can be used:

Linear Splines: These are piecewise linear functions that ensure continuity at the knots but may have sharp corners, as the first derivative is not continuous.

Quadratic Splines: These use quadratic polynomials, offering better smoothness than linear splines but less flexibility than cubic splines.

Cubic Splines: These use cubic polynomials, providing a smooth curve with continuous first and second derivatives at the knots. Cubic splines are commonly used because they offer a good balance between flexibility and smoothness.

The most common form of splines is the cubic spline; an example can be seen in figure 2.2. A cubic polynomial has four free constants per segment, which provide enough degrees of freedom to ensure the interpolant is continuously differentiable with a continuous second derivative across the interval. The cubic spline construction does not require the interpolant's derivatives to match the function it approximates, even at the nodes [8]. Given a function f defined on an interval, $[a, b]$, and a set of nodes $a = x_0 < x_1 < \dots < x_n = b$, a cubic spline interpolant S of f is a function that satisfies the following conditions:

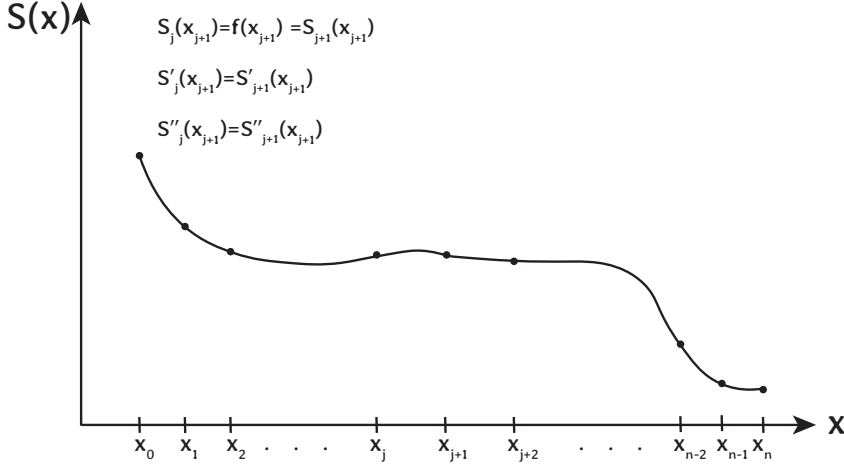


Figure 2.2: Example of a cubic spline fit to data, illustrating piecewise polynomial segments joined at the knots.

- (a) $S(x)$ is a cubic polynomial, denoted $S_j(x)$, on the subinterval $[x_j, x_{j+1}]$ for each $j = 0, 1, \dots, n-1$;
- (b) $S_j(x_j) = f(x_j)$ and $S_j(x_{j+1}) = f(x_{j+1})$ for each $j = 0, 1, \dots, n-1$;
- (c) $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ for each $j = 0, 1, \dots, n-2$; (Implied by (b).)
- (d) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ for each $j = 0, 1, \dots, n-2$;
- (e) $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ for each $j = 0, 1, \dots, n-2$;
- (f) One of the following sets of boundary conditions is satisfied:
 - (i) $S''(x_0) = S''(x_n) = 0$ (natural (or free) boundary);
 - (ii) $S'(x_0) = f'(x_0)$ and $S'(x_n) = f'(x_n)$ (clamped boundary).

Piecewise polynomials offer a simple but robust model for data [8]. Dividing the data into smaller segments, with a polynomial independently fitted to each segment, ensures that changes or noise in one part of the data have a limited effect on the overall model. This is unlike high-degree polynomials, where a small change in one data point can influence the entire polynomial. Additionally, high-degree polynomials are susceptible to Runge's phenomenon [16], i.e. wild

oscillations, especially at the edges of the interval, making the model unstable and unreliable. By employing lower-degree polynomials within each segment, splines avoid these oscillations, resulting in a smoother and more stable curve that better reflects the underlying trends in data.

Shape Language Modeling (SLM)

MATLAB offers several tools for Piecewise Polynomial Fitting, one of which is the Shape Language Modeling (SLM) toolbox [19]. SLM allows for flexible shape modeling with control over smoothness, continuity, and other properties. Key features include:

Piecewise Polynomial Fitting: Fitting splines to data using piecewise polynomial segments.

Constraint Handling: Applying constraints, such as monotonicity, convexity, and boundary conditions, to the fitted curves

Interactive Tools: Visualize and interactively adjust fits using GUI tools.

Robust Fitting: Provides reliable fits even with noisy or challenging datasets.

Customizability: Customize the fitting process through various options and parameters.

2.2.2 Nelder–Mead

One of the common optimization methods is the Nelder–Mead algorithm [58]. Also known as the downhill simplex method, Nelder–Mead is widely used to minimize or maximize derivative-free objective functions in multidimensional parameter spaces [50].

For a problem with n parameters, the algorithm begins with a simplex consisting of $n + 1$ points. At each iteration, these points are ranked according to their objective-function values, from best to worst. The worst point is then replaced using one of four operations: reflection, expansion, contraction, or shrinkage. Reflection moves the worst point through the centroid of the remaining points. If the reflected point provides a substantial improvement, expansion continues further in the same direction. If reflection is unsuccessful, contraction moves the point closer to the centroid. If contraction also fails, the simplex is shrunk

towards the best point. These steps are repeated until the simplex size or the differences between the objective-function values become sufficiently small.

Four scalar parameters define the behavior of the Nelder–Mead method:

Reflection coefficient (ρ): Determines the reflection of the simplex through its centroid.

Expansion coefficient (χ): Controls the expansion beyond the reflected point.

Contraction coefficient (γ): Governs the contraction towards the best point.

Shrinkage coefficient (σ): Determines the shrinkage of the simplex towards the best point when necessary.

In MATLAB, `fminsearch` [76, 51] is a built-in function used for unconstrained optimization of a scalar function of several variables. It employs the Nelder–Mead simplex (direct search) algorithm to find the minimum of the function without needing gradient information.

The methods described above, stochastic wave modeling, spline fitting, and derivative-free optimization, address the wave energy applications of Papers I and II, where the goal is to optimize a physical system under environmental variability. The remainder of this chapter introduces the machine learning methods used in Papers III, IV, and V, where the goal shifts to learning structure directly from high-dimensional remote sensing data.

2.3 Extreme Gradient Boosting (XGBoost)

While Papers I and II use more classical statistical and optimization tools, Papers III, IV, and V rely more heavily on machine learning methods, ranging from simple tree-based models described in this section to more complex neural networks in section 2.4.

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable [10], used in Papers III and IV. It belongs to a family of ensemble learning methods that combine the predictions of multiple weak learners, typically decision trees, to create a single strong learner. A decision tree makes predictions by recursively splitting the data based on feature values, creating a sequence of simple decision rules that lead to a final prediction. The core principle of XGBoost is gradient boosting [25], where trees

are built sequentially, with each new tree attempting to reduce the errors of the existing ensemble. The model predicts an output by summing the predictions of K trees:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

where $\mathcal{F}$ are all trees in the model, $f_k \in \mathcal{F}$ is the output of the k -th regression tree and K is the total number of trees.

XGBoost extends traditional gradient boosting by incorporating an explicit regularization term into the objective function. The objective consists of a training loss that measures the difference between the observed values y_i and the predictions $\hat{y}_i$, together with a regularization term that penalizes overly complex trees:

$$\mathcal{L} = \sum_i L(y_i, \hat{y}_i) + \sum_k \Omega(f_k),$$

where the regularization term is defined as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|_2^2.$$

Here, T is the number of leaves in the tree, where a leaf is a terminal node reached after following the tree’s splitting rules and represents the final prediction for the samples assigned to it. The parameter γ controls the penalty for adding new leaves, while λ controls the L_2 regularization applied to the leaf weights. The vector w contains the prediction weights assigned to the leaves, such that a sample follows the decision rules of the tree until it reaches a leaf, whose corresponding weight contributes to the final prediction. Thus, γ primarily controls the tree structure by discouraging unnecessary splits and additional leaves, while λ limits the magnitude of the leaf predictions. Together, these regularization terms control model complexity and help reduce overfitting.

In forest disturbance studies, XGBoost is effective at handling static features such as topography (slope, aspect), soil characteristics, and stand structure. The XGBoost decision-tree structure can capture interactions between these features more effectively than linear models [10].

Although the XGBoost model used in Paper III generalizes well across space, it is less robust across time. Training on both satellite data and static features from a specific year likely overfitted to the specific spectral conditions of that year, harming performance when the model was applied to a different growing season. This limitation motivates the use of deep generative models, which can learn more generalizable latent representations of temporal dynamics and better handle variability across seasons.

2.4 Deep Generative Models

Generative models are a class of ML methods that learn the underlying data distribution, enabling them to generate new samples that resemble the training data. They can model the joint distribution of observations X and conditioning variables Y , $P(X, Y)$, or, in conditional generative models, the conditional distribution $P(X | Y)$ [29].

2.4.1 Variational autoencoder (VAE)

A commonly used generative model is the VAE [49], and to understand it, it is first necessary to grasp the principles of one of its building blocks, autoencoder (AE). An AE is a neural network trained in an unsupervised manner to reproduce its input as the output [68]. It consists of two main components (see also figure 2.3):

Encoder ($h = f(x)$): Maps the high-dimensional input x to a lower-dimensional latent representation h . The encoder is typically implemented as a neural network consisting of multiple layers that progressively transform the input into a compact representation.

Decoder ($\hat{x} = g(h)$): Reconstructs the input from the latent representation h . Similar to the encoder, the decoder is typically implemented as a multi-layer neural network that transforms the latent representation back to the original input space.

Both the encoder and decoder are commonly constructed using standard feed-forward neural network architectures, with their parameters learned jointly during training to minimise the reconstruction error, $x - \hat{x}$, [29].

Encoder–decoder architectures are widely used for representation learning and reconstruction tasks. By constraining the information flow through a so-called bottleneck, the AE is encouraged to learn the most salient structure in the data. However, standard AEs often overfit the latent space, effectively memorizing the training data rather than learning a smooth and meaningful latent distribution [29].

A standard AE is not considered a generative model because it learns a deterministic mapping between the input and a latent representation without imposing any probability distribution on the latent space. Consequently, new latent samples cannot be drawn and decoded to generate meaningful data. To that end,

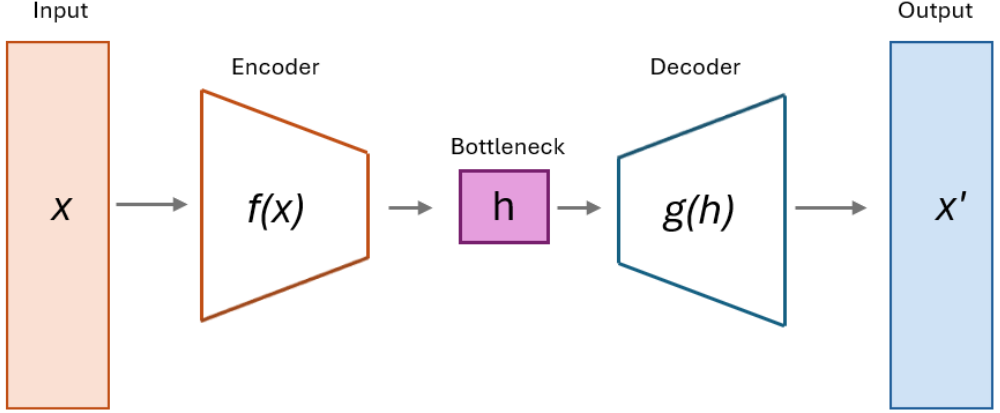


Figure 2.3: General architecture of an AE, illustrating the encoder that maps the input data to a lower-dimensional latent representation and the decoder that reconstructs the input from the learned latent representation.

its probabilistic counterpart, the VAE, uses a probabilistic and Bayesian perspective to approximate the posterior distribution of the latent variables given the input. Since the true posterior, $p(z|x)$, is generally intractable, the encoder learns a variational approximation $q_\phi(z|x)$, typically modeled as a Gaussian distribution parameterized by mean, μ , and variance, σ^2 , [12]. Specifically, the encoder learns:

$$q_\phi(z|x) = \mathcal{N}(z|\mu(x), \sigma^2(x)),$$

where ϕ denotes the encoder parameters, typically neural-network weights, to distinguish the AE bottleneck, which is fixed, from the VAE bottleneck, which is random; we denote the random latent variables as z rather than h .

This means that each input is mapped to a probability distribution rather than a single latent vector, thereby capturing uncertainty in the latent representation. During training, the approximate posterior is encouraged to remain close to a predefined prior distribution, usually a standard Gaussian, $p(z) = \mathcal{N}(0, I)$, by minimizing the Kullback–Leibler (KL) divergence [15], which quantifies how much a probability distribution differs from a reference distribution. This Bayesian formulation regularizes the latent space, forcing nearby points to correspond to similar data, resulting in a smooth and continuous latent space that supports meaningful interpolation and the generation of new samples.

Loss Function

The VAE optimizes the evidence lower bound (ELBO):

$$\mathcal{L} = \underbrace{\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]}_{\text{Reconstruction}} - \underbrace{D_{KL}(q_\phi(z|x) \parallel p(z))}_{\text{Regularization}},$$

where $p_\theta(x|z)$ is the decoder, or generative distribution, that creates an output $\hat{x}$, given latent states z , with trainable decoder parameters θ . By evaluating the generative distribution at x , the reconstruction loss ensures that the output $\hat{x}$ is close to the input x in some suitable error metric (e.g., MSE or cross-entropy), and the KL Divergence forces the learned latent distribution towards the standard Gaussian prior.

Reparameterization Trick

Since sampling from a probability distribution is inherently non-differentiable, directly drawing a latent vector from $q_\phi(z|x)$ would prevent gradients from propagating through the encoder during training. To overcome this issue, the VAE employs the reparameterization trick (also known as stochastic backpropagation), which expresses the stochastic latent variable as a deterministic transformation of a random variable sampled from a fixed distribution. Specifically, the latent variable is computed as

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I),$$

where $\odot$ denotes element-wise multiplication, μ and σ are the outputs of the encoder network, and ϵ is sampled independently from a standard multivariate Gaussian distribution. This formulation separates the source of randomness (ϵ) from the learnable parameters (μ and σ), making the sampling operation differentiable with respect to the encoder parameters. Consequently, gradients can be propagated through the network using a standard learning technique, i.e., backpropagation, enabling efficient end-to-end optimization [49, 66].

2.4.2 Conditional Variational Autoencoder

The CVAE extends the VAE by incorporating side information y (labels, class, metadata, or context) in both encoding and generation. This enables controlled generation rather than relying on unconditional sampling [73]. Figure 2.4 gives an overview of a CVAE.

The CVAE models the data conditional on the side information,

$$p(x|y)$$

with a conditional latent distribution, $q_\phi(z|x, y)$, and conditional reconstruction distribution $p_\theta(x|z, y)$. The encoder takes both data and labels, (x, y) , and outputs parameters for the latent distribution, $\mu(x, y)$ and $\sigma(x, y)$, while the decoder takes latent states and labels (z, y) to reconstruct or generate $\hat{x}$. Conditioning is typically implemented by concatenating y with the input or latent vector or via embedding layers.

To derive the loss function used in training CVAEs, we consider the log-probability of an input x given y under the parameterized distribution, obtained by multiplying and dividing by $p_\theta(x, z|y)$ and $q_\phi(z|x, y)$

$$\begin{aligned} \log[p_\theta(x|y)] &= \log \frac{p_\theta(x|y) p_\theta(x, z|y) q_\phi(z|x, y)}{p_\theta(x, z|y) q_\phi(z|x, y)} \\ &= \log \frac{p_\theta(x|y) q_\phi(z|x, y)}{p_\theta(x, z|y)} + \log \frac{p_\theta(x, z|y)}{q_\phi(z|x, y)} \\ &= \log \frac{q_\phi(z|x, y)}{p_\theta(z|x, y)} + \log \frac{p_\theta(x|z, y) p_\theta(z|y)}{q_\phi(z|x, y)} \\ &= \log \frac{q_\phi(z|x, y)}{p_\theta(z|x, y)} + \log[p_\theta(x|z, y)] - \log \frac{q_\phi(z|x, y)}{p_\theta(z|y)}. \end{aligned}$$

Since this identity holds for any z , multiplying both sides by $q_\phi(z|x, y)$ and integrating over z yields the expectation under $z \sim q_\phi(z|x, y)$:

$$\begin{aligned} \log[p_\theta(x|y)] &= \mathbb{E}_{q_\phi} \left[\log \frac{q_\phi(z|x, y)}{p_\theta(z|x, y)} \right] + \mathbb{E}_{q_\phi} [\log p_\theta(x|z, y)] - \mathbb{E}_{q_\phi} \left[\log \frac{q_\phi(z|x, y)}{p_\theta(z|y)} \right] \\ &= D_{\text{KL}}[q_\phi(z|x, y) \parallel p_\theta(z|x, y)] + \mathbb{E}_{q_\phi} [\log p_\theta(x|z, y)] \\ &\quad - D_{\text{KL}}[q_\phi(z|x, y) \parallel p_\theta(z|y)]. \end{aligned} \tag{2.7}$$

As in the VAE, the posterior $p_\theta(z | x, y)$ is intractable. Since the first KL divergence is always non-negative, subtracting it yields a quantity that is less than or equal to the conditional log-likelihood. This quantity therefore forms a lower bound on $\log p_\theta(x | y)$, known as the conditional ELBO:

$$\begin{aligned} \log[p_\theta(x|y)] &\geq \log[p_\theta(x|y)] - D_{\text{KL}}[q_\phi(z|x, y) \parallel p_\theta(z|x, y)] \\ &= \underbrace{\mathbb{E}_{q_\phi(z|x, y)} [\log p_\theta(x|z, y)]}_{\text{Reconstruction}} - \underbrace{D_{\text{KL}}(q_\phi(z|x, y) \parallel p_\theta(z|y))}_{\text{Regularization}}, \end{aligned} \tag{2.8}$$

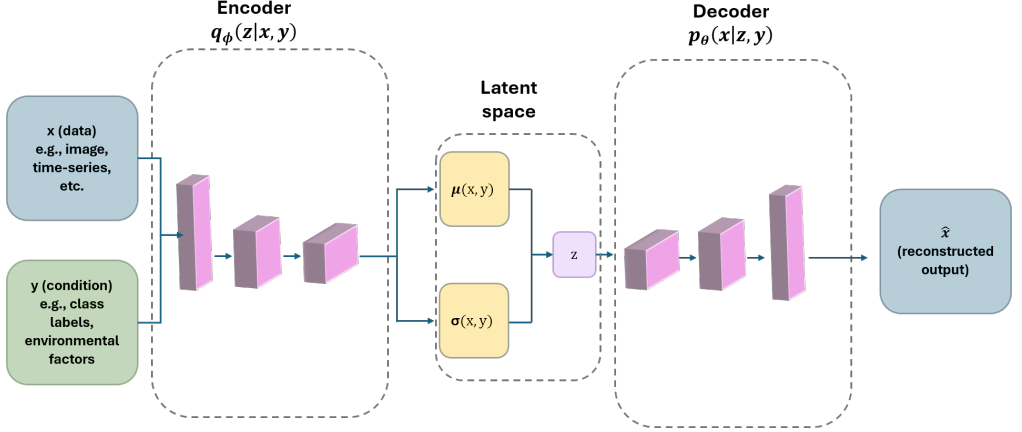


Figure 2.4: Overview of the CVAE architecture. The CVAE incorporates label information into the latent space, enforcing a structured, label-conditioned representation of the learned latent space.

where the prior $p_\theta(z|y)$ is typically chosen to be a standard normal $\mathcal{N}(0, I)$, or a class-conditional Gaussian $\mathcal{N}(\mu_y, \sigma_y^2 I)$ when label information is available.

The loss function is the negative conditional ELBO, decomposed into a reconstruction loss $\mathcal{L}_{\text{rec}}$ and a KL divergence loss $\mathcal{L}_{\text{KL}}$:

$$\mathcal{L}(x, y, \theta, \phi) = \underbrace{-\mathbb{E}_{q_\phi(z|x, y)}[\log p_\theta(x|z, y)]}_{\mathcal{L}_{\text{rec}}} + \underbrace{D_{\text{KL}}(q_\phi(z|x, y) \parallel p_\theta(z|y))}_{\mathcal{L}_{\text{KL}}}.$$

The reconstruction term ensures that $\hat{x}$ matches x given y , while the KL term regularizes the latent space as in the VAE. Conditioning on y supports both targeted generation of a specific class and improves reconstruction quality by providing contextual information to both the encoder and the decoder.

A natural extension replaces the standard isotropic prior $\mathcal{N}(0, I)$ with a class-conditional prior $\mathcal{N}(\mu_y, I)$, where $\mu_y \in \mathbb{R}^d$ is a learned or fixed mean vector associated with class y . The KL term becomes:

$$D_{\text{KL}}(q_\phi(z|x, y) \parallel \mathcal{N}(\mu_y, I)),$$

which encourages the encoder to map samples from class y to a region of latent space centered on μ_y rather than on the global origin.

Cosine-similarity regularisation An alternative, or complementary, approach uses cosine similarity to enforce angular separation in the latent space. For a mini-batch, i.e., a subset of training samples processed together during

one training iteration, with encoded means μ_i , a contrastive loss rewards high similarity between same-class pairs and penalizes it for different-class pairs:

$$\mathcal{L}_{\text{cos}} = -\frac{1}{|\mathcal{S}^+|} \sum_{(i,j) \in \mathcal{S}^+} \cos(\mu_i, \mu_j) + \frac{1}{|\mathcal{S}^-|} \sum_{(i,j) \in \mathcal{S}^-} \cos(\mu_i, \mu_j),$$

where $\mathcal{S}^+$ and $\mathcal{S}^-$ denote the sets of positive (same-label) and negative (different-label) pairs, and $\cos(u, v) = u^\top v / (\|u\| \|v\|)$. Unlike Euclidean distance, cosine similarity is invariant to the magnitude of the latent codes and is therefore well-suited to high-dimensional spaces where norms tend to concentrate around a narrow range. Combined with the class-conditional prior, this encourages a latent manifold, i.e., a region of the latent space occupied by the learned representations, in which the angular position encodes class identity, supporting both structured generation and discriminative ability in downstream tasks.

The full training objective combines reconstruction, KL regularisation, cluster regularisation, and cosine contrastive loss.

2.4.3 U-Net

U-Net is another widely used encoder–decoder architecture originally proposed for biomedical image segmentation [69]; see figure 2.5 for an illustration of the general U-Net architecture. Its strength lies in combining contextual information from deep layers with high-resolution spatial details from shallow layers through so-called skip connections, making the U-net particularly effective for image reconstruction and dense prediction tasks. The architecture consists of two parts: an encoder (contracting path) that progressively reduces spatial resolution while increasing feature abstraction, and a decoder (expanding path) that gradually restores spatial resolution to reconstruct the output. Skip connections are a defining component of U-Nets, transferring high-resolution features from the encoder to the decoder and thus preserving spatial details, improving gradient flow, and reducing information loss caused by pooling operations, which progressively reduce the spatial resolution of feature maps to capture higher-level contextual information. However, standard skip connections transfer encoder features directly to the decoder without explicitly selecting which features are relevant. Consequently, irrelevant or noisy features may also be propagated. This can be particularly problematic in multi-modal settings, where information from different data sources or modalities, such as VNIR and SWIR imagery, is processed or combined and may differ in quality and information content. Attention, discussed in section 2.5 below, provides a mechanism for selectively emphasizing relevant features while suppressing less informative or noisy ones.

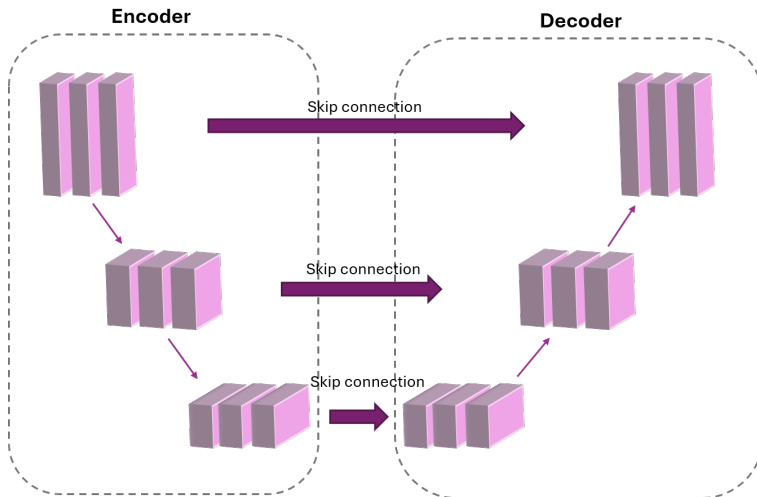


Figure 2.5: Overall architecture of the U-Net. The network follows an encoder-decoder structure in which the encoder progressively extracts hierarchical feature representations while reducing spatial resolution, and the decoder progressively restores the spatial resolution to produce the final output. Skip connections transfer feature maps from corresponding encoder stages to the decoder, preserving fine-scale spatial information that may otherwise be lost due to down-sampling.

To address multi-modal data, the U-Net can be extended into a dual- or multi-path architecture, where separate encoders process different inputs. This design allows the model to learn modality-specific representations, avoids early mixing of incompatible features, and enables controlled fusion at later stages [17].

2.5 Attention

The so-called attention mechanism [81] enables a model to dynamically focus on the most relevant parts of an input. Instead of processing all features equally, attention assigns adaptive weights that highlight informative components and suppress less useful ones.

In a neural network (such as a CNN, CVAE, or U-Net), attention operates on intermediate feature maps, which are learned representations of the input produced by preceding network layers. A feature map contains information extracted from the input, with different channels representing different learned features, ranging from simple patterns such as edges and textures in early layers to more complex structures in deeper layers. Given a feature map $X \in \mathbb{R}^{H \times W \times C}$, where H and W represent its spatial dimensions and C is the number of feature channels, the spatial dimensions are first flattened to obtain a sequence of $N = H \times W$ tokens, yielding $X \in \mathbb{R}^{N \times C}$. Each token therefore represents the

learned features at one spatial location. From these tokens, three elements are computed through learned linear projections: the *query* Q , *key* K , and *value* V . The query represents what the network is looking for, the key represents the information available at each position, and the value represents the actual feature content to be aggregated based on the resulting attention weights.

The output is computed as a weighted combination of values, where the weights are determined by the similarity between queries and keys:

$$\text{Attention}(Q, K, V) = \psi(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

where $\psi(\dots)$ denotes the attention operator, d_k is the projection dimension of the queries and keys, and d_v is the projection dimension of the values. In multi-head attention, the feature representation is divided into h attention heads, where each head is an independent attention operation with its own learned query, key, and value projections. This allows different heads to focus on different relationships or patterns in the input simultaneously. In many implementations, $d_k = d_v = d_{\text{model}}/h$, where d_{model} is the overall feature dimension and h is the number of attention heads. Dividing the query–key dot products by $\sqrt{d_k}$ prevents them from becoming excessively large as d_k increases, which could cause the softmax function to saturate and produce very small gradients. The softmax function then converts the similarity scores for each query into normalized attention weights, defined as $\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$. These weights are non-negative and sum to 1, and can therefore be interpreted as a probability distribution describing how strongly a given query attends to each key. A larger weight indicates that the information associated with that key is considered more relevant to the query. The resulting weights are subsequently used to compute a weighted combination of the corresponding values.

In self-attention, Q , K , and V are all derived from the same input feature map X , allowing the model to capture dependencies between different positions within the same representation. In cross-attention, Q is derived from one feature space (e.g., a decoder layer in a U-Net), while K and V are derived from another (e.g., an encoder layer connected through a skip connection), allowing information from one representation to selectively guide another [81].

The computational cost is determined by the image dimensions. With $Q, K \in \mathbb{R}^{N \times d_k}$, the matrix multiplication QK^T requires $O(N^2)$ computations with respect to the number of spatial tokens N and produces an $N \times N$ similarity matrix. To address this quadratic bottleneck for high-resolution images, linearized attention approximations can be employed [47]. In standard attention, the softmax-based similarity between queries and keys requires pairwise inter-

actions between all tokens, preventing the computation from being directly rearranged to avoid the $N \times N$ attention matrix. Linearized attention instead expresses the query–key similarity using a kernel function $\phi(\cdot)$, i.e., a feature transformation applied independently to the query and key vectors such that their similarity can be computed from the transformed representations. This formulation allows the order of matrix multiplications to be rearranged using the associative property, avoiding explicit construction of the full pairwise similarity matrix:

$$\text{Attention}(Q, K, V) \approx \phi(Q) (\phi(K)^T V).$$

By computing $C = \phi(K)^\top V \in \mathbb{R}^{d_k \times d_v}$ first (cost $O(Nd_k d_v)$), and then $\phi(Q) C \in \mathbb{R}^{N \times d_v}$ (cost $O(Nd_k d_v)$), the total complexity becomes $O(Nd_k d_v)$. Since $d_k, d_v \ll N$ in practice, this reduces the computational cost from $O(N^2)$ to only $O(N)$. This approximation retains the non-negativity and weighting properties of standard attention while making it feasible for high-resolution images.

Attention mechanisms can operate along multiple dimensions depending on the structure of the data:

Channel Attention: Focuses on relationships between feature channels ($1 \times 1 \times C$). It reweights channels to emphasize important spectral or semantic responses. This is useful in multi-spectral data where different bands contribute differently.

Spatial Attention: Operates across spatial locations ($H \times W$). It assigns an importance weight to each pixel [37]:

$$x_{\text{att}}(i, j) = w(i, j) \cdot x(i, j),$$

where $w(i, j) \in [0, 1]$ is a learned spatial attention map that indicates the importance of pixel (i, j) . This allows the model to highlight relevant regions (e.g., edges, objects) and to suppress noise [84].

Temporal Attention: Applies to sequential or time-series data, where dependencies exist across time steps. Given a sequence $\{x_t\}_{t=1}^T$, temporal attention learns weights over time [27]:

$$x_{\text{att}} = \sum_{t=1}^T \alpha_t x_t, \quad \text{where} \quad \sum_t \alpha_t = 1,$$

and α_t are normalized attention weights assigned to each time step. Temporal attention is particularly useful in applications such as satellite time series analysis, where certain observations (e.g., seasonal peaks or anomalies) are more informative than others.

In Paper V, channel and spatial attention mechanisms are combined within the dual-encoder architecture to enable asymmetric cross-spectral feature exchange between visible and near-infrared (VNIR) and SWIR branches. This allows a U-Net model to reconstruct missing or corrupted remote sensing data in the SWIR spectral band by querying the available VNIR features.

Chapter 3

Waves and Wave Energy

Global energy demand has increased significantly due to population and industrial growth during the past 300 years. The consumption of fossil fuels has been the primary source of energy since the Industrial Revolution [67]. Due to the numerous downsides, such as global warming and air pollution, associated with burning fossil fuels, there has been a significant shift towards renewable energy in recent years.

Renewable energy is generated from natural resources that are continuously replenished [63]. Despite some drawbacks, using renewable energy sources has several clear benefits. These sources are sustainable and can be harnessed without depleting the Earth's resources.

Ocean energy, or marine energy, is one source of renewable energy that captures the kinetic or thermal energy of seawater to generate electricity or heat. The world's oceans are in constant motion, with currents, tides, and waves producing significant amounts of kinetic energy. This dynamic energy, resulting from the movement of water, can be captured and converted into electricity. One of the main advantages of wave energy is that the energy in waves is more concentrated than in other forms of renewable energy, such as wind or solar [21]. Since ocean waves are primarily generated by the gravitational pull of the moon and consistent wind patterns, wave energy is relatively predictable, enabling short-term forecasting of energy availability [57]. Despite having great potential, harnessing ocean energy still faces technical and infrastructure challenges [41], with installed capacity remaining low and roughly constant from 2015 to 2023 [45]. As wave energy is still underdeveloped, the cost of energy extraction remains high.

3.1 Ocean Waves

Ocean surface waves are generated primarily by wind acting on the sea surface, transferring energy to the water through shear stress and pressure fluctuations. Waves come in many sizes, from small ripples to large, crashing waves.

Understanding ocean waves is crucial because they have a significant impact on various aspects of coastal environments and marine life, as well as navigation. Waves shape coastlines, move sand and rocks, and influence the habitats of many sea creatures. They are also important for humans, as waves impact shipping routes, recreational activities like surfing, and even the generation of renewable energy through wave power.

In this section, we will explore ocean waves, the various kinds of waves that exist, and the reasons they are important. The following quantities, see also figure 3.1, characterize the properties of ocean waves and are used throughout this thesis:

Crest: The highest point of a wave.

Trough: The lowest point of a wave.

Height (H): The vertical distance between the crest and the trough. Significant wave height (H_s) is the average wave height of the top one-third of the highest waves.

Amplitude (A): The distance from the still water level to the wave crest or trough, essentially half the wave height.

Energy period (T_e): The time it takes for the average energy of the wave system to pass a fixed point.

Peak period (T_p): The time interval between the wave crests of the most energetic waves in a wave spectrum. It corresponds to the period at which the maximum energy occurs.

Mean zero-crossing period (T_z): The average time interval between successive crossings of the water surface level at zero (mean sea level) in the same direction (either upward or downward).

Frequency (f): The number of waves passing a fixed point in a certain amount of time (usually a second).

Velocity (v): The speed at which the waveform is moving, typically measured in meters per second.

Wavelength (L): The horizontal distance between two successive crests (or troughs). Indeed, it is the distance over which the wave's shape repeats. The relationship between Wavelength (L), period (T), and wave velocity (v) can be described by the formula:

$$v = \frac{L}{T},$$

Steepness: The ratio of significant wave height to wavelength, either defined with energy period, T_p , or mean zero-crossing period, T_z , as:

$$S_p = \frac{2\pi H_s}{gT_p^2}, \quad S_z = \frac{2\pi H_s}{gT_z^2}. \quad (3.1)$$

Phase Velocity (v_p): How fast individual wave crests move.

Group Velocity (v_g): How fast energy or information is transmitted through the medium.

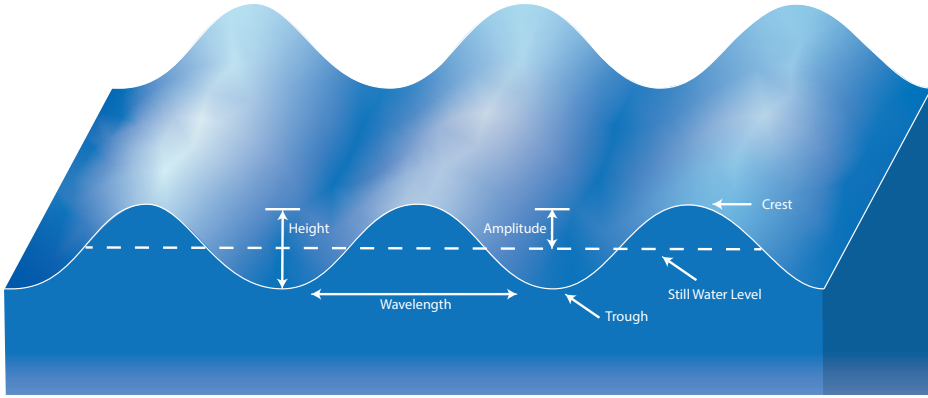


Figure 3.1: Illustration of key wave characteristics.

There are a variety of criteria for categorizing ocean waves. One of the most intuitive is based on the wave period or the associated wavelength [77] and consists of eight main wave types described in table 3.1.

Table 3.1: Classification of ocean waves by period, wavelength, restoring force, and primary physical behaviour [77].

Wave Type	Period	Wavelength	Restoring Force	Main Characteristics
Capillary waves	< 0.1 s	< 1.7 cm	Surface tension	First ripples generated by light winds; energy propagates faster than individual crests.
Gravity waves (wind sea)	1–25 s	1.5–900 m	Gravity	Generated by local winds; irregular sea state with short periods and large wave heights during storms.
Swell ^a	13–24 s	260–900 m	Gravity	Long, regular waves that propagate far from their generation area with little energy loss.
Infragravity waves	20 s–5 min	Hundreds of meters to kilometers	Gravity	Produced by nonlinear interactions of gravity waves; important for coastal processes and harbor oscillations.
Tsunamis	1–20 min	Few–hundreds of km	Gravity	Generated by earthquakes, landslides, or volcanic activity; small offshore but amplify near the coast.
Seiches	Seconds–hours	Basin-dependent	Gravity	Standing waves in enclosed or semi-enclosed basins caused by atmospheric disturbances.
Storm surges	Hours–days	Basin scale	Gravity	Sea-level rise caused by persistent low atmospheric pressure and strong winds; major source of coastal flooding.
Tides	12.25 or 24.5 h	Ocean basin scale	Lunar and solar gravity	Periodic sea-level oscillations driven by the gravitational attraction of the Moon and Sun.

^aswell and wind sea overlap in period but differ in their generation mechanism and spatial coherence.

3.2 Wave Energy

Interest in harnessing wave energy dates back at least to 1799, when the French inventor Girard & Son patented a machine designed to convert wave energy into mechanical power for mills and pumps. From 1855 to 1973, there were 340 patents related to the development of wave-powered devices [52].

Some European countries have extensively researched and developed wave energy, especially after the 1973 oil crisis. Several research programs with govern-

ment and private support began in different countries based on their exploitable wave power resources, most notably the United Kingdom, Portugal, Ireland, Norway, Sweden, and Denmark. The aim was to develop commercially viable wave power conversion technologies in the medium and long term [13].

Wind waves are more common than other types of waves. Waves created by the wind move across the ocean surface and continue long after the wind has died off. This makes wave energy the perfect partner for wind energy, as it extends power production significantly. When waves travel slower than the wind speed just above the water, energy is transferred from the wind to the waves. The size and period of the resulting waves depend on the amount of transferred energy, which is determined by the wind speed, the duration of the wind (typically over the course of days), and the distance over the ocean that the wind travels (the fetch) [44]. The amount of wave energy fluctuates over short and long periods, with the Northern Hemisphere experiencing greater seasonal variation [44]. The best wave resources are found in areas where strong winds have traveled over long distances. In Europe, this is along the western coasts, which are at the end of a long fetch (the Atlantic Ocean). As waves approach the coastline, their energy decreases due to friction with the seabed. Therefore, waves in deeper, well-exposed waters offshore have the greatest energy potential.

It is estimated that the total theoretical wave energy potential is 32,000 TWh/yr [56], which is substantially more than the currently installed capacity of about 500 MW.

3.2.1 Wave Power Plants

Capturing the energy of waves and converting it into electricity requires advanced technologies. WECs are among these technologies and can be categorized based on the basic concepts defining their operation (see section 3.2.2) and the method used to convert wave energy into electrical energy (see section 3.2.3). They are designed based on different environmental situations, such as water depth and location (shoreline, near shore, or offshore) [18].

The mechanism by which WECs convert absorbed energy into usable energy is known as the PTO system. In this section, we begin by studying different types of WEC and PTO systems. Subsequently, we delve into the specifics of the WEC utilized in this study.

3.2.2 Wave Energy Converters

WECs capture and convert the kinetic and potential energy of ocean waves into electrical energy through a series of steps involving wave capture, energy conversion, PTO, and electricity generation.

WECs can be categorized based on the underlying basic concepts that define the system's operation [6]:

Overtopping devices: Overtopping devices make use of the potential energy of ocean waves. They collect water from incident waves in an elevated reservoir. As the reservoir water is released back into the sea, it passes through a turbine, driving a generator to produce electricity.

Oscillating bodies: Oscillating bodies float in water, moving with the waves and capturing the wave's kinetic energy. Oscillating bodies are categorized into three main groups based on their degrees of freedom:

- Heaving devices use the up-and-down motion of waves.
- Pitching devices use the tilting motion of waves.
- Surging devices use the longitudinal motion of waves.

Oscillating water column (OWC): OWC systems work by harnessing the up-and-down motion of ocean waves. This motion drives the oscillation of water within a chamber. The motion of the water is then used to generate power through different PTO systems. OWCs are designed for installation in both onshore and offshore locations. Onshore installations require a minimum depth of 10 *m*, while offshore installations need to be placed at depths ranging from 40 *m* to 80 *m*. OWC is typically the favored system for use on western shores due to the more consistent and predictable wave patterns in these regions [20].

3.2.3 Power Take-off

The PTO system is crucial because it determines the efficiency of converting wave power into electricity, affecting the mass, size, and structural dynamics of WECs as well as the levelized cost of energy (LCoE) [60] of the entire system. There are different types of PTO systems, and choosing a PTO for a particular WEC depends on its type. The four broad types of conversion mechanisms employed in WECs [23] are:

High-pressure oil hydraulic PTOs: The buoy or body of the WEC oscillates due to wave forces, activating a hydraulic ram or piston that pumps hydraulic oil at high pressures. The oil flows through check valves, which ensure one-way fluid flow regardless of piston motion. The hydraulic fluid then collects in an accumulator tank before being used to drive a hydraulic motor coupled to an electric generator. The accumulator serves as an energy storage and dampener, absorbing pressure fluctuations and smoothing the fluid flow to the motor. Without an accumulator, the flow rate would be highly unsteady, leading to irregular electrical output. Controlling the flow into the motor is crucial for determining energy output and overall system efficiency across various sea states.

Air turbines: An air turbine driving a generator is common in OWC WECs. The airflow that drives the turbine is generated by the OWC inside the chamber. As an incoming wave raises the water level within the chamber, the volume of air above the water surface decreases, compressing the trapped air and forcing it through the turbine duct toward the atmosphere. Conversely, as the wave retreats and the water level falls, the air volume within the chamber increases and creates a pressure difference that draws atmospheric air back through the turbine into the chamber. Consequently, the oscillatory motion of the water column produces a bidirectional airflow through the turbine.

The bidirectional flow of air makes conventional turbines unsuitable for OWC WECs. This issue is addressed by using self-rectifying turbines. Wells turbines, Dennis-Auld turbines, and impulse turbines have been used in many applications, with Wells turbines being the most common among air turbines in OWCs.

Low-head water turbines: Hydraulic turbines coupled with generators are common in overtopping WECs, drawing from their long history in hydroelectric power generation. These turbines are categorized into impulse turbines and reaction turbines. Impulse turbines convert the kinetic energy of water into mechanical energy; the water pressure is converted to velocity as it passes through nozzles, and the high-speed water then hits bucket-shaped blades, causing the turbine to rotate. Reaction Turbines use the pressure changes as water flows through the rotor blades. The pressure difference across the blades drives the rotation.

Linear generators: This type of PTO system is a linear permanent magnet generator. This generator consists of a stator and a translator, converting the motion of the WEC directly into electricity. These generators are

directly coupled to the WEC’s float, eliminating the need for turbines and/or hydraulic systems.

3.2.4 WEC in this study

The WEC examined in this study is WaveEL 4.0, a prototype owned, designed, and developed by the company Waves4Power. Figure 3.2 illustrates this WEC with its dimensions. WaveEL 4.0 incorporates an OWC with a hydraulic PTO system, where the hydraulic pressure is generated by a piston moving within a water-filled tube.

A previous version, WaveEL 3.0, was installed near Ålesund, Norway, in 2016 and operated until November 2017, generating energy for the Norwegian grid during this period. The primary distinction between WaveEL 3.0 and 4.0 is the length of the water tube, which was increased from 18.5 m to 34 m, enhancing its energy-capturing capabilities but requiring deeper water to operate.

3.3 Wave Modelling

Wind-induced waves are among the most complex phenomena in nature, arising from nonlinear interactions between wind, water, and bathymetry that make accurate numerical modeling challenging. Wave modeling refers to numerical methods used to represent sea conditions and forecast wind-generated wave behavior [31]. Numerical wave models can be categorized based on their complexity, applicability to offshore or nearshore waters, use of parametric or spectral wave data, or other criteria [31].

Regular (monochromatic) waves are idealized, periodic unidirectional wave trains with constant amplitude, wavelength, and period, described by a single harmonic [62].

3.3.1 Linear waves

Linear wave theory¹ assumes that waves have small amplitudes compared to their wavelength and the depth of the water. Under this assumption, the free surface boundary conditions can be linearized. The theory ignores more complex, nonlinear effects, such as waves becoming steep or breaking. Figure 3.3(a)

¹Also referred to as Airy wave theory.

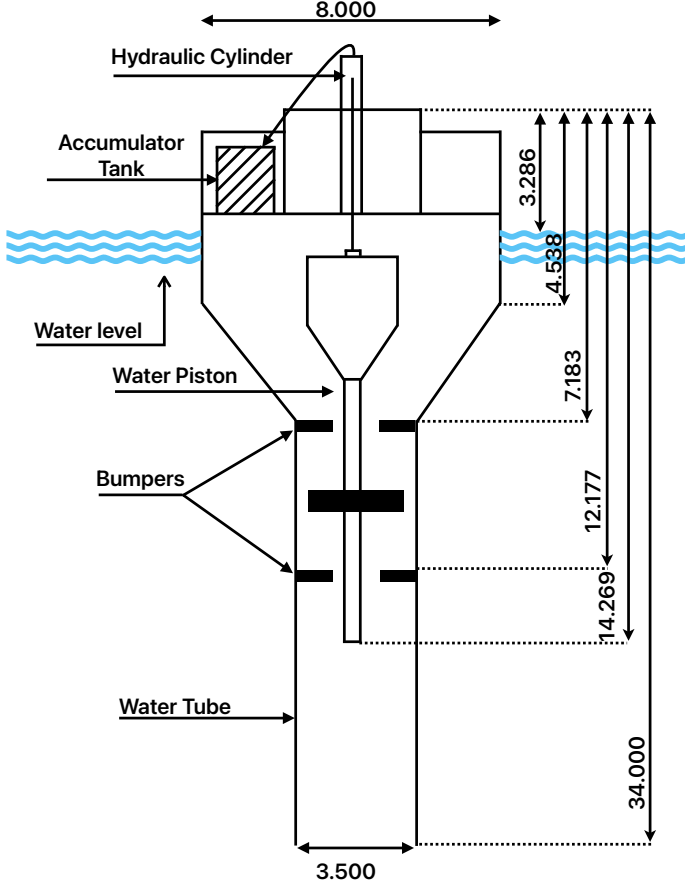


Figure 3.2: Diagram of the WaveEL 4.0 buoy used in this study, with key dimensions (in meters) and components.

shows a scheme of linear waves. The basic linear wave equation for a sinusoidal wave traveling in the x -direction is [74].

$$\zeta(x, t) = A \cos(kx - \omega t)$$

Here $\zeta(x, t)$ is the wave elevation at position x and time t . A is the wave amplitude (half of the wave height, H); $k = 2\pi/L$ is the wave number, which is defined as the number of wavelengths per unit distance; L is the wavelength; and ω is the angular frequency (radians per second). The relationship between angular frequency and wave period, T , is $\omega = 2\pi/T = 2\pi f$.

The dispersion relation in wave theory states that the wave frequency ω is related

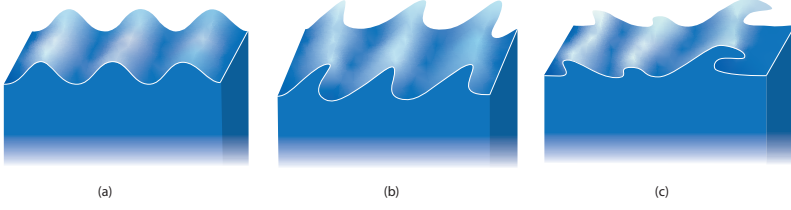


Figure 3.3: Schematic description of different types of waves: a) linear, b) non-linear, and c) random

to wave number k as [74],

$$\omega^2 = gk \tanh(kd),$$

where d is the water depth and g is the acceleration of gravity. For deep and shallow water, this can be approximated as [74]:

$$\begin{cases} \tanh(kd) \approx 1 & \text{for deep water } d \gg L \\ \tanh(kd) \approx kd & \text{for shallow water } d \ll L \end{cases}$$

3.3.2 Spectral Representation of Irregular Wave

An irregular sea wave, often referred to as a random wave, is a type of ocean wave that does not follow a regular, periodic pattern, as seen in figure 3.3(c). Unlike regular waves, which have consistent wavelength, amplitude, and period, irregular waves are characterized by varying heights, periods, and directions. This randomness results from the complex and dynamic interactions between many factors such as wind, currents, tides, and topography [74].

The stochastic process framework introduced in section 2.1 provides the mathematical foundation for modeling irregular waves. Specifically, the sea surface elevation $\zeta(t)$ at a fixed point is treated as a realization of a zero-mean stationary Gaussian process. Its statistical structure is fully characterized by its autocovariance function $r(\tau)$, or equivalently, by its PSD $S(\omega)$ via the Wiener–Khinchin relation equation (2.2). The stochastic simulation formula equation (2.5) then shows how individual wave realizations can be generated from a prescribed $S(\omega)$: each realization draws independent Rayleigh-distributed amplitudes A_k and uniform random phases θ_k , producing a different but statistically equivalent wave-elevation time series. Then the sampling is from:

$$Z(t) = \sum_{k=0}^{N-1} \underbrace{\sqrt{S(\omega_k) \Delta\omega} R_k}_{=A_k} \cos(\omega_k t + \theta_k),$$

The key question that remains is: what form should $S(\omega)$ take for a given sea state? Through decades of field measurements, empirical spectral models have been established that relate $S(\omega)$ to observable sea-state parameters, primarily the significant wave height H_s and the peak period T_p . The three main families used in wave modelling are the PM (or Bretschneider), JONSWAP, and Ochi-Hubble spectra [46], described below. Once a spectrum is selected and its parameters fixed, the simulation method of section 2.1.2 produces time series whose statistical properties match the target sea state.

Before defining the specific spectral families, it is useful to introduce spectral moments, defined using frequency and not angular frequency, as

$$m_n = \int_0^\infty f^n S(f) df.$$

These moments provide a compact link between the spectrum and the physical wave parameters used to parameterize it. In particular, H_s and T_p , the two parameters that appear in the PM and JONSWAP formulas, can both be expressed in terms of spectral moments as shown below. Some common spectral moments and their interpretation include:

- m_0 : The total energy of the wave field and is related to significant wave height as $H_s = 4\sqrt{m_0}$.
- m_1 : The ratio m_1/m_0 gives the mean frequency of the wave spectrum.
- m_2 : Related to the spectral width or spread of wave frequencies.
- m_{-1} : The ratio m_0/m_{-1} is related to the energy period T_e of the dominant wave system.

Pierson–Moskowitz Spectrum

The PM spectrum [64] is based on observations from the North Atlantic and is among the most popular spectra for wave modeling. It is based on the idea that, given enough space and time, waves and wind will reach an equilibrium, resulting in a fully developed sea. In this context, *enough time* is about ten thousand wave periods, and *enough space* is about ten thousand wavelengths on each side. Thus, the PM spectrum is appropriate for fully developed seas. Often the generalized Pierson–Moskowitz (GPM), or Bretschneider, spectrum is used, with spectral density given by [46] as:

$$S(\omega) = m_0 \frac{5}{\omega_p} \left(\frac{\omega_p}{\omega} \right)^5 \exp \left(-\frac{5}{4} \left(\frac{\omega}{\omega_p} \right)^{-4} \right). \quad (3.2)$$

Here m_0 is the wave energy or variance, and the spectral peak, ω_p , is the peak angular frequency, with corresponding peak period $T_p = 2\pi/\omega_p$. The relation between peak and energy period is given in Paper I as $T_e = 0.86 T_p$, establishing a link between m_{-1} and T_p through T_e .

Figure 3.4 shows an example of a PM spectrum with $H_s = 6.25$ and $T_e = 10.5$ and waves simulated based on that spectrum.

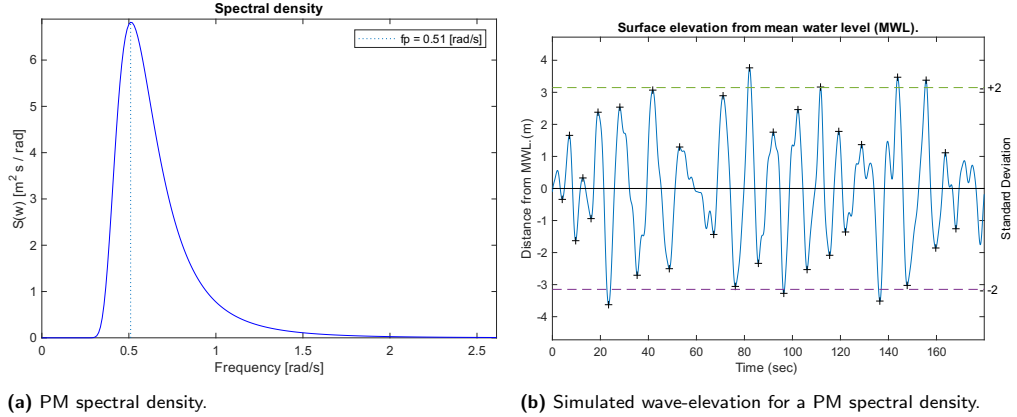


Figure 3.4: Illustration of a PM spectral density and corresponding wave simulation for $H_s = 6.25$ and $T_e = 10.5$ ($T_p \approx 12.2$ s using $T_e = 0.86 T_p$).

The GPM spectrum in equation (3.2) is a widely used model for the spectrum of waves measured at a point (without regard to wave direction), and Paper I uses a truncated version of this spectrum. Slight changes in the definition give other spectra, including the JONSWAP.

JONSWAP spectrum

In 1973, Hasselmann et al. [33] analysed data from the JONSWAP and discovered that the wave spectrum never reaches full development. It continues to evolve through non-linear wave-wave interactions over long periods and distances. Therefore, an additional factor was incorporated into the PM spectrum to better align with their measurements. The JONSWAP spectrum [46] using the same parameters as in equation (3.2) is

$$S(\omega) = m_0 \frac{5}{\omega_p} \left(\frac{\omega_p}{\omega} \right)^5 \exp \left(-\frac{5}{4} \left(\frac{\omega}{\omega_p} \right)^{-4} \right) \gamma^{\exp \left[-\frac{(\omega - \omega_p)^2}{2\sigma^2 \omega_p^2} \right]}.$$

In addition to m_0 and ω_p from equation (3.2) we now have a peak enhancement, γ , and a peak width σ commonly taken as:

$$\sigma = \begin{cases} 0.07 & \text{if } \omega \leq \omega_p \\ 0.09 & \text{if } \omega > \omega_p \end{cases}$$

In fact, the JONSWAP spectrum is a PM spectrum multiplied by an extra peak enhancement factor, $\gamma^{\exp\left[-\frac{(\omega-\omega_p)^2}{2\sigma^2\omega_p^2}\right]}$. A value of $\gamma = 1$ recovers the PM spectrum. Values between 1 and 3.3 represent intermediate stages of wave development, with higher values indicating a sharper, more pronounced spectral peak. It is important to note that the JONSWAP spectrum is assumed to be particularly suitable for the North Sea [7], which is too small to allow waves to develop fully.

The significant wave height for the JONSWAP spectrum is given in [46] as

$$H_s = (1.555 + 0.2596\gamma - 0.02231\gamma^2 + 0.001142\gamma^3) \left[\frac{g\sqrt{\alpha}}{(2\pi\omega_p)^2} \right],$$

where α is the JONSWAP energy scale parameter (also known as the Phillips constant), which controls the overall energy level of the spectrum and is typically determined empirically from the fetch and wind speed.

Ochi-Hubble spectrum

The Ochi-Hubble spectrum is characterized by summing two spectral components, which represent different sea states or wave systems, such as wind sea and remote swell [46], allowing for more accurate modeling of the combined wave energy that accounts for both local and distant wave systems [61]. The double peak spectral model is given as a sum of two spectra, each closely related to the GPM, with six parameters [46]:

$$S(\omega) = \frac{1}{4} \sum_{j=1,2} \left(\frac{4\lambda_j + 1}{4\omega_{pj}^4} \right)^{\lambda_j} \frac{H_{sj}}{\omega^{4\lambda_j+1}} \exp \left[-\frac{4\lambda_j + 1}{4} \left(\frac{\omega_{pj}}{\omega} \right)^4 \right]$$

3.4 Analysis of the WEC

In this section, we review the tools used in papers I and II to analyse waves and the WEC.

The omnidirectional spectra described above characterize wave energy as a function of frequency only, implicitly assuming that all energy propagates in a single direction, a long-crested sea. Real ocean waves, however, are short-crested: energy arrives simultaneously from a spread of directions, producing a more complex and realistic sea surface [35]. A short-crested sea is described by a directional spectrum $S(\omega, \varphi) = S(\omega) D(\varphi)$, where $D(\varphi)$ is a directional spreading function satisfying $\int_{-\pi}^{\pi} D(\varphi) d\varphi = 1$ [46, 35]. In Papers I and II, the long-crested formulation is adopted, consistent with the one-dimensional heave dynamics of the WaveEL 4.0 WEC and the available sea-state data from Billia Croo. Extending the framework to short-crested seas would require directional sea-state data and a multi-directional WEC response model, and represents a natural avenue for future work.

3.4.1 Wave Analysis for Fatigue and Oceanography (WAFO)

WAFO [7] is a MATLAB toolbox for statistical modeling, simulation, and analysis of random waves and their distributional properties. Given the parametric form of a wave spectrum, WAFO can be used to simulate waves. Several different spectral densities are included in WAFO, ranging from those outlined in section 3.3.2 to models for bimodal spectra and finite water depth [7]. Simulation is performed using the methods outlined above in section 2.1.2 of chapter 2.

3.4.2 Simulink

Simulink is a powerful tool developed by MathWorks for multidomain simulation and model-based design [75]. A Simulink model, developed by Waves4Power, of the Waves4Power buoy is used to investigate buoy behavior under varying wave conditions. A detailed description of the model is provided in the appendix of Paper I. The Simulink model of the buoy incorporates various physical characteristics and environmental conditions, such as buoy shape, material, mass, volume, water density, and wave characteristics. The simulated waves are used as input to the model to investigate its behavior and power generation under different conditions. The main aim of Papers I and II is to optimize the parameters of the Simulink model.

3.5 Paper I: Optimizing the hydraulic power take-off system in a wave energy converter (Published)

This paper presents a simulation-driven framework for optimizing the operating pressure of the hydraulic accumulator tank in a WEC equipped with a hydraulic PTO system. The proposed methodology integrates simulation of irregular waves, dynamic system modeling, and statistical learning techniques to investigate the relationship between sea-state characteristics, accumulator pressure, and generated power.

The primary scientific contribution of this work is the identification of wave steepness as the dominant sea-state parameter governing the optimal accumulator tank pressure. While sea states are conventionally characterized by both significant wave height and energy period, the results demonstrate that wave steepness alone provides sufficient information for predicting the pressure that maximizes power generation. This finding simplifies the optimization and design of hydraulic PTO systems by reducing the dimensionality of the operating space from two environmental variables to a single physically meaningful parameter.

A second contribution is the development of a statistical optimization framework that combines dynamic simulations with concave regression to determine the pressure corresponding to maximum power output for a wide range of sea states. Building on these optimal operating points, a regression model is developed to predict the optimal accumulator pressure directly from wave steepness, eliminating the need for computationally intensive optimization during operation.

Furthermore, the study demonstrates the potential of adaptive pressure control for improving energy capture. By continuously adjusting the accumulator pressure according to prevailing wave conditions, the proposed approach provides a practical basis for enhancing the efficiency and performance of hydraulic PTO systems. The results therefore contribute not only to the understanding of the interaction between sea-state characteristics and hydraulic energy conversion, but also to the development of intelligent control strategies for future WECs.

3.6 Paper II: The effects of sea-state on optimal generator parameters for a wave energy converter (Published)

The performance of WECs is highly dependent on the design of the PTO system and the control strategy employed to adapt to changing sea conditions. This paper presents a simulation-based framework for identifying the optimal generator operating parameters for a hydraulic PTO system under a wide range of sea states. Using irregular waves generated from PM spectra and a detailed nonlinear Simulink model of the WEC, the study determines which generator configurations maximize electrical power production.

The main scientific contribution of this study is an extension to Paper I, in which we show that wave steepness is the key sea-state parameter determining the optimal operating point of the generator. While sea states are typically characterized by significant wave height and wave period, the findings indicate that wave steepness by itself encapsulates the critical features needed for generator optimization. This insight allows sea states to be clustered into representative operating regions and greatly streamlines the development of control laws for the PTO system.

A second key contribution is the formulation of a structured optimization framework that integrates high-fidelity dynamic simulations with statistical evaluation to identify optimal generator settings over a wide spectrum of operating conditions. Using these findings, the study introduces a control scheme that modifies the generator speed in response to the current wave steepness, offering a practical solution for adaptive PTO control.

Finally, the paper shows that adjusting control strategies to the prevailing sea state can substantially increase electrical power output relative to using fixed operating settings. Consequently, the results deliver both physical understanding of how sea-state properties affect hydraulic PTO behavior and practical recommendations for designing adaptive control schemes. Moreover, these findings lay the groundwork for future optimization of wave farms and fully integrated wave-to-wire energy conversion systems.

Chapter 4

Remote Sensing

In this chapter, the definitions and remote sensing concepts used in Papers III and IV for detection of bark beetle attacks and in Paper V for restoring corrupted remote sensing images are presented.

4.1 Fundamentals of Remote Sensing

Remote sensing is the acquisition of information about the Earth's surface without direct physical contact, through the measurement of electromagnetic radiation reflected or emitted by targets [71]. The fundamental principle in remote sensing is the interaction between incoming radiation and surface materials, which determines how energy is absorbed, transmitted, or reflected at different wavelengths. As illustrated in figure 4.1, the electromagnetic spectrum spans a wide range of wavelengths, from gamma rays to radio waves. Optical remote sensing focuses primarily on the visible spectrum (VIS), red edge (RE), near-infrared (NIR), and SWIR regions [71].

Different surface types exhibit distinct spectral signatures across these regions, enabling their identification and characterization. For instance, vegetation, soil, and water each reflect and absorb radiation differently depending on wavelength. Optical satellite sensors exploit these differences to identify land cover and derive biophysical and biochemical properties of the land surface [71].

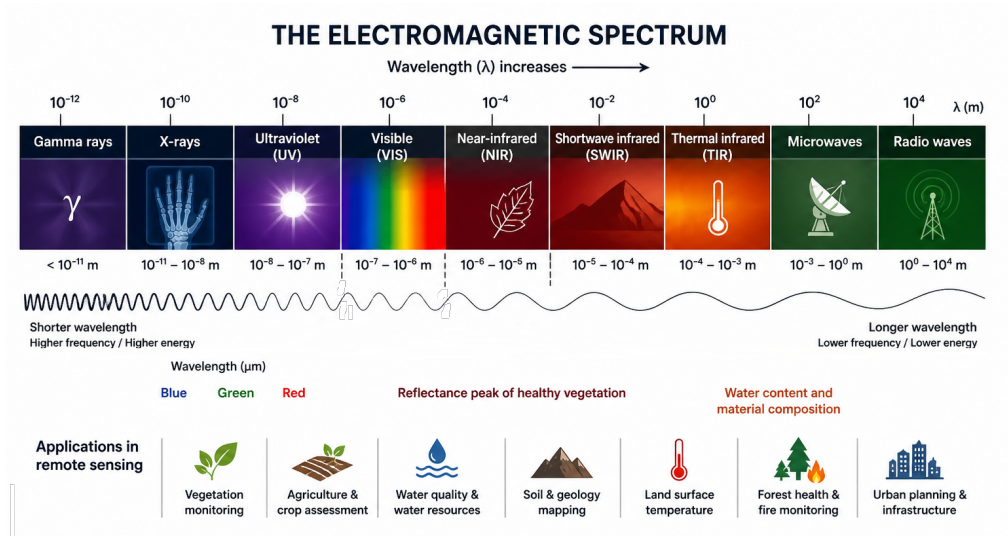


Figure 4.1: Electromagnetic spectrum illustrating the range of wavelengths used in remote sensing, highlighting the visible, NIR, and SWIR regions commonly utilized in optical satellite imagery.

4.1.1 Passive vs Active Remote Sensing

Remote sensing systems are classified based on the source of radiation:

Passive Remote Sensing: Relies on natural energy sources, mainly the Sun. The sensor records radiation reflected (or emitted) by the Earth's surface. An example is optical imagery from satellites and drones. Passive sensors are dependent on illumination conditions and affected by cloud cover and atmospheric interference, but provide rich spectral information for vegetation analysis.

Active Remote Sensing: The sensor emits its own energy and measures the returned signal after interacting with the target (e.g., synthetic aperture radar (SAR)). Such systems are independent of solar illumination, capable of penetrating cloud cover, and primarily sensitive to surface structure and dielectric properties rather than spectral reflectance.

This thesis employs passive optical remote sensing, as vegetation health and stress are strongly expressed in the spectral reflectance signatures of forest canopies.

4.1.2 Components of a Remote Sensing System

A remote sensing system can be described through four main components that together form the observation chain from energy source to recorded measurement:

Radiation (Electromagnetic Energy) : The energy used to illuminate the target [71].

Target (Object of Interest) : The surface being observed, such as vegetation, soil, or water.

Sensor : A device that records the reflected or emitted radiation from the target.

Platform : The carrier of the sensor. Platforms include satellite-based systems (e.g., the Copernicus Sentinel-2 mission or the Landsat programme), airborne systems such as manned aircraft, and unmanned aerial vehicles (UAVs).

In this work, data from both satellite (Sentinel-2) and drone-based platforms are used. Satellites provide large-scale, repeated coverage, while drones provide very high spatial resolution and flexible data acquisition at local scales [71].

4.1.3 Spectral Indices for Vegetation Monitoring

Spectral indices combine reflectance from multiple bands to enhance specific vegetation properties while reducing external noise such as illumination effects or soil background.

The indices used in this study capture complementary aspects of vegetation condition:

Normalized difference vegetation index (NDVI) [70]:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$$

Measures overall vegetation health and biomass, decreasing with chlorophyll loss and structural degradation. It leverages the unique spectral signature of healthy plants, where chlorophyll absorbs red light, while cell structures reflect NIR light. An issue with NDVI is that the signal saturates for dense vegetation.

Normalized difference red-edge index (NDRE) [4]:

$$\text{NDRE} = \frac{\text{NIR} - \text{RE}}{\text{NIR} + \text{RE}}$$

Unlike NDVI, which uses visible red light, NDRE utilizes the RE, the transition zone between visible red and NIR light. As RE light penetrates leaves and deep vegetation canopies much more effectively than red light, NDRE is highly sensitive to the nitrogen and chlorophyll status of vegetation.

Normalized difference water index (NDWI) [26]:

$$\text{NDWI} = \frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$$

Here SWIR1 refers to the band centred at approximately $1610nm$ (Sentinel-2, Band 11), which is sensitive to vegetation water content. NDWI is effective for identifying moisture stress. Higher values indicate higher canopy water content and healthier vegetation, whereas lower values indicate water stress.

Normalized burn ratio (NBR) [34]:

$$\text{NBR} = \frac{\text{NIR} - \text{SWIR2}}{\text{NIR} + \text{SWIR2}}$$

Here SWIR2 refers to the band centred at approximately $2190nm$ (Sentinel-2, Band 12). NBR is sensitive to both moisture and structural changes, making it useful for detecting infestation effects. It works similarly to NDVI but emphasizes the NIR and SWIR bands, where healthy vegetation shows high reflectance.

Distance red and shortwave infrared (DRS) [38]:

$$\text{DRS} = \sqrt{\text{Red}^2 + \text{SWIR}^2}$$

Combines red and SWIR information to capture chlorophyll degradation and moisture stress. When a spruce tree undergoes early attack by bark beetles, its chlorophyll absorption weakens (causing red reflectance to rise) and its canopy moisture drops (causing SWIR reflectance to spike). Because both values increase simultaneously, the Euclidean distance (DRS) grows significantly.

Normalized distance red and shortwave infrared (NDRS) [38]:

$$\text{NDRS} = \frac{\text{DRS} - \text{DRS}_{\min}}{\text{DRS}_{\max} - \text{DRS}_{\min}}$$

Normalizing DRS by the minimum, $\text{DRS}_{\min}$, and maximum, $\text{DRS}_{\max}$, values observed across the dataset (or time series) maps these stress values onto a standardized scale. This enables easy spotting of anomalous, infested pixels before the canopy turns brown; hence, NDRS provides a normalized and more comparable measure across time and space.

Together, these indices provide a compact representation of vegetation condition, capturing pigment dynamics (red, green, and blue (RGB) and RE bands), structural integrity (NIR), and moisture status (SWIR). When combined with raw spectral bands in a temporal framework, they enable the detection of both early and advanced stages of bark beetle infestation.

4.2 Sensor Platforms

4.2.1 Sentinel-2

In recent years, advances in aerial and satellite technology have made satellite imagery a primary tool for environmental monitoring. Sentinel-2, see figure 4.2, is among the most widely used satellite platforms for environmental monitoring, offering free and open access as part of the European Space Agency (ESA)'s Copernicus programme, combined with high spatial resolution, frequent revisit times, and vegetation-sensitive spectral bands.

As an optical imaging system, Sentinel-2 captures reflected solar radiation across 13 spectral bands, enabling detailed observation of the Earth's surface. Optical imagery records information in the visible, NIR, and SWIR regions of the electromagnetic spectrum. This spectral richness allows for the precise characterization of vegetation, soil, and water bodies based on their reflectance properties. One disadvantage of optical imagery is that it is significantly affected by atmospheric conditions such as clouds, haze, and shadows, which can introduce missing or noisy data. Despite these challenges, preprocessing steps such as atmospheric correction and cloud masking help improve data quality [22].

Sentinel-2 consists of two twin satellites flying in the same polar orbit but 180° apart. This allows for a revisit time of 5 days at the equator and more frequently at higher latitudes. This frequency is critical because forest disturbances like

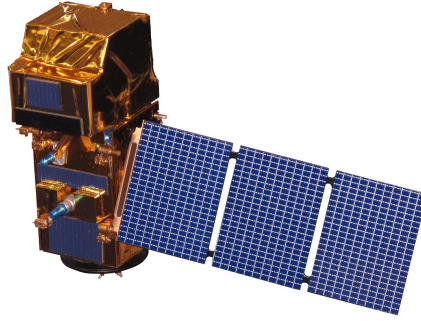


Figure 4.2: Model of a Sentinel-2 satellite.

bark beetle attacks can progress rapidly; missing a few weeks of data could mean missing the early attack phase entirely.

The onboard imaging sensor, Multispectral Instrument (MSI), captures 13 spectral bands [22] at different spatial resolutions:

10-meter resolution : Includes the visible bands (VIS) and one NIR band.

20-meter resolution : Includes three RE bands, two SWIR bands (SWIR1, SWIR2) and also one NIR band, which are essential for detecting vegetation stress and moisture levels.

60-meter resolution : Primarily used for atmospheric corrections, such as detecting clouds and water vapor.

The most relevant bands for vegetation monitoring are:

VIS: Sensitive to pigment content, especially chlorophyll absorption.

NIR : Sensitive to vegetation structure and biomass.

SWIR: Sensitive to water content and moisture stress.

RE: Particularly important for detecting early vegetation stress through subtle shifts in chlorophyll content.

The combination of high spatial resolution, short revisit time, and vegetation-sensitive spectral bands has made Sentinel-2 a standard for monitoring forest health. The short revisit times generate a dense time series for each pixel, which is crucial for distinguishing gradual stress signals from normal seasonal variation and identifying early-stage bark beetle attacks.



Figure 4.3: From left to right: The MAIA camera, the MAIA ILS, the DJI Matrice 210 RTK UAV, and the radiometric reference targets. Images from [40]

4.2.2 Unmanned Aerial Vehicles (UAVs)

UAVs, commonly referred to as drones, have become an important platform for remote sensing due to their high spatial resolution, operational flexibility, and relatively low operating cost. Compared to satellites and manned airborne platforms, UAVs can acquire imagery on demand and provide centimeter-level spatial resolution, making them particularly suitable for local monitoring of vegetation health, forest structure, and environmental changes. Depending on the application, UAVs can be equipped with a variety of sensors, including RGB, multispectral, hyperspectral, thermal, and LiDAR sensors.

Paper V analyzes UAV-based multispectral and hyperspectral imagery collected by two different UAVs over Remningstorp Experimental Forest in southern Sweden, a managed forest dominated by Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*).

MAIA S2 Multispectral sensor

The multispectral imagery was acquired using a MAIA S2 sensor mounted on a DJI Matrice 210 RTK UAV, shown in figure 4.3. The sensor captures nine spectral bands spanning VNIR 443 – 865 nm . Flights were conducted at an altitude of 120 m , producing a ground sampling distance (GSD) of approximately 6 cm . The acquired images were processed to generate orthomosaics through geometric and radiometric corrections and converted to surface reflectance using empirical line calibration with ground reference panels. Further details of the multispectral dataset are provided in [39].



Figure 4.4: From left to right: The Freefly Alta X drone with the SPECIM AFX17 hyperspectral sensor used for data acquisition, and ground calibration targets used for post-processing reflectance estimation [14].

SPECIM AFX17 Hyperspectral UAV

The hyperspectral imagery was acquired using a SPECIM AFX17 sensor mounted on a Freefly Alta X UAV, shown in figure 4.4. The sensor operates in the SWIR region ($1000 - 1700\text{ nm}$) and records 112 spectral bands. Flights were performed at an altitude of 100 m , resulting in a GSD of approximately 10.8 cm . Data were collected using a pushbroom imaging system with multiple parallel flight lines [14]. Radiometric correction, georeferencing, and conversion to reflectance were carried out using sensor calibration coefficients and ground reference panels. Additional details regarding data acquisition and preprocessing can be found in [14].

4.3 Bark Beetles and Their Spectral Response

Bark beetles are insects that live and reproduce beneath the bark of trees and shrubs [78]. While many species contribute to the natural decomposition of weakened or dead trees, large-scale outbreaks driven by drought, warming temperatures, and other climate-related stressors can cause extensive forest mortality [72]. Bark beetle infestations disrupt the physiological processes of trees, producing characteristic changes in their spectral signatures that can be detected using remote sensing. Bark beetles infest trees by boring through the bark and weakening the water transport, causing progressive physiological stress, in-

cluding reduced water content, chlorophyll degradation, cellular damage, and eventual needle discoloration [78, 79, 30, 1].

To understand how these physiological changes can be detected remotely, it is first necessary to consider the spectral properties of healthy vegetation. When electromagnetic radiation reaches a plant canopy, it is reflected, absorbed, or transmitted depending on leaf pigments, internal structure, and water content, producing a characteristic spectral signature [80, 28, 32, 3]. Healthy vegetation exhibits strong chlorophyll absorption in the blue and red portions of the VIS spectrum, relatively high reflectance in the green region, high reflectance in the NIR region due to multiple scattering within the spongy mesophyll (i.e. cells inside the leaves), and pronounced water-absorption features in the SWIR region. Consequently, VIS reflectance primarily reflects pigment concentration, NIR reflectance is sensitive to leaf and canopy structure, and SWIR reflectance provides information about vegetation water content. Sentinel-2 is particularly well suited for monitoring these properties through dedicated bands, including a NIR band and two SWIR bands.

The physiological changes induced by bark beetle infestation are reflected in the spectral response of the canopy and evolve throughout the infestation process. During the initial *green attack* stage, the crown remains visually healthy, but reduced water transport caused by beetle activity and fungal colonization leads to moisture stress [1]. Since water strongly absorbs SWIR radiation, declining water content increases SWIR reflectance, making this spectral region particularly effective for early detection. Spectral indices incorporating SWIR bands, such as the NDWI and the NBR, are therefore widely used to identify early infestations [48, 40]. The DRS index, which combines red and SWIR reflectance, was specifically designed to capture the simultaneous increase in both bands during the green attack stage [38].

As moisture stress intensifies, chlorophyll content gradually decreases, producing subtle changes in the RE region (approximately $680 - 750\text{ nm}$). The characteristic sharp increase in reflectance between the red and NIR wavelengths, known as the red-edge, becomes less pronounced and shifts toward shorter wavelengths [36]. Sensors equipped with dedicated RE bands, such as Sentinel-2, are therefore particularly sensitive to these early physiological changes and enable detection before visible crown discoloration occurs [2].

As the infestation progresses to the *red attack* stage, structural degradation of the needles becomes dominant. The collapse of the internal leaf structure reduces NIR reflectance, while continued chlorophyll degradation increases reflectance in the red portion of the spectrum [80]. Consequently, vegetation

indices such as the NDVI decline substantially, making the infestation readily detectable by optical remote sensing [65, 40]. Papers III and IV address the issue of detecting bark beetle attack based on Sentinel-2 data.

4.4 Multispectral Image Fusion

The spectral responses discussed above highlight the importance of the SWIR region for detecting the early physiological effects of bark beetle infestation. In particular, changes in vegetation water content can become apparent in SWIR reflectance before visible discoloration occurs. However, SWIR imagery is often affected by noise and sensor-related artifacts. Multispectral image fusion provides a potential approach to address this limitation by exploiting complementary information from VIS and NIR imagery to reconstruct corrupted SWIR observations.

Multispectral image fusion in remote sensing combines information from different wavelength regions to improve image quality and interpretation. In particular, VNIR bands can provide detailed spatial information on edges, textures, and vegetation structure, whereas the SWIR bands contain valuable spectral information related to vegetation water content, material properties, and canopy biochemical characteristics. These complementary properties make VNIR imagery a useful source of guidance for SWIR reconstruction.

The main challenge is to exploit this complementary information without compromising either spatial detail or spectral fidelity. Effective fusion therefore requires models that can selectively transfer relevant structural information from the VNIR imagery, while preserving the spectral characteristics of the SWIR observations [82]. This task becomes particularly challenging when the spectral bands are acquired by different sensors, as differences in sensor characteristics can introduce geometric and radiometric inconsistencies.

Accurate sensor fusion is particularly important for forest monitoring, where small spectral changes in the SWIR bands that are associated with vegetation moisture stress may contain important early-warning information about bark beetle infestations [40]. Paper V proposes a sensor fusion method for VNIR and SWIR imagery obtained by two different UAVs.

4.5 Paper III: Detection of Bark Beetle Attacks Using Time-Aggregated Satellite Data With Machine Learning (Published)

This paper presents a machine learning framework for the early detection of bark beetle attacks by integrating temporally aggregated Sentinel-2 satellite imagery with static environmental and forestry data. The proposed approach combines multi-temporal remote sensing observations with site-specific susceptibility information using an XGBoost classifier, enabling both accurate prediction and interpretation of the factors influencing bark beetle infestations.

The primary scientific contribution of this work is the demonstration that combining static environmental variables with cumulative Sentinel-2 observations significantly improves the early detection of bark beetle attacks compared with using either data source independently. While static features capture long-term site susceptibility, satellite observations provide complementary information on the temporal evolution of vegetation stress, resulting in a more robust and accurate predictive model, particularly during the early stages of infestation when spectral symptoms remain subtle.

A second contribution is the systematic temporal analysis of Sentinel-2 observations to identify the most informative periods for bark beetle detection. By training models using cumulative weekly observations and analyzing feature importance over time, the study reveals that only a limited number of weeks contribute most of the predictive information. In particular, weeks corresponding to peak swarming activity, the period preceding completion of the second generation, and later stages of canopy stress were found to be especially informative. These findings provide valuable insight into the temporal dynamics of bark beetle infestations and can support the design of more efficient monitoring strategies.

Furthermore, the study provides an interpretable analysis of feature importance across both satellite and static data sources. Among the Sentinel-2 variables, the RE and Blue bands were identified as the most influential when static data were incorporated, while land-cover and landform variables emerged as the most important static predictors. The analysis also demonstrates that incorporating static information changes the relative importance of several spectral bands, highlighting the complementary nature of environmental and remote sensing data.

4.6 Paper IV: Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series (Submitted)

This paper presents a semi-supervised deep learning framework for the early detection of European spruce bark beetle outbreaks using Sentinel-2 time-series data and static environmental information. The proposed approach employs a CVAE that leverages environmental risk estimates as proxy supervision, enabling outbreak detection without requiring extensive ground-truth annotations of attacked trees. By integrating temporal satellite observations with environmental susceptibility information, the framework learns discriminative latent representations that distinguish healthy from attacked forest areas.

The primary scientific contribution of this work is the development of a semi-supervised CVAE for bark beetle detection in data-scarce environments. Unlike conventional supervised machine learning approaches that rely on large labeled datasets, the proposed method conditions the latent representation on risk levels derived from static environmental features and a limited set of labeled data. This allows the model to learn the temporal dynamics of forest health while effectively separating healthy and attacked pixels in the latent space without requiring direct temporally aligned same-year ground-truth annotations of attacks.

A second contribution is integrating multi-temporal Sentinel-2 observations with static environmental features within a unified probabilistic framework. Multi-spectral Sentinel-2 observations characterize the temporal evolution of forests, and static environmental features provide complementary information on the susceptibility of forest areas to infestation. This joint framework allows the model to capture where attacks are more likely to happen and how forest spectral characteristics respond to infestation.

Furthermore, the study demonstrates that meaningful outbreak detection can be achieved with limited labeled data while achieving competitive classification performance relative to fully supervised baselines. The experimental results show that the learned model’s latent representations, which were trained only on the first year, effectively discriminate between healthy and attacked forest areas across two consecutive years, highlighting the robustness and generalization capability of the proposed approach. The consistently high recall further indicates that the model is particularly suitable for early-warning applications, where detecting as many potential infestations as possible is often more important than minimizing false positives.

The work also provides insight into the temporal behavior of the proposed model by analyzing its performance as additional Sentinel-2 observations become available throughout the growing season. The results show that prediction accuracy improves as vegetation stress develops over time, reflecting the progressive nature of bark beetle infestations and demonstrating the value of continuous time-series observations for forest health monitoring.

4.7 Paper V: Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction

This paper presents an asymmetric dual-path U-Net for restoring corrupted SWIR imagery by exploiting complementary information from clean VNIR images. The proposed framework is motivated by the observation that VNIR imagery generally provides reliable spatial structure and texture, whereas SWIR imagery contains valuable spectral information but is more susceptible to sensor artifacts, missing regions, and degradations. By explicitly modeling the interaction between these complementary modalities, the proposed architecture reconstructs high-quality SWIR images while preserving both spectral fidelity and spatial detail.

The primary scientific contribution of this work is the development of a novel asymmetric dual-path architecture for cross-spectral image restoration. Unlike conventional multimodal fusion networks that allow bidirectional feature exchange, the proposed model employs asymmetric cross-spectral attention in which VNIR features guide the refinement of SWIR representations while the VNIR stream remains unchanged. This design prevents corrupted SWIR information from contaminating the cleaner VNIR representation, enabling more reliable feature fusion throughout the network.

A second contribution is the introduction of a decoder architecture that combines SWIR-guided spatial attention gates with direct VNIR feature projections. The spatial attention gates selectively filter VNIR skip features according to the restoration requirements of the SWIR decoder, while the direct VNIR projections provide a stable pathway for transferring structural information at multiple spatial resolutions. Together, these components improve the reconstruction of fine image structures, edges, and textures without compromising the spectral characteristics of the restored SWIR image.

Furthermore, the paper proposes a composite restoration objective specifically

designed for cross-spectral image restoration. The loss function integrates robust pixel reconstruction, structural similarity, gradient consistency, VNIR-guided edge preservation, texture recovery, and distribution-based constraints to jointly optimize radiometric accuracy, structural fidelity, and high-frequency detail. A two-stage training strategy further improves restoration quality by first learning global image reconstruction and subsequently fine-tuning the model for localized inpainting of block-like corruptions.

Extensive experiments demonstrate the effectiveness and robustness of the proposed framework. The model consistently produces substantial improvements in image quality while preserving structural similarity, edge information, and radiometric consistency. Evaluation on an independent forest region further demonstrates that the learned cross-spectral representations generalize well to unseen geographical areas, highlighting the practical applicability of the proposed approach for remote sensing image restoration.

References

- [1] H. Abdullah, R. Darvishzadeh, A. K. Skidmore, T. A. Groen, and M. Heurich. European spruce bark beetle (*ips typographus*, l.) green attack affects foliar reflectance and biochemical properties. *International journal of applied earth observation and geoinformation*, 64:199–209, 2018.
- [2] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, and M. Heurich. Timing of red-edge and shortwave infrared reflectance critical for early stress detection induced by bark beetle (*ips typographus*, l.) attack. *International Journal of Applied Earth Observation and Geoinformation*, 82:101900, 2019.
- [3] M. Aliyu and I. Y. Chedi. Analyses of vegetation spectral characteristics for accurate identification. *International Journal of Research and Innovation in Applied Science*, 10(7):1168–1175, 2025.
- [4] E. Barnes, T. Clarke, S. Richards, P. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, C. Choi, E. Riley, T. Thompson, et al. Co-incident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In *Proceedings of the fifth international conference on precision agriculture, Bloomington, MN, USA*, volume 1619, 2000.
- [5] R. Bracewell. *The Fourier Transform And Its Applications*. McGraw-Hill, 2000.
- [6] T. K. Brekken, B. A. Batten, and E. A. Amon. From blue to green [ask the experts]. *IEEE Control Systems Magazine*, 31(5):18–24, 2011.
- [7] P. Brodtkorb, P. Johannesson, G. Lindgren, I. Rychlik, J. Rydén, and E. Sjö. WAFO - a Matlab toolbox for the analysis of random waves and loads. In *Proc. 10'th Int. Offshore and Polar Eng. Conf., ISOPE, Seattle, USA*, volume III, 2000.

- [8] R. L. Burden and J. D. Faires. *Numerical Analysis*. Brooks/Cole, 6th edition, 1997.
- [9] G. Camps-Valls et al. Deep learning for the earth sciences: A comprehensive review. *Nature Reviews Earth & Environment*, 2:1–18, 2021.
- [10] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [11] Z. Chen, J. Yuan, J. Liu, et al. Artificial intelligence-aided new paradigm of environmental research. *AI Environ.*, 1(1):23–32, 2026. doi: 10.66178/aie-0026-0004.
- [12] L. P. Cinelli, M. A. Marins, E. A. B. Da Silva, and S. L. Netto. *Variational methods for machine learning with applications to deep networks*, volume 15. Springer, 2021.
- [13] A. Clément, P. McCullen, A. Falcão, A. Fiorentino, F. Gardner, K. Hammarlund, G. Lemonis, T. Lewis, K. Nielsen, S. Petroncini, M.-T. Pontes, P. Schild, B.-O. Sjöström, H. C. Sørensen, and T. Thorpe. Wave energy in europe: current status and perspectives. *Renewable and Sustainable Energy Reviews*, 6(5):405–431, 2002. ISSN 1364-0321. doi: 10.1016/S1364-0321(02)00009-6. URL <https://www.sciencedirect.com/science/article/pii/S1364032102000096>.
- [14] L. H. E. Cosimo, J. G. Lopez, E. Lindberg, and L. Huo. Linking hyper-spectral remote sensing and tree hydraulic functioning for pre-emergence detection of bark beetle infestations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 237:628–651, 2026. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2026.04.038. URL <https://www.sciencedirect.com/science/article/pii/S092427162600211X>.
- [15] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The annals of probability*, pages 146–158, 1975.
- [16] G. Dahlquist and Å. Björck. *Numerical methods*. Courier Corporation, 2003.
- [17] D. Dai, C. Dong, Q. Yan, Y. Sun, C. Zhang, Z. Li, and S. Xu. I2u-net: A dual-path u-net with rich information interaction for medical image segmentation. *Medical Image Analysis*, 97:103241, 2024.
- [18] A. F. de O. Falcão. Wave energy utilization: A review of the technologies. *Renewable and Sustainable Energy Reviews*, 14(3):899–918, 2010.

- ISSN 1364-0321. doi: 10.1016/j.rser.2009.11.003. URL <https://www.sciencedirect.com/science/article/pii/S1364032109002652>.
- [19] J. D’Errico. shape language modelling, 2024. URL <https://ch.mathworks.com/matlabcentral/fileexchange/24443-slm-shape-language-modeling>.
 - [20] I. Dincer, M. A. Rosen, and F. Khalid. 3.8 ocean (marine) energy production. In I. Dincer, editor, *Comprehensive Energy Systems*, pages 335–379. Elsevier, Oxford, 2018. ISBN 978-0-12-814925-6. doi: 10.1016/B978-0-12-809597-3.00316-3. URL <https://www.sciencedirect.com/science/article/pii/B9780128095973003163>.
 - [21] B. Drew, A. Plummer, and M. Sahinkaya. A review of wave energy converter technology. *Proceedings of The Institution of Mechanical Engineers Part A-journal of Power and Energy - PROC INST MECH ENG A-J POWER*, 223:887–902, 12 2009. doi: 10.1243/09576509JPE782.
 - [22] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al. Sentinel-2: Esa’s optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.
 - [23] İ. Ö. Erselcan and A. Kükner. A review of power take-off systems employed in wave energy converters. *Journal of Naval Science and Engineering*, 10(1):32–44, 2014.
 - [24] European Marine Energy Centre. Billia Croo wave data base, 2022.
 - [25] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.
 - [26] B. Gao. NdwI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266, 1996. ISSN 0034-4257. doi: 10.1016/S0034-4257(96)00067-3. URL <https://www.sciencedirect.com/science/article/pii/S0034425796000673>.
 - [27] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata. Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12325–12334, 2020.
 - [28] E. P. Glenn, A. R. Huete, P. L. Nagler, and S. G. Nelson. Relationship between remotely-sensed vegetation indices, canopy attributes and plant

- physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors*, 8(4):2136–2160, 2008.
- [29] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
 - [30] Government of British Columbia. Spruce beetle — biology, 2024. URL https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/forestry/forest-health/bark-beetles/spruce_beetle_biology.pdf. Accessed 2026-05-11.
 - [31] E. Guisado-Pintado. 17 - shallow water wave modelling in the nearshore (swan). In D. W. Jackson and A. D. Short, editors, *Sandy Beach Morphodynamics*, pages 391–419. Elsevier, 2020. ISBN 978-0-08-102927-5. doi: 10.1016/B978-0-08-102927-5.00017-5. URL <https://www.sciencedirect.com/science/article/pii/B9780081029275000175>.
 - [32] M. P. Hansen and D. S. Malchow. Overview of swir detectors, cameras, and applications. In *Proceedings of SPIE — The International Society for Optical Engineering*, volume 6939, pages 94–104. SPIE, 2008.
 - [33] K. Hasselmann, T. P. Barnett, E. Bouws, H. Carlson, D. E. Cartwright, K. Enke, J. Ewing, A. Gienapp, D. Hasselmann, P. Kruseman, et al. Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A*, 1973.
 - [34] C. Hechtel, S. Hauser, A. Schmitt, M. Heurich, and A. Wendleder. Kennaugh elements allow early detection of bark beetle infestation in temperate forests using sentinel-1 data. *Forests*, 16(8), 2025. ISSN 1999-4907. doi: 10.3390/f16081272. URL <https://www.mdpi.com/1999-4907/16/8/1272>.
 - [35] L. H. Holthuijsen. *Waves in Oceanic and Coastal Waters*. Cambridge University Press, 2007. doi: 10.1017/CBO9780511618536.
 - [36] D. Horler, M. Dockray, and J. Barber. The red edge of plant leaf reflectance. *International journal of remote sensing*, 4(2):273–288, 1983.
 - [37] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
 - [38] L. Huo, H. J. Persson, and E. Lindberg. Early detection of forest stress from european spruce bark beetle attack, and a new vegetation

- index: Normalized distance red & swir (ndrs). *Remote Sensing of Environment*, 255:112240, 2021. ISSN 0034-4257. doi: 10.1016/j.rse.2020.112240. URL <https://www.sciencedirect.com/science/article/pii/S0034425720306131>.
- [39] L. Huo, L. H. E. Cosimo, A. Bozzini, S. Bijou, R. A. de Oliveira, J. Suomalainen, E. Vennervirta, N. Koivumäki, R. Näsi, L. Kupková, M. Faccoli, and E. Honkavaara. A cross-european assessment on the pre-emergence detection of trees attacked by spruce bark beetle using uav imagery. *Remote Sensing of Environment*, 342:115445, 2026. ISSN 0034-4257. doi: 10.1016/j.rse.2026.115445. URL <https://www.sciencedirect.com/science/article/pii/S0034425726002154>.
- [40] L. Huo et al. Assessing the detectability of european spruce bark beetle infestations using satellite data. *Remote Sensing*, 2023. URL <https://pub.epsilon.slu.se/30460/1/huo-l-et-al-20230322.pdf>.
- [41] A. M. Hussain A, Arif SM. Emerging renewable and sustainable energy technologies: State of the art. *Renew. sustain. energy rev.*, 71:12–28, 2017. ISSN 1364-0321. doi: 10.1016/j.rser.2016.12.033.
- [42] Š. Hýsek, R. Löwe, and M. Turčáni. What happens to wood after a tree is attacked by a bark beetle? *Forests*, 12(9):1163, 2021.
- [43] IPCC. Summary for policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1–31. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2021. doi: 10.1017/9781009157896.001. URL https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf.
- [44] IPCC. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022. doi: 10.1017/9781009157926. URL https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf.
- [45] IRENA. Renewable capacity statistics 2025. Technical report, International Renewable Energy Agency, 2025.

- [46] ITTC. The specialists committee on waves, final report and recommendations to the 23rd ITTC. In *Proceedings of the 23rd ITTC*, volume II, pages 505–736, 2002.
- [47] A. Katharopoulos, A. Vyas, N. Pappas, and F. Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pages 5156–5165. PMLR, 2020.
- [48] D. M. Kim, H. Zhang, H. Zhou, T. Du, Q. Wu, T. C. Mockler, and M. Y. Berezin. Highly sensitive image-derived indices of water-stressed plants using hyperspectral imaging in swir and histogram analysis. *Scientific reports*, 5(1):15919, 2015.
- [49] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [50] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998. doi: 10.1137/S1052623496303470. URL <https://doi.org/10.1137/S1052623496303470>.
- [51] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.
- [52] J. Leishman and G. Scobie. The development of wave power: a techno-economic study. Technical report, Department of Industry: NATIONAL ENGINEERING LABORATORY, 1976.
- [53] X. Li, Y. Wang, H. Chen, et al. Artificial intelligence in environmental monitoring: Opportunities and challenges. *Environmental Science and Ecotechnology*, 20:100321, 2024.
- [54] G. Lindgren, H. Rootzén, and M. Sandsten. Stationary stochastic processes. *Theory and Applications; Texts in Statistical Science Series; CRC Press: Boca Raton, FL, USA*, page 347, 2013.
- [55] E. Mackay. Resource assessment for wave energy. In A. Sayigh, editor, *Comprehensive Renewable Energy*, volume 8, pages 11–72. Elsevier, Oxford, 2012.
- [56] G. Mork, S. Barstow, A. Kabuth, and M. T. Pontes. Assessing the global wave energy potential. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 49118, pages 447–454, 2010.

- [57] D. S. Nandi. Wave energy - the renewable energy of the future. *Quest Journals: Journal of Research in Humanities and Social Science*, 11(2): 275–281, 2023. ISSN 2321-9467. URL <https://www.questjournals.org/jrhss/papers/vol11-issue2/1102275281.pdf>.
- [58] J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [59] M. Neshat, M. Nezhad, N. Sergiienko, S. Mirjalili, G. Piras, and D. Garcia. Wave power forecasting using an effective decomposition-based convolutional Bi-directional model with equilibrium Nelder-Mead optimiser. *Energy*, 256:124623, 2022. doi: 10.1016/j.energy.2022.124623.
- [60] S. Ocean. Ocean energy: cost of energy and cost reduction opportunities. Technical report, SI OCEAN, 2013.
- [61] M. K. Ochi and E. N. Hubble. Six-parameter wave spectra. In *Proceedings of the 15th International Conference on Coastal Engineering*, pages 301–328. American Society of Civil Engineers, Honolulu, Hawaii, USA, 1976.
- [62] S. C. on Modelling of Environmental Conditions of the 28th ITTC. Laboratory modeling of waves: Regular, irregular and extreme events. Technical report, International Towing Tank Conference (ITTC), Switzerland, 2017. URL <https://www.ittc.info/media/8095/75-02-07-012.pdf>.
- [63] C. Park and M. Allaby. *A Dictionary of Environment and Conservation*. Oxford University Press, 2017. ISBN 9780191826320. doi: 10.1093/acref/9780191826320.001.0001.
- [64] W. J. Pierson Jr and L. Moskowitz. A proposed spectral form for fully developed wind seas based on the similarity theory of sa kitaigorodskii. *Journal of geophysical research*, 69(24):5181–5190, 1964.
- [65] A. Reichmuth, L. Henning, N. Pinnel, M. Bachmann, and D. Rogge. Early detection of vitality changes of multi-temporal norway spruce laboratory needle measurements—the ring-barking experiment. *Remote Sensing*, 10(1):57, 2018.
- [66] D. J. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014.
- [67] H. Ritchie. How have the world’s energy sources changed over the last two centuries? Technical report, Global Change Data Lab, 2021. URL <https://ourworldindata.org/global-energy-200-years>.

- [68] L. Rokach, O. Maimon, and E. Shmueli. Machine learning for data science handbook. *Springer* [https://doi.org/10, 1007:978-3, 2023](https://doi.org/10.1007:978-3, 2023).
- [69] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [70] J. Rouse, J. W., D. W. Deering, R. H. Haas, and J. A. Schell. Monitoring vegetation systems in the great plains with erts. Technical report, Texas A&M University, 1974. URL <http://hdl.handle.net/2060/19740022614>. NASA Technical Report.
- [71] R. A. Schowengerdt. *Remote sensing: models and methods for image processing*. elsevier, 2006.
- [72] R. Seidl, D. Thom, M. Kautz, D. Martin-Benito, M. Peltoniemi, G. Vacchiano, J. Wild, D. Ascoli, M. Petr, J. Honkaniemi, et al. Forest disturbances under climate change. *Nature climate change*, 7(6):395–402, 2017.
- [73] K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28, 2015.
- [74] R. H. Stewart. *Introduction to physical oceanography*. Texas A&M University, 2008.
- [75] The MathWorks, Inc. Simulink, 2022. URL <https://se.mathworks.com/products/simulink.html>.
- [76] The MathWorks, Inc. fminsearch - find minimum of unconstrained multi-variable function, 2024. URL <https://ch.mathworks.com/help/matlab/ref/fminsearch.html>.
- [77] A. Toffoli and E. M. Bitner-Gregersen. Types of ocean surface waves, wave classification. *Encyclopedia of maritime and offshore engineering*, pages 1–8, 2017.
- [78] University of Maryland Extension. Conifer bark beetles on trees and shrubs. *UMD Extension*, 2024. URL <https://extension.umd.edu/resource/conifer-bark-beetles-trees-and-shrubs>. Accessed 2026-05-11.
- [79] USDA Forest Service. Introduction to bark beetles. *USDA Forest Service*, 2010. URL <https://www.fs.usda.gov/sites/nfs/files/legacy-media/r02/2pager%20-%20intro%20bark%20beetles.pdf>. Accessed 2026-05-11.

- [80] S. L. Ustin and E. M. Middleton. Current and near-term advances in earth observation for ecological applications. *Ecological Processes*, 10(1):1, 2021.
- [81] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [82] G. Vivone. Multispectral and hyperspectral image fusion in remote sensing: A survey. *Information Fusion*, 89:405–417, 2023.
- [83] N. Wiener. Generalized harmonic analysis. *Acta Mathematica*, 55:117–258, 1930.
- [84] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
- [85] M. A. Wulder, D. P. Roy, V. C. Radeloff, T. R. Loveland, M. C. Anderson, D. M. Johnson, S. Healey, Z. Zhu, T. A. Scambos, N. Pahlevan, M. Hansen, N. Gorelick, C. J. Crawford, J. G. Masek, T. Hermosilla, J. C. White, A. S. Belward, C. Schaaf, C. E. Woodcock, J. L. Huntington, L. Lymburner, P. Hostert, F. Gao, A. Lyapustin, J.-F. Pekel, P. Strobl, and B. D. Cook. Fifty years of landsat science and impacts. *Remote Sensing of Environment*, 280:113195, 2022. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2022.113195>. URL <https://www.sciencedirect.com/science/article/pii/S0034425722003054>.

Author contributions

Shokoufa Zeinali's contributions to each of the papers included in the thesis are mentioned below:

Paper I: Optimizing the hydraulic power take-off system in a wave energy converter (Published)

The Simulink model was provided by Jan Forsberg from Waves4Power. The design of the study and the selection of target variables was done in collaboration with the company. I implemented the simulation study and performed the analysis of the results. Theoretical derivations and proofs were done in collaboration with co-authors (mainly Magnus). I wrote most of the first draft (the appendix describing the Simulink model was written in close collaboration with Jan). The final revision was done in collaboration with co-authors.

Paper II: The effects of sea-state on optimal generator parameters for a wave energy converter (Published)

Again, the Simulink model was provided by the company Waves4Power, Information regarding hydrodynamics was provided by the co-authors from Chalmers. Several revisions were made to the initial Simulink model based on discussion between Jan and me to account for artifacts that arose during the initial simulation studies. I designed, implemented, and analyzed the simulation study. I also wrote the first draft, and took main responsibility for the revisions.

Paper III: Detection of Bark Beetle Attacks Using Time-Aggregated Satellite Data With Machine Learning (Published)

Per-Ola Olsson provided the Sentinel-2 and static feature data, along with important insights into the problem. I was responsible for designing and implementing the classification. The work underwent several rounds of revision in collaboration with my co-authors.

Paper IV: Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series (Submitted)

As with Paper III, data and biological insight into the problem was provided by Per-Ola Olsson. The main ideas for the classification are mine, but they were refined in discussion with the co-authors, particularly Ted Krovall and Johan Lindström. I implemented everything and carried out all the experiments and analysis. I wrote most of the initial draft, with co-authors contributing to several rounds of revision.

Paper V: Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction

Langning Hou provided the drone imagery, and several meetings between Langning and me resulted in the final problem formulation. I was responsible for designing and implementing the neural network used for the reconstruction. The design of the neural network was refined in discussion with Ted Krovall and Johan Lindström. I also wrote the first draft, and took main responsibility for the revisions.

Paper I





Research paper

Optimizing the hydraulic power take-off system in a wave energy converter

Shokoufa Zeinali ^{a,*}, Magnus Wiktorsson ^a, Jan Forsberg ^{b,c}, Georg Lindgren ^a, Johan Lindström ^a^a Mathematical Statistics, Centre for Mathematical Sciences, Lund University, 22100, Lund, Sweden^b Environmental and Energy Systems, Department of Engineering and Chemical Sciences, Karlstad University, 65188, Karlstad, Sweden^c W4P Waves4Power AB, 41292, Gothenburg, Sweden

ARTICLE INFO

Keywords:

Bretschneider spectrum
Wave steepness
Simulation study
Uncertainty analysis
Concave regression
Semi-parametric bootstrap

ABSTRACT

This study aims to determine the optimal pressure for the accumulator tank in a wave energy converter (WEC) with hydraulic power take-off (PTO) to maximize energy generation. A simulation-based approach was employed to analyze the buoy's performance with wave data generated from the Wafu toolbox using the generalized Pierson–Moskowitz (GPM) spectrum. These data were then fed into a Simulink model of the WEC system. Using output from the Simulink model, concave regression was applied to model the relationship between tank pressure and power output. Subsequently, a regression analysis was performed to model the optimal pressure as a function of wave steepness. The findings indicate a strong dependency of optimal pressure on wave steepness, suggesting that adjusting the accumulator tank pressure according to wave conditions can significantly enhance the WEC's energy output. This research contributes to the ongoing efforts to improve the efficiency of WECs and provides valuable insights for developing adaptive control strategies in WEC systems.

1. Introduction

The continued growth of renewable energy plays a key role in reducing greenhouse gas emissions and limiting climate change. A review of emerging renewable energy (Hussain et al., 2017) lists marine power, i.e., ocean and tidal-based, as having great potential but still facing technical and infrastructure challenges.

In this paper, we aim to address some of the technical challenges. Specifically, those related to selecting optimal design parameters for a wave energy converter (WEC) while accounting for uncertainties both among and within different sea states.

1.1. Wave energy converters

A WEC extracts electrical power from ocean waves by converting wave energy to electrical energy. Several different designs for WECs exist, and Ahamed et al. (2020) provides an overview. The power take-off (PTO) system for a WEC converts the absorbed mechanical energy to electric energy, and depends on both the internal workings of the said WEC and the sea state in which the WEC acts. Optimizing design and operating rules for WECs is a multi-faceted matter that needs to balance energy absorption against durability and robustness. Flexibility against variability in sea state can be built into the design or achieved through intelligent but sensitive regulation (Falnes and Kurniawan, 2020).

1.1.1. Effects of sea-state

Previous studies have investigated the effect of different sea state climates on the economy and durability of WECs (e.g. Mackay, 2012; Neshat et al., 2022). Further, simple WEC models have been used in detailed studies of how individual wave characteristics, such as wave geometry and groupiness, influence WEC efficiency (Saulier et al., 2011; Lindgren, 2015; Wang, 2018, 2019, 2020, 2021; Wang and Wang, 2018). More elaborate WEC design studies of heaving buoys are reported by Ahmed et al. (2022), Azam et al. (2022), using similar low-energy sea state as in the Wang-studies.

Response studies combine sea state and wave characteristics with the hydrodynamics of the WEC as well as the PTO system used for energy conversion; many early response studies use simple harmonic waves. Works with more complex wave models are Zurkinden et al. (2014) use of a toy buoy in irregular waves; physical experiments in a sea state with bimodal spectrum (Huang et al., 2022); and mathematical analysis of non-linear effects of different buoy shapes in single frequency sea state (Kim et al., 2021).

1.1.2. Hydraulic PTO systems

Hydraulic PTO systems (Ahamed et al., 2020, Sec. 2.1) generally consist of a hydraulic cylinder that pressurizes a hydraulic medium in an accumulator tank which is used to drive a generator. The hydraulic

* Corresponding author.

E-mail addresses: shokoufa.zeinali@matstat.lu.se (S. Zeinali), magnus.wiktorsson@matstat.lu.se (M. Wiktorsson), jan.forsberg@fluidynamik.se, jan.forsberg@kau.se (J. Forsberg), georg.lindgren@matstat.lu.se (G. Lindgren), johan.lindstrom@matstat.lu.se (J. Lindström).

<https://doi.org/10.1016/j.oceaneng.2024.118636>

Received 5 October 2023; Received in revised form 18 June 2024; Accepted 30 June 2024

Available online 14 July 2024

0029-8018/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

cylinder can utilize both translational and rotational wave energy, allowing for use in many different types of WECs. A key feature of the hydraulic system is the target pressure in the accumulator tank. A too-low pressure limits the energy uptake from the waves by providing too little resistance in the hydraulic cylinder, while a too-high pressure will counteract the wave energy, locking the cylinder and leading to zero energy production (Babarit et al., 2009).

Even if the accumulator pressure varies during operation, the optimal design pressure is of importance since it determines the system's potential efficiency, and informs design choices regarding the hydraulic cylinder, size of the accumulator tank, and generator characteristics. Further, the variability in optimal pressure across different sea states will determine how often the WEC can achieve optimal performance and which locations might be suitable for deploying a specific WEC.

1.1.3. Optimization of PTO systems

Optimization of hydraulic PTO systems has recently been done with various techniques and for a number of different WEC constructions. In Jusoh et al. (2020) a 10 kW PTO system at the coastline is optimized using genetic algorithms, Amini et al. (2022) uses several different optimization methods to optimize parameters of a floating PTO system. In Xue et al. (2023) an eccentric rotating wave energy converter is optimized using Gaussian process regression. In Liu et al. (2024) the accumulator pressure for a floating PTO system is optimized. Falcão (2007) designed a simple control system for a WEC in which the control law depends on significant wave height. García-Rosa et al. (2009) used equations buoy motion to link maximum absorption to buoy parameters, wave height, and the frequency of incident waves. These studies consider different WEC systems and a wide variety of optimization methods as well as aspects of the sea states for the optimal running condition. In this study, we aim to expand on the link between sea states and optimal operating parameters, by considering a large range of different sea states and we will attempt to identify relationships between the optimal accumulator pressure and the realized wave conditions.

1.1.4. WEC in this study

The WEC considered in this study is WaveEl 4.0 owned, designed and developed by Waves4Power, see Figs. 1 and 3. This WEC is of a two-body type with a hydraulic PTO system. One body of the WEC is an oscillating water column (OWC), where the hydraulic pressure is provided by a moving piston in a water tube. A previous version of the WEC in question, WaveEl 3.0, was installed by Waves4Power in 2016 close to the Runde Environment Center AS, near Ålesund in Norway. The main difference between the WaveEl 3.0 and 4.0 buoys is the lengthening of the water tube from 18.5 to 34 meter. The first installation of the 3.0 buoy produced energy for the Norwegian grid between February 2016 and November 2017. Safety studies of extreme loading and fatigue for the system were reported by Gustafsson (2016), Johannesson et al. (2022) and Shao et al. (2023) carried out a detailed performance analysis of WaveEl 3.0, focusing on mooring forces and using a simplified PTO model. WECs of the OWC type are regarded as the most reliable and efficient WEC solutions available, (Falcão and Henriques, 2016; Kasacik et al., 2020; Rosati et al., 2022).

1.2. Study aims

The present study aims to find the optimum pressure of the accumulator tank at which the Waves4Power WEC buoy generates maximum energy in different sea states. The different sea states used here are based on wave statistics from the European Marine Energy Centre at Billia Croo, Orkney (European Marine Energy Centre, 2022). The wave characteristics are described in Section 2.1, and the frequency of different sea states is given in Table 1 of Appendix A.

Following Penalba and Ringwood (2016), the buoy is represented by a Simulink model including all the linear and non-linear steps between

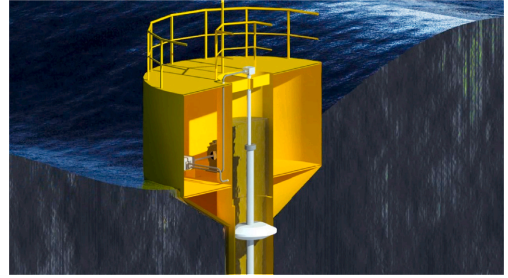


Fig. 1. Cross-sectional image of an oscillating water column (OWC) type wave energy converter (WEC) buoy, with a hydraulic PTO. Illustrating the water tube and moving piston that generates the hydraulic pressure for the PTO system.

the wave movements and the electric power output from the generator, as well as a number of adjustable parameters. The wave input to the Simulink model is represented by Monte Carlo-generated irregular linear Gaussian waves in different sea states, i.e. with different energy spectra. Section 2.2 describes the simulation procedure and Simulink model.

As both the input and output are stochastic, any optimization should be done over time, accounting for both variability among different sea states and uncertainties within a given sea state. Results of the parameter optimization for the buoy in this study are given in Section 3 and concluding remarks can be found in Section 4.

2. Material and methods

2.1. Wave models and characteristics

For this study, we used waves generated from the generalized Pierson Moskowitz (GPM) or Bretschneider spectrum according to the ITTC-standard (ITTC, 2002, Appendix A), with wave variance m_0 and spectral peak $\omega_p = 2\pi/T_p$; see Fig. 2 for an example. We truncated the spectrum at $5\omega_p$ which preserves 99.8% of the spectral energy and results in the following spectrum,

$$S_{PM}(\omega) = \begin{cases} m_0 \frac{5}{\omega_p} \left(\frac{\omega}{\omega_p} \right)^{-5} e^{-\frac{5}{4} \left(\frac{\omega}{\omega_p} \right)^{-4}}, & 0 \leq \omega \leq 5\omega_p, \\ 0, & \omega > 5\omega_p. \end{cases} \quad (1)$$

The significant wave height is four times the standard deviation of the wavelength, or $H_s = 4\sqrt{m_0}$, and the spectral moments are, see Appendix B.1 for a derivation,

$$\begin{cases} m_n = \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \Gamma \left(1 - \frac{n}{4} \right), & \text{standard spectrum,} \\ m_n^{\text{trunc}} = \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \Gamma \left(1 - \frac{n}{4}, \frac{1}{500} \right), & \text{truncated spectrum.} \end{cases} \quad (2)$$

Here $\Gamma(a)$ is the standard gamma function (Eq. 6.1.1 on p. 255 in Abramowitz and Stegun, 1965) and $\Gamma(a, b)$ is the upper incomplete gamma function (Eq. 6.5.3 on p. 260 in Abramowitz and Stegun, 1965).

The European Marine Energy Centre (2022) data for Billia Croo (cf. Appendix A) is presented using significant wave heights, H_s and energy periods, T_e , and we need to establish a relation between the energy period, T_e , and the peak period, $T_p = 2\pi/\omega_p$, in a truncated GPM-spectrum. The energy period can be computed from the spectral moments as

$$T_e = 2\pi \frac{m_{-1}}{m_0}. \quad (3)$$

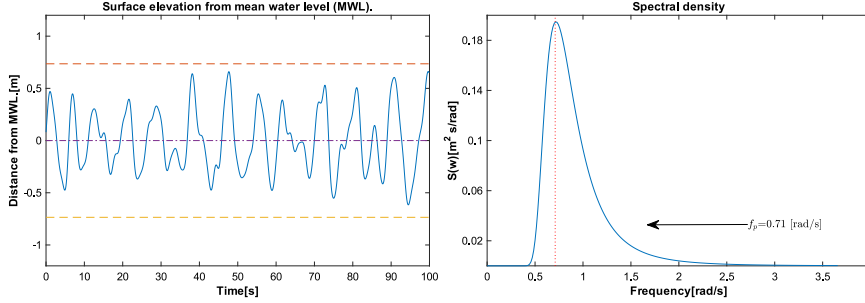


Fig. 2. Simulated waves (left) from a generalized Pierson Moskowitz (GPM) (or Bretschneider) spectrum (right) with significant wave height, $H_s = 1.25(m)$, and energy period, $T_e = 7.5(s)$. The waves have been simulated for 100s using a step size of $\Delta t = 0.1s$. The dashed lines in the left panel are ± 2 standard deviations of wave height (significant wave height is four standard deviations) and the vertical dashed line in the right pane is the peak frequency ($f_p = 2\pi/T_p \approx 0.86 \cdot 2\pi/T_e$, see Eq. (4)).

Using the truncated spectral moments in Eq. (2) this expands to

$$T_e = 2\pi \frac{m_1^{\text{trunc}}}{m_0^{\text{trunc}}} = \left(\frac{4}{5}\right)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}, \frac{1}{500}\right) e^{\frac{1}{500}} T_p \approx 0.86 T_p, \quad (4)$$

where m_n^{trunc} are given by Eq. (2).

2.1.1. Energy in the waves

We will need the energy content in the waves to evaluate the efficiency of the buoy. According to [Falnes and Kurniawan \(2020, Eq. 4.188 on p. 91\)](#) the energy transport per meter wave crest per second in deep water is

$$J = \frac{g^2 \rho}{2} m_{-1} = \frac{g^2 \rho}{64\pi} T_e H_s^2. \quad (5)$$

where g is the standard gravity, ρ is the water density, m_{-1} is the spectral moment with $n = -1$, we have used Eq. (3) for energy period, T_e , and that the significant wave height is $H_s = 4\sqrt{m_0}$ to express the energy using these quantities (see Eq. 4.190 in [Falnes and Kurniawan, 2020, p. 91](#)). It is important to note that Eq. (5) does not depend on the truncation of the spectrum.

2.1.2. Steepness of waves

Wave steepness is an important characteristic for many ocean engineering applications, ranging from ship safety ([Toffoli et al., 2005](#)) to coastal structures ([Van der Meer, 1999](#)). The steepness can be defined using wave height in combination with either energy period, T_e , or mean zero-crossing period, T_z , ([Myrhaug, 2018](#)). Resulting in two slightly different formulas for mean wave steepness, S_p , and significant wave steepness, S_s ,

$$S_p = \frac{2\pi H_s}{g T_e^2}, S_s = \frac{2\pi H_s}{g T_z^2}. \quad (6)$$

Here T_z is the mean zero-crossing period defined from the spectral moments as

$$T_z = 2\pi \sqrt{\frac{m_0}{m_2}}$$

Using the computations of the (truncated) spectral moments in Eq. (2) we get that (see [Appendix B.2](#) for derivation)

$$S_s = \begin{cases} \sqrt{\frac{5}{4}} \Gamma\left(\frac{1}{2}\right) \frac{2\pi H_s}{g T_e^2} \\ \exp\left(\frac{1}{500}\right) \sqrt{\frac{5}{4}} \Gamma\left(\frac{1}{2}, \frac{1}{500}\right) \frac{2\pi H_s}{g T_e^2} \end{cases} \quad (7)$$

$$\approx \begin{cases} 1.98 S_p, & \text{standard spectrum,} \\ 1.89 S_p, & \text{truncated spectrum.} \end{cases}$$

Thus, the two steepness definitions only differ by a multiplicative constant for the (truncated) GPM spectrum.

2.2. Simulation study

The Waves4Power WEC buoy is modeled as a Simulink point absorber driven by a single time-series of generated or measured waves. The simulation study of the buoy consists of several steps: (1) Generating waves from the truncated GPM spectrum using different sea states, see Section 2.2.1; (2) Prefiltering the waves to account for the physical buoy geometry in relation to the waves, see Section 2.2.2; (3) Feeding the filtered waves into the Simulink model, see Section 2.2.3; and (4) Processing of the simulation results is described in Section 2.2.4. Then, in Section 3, the optimization methods described in Section 2.3 are applied to the simulation results to determine optimal WEC parameters and investigate how the parameters vary between different sea states.

2.2.1. Generation of irregular input waves

In this study, we generate the input waves from a linear Gaussian time-series (t, Z_t) with N samples at sampling times $t = 0, \Delta_t, 2\Delta_t, \dots, T$ where $\Delta_t = \frac{T}{N-1}$ and Z_t denoting the height of the undisturbed sea surface at the buoy location. The time-series Z_t will be simulated as a Fourier sum

$$Z_t = \sum_{i=0}^{N-1} \sqrt{S(\omega_i) \Delta\omega} R_i \cos(\omega_i t + \theta_i) \quad (8)$$

where the frequency steps are $\Delta\omega = \frac{2\pi}{N-1}$; $\omega_i = i\Delta\omega$; $S(\omega_i)$ is the truncated GPM spectrum from Eq. (1); θ_i are independent random phases, uniformly distributed in $(0, 2\pi)$; and R_i are Rayleigh distributed amplitudes with probability density

$$f(r) = r \exp(-r^2/2), r \geq 0. \quad (9)$$

The sampling in Eq. (8) is implemented in the MATLAB toolbox *Wafo* (Wave Analysis for Fatigue and Oceanography; [Brodtkorb et al., 2000](#)), which is freely available online ([WAFO-group, 2017](#)). The *Wafo* routine `spec2sdat` can simulate with (almost) any time-interval $[0, T)$ and sampling interval Δ_t . For small Δ_t , the spectrum may be extended and, if needed, padded with zeros at high frequencies. For large T , the spectrum will be interpolated with a small frequency step $\Delta\omega$ to avoid periodicity.

2.2.2. Prefiltering of the waves

The buoy acts as a point absorber in the sea state with dominating waves much longer than the buoy dimension. For a realistic irregular sea state with a fair amount of short waves, the shape and dimensions of the horizontal sections of the buoy need to be taken into account. The studied buoy has a ring-shaped upper part on top of a conical section attached to an open-ended bottom pipe. This hollow pipe extends through the cone and all the way to the top of the construction.

Under modest sea conditions, the average buoyancy force over the ring-shaped section can be used to approximate the total buoyancy force. Before feeding the simulated waves into the Simulink model, they will be prefiltered to take account of this hydrostatic averaging. Hydrodynamic forces from the added mass and damping due to the, customarily submerged, conical section are taken care of in the Simulink model and will not be included in the averaging.

We use a strictly linear wave model, treating each frequency and wave direction as independent and additive, resulting in a final wave that is a sum of harmonic waves, see Eq. (8). Consider one pure harmonic unidirectional wave

$$W(x, t) = A_w \cos(\kappa x - \omega t) = A_w \cos\left(\frac{\omega^2}{g} x - \omega t\right),$$

where we used that $\kappa = \omega^2/g$, moving in the positive x -direction over a ring-shaped region (annulus) with outer radius R_2 , inner radius R_1 and centered at the origin $x = y = 0$. The wave amplitude is A_w , and we seek a *phantom amplitude* $\tilde{A}_w$ such that a wave $\tilde{A}_w \cos(\kappa x - \omega t)$ would give the same hydrostatic buoyancy variation at the origin as the original wave gives on average over the ring.

The maximal buoyancy occurs when a crest is located at the origin which occurs at time $t = 0$ (and $t = \frac{2\pi}{\omega}, \frac{4\pi}{\omega}, \dots$). Then the wave height at a point $(r \cos \theta, r \sin \theta)$ is $A_w \cos\left(\frac{\omega^2}{g} r \cos \theta\right)$ with the average height over the ring being equal to

$$\frac{A_w}{2\pi(R_2^2 - R_1^2)} \int_{r=R_1}^{R_2} \int_{-\pi}^{\pi} \cos\left(\frac{\omega^2}{g} r \cos \theta\right) r \, d\theta \, dr,$$

which is the maximum hydrostatic buoyancy. When the trough is at the origin, the average amplitude changes sign, giving the minimum buoyancy.

Thus, the corresponding *phantom amplitude* can be expressed as a linear filter, $\tilde{A}_w = A_w \times H(\omega)$, where the ring acts as a filter upon the original waves with the frequency response function

$$\begin{aligned} H(\omega) &= \frac{1}{2\pi(R_2^2 - R_1^2)} \int_{r=R_1}^{R_2} \int_{-\pi}^{\pi} \cos\left(\frac{\omega^2}{g} r \cos \theta\right) r \, d\theta \, dr \\ &= \frac{1}{(R_2^2 - R_1^2)} \int_{R_1}^{R_2} J_0\left(\frac{\omega^2 r}{g}\right) r \, dr \\ &= \frac{2g}{\omega^2 (R_2^2 - R_1^2)} \left(R_2 J_1\left(\frac{R_2 \omega^2}{g}\right) - R_1 J_1\left(\frac{R_1 \omega^2}{g}\right) \right). \end{aligned} \quad (10)$$

Here J is a Bessel function of the first kind (see e.g. Ch. 9 in Abramowitz and Stegun, 1965, with the two integrals given by Eq. 9.1.18 and 9.1.30); R_1 and R_2 are the inner and outer radii of the ring, respectively. Note, that Eq. (10) gives the amplitude response for each frequency and each direction in a multi-directional sea.

For simulation purposes, it is useful to formulate the filter in the time domain, since we may want to generate waves by many different methods, and not only using Fourier methods in the frequency domain. The impulse response function for the filter is computed using the inverse Fourier transform of the frequency filter. Using the results from (Eq. 3.691.7 on p. 419 in Gradshteyn et al., 2015), a change of variable, and then the results from (Eq. 3.775.7-8 on p. 450 in Gradshteyn et al., 2015) together with the relation between the Bessel J and Y functions we obtain the impulse response

$$h(t) = \frac{g|t|}{48(R_2^2 - R_1^2)} (f(t, R_2) - f(t, R_1)) \quad (11)$$

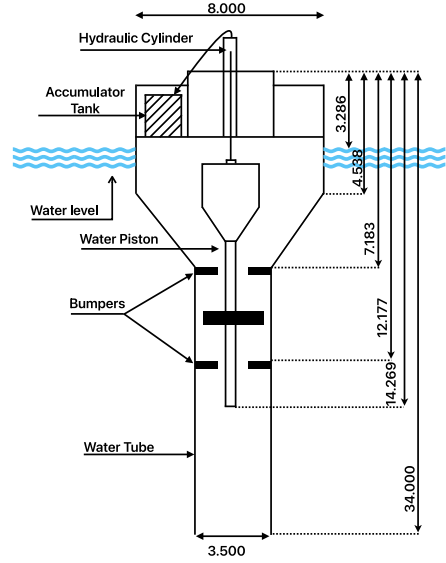


Fig. 3. Diagram of the WaveEl 4.0 buoy used in this study, with key dimensions and components.

where

$$\begin{aligned} f(t, x) &= \left(\frac{t^4 g}{2x} + 8x \right) \left(J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right)^2 - J_{\frac{1}{4}}\left(\frac{t^2 g}{8x}\right)^2 \right) \\ &\quad - 2gt^2 \left(J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) J_{\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) \right. \\ &\quad \left. + J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) J_{\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) \right) \\ &\quad + \frac{t^4 g}{2x} \left(J_{\frac{3}{4}}\left(\frac{t^2 g}{8x}\right)^2 - J_{-\frac{3}{4}}\left(\frac{t^2 g}{8x}\right)^2 \right). \end{aligned} \quad (12)$$

This time filter will be used to prefilter any generated waves before sending them as input to the Simulink model.

2.2.3. Simulink model

The physical buoy, see diagram in Fig. 3, is modeled using a Simulink model; see Fig. 4 for a schematic figure, a complete model description with all equations is given in Appendix C, including the full Simulink diagram Fig. 12. In the following brief description, main blocks of the Simulink model are marked in *italics*.

The *buoy* consists of the buoy hull with a diameter of eight meters enclosing an open-ended tube in which the *water piston* floats. The relation between the masses of the buoy and the enclosed tube water is about 1:3 and buoy draught, due to the tube, is about thirty meters. The moorings are almost horizontal and soft, to minimize vertical forces that influence buoy movements.

The vertical stroke of the water piston is restricted to four meters by rubber *bumpers* fixed to the buoy hull, above and below the water piston; the bumpers also prevent damage in extreme sea states. There is a leaking channel between the periphery of water piston and the tube wall, which widens at the ends of the stroke. The pressure-drop across the water piston, due to the flow in the leaking channel, excites forces on the tube water. This couples tube water to the buoy hull, influencing

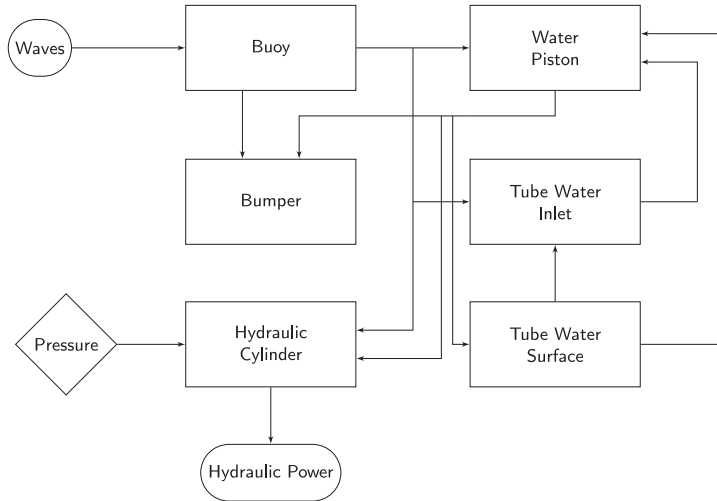


Fig. 4. A schematic figure of the Simulink model for the WEC buoy, illustrating key components of the model and how they interact. Rounded rectangles indicate input and output values for the model, diamonds fixed parameters, and rectangles the internal blocks of the model. The arrows indicate connections between components. The complete Simulink diagram can be found in Fig. 12 in Appendix C.

the resonance period of the buoy. The tube water is modeled by the *tube water inlet* and *tube water surface* blocks.

A *hydraulic cylinder* fixed to the buoy hull with its hydraulic piston fixed to the floating water piston connects the inertia of the tube water to the buoy hull. As a wave approaches, the buoy and hydraulic cylinder are lifted, while the hydraulic piston remains attached to the floating water piston. The hydraulic cylinder uses three rectifying valves to direct the resulting hydraulic flow from a low-pressure tank to an accumulator tank. The latter opens at 2 MPa and has a maximum pressure set to 25 MPa, by a relief-valve. If the accumulator pressure is larger than the pressure inside the hydraulic cylinder the water piston remains locked to the buoy hull, with no wave power absorbed. Otherwise, the movement of the water piston excites a braking force on the buoy hull, with amplitude given by the accumulator pressure. It seems like maximum wave power is absorbed when the water-piston is locked to the buoy hull during 40%–60% of the time.

Electrical power is produced by a hydraulic motor driving a generator, the motor is feed from the accumulator tank with flow that returns to the low-pressure tank. Two main control laws exist for the generator, either a constant speed is maintained for the hydraulic motor, allowing the pressure in the accumulator to fluctuate; or the controller varies the motor speed to maintain a roughly constant pressure in the accumulator. Regardless of the control law, the uptake of mechanical energy from the waves will depend on the pressure in the accumulator tank.

Since this study focuses on the uptake of mechanical energy, the Simulink model used here omits the electrical generator and uses a fixed pressure in the accumulator tank. The pressure is a model parameter and can be adjusted to investigate how differences in pressure and sea state interact to generate *hydraulic power*.

2.2.4. Simulation results

Our aim is to investigate which pressures in the accumulator tank provide optimal hydraulic power in different sea states and to characterize the variability in output; as such we need to run the Simulink model repeatedly. However, the model is computationally expensive to

run and provides large output data sets, resulting in both computational and storage challenges. Especially since one aim is to provide a method that could be run in a reasonable time on standard computers. After a few initial runs at the most representative sea states and considerations of reasonable pressures from a design and materials perspective, it was decided to run the model for 78 representative sea states in Table 1 of Appendix A, representing 92.4% of all occurrences and using 14 pressure values: 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, and 17 MPa. The accumulator tank relief-valve opens at 25 MPa, making this the highest possible pressure. However, in practice some margin is needed between this upper limit and the desired operational range. Initial exploratory runs indicated that 17 MPa constituted a reasonable upper limit for the operational range.

For each sea state and pressure combination, we generated 650 s of wave data at a 10 Hz sampling rate (Section 2.2.1) giving a time-series with 6500 samples. The 650 s duration of simulated waves is a trade-off between reducing variance and managing computational cost. Although longer simulation times would decrease variance, they would also significantly increase the computational resources needed. The waves were then prefiltered (Section 2.2.2) and used as input to the Simulink model (Section 2.2.3 and Appendix C) providing time series of generated power. Due to the adaptive solvers in Simulink, the output series is non-equidistantly sampled. The data output was realigned to the same 10 Hz sampling points as the waves using linear interpolation. We discarded a burn-in period consisting of the first 50 s (or 500 samples) to remove initial transients. The generated power was integrated over time, giving cumulative power during each period.

The non-linear behavior of the system causes significant variations in power output; see Fig. 5. To quantify the variability, we divided the ten-minute power series into five two-minute blocks, computing the average power for each block. The time span of two minutes gives five averages that are only weakly correlated. This is done separately for each of the fourteen pressures, giving 70 data points (5 for each pressure) for each sea-state, see Fig. 6(a).

The whole procedure was repeated for the 78 sea states. The waves are simulated independently for each pressure and each sea state.

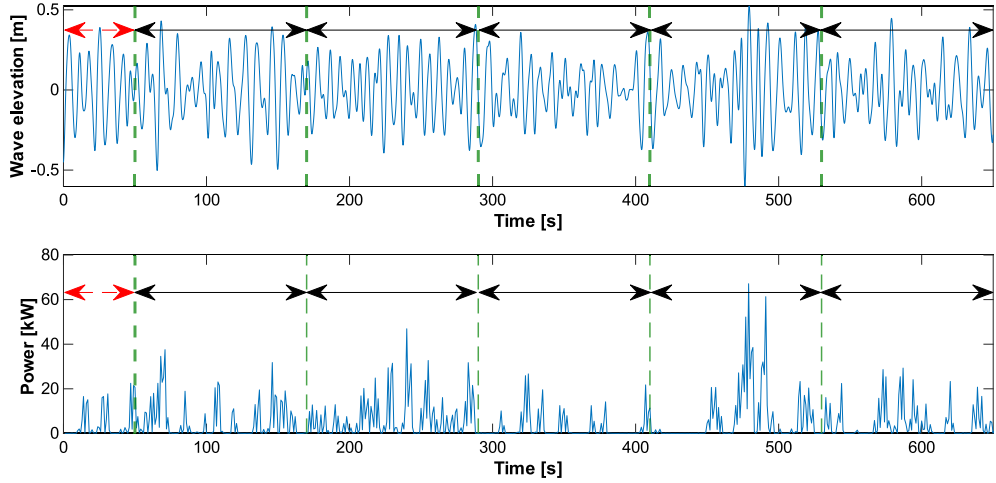


Fig. 5. Example of a simulated wave, top, and corresponding generated instantaneous hydraulic power from the Simulink model (the signal HydraulicPower [kW] in the Simulink diagram shown in Fig. 12), bottom, using a wave from the $H_s = 0.75$, $T_p = 7.5$ sea-state. The vertical dashed green lines indicate the data division into an initial 50 s burn-in (red dashed double arrow) that is discarded and the subsequent 2-minute averaging blocks (black solid double arrows).

2.3. Optimization method

Given the simulation results for all 78 sea states and 14 pressure levels, the maximum power and corresponding pressure are found using a concave regression. The regression of power against pressure is fitted using the MATLAB Shape Language Modeling toolbox (D'Errico, 2023). To account for the difference in variance at each pressure level, the observations are weighted by the inverse variances of the data at each pressure. The maximum power and optimal pressure are then extracted from the fitted curves, an example can be seen in Fig. 6(a).

The empirical relationship between optimal pressure and steepness is then estimated using non-linear regression with uncertainties obtained using semi-parametric bootstrap (see e.g. Givens and Hoeting, 2012, p. 290).

3. Results and discussion

Given the simulation results for all 78 wave states and 14 pressure levels, we now seek to determine optimal pressures for each wave state (Section 3.1) and how the optimal pressures depend on wave states (Section 3.2).

3.1. Optimal pressure for each sea state

Basing the maximum power and corresponding optimal pressure for each wave state on the raw data leads to noisy estimates due to the substantial variability among the 2-minute averages. Instead, the 70 data points, five observations of 2-minute average power at 14 different pressure levels, are modeled using a concave regression (see Section 2.3).

Using a fitted curve has the advantage of allowing maxima that lie between the 14 pressure values used in the simulation, see Fig. 6(a). In most cases, the maxima derived from the raw data and fitted curves are similar, but the fitted curve provides more stable estimates, being less sensitive to occasional large or small 2-minute averages.

In order to verify that we have captured the essential relationship between optimal pressure and optimal power, we re-simulated the model for each wave state at the optimal pressure and compared the

resulting power output to that predicted by the concave regression, see Fig. 6(b). The results show good agreement, indicating that we have satisfactorily captured the key features of the relationship.

3.2. Optimal pressure as a function of steepness

From the optimal pressure and observed variability, we now seek a relation between the optimal pressure and the sea states. An exploratory analysis of the 78 wave states, Figs. 7 and 6(b), showed that both optimal pressure and optimal power increase with increasing mean wave steepness, S_p , Eq. (6).

The relation between optimal pressure, p_{opt} , and mean wave steepness, S_p , was described empirically by the nonlinear model

$$p_{opt} = a + b(1 - \exp(-c \cdot S_p)), \quad (13)$$

with coefficients estimated using non-linear least squares as $a = -5.76 (\pm 0.84)$, $b = 18.11 (\pm 0.75)$, and $c = 138.51 (\pm 13.11)$, with standard deviations given in parenthesis.

To assess the uncertainty of the estimates, we use a semi-parametric bootstrap (Givens and Hoeting, 2012) where we re-sample the standardized residuals (according to the locally estimated standard deviations from the five observations at each pressure) from the original concave regression and create new synthetic data for which we again do concave regression, obtain optimal pressures and refit the non-linear model in Eq. (13). We repeat this 300 times for each of the 78 sea states and compute a bias-corrected version of the non-linear curve, the parameter uncertainties reported above, as well as upper and lower 2.5% quantiles for the curve, see Fig. 7. R^2 of this model is 0.76. Note that we here only assess the uncertainty of the fitted non-linear model. We are not trying to capture the measurement variability in the data points themselves.

To evaluate the empirical model we compared the optimal pressure for each sea state obtained from the concave regression (see Fig. 6(a)) with those predicted by the model in Eq. (13). This model also captures the most important features of the data; see Fig. 8. The correlation coefficient between the actual optimum pressure and the steepness-based optimum pressure is 0.92 with a p -value of $9e-35$.

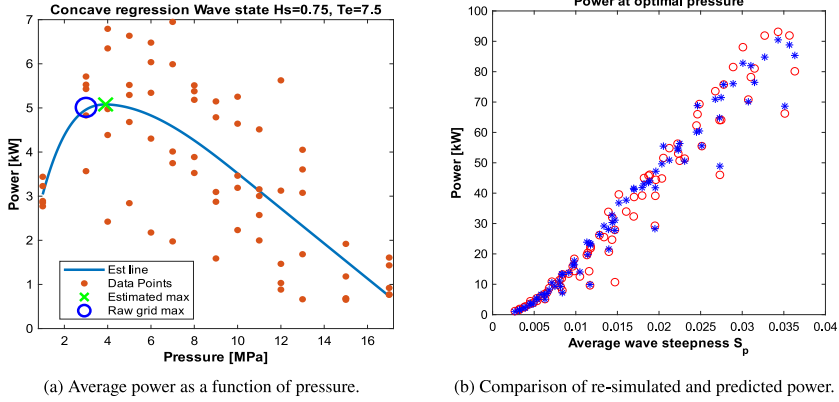


Fig. 6. The left pane shows the concave regression (blue line) fitted to the 70 observations of 2-minute average power (red dots) for the $H_s = 0.75$, $T_e = 7.5$ sea-state. The optimal power, based on either an average of the five observations at each pressure (blue circle) or the fitted concave function (green X), is also shown. The right pane shows re-simulated power at optimal pressures (red circles) for each sea state compared to the optimal power predicted by the concave regression (blue stars). The sea-states have been ordered by wave steepness according to Eq. (6).

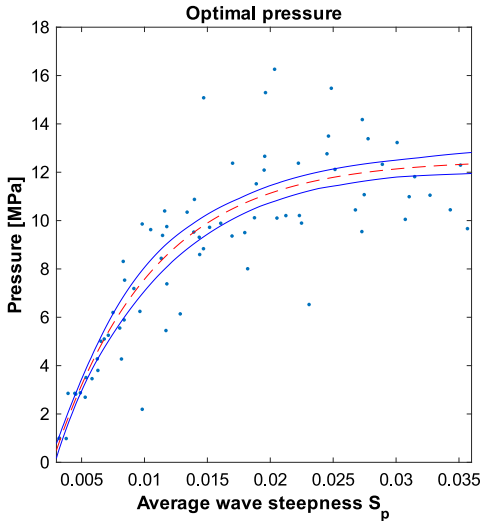


Fig. 7. Curve fitted for the relation between average wave steepness and optimal pressure according to Eq. (13). Both the fitted curve (red dashed line) and bootstrapped 2.5% quantiles for the fitted model (blue solid line) are shown.

3.3. Energy efficiency as a function of sea state

The theoretical energy efficiency of the WEC can now be computed by comparing the obtained maximum hydraulic power at each sea state with the energy content in the waves from Eq. (5), see Fig. 9. For this WEC configuration, we find very high efficiencies, close to 90%, for short, shallow waves, with the efficiency decreasing for longer energy periods, T_e , and large significant wave heights, H_s . The overall efficiency is 24%, when weighting the sea states according to their frequency of occurrence (Table 1 in Appendix A).

In a large review of energy efficiency for WECs (Aderinto and Li, 2019), previous studies of ten different OWC-type systems were compared (as well as several other types of WEC systems). Six of the ten studies reported efficiencies around 20–25%, with three studies reporting lower ($\leq 5\%$) and one reporting higher (40%) efficiencies (Aderinto and Li, 2019, Fig. 3a). Our overall efficiency is thus well in line with previous studies.

4. Conclusion

This study has illustrated the effectiveness of combining statistical modeling with a simulation-based analysis of the hydraulic PTO-system in a WEC. By dividing the simulations into shorter temporal blocks and computing the average generated power for each block, we can quantify both the average and variability in generated power. Using this data, the optimal pressure in the hydraulic accumulator tank that yields the highest power uptake was identified for a number of different sea states (cf. Appendix A). Comparing across sea states, we then identified wave steepness, Eq. (6), as the critical feature of the waves for explaining both optimal pressure (Fig. 7) and generated power (Fig. 6(b)). The overall efficiency of the WEC, at 24% when averaged across sea states, was found to be in line with previous studies (Aderinto and Li, 2019).

By identifying wave steepness as a crucial factor for modeling optimal pressure in hydraulic PTO-systems, this research contributes to a better insight into the interplay between sea state characteristics and hydraulic power generated by the WEC. The insights earned from this study can be used to improve the efficiency and performance of WEC systems. Significant wave height and energy period are used to design for the structural integrity of the system and to get an appropriate natural period of the buoy hull. However, for this WEC, it turns out that only considering wave steepness, instead of significant wave height and energy period, gives sufficient accuracy when designing certain parts of the PTO-system, reducing the design problem from two to one degree of freedom.

4.1. Further investigation

A natural extension of this study would be to use the obtained relationship between wave steepness and optimal pressure to design better control laws for the electrical generator in the buoy. Observations or

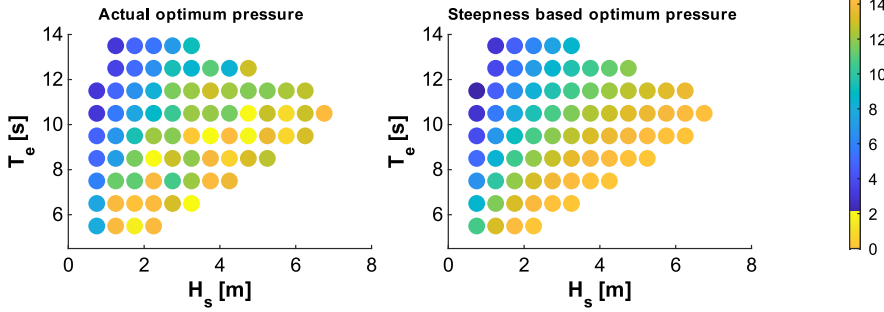


Fig. 8. Optimal pressure based on different concave regressions for each sea state (left) or obtained from the empirical relationship with steepness in Eq. (13) (right).

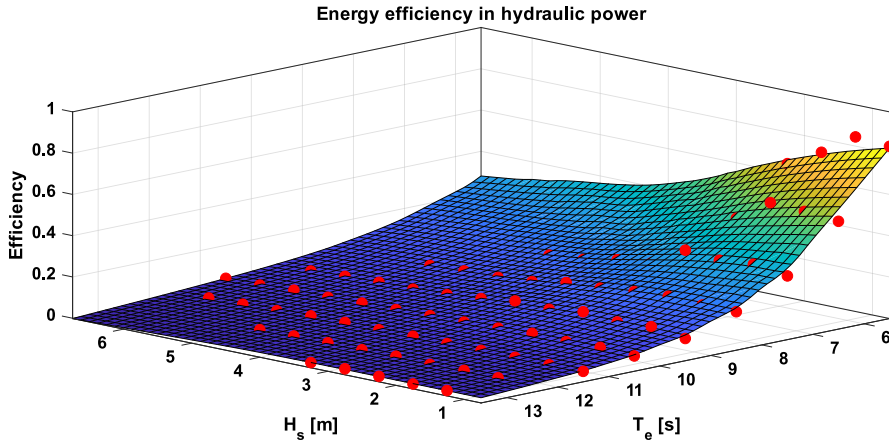


Fig. 9. Power efficiency as a function of sea state. This would imply an average efficiency of about 24% across all sea states when states are weighed according to occurrence.

short-term predictions of the local sea state could be used to compute expected wave steepness, allowing the control logic to better adapt to both current and future sea states; leading to better energy extraction.

CRedit authorship contribution statement

Shokoufa Zeinali: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Magnus Wiktorsson:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jan Forsberg:** Writing – original draft, Validation, Investigation, Formal analysis, Conceptualization. **Georg Lindgren:** Writing – original draft, Software, Formal analysis, Conceptualization. **Johan Lindström:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: As mentioned in the article, the buoy analyzed here is owned and operated by the company Waves4Power. The research was funded by

the Swedish Energy Agency with a grant awarded jointly to Lund University and the company. Jan Forsberg is the only author with a financial stake in the company.

Acknowledgments

We thank the editor and the referee for helpful comments that improved this study. This research was funded by grant 50197-1 from the Swedish Energy Agency and performed in collaboration with the company Waves4Power.

Appendix A. Wave state occurrences

For a detailed account of the wave states see Table 1.

Appendix B. Derivation of wave characteristics

B.1. Spectral moments

Using the standard (untruncated) generalized Pierson Moskowitz (GPM) or Bretschneider spectrum (ITTC, 2002, Appendix A), the spectral moments are defined as

$$m_n = \int_0^\infty \omega^n m_0 \frac{5}{\omega_p} \left(\frac{\omega}{\omega_p} \right)^{-5} e^{-\frac{5}{4} \left(\frac{\omega}{\omega_p} \right)^{-4}} d\omega.$$

Table 1

The frequency, in %, of the wave states used in this study. The wave statistics from the European Marine Energy Centre at Billia Croo, Orkney (European Marine Energy Centre, 2022) contain a total of 126 different wave states. For this study, we have considered a representative sub-sample of 78 states, covering 92.4% of all occurrences.

Hs/Te	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5
6.75						0.11			
6.25					0.11	0.23	0.16		
5.75					0.24	0.36	0.16		
5.25				0.17	0.43	0.41	0.24		
4.75				0.35	0.72	0.58	0.30	0.14	
4.25			0.14	0.71	0.95	0.68	0.37	0.17	
3.75			0.43	1.27	1.24	0.80	0.42	0.20	
3.25		0.13	0.96	1.73	1.67	1.10	0.61	0.27	0.12
2.75		0.52	1.75	2.35	1.96	1.21	0.62	0.30	0.14
2.25	0.16	1.43	2.62	3.10	2.57	1.43	0.76	0.36	0.15
1.75	0.81	2.48	3.18	3.27	2.43	1.53	0.81	0.37	0.15
1.25	2.18	3.97	4.71	4.06	2.65	1.43	0.61	0.24	0.10
0.75	3.16	4.65	4.49	2.77	1.40	0.64	0.23		

With the change of variable

$$x = \frac{5}{4} \left(\frac{\omega_p}{\omega} \right)^4$$

we get the integral

$$m_n = \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \int_0^\infty x^{(1-n/4)-1} e^{-x} dx \quad (14)$$

$$= \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \Gamma \left(1 - \frac{n}{4} \right),$$

with $\Gamma(a)$ being the standard gamma function (see e.g. Eq. 6.1.1 on p. 255 in Abramowitz and Stegun, 1965). The spectral moments are finite only for $n < 4$. If we instead use the truncated spectrum in Eq. (1) we get

$$m_n^{\text{trunc}} = \int_0^{5\omega_p} \omega^n m_0 \frac{5}{\omega_p} \left(\frac{\omega}{\omega_p} \right)^{-5} e^{-\frac{5}{4} \left(\frac{\omega}{\omega_p} \right)^4} d\omega.$$

With the same change of variables as before, yielding

$$m_n^{\text{trunc}} = \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \int_{\frac{1}{500}}^\infty x^{(1-n/4)-1} e^{-x} dx \quad (15)$$

$$= \omega_p^n m_0 \left(\frac{5}{4} \right)^{n/4} \Gamma \left(1 - \frac{n}{4}, \frac{1}{500} \right),$$

with $\Gamma(a, b)$ being the upper incomplete gamma function (see e.g. Eq. 6.5.3 on p. 260 in Abramowitz and Stegun, 1965). The spectral moments are finite for all n for the truncated GPM spectrum.

B.2. Wave steepness

To show the equality in Eq. (7) we use the definition of $T_z = 2\pi \sqrt{\frac{m_0}{m_2}}$ which, for the truncated spectrum, gives

$$S_s = \frac{2\pi H_s}{g T_z^2} = \frac{H_s}{2\pi g} \frac{m_2^{\text{trunc}}}{m_0^{\text{trunc}}}.$$

Using the formula for truncated spectral moments in Eq. (15) and the equality $T_p = \frac{2\pi}{\omega_p}$ now gives

$$S_s = \frac{H_s}{2\pi g} \frac{\omega_p^2 m_0 \sqrt{\frac{5}{4}} \Gamma \left(\frac{1}{2}, \frac{1}{500} \right)}{m_0 \Gamma \left(1, \frac{1}{500} \right)}$$

$$= \frac{H_s}{2\pi g} \left(\frac{2\pi}{T_p} \right)^2 \frac{\sqrt{\frac{5}{4}} \Gamma \left(\frac{1}{2}, \frac{1}{500} \right)}{e^{-\frac{1}{500}}} \approx 1.89 S_p.$$

The corresponding result for the standard (untruncated) spectrum, using the spectral moments in Eq. (14), is

$$S_s = \frac{2\pi H_s}{g T_p^2} \sqrt{\frac{5}{4}} \Gamma \left(\frac{1}{2} \right) \approx 1.98 S_p.$$

Table 2

Parameters of the Simulink model, their descriptions and units.

Symbol	Description	Unit
m_1	buoy mass	kg
μ	coefficient added mass	
ω	angular frequency	$\frac{rad}{s}$
V_1	displaced volume	m^3
L_1	draught of the buoy hull	m
ϵ	coefficient added damping	
ρ	density of water	$\frac{kg}{m^3}$
g	gravity constant	$\frac{m}{s^2}$

Appendix C. Simulink model

This section contains a detailed description of the Simulink model along with the corresponding block-diagram Fig. 12, Table 4 lists all signals in the model, and Table 2 summarizes some additional notation used in the equations.

C.1. Buoy equations

C.1.1. Heave motion

The motion of the buoy depends on mass forces and boundary forces. Gravity excites mass forces, and the waves work on the boundaries of the buoy-tube. The linear Froude-Krylov wave force F_{FK} is used to model the acting forces of the waves. The positive direction of vectors is in the direction opposite to gravity. The dynamics of the heave motion of the free-floating buoy are expressed as

$$(m_1 + \mu(\omega)\rho V_1) \ddot{z}_1 + \epsilon(\omega)\rho V_1 \dot{z}_1 + \rho g V_1 (L_1 - z_1 + z_2) - m_1 g =$$

$$= \begin{cases} 0, & \text{in calm water with} \\ F_{FK} + F_{31} + F_{b1}, & \text{sea surface elevation } z = 0, \\ & \text{in waves.} \end{cases}$$

Here F_{31} and F_{b1} represent the acting forces of the PTO system, from the hydraulic cylinder and bumpers respectively. Sub-index 1 is connected to the buoy hull, 2 is used with tube water and 3 is connected to the floating water piston and the hydraulic cylinder inside the tube. Index b is connected to rubber bumpers at end-stops, absorbing kinetic energy when the water piston hits the ends of its stroke. The angular frequency ω is considered constant and set to the natural frequency, ω_0 , of the buoy. Values of μ and ϵ are based on estimated from early model tests.

The Froude-Krylov force, F_{FK} , is approximated by

$$F_{FK} = \mu(\omega)\rho V_1 \ddot{z} + \epsilon(\omega)\rho V_1 \dot{z}.$$

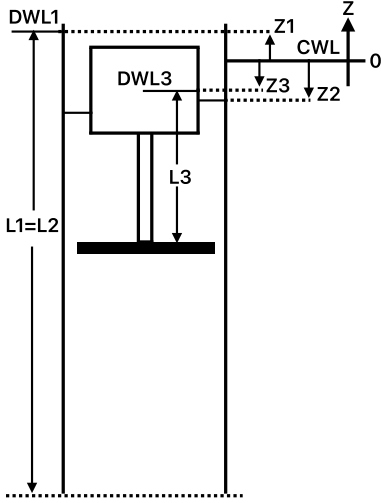


Fig. 10. Schematic diagram of the acceleration tube, with calm water line (CWL) as position reference for the positions of: sea-level z , buoy hull z_1 , tube-water z_2 , and water piston z_3 . For the buoy used in this paper the water tube has length $L_1 = L_2 = 30.7$ meters and the water piston is $L_3 = 6.4$ meters.

C.1.2. Dynamics of tube water inlet

Conservation laws of mass and vertical momentum are used to model the dynamics of the tube water in heave motion. The water piston is floating in the tube water and will excite internal boundary forces on the tube water. Fig. 10 illustrates the positions of the different components relative to the calm water line (CWL): z_1 is the buoy design water line (DWL1), z_2 is the surface of tube-water, and z_3 is the design water line of the water piston (DWL3).

The dynamics of the vertical momentum of tube-water is estimated by the relation:

$$\frac{dM_2}{dt} = \dot{M}_{2i} - \dot{M}_{2o} + \sum F_{2i}.$$

The tube water volume V_2 and displaced water piston volume V_3 depend on z_2 and z_1 , respectively z_3 and z_2 . The vertical momentum of the tube water is set to

$$M_2(t) = \rho V_2 (L_2 + z_2 - z_1) \dot{z}_2 - \rho V_3 (L_3 + z_2 - z_3) \dot{z}_3$$

The momentum flux $\dot{M}_{2i}$ at the lower end (inlet) and $\dot{M}_{2o}$ at tube-water surface (outlet) are

$$\dot{M}_{2i} = \rho A_2 \lambda (z_{2i} - z_1) \dot{z}_{2i}, \dot{M}_{2o} = 0.$$

The dynamics of the tube water are excited by mass and boundary forces; mass force from gravity, boundary force from static pressure at the lower end inlet, viscous losses at the lower end inlet, and the reactive force from the pressure drop across the internal boundary of the water-piston.

$$F_{2g} = -\rho g V_2 (L_2 + z_2 - z_1)$$

$$F_{2pi} = \rho g A_2 \lambda (L_2 - z_1)$$

$$F_{2iloss} = -A_2 \zeta_{iloss} \frac{1}{2} \rho |\dot{z}_{2i}| \dot{z}_{2i}$$

$$F_{23} = -A_3 \zeta_{dp} \frac{1}{2} \rho |w - \dot{z}_3| (w - \dot{z}_3)$$

$$w = \frac{A_2}{A_{2a} - A_3} (\dot{z}_{2i} - \dot{z}_3)$$

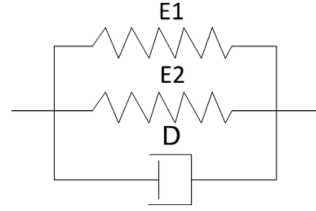


Fig. 11. Modified linear viscoelastic Kelvin-Voigt model.

Here the parameters ζ_{iloss} and ζ_{dp} are estimated and treated as constant. The area A_{2a} represents a variable tube area within the stroke of the hydraulic cylinder and water piston. The area A_{2a} increases close to both end-stops. The velocity w corresponds to the average velocity in the leaking channel between the water piston and tube-wall.

C.1.3. Rubber bumper model

The survival of the water piston system depends on how forces are handled at the two end-stops of the piston stroke. Rubber bumpers are used in the actual buoy-water piston design. The bumper is designed to absorb fast and heavy strokes. The initial impact generates a great load on the support of the bumper, the higher the water piston velocity the greater the load.

Rubber is highly non-linear, and a rigorous analysis needs Finite Element Method (FEM) methods. However, there are lumped models that may be useful in practice. A modified Kelvin-Voigt model is presented in Fig. 11, to include shear stress and viscoelasticity in transient motions of the bumper. Part of the internal shear work in the rubber material is stored as elastic energy E2 and the rest is lost into heat in the dashpot (D). The upper branch represents the stored elastic energy or stress in axial deformation. The shear velocity has components in both the x - or r -direction. Shear velocities in the material are larger than those in the axial x -direction. This is compensated for in Eq. (16) where the rubber is considered incompressible.

$$\dot{E}_{sh} \approx \dot{E} \sqrt{\frac{4V_0}{\pi(x_0 + \epsilon)^3} + 1} \quad (16)$$

The dynamic equation to be solved for one of N bumpers is:

$$\ddot{E} = \frac{1}{m} (F_e(\epsilon) + F_{sh}(\epsilon) + F_D(\dot{\epsilon}))$$

Here m is the mass of the water piston, while V_0 and x_0 represent the volume and length of the unloaded bumper. We need initial values of axial strain $\epsilon(0)$ and strain velocity $\dot{\epsilon}(0)$ to solve the dynamic equation. The energies $E1$ and $E2$ are estimated from static analysis of the rubber. Relations are used in a simplified uniaxial load case, like ours, as the rubber is free to expand radially. Let the axial compression be x and the unloaded length of the cylindrical bumper be x_0 . The extension ratio λ is defined as:

$$\lambda = \frac{x_0 - x}{x_0} = 1 - \frac{x}{x_0}.$$

The elastic free energy of the rubber may be written

$$4A_e = \frac{v k T}{2} \left(\lambda^2 + \frac{2}{\lambda} - 3 \right) \quad (17)$$

Here v is the number of polymer chains, k is Boltzmann's constant and T is the temperature in Kelvins (of the seawater). The equation is usable in the case of uniaxial deformation of incompressible rubber. The incompressibility leads to radial expansion of the cylindrical bumper, and we will add the elastic energy from the shear

$$4A_{sh} = \frac{v k T}{6} \left(2\lambda + \frac{1}{\lambda^2} - 3 \right) \quad (18)$$

Forces excited by very slow (almost static) changes in deformation or extension ratio may be expressed by means of the change in elastic free energy ΔA , when the bumper length $x_0 - x$ changes, $0 < x \leq x_0$ for compression

$$F = \frac{\partial \Delta A}{\partial (x_0 - x)} = \frac{\partial \Delta A}{\partial \lambda} \frac{\partial \lambda}{\partial (x_0 - x)} \quad (19)$$

The force components appear when we apply Eq. (19) to the energies of Eq. (17) and Eq. (18) as one elastic axial force F_e and one elastic shear force F_{sh}

$$\begin{aligned} F_e &= \frac{\nu k T}{x_0} \left(\lambda - \frac{1}{\lambda^2} \right) \\ F_{sh} &= \frac{\nu k T}{3x_0} \left(1 - \frac{1}{\lambda^3} \right) \\ F &= F_e + F_{sh} \end{aligned} \quad (20)$$

The elastic shear is due to the incompressibility of rubber, and the bumper expands radially when compressed. Given material choices a reasonable estimate of the maximal stored elastic energy in the bumper is $\nu k T = 4300$ J.

Eq. (20) above uses λ as an independent variable, which must be transformed to ϵ for which we obtain

$$\lambda = 1 + \frac{\epsilon}{x_0}$$

The dashpot coefficient D is estimated by iterated simulations with known test examples using the bumper datasheets from the manufacturer. Bumpers are activated when

$$|z_3 - z_1| \geq \frac{L_c}{2} - x_0$$

where L_c is the stroke of the hydraulic piston.

C.1.4. Simulink model for the oil-hydraulic piston

The hydraulic cylinder has two oil sub-volumes: upper (with low-pressure inlet) and lower (with high-pressure outlet). The volumes are sometimes connected and approximately at the same pressure, and sometimes separated and at different pressures. The pressure drop across the valve in the piston that connects the two sub-volumes determines which state is activated. Non-linear elements, like valves, cause numerical problems when changing state. Their opening and closing will easily cause numerical oscillations. Compressibility of the hydraulic fluid introduces dead zones that may damp out or reduce such numerical oscillations.

The larger section area of the piston is set to A in the upper sub-volume V_1 , while the piston rod section area is set to a . This means that the piston section area of the lower sub-volume V_2 will equal the difference $A - a$, where $A = 2a$. After a vertical move dz we get volume changes:

$$\begin{aligned} dV_1 &= -Adz \\ dV_2 &= (A - a)dz \\ dV &= dV_1 + dV_2 = -adz \end{aligned}$$

The hydraulic cylinder has two external pressure references p_{suc} and p_{acc} , from the suction tank and the accumulators respectively. Both pressure references couple to the hydraulic cylinder through rectifying valves. Consider the state of a pressure-controlled valve, then

$$\begin{aligned} p_1 < p_{suc} &\rightarrow \text{suction valve open(s)} \\ p_{acc} < p_2 &\rightarrow \text{outlet valve open(s)} \\ p_2 < p_1 &\rightarrow \text{internal piston valve open(s)} \end{aligned} \quad (21)$$

This means that we omit the inertia of oil and moving valve components. If conditions in Eq. (21) are not met, valves close immediately or will stay closed. The non-linear conditions of Eq. (21) means that the state of the valves must be known at the beginning of each time step in the simulation.

Table 3

Different combinations of the three valves being open (1) or closed (0) and the corresponding cases.

Valve \ Case	I	II	III	IV	V	VI	VII	VIII
Outlet	0	0	0	1	1	1	1	1
Piston	0	0	1	0	0	1	1	1
Inlet	0	1	0	0	1	0	1	1

Suppose that the oil has bulk modulus K . The oil pressure p will change (dp) if enclosed oil (connected valves closed) of volume V changes its volume by dV :

$$dp = -\frac{K}{V} dV$$

We will use the conservation law on the energy of the oil inside the hydraulic cylinder, to establish a relation between different states and piston forces. A general relation between added heat dQ , work dW done by the oil, changes dE in the energy of the oil, and work done on open boundaries (the valves) will be used. The specific energy e , energy per unit mass of the oil, includes specific enthalpy, kinetic, and potential energy.

$$e = u + pv + \frac{w^2}{2} + gz, \left[\frac{J}{kg} \right] \text{ or } \left[\frac{m^2}{s^2} \right] \quad (22)$$

The order of the last two terms, the kinetic and potential energy respectively, is a couple of tens smaller than that of the two first. The two smaller terms are omitted from Eq. (22). The remaining terms u and pv represent inner (thermal) energy and potential energy of pressure respectively. Temperature changes of the oil are supposed to be small, and the specific inner energy u will be almost constant or vary very slowly, while changes in pressure may be both large and rapid. The temperature and pressure of in-flowing oil is supposed to be the same as that of oil inside the hydraulic cylinder. Reynolds transport theorem for energy gives us the following general relation

$$\dot{Q} - \dot{W} = \frac{\partial}{\partial t} \int_{V_0} \rho e dV + \int_{S_0} \rho e (\vec{U} \cdot d\vec{S}) \quad (23)$$

The equation handles both closed and open systems. To the left in Eq. (23) the first term is external thermal power or heat added to the oil system and the second term is mechanical power delivered by the oil to moving boundaries (surfaces of the piston). To the right the first term represents the explicit time derivative of the oil energy in the hydraulic cylinder and the last term represents energy flux through oil flow at inlet and outlet boundaries and outlet valves, if open. The right-hand side of Eq. (23) equals zero when the state of the oil is constant (T, p constant) without moving boundaries. If we look at an infinitesimal change dV of fluid volume in the open subsystem we may rewrite Eq. (23) as

$$-dW + pdV = 0 \Rightarrow -dW = -pdV \quad (24)$$

The first term in Eq. (24) represents work done on the piston by the oil and the second is external oil flow acting on the boundaries during in- or outflow. If the three valves work independently there are $2^3 = 8$ different configurations of the valves in the hydraulic cylinder. The upper sub-volume V_1 may represent a closed or an open sub-system independently of V_2 if the valve on the piston is closed. The same is true for the lower sub-volume V_2 . When the valve on the piston is open the states of oil in V_1 and V_2 are considered identical and the hydraulic cylinder is represented by one volume, closed or open. All combinations of the states of valves do not have equal probability. The last case (VIII) in Table 3 has low probability but may occur if kinetic energy is left in the rotating generator, the accumulators are empty, and the hydraulic cylinder cannot feed the hydraulic motor with enough oil. The most important cases of Table 3 are IV–VII.

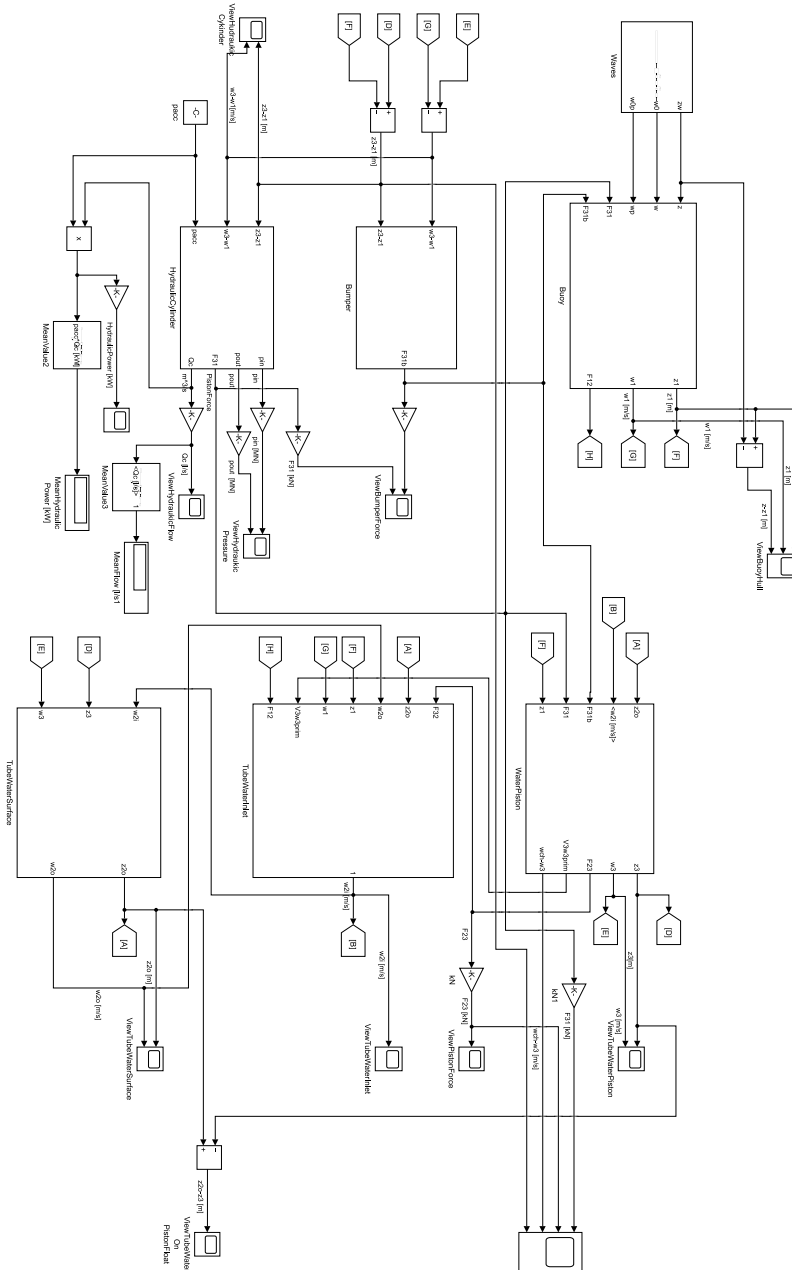


Fig. 12. The complete Simulink diagram, descriptions of all signals can be found in Table 4. Model description along with a simplified schematic diagram are provided in Section 2.2.3 and Fig. 4.

Table 4

Description of all signals in the Simulink model in Fig. 12. Model description along with a simplified schematic diagram are provided in Section 2.2.3 and Fig. 4. The table gives signal name, brief description and which blocks the signal connects. For cases where the signal has different names between blocks the new input name is listed in parenthesis in the To-column. Signals with empty From-column are either model parameters to set or combinations of multiple inputs; empty To-column indicate Simulink outputs for model evaluation.

Name	Description	From	To
pacc	Assumed constant pressure in the accumulator tank which is applied as an external pressure to the valves of the hydraulic cylinder. This is the parameter to be optimized (Section 2.2.4).		HydraulicCylinder
z0	Vertical position of generated (Section 2.2.1) or measured sea surface.	WAFO	Waves
zw	Vertical position of sea surface after prefiltering (Section 2.2.2).	Waves	Buoy (z)
w0	Vertical velocity of the sea surface, derivative of zw.	Waves	Buoy (w)
w0p	Vertical acceleration of the sea surface, derivative of w0.	Waves	Buoy (wp)
z1	Vertical position of the design waterline (DWL) of the buoy relative the reference level of calm water line (CWL).	Buoy	TubeWaterInlet
w1	Vertical velocity of the buoy-hull, derivative of z1.	Buoy	WaterPiston
F12	Hydrodynamic force between buoy hull and tube water.	Buoy	TubeWaterInlet
F31b	Force impulse from the water-piston, hitting the bumpers and acting on the buoy hull. This force travels through the bumper rubber material and hits the bumper support on the buoy hull.	Bumper	Buoy WaterPiston
w2i	Flow velocity in the "inlet" section at the lower end of the acceleration tube.	TubeWaterInlet	TubeWaterSurface WaterPiston
z2o	Position of tube-water surface relative CWL.	TubeWaterSurface	TubeWaterInlet
w2o	Velocity of the surface of the tube-water, derivative of z2o.	TubeWaterSurface	WaterPiston TubeWaterInlet
z3	Position of DWL of the water piston, relative the CWL.	WaterPiston	TubeWaterSurface
w3	Velocity of the water-piston, derivative of z3.	WaterPiston	TubeWaterSurface
wch-w3	Difference between flow velocity in the tapered channel, between water-piston and tube wall, and the velocity of water-piston.	WaterPiston	
V3w3prim	Accelerating, displaced water-piston volume [m ³ /s ³].	WaterPiston	TubeWaterInlet
F23	Hydrodynamic force between water-piston and tube water.	WaterPiston	TubeWaterInlet (F32)
F31	Force of the hydraulic piston inside the hydraulic cylinder, acting on the buoy-hull.	HydraulicCylinder	Buoy
pin	Internal pressure of the oil volume above the hydraulic piston, inside the hydraulic cylinder, connected to the oil inlet valve.	HydraulicCylinder	WaterPiston
pout	Internal pressure of the oil volume below the hydraulic piston, inside the hydraulic cylinder, connected to the oil outlet valve and external accumulator or pressure	HydraulicCylinder	
Qc	Flow through the outlet valve of the hydraulic cylinder that feeds the energy converting system. Generated power is given by $Qc \cdot pacc$	HydraulicCylinder	
z3-z1	Relative position between water-piston and buoy-hull, i.e. where the water-piston is in the working area of the acceleration tube. This also represents the position of the hydraulic piston inside the hydraulic cylinder		Bumper HydraulicCylinder
w3-w1	Relative velocity between water-piston and buoy-hull, same as the relative velocity between the hydraulic piston and cylinder.		Bumper HydraulicCylinder

Energy balance in the piston material gives the relation between work dW added to the piston from waves and work added from oil, dW_1 and dW_2 at a change dz in the position of the piston.

$$dW_w + dW_1 + dW_2 = 0 \Rightarrow F_w dz = -dW_1 - dW_2 \quad (25)$$

F_w represents wave forces on the piston. We ignore elastic energy stored in the piston material, in this case. Each state of valves in Table 3 gives a unique relation of Eq. (25). A valve may close immediately, as mentioned earlier. Some states of the table may jump directly to other states according to conditions in Eq. (21) and depending on the sign of the vertical piston displacement dz . Changes dW_1 and dW_2 are computed using Eq. (24) at the displacement dz . The force of the hydraulic cylinder F_{31} acting on the buoy, is

$$F_{31} = a(p_1 - 2p_2)$$

or

$$F_{31} = \begin{cases} ap_{acc} & \text{if } z > 0 \\ a(-p_{acc} + 2p_{suc}) & \text{if } z < 0 \end{cases}$$

depending on which valves are open.

C.1.5. Equations of the water piston, used in the simulink

The water-piston is floating in the tube water and excited by displacement and the previously described forces F_{31} from the hydraulic cylinder and F_{31b} from rubber bumpers. Added to these forces is the force due to the pressure drop across the submerged water-piston, F_{23} , which depends on the difference in speed between tube-water and water-piston.

The dynamics of the water is described as

$$m_3 \ddot{z}_3 + \rho g V_3 (L_3 + z_{20} - z_3) - m_3 g = -F_{31} - F_{31b} + F_{23} \quad (26)$$

The tube water passes the water piston in the thin channel at the tube wall. The difference between the fluid velocity w in the channel and the velocity of the water piston is

$$w - w_3 = \frac{A_2 w_{2i} - A_{2a} w_3}{A_{2a} - A_3} \quad (27)$$

where A_2 , A_{2a} , and A_3 are section areas of the tube at the lower end, of the variable tube area at the actual position of the water-piston and of the latter, respectively. This pressure drop excites a force F_{23} on the

water-piston

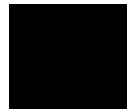
$$F_{23} = \frac{1}{2} A_3 \rho \zeta_c |w - w_3| (w - w_3) \quad (28)$$

where ζ_c is a loss-coefficient.

References

- Abramowitz, M., Stegun, I., 1965. *Handbook of Mathematical Functions*. Dover Publications, New York.
- Aderinto, T., Li, H., 2019. Review on power performance and efficiency of wave energy converters. *Energies* 12, <http://dx.doi.org/10.3390/en1224329>, URL <https://www.mdpi.com/1996-1073/12/22/4329>.
- Ahamed, R., McKee, K., Howard, I., 2020. Advancements of wave energy converters based on power take off (pto) systems: A review. *Ocean Eng.* 204, 107248. <http://dx.doi.org/10.1016/j.oceaneng.2020.107248>.
- Ahmed, A., Wang, Y., Azam, A., Zhang, Z., 2022. Design and analysis of the bulbous-bottomed oscillating resonant buoys for an optimal point absorber wave energy converter. *Ocean Eng.* 263, 112443.
- Amini, E., Mehdipour, H., Faraggiana, E., Golbah, D., Mozaffari, S., Bracco, G., Neshat, M., 2022. Optimization of hydraulic power take-off system settings for point absorber wave energy converter. *Renew. Energy* 194, 938–954. <http://dx.doi.org/10.1016/j.renene.2022.05.164>, URL <https://www.sciencedirect.com/science/article/pii/S09560148122008242>.
- Azam, A., Ahmed, A., Li, H., Tairab, A.M., Jia, C., Li, N., Zhang, Z., 2022. Design and analysis of the optimal spinning top-shaped buoy for wave energy harvesting in low energy density seas for sustainable marine aquaculture. *Ocean Eng.* 255, 111434. <http://dx.doi.org/10.1016/j.oceaneng.2022.111434>, URL <https://www.sciencedirect.com/science/article/pii/S0029801822008150>.
- Babari, A., Guglielmi, M., Clément, A.H., 2009. Decutching control of a wave energy converter. *Ocean Eng.* 36, 1015–1024. <http://dx.doi.org/10.1016/j.oceaneng.2009.05.006>.
- Brodtkorb, P., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E., 2000. Wafo - A matlab toolbox for analysis of random waves and loads. In: *Proceedings of the 10th International Offshore and Polar Engineering Conference*, pp. 343–350.
- DÉrrico, J., 2023. SIm — shape language modeling. URL <https://www.mathworks.com/matlabcentral/fileexchange/24443-sim-shape-language-modeling>. version 1.14. Retrieved from the MATLAB Central File Exchange.
- European Marine Energy Centre, 2022. Billia Croo wave data base.
- Falcão, A.F.d.O., 2007. Modelling and control of oscillating-body wave energy converters with hydraulic power take-off and gas accumulator. *Ocean Eng.* 34, 2021–2032.
- Falcão, A., Henriques, J., 2016. Oscillating-water-column wave energy converters and air turbines: A review. *Renew. Energy* 85, 1391–1424.
- Falnes, J., Kurniawan, A., 2020. *Ocean Waves and Oscillating Systems: Linear Interactions Including Wave-Energy Extraction*, vol. 8, Cambridge University Press.
- García-Rosa, P.B., Cunha, J.P., Lizarralde, F., Estefen, S.F., Costa, P.R., 2009. Efficiency optimization in a wave energy hyperbaric converter. In: *2009 International Conference on Clean Electrical Power*. IEEE, pp. 68–75.
- Givens, G.H., Hoeting, J.A., 2012. *Computational Statistics*, vol. 703, John Wiley & Sons.
- Gradshteyn, I., Ryzhik, I., Zwillinger, D., Moll, V., 2015. *Table of Integrals, Series, and Products*, eighth ed. Academic Press, Amsterdam.
- Gustafsson, E., 2016. *Extreme Loading and Fatigue Analysis of a Wave Energy Device*.
- Huang, Y.J., Liu, Y.J., Xue, G., Xue, Y.F., 2022. Hydrodynamic response and power performance of a heave and pitch buoy wave energy converter under bimodal ochi-hubble wave spectrum. *China Ocean Eng.* 36, 28–37. <http://dx.doi.org/10.1007/s13344-022-0002-2>.
- Hussain, A., Arif, S.M., Aslam, M., 2017. Emerging renewable and sustainable energy technologies: State of the art. *Renew. Sustain. Energy Rev.* 71, 12–28. <http://dx.doi.org/10.1016/j.rser.2016.12.033>.
- ITTC, 2002. The specialists committee on waves, final report and recommendations to the 23rd ITTC. In: *Proceedings of the 23rd ITTC*. pp. 505–736.
- Johannesson, P., Lang, X., Johnsson, E., Ringsberg, J., 2022. Mechanical reliability analysis of flexible power cables for marine energy. *J. Mar. Sci. Eng.* 10, 716. <http://dx.doi.org/10.3390/jmse10060716>.
- Jusoh, M., Ibrahim, M., Daud, M., Yusop, Z., Albani, A., Rahman, S., Mohad, S., 2020. Parameters estimation of hydraulic power take-off system for wave energy conversion system using genetic algorithm. In: *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, p. 012129.
- Kasacik, D., Stratigaki, V., Wu, M., Cappiotti, L., Simonetti, L., Troch, P., Crespo, A., Altomare, C., Dominguez, J., Hall, M., Gómez-Gesteira, M., Canelas, R.B., Stansby, P., 2020. Efficiency and survivability of a floating oscillating water columns wave energy converter moored to the seabed: An overview of the EsfOWC MaRINET2 database. *Water* 12, 1–21.
- Kim, S.J., Koo, W., Kim, M.H., 2021. The effect of geometrical buoy shape with nonlinear Froude-Krylov force on a heaving buoy point absorber. *Int. J. Nav. Arch. Ocean* 13, 86–101.
- Lindgren, G., 2015. Asymmetric waves in wave energy systems analysed by the stochastic Gauss-Lagrange model. *Proc. Est. Acad. Sci.* 64, 291–296.
- Liu, Z., Liu, S., Chen, W., Yang, Y., Feng, G., 2024. Implementation and optimization of hydraulic wave energy generation system. *PLoS One* 19, e0293209.
- Mackay, E., 2012. Resource assessment for wave energy. In: Sayigh, A. (Ed.), *In: Comprehensive Renewable Energy*, vol. 8, Elsevier, Oxford, pp. 11–72.
- Myrhaug, D., 2018. Some probabilistic properties of deep water wave steepness. *Oceanologia* 60, 187–192. <http://dx.doi.org/10.1016/j.oceano.2017.10.003>.
- Neshat, M., Nezhad, M., Sergienko, N., Mirjalili, S., Piras, G., Garcia, D., 2022. Wave power forecasting using an effective decomposition-based convolutional Bi-directional model with equilibrium Nelder-Mead optimiser. *Energy* 256, 124623. <http://dx.doi.org/10.1016/j.energy.2022.124623>.
- Penalba, M., Ringwood, J., 2016. A review of wave-to-wire models for wave energy converters. *Energies* 9, 1–45. <http://dx.doi.org/10.3390/en9070506>.
- Rosati, M., Henriques, J., Ringwood, J., 2022. Oscillating water column wave energy converters: A critical review of numerical modelling and control. *Energy Conversion Manag.* X 16, 100322.
- Sautier, J.B., Clément, A., de Falcão, A., Pontes, T., Prevosto, M., Ricci, P., 2011. Wave groupiness and spectral bandwidth as relevant parameters for the performance assessment of wave energy converters. *Ocean Eng.* 38, 130–147.
- Shao, X., Yao, H.D., Ringsberg, J., Li, Z., Johnson, E., Fredrikson, G., 2023. A comparison of the performance and characteristics of two generations waves4power wave energy converters. In: *Proceedings of the 5th International Conference on Renewable Energies Offshore*. RENEW 2022, pp. 277–284.
- Toffoli, A., Lefèvre, J., Bitner-Gregersen, E., Monbaliu, J., 2005. Towards the identification of warning criteria: Analysis of a ship accident database. *Appl. Ocean Res.* 27, 281–291. <http://dx.doi.org/10.1016/j.apor.2006.03.003>.
- Van der Meer, J., 1999. Design of concrete armour layers. In: *Proceedings of the International Conference on Coastal Structures*. pp. 213–221.
- WAFO-group, 2017. *Wafo – A Matlab toolbox for analysis of random waves and loads*. <https://github.com/wafo-project>.
- Wang, Y., 2018. A novel simulation method for predicting power output of wave energy converters. *Appl. Ocean Res.* 80, 37–48.
- Wang, Y., 2019. Comparison of a Lagrangian and a Gaussian model for power output predictions in a random sea. *Renew. Energy* 134, 426–435.
- Wang, Y., 2020. Predicting absorbed power of a wave energy converter in a nonlinear mixed sea. *Renew. Energy* 153, 362–374.
- Wang, Y., 2021. A second order random wave model for predicting the power performances of a wave energy converter. *Acta Oceanol. Sin.* 40, 127–135.
- Wang, Y., Wang, L., 2018. Towards realistically predicting the power outputs of wave energy converters: Nonlinear simulation. *Energy* 144, 120–128.
- Xue, G., Zhang, Z., Qin, J., Huang, S., Liu, Y., 2023. Control parameters optimization of accumulator in hydraulic power take-off system for eccentric rotating wave energy converter. *J. Mar. Sci. Eng.* 11, 792.
- Zurkinden, A.S., Ferri, F., Kofoed, J.P., Kramer, M.M., 2014. Nonlinear numerical modeling and experimental testing of a point absorber wave energy converter. *Ocean Eng.* 78, 11–21.

Paper II



The effects of sea-state on optimal generator parameters for a wave energy converter

S. Zeinali, M. Wiktorsson, J. Lindström & G. Lindgren

Division of Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden

J. Forsberg

Environmental and Energy Systems, Karlstad University, Sweden

X. Shao & J.W. Ringsberg

*Division of Marine Technology, Department of Mechanics and Maritime Sciences
Chalmers University of Technology, Sweden*

ABSTRACT: The power take-off (PTO) system and control laws for wave energy converters (WECs) are crucial for system performance. Determining optimal PTO parameters based on different sea states is key for formulating control laws, identifying optimal WEC locations, and optimizing wave-to-wire performance. This study focuses on an oscillating water column (OWC) type WEC developed by Waves4Power using a buoy and piston in a water tube to generate hydraulic pressure in an accumulator tank for an electric generator. For 78 different sea states, optimal generator parameters maximizing electrical output are identified using waves simulated from Pierson-Moskowitz (PM) spectra and a Simulink model, including non-linear PTO effects. Results show a strong dependency between optimal generator speed and sea state, suggesting control laws based on sea state can enhance energy production. Wave steepness is the most important characteristic, hence four representative sea states based on their steepness are used to illustrate results.

1 INTRODUCTION

In recent years, there has been an increased interest in harnessing renewable energy, in line with countries' objectives to transition towards renewable energy sources. An analysis of different renewable energy sources shows a significant increase in the global use of renewable energy in recent years. Hydropower, solar energy, and wind power are the most common sources of renewable electricity, followed by other sources including wave, geothermal, biomass, waste, and tidal (Energy Institute 2023).

Wave energy offers a predictable and sustainable source of energy that has minimal environmental impact, providing enhanced energy security. Although wave energy technology has potential, it is still in an early stage of development compared to other renewable energy sources (Hussain et al. 2017).

A central issue in the field of wave energy is improving the efficiency and effectiveness of Wave Energy Converters (WECs). Much research has been devoted to different aspects of WEC design, including structural design, response to different sea states and optimization of different Power Take-Off (PTO) sys-

tems. The following subsections provide an overview of existing works.

1.1 Wave Energy Converters

A WEC generates electricity from ocean waves by converting wave motion into electricity. The conversion to electricity is performed by a PTO system and several different PTO systems exist (see e.g. Ahamed et al. 2020, Lin et al. 2015 for overviews and comparisons). Zhang et al. (2021) utilized both qualitative and quantitative methods to identify the most suitable technical scheme for WECs. The study includes a comprehensive review and analysis of the technology principles of wave energy generation and devices, which can be utilized to develop WECs.

1.1.1 Effects of Sea States

Earlier studies have shown how different sea states affect the cost and life of WECs as well as how efficiently these devices can turn wave energy into electricity (Mackay 2012, Castellucci et al. 2017, Falnes 2020, Neshat et al. 2022).

Several authors have extensively explored the efficiency of WECs using simple models, in studies that examined how various wave features, including wave shape and grouping, affected the power performance of WECs (Saulnier et al. 2011, Lindgren 2015, Wang 2018, Wang & Wang 2018, Wang 2020). Further, more complex WEC models that include heaving buoy motion have been studied by Ahmed et al. (2022) and Azam et al. (2022). These studies used similar low-energy sea states as the Wang & Wang (2018) studies.

1.1.2 Power Take-Off Optimization

A PTO system is a crucial component in WECs, converting the mechanical energy of waves into electric energy. The system is essential for the efficient operation of WECs, as it directly influences their ability to harness wave power (Wang & Lu 2018). Hydraulic PTO systems, such as the one considered in this study, convert the mechanical energy from the waves into hydraulic energy that drives an electric generator.

An important part of this conversion process is the generator, particularly the speed at which it operates. The generator's speed, measured in Revolutions Per Minute (RPM), is a key factor in the system's overall efficiency. The optimal speed of the generator is significantly influenced by the prevailing sea states (see e.g. Falcão 2007). Achieving the right RPM is vital for maximizing the electrical energy output, thereby enhancing the efficiency of the WEC system.

Previous studies have shown the importance of RPM in WECs efficiency. Lekube et al. (2016) tried to find the corresponding operating point to maximize the delivered wave-to-wire power, and their results indicate that finding the optimal RPM gives maximum power generation. Falcão et al. (2017) showed that implementing a control strategy that adjusts the generator's speed can optimize the energy conversion process, resulting in higher electrical output while preventing the overloading of system components. Further work on optimization of PTO systems parameters include work by Jusoh et al. (2020), Amini et al. (2022), and Liu et al. (2024), where the last two consider floating PTOs of floating WECs.

1.2 WEC in this Study

The WEC considered in this study (WaveEL 4.0, a prototype manufactured by Waves4Power, see Fig. 1), is a buoy of an oscillating two-body type with a hydraulic PTO system. A brief description of the buoy is given in Section 2.4, for a more detailed analysis of the buoy hydrodynamics and anchoring see Shao et al. (2023). One body of the WEC is an oscillating water column (OWC), where the hydraulic pressure is provided by a moving piston in a water tube. WECs of the OWC type are regarded as the most reliable and efficient WEC solutions available (Falcão & Henriques 2016, Kasacik et al. 2020, Rosati et al. 2022).

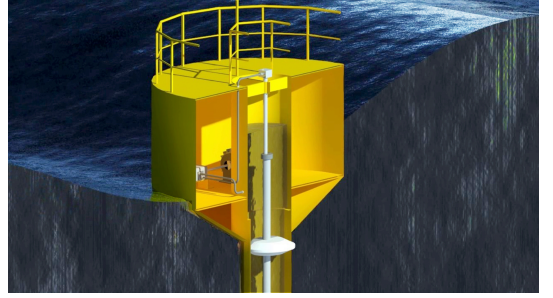


Figure 1: Cross-sectional illustration of an oscillating water column (OWC) type wave energy converter (WEC) buoy, featuring a water tube and moving piston that generate hydraulic pressure for the PTO system.

The earlier version, WaveEl 3.0, was installed near Ålesund, Norway, in 2016 and operated until November 2017. The main difference between the two versions is the extension of the water tube from 18.5 to 34 meters. The first installation of WaveEl 3.0 generated energy for the Norwegian grid. Previous studies of the WaveEL buoys include loading and fatigue in the mooring system as well as optimization of the accumulator pressure in the hydraulic PTO system (Gustafsson 2016, Johannesson et al. 2022, Shao et al. 2023, Zeinali et al. 2023, Zeinali et al. 2024).

1.3 Objective of study and outline

In this research, the aim is to optimize the generator's speed across different sea states, which are based on wave statistics from the European Marine Energy Centre (2022) at Billia Croo, Orkney. Our findings emphasize that the optimal speed of the generator depends on the sea state. Further, the speed strongly affects the energy output of the WEC, and thus the efficient and cost-effective production of wave energy.

Section 2 provides an overview of the methods used, including key aspects of wave modeling (Section 2.1) as well as simulation of waves (Section 2.3) and power output from the WEC (Section 2.4). The electrical power output is then optimized in Section 2.5, by finding optimal generator speeds for different sea states. In Section 3, the results on optimal generator speeds and configurations are presented and Section 4 concludes with a summary of findings.

2 MATERIALS AND METHODS

2.1 Waves

Sea state statistics used in this paper were selected based on wave statistics from the European Marine Energy Centre (2022) at Billia Croo, Orkney. The dataset is characterized by significant wave heights, H_s , energy periods, T_e , and frequency of occurrence. For this study, we have considered a representative

sub-sample of 78 states, covering 92.4% of all occurrences.

Using these states, waves were simulated using a generalized Pierson Moskowitz (GPM) or Bretschneider spectrum according to the ITTC-standard (ITTC 2002). The choice of spectrum is motivated by the location of the buoy at Runde, which is on the North Atlantic Coast of Norway. A GPM spectrum is suitable for fully developed seas where the waves have had enough time and distance to reach a stable, mature state such as in the North Atlantic (Pierson & Moskowitz 1964). For more enclosed waters, such as the North Sea, a JONSWAP spectrum might be more appropriate (Hasselmann et al. 1973). The JONSWAP spectrum requires an additional peakedness parameter which is not provided in the Billia Croo dataset.

We truncated the spectrum at $5\omega_p$ which preserves 99.8% of the spectral energy and results in the following spectrum,

$$S_{PM}(\omega) = \begin{cases} \frac{5m_0}{\omega_p} \left(\frac{\omega}{\omega_p}\right)^{-5} e^{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^4}, & 0 \leq \omega \leq 5\omega_p, \\ 0, & \omega > 5\omega_p. \end{cases} \quad (1)$$

Since the data for Billia Croo is presented using significant wave heights, H_s , and energy periods, T_e , we need to establish a relation between the energy period, T_e , and the peak period, $T_p = 2\pi/\omega_p$, in a truncated GPM-spectrum. The energy period can be computed from the spectral moments as

$$T_e = 2\pi \frac{m_{-1}}{m_0}, \quad (2)$$

where the spectral moments are given by

$$\begin{aligned} m_n &= \int_0^{5\omega_p} \omega^n m_0 \frac{5}{\omega_p} \left(\frac{\omega}{\omega_p}\right)^{-5} e^{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^4} d\omega \\ &= \omega_p^n m_0 \left(\frac{5}{4}\right)^{n/4} \Gamma\left(1 - \frac{n}{4}, \frac{1}{500}\right), \end{aligned} \quad (3)$$

with $\Gamma(a, b)$ being the upper incomplete gamma function (see e.g. Eq. 6.5.3 on p. 260 in Abramowitz & Stegun 1965). The spectral moments are finite for all n for the truncated GPM spectrum. Using this we obtain the following relation between T_p and T_e ,

$$T_e = \left(\frac{4}{5}\right)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}, \frac{1}{500}\right) e^{\frac{1}{500}} T_p \approx 0.86 T_p, \quad (4)$$

2.2 Wave steepness

Wave steepness is an important characteristic of the sea state. The steepness can be defined using significant wave height H_s , in combination with either energy period, T_p , or mean zero-crossing period, T_z , (Myrhaug 2018). We will here use the definition based on energy period,

$$S_p = \frac{2\pi H_s}{g T_p^2}. \quad (5)$$

2.3 Simulation of waves

In this study, we generate the input waves from a linear Gaussian time-series (t, Z_t) with N samples at sampling times $t = 0, \Delta_t, 2\Delta_t, \dots, T$ where $\Delta_t = \frac{T}{N-1}$ and Z_t denoting the height of the undisturbed sea surface at the buoy location. The time-series, Z_t , will be simulated as a Fourier sum

$$Z_t = \sum_{i=0}^{N-1} \sqrt{S(\omega_i) \Delta_\omega} R_i \cos(\omega_i t + \theta_i) \quad (6)$$

where the frequency steps are $\Delta_\omega = \frac{2\pi}{N-1}$; $\omega_i = i\Delta_\omega$; $S(\omega_i)$ is the truncated GPM spectrum from Eq. (1); θ_i are independent random phases, uniformly distributed in $(0, 2\pi)$; and R_i are Rayleigh distributed amplitudes with probability density

$$f(r) = r \exp(-r^2/2), \quad r \geq 0. \quad (7)$$

The sampling in Eq. (6) is implemented in the MATLAB toolbox Wafo (Wave Analysis for Fatigue and Oceanography; see Brodtkorb et al. 2000), which is freely available online (WAFO-group 2017). The Wafo routine spec2sdat can simulate with (almost) any time-interval $[0, T)$ and sampling interval Δ_t . For small Δ_t the spectrum may be extended and, if needed, padded with zeros at high frequencies. For large T , the spectrum will be interpolated with a small frequency step Δ_ω to avoid periodicity.

For each sea state, a time series of 650 seconds with a frequency of 10 Hz is generated. The first 50 seconds will be used as a burn-in period for the Simulink simulations.

2.3.1 Prefiltering of the Waves

The buoy acts as a point absorber in when the dominating waves are much longer than the buoy dimension. For a realistic irregular sea state with a fair amount of short waves, the shape and dimension of the horizontal section of the buoy need to be taken into account. The studied buoy has a ring-shaped upper part on top of a conical section attached to an open-ended bottom pipe. This hollow pipe extends through the cone and all the way to the top of the construction (Fig. 1).

Under modest sea conditions, the average buoyancy force over the ring-shaped section can be used to approximate the total buoyancy force. Before feeding the simulated waves into the Simulink model, the waves will be prefiltered to account for this hydrostatic averaging. Hydrodynamic forces from the added mass and damping due to the, customarily submerged, conical section are taken care of in the Simulink model and will not be included in the averaging.

2.4 Simulink Model for the WEC

The WEC in this study, see Section 1.2, is modeled as a Simulink object with a simplified block diagram presented in Fig. 2. In the following description, blocks are highlighted in *italics*.

The *buoy* consists of the buoy hull with a diameter of eight meters enclosing an open-ended tube in which the *water piston* floats. The relation between the masses of the buoy and the enclosed tube water is about 1:3 and buoy draught, due to the tube, is about thirty meters. The moorings are almost horizontal and soft, to minimize vertical forces that can influence buoy movements.

The vertical stroke of the water piston is restricted to four meters by rubber *bumpers* fixed to the buoy hull, above and below the water piston; the bumpers also prevent damage in extreme sea states. There is a leaking channel between the periphery of the water piston and the tube wall, which widens at the ends of the stroke. The pressure drop across the water piston, due to the flow in the leaking channel, excites forces on the tube water. This couples tube water to the buoy hull, influencing the resonance period of the buoy. The tube water is modeled by the *tube water inlet* and *tube water surface* blocks.

A *hydraulic system* with the hydraulic cylinder fixed to the buoy hull and the hydraulic piston fixed to the floating water piston connects the inertia of the tube water to the buoy hull. As a *wave* approaches, the buoy and hydraulic cylinder are lifted, while the hydraulic piston remains attached to the floating water piston. The hydraulic system uses three rectifying valves to direct the resulting hydraulic flow from a low-pressure tank to an accumulator tank. The tank opens at 2 MPa and has a maximum pressure set to 25 MPa, by a relief valve. If the accumulator pressure is larger than the pressure inside the hydraulic cylinder the water piston remains locked to the buoy hull, with no wave power absorbed. Otherwise, the movement of the water piston excites a braking force on the buoy hull, with amplitude given by the accumulator pressure. It seems like maximum wave power is absorbed when the water piston is locked to the buoy hull 40–60% of the time.

Electrical power is produced by a hydraulic motor driving a *generator*, the motor is fed from the accumulator tank by a flow that returns to the low-pressure tank. A *speed controller* maintains constant speed (RPM) for the generator, allowing the pressure in the accumulator to fluctuate. The uptake of mechanical energy from the waves will depend on the pressure in the accumulator tank. A key question is what target RPM should be used by the controller to maximize energy uptake for different sea states.

The Simulink model used here (see Zeinali et al. (2024) for a longer description) focuses on interactions within the buoy. This includes nonlinearities due to 1) water flow in the open-ended tube, both at the in-

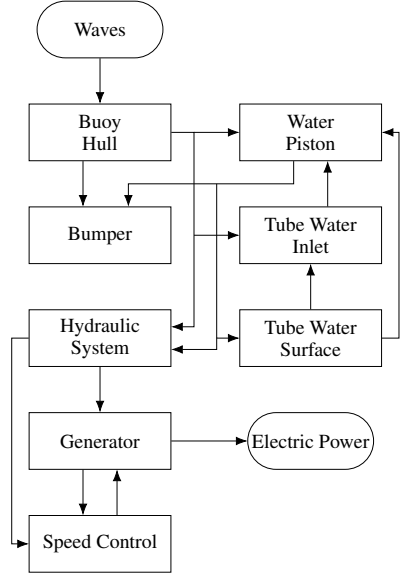


Figure 2: A schematic figure of the Simulink model for the WEC buoy, illustrating key components of the model and how they interact. Rounded rectangles indicate input and output values for the model and rectangles are the internal blocks of the model. The arrows indicate connections between components.

let and in the leaking channel between the water piston and tube wall; 2) elasticity in the rubber bumpers when the water piston reaches its extreme range of motion; 3) compressibility effects for the hydraulic fluid; and 4) opening and closing of valves controlling the flow of hydraulic flow in and out of the accumulator tank. A simplified model, considering only vertical heave motion and ignoring mooring effects, is used for the hydrodynamics of the buoy hull. A companion paper (Shao et al. 2024) investigates the magnitude of mooring effects on the buoy when considering increasingly complex hydrodynamic models.

2.5 Optimization

A total of 78 wave time series were fed into the Simulink model. Our setup has three combinations of running conditions for the generators, a 50 kW generator, a 100 kW generator, or both generators (150 kW). For each combination of sea states and generator configuration, we optimized the generator's speed by maximizing the average power over the last 10 minutes of the simulation for each of the states. The first 50 seconds were discarded as burn-in to eliminate transient effects. The averaging period of 10 minutes was selected as a trade-off between computational cost and accuracy of the estimated power. The uncertainty was considered sufficiently small by assessing the variances in average power across the time period.

In this study, MATLAB's built-in function `fminbnd` was employed to optimize the speed of the generator (MathWorks 2024), with a focus on max-

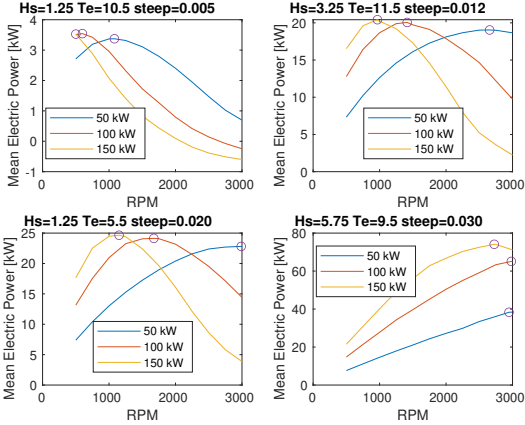


Figure 3: Average generated electric power as a function of generator speed for four representative sea states with increasing steepness. Circles show the maximum electric power based on the optimized speed for each of the three generators configurations, as indicated by the different colors.

imizing the electric power generation. `fminbnd` was chosen for its efficiency and robustness in optimizing univariate functions on bounded intervals as well as its ability to handle noisy or complex objective functions (Press 2007, Ch. 10.2).

2.6 Common Random Numbers

As previously mentioned, 78 sea states were utilized for this study. Employing distinct wave data for each tested speed could result in high variance, which in turn could lead to unstable optimization. To mitigate this variability, the approach of common random numbers (CRN) was employed, meaning that one shared set of wave data was utilized for all speed values. CRN is a technique widely used in statistics to reduce variances and provide smoother outputs, as can be seen in Fig. 3.

3 RESULTS

There are three combinations of running conditions for the generators, only the 50 kW generator, only the 100 kW generator, or both generators (150 kW). Both generators have an allowed RPM range of 500 to 3000 min^{-1} .

The results in Fig. 4, illustrate how steepness is related to optimal speed for each generator configuration. As can be seen in Fig. 4, using only the 50 kW generator for a steepness of more than 0.01 results in the RPM being set to its maximum value, which shows the necessity of a larger generator. The same argument applies for the 100 kW generator when the steepness exceeds 0.02.

Based on the results in Fig. 4 we want to design a simple rule of thumb for which generator to use at different wave steepnesses. For low steepness, there is no

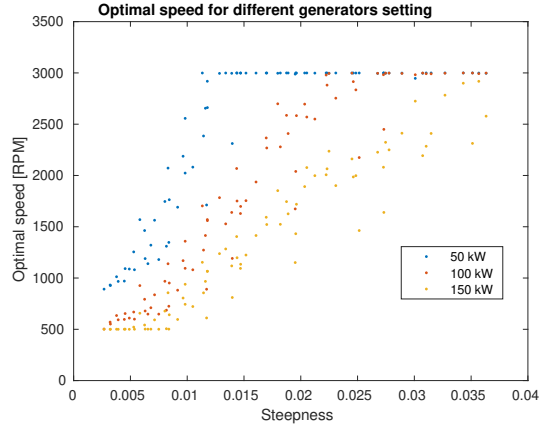


Figure 4: Optimal speed as a function of steepness for each of the 78 sea states, across all generator settings. We see that there are sea states such that if we run either the 50 kW or the 100 kW generators separately we reach the maximum allowed long-term RPM, which is 3000 for our generators. Therefore we need the 150 kW setup, which is accomplished by running both generators in parallel.

significant gain in average power by running the bigger generator (or both generators). For medium steepness, it is enough to run the 100 kW generator and for the high steepness states we need to run both generators to fully utilize the available energy in the waves. For simplicity, we want a rule based on only steepness for each running condition. The cut-off points are selected so that the optimal RPM is in a reasonable range for each case. We then obtain three optimal generators running configurations (Eq. (8)) as a function of wave steepness:

$$\begin{cases} 50\text{kW} & 0 < \text{steepness} \leq 0.01 \\ 100\text{kW} & 0.01 < \text{steepness} \leq 0.02 \\ 150\text{kW} & 0.02 < \text{steepness} \end{cases} \quad (8)$$

Depending on wave steepness, one of the generator configurations should be selected, and the optimal speed is then determined based on that configuration, as illustrated in Fig. 5.

In Fig. 6 we see the optimal average electric power obtained by the above rule of thumb, as well as the average accumulator pressure as a function of wave steepness. Recall that the accumulator pressure excites a breaking force on the buoy hull through the water piston. Higher steepness is in many cases, for this dataset, associated with more energetic waves (although, there is no direct equality between steepness and energy content). This translates into larger accumulator pressure which provides the increased breaking forces needed to absorb the energy in those waves. It is clear that the steepness is a very important wave characteristic when evaluating the performance of our WEC.

Even though the accumulator tank and generator do a great job of smoothing out the highly varying in-

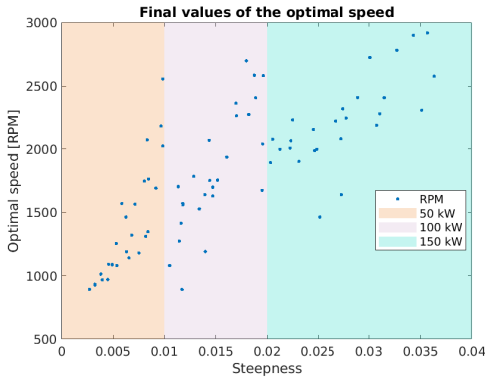


Figure 5: Optimal speed for each of the 78 sea states, based on wave steepness and selected generator configuration.

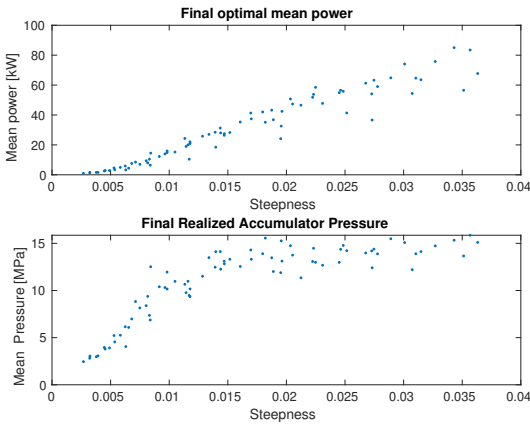


Figure 6: The upper panel shows the mean generated electric power and the lower panel shows the resulting mean accumulator pressure, both as a function of steepness.

coming wave energy, there are still substantial power fluctuations as can be seen in Fig. 7. A more detailed suggestion on how to address this can be found in the conclusions.

4 CONCLUSIONS

In this paper, our goal was to design a control system and find optimal speed values for the generators, to maximize the generated electric power. We identified the optimal speed and generator configuration by dividing the sea states into three regions based on wave steepness. This emphasizes the need for precise control mechanisms to adapt to varying sea states.

Our research highlights the crucial role of a control design that can adapt to different sea states when optimizing PTO systems in WECs to maximizing power production. Our findings contribute to enhancing the sustainability and competitiveness of renewable energy sources by improving the performance of WECs. In order to provide energy to the power grid in re-

alistic settings, further smoothing of the electrical output is likely to be needed. Using a park of several WECs will partly stabilize the output by averaging across multiple sources. The optimal configuration for the relative placement of the WECs is an important design feature in this case (Shao et al. 2023). Alternatively, an optimization function that penalizes high variability in the generated power could be considered, possibly in conjunction with optimization based on both wave height and period instead of only steepness, preliminary work in this direction has been done by Augustsson (2024). Further smoothing can be done by adding power electronics and/or batteries to the construction. For maintenance reasons preferable where the cables reach the coast. These final parts are interesting extensions, but outside the scope of this paper.

ACKNOWLEDGMENTS

This research was funded by grant 50197-1 from the Swedish Energy Agency and performed in collaboration with the company Waves4Power.

REFERENCES

- Abramowitz, M. & I. Stegun (1965). *Handbook of mathematical functions*. New York: Dover Publications.
- Ahamed, R., K. McKee, & I. Howard (2020). Advancements of wave energy converters based on power take off (pto) systems: A review. *Ocean Engineering* 204, 107248.
- Ahmed, A., Y. Wang, A. Azam, & Z. Zhang (2022). Design and analysis of the bulbous-bottomed oscillating resonant buoys for an optimal point absorber wave energy converter. *Ocean Eng.* 263, 112443.
- Amini, E., H. Mehdipour, E. Faraggiana, D. Golbazi, S. Mozafari, G. Bracco, & M. Neshat (2022). Optimization of hydraulic power take-off system settings for point absorber wave energy converter. *Renewable Energy* 194, 938–954.
- Augustsson, M. (2024). Unwavering potential — optimizing and estimating the power output from a wave energy converter. Master's thesis, Lund University, Lund.
- Azam, A., A. Ahmed, H. Li, A. M. Tairab, C. Jia, N. Li, & Z. Zhang (2022). Design and analysis of the optimal spinning top-shaped buoy for wave energy harvesting in low energy density seas for sustainable marine aquaculture. *Ocean Eng.* 255, 111434.
- Brodtkorb, P., P. Johannesson, G. Lindgren, I. Rychlik, J. Rydén, & E. Sjö (2000). Wafo - a matlab toolbox for analysis of random waves and loads. In *Proceedings of the 10th International Offshore and Polar Engineering Conference*, Volume III, pp. 343–350.
- Castellucci, V., J. García-Terán, M. Eriksson, L. Padman, & R. Waters (2017). Influence of sea state and tidal height on wave power absorption. *IEEE Journal of Oceanic Engineering* 42(3), 566–573.
- Energy Institute (2023). Statistical review of world energy (2023) – other renewables (including geothermal and biomass) [dataset]. <https://www.ourworldindata.org>. Major processing by Our World in Data. Original data from Energy Institute, “Statistical Review of World Energy.”
- European Marine Energy Centre (2022). Billia Croo wave data base.
- Falcão, A. F. d. O. (2007). Modelling and control of oscillating-body wave energy converters with hydraulic power take-off

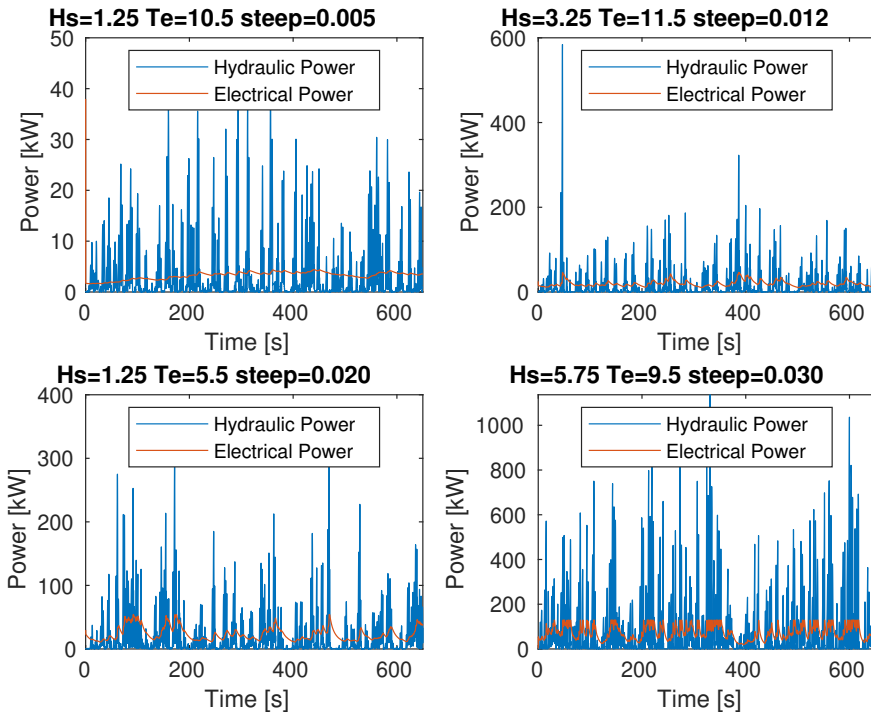


Figure 7: The energy captured by the hydraulic system is fed into the accumulator tank and is thereby smoothed out before being converted to electricity by the generator. We clearly see the smoothing effect of the accumulator when comparing hydraulic and electrical power for four representative sea states with increasing steepness.

- and gas accumulator. *Ocean engineering* 34(14-15), 2021–2032.
- Falcão, A. & J. Henriques (2016). Oscillating-water-column wave energy converters and air turbines: A review. *Renewable Energy* 85, 1391–1424.
- Falcão, A., J. Henriques, & L. Gato (2017). Rotational speed control and electrical rated power of an oscillating-water-column wave energy converter. *Energy* 120, 253–261.
- Falnes, K. (2020). *Ocean waves and oscillating systems: linear interactions including wave-energy extraction*, Volume 8. CUP.
- Gustafsson, E. (2016). Extreme loading and fatigue analysis of a wave energy device. Master's thesis, KTH Royal Institute of Technology, Stockholm.
- Hasselmann, K., T. Barnett, E. Bouws, H. Carlson, D. Cartwright, K. Enke, J. Ewing, H. Gienapp, D. Hasselmann, P. Kruseman, A. Meerburg, P. Muller, D. Olbers, K. Richter, W. Sell, & H. Walden (1973). Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). *Deutschen Hydrographischen Zeitschrift* 8, 1–95.
- Hussain, A., S. Arif, & M. Aslam (2017). Emerging renewable and sustainable energy technologies: State of the art. *Renew. sustain. energy rev.* 71, 12–28.
- ITTC (2002). The specialists committee on waves, final report and recommendations to the 23rd ITTC. In *Proceedings of the 23rd ITTC*, Volume II, pp. 505–736.
- Johannesson, P., X. Lang, E. Johnsson, & J. Ringsberg (2022). Mechanical reliability analysis of flexible power cables for marine energy. *J. Mar. Sci. Eng.* 10, 716.
- Jusoh, M., M. Ibrahim, M. Daud, Z. Yusop, A. Albani, S. Rahman, & S. Mohad (2020). Parameters estimation of hydraulic power take-off system for wave energy conversion system using genetic algorithm. In *IOP Conference Series: Earth and Environmental Science*, Volume 463, pp. 012129. IOP Publishing.
- Kasacik, D., V. Stratigaki, M. Wu, L. Cappietti, I. Simonetti, P. Troch, A. Crespo, C. Altomare, J. Dominguez, M. Hall, M. Gómez-Gesteira, R. B. Canelas, & P. Stansby (2020). Efficiency and survivability of a floating oscillating water columns wave energy converter moored to the seabed: An overview of the EsfOWC MaRINET2 database. *Water* 12, 1–21.
- Lekube, J., A. J. Garrido, & I. Garrido (2016). Rotational speed optimization in oscillating water column wave power plants based on maximum power point tracking. *IEEE Transactions on Automation science and Engineering* 14(2), 681–691.
- Lin, Y., J. Bao, H. Liu, W. Li, L. Tu, & D. Zhang (2015). Review of hydraulic transmission technologies for wave power generation. *Renewable and Sustainable Energy Reviews* 50, 194–203.
- Lindgren, G. (2015). Asymmetric waves in wave energy systems analysed by the stochastic Gauss-Lagrange model. *Proc. Est. Acad. Sci.* 64, 291–296.
- Liu, Z., S. Liu, W. Chen, Y. Yang, & G. Feng (2024). Implementation and optimization of hydraulic wave energy generation system. *Plos one* 19(2), e0293209.
- Mackay, E. (2012). Resource assessment for wave energy. In A. Sayigh (Ed.), *Comprehensive Renewable Energy*, Volume 8, pp. 11–77. Oxford: Elsevier sci.
- MathWorks (2024). fminbnd Documentation. [Online; accessed Date].
- Myrhaug, D. (2018). Some probabilistic properties of deep water wave steepness. *Oceanologia* 60(2), 187–192.
- Neshat, M., M. Nezhad, N. Sergienko, S. Mirjalili, G. Piras, & G. DA. (2022). Wave power forecasting using an

- effective decomposition-based convolutional Bi-directional model with equilibrium Nelder-Mead optimiser. *Energy J.* 256, 124623.
- Pierson, W. & L. Moskowitz (1964). A proposed spectral form for fully developed wind seas based on the similarity theory of s. a. kitaigorodskii. *Journal of Geophysical Research (1896-1977)* 69(24), 5181–5190.
- Press, W. H. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press.
- Rosati, M., J. Henriques, & J. Ringwood (2022). Oscillating water column wave energy converters: A critical review of numerical modelling and control. *Energy Conversion and Management: X* 16, 100322.
- Saulnier, J., A. Clément, A. Falcão, T. Pontes, M. Prevosto, & P. Ricci (2011). Wave groupiness and spectral bandwidth as relevant parameters for the performance assessment of wave energy converters. *Ocean Eng.* 38, 130–147.
- Shao, X., J. W. Ringsberg, H.-D. Yao, E. Johnson, J. Forsberg, S. Zeinali, J. Lindström, & M. Wiktorsson (2024). A comparison of approaches integrating power take-off systems into wave energy converters simulations. In *submitted to RENEW 2024*.
- Shao, X., J. W. Ringsberg, H.-D. Yao, Z. Li, E. Johnson, & G. Fredriksson (2023). A comparison of two wave energy converters' power performance and mooring fatigue characteristics – one wec vs many wecs in a wave park with interaction effects. *Journal of Ocean Engineering and Science* 8(4), 446–460. Ocean Engineering for Future Marine Structures and Applications.
- Shao, X., H.-D. Yao, J. Ringsberg, Z. Li, E. Johnson, & G. Fredrikson (2023). A comparison of the performance and characteristics of two generations waves4power waveel wave energy converters. In *Proceedings of The 5th International Conference on Renewable Energies Offshore (RENEW 2022)*, pp. 277–284.
- WAFO-group (2017). Wafo – a Matlab toolbox for analysis of random waves and loads. <https://github.com/wafo-project>.
- Wang, D. & K. Lu (2018). Design optimization of hydraulic energy storage and conversion system for wave energy converters. *Protection and Control of Modern Power Systems* 3, 1–9.
- Wang, Y. (2018). A novel simulation method for predicting power outputs of wave energy converters. *Applied Ocean Research* 80, 37–48.
- Wang, Y. (2020). Predicting absorbed power of a wave energy converter in a nonlinear mixed sea. *Renew. Energ.* 153, 362–374.
- Wang, Y. & L. Wang (2018). Towards realistically predicting the power outputs of wave energy converters: Nonlinear simulation. *Energy* 144, 120–128.
- Zeinali, S., M. Wiktorsson, J. Forsberg, G. Lindgren, & J. Lindström (2023). Optimizing the hydraulic power take-off system in a wave energy converter. In *NordStat 2023*, Gothenburg, Sweden.
- Zeinali, S., M. Wiktorsson, J. Forsberg, G. Lindgren, & J. Lindström (2024). Optimizing the hydraulic power take-off system in a wave energy converter. Accepted for publication in *Ocean Engineering*.
- Zhang, Y., Y. Zhao, W. Sun, & J. Li (2021). Ocean wave energy converters: Technical principle, device realization, and performance evaluation. *Renewable and Sustainable Energy Reviews* 141, 110764.

Paper III



Article

Detection of Bark Beetle Attacks Using Time-Aggregated Satellite Data with Machine Learning

Shokoufa Zeinali ^{1,*} , Per-Ola Olsson ² , Ted Kronvall ¹ , Magnus Wiktorsson ¹  and Johan Lindström ¹ 

¹ Centre for Mathematical Sciences, Lund University, Märkesbacken 4, 223 62 Lund, Sweden; ted.kronvall@matstat.lu.se (T.K.); magnus.wiktorsson@matstat.lu.se (M.W.); johan.lindstrom@matstat.lu.se (J.L.)

² Department of Earth and Environmental Sciences, Lund University, Sölvegatan 12, 223 62 Lund, Sweden; per-ola.olsson@mgeo.lu.se

* Correspondence: shokoufa.zeinali@matstat.lu.se

Highlights

What are the main findings?

- Combining weekly Sentinel-2 data with static geo- and forestry features using XGBoost significantly improves early bark beetle detection compared to using either data source alone and outperforms Random Forest and prior studies.
- Key spectral bands (Red-edge 3, blue) and specific weeks after main swarming (19, 25, 31, 33) are most informative; static features (land cover, landforms) enhance performance and shift feature importance.

What are the implications of the main findings?

- Integrating temporal satellite data with static environmental information is critical for accurate early detection of bark beetle attacks.
- Identifying important weeks and features supports the development of more effective, timely early warning and monitoring systems.

Abstract

In this study, we explored how early bark beetle attacks can be detected using weekly aggregated Sentinel-2 data in combination with static data, such as geo- and forestry data. We used an XGBoost classifier, known for its strength and reliability in classification, and compared three sets of data: static data only, satellite data only, and using both together. Having trained the models on cumulative weekly data, we were able to track changes in model performance and feature importance over time, identifying key weeks for the detection of bark beetle attacks. A systematic overview of feature importance identified the Red-edge 3 and blue Sentinel-2 bands as the most important when combined with static data; it also showed changes in feature importance compared to using satellite-only data, e.g., adding static features reduced the importance of red and red-edge 2 bands. Among the static features, land cover and landforms were the most important. Evaluating the temporal features for the combined model highlighted certain weeks as containing key information for detection: week 19, which was the main swarming week; week 25, which is 6 weeks after swarming and just before the second generation is completed; and weeks 31 and 33, more than 3 months after the tree was attacked and well after the new generation has swarmed. The study shows that combining static features with cumulative Sentinel-2, accumulated across weeks, are all important for improving the detection of bark beetle attacks, and that such ideas form an important part of early warning systems.



Academic Editor: Michael Sprintsin

Received: 28 April 2026

Revised: 17 June 2026

Accepted: 24 June 2026

Published: 6 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: detecting insect infestations; XGBoost; static features; time series analysis; forest health monitoring; Sentinel-2

1. Introduction

Bark beetle outbreaks represent one of the most severe disturbances affecting coniferous forests, posing significant threats to forest health, ecosystem services, the economic viability of forestry operations, as well as carbon sequestration [1–4]. The European spruce bark beetle (*Ips typographus*) is a phloem-feeding insect whose life cycle and outbreak dynamics are strongly influenced by temperature. Adult beetles locate and attack suitable Norway spruce (*Picea abies*) trees, mate, and lay eggs under the bark; the larvae feed and develop in the phloem, and the brood adults emerge to colonize new hosts. Successful colonization is facilitated by mass attack and, in many cases, by associated fungi that help overcome host defenses [5]. Temperature-dependent development determines the timing of key life-cycle stages such as swarming, mating, and larval development, making the species highly sensitive to inter-annual climatic variability. In southern Sweden, a new generation of bark beetles usually swarm around eight weeks after a tree was colonized [6].

In Europe, the European spruce bark beetle (*Ips typographus*) is estimated to have accounted for 8% of mortality due to natural disturbances in Norway spruce (*Picea abies*) during the period 1850 to 2000 [7]; this proportion increased after 2000 [8]. Recent bark beetle outbreaks have killed nearly 31 million m³ of spruce trees in Sweden, 50 million m³ in Czechia, and 86 million m³ in Germany [9,10]. Unlike the extensive, contiguous outbreak patterns often reported in Central Europe [11], bark beetle damage in Sweden was more spatially dispersed, with many attacked trees scattered across the landscape.

Several studies have investigated the ability of satellite-based remote sensing data to detect beetle damage and attacked trees. Sentinel-2 gave better results than Landsat for early detection of bark beetle-infested trees in Germany [12] and Czechia [13]. Regarding the utility of different spectral bands, ref. [14] developed a new vegetation index, NDRS (normalized distance Red and SWIR), while [15] obtained the best detection results with the tasseled cap wetness (TCW) index. Various areal-based hyperspectral platforms have been used as an alternative to satellites, e.g., [16–18].

From a forest management perspective, it is important to detect attacked trees as early as possible. Several studies (e.g., [15,19–21]) have investigated early detection and the ability to separate trees in the early (green) and late (red) stages of attack. While evaluating how model performance changes with additional temporal observations, these studies do not systematically evaluate which weeks are the most important for detection, using, e.g., feature importance.

The high accuracies for early detection might reflect the detection of stressed trees rather than the true symptoms of early attacks, since there might have been differences between healthy and green attack pixels before an attack [13,14,20]. This has inspired the use of static forest features to identify susceptible trees [22,23], with promising results. Ref. [24] combined static features with Sentinel-2 to detect trees at risk one year before the attack. Using Random Forest models, authors of [24] found a small (5%) increase when adding Sentinel-2 data to static features. Detection accuracy was the same before and after the attack, indicating that the model identified stressed/susceptible trees. Using only Sentinel-2 from single dates, authors of [24] did not investigate how increased availability of Sentinel-2 during the growing seasons affects detection, or which time periods have the greatest influence on detection.

In this paper, we expand on the single-date analysis in [24] by using time-aggregated satellite data combined with static features (data are described in Section 2). We also replace random forests with the computationally more efficient XGBoost (see [25] and methodology in Section 3). The aim of the analysis (Section 4) is to:

1. Systematically evaluate how the temporal information in the satellite data contributes to detecting bark beetle attacks;
2. Investigate which individual weeks are the most important for detection and relate these weeks to the time of bark-beetle swarming as observed by pheromone traps;
3. Determine when time-aggregated satellite data is informative enough to outperform static features for detection; and
4. Investigate the importance of individual spectral bands and static features for detection.

The methodology of this study is deliberately based on well-established machine learning techniques, such as XGBoost and standard pre-processing strategies, to ensure that observed differences in performance are primarily driven by the structure and temporal availability of the input data rather than model complexity. The main contribution of this work lies in the systematic evaluation of how temporally accumulating multi-spectral satellite observations influence the detectability of bark beetle infestations. In particular, our experimental design enables controlled comparisons across different data configurations, including static-only predictors, satellite-only temporal features, and their combination. To analyze how the predictive performance evolves over time and how early reliable detection can be achieved the number of weekly satellite observations seen by the model is gradually increased.

2. Study Area and Datasets

The study area is $180 \times 180 \text{ km}^2$ and located in southeastern Sweden (Figure 1). The annual mean temperature is 6 to 8 °C and the average summer temperature is 15 to 16 °C for the period 1991–2020 [26]. The area is dominated by forests with Norway spruce (*Picea abies* (L.) Karst) and Scots pine (*Pinus sylvestris* L.) as the most common species, with only smaller patches of deciduous forests [27]. The hot drought in 2018 triggered the largest documented bark beetle outbreak in Sweden, with 31 million m³ of spruce trees affected in the south between 2018 and 2022 [10] and strongly affecting the study area.

2.1. Sentinel-2 Data

The satellite images used in this article are derived from Sentinel-2 with a spatial resolution of 10–60 m and a revisiting period of 2–3 days across the study area [28]. A total of 21 weeks of satellite data were retained, covering calendar weeks 15 to 43 (April to late October) of 2018. The images were downloaded as level 1C and preprocessed with the Framework for Operational Radiometric Correction for Environmental Monitoring [29] (FORCE ver. 3.6.5) to calculate surface reflectance and perform image-to-image registration [30]. The preprocessing followed the same steps as in [24]. Only bands with a spatial resolution of 10 and 20 m were included (see Table 1), and all bands with 20 m resolution were resampled to 10 m with nearest neighbor interpolation. The study area was divided into four tiles as depicted in Figure 1. To ensure consistency across tiles, only the calendar weeks with available data in all tiles were included in the analysis. The distribution of available weeks per tile is shown in Figure 2 and correlation between bands can be found in Figure A1 of Appendix A.

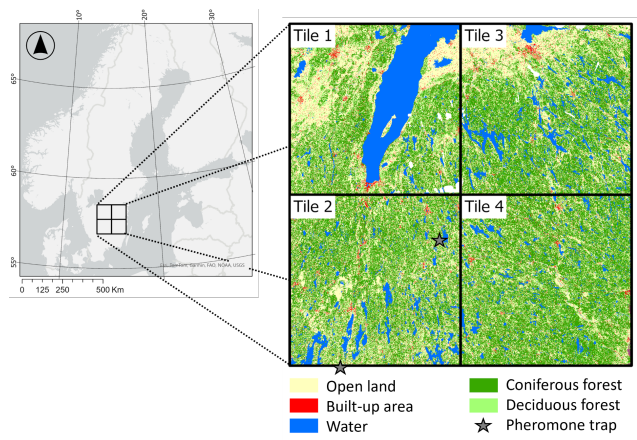


Figure 1. Location of the study area in Sweden divided into four tiles (left) and land cover [31] in the four tiles (right).

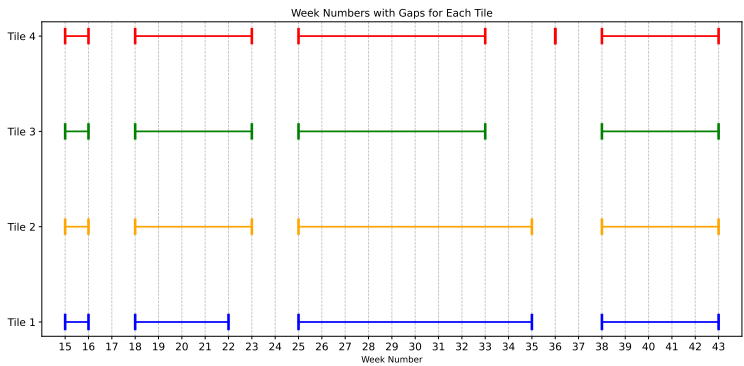


Figure 2. Available weeks of Sentinel-2 data with sufficient cloud-free observations for each tile before preprocessing, where each line and colour represents a different tile, segments show continuous weeks, gaps indicate missing weeks, and vertical bars mark the first and last week of each sequence.

Table 1. Band number, name, spatial resolution, and abbreviations for the Sentinel-2 bands used in this study.

Sentinel-2 Band	Name	Original Spatial Resolution (m)	Abbreviation
2	Blue	10	Blue
3	Green	10	Green
4	Red	10	Red
5	Red-edge 1	20	RE1
6	Red-edge 2	20	RE2
7	Red-edge 3	20	RE3
8	Near Infrared	10	BNR
8a	Narrow NIR	20	NIR
11	Short Wave Infrared 1	20	SWIR1
12	Short Wave Infrared 2	20	SWIR2

2.2. Static Data

We also used geo-spatial, topography, and forestry features that are known to influence the risk of bark beetle infestations [22,24] in this article. Since topography is static and the

forestry parameters change slowly (data is typically updated every few years), we refer to those variables as static, in contrast to the dynamic satellite data that follow the within-season changes in phenology. The static features used here are listed below; they were all pre-processed to the same extent and spatial resolution as the Sentinel-2 data (see [22] for further details). Correlation between the features is illustrated in Figure A2 of Appendix A. These features are not direct spectral indicators of attack symptoms; rather, they provide contextual information on site conditions, stand structure, and forest composition that may show infestation risk and susceptibility of trees.

aspect: The direction that a slope faces.

basalArea: A measure of the cross-sectional area of tree trunks in a forest.

biomass: Total mass of living vegetation in an area.

CHM (Canopy Height Model): Height of vegetation canopies.

DEM (Digital Elevation Model): Elevation data of the terrain, indicating the height above sea level.

Edge: Proximity of a location to the edge of a forest or habitat.

LandC (Land Cover): Type of land cover.

LandForm: Characterizes the shape of the terrain.

slope: The steepness of the terrain.

soilType: Classification of soil.

soilWetness: Moisture content of the soil. Static variable partly derived from a topographic wetness index.

spruceVolume: Volume of spruce trees.

2.3. Response Data

Bark beetle damage data were obtained from Sveaskog, the largest forest owner in Sweden. The location and date of harvest for trees killed by bark beetles were collected by harvester machines; the data do not include the date of attack. We included only trees harvested from April to June 2019, as they were most likely attacked in 2018. The harvester dataset was rasterized to the same 10×10 m spatial resolution and extent as the FORCE-processed Sentinel-2 data, with pixels containing at least one attacked tree labeled as attacked, giving a binary response variable (see [24] for further details).

Limited data on forest stands without bark beetle-infested trees, here labeled healthy data, were available as inventory data obtained from the Swedish forest agency. To increase the amount of data on healthy forest stands, we utilized national land cover data with a spatial resolution of 10×10 m [31] and a property map [32] in combination with harvester data on attacked trees. The harvester machines were brought to the forest by the management to remove bark beetle-killed trees; hence, we made the assumption that forest stands without harvested trees were healthy if they were located inside a forest estate where trees had been attacked and harvested in other stands. To focus on spruce trees, we only included the land cover classes spruce forest, mixed coniferous forest, and mixed forest when deriving the healthy data. The identified pixels were added as healthy data to the binary raster. The resulting dataset is highly imbalanced; the total number of samples and the percentage of attacked samples across different tiles are shown in Table 2.

Bark beetle swarming data, based on pheromone traps, from the Swedish Forest Agency [33] were used to estimate the timing of attacks (see [24]). Two stations with pheromone traps were located in the study area, one on the southern boundary (Lagan) and one in the central part (Nässjö) of the area (Figure 1). The traps were emptied on a weekly

basis, and in 2018 there was a distinct peak in the swarming data in week 19, indicating that the trees were most likely attacked at that time (the swarming data are included in Figure 3 in the results). In addition, we checked the two pheromone traps closest to the study area; one just west of the area and another near the city of Örebro around 90 km north of the study area, and both followed the same pattern with a very distinct peak during week 19, indicating that week 19 was the main swarming week over the entire study area.

Table 2. Total number of samples and proportion of attacked samples in each tile.

Tile Id	Total Number of Samples	Percentage of Attacked Samples
1	19,882	3.8%
2	28,204	5.2%
3	43,911	0%
4	71,474	6.7%
	163,471	4.3%

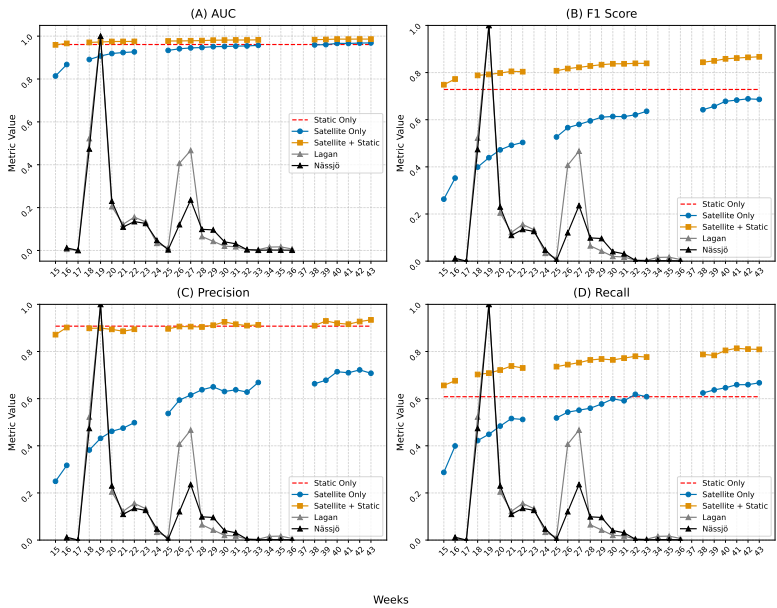


Figure 3. Evaluation metrics over time using: only static data (dashed Red line), only satellite data (blue line/circles), static and satellite data (yellow line/squares). The swarming data for the pheromone traps at Lagan and Nässjö are shown with gray/black lines and triangles; values are scaled so that 1 corresponds to the largest number of captured bark beetles (6000 at Lagan and 11,885 at Nässjö).

3. Methodology

3.1. Sentinel-2 Data Preparation

To obtain a consistent temporal resolution and reduce the effects of missing data and noise—such as cloud cover, cloud shadows, or snow—weekly maxima were computed for each pixel. Temporal aggregation of satellite data using maxima is commonly used in environmental sciences, since maxima have been shown to better indicate ecological processes compared to averages or minima [34]. Additionally, clouds (and cloud shadows)

tend to bias vegetation indices towards lower values [35]. The resulting consistent temporal resolution, consisting of weekly data, facilitates our trend analysis and anomaly detection.

After aggregation, weeks with more than 10% missing values across the study area were dropped. For weeks with less than 10% missing values, the remaining missing values were imputed using the average of the previous three weeks. For the first three weeks in the time series, imputation was performed using forward and backward filling techniques. First, forward filling propagated the first available observation forward, while backward filling assigned the earliest valid observation backward to fill any initial missing values.

3.2. Balancing Response Data

The response dataset is highly imbalanced (see Section 2.3), potentially biasing any classifier towards healthy pixels. We therefore used the Synthetic Minority Oversampling Technique (SMOTE, [36]) to balance the training dataset by creating synthetic examples for the minority class and dropping a proportion of the majority class. It is important to note that the validation data were left untouched, so that the data imbalance is correctly represented when validating the models.

Given a matrix of features, X_{train} , and a vector of labels, y_{train} , SMOTE takes a random sample, x_i , from the minority class and finds the k -nearest neighbors in the minority class. Here, we used $k = 5$ neighbors. Synthetic samples for the minority class were then generated by linearly interpolating between the sample, x_i , and each of the k -nearest neighbors, x_{nn} :

$$x_{new} = x_i + \lambda \cdot (x_{nn} - x_i), \quad (1)$$

where λ is a random number between 0 and 1. This process was repeated, e.g., new random samples x_i were drawn, until the target number of synthetic minority samples was reached. Pseudo-code for the method is given in Algorithm 1.

Algorithm 1 SMOTE Resampling Algorithm

Require:

- 1: X_{train} : Feature matrix of training data
 - 2: Y_{train} : Labels of training data
 - 3: **while** more synthetic samples are needed **do**
 - 4: Randomly pick a minority class sample x_i from X_{train}
 - 5: Find the k nearest neighbors of x_i within the minority class
 - 6: **for each** x_{nn} from the k neighbors **do**
 - 7: Generate a random number $\lambda \sim U(0, 1)$
 - 8: Compute synthetic sample:
 - 9: $x_{new} = x_i + \lambda \cdot (x_{nn} - x_i)$
 - 10: Add x_{new} to X_{new}
 - 11: Assign label of minority class to Y_{new}
 - 12: **end for**
 - 13: **end while**
 - 14: Combine the original data and synthetic samples $\hat{X} = \{X_{train}, X_{new}\}$ and $\hat{Y} = \{Y_{train}, Y_{new}\}$.
 - 15: Return the resampled $\hat{X}$ and $\hat{Y}$
-

By interpolating between points in this way, SMOTE creates a more general representation of the minority class, helping to enhance the training of the classifier by providing more examples to learn from.

3.3. XGBoost

XGBoost (Extreme Gradient Boosting), introduced by [25], is a scalable and efficient implementation of gradient boosting decision trees, widely used for classification and regression tasks. Unlike Random Forest, which builds an ensemble of trees in parallel, XGBoost constructs trees sequentially, where each new tree is trained to correct the errors made by the previously built trees. The final prediction is obtained by aggregating the outputs of all trees, typically using a weighted sum for regression or logistic transformation for binary classification. This boosting framework helps minimize both bias and variance, often achieving superior predictive performance compared to Random Forest and other bagging-based methods.

In XGBoost, each tree is built by greedily selecting feature splits that minimize an objective function that combines the training loss with a regularization term that penalizes model complexity. The objective function at iteration t is generally expressed as:

$$\mathcal{L}^{(t)}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

where $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ is the training loss for training sample i given the previous prediction $\hat{y}_i^{(t-1)}$ and the new tree $f_t(x_i)$, and $\Omega(f_t)$ is the regularization term for the new tree f_t . A common loss function for binary classification is the logistic, also known as binary cross-entropy, loss:

$$l(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

The algorithm uses a second-order (Newton) approximation of the loss function. By applying a Taylor expansion to the loss function around the current prediction $\hat{y}_i^{(t-1)}$, the objective function can be simplified and optimized in a greedy manner.

To determine the best splits, XGBoost computes the gain, or improvement, in the objective function at each potential split. For a split that divides a node into left (I_L) and right (I_R) child nodes, the gain is calculated as:

$$\text{Gain} = \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I_L \cup I_R} g_i)^2}{\sum_{i \in I_L \cup I_R} h_i + \lambda} - \gamma$$

where g_i is the first derivative (gradient) of the loss function with respect to the model's prediction for the i th training example and h_i is the second derivative (Hessian) of the loss function with respect to the model's prediction for the i th training example. The **feature importance** is then quantified by the total gain contributed by each feature across all trees in the model. Features that frequently appear in high-gain splits are considered more important, as they contribute more significantly to reducing the model's error.

We used the XGBClassifier from the **XGBoost** (ver. 3.0.2) Python library [37] and Python ver. 3.10. The original satellite imagery was provided in raster format, where each pixel represents a spatial observation, and each raster layer corresponds to a spectral band or derived variable. Data were extracted from these spatial rasters and transformed into a tabular dataset, with one observation per row, one column for each explanatory variable, and one indicator column containing the outcome for each observation (attacked/health). The model was trained with 450 boosting rounds, a logistic loss, and the *average precision* evaluation metric. A fixed random seed was used to ensure reproducibility. Hyperparameters were selected to balance model complexity and performance while preventing overfitting.

3.4. Evaluation

To evaluate model performance on our imbalanced dataset, we utilized metrics that focus on the minority class (attacked pixels). While traditional accuracy can be misleading in imbalanced contexts, we include it here to allow for direct comparison with existing literature. Table 3 summarizes the evaluation metrics used, their mathematical definitions, and their specific relevance to bark beetle detection.

Table 3. Summary of evaluation metrics used to assess model performance for beetle attack detection. When comparing the prediction of an attack with the actual outcome, we divide the results into the classical cases: TP—true positive; TN—true negative; FP—false positive; and FN—false negative.

Metric	Formula	Ecological Significance
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Included for comparison with other studies.
Precision	$\frac{TP}{TP+FP}$	Minimizes false alarms to ensure resources are not wasted on healthy trees.
Recall	$\frac{TP}{TP+FN}$	Ensures that the maximum number of actual infestations are detected.
F1-Score	$2 \times \frac{Prec \cdot Rec}{Prec + Rec}$	Provides a single measure that penalizes poor performance in either metric.
AUC-ROC	Area under Curve	AUC-ROC (Area Under the Receiver Operating Characteristic curve) measures the model's ability to distinguish between healthy and attacked pixels across all possible classification thresholds.

A calibration curve (or reliability diagram, [38]) is used to assess how well the predicted probabilities from a model reflect the true likelihood of an event. In other words, it measures the agreement between predicted probabilities and observed outcomes. If a model is perfectly calibrated, predictions of probability p should correspond to the observed frequency of the event p . To create a calibration curve, predicted probabilities are typically binned (in this paper we use 10 bins), and for each bin, the mean predicted probability is plotted against the observed fraction of positive outcomes. Deviations from the diagonal line $y = x$ indicate miscalibration.

The Brier score [39] is a numerical measure of calibration and accuracy for probabilistic predictions. It quantifies the mean squared difference between predicted probabilities and actual binary outcomes. Mathematically, for N predictions:

$$\text{Brier Score} = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2$$

where f_i is the predicted probability for observation i and $o_i \in \{0, 1\}$ is the actual outcome. Brier Score ranges from 0 (perfect prediction) to 1 (worst prediction for binary outcomes). An advantage of using the Brier score is that it evaluates both the calibration and refinement of predicted probabilities.

AUC-ROC (Area Under the Receiver Operating Characteristic curve) quantifies the model's ability to discriminate between the two classes across all possible classification thresholds. The ROC curve is constructed by varying the decision threshold and plotting the True Positive Rate against the False Positive Rate. The area under the curve (AUC) can be interpreted as the probability that the model assigns a higher score to a randomly selected attacked pixel than to a randomly selected healthy pixel. An AUC value of 0.5 indicates random performance, while a value of 1.0 corresponds to perfect separability between classes.

To evaluate the contribution of the satellite data and identify the most informative time periods, we began with the initial week and progressively added data from subse-

quent weeks, retraining the XGBoost model at each step to incorporate the new temporal information. This incremental approach allowed us to determine whether certain weeks are particularly important for detecting attacks. Important features of each model were also extracted. For the model with only static features, feature importance was easy to extract. For the models trained on only Sentinel-2 data or a combination of static and Sentinel-2 data, feature importance was found by averaging the importance of each spectral band across all available weeks used in the final model.

4. Results

Following data pre-processing (Section 2) and augmentation (Section 3.2), we trained the XGBoost model using three combinations of the datasets: (1) only static features, (2) only Sentinel-2 data, and (3) a combination of the two. For each Sentinel-2 tile, we standardized the spectral information to mean zero and variance one. Standardization was performed separately for each band and each week.

For all models, we applied stratified five-fold cross-validation to divide the dataset into training and test folds (20% of the data were used in each test fold). The stratification maintained consistent class distributions in all folds, ensuring that the class imbalance ratio is preserved in each split, reduced sampling bias, and provided a more reliable estimate of model generalization under imbalanced conditions. Within each fold, SMOTE was applied only to the training partition, while the corresponding test partition remained untouched and preserved its original distribution. This procedure prevents data leakage from synthetic samples into the evaluation phase and allows the reported metrics to reflect performance on truly unseen, imbalanced data. Cross-validation further ensured that the evaluation was robust to the choice of train–test split and provided stable estimates of model performance.

Our best model, which combined static features and satellite data, achieved an overall accuracy of 98.74%, representing a significant improvement compared to previous studies. However, the main contribution of this paper is the following analysis of feature importance and model performance. A comparison with other classifiers as well as an ablation study comparing different missing-value imputation strategies are provided in Appendices B and C.

4.1. Feature Importance

Feature importance for the different models is illustrated in Figure 4. Among the static features, land cover was by far the most influential predictor, followed by canopy height (CHM) and soil type. When combining satellite data and static features, land cover is still the most important static feature, but it is much less dominant, with the combined model giving a markedly higher weight to landforms, soil type, and spruce volume.

For the model using only Sentinel-2 data, the Red, Blue, and Red-edge bands are the most important. However, the difference between the most and least important bands is small. For the combined model, the Red-edge 3, Blue, and Near Infrared bands are the most important. The combined model also shows a larger difference in feature importance between bands, with the Red-edge 3 band being the most important by a wide margin.

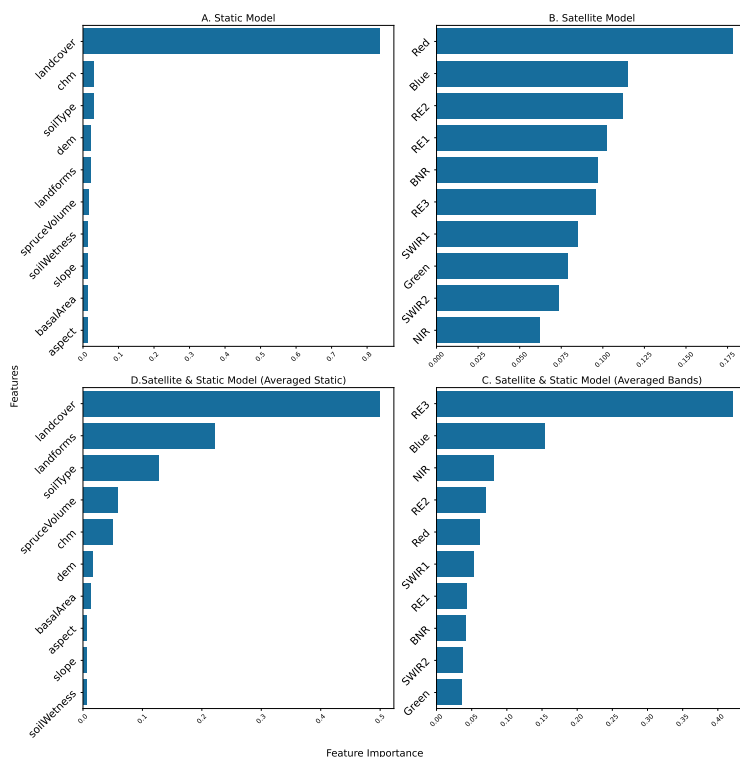


Figure 4. The 10 most important features: (A) using static data, (B) using satellite data, and (C,D) using both satellite and static data, where the mean importance is calculated by averaging the importance of each spectral band across all available weeks used in the final model. Note that feature importance is relative, and it is difficult to compare absolute values between models.

4.2. Model Performance over Time

Following the pre-processing in Section 3.1, 22 weeks of satellite data were retained, covering calendar weeks 15 to 43 (mid-April to late October), with some weeks missing due to cloud cover (see Figure 2). The improvement in model evaluation metrics and ROC curves as more weeks are incrementally added is shown in Figures 3 and 5. As expected, model performance increases as additional satellite data become available.

When satellite data are added to the static features, we generally expect the model's performance to improve or at least remain stable, since it has access to more information. This expectation is supported by our results: the combined model either outperforms (for AUC, F1 score, and recall) or performs as well as (for precision) the static-only model throughout the study period.

In contrast to this, the satellite-only model struggles to match the other models. It reaches parity with the static-only model either late, around week 33 (for AUC and recall) or not at all (for F1 score and precision).

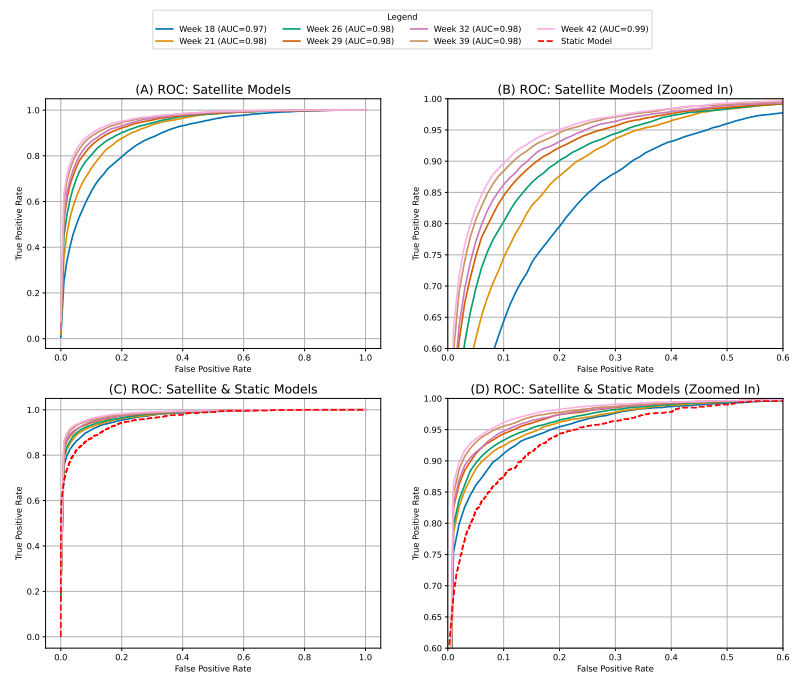


Figure 5. ROC curves comparing the performance of models trained with different input data types across multiple weeks. To improve legibility, only every 3rd week is shown. ROC curves for models trained using only satellite data (A) and zoomed-in view of these ROC curves (B). ROC curves for models trained using both static and satellite data (C), with the static-only model shown as a dashed red line, and zoomed-in view of these ROC curves (D).

Comparison of the Satellite Only and Combined Models

To assess whether the differences between the combined and satellite-only model are statistically significant, we used bootstrap resampling, calibration curves, and Brier scores.

Bootstrap was carried out on the unseen test data by repeatedly drawing samples with replacement and computing the difference in evaluation metrics between the two models. The average paired differences and 95% confidence intervals show that the combined model is significantly better across all metrics; AUC values are shown in Figure 6, and the other metrics can be found in Figure A6 of Appendix D.

Calibration curves were used to examine the relationship between predicted and observed outcomes. A well-calibrated model should produce probabilities that closely match the true frequencies. The calibration curves and Brier scores (see legend) in Figure 7 clearly show that the combined models consistently outperform the satellite-only models. The Brier score directly evaluates how close the predicted probabilities are to the truth, making it particularly useful when working with imbalanced datasets and risk estimation.

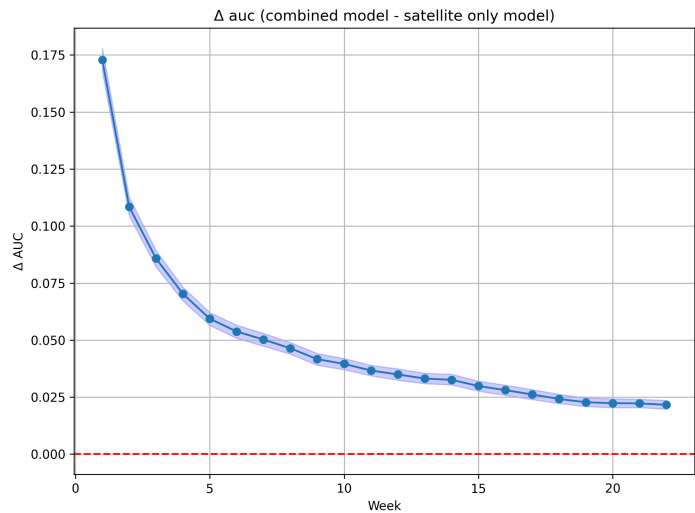


Figure 6. Difference between bootstrapped AUC estimates and paired 95% confidence intervals on the unseen test data. Values above zero (the red dashed line) indicate that the combined model significantly outperforms the satellite-only model.

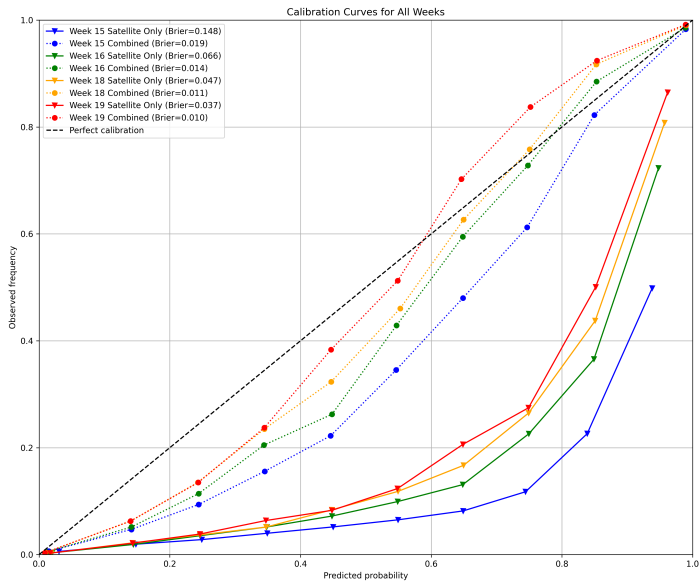


Figure 7. Calibration curves of some representative weeks for satellite-only and combined models; values closer to the 1-to-1 line are better. Brier scores for each model are presented in the legend; lower values are better.

4.3. Most Important Weeks

We used feature importance to systematically evaluate how temporal information contributes to detecting bark beetle attacks and to identify the most important weeks.

For the combined model, using both static features and satellite data, we averaged the importance scores for features belonging to the same week within each model, following the same principle used for band-level aggregation in Section 4.2.

The five most important weeks per model are illustrated using a heatmap; see Figure 8, where the models are represented by columns and the swarming data shown to the right. Several temporal patterns are obvious in the heatmap. Week 19 (swarming week) consistently emerges as highly important, suggesting that it captures early signals of bark beetle activity. Week 25 (six weeks after swarming), and to a lesser extent week 27, initially show high importance. Week 27 is eight weeks after the main swarming, and the second peak in the swarming data suggests that this is when the second generation was swarming. Once data from the highly important weeks 31 and 33 are included in the modeling, the importance of weeks 25 and 27 declines, in contrast to week 19, which remains relevant throughout the period.

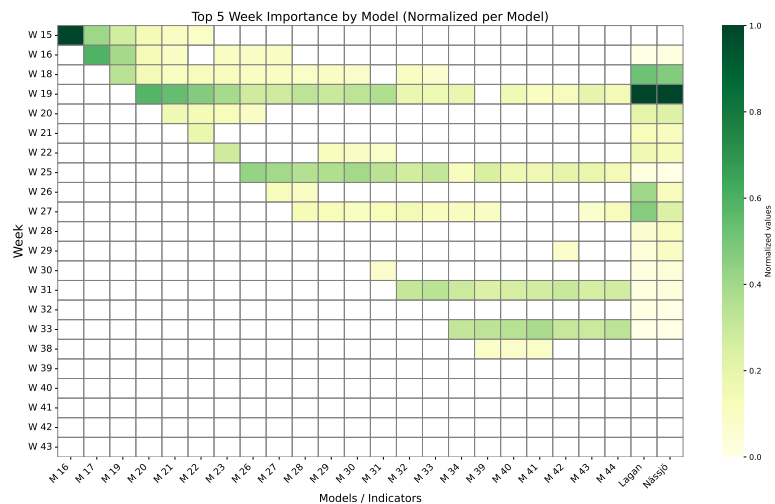


Figure 8. Heatmap showing the five most important weeks per model. The color intensity represents the normalized importance of each week’s features (rows) within the respective model (columns). Note that the lower left part of the heatmap is empty since models are not allowed to use future weeks, i.e., Model 32 only uses data from week 15 to 32. Normalized swarming data are also shown.

4.4. Random Forest as an Alternative to XGBoost

Random Forest is a popular method for classification of satellite data, and for completeness, the analysis was also run using Random Forests. Since the Random Forest achieved similar or worse results at substantially higher computational cost, those results are only briefly presented in Appendix B.

5. Discussion

This study investigates the early detection of bark beetle attacks using a combination of Sentinel-2 satellite imagery and static environmental features within an XGBoost-based framework. The results demonstrate that integrating multi-source data consistently improves predictive performance compared to models relying solely on either static or satellite-derived features. This highlights the complementary nature of structural landscape information and temporally evolving spectral signals. The accuracy of 98.74% obtained by our best model compares favorably to other similar studies [12–15,19,20,24] that report

accuracies in the range of 36% to 96%. However, comparisons of accuracies between studies should be taken with a grain of salt due to differences in datasets and experimental settings.

An important characteristic of our study area is that the bark beetle infestation was less intense and more spatially dispersed compared to the large, spatial contiguous outbreaks common in Central Europe [11]. Thus, the attacked trees in this study often occur as isolated individuals or in small groups, likely resulting weak spectral signals within each Sentinel-2 pixel.

A key finding is that static features provide strong predictive power for the susceptibility to attack in the early stages of the infestation, when spectral changes in vegetation are still weak or not yet observable. Variables such as land cover, landforms, and canopy structure likely capture predisposing environmental conditions that influence tree susceptibility to attack. In particular, land cover remains the most influential feature across all models, although its dominance decreases when satellite data are included. This suggests that while static conditions define baseline vulnerability, temporal spectral observations progressively refine detection by capturing physiological stress responses.

As the season progresses and more Sentinel-2 observations become available, spectral information increasingly contributes to model performance. The Red-edge and Red bands emerge as particularly important, which is consistent with previous studies linking these wavelengths to chlorophyll content and vegetation stress [16,18,40,41]. The strong contribution of Red-edge 3 in the combined model further suggests that subtle physiological changes are better captured when multi-source information is jointly exploited. The Blue band shows moderate importance, aligning with mixed findings in the literature, where its contribution appears to be context-dependent [15,20,24,40].

Temporal analysis of feature importance reveals that specific weeks (19, 25, 31, and 33) contain disproportionately high predictive value. Week 19 corresponds to the peak swarming period, where no visible attack symptoms are expected; its importance likely reflects pre-dispositional stress conditions that increase susceptibility. Week 25, just before the new generation leaves the attacked tree, shows a strong contribution to detection performance, which aligns with reported delays between infestation and observable spectral response in previous studies [17,21]. However, this period is also affected by missing observations in weeks 23–24, which may amplify the apparent importance of week 25.

The combined model shows improved stability in recall during early weeks compared to the satellite-only model, indicating that static features help mitigate missing or weak temporal signals. This suggests that static data provide a regularizing effect, improving robustness under incomplete or noisy temporal observations. Later weeks (31 and beyond) contribute less to performance improvement, likely because most discriminative information has already been captured earlier in the temporal sequence.

Despite these promising results, several limitations should be noted. First, the reliance on SMOTE for balancing may introduce synthetic patterns that do not fully reflect ecological variability. Second, missing satellite observations in certain weeks may bias temporal importance interpretation. Third, while XGBoost provides interpretable feature importance estimates, these should not be interpreted as causal relationships. Finally, Sentinel-2's spatial and spectral resolution limits the ability to detect very early physiological stress signals at the individual tree level.

Overall, the findings demonstrate that combining static and temporal remote sensing data improves both predictive performance and temporal robustness, supporting the value of multi-source data fusion for operational forest monitoring systems.

6. Conclusions

This study shows that combining Sentinel-2 satellite imagery with static environmental features and an XGBoost-classifier results in a model that excels at detecting early bark beetle attacks. The results demonstrate that the model trained on Sentinel-2 satellite imagery and static environmental features consistently outperforms models based on either data source alone, particularly in early detection scenarios when the spectral signals are weak.

Static features provide strong baseline predictive capability by capturing site-specific susceptibility, while Sentinel-2 data become increasingly informative as vegetation stress develops over time. The analysis further indicates that only a subset of the temporal observations carries the majority of the predictive information, highlighting the importance of key phenological and infestation-related periods. Temporal feature importance analysis identified weeks 19, 25, 31, and 33 as particularly informative for prediction. The high importance of week 19, which corresponds to the peak swarming period, suggests that pre-dispositional stress conditions that influence attack susceptibility may already be detectable before visible infestation symptoms appear. Week 25 also contributed substantially to model performance, which is consistent with the expected delay between infestation and observable spectral responses reported in previous studies. In addition, the increased importance of later-season observations reflects the progression of infestation-related canopy stress and discoloration.

Overall, the proposed multi-source framework offers a robust and interpretable approach for monitoring bark beetle infestations. It has potential for operational use in forest management systems, where early detection is critical for reducing ecological and economic damage. Future work should focus on improving robustness to missing temporal observations and exploring more advanced spatio-temporal learning methods for further performance gains.

Author Contributions: Conceptualization, P.-O.O., T.K., M.W. and J.L.; methodology, S.Z., P.-O.O., T.K., M.W. and J.L.; software, S.Z.; validation, S.Z.; formal analysis, S.Z., P.-O.O. and T.K.; investigation, S.Z.; resources, P.-O.O.; data curation, P.-O.O.; writing—original draft preparation, S.Z., P.-O.O. and J.L.; writing—review and editing, T.K. and M.W.; visualization, S.Z.; supervision, T.K., M.W. and J.L.; project administration, J.L.; funding acquisition, P.-O.O. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: SZ and JL have received financial support from the Swedish Strategic Research Area (SRA) MERGE (Modelling the Regional and Global Earth System). The study is also a contribution to the SRA BECC (Biodiversity and Ecosystem Services in a Changing Climate) and to the Lund University profile area Nature-based future solutions. PO was funded by the Swedish National Space Agency, grant Dnr. 2020-00173.

Data Availability Statement: Due to the data use agreement signed with the data provider, the bark beetle attack data cannot be shared publicly. Access to these data is restricted. Sentinel-2 data are freely available; however, without the restricted bark beetle dataset, they cannot be used to fully reproduce the results of this study.

Acknowledgments: Sveaskog, the largest forest owner in Sweden, and the forest owner association Södra are acknowledged for providing harvester data, and the Swedish Forest Agency (Skogsstyrelsen) and the National Forest Inventory (NFI) for providing bark beetle presence data.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Correlation Between Features

This section presents the correlation matrices of spectral bands and static features in Figure A1 and Figure A2, respectively. To quantify relationships between spectral bands, temporal observations were first aggregated by computing the mean correlation values over time.

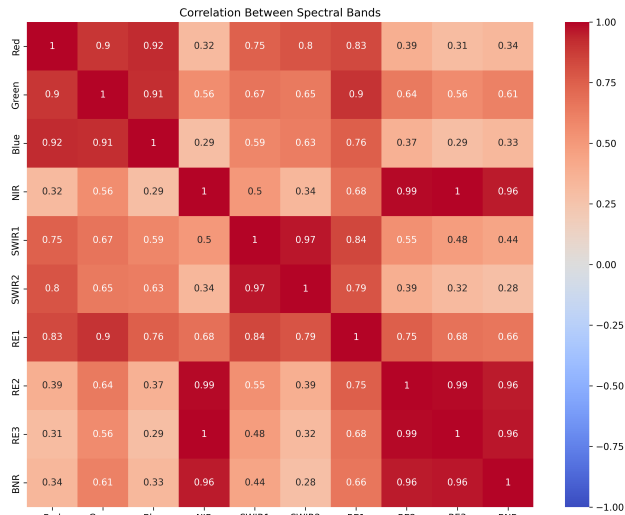


Figure A1. Correlation between bands averaged over the temporal dimension.

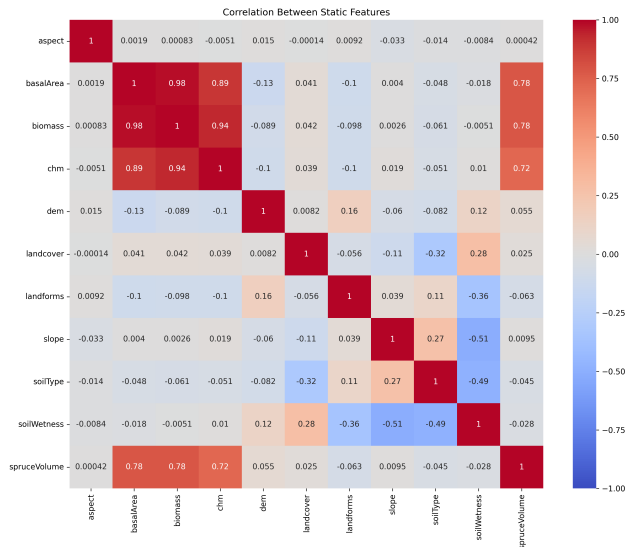


Figure A2. Correlation between static features.

Appendix B. Comparison with Other Classifiers

Appendix B.1. Random Forest

To assess the use of the XGBoost method, the same analysis was performed using Random Forests (RFs); see evaluation metrics in Figure A3. Both methods achieve similar AUC values, suggesting comparable ability to distinguish between pixels with attacked and healthy trees. However, the computational cost of RFs is substantially higher, and the evolution of precision, recall, and F1 score across weeks highlights clear differences in how the methods balance false positives and false negatives.

For both RF and XGBoost, precision increases with more data, indicating improved reliability of positive predictions. However, RF increasingly favors labeling pixels as healthy, reducing false positives at the cost of missing a growing share of attacked trees. Consequently, recall and F1 score decrease or stagnate, further emphasizing the differences between the XGBoost- and RF-based models.

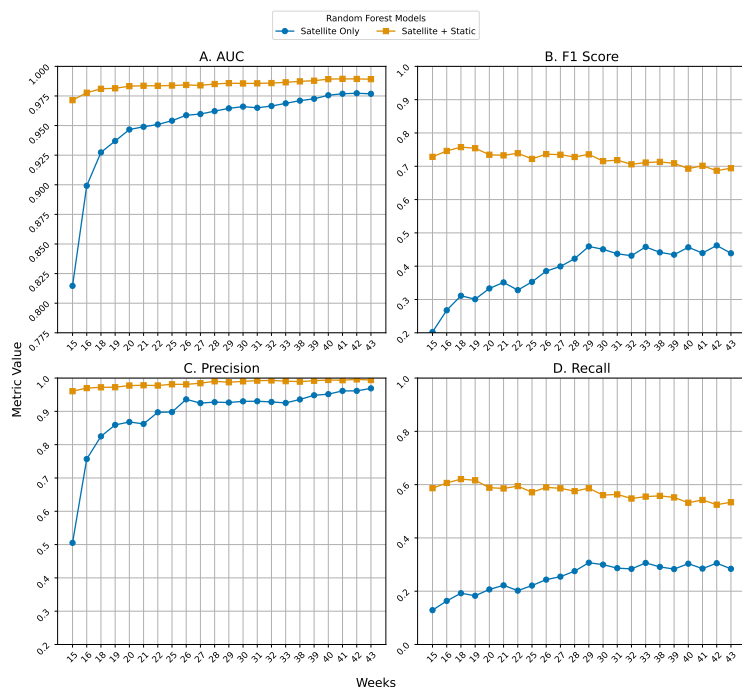


Figure A3. Evaluation metrics of Random Forest models over time using satellite data (blue) and combined satellite and static data (yellow). Overall performance is lower than that of XGBoost (see Figure 3).

Appendix B.2. Deep Learning

The performance of the proposed framework was also compared against a recurrent neural network (RNN)-based [42] baseline. The baseline model consists of a bidirectional Gated Recurrent Unit (GRU) network designed to capture temporal dependencies within the multispectral satellite time series.

The GRU architecture has two stacked recurrent layers with 128 hidden units per direction. By operating bidirectionally, the model processes the temporal sequence in both forward and backward directions, enabling it to learn relationships from both past and

future observations within the input sequence. Dropout regularization is applied between recurrent layers to reduce overfitting.

The output hidden states from all time steps are aggregated using temporal mean pooling. This representation is then passed through a fully connected classification head consisting of two linear layers with ReLU activation and additional dropout regularization. The final layer uses sigmoid activation function to output a single probability.

The RNN-model captured the same temporal patterns as the XGBoost model—see Figure A4—confirming the importance of temporal information for detecting bark beetle-attacked forest stands. However, XGBoost consistently outperformed the RNN across all evaluation metrics.

Several factors may explain these differences. Firstly, the available dataset contains a relatively limited number of attacked samples despite the use of SMOTE. Tree-based ensemble methods, such as XGBoost, are well known for their robustness on limited training samples, whereas deep learning models typically require larger datasets to fully exploit their representational capacity. Secondly, XGBoost can effectively capture non-linear relationships and feature interactions without having to learn any long-term temporal representations. Thirdly, XGBoost is resistant to overfitting through the use of regularization, tree pruning, and ensemble averaging, making it particularly suitable for noisy data with missing-value imputation and class imbalances.

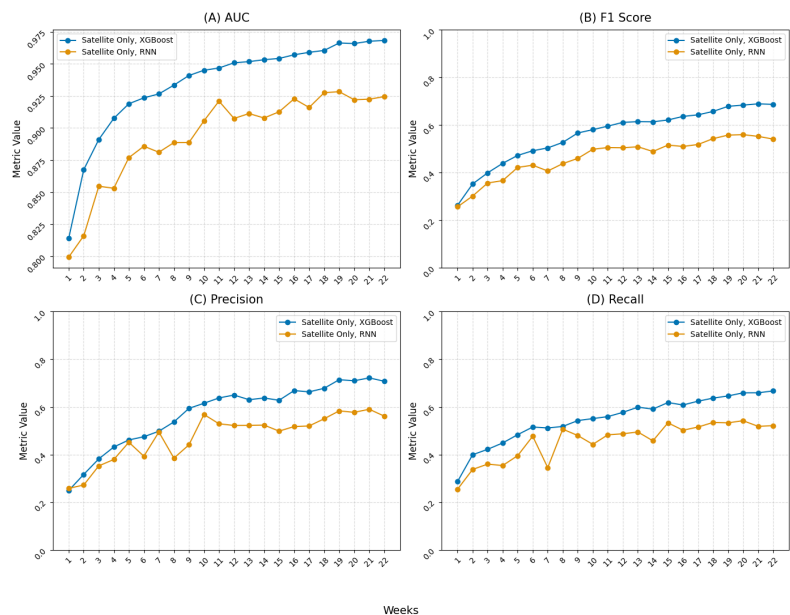


Figure A4. Comparing evaluation metrics of the RNN-deep learning approach to the XGBoost model.

Appendix C. Ablation Study

As discussed earlier, the dataset contains a substantial proportion of missing values. The total number of samples available in each tile is reported in Table 2. When all pixels containing at least one missing observation along the temporal dimension are removed prior to the weekly maximum aggregation step, fewer than ten samples remain in each tile.

This drastic reduction in sample size highlights the necessity of missing-value imputation before any further analysis.

To assess the impact of different strategies for handling missing values and class imbalances, we conducted an ablation study using several different pre-processing approaches; see Figure A5. Specifically, we compared: (1) the original pipeline employing SMOTE for class balancing and temporal forward filling of missing values, (2) spatial, instead of temporal, imputation of the missing-values, and (3) random down-sampling of the majority class (healthy pixels) to match the minority class (attacked pixels). The results indicate that spatial imputation performs poorly due to the sparsity of the available observations, which limits the effectiveness of neighboring pixels for reconstructing missing values. Similarly, down-sampling substantially reduces the amount of training data, degrading model performance. Overall, the original SMOTE-based approach with temporal forward filling provides the most favorable balance between preserving information and addressing class imbalance.

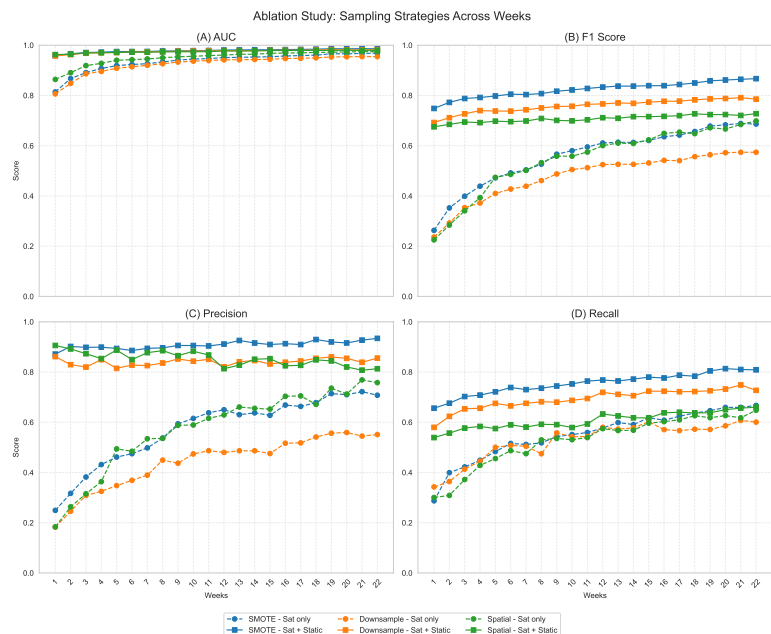


Figure A5. Ablation study evaluating alternative strategies for handling missing values and class imbalance. The baseline approach combines temporal imputation with SMOTE-based oversampling of the minority class. Alternative configurations include spatial imputation and random downsampling of the majority (healthy) class. Spatial imputation yields inferior performance due to the sparsity of observations, while downsampling leads to substantial information loss by discarding a large fraction of healthy samples. The SMOTE-based approach achieves the best overall performance.

Appendix D. Comparison of the Temporal Model

To assess the differences in F1 score, recall, and precision between the satellite-only and combined models, the bootstrap approach described in Section Comparison of the Satellite Only and Combined Models was applied. The mean differences in F1 score, recall, and precision, along with the corresponding 95% confidence intervals, are presented in

Figure A6. This shows that the combined models consistently outperform the satellite-only models across all weeks and metrics.

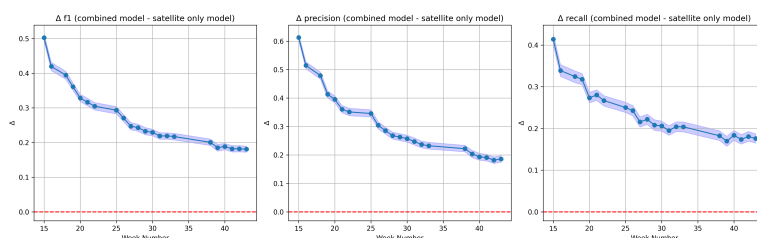


Figure A6. Difference between bootstrapped AUC estimates and paired 95% confidence intervals on the unseen test data. Values above zero (the red dashed line) indicate that the combined model significantly outperforms the satellite-only model.

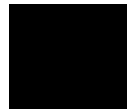
References

- Hysek, Š.; Löwe, R.; Turčáni, M. What happens to wood after a tree is attacked by a bark beetle? *Forests* **2021**, *12*, 1163. [\[CrossRef\]](#)
- Robbins, Z.J.; Xu, C.; Jonko, A.; Chitra-Tarak, R.; Fettig, C.J.; Costanza, J.; Mortenson, L.A.; Aukema, B.H.; Kueppers, L.M.; Scheller, R.M. Carbon stored in live ponderosa pines in the Sierra Nevada will not return to pre-drought (2012) levels during the 21st century due to bark beetle outbreaks. *Front. Environ. Sci.* **2023**, *11*, 1112756. [\[CrossRef\]](#)
- Kurz, W.A.; Dymond, C.C.; Stinson, G.; Rampley, G.J.; Neilson, E.T.; Carroll, A.L.; Ebata, T.; Safranyik, L. Mountain pine beetle and forest carbon feedback to climate change. *Nature* **2008**, *452*, 987–990. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hlásny, T.; König, L.; Krokene, P.; Lindner, M.; Montagné-Huck, C.; Müller, J.; Qin, H.; Raffa, K.F.; Schelhaas, M.J.; Svoboda, M.; et al. Bark Beetle Outbreaks in Europe: State of Knowledge and Ways Forward for Management. *Curr. For. Rep.* **2021**, *7*, 138–165. [\[CrossRef\]](#)
- Raffa, K.F.; Grégoire, J.C.; Staffan Lindgren, B. Chapter 1—Natural History and Ecology of Bark Beetles. In *Bark Beetles*; Vega, F.E., Hofstetter, R.W., Eds.; Academic Press: San Diego, CA, USA, 2015; pp. 1–40. [\[CrossRef\]](#)
- Öhrn, P.; Långström, B.; Lindelöw, Å.; Björklund, N. Seasonal flight patterns of *Ips typographus* in southern Sweden and thermal sums required for emergence. *Agric. For. Entomol.* **2014**, *16*, 147–157. [\[CrossRef\]](#)
- Schelhaas, M.J.; Nabuurs, G.J.; Schuck, A. Natural disturbances in the European forests in the 19th and 20th centuries. *Glob. Change Biol.* **2003**, *9*, 1620–1633. [\[CrossRef\]](#)
- Senf, C.; Pflugmacher, D.; Zhiqiang, Y.; Sebald, J.; Knorn, J.; Neumann, M.; Hostert, P.; Seidl, R. Canopy mortality has doubled in Europe's temperate forests over the last three decades. *Nat. Commun.* **2018**, *9*, 4978. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fernandez-Carrillo, A.; Patočka, Z.; Dobrovolský, L.; Franco-Nieto, A.; Revilla-Romero, B. Monitoring bark beetle forest damage in Central Europe. A remote sensing approach validated with field data. *Remote Sens.* **2020**, *12*, 3634. [\[CrossRef\]](#)
- Schroeder, M. *Granbarkborrens förökningsframgång i dödade träd under sommaren 2022 i sydöstra Småland, Värmland och Uppland/Västmanland*; Sveriges lantbruksuniversitet (SLU): Uppsala, Sweden, 2023.
- Washaya, P.; Modlinger, R.; Tyser, D.; Hlásny, T. Patterns and impacts of an unprecedented outbreak of bark beetles in Central Europe: A glimpse into the future? *For. Ecosyst.* **2024**, *11*, 100243. [\[CrossRef\]](#)
- Abdullah, H.; Skidmore, A.K.; Darvishzadeh, R.; Heurich, M. Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (*Ips typographus*, L.) compared with Landsat-8. *Remote Sens. Ecol. Conserv.* **2019**, *5*, 87–106.
- König, S.; Thonfeld, F.; Förster, M.; Dubovyk, O.; Heurich, M. Assessing combinations of Landsat, Sentinel-2 and Sentinel-1 time series for detecting bark beetle infestations. *Gisci. Remote Sens.* **2023**, *60*, 2226515. [\[CrossRef\]](#)
- Huo, L.; Persson, H.J.; Lindberg, E. Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS). *Remote Sens. Environ.* **2021**, *255*, 112240. [\[CrossRef\]](#)
- Bárta, V.; Lukeš, P.; Homolová, L. Early detection of bark beetle infestation in Norway spruce forests of Central Europe using Sentinel-2. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102335. [\[CrossRef\]](#)
- Fassnacht, F.E.; Latifi, H.; Ghosh, A.; Joshi, P.K.; Koch, B. Assessing the potential of hyperspectral imagery to map bark beetle-induced tree mortality. *Remote Sens. Environ.* **2014**, *140*, 533–548. [\[CrossRef\]](#)
- Huo, L.; Koivumäki, N.; Oliveira, R.A.; Hakala, T.; Markelin, L.; Näsi, R.; Suomalainen, J.; Polvivaara, A.; Junttila, S.; Honkavaara, E. Bark beetle pre-emergence detection using multi-temporal hyperspectral drone images: Green shoulder indices can indicate subtle tree vitality decline. *ISPRS J. Photogramm. Remote Sens.* **2024**, *216*, 200–216. [\[CrossRef\]](#)

18. Huo, L.; Lindberg, E.; Bohlin, J.; Persson, H.J. Assessing the detectability of European spruce bark beetle green attack in multispectral drone images with high spatial-and temporal resolutions. *Remote Sens. Environ.* **2023**, *287*, 113484. [CrossRef]
19. Dalponte, M.; Solano-Correa, Y.T.; Frizzera, L.; Gianelle, D. Mapping a European Spruce Bark Beetle Outbreak Using Sentinel-2 Remote Sensing Data. *Remote Sens.* **2022**, *14*, 3135. [CrossRef]
20. Candotti, A.; De Giglio, M.; Dubbini, M.; Tomelleri, E. A Sentinel-2 Based Multi-Temporal Monitoring Framework for Wind and Bark Beetle Detection and Damage Mapping. *Remote Sens.* **2022**, *14*, 6105. [CrossRef]
21. Jamali, S.; Olsson, P.O.; Müller, M.; Ghorbanian, A. Early Detection of Forest Bark Beetle Attack Using Time Series Spatial Variability of Spectral Indexes from Sentinel-2. In Proceedings of the IGARSS 2024—2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 10157–10160. [CrossRef]
22. Müller, M.; Olsson, P.O.; Eklundh, L.; Jamali, S.; Ardö, J. Features predisposing forest to bark beetle outbreaks and their dynamics during drought. *For. Ecol. Manag.* **2022**, *523*, 120480. [CrossRef]
23. Zhao, P.; Olsson, P.O.; Öhrman Wellendorf, A.; Müller, M.; Mansourian, A. Exploring the spatial heterogeneity of bark beetle infestation risk factors with geographically weighted regression and random forest. *Scand. J. For. Res.* **2025**, *40*, 188–204. [CrossRef]
24. Olsson, P.O.; Zhao, P.; Müller, M.; Mansourian, A.; Ardö, J. Combining Sentinel-2 Data and Risk Maps to Detect Trees Predisposed to and Attacked by European Spruce Bark Beetle. *Remote Sens.* **2024**, *16*, 4166. [CrossRef]
25. Chen, T.; Guestin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: New York, NY, USA, 2016; KDD' 16; pp. 785–794. [CrossRef]
26. Swedish Meteorological and Hydrological Institute. *Climate Normals*; Swedish Meteorological and Hydrological Institute: Norrköping, Sweden, 2025.
27. Nilsson, P.; Roberge, C.; Fridman, J. *Skogsdata 2021: Aktuella uppgifter om de svenska skogarna från SLU Riksskogstaxeringen*; Sveriges lantbruksuniversitet (SLU): Uppsala, Sweden, 2021.
28. European Space Agency. *Copernicus Open Access Hub*; European Space Agency: Paris, France, 2015.
29. Frantz, D. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sens.* **2019**, *11*, 1124. [CrossRef]
30. Rufin, P.; Frantz, D.; Yan, L.; Hostert, P. Operational Coregistration of the Sentinel-2A/B Image Archive Using Multitemporal Landsat Spectral Averages. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 712–716. [CrossRef]
31. Naturvårdsverket. *National Land Cover Database (NMD)*; Naturvårdsverket: Stockholm, Sweden, 2021.
32. Lantmäteriet. *GSD-Terrängkartan, Vektor*, version 5.10; Lantmäteriet: Gävle, Sweden, 2021.
33. Skogsstyrelsen. *Svärningsövervakning*. Available online: <https://www.skogsstyrelsen.se/svarningsovervakning> (accessed on 23 June 2026).
34. Yang, Y.; Xu, J.; Hong, Y.; Lv, G. The dynamic of vegetation coverage and its response to climate factors in Inner Mongolia, China. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 357–373. [CrossRef]
35. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
36. Nitesh, V.C. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321. [CrossRef]
37. Cho, H.; XGBoost Contributors. XGBoost: Extreme Gradient Boosting (ver. 3.0.2). 2025. Available online: <https://pypi.org/project/xgboost/3.0.2/> (accessed on 23 June 2025).
38. Niculescu-Mizil, A.; Caruana, R. Predicting good probabilities with supervised learning. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 625–632.
39. Glenn, W.B. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [CrossRef]
40. Kautz, M.; Feurer, J.; Adler, P. Early detection of bark beetle (*Ips typographus*) infestations by remote sensing—A critical review of recent research. *For. Ecol. Manag.* **2024**, *556*, 121595. [CrossRef]
41. Huo, L.; Lindberg, E.; Fransson, J.E.; Persson, H.J. Comparing spectral differences between healthy and early infested spruce forests caused by bark beetle attacks using satellite images. In Proceedings of the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium; IEEE: New York, NY, USA, 2022; pp. 7709–7712.
42. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Paper IV



Graphical Abstract

Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series

Shokoufa Zeinali, Per-Ola Olsson, Ted Kronvall, Magnus Wiktorsson, Johan Lindström

Highlights

Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series

Shokoufa Zeinali, Per-Ola Olsson, Ted Kronvall, Magnus Wiktorsson, Johan Lindström

- Semi-supervised CVAE integrating environmental priors and spectral dynamics.
- Improved cross-year generalization over traditional Machine Learning benchmarks.

Learning to Disentangle Forest Health: A Conditional Variational Autoencoder Approach to Satellite Time-Series

Shokoufa Zeinali^a, Per-Ola Olsson^b, Ted Kronvall^a, Magnus Wiktorsson^a,
Johan Lindström^a

^a*Centre for Mathematical Sciences, Lund University, Märkesbacken 4, Lund, Sweden*

^b*Dept of Earth and Environmental Sciences, Lund University, Sölvegatan
12, Lund, Sweden*

Abstract

Outbreaks by the European spruce bark beetle (*Ips typographus*) cause massive ecological and economic damage across Europe. Detecting these attacks early is a significant challenge because ground-truth data of which trees are attacked and when are often scarce or unavailable.

This study presents a Conditional Variational Autoencoder (CVAE) designed to detect outbreaks by combining Sentinel-2 time-series data with static environmental features. Unlike traditional machine learning models that require thousands of labeled examples, the CVAE operates in a semi-supervised setting, meaning it does not require temporally aligned same-year labels. We use static environmental features to identify the susceptibility (risk levels) of specific forest areas. Then, by conditioning a Variational Autoencoder model on these risk levels, we train it to reconstruct forest dynamics for healthy and attacked situations. This allows us to effectively separate healthy and attacked areas within the model’s latent space.

Our model was trained on Sentinel-2 data from 2018 and evaluated on data from 2018 and 2019. Our results demonstrate that the model effectively separates healthy and attacked pixels, highlighting its potential as a powerful tool for early detection of bark beetle outbreaks in data-scarce environments.

Keywords: Conditional Variational AutoEncoder, Sentinel-2, Bark Beetle Attack Detection

1. Introduction

Bark beetle outbreaks are among the most destructive biotic disturbances affecting coniferous forests. The European spruce bark beetle (*Ips typographus*) is a phloem-feeding insect whose life cycle and outbreak dynamics are strongly influenced by temperature. Adult beetles locate and attack suitable Norway spruce (*Picea abies*) trees, mate, and lay eggs beneath the bark. The larvae then feed and develop in the phloem, after which the new generation of adults emerges and disperses to colonize new host trees. Successful colonization is facilitated by coordinated mass attacks and, in many cases, by associated fungi that contribute to overcoming host defenses [1]. Temperature-dependent development regulates the timing of important life-cycle stages, including swarming, mating, and larval development, making the species highly responsive to inter-annual climatic variability. In southern Sweden, a new generation of bark beetles typically begins to swarm approximately eight weeks after a tree has been colonized, with the first swarm usually in May [2].

The impacts of bark beetle outbreaks extend beyond tree mortality. They can affect overall ecosystem health, disrupt timber production, and reduce the capacity of forests to store carbon through carbon sequestration [3, 4, 5, 6]. In Europe, the European spruce bark beetle (*Ips typographus*) is one of the main threats to Norway spruce forests. Historically, bark beetle disturbances accounted for approximately 8% of tree mortality caused by natural disturbances between 1850 and 2000 [7]. However, in recent decades, the extent and severity of outbreaks have increased substantially, with climate change contributing to more favorable conditions for beetle development and population growth [8].

The scale of the destruction is massive. For example, in Sweden, recent outbreaks have killed nearly 31 million m^3 of spruce trees, and approximately 86 million m^3 have been lost in Germany; in the Czech Republic, over 50 million m^3 have been destroyed. These losses cost billions of euros in timber value and permanently alter European forests [9, 10, 11]. Given the magnitude of these impacts, there is a critical need for methods that can detect bark beetle attacks at an early stage, before visible symptoms become apparent and management options become limited.

To address this need, we present a time-aware risk-conditioned variational autoencoder (VAE) model for the early detection of bark beetle attacks on spruce trees in Sweden, using Sentinel-2 imagery. The model extends a stan-

dard VAE by replacing the typical fixed latent space prior with a conditional Gaussian prior. The mean of this prior is informed by each pixel’s susceptibility to attack using static features of trees, which helps to define the expected "healthy" state of the latent space. To improve the discriminative power of the model, a cosine similarity loss [12] is introduced during training. This loss encourages the separation of the latent space representations of healthy and attacked trees into distinct clusters. Using these clusters, the model can identify early signs of attack before visible symptoms appear, which is critical for timely forest management.

An additional advantage of our approach is the ability to train in a semi-supervised setting, not requiring temporally aligned same-year labels. This enables the effective use of abundant unlabeled satellite data while still benefiting from limited expert labels when they are available. This flexibility improves both the sensitivity of early detection and the scalability of forest monitoring systems.

2. Related work

Recent studies have demonstrated the effectiveness of satellite-based remote sensing for frequent large-scale monitoring. While traditional multispectral data from Landsat-8 and Sentinel-2 are standard, the choice of spectral indicators and modeling techniques remains a critical factor for early detection.

2.1. Spectral and Physiological Indicators

Vegetation indices (VIs) are commonly used to detect changes in forest health. Research by Abdullah et al. [13, 14] and Huo et al. [15] consistently identifies the Red-edge and SWIR bands as the most sensitive to early physiological stress. These bands capture declines in chlorophyll and water content before symptoms become visible to the human eye. Notably, the Normalized Difference Red and SWIR (NDRS) index developed by Huo et al. [15] achieved detection accuracies of 0.80–0.88 prior to visible attacks, outperforming established indices such as RDI and DSWI. Our suggested model integrates NDRS as a key feature for identifying early-stage infestations.

2.2. Remote Sensing for Bark Beetle Detection

Time-series analysis (TSA) is widely used to model forest phenology and isolate disturbances. Using Sentinel-2, Bárta et al. [16] and Dalponte et al.

[17] achieved accuracies of 78% and 83% in identifying infested trees by monitoring deviations from a healthy baseline.

In Southern Sweden, several studies have validated the precision of Sentinel-2 for local conditions. Jamali et al. [18, 19] applied kernel-based spatial analysis and the Detecting Breakpoints and Estimating Segments in Time-series (DBEST) algorithm, reaching accuracies between 83.8% and 87.9%. Furthermore, Olsson et al. [20] combined Sentinel-2 data with geospatial risk maps to detect both predisposed and attacked trees, highlighting the importance of Blue, Red-edge, and SWIR bands in early detection. These regional findings provide a robust baseline for the environmental conditions analyzed in this paper.

2.3. Deep Learning and Anomaly Detection

Recent research has increasingly turned to Deep Learning (DL) to capture complex temporal dependencies. Models such as Long Short-Term Memory (LSTM) networks have been used to track seasonal patterns and detect anomalies over month-long horizons [21, 22].

For unsupervised change detection, Autoencoders (AEs) have shown significant promise. Zhou et al. [23] utilized time-series of the NBR index within an AE framework to achieve 90.3% accuracy, while Saha et al. [24] proposed LSTM encoder-decoders to detect anomalies without prior assumptions on timing. Most recently, Andresini et al. [25] introduced PHANTASM, which uses spectral-spatial deep embeddings from pre-trained U-Nets to map forest dieback. While these methods reduce the need for labeled data, they often treat all trees as having equal prior probability of attack, potentially ignoring known susceptibility factors.

Andresini et al. [26] introduced the XGBoost-based SILVIA for bark beetle mapping in France. Building on this, DIAMANTE [27] uses U-Net-based semantic segmentation to combine spatial context and spectral information from both Sentinel-1 and Sentinel-2. Schiller et al. [28] showed that transformer architectures applied to Sentinel-2 time-series can use long-range temporal dependencies to detect forest disturbances across Central Europe, outperforming LSTM-based approaches on multi-year datasets. Raspović et al. [29] applied Random Forest (RF), XGBoost, CNN-LSTM, and Temporal Convolutional Neural Network (TempCNN) to Sentinel-2 images from Croatia for early detection of bark beetle attacks, with TempCNN having the highest accuracy.

Schiller and Fassnacht [30] compared four near-real-time change detection methods (FORDEAD, S-CCD, BFAST Monitor, and TimesNet RSAD) on dated bark beetle infestations in German spruce monocultures using Sentinel-2. None achieved satisfactory accuracy within the critical 10-week management window. Extending the window to 14 weeks, the deep-learning-based TimesNet RSAD achieved the highest accuracy ($>80\%$). Schiller et al. [31] reached a similar conclusion using an LSTM Autoencoder trained on over 300,000 undisturbed forest time-series for reconstruction-based anomaly detection. Evaluated on a geographically independent dataset with known infestation dates, the model achieved only moderate performance within 10 weeks (accuracy = 11.8%), improving substantially at 13 weeks (accuracy = 81.5%).

Each of these methods have their own challenges. Classical methods do not incorporate ecological susceptibility knowledge, supervised temporal models require temporally aligned same-year labels, and patch-based architectures such as U-Net require spatially contiguous labeled regions. In our setting, with ground truth collected retrospectively at sparse point locations, the labels lack both temporal alignment and spatial continuity. Unsupervised reconstruction-based methods avoid labeled attacked data but do not condition their detection on known susceptibility factors. Our CVAE framework addresses this combination of constraints by operating at the pixel level with temporally misaligned proxy labels derived from static environmental features, conditioning the latent space on ecological susceptibility through a class-conditional prior, and enabling deployment in future seasons without new annotation effort. To our knowledge, no existing approach simultaneously satisfies all three of these operational requirements.

3. Data and preprocessing

In this section, we describe the datasets used and detail the preprocessing procedures applied.

3.1. Study area

Encompassing an area of $90 \times 180 \text{ km}^2$ in southeastern Sweden, the study area, as detailed in Fig. 1, was divided into two tiles. The region experiences an annual mean temperature of 6 to 8 °C, with summer averages ranging from 15 to 16 °C [32, based on data from 1991-2020]. The landscape is predominantly forested, with Norway spruce (*Picea abies* (L.) Karst) and

Scots pine (*Pinus sylvestris* L.) being the most common species, and only minor patches of deciduous forests [33]. A significant portion of the largest recorded bark beetle outbreak in Sweden, triggered by the drought in 2018, occurred within the study area. This outbreak affected an estimated 31 million m³ of spruce trees in southern Sweden between 2018 and 2022 [10].

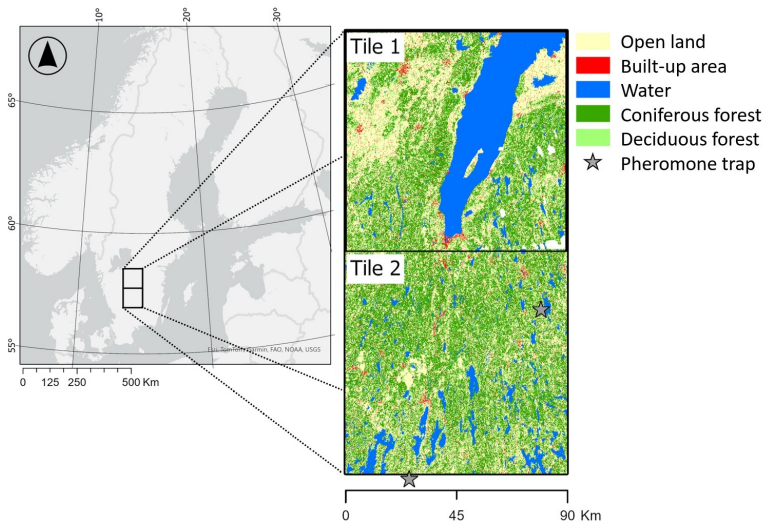


Figure 1: Location of the study area in Sweden, divided into two tiles. Source of the land cover data (left): [Naturvårdsverket \(2018\)](#).

3.2. Response Data — Bark Beetle Damage Records

The response variable in this study was binary data derived from bark beetle damage records provided by Sveaskog, Sweden’s largest forest owner. These records, collected via harvester machines, include geographic coordinates and harvest dates of attacked trees from April 2019 to March 2021. The exact infestation dates are unavailable. Healthy stands were identified using inventory data from the Swedish Forest Agency (Skogsstyrelsen) and supplemented with maps of national land cover [34] and properties [35]. Stands with no harvested trees, located in estates where other stands were attacked, were assumed to be healthy. For further details, see Olsson et al.

[20]. The response data was highly imbalanced, with about 96% healthy pixels. The class imbalance was handled by downsampling of the majority class; see Section 3.5 for details.

3.3. Static features

The static data utilized in this study include topography, soil type, soil wetness, and forestry data (Table 1). Since this information changes slowly over time, it is regarded as static. These static features are used to identify areas that are more susceptible to bark beetle attacks. The static data is described in detail by Müller et al. [36].

Table 1: Description of static features included in the study

Feature	Description
aspect	The direction that a slope faces.
basalArea	A measure of the cross-sectional area of tree trunks in a forest.
biomass	Total mass of living vegetation in an area.
chm (Canopy Height Model)	Height of vegetation canopies.
DEM (Digital Elevation Model)	Elevation data of the terrain, indicating the height above sea level.
Edge	Proximity of a location to the edge of a forest or habitat.
LandC (Land Cover)	Type of land cover.
LandForm	Characterizes the shape of the terrain.
slope	The steepness of the terrain.
soilType	Classification of soil.
soilWetness	Moisture content of the soil. Static variable partly derived from a topographic wetness index.
spruceVolume	Volume of spruce trees.

3.4. Sentinel-2 data

This study uses Sentinel-2 data to capture both spectral and biophysical characteristics of forests. The Sentinel-2 data were obtained as Level 1C, and the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE, ver. 3.6.5, [37]) was used to calculate surface

reflectance. Then the Time-Series Analysis module [38] was applied to derive time-series of individual wavelength bands. Bands with 20 m spatial resolution were split into four 10 m pixels without modifying the pixel values.

To enhance the analytical capacity of the model, six vegetation and water indices that highlight biophysical properties of the landscape were computed from the raw bands. See Table 2 for a summary of bands and indices used in this study. The definitions of the six derived indices are given in Table 3.

Table 2: Sentinel-2 bands and calculated indices used in this study.

Name / Description	Abbreviation	Resolution
Raw Sentinel-2 Bands		
Blue	Blue	10 m
Green	Green	10 m
Red	Red	10 m
Red Edge 1	RE1	20 m
Red Edge 2	RE2	20 m
Red Edge 3	RE3	20 m
Broad Near-Infrared	BNR	10 m
Near-Infrared	NIR	20 m
Shortwave Infrared 1	SWIR1	20 m
Shortwave Infrared 2	SWIR2	20 m
Calculated Indices		
Normalized Difference Vegetation Index	NDVI	10 m
Normalized Difference Water Index	NDWI	10 m
Normalized Difference Red Edge	NDRE	10 m
Normalized Burn Ratio	NBR	10 m
Distance Red-SWIR	DRS	10 m
Normalized DRS	NDRS	10 m

Satellite data, particularly when collected daily, can contain missing values due to cloud cover, atmospheric interference, or sensor issues. To address this, we employed a three-step strategy to ensure a complete and reliable time-series, depicted in Fig. 2.

Before dealing with missing data, all images are normalized to mean-zero and variance-one; this removes any large-scale seasonal trend in the data while maintaining the contrast between spectral bands and indices at each time point. While removing large-scale seasonal trends can seem counter-

Table 3: Definition of the spectral indices used. To avoid division by zero, a small constant, $\epsilon = 1 \times 10^{-6}$, was added to the denominators.

Index	Formula	Reference
NDVI	$\frac{NIR-RED}{NIR+RED+\epsilon}$	Rouse et al. [39]
NDWI	$\frac{NIR-SWIR}{NIR+SWIR+\epsilon}$	Gao [40]
NDRE	$\frac{NIR-RE1}{NIR+RE1+\epsilon}$	Barnes et al. [41]
NBR	$\frac{NIR-SWIR2}{NIR+SWIR2+\epsilon}$	Hecht et al. [42]
DRS	$\sqrt{RED^2 + SWIR1^2}$	Huo et al. [15]
NDRS	$\min(\max(\frac{DRS-q_{0.05}}{q_{0.95}-q_{0.05}}, 0), 1)$	Huo et al. [15]

intuitive, the normalized images ensure that changes in timing of the attacks between years or differences in tree species composition between tiles do not adversely affect detection. The normalization prevents the model from confusing reflectance differences caused by species composition with those caused by attacks.

To handle missing data, any acquisition dates with more than 60% missing pixels were discarded. For acquisition dates with less than 60% missing pixels, the missing pixels were forward filled using the most recent valid observation for that pixel (i.e. data is imputed forward in time to maintain causality). The irregular acquisition dates were then mapped to uniform 5-day intervals, using the earliest available image in each interval. Intervals with no available images were zero-filled, because the normalization above zero-filling is equivalent to mean imputation. A Mann-Whitney test showed no significant differences in the frequency of zero-filling between healthy and attacked pixels in the training set. This is consistent with the scene-level nature of the zero-fill procedure, since entire acquisition dates are discarded; the zero-filling reflects landscape-wide cloud cover events that affect all pixels.

3.5. Train-Test Split

The dataset was divided into spatially blocked training and test subsets; a design meant to respect the natural patch structure of forest stands and bark beetle outbreaks. We applied DBSCAN [43] clustering to all 2,250 labeled attacked pixels, grouping them into 73 spatial clusters with a median size of 20 pixels. Each cluster corresponds roughly to a forest stand or an area of

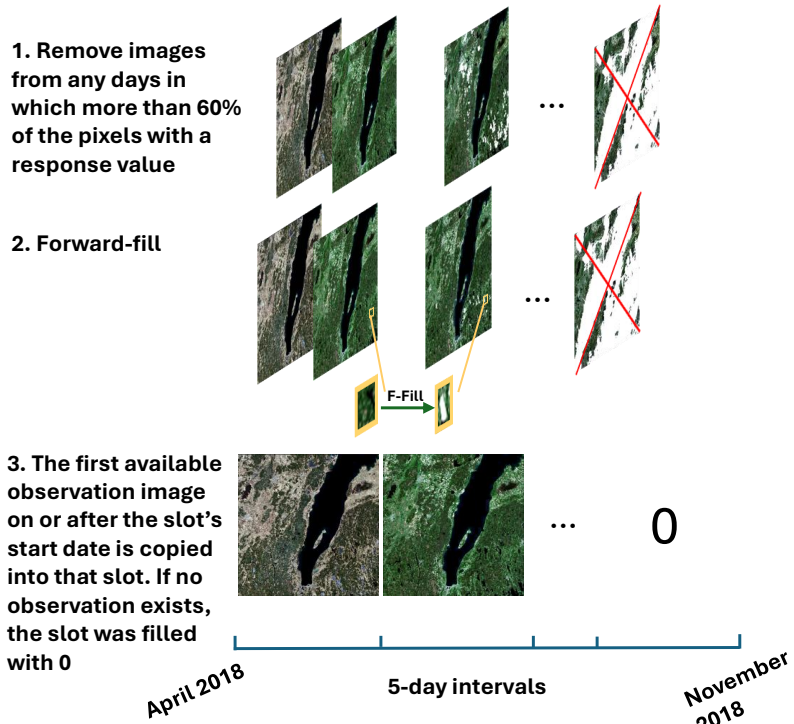


Figure 2: Data Preprocessing Workflow: 0. Images are normalized to mean zero and variance one. 1. Filtering: Sentinel-2 images with $> 60\%$ missing pixels were discarded to ensure data reliability. 2. Imputation: Remaining missing values were filled using forward-fill to maintain temporal continuity. 3. Temporal Alignment: To enable tile merging, the irregular acquisition dates were mapped onto a uniform 5-day interval calendar. This 5-day aggregation ensures a consistent time series for vegetation index calculation and seamless spatial analysis. The intervals with no observation were filled with 0; due to normalization of images, the zero-filling corresponds to mean imputation.

a contiguous bark beetle outbreak. By design, see Section 3.2, the healthy pixels are already grouped into stands, and each stand was seen as a spatial cluster. The clusters were assigned 80/20 to training and test sets. Healthy clusters were assigned to the same partition as the nearest attacked pixel and downsampled to maintain a 1.5:1 healthy-to-attacked ratio within each set. The final split contains 4,137 training pixels (1,655 attacked, 2,482 healthy) and 1,487 test pixels (595 attacked, 892 healthy). The same test set is used for both 2018 and 2019, with training using only data from 2018.

Empirical semivariograms computed on the Sentinel-2 images showed autocorrelation ranges of around 900 to 11,000 m (median 2,600 m) across all spectral features. We verified spatial separation by computing the nearest-neighbour distance from each test pixel to its closest training pixel. The median separation is 3,700 m, exceeding the empirically estimated median autocorrelation range. In total, 63.9% of test pixels are more than 2,600 m from any training pixel.

4. Method

In this section, we go through the details of the methods used.

4.1. XGBoost-based risk estimation from static features

To guide the latent space of our CVAE, we require information about the risk of bark beetle attack for each pixel. Since ground truth labels are often limited or unavailable in satellite-based datasets, we instead generated proxy risk estimates using an XGBoost model trained solely on static features described in 3.3.

XGBoost was selected over alternative classifiers such as Random Forest based on the results in Zeinali et al. [44], which performed a detailed comparison of XGBoost and Random Forest on the same static feature set and study area, together with a full analysis of feature importance. In that study, XGBoost performed equally to or better than Random Forest at a lower computational cost. Among the twelve static features, land cover was by far the most influential predictor of bark beetle susceptibility, followed by canopy height (CHM) and soil type. These findings are consistent with known risk factors for *Ips typographus* attack, where coniferous land cover type, dense canopy structure, and soil conditions associated with drought stress are primary determinants of stand vulnerability.

XGBoost was trained to predict whether a pixel corresponds to a low-risk (healthy) or high-risk (attacked) area using the response variables described in Section 3.2. Once trained, the XGBoost model was used to assign predicted risk labels $y \in \{0, 1\}$ across the entire dataset. These predicted labels serve as a soft proxy for attack risk and are used to condition the VAE’s latent space via a class-conditional prior. Importantly, we avoid using ground truth labels directly, which allows the model to be trained on large regions of satellite data where only static features are available. This design improves scalability and makes the approach practical for real-world forest monitoring, where manual labeling is sparse or infeasible. The predicted risk labels are incorporated into the CVAE, described next.

4.2. Conditional Variational Autoencoder (CVAE)

VAE is a generative model that learns a probabilistic mapping between input data and output by a continuous, lower-dimensional latent space [45]. Unlike traditional autoencoders that learn a direct mapping, VAEs learn the parameters (mean μ and variance σ^2) of a probability distribution (typically Gaussian) in the latent space for each input data point. This probabilistic encoding, coupled with a reparameterization trick, which is explained in Appendix B.5, enables the generation of new and similar data points by sampling from the learned latent distributions. The vanilla VAE objective function comprises two main components: reconstruction loss to measure how accurately the decoder reconstructs the input data from the sampled latent variable, and Kullback-Leibler (KL) divergence to regularize the latent space by minimizing the divergence between the learned latent distribution and a prior distribution, the standard Gaussian distribution.

4.2.1. CVAE Architecture and Latent Space

Our model extends the standard Variational Autoencoder (VAE) by conditioning the latent space on discrete class labels, y , derived from the static feature model (Section 4.1). Unlike a standard VAE that regularizes toward a unit Gaussian, we define the latent space prior as $p(z|y)$. This encourages the model to map healthy and attacked samples toward distinct, learnable centroids, μ_y , enhancing class separability while remaining end-to-end trainable.

4.2.2. Encoder-Decoder Structure

The model follows an encoder-bottleneck-decoder topology designed for multivariate time-series $x \in \mathbb{R}^{D \times F}$, where D represents the number of 5-day

intervals and F denotes the number of spectral features (Section 3.4). A detailed visualization of the architecture is provided in Fig. 3, with specific layer details in Appendix B.

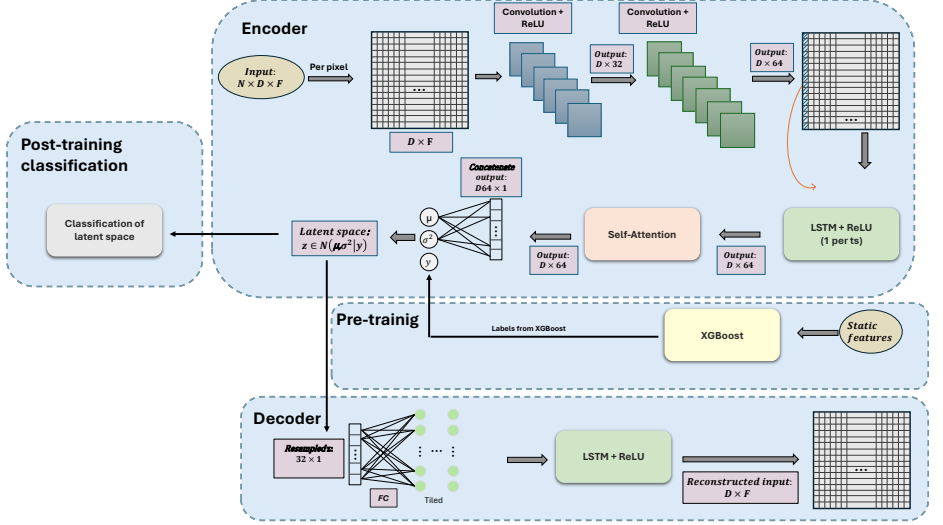


Figure 3: Architecture of the CVAE used in this paper; for details on each layer, see Appendix B.

Encoder: To capture multi-scale temporal dependencies, the encoder integrates three distinct modules:

Temporal Convolution: A 1D-CNN block to model short-term temporal dependencies in the spectral data.

Recurrent Processing: An LSTM network [46] processes these features to maintain a global temporal context across the time-series.

Attention Mechanism: A Multihead Self-Attention layer [47] assigns weights to specific 5-day intervals, allowing the model to prioritize critical phenological windows.

The resulting representation is projected into a 32-dimensional latent space, defined by the distribution $q_\phi(z|x) = \mathcal{N}(z; \mu, \text{diag}(\sigma^2))$.

Decoder: The decoder mirrors the encoder’s complexity to reconstruct the original time-series from the sampled latent vector z . The reconstruction loss, combined with the class-aware KL divergence, ensures that the latent space is both representative of the input signal and discriminative for early-stage infestation detection.

4.3. Loss function

The loss function used in this paper follows the standard VAE formulation with a reconstruction and KL divergence term, and augments it with a class-conditional prior and a cosine similarity-based regularization to impose additional structure on the latent space:

Reconstruction Loss The reconstruction loss measures the discrepancy between the input time-series x and its reconstruction $\hat{x}$ produced by the decoder. This term encourages the model to accurately reproduce the input data. We employ the mean squared error (MSE) as the reconstruction objective:

$$\mathcal{L}_{\text{recon}} = \|x - \hat{x}\|^2.$$

Conditional Kullback–Leibler Divergence To enforce a class-aware structure in the latent space, we introduce a conditional KL divergence term based on the proxy class labels obtained from the static XGBoost model described in Section 4.1. Rather than regularizing the approximate posterior towards a standard Gaussian prior, we adopt a class-conditional prior distribution

$$p(z|y) \sim \mathcal{N}(\mu_y, I),$$

where $y \in \{0, 1\}$ denotes the predicted class label corresponding to healthy or attacked pixels, μ_y is a learnable class-specific mean vector, and the covariance matrix is fixed to the identity matrix for simplicity.

The approximate posterior distribution produced by the encoder is given by

$$q(z|x) = \mathcal{N}(\mu, \Sigma),$$

where Σ is a diagonal covariance matrix. The KL divergence between the approximate posterior and the class-conditional prior admits the following closed-form expression:

$$D_{KL}(N(\mu, \Sigma) || N(\mu_y, I)) = \frac{1}{2} \sum_{i=1}^d (\sigma_i^2 + (\mu_i - \mu_{y,i})^2 - 1 - \log(\sigma_i^2)).$$

This formulation encourages latent representations belonging to the same class to cluster around a shared, learnable centroid, while promoting separation between different classes. Additional computational details are provided in [Appendix B.4](#).

Cosine Similarity Regularization To further enhance the discriminative structure of the latent space, we incorporate an additional cosine similarity-based regularization term following Cai et al. [48]. The cosine similarity loss consists of two components: a positive term that minimizes the angular distance between a latent vector z and the mean of its corresponding class μ_y , and a negative term that penalizes high similarity between z and the means of all incorrect classes $\mu_{y'}$, where $y' \neq y$. The loss is defined as:

$$\mathcal{L}_{\text{Cosine}} = (1 - \cos(z, \mu_y)) + \sum_{y' \neq y} \max(0, \cos(z, \mu_{y'}) - \alpha),$$

where α is a margin hyperparameter. In this study, we set $\alpha = 0.5$. This value imposes a moderate separation constraint between latent representations from different classes, encouraging discriminative structure without excessively restricting the latent space or compromising reconstruction performance. Smaller values of α produce weaker class separation, whereas larger margins risk over-constraining the latent representation.

The model is optimized using a composite of the above loss functions:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + D_{\text{KL}} + \mathcal{L}_{\text{Cosine}},$$

To train this model, we used the Adam optimizer [49] with a learning rate of 10^{-3} and 500 epochs.

4.4. Classification

Post-training, the model provides latent representations of the time-series as Gaussian parameters (μ, σ^2) for each sample. To evaluate latent separability, we leverage the class centroids optimized during training as reference points. Specifically, for each test sample x , we compute the Kullback-Leibler (KL) divergence between its latent distribution $q(z|x)$ and the respective centroids for healthy (KL_{healthy}) and attacked (KL_{attacked}) groups (see [Section 5.1](#)).

From these divergences, we define a relative risk score:

$$\text{Risk Score} = \frac{KL_{\text{attacked}}}{KL_{\text{healthy}} + KL_{\text{attacked}}}$$

where higher scores indicate a greater likelihood of an attack. To separate these classes, we fit a two-component Gaussian Mixture Model (GMM) directly to the distribution of risk scores ([Appendix C.2](#)). The decision threshold is determined analytically as the intersection point of the two Gaussian densities:

$$p_0(x) = p_1(x)$$

where $p_0(x)$ and $p_1(x)$ are the probability density functions of the healthy and attacked components, respectively. This fully unsupervised approach utilizes the intrinsic probabilistic structure of the CVAE latent space, bypassing the need for manual threshold tuning or external clustering algorithms like k -means [\[50\]](#).

5. Results

To evaluate the model’s ability to distinguish between attacked (positive class) and healthy (negative class) pixels, we employ four standard metrics: accuracy, precision, recall, and F1-score. These metrics are calculated on the test data of the time series of 2018, 2019, and static features.

5.1. Classification performance

We compare our model’s performance on the 2018 and 2019 test sets against two baselines. First, a simple XGBoost model trained only on static features, and second, a fully supervised XGBoost model trained on both 2018 satellite time-series and static features ([Section 3.2](#)). The spatially blocked training and test splits are described in [Section 3.5](#).

For the 2019 evaluation, the model trained on 2018 data is applied directly to 2019 Sentinel-2 time-series data without any retraining. Crucially, the 2019 evaluation uses the same test set as used in 2018. Thus, the 2019 test pixels are spatially separated from 2018 training pixels due to the spatial blocking. The response labels for the 2019 evaluation are the same binary healthy/attacked labels derived from the Sveaskog harvest records as used in 2018.

As shown in Table 4, the supervised XGBoost baseline performed very well on the 2018 test data, reaching an F1-score of 0.71. However, its performance crashed when applied to 2019 data; the F1-score dropped to 0.13, and recall fell to 0.07. Even with strong regularization, the supervised model memorized patterns specific to 2018 that are not transferable to 2019.

In contrast, our CVAE model remains more stable. While its 2018 performance is slightly lower than the supervised baseline, it generalizes much better to 2019, maintaining an F1-score of 0.73. This stability is due to two main reasons: our CVAE is guided by static features that do not change significantly between years, rather than relying on year-specific time-series labels, and it must reconstruct the input data while separating classes. This dual task prevents the model from overfitting to the noise of a specific year. These results show that, while supervised models might be more accurate in the short term, the CVAE is more reliable for long-term monitoring across different years.

The metrics reported in Table 4 are computed on test sets rebalanced to a 1.5:1 healthy-to-attacked ratio, which was necessary to obtain stable performance estimates given the severe natural class imbalance of approximately 96:4 in the study landscape. It is important to contextualise how the reported metrics would be expected to change under the natural distribution. Accuracy would increase, since a trivial classifier predicting every pixel as healthy would achieve 96% accuracy. This makes accuracy an uninformative metric in the operational setting, and we do not recommend it as the primary basis for evaluation. Precision would decrease substantially, with 96% healthy pixels; even a model with strong discriminative ability would generate a large absolute number of false positives, considerably suppressing precision. For example, if the model flags 5% of healthy pixels as attacked across a landscape where healthy pixels outnumber attacked pixels by 24 to 1, false positives would dominate the predicted positive set and precision could fall well below 0.5. Recall is unaffected by class distribution, since it measures only the proportion of correctly identified attacked pixels, regardless of how many healthy pixels exist. The recall values reported in Table 4 therefore transfer directly to the operational setting without adjustment. F1-score, which is the harmonic mean of precision and recall, would decrease under the natural distribution in proportion to the drop in precision.

To assess how well the GMM has separated the risk scores, the distribution of risk scores is shown in Fig. 4. The two distinct peaks indicate good separation between health and attacked pixels.

Table 4: Classification Performance Comparison: Proposed Model vs. Baselines. Static + ts refers to the models trained on static features plus Sentinel-2 time-series. All models are trained on 2018 data and evaluated on either 2018 or 2019.

Model	Data/Year	Accuracy	F1-score	Precision	Recall
<i>XGBoost</i>	Static Only	0.66	0.67	0.55	0.87
	Static + ts (2018)	0.82	0.82	0.81	0.84
	Static + ts (2019)	0.62	0.13	0.68	0.07
<i>Proposed CVAE</i>	ts (2018)	0.71	0.71	0.73	0.74
	ts (2019)	0.73	0.73	0.71	0.70

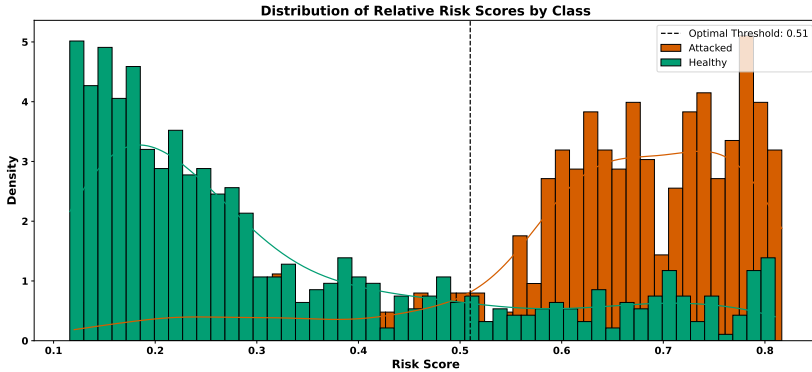


Figure 4: Distribution of the GMM-derived risk scores, illustrating the separation of observations across the classes.

5.2. Effect of reduced temporal information

To investigate the effect of limited temporal information on detection performance, we iteratively masked the most recent days with 0 and re-evaluated the model. This reflects an operational setting where only data up to a certain time-point is known, with future unknown values being replaced by zeros (i.e. mean-imputation given the normalization of images). Fig. 5 shows how the evaluation metrics change as fewer days are available in 2018. To estimate the timing of the attack, we used Bark beetle swarming data, based

on pheromone traps, from the Swedish Forest Agency [51]. Two pheromone traps were located within the study area: one at the southern edge (Lagan) and one in the central region (Nässjö) of the area (Fig. 1). The traps were checked and emptied weekly, and in 2018 the swarming data showed a pronounced peak in week 19, suggesting that this was the most likely period of tree attack. The normalized swarming data are shown in Fig. 5 and Fig. 6.

A consistent increase in classification performance is observed when more days become available. In particular, the model mainly predicts pixels as healthy at the beginning of the study period, reflecting the real-world situation when relatively few trees were under attack. There is a jump in May, which is when attacks usually occur according to Öhrn [52]. This prediction pattern explains the behavior of the evaluation metrics: precision remains high because when the model predicts an attack, it is usually correct (few false positives). Accuracy also stays high since the majority class (healthy trees) is correctly identified. However, recall is low (large number of false negatives) and then jumps sharply when the model starts detecting truly attacked pixels in May. This leads to a similar behavior in the F1-score, which balances precision and recall.

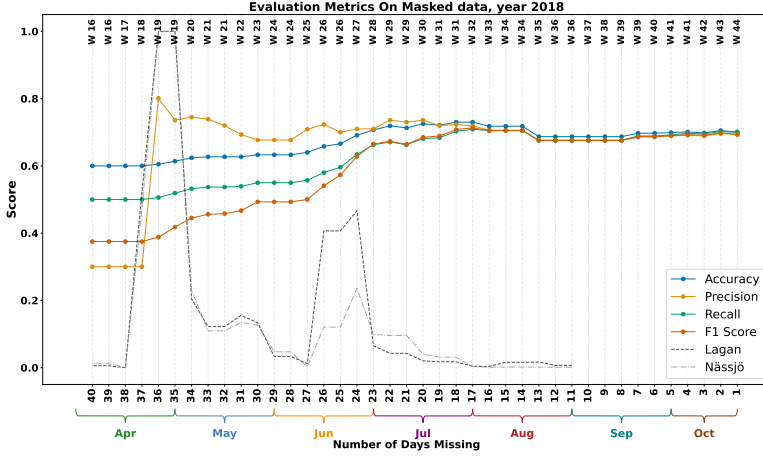


Figure 5: Performance of the proposed model versus the number of available days, for 2018. Here, days mean 5-day intervals. The swarming activity recorded in the pheromone traps at Lagan and Nässjö is represented by gray/black lines; the data are normalized so that a value of 1 equals the highest number of trapped bark beetles (6,000 at Lagan and 11,885 at Nässjö).

Fig. 6 illustrates the results of temporal masking for 2019. Early in the season, the model had difficulty detecting attacked pixels, primarily due to the limited number of available satellite images. However, as the season progressed and more observations became available, the model's detection performance improved considerably.

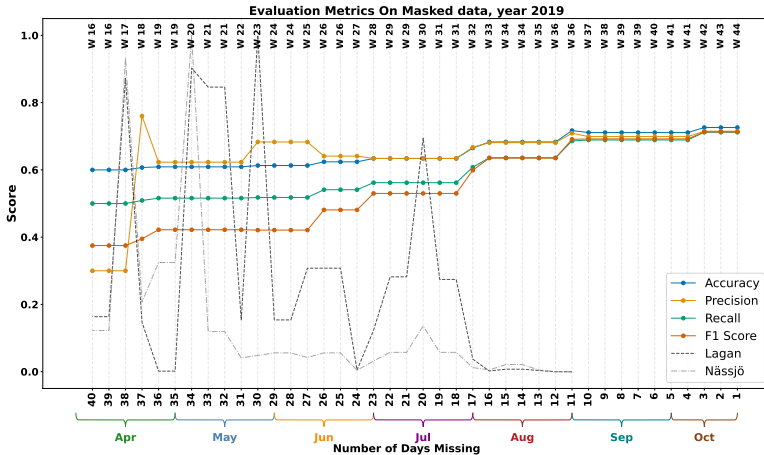


Figure 6: Performance of the proposed model versus the number of available days, for 2019. The swarming data recorded in the pheromone traps at Lagan and Nässjö are represented by gray/black lines; the data are normalized so that a value of 1 equals the highest number of trapped bark beetles (2,600 at Lagan and 8,833 at Nässjö).

5.3. Clustering of latent space in test data

To explore the separability of the CVAE latent space, we applied dimensionality reduction techniques on the latent means of the test data. The goal was to see whether the classes generated by this model could be distinguished in lower-dimensional projections. To this end, t-SNE was used (see [Appendix C](#) for details).

For t-SNE, we used 5 components to retain multiple latent directions, although for visualization, we only show the components that distinguish the classes clearly in the first row of Fig. 7. To complement the visual assessment of the latent space, class separability was quantified using the Davies–Bouldin index [53], where lower values indicate more compact and better-separated clusters. The index decreased from 1.777 for t-SNE components 1–2 to 1.467 for components 1–3 and 1.283 for components 1–4, indicating the clearest separation in the latter projection.

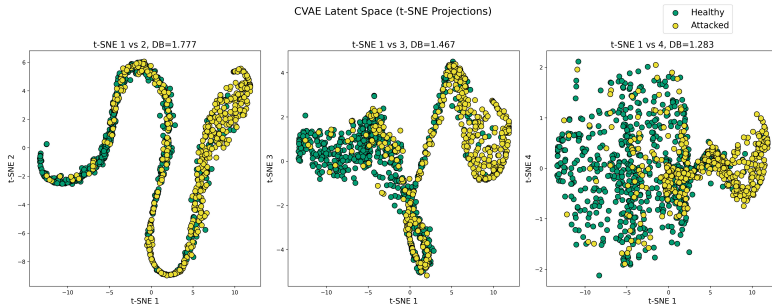


Figure 7: Visualization of the CVAE latent space using t-SNE projections. The figure shows t-SNE projections (components 1 vs 2, 1 vs 3, and 1 vs 4).

6. Conclusion

Using Sentinel-2 data, Bárta et al. [16], Dalponte et al. [17], and Jamali et al. [18, 19] reported accuracies from 78% to 87.9% when detecting infested trees by identifying differences from healthy trees. Olsson et al. [20], on the other hand, did not use time-series data. Instead, they used data from a single date and compared it across one year. They combined spectral data with geospatial risk maps and showed that Blue, Red-edge, and SWIR bands are important for early detection. In deep learning, Zhou et al. [23] used NBR time-series with an autoencoder and achieved a higher accuracy of 90.3%. These results provide a useful baseline for the conditions studied in this paper.

In this study, a Conditional Variational Autoencoder (CVAE) was used in a semi-supervised setting to detect bark beetle outbreaks. The semi-supervised aspect arises from the use of proxy class labels, obtained from static environmental features, which guide the model without requiring direct temporally aligned same-year ground-truth annotations of attacks. The model then learns how vegetation normally changes over time and predicts future states based on past observations and these environmental features, capturing both short-term changes and longer-term trends in forest health.

The CVAE achieved an accuracy of 0.71 (F1-score: 0.71, precision: 0.73, recall: 0.74) in 2018 and 0.73 (F1-score: 0.73, precision: 0.71, recall: 0.70) in 2019. These results are similar to those from traditional methods, although slightly lower than the deep learning approach by Zhou et al. [23]. It should

be noted that a direct comparison of accuracies is difficult due to differences in study design, available data, and class imbalances. However, an important advantage is that the CVAE does not need temporally-aligned labeled training data.

Compared to previous studies, a key strength of this work is the use of full time-series data. While Olsson et al. [20] showed that single-date data can be useful, the CVAE uses changes over time. The temporal masking studies (Fig. 5, Fig. 6) also show that the method can be used operationally with only data up to a given time point being available. This makes it easier to detect early signs of attack, which are often small and hard to see in single images.

The 2018 results also show that the model depends on how much data is available. When only a few days of data are used, the model has low recall but high precision. This means that it misses many early attacks but is usually correct when it predicts an attack. As more data becomes available, recall improves. This matches the real behavior of bark beetle outbreaks, where early changes are small and become clearer over time.

In 2019, the model shows stable performance. Both precision and recall are relatively high once enough data is available. This may be because there were fewer or less complex attacks that year, making detection easier.

Overall, the CVAE provides a useful alternative for monitoring forest health, especially when labeled data is limited. Even though it does not reach the highest accuracy, it performs well and can be used for large-scale early detection by combining time-series data with environmental information.

Appendix A. Sensitivity studies

Appendix A.1. Evaluating healthy pixels

To further validate the model, we provide it with data from a completely unseen source that contains only healthy samples in 2018. Since we know the ground truth is healthy, any high-risk score generated by the model is a mistake. This test measures the model’s baseline noise level and its ability to correctly identify normal behavior without being influenced by the presence of a second class. The method of getting risk scores is described in Section 5.1. The resulting distribution of risk scores, visualized in Fig. A.8, demonstrates that the model consistently assigns low-risk values to healthy samples.

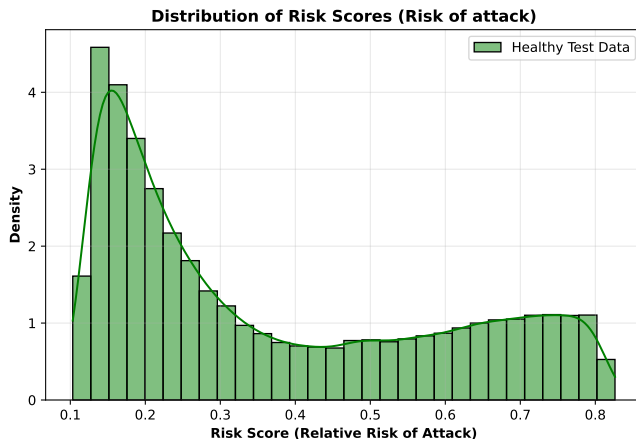


Figure A.8: Histogram of risk of attack in new health-only data.

Appendix A.2. Ablation study

To assess the contribution of each component of the proposed architecture, we evaluated three ablation variants against the full CVAE on the 2018 and 2019 test sets. All variants share the same CNN-LSTM-Attention encoder and are evaluated under identical train-test splits. The results are summarised in Table A.5. The isotropic VAE, which removes both the class-conditional prior and the cosine similarity loss and regularises the latent space toward a standard Gaussian $N(0, I)$, achieves near-perfect recall (0.99 in 2018, 1.00 in 2019) but very low precision (0.40 and 0.40 respectively), resulting in an F1-score of 0.57 in both years. This behaviour indicates that without class conditioning, the model cannot form discriminative latent clusters and effectively classifies almost all pixels as attacked to maximise recall. While the cross-year stability is high, it reflects a degenerate solution rather than genuine generalisation.

Removing only the cosine similarity loss while retaining the class-conditional prior produces the weakest overall result, with F1-scores of 0.38 in 2018 and 0.49 in 2019. Both recall and precision are substantially lower than the full model, indicating that the cosine similarity regularisation plays a critical role

in forming well-separated latent clusters. Without it, the class-conditional prior alone is insufficient to achieve reliable class separation, likely because the KL divergence term provides only a soft pull toward class centroids without explicitly penalising angular overlap between healthy and attacked representations. Notably, 2019 performance is higher than 2018 for this variant, which may reflect that the latent space without cosine regularisation is less sensitive to year-specific spectral patterns and therefore generalises more easily, albeit at a lower absolute level.

The direct CNN-LSTM-Attention classifier, trained in a fully supervised manner on ground truth labels, achieves competitive 2018 performance (F1: 0.75) but shows a substantial drop in 2019 (F1: 0.54), with recall falling from 0.71 to 0.42. This cross-year degradation mirrors the pattern observed for the supervised XGBoost baseline in Table 4 and confirms that fully supervised temporal models tend to overfit to year-specific spectral dynamics.

The full CVAE achieves the best overall performance across both years and the best cross-year performance among all the tested variants.

Table A.5: Ablation study results on the 2018 and 2019 test sets. All models use the same CNN-LSTM-Attention encoder and are evaluated on the same parent-blocked train/test split. The threshold for each variant is selected on the 2018 training set via a precision-recall curve and applied without modification to both years.

Model	Year	Accuracy	Precision	Recall	F1
VAE Isotropic prior	2018	0.40	0.40	0.99	0.57
	2019	0.40	0.40	1.00	0.57
CVAE without cosine loss	2018	0.25	0.28	0.57	0.38
	2019	0.34	0.35	0.81	0.49
CNN-LSTM-Attn Classifier (supervised)	2018	0.81	0.79	0.71	0.75
	2019	0.71	0.76	0.42	0.54
Full CVAE (proposed)	2018	0.71	0.71	0.73	0.74
	2019	0.73	0.73	0.71	0.70

Appendix B. Model details

In this section, we explain the details of the blocks used in the proposed model.

Appendix B.1. Temporal CNN

The model starts with a 1D Convolutional Neural Network (CNN) block LeCun et al. [54], where the input is the multivariate time-series of $D = 41$ five-day intervals for each pixel. The convolutional layers act as temporal feature extractors, scanning across the time dimension to identify local patterns and short-term dependencies between spectral bands. In the context of multispectral satellite data, this allows the model to capture subtle short-term variations in reflectance values that may be indicative of early stress responses. The CNN block consists of two 1D convolutional layers, each followed by a ReLU activation function. Both layers use a kernel size of 3 and padding of 1, preserving the temporal dimension of D throughout. For an input sequence of $\mathbf{x} \in \mathbb{R}^{D \times F}$, the output of the convolution at time t with kernel size 3 and filter j is given by:

$$h_t^{(j)} = \text{ReLU}\left(\sum_{f=1}^F \sum_{\tau=0}^2 W_{\tau,f}^{(j)} x_{t+\tau,f} + b^{(j)}\right),$$

where $W^{(j)} \in \mathbb{R}^{k \times F}$ and $b^{(j)}$ are the learnable weights and bias of filter j , and the ReLU activation function is

$$\text{ReLU}(x) = \max(0, x).$$

The first layer maps the 16 input channels to 32 feature maps, and the second layer maps the 32 to 64. The ReLU activation does not change the dimensions. A kernel size of 3 at 5-day temporal resolution corresponds to a 15-day temporal window per layer, which is reasonable since the bark beetle’s green attack phase produces detectable spectral changes during two to six weeks [52]. Hence, the output of the first convolutional layer for time t is $h^{(1)} \in \mathbb{R}^{32 \times D}$. The output of the second convolutional layer is $h^{(2)} \in \mathbb{R}^{64 \times D}$. The final output of the CNN block is then transposed to $\mathbf{h} \in \mathbb{R}^{D \times 64}$ for use in the subsequent layers.

Appendix B.2. LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) well-suited for modeling sequential data and capturing long-range dependencies. The output from the CNN block is passed to an LSTM network [46]. It processes the feature sequences from the CNN, maintaining an internal memory state that summarizes the temporal context up to each time

step. The LSTM receives the full $D \times 64$ CNN feature sequence and returns the complete sequence of hidden states of shape $(D, 64)$, one per time step.

At each time step t , given an input vector $\mathbf{x}_t \in \mathbb{R}^{64}$ (the CNN features), the LSTM updates its hidden state and cell state $\mathbf{h}_t \in \mathbb{R}^{64}$ and cell state $\mathbf{c}_t \in \mathbb{R}^{64}$ using the following gating mechanism:

$$\begin{aligned}\mathbf{f}_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\ \mathbf{i}_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ \mathbf{o}_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ \tilde{\mathbf{c}}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c), \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh \mathbf{c}_t\end{aligned}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function, $\odot$ denotes element-wise multiplication, and W , U and b are learnable parameters. The output of the LSTM across the temporal dimension has a dimension of $D \times 64$, where $D = 41$ is the number of 5-day intervals as before, and 64 is the hidden dimension.

Appendix B.3. Multihead Self-Attention

Following the LSTM, a Multihead Self-Attention mechanism is applied [47]. The multihead self-attention layer receives the full $D \times 64$ LSTM output as query, key, and value, producing a context-aware attended sequence of the same shape $D \times 64$. This allows the model to assign different weights to different 5-day intervals across the full season when forming the latent representation.

The self-attention process for a single head can be broken down into three main steps:

1. **Projecting into Query, Key, and Value Spaces:** For each hidden state vector h_i from the LSTM's output, it is linearly projected into three distinct vectors: a **query** (q_i), a **key** (k_i), and a **value** (v_i). This is achieved by multiplying the hidden state by three different learned weight matrices (W^Q , W^K , and W^V).

$$\begin{aligned}q_i &= h_i W^Q \\ k_i &= h_i W^K \\ v_i &= h_i W^V\end{aligned}$$

2. **Calculating Attention Scores:** A score is calculated for each query vector (q_i) against all key vectors (k_j). This is typically done using the dot product, which measures the similarity between the two vectors. The scores are then scaled by the square root of the key dimension, d_k , to stabilize training.

$$\text{Score}_{ij} = \frac{q_i \cdot k_j}{\sqrt{d_k}}$$

These scaled scores are then normalized using a softmax function to produce the final attention weights α_{ij} :

$$\alpha_{ij} = \frac{e^{\text{Score}_{ij}}}{\sum_{l=1}^D e^{\text{Score}_{il}}}$$

3. **Computing the Weighted Sum:** The attention weights (α_{ij}) are used to compute a weighted sum of all the value vectors (v_j). This creates a new, contextualized vector for each position in the sequence.

$$o_i = \sum_{j=1}^D \alpha_{ij} v_j$$

The result is a new sequence of context-aware vectors, $\mathbf{O}_{\text{Attention}} \in \mathbb{R}^{D \times H'}$, where the attention weights have focused the representation on the most relevant information. The value of H' is the hidden dimension divided by the number of attention heads ($H' = 64/16 = 4$).

Appendix B.3.1. Multi-Head Attention

The process described above is for a single attention head. The ‘MultiheadAttention’ module runs this in parallel using r different heads, each with its own set of learned projection matrices. The outputs from all heads are concatenated and passed through a final linear layer to produce the final attention block output, which has the same dimensions as the LSTM output, $(D, 64)$. That is

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_r) W^O,$$

where head_i is the result of a single attention head and W^O is a learned weight matrix.

Appendix B.4. Kullback-Leibler (KL) divergence

The conditional KL divergence function in the model calculates the Kullback-Leibler (KL) divergence between the learned latent distribution for an input data point and a class-specific prior distribution. Specifically, it computes

$$KL(\mathcal{N}(\mu, \Sigma) || \mathcal{N}(\mu_y, I)),$$

where $\mathcal{N}(\mu, \Sigma)$ is the approximate posterior distribution learned by the encoder, with mean vector μ and a diagonal covariance matrix, Σ ; and $\mathcal{N}(\mu_y, I)$ is the class-specific prior distribution, with mean vector μ_y (the learned mean for class y) and an identity covariance matrix, I .

Let's denote the two multivariate Gaussian distributions as $P \sim \mathcal{N}(\mu_0, \Sigma_0)$ and $Q \sim \mathcal{N}(\mu_1, \Sigma_1)$. The general formula for the KL divergence between two multivariate Gaussian distributions is

$$KL(P||Q) = \frac{1}{2} \left(\text{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T \Sigma_1^{-1} (\mu_1 - \mu_0) - k + \ln \left(\frac{\det(\Sigma_1)}{\det(\Sigma_0)} \right) \right),$$

where k is the dimension of the random variable (in our case the size of the latent space, i.e. $k = 32$).

For our specific distributions, we now have

$$\begin{aligned} \mu_0 &= \mu, & \Sigma_0 &= \text{diag}(\sigma_i^2), & \det(\Sigma_0) &= \prod_{j=1}^k \sigma_j^2, \\ \mu_1 &= \mu_y, & \Sigma_1 &= \Sigma_1^{-1} = I, & \det(\Sigma_1) &= 1, \end{aligned}$$

where $\text{diag}(\sigma_i^2)$ is a diagonal matrix with elements σ_1^2 to σ_k^2 on the diagonal. This allows us to simplify the terms of the general KL divergence formula:

$$\begin{aligned} \text{tr}(\Sigma_1^{-1}\Sigma_0) &= \text{tr}(\Sigma_0) = \sum_{j=1}^k \sigma_j^2, \\ (\mu_1 - \mu_0)^T \Sigma_1^{-1} (\mu_1 - \mu_0) &= (\mu_y - \mu)^T I (\mu_y - \mu) = \sum_{j=1}^k (\mu_j - \mu_{y,j})^2, \\ \ln \left(\frac{\det(\Sigma_1)}{\det(\Sigma_0)} \right) &= \ln(1) - \ln \left(\prod_{j=1}^k \sigma_j^2 \right) = - \sum_{j=1}^k \log \sigma_j^2. \end{aligned}$$

Combining these terms gives us the KL divergence as

$$\begin{aligned} KL(\mathcal{N}(\mu, \sigma^2) || \mathcal{N}(\mu_y, I)) &= \frac{1}{2} \left(\sum_{j=1}^k \sigma_j^2 + \sum_{j=1}^k (\mu_j - \mu_{y,j})^2 - k - \sum_{j=1}^k \log \sigma_j^2 \right) = \\ &= \frac{1}{2} \sum_{j=1}^k \left(\sigma_j^2 + (\mu_j - \mu_{y,j})^2 - 1 - \log \sigma_j^2 \right) \end{aligned}$$

Appendix B.5. Reparameterization trick

The reparameterization trick separates the randomness from the model’s parameters [45, 55]. Instead of directly sampling a latent variable z from the learned distribution $\mathcal{N}(\mu, \sigma^2)$, the process is rewritten as a deterministic function of the mean μ , the standard deviation σ , and a random sample ϵ drawn from a standard normal distribution $\mathcal{N}(0, 1)$.

The latent variable z is computed as:

$$z = \mu + \epsilon \cdot \exp\left(\frac{1}{2} \log(\sigma^2)\right).$$

The randomness is now isolated in ϵ , which is external to the network. Allowing the gradients to pass through the deterministic parts of the equation to the parameters μ and σ , which can be updated using gradient descent.

Appendix C. Dimension reduction methods

Appendix C.1. t-SNE

High-dimensional data often lie on lower-dimensional manifolds that are not easily visible in raw space. t-SNE is a nonlinear dimensionality reduction technique specifically designed to visualize high-dimensional data in two or three dimensions while preserving local neighborhood relationships [56]. It is particularly useful for revealing clusters that may not be visible in linear projections. t-SNE converts pairwise similarities between data points into probabilities and tries to match these probabilities between the high-dimensional space and the low-dimensional embedding.

- In the original space, nearby points should have high similarity (high probability of being neighbors).
- In the embedding space, the algorithm arranges points so that these neighborhood probabilities are reproduced as closely as possible.

This is achieved by minimizing a divergence measure between the two probability distributions:

For a dataset of points $\{x_1, x_2, \dots, x_N\}$ where each $x_i \in \mathbb{R}^d$, Similarity of point x_j to x_i is modeled as a conditional probability:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$

Here σ_i^2 is a variance parameter chosen so that the perplexity of the distribution matches a user-defined value (perplexity controls the effective number of neighbors). The joint probability is then symmetrized:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N}$$

Thus, p_{ij} is high if x_j and x_i are close in the original space. For $\{y_1, y_2, \dots, y_N\}$ low-dimensional representations, similarities are modeled with a Student t-distribution with one degree of freedom (a Cauchy distribution):

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}$$

This heavy-tailed distribution solves the “crowding problem”: in low dimensions, it prevents distant points from being forced too close together. The algorithm minimizes the Kullback–Leibler divergence between the two distributions:

$$KL(P \parallel Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

If two points are close in high-dimensional space (p_{ij} large), but not close in the embedding (q_{ij} small), the KL term heavily penalizes the mismatch.

Conversely, distant points (low p_{ij}) exert little influence, making t-SNE primarily focused on preserving local structure.

Optimization is performed using gradient descent, often with momentum and early exaggeration to help clusters form.

Appendix C.2. Gaussian Mixture Models

Gaussian Mixture Models (GMMs) [57] are a probabilistic approach for clustering and density estimation. They assume that the data is generated from a mixture of several Gaussian distributions, each representing a latent

component or cluster. Unlike linear techniques, GMMs provide a flexible model capable of capturing multimodal and overlapping distributions.

Let the dataset consist of N samples $\{x_1, x_2, \dots, x_N\}$, where $x_i \in \mathbb{R}^d$. In a GMM with K components, the probability density of x_i is given by:

$$p(x_i) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k),$$

where π_k is the mixture weight of the k -th component, with $\sum_{k=1}^K \pi_k = 1$ and $\pi_k \geq 0$; $\mu_k \in \mathbb{R}^d$ is the mean vector of the k -th Gaussian; $\Sigma_k \in \mathbb{R}^{d \times d}$ is the covariance matrix of the k -th Gaussian; and $\mathcal{N}(x \mid \mu, \Sigma)$ denotes a multivariate Gaussian density.

Each data point x_i is assumed to belong to an unknown class with posterior probabilities that x_i belongs to class k given by

$$\gamma_{ik} = P(z_i = k \mid x_i) = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}.$$

The parameters $\{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$ are estimated using the Expectation-Maximization (EM) algorithm [58]:

E-step: Compute class probabilities γ_{ik} using the current parameters.

M-step: Update parameters by maximizing the expected complete-data log-likelihood:

$$\mu_k = \frac{\sum_{i=1}^N \gamma_{ik} x_i}{\sum_{i=1}^N \gamma_{ik}}, \quad \Sigma_k = \frac{\sum_{i=1}^N \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^\top}{\sum_{i=1}^N \gamma_{ik}}, \quad \pi_k = \frac{1}{N} \sum_{i=1}^N \gamma_{ik}.$$

These steps are iterated until convergence.

In the case of two components ($K = 2$), the decision boundary lies at the point where the two Gaussian densities intersect.

AI statement

AI was used solely to polish the text for clarity, grammar, and readability.

References

- [1] K. F. Raffa, J.-C. Grégoire, B. Staffan Lindgren, Chapter 1 - natural history and ecology of bark beetles, in: F. E. Vega, R. W. Hofstetter (Eds.), *Bark Beetles*, Academic Press, San Diego, 2015, pp. 1–40. URL: <https://www.sciencedirect.com/science/article/pii/B9780124171565000010>. doi:<https://doi.org/10.1016/B978-0-12-417156-5.00001-0>.
- [2] P. Öhrn, B. Långström, Å. Lindelöw, N. Björklund, Seasonal flight patterns of *ips typographus* in southern sweden and thermal sums required for emergence, *Agricultural and Forest Entomology* 16 (2014) 147–157. URL: <https://resjournals.onlinelibrary.wiley.com/doi/abs/10.1111/afe.12044>. doi:<https://doi.org/10.1111/afe.12044>.
- [3] Š. Hýsek, R. Löwe, M. Turčáni, What happens to wood after a tree is attacked by a bark beetle?, *Forests* 12 (2021) 1163.
- [4] Z. J. Robbins, C. Xu, A. Jonko, R. Chitra-Tarak, C. J. Fettig, J. Costanza, L. A. Mortenson, B. H. Aukema, L. M. Kueppers, R. M. Scheller, Carbon stored in live ponderosa pines in the sierra nevada will not return to pre-drought (2012) levels during the 21st century due to bark beetle outbreaks, *Frontiers in Environmental Science* 11 (2023) 1112756.
- [5] W. A. Kurz, C. C. Dymond, G. Stinson, G. J. Rampley, E. T. Neilson, A. L. Carroll, T. Ebata, L. Safranyik, Mountain pine beetle and forest carbon feedback to climate change, *Nature* 452 (2008) 987–990. URL: <http://dx.doi.org/10.1038/nature06777>. doi:http://www.nature.com/nature/journal/v452/n7190/supinfo/nature06777_S1.html.
- [6] T. Hlásny, L. König, P. Krokene, M. Lindner, C. Montagné-Huck, J. Müller, H. Qin, K. F. Raffa, M.-J. Schelhaas, M. Svoboda, H. Viriri, R. Seidl, Bark beetle outbreaks in europe: State of knowledge and ways forward for management, *Current Forestry Reports* 7 (2021) 138–165. URL: <https://doi.org/10.1007/s40725-021-00142-x>. doi:[10.1007/s40725-021-00142-x](https://doi.org/10.1007/s40725-021-00142-x).
- [7] M.-J. Schelhaas, G.-J. Nabuurs, A. Schuck, Natural disturbances in the european forests in the 19th and 20th centuries, *Global Change*

- Biology 9 (2003) 1620–1633. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2486.2003.00684.x>. doi:10.1046/j.1365-2486.2003.00684.x.
- [8] C. Senf, D. Pflugmacher, Y. Zhiqiang, J. Sebal, J. Knorn, M. Neumann, P. Hostert, R. Seidl, Canopy mortality has doubled in europe’s temperate forests over the last three decades, *Nature Communications* 9 (2018) 4978. URL: <https://doi.org/10.1038/s41467-018-07539-6>. doi:10.1038/s41467-018-07539-6.
 - [9] A. Fernandez-Carrillo, Z. Patočka, L. Dobrovolný, A. Franco-Nieto, B. Revilla-Romero, Monitoring bark beetle forest damage in central europe. a remote sensing approach validated with field data, *Remote Sensing* 12 (2020) 3634.
 - [10] M. Schroeder, Granbarkborrens förökningsframgång i dödade träd under sommaren 2022 i sydöstra småland, värmland och uppland/västmanland, 2023.
 - [11] S. Bundesamt, Forest damage: Logging of timber damaged by insect infestation grew more than tenfold within five years, *Press Release* (2021) 050.
 - [12] G. Salton, M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1983. URL: <https://books.google.se/books?id=7f5TAAAMAAJ>.
 - [13] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, M. Heurich, Sentinel-2 accurately maps green-attack stage of european spruce bark beetle (*ips typographus*, l.) compared with landsat-8, *Remote sensing in ecology and conservation* 5 (2019) 87–106.
 - [14] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, M. Heurich, Timing of red-edge and shortwave infrared reflectance critical for early stress detection induced by bark beetle (*ips typographus*, l.) attack, *International Journal of Applied Earth Observation and Geoinformation* 82 (2019) 101900.
 - [15] L. Huo, H. J. Persson, E. Lindberg, Early detection of forest stress from european spruce bark beetle attack, and a new vegetation index:

Normalized distance red & swir (ndrs), *Remote Sensing of Environment* 255 (2021) 112240. URL: <https://www.sciencedirect.com/science/article/pii/S0034425720306131>. doi:<https://doi.org/10.1016/j.rse.2020.112240>.

- [16] V. Bárta, P. Lukeš, L. Homolová, Early detection of bark beetle infestation in norway spruce forests of central europe using sentinel-2, *International Journal of Applied Earth Observation and Geoinformation* 100 (2021) 102335. URL: <https://www.sciencedirect.com/science/article/pii/S0303243421000428>. doi:<https://doi.org/10.1016/j.jag.2021.102335>.
- [17] M. Dalponte, Y. T. Solano-Correa, L. Frizzera, D. Gianelle, Mapping a european spruce bark beetle outbreak using sentinel-2 remote sensing data, *Remote Sensing* 14 (2022). URL: <https://www.mdpi.com/2072-4292/14/13/3135>. doi:[10.3390/rs14133135](https://doi.org/10.3390/rs14133135).
- [18] S. Jamali, P.-O. Olsson, A. Ghorbanian, M. Müller, Examining the potential for early detection of spruce bark beetle attacks using multi-temporal sentinel-2 and harvester data, *ISPRS Journal of Photogrammetry and Remote Sensing* 205 (2023) 352–366. URL: <https://www.sciencedirect.com/science/article/pii/S0924271623002873>. doi:<https://doi.org/10.1016/j.isprsjprs.2023.10.013>.
- [19] S. Jamali, P.-O. Olsson, M. Müller, A. Ghorbanian, Kernel-based early detection of forest bark beetle attack using vegetation indices time series of sentinel-2, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024) 12868–12877. doi:[10.1109/JSTARS.2024.3425795](https://doi.org/10.1109/JSTARS.2024.3425795).
- [20] P.-O. Olsson, P. Zhao, M. Müller, A. Mansourian, J. Ardö, Combining sentinel-2 data and risk maps to detect trees predisposed to and attacked by european spruce bark beetle, *Remote Sensing* 16 (2024) 4166. URL: <https://www.mdpi.com/2072-4292/16/22/4166>.
- [21] M. Kirsch, J. Wernicke, P. Datta, C. Preisach, Early detection of forest calamities in homogeneous stands—deep learning applied to bark-beetle outbreaks, *arXiv preprint arXiv:2503.12883* (2025).

- [22] Y.-L. Kong, Q. Huang, C. Wang, J. Chen, J. Chen, D. He, Long short-term memory neural networks for online disturbance detection in satellite image time series, *Remote Sensing* 10 (2018) 452.
- [23] G. Zhou, M. Liu, X. Liu, An autoencoder-based model for forest disturbance detection using landsat time series data, *International Journal of Digital Earth* 14 (2021) 1087–1102.
- [24] S. Saha, F. Bovolo, L. Bruzzone, Change detection in image time-series using unsupervised lstm, *IEEE geoscience and remote sensing letters* 19 (2020) 1–5.
- [25] G. Andresini, A. Appice, D. Malerba, V. Recchia, Deep change vector analysis to map bark beetle outbreaks in open sentinel-2 data, in: 2025 International Joint Conference on Neural Networks (IJCNN), 2025, pp. 1–8. doi:[10.1109/IJCNN64981.2025.11229013](https://doi.org/10.1109/IJCNN64981.2025.11229013).
- [26] G. Andresini, A. Appice, D. Malerba, An explainable framework to map bark beetle infestation in sentinel-2 images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023) 10050–10066.
- [27] G. Andresini, A. Appice, D. Ienco, V. Recchia, Diamante: A data-centric semantic segmentation approach to map tree dieback induced by bark beetle infestations via satellite images, *Journal of intelligent information systems* 62 (2024) 1531–1558.
- [28] C. Schiller, J. Költzow, S. Schwarz, F. Schiefer, F. E. Fassnacht, Forest disturbance detection in central europe using transformers and sentinel-2 time series, *Remote Sensing of Environment* 315 (2024) 114475.
- [29] L. Raspović, B. Radun, I. Tekić, I. Odak, I. Tomljenović, Sentinel-2 and machine learning approaches for early detection of forest pest outbreaks in croatia, in: 2nd International Conference of Environmental Remote Sensing and GIS, University of Zagreb-Faculty of Geodesy, 2026, pp. 333–336.
- [30] C. Schiller, F. E. Fassnacht, Comparative study of near real-time monitoring algorithms for early detection of bark beetle infestations in germany with sentinel-2, *European Journal of Remote Sensing* 59 (2026) 2662660.

- [31] C. Schiller, J. May, R. Klinke, F. E. Fassnacht, Early detection of bark beetle infestations in central europe using deep learning-based reconstructions of irregular sentinel-2 time series, *Forestry: An International Journal of Forest Research* 99 (2026) cpaf053.
- [32] Swedish Meteorological and Hydrological Institute, Climate normals, 2025. URL: <https://www.smhi.se/klimat/klimatet-da-och-nu/normalkartor/normal/>.
- [33] P. Nilsson, C. Roberge, J. Fridman, Skogsdata 2021: aktuella uppgifter om de svenska skogarna från slu riksskogstaxeringen, 2021.
- [34] Naturvårdsverket, National land cover database (nmd), 2021. URL: <https://www.naturvardsverket.se/en/services-and-permits/maps-and-map-services/national-land-cover-database/>.
- [35] Lantmäteriet, Gsd-terrängkartan, vektor. version 5.10, 2021. URL: <https://www.lantmateriet.se/globalassets/geodata/geodataprodukter/produktlista/terrshmi.pdf>.
- [36] M. Müller, P. O. Olsson, L. Eklundh, S. Jamali, J. Ardö, Features predisposing forest to bark beetle outbreaks and their dynamics during drought, *Forest Ecology and Management* 523 (2022) 120480. URL: <https://www.sciencedirect.com/science/article/pii/S0378112722004741>. doi:<https://doi.org/10.1016/j.foreco.2022.120480>.
- [37] D. Frantz, Force—landsat+ sentinel-2 analysis ready data and beyond, *Remote Sensing* 11 (2019) 1124.
- [38] D. Frantz, FORCE — Framework for Operational Radiometric Correction for Environmental Monitoring: Time Series Analysis, Read the Docs, 2023. URL: <https://force-eo.readthedocs.io/en/latest/components/higher-level/tsa/index.html>, version 3.10.1.
- [39] J. Rouse, J. W., D. W. Deering, R. H. Haas, J. A. Schell, Monitoring vegetation systems in the Great Plains with ERTS, Technical Report, Texas A&M University, 1974. URL: <http://hdl.handle.net/2060/19740022614>, nASA Technical Report.

- [40] B. Gao, NdwI—a normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment* 58 (1996) 257–266. URL: <https://www.sciencedirect.com/science/article/pii/S0034425796000673>. doi:[https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).
- [41] E. M. Barnes, T. Clarke, S. E. Richards, P. D. Colaizzi, J. Haberland, M. Kostrzewski, P. M. Waller, C. Y. Choi, E. Riley, T. L. Thompson, R. J. Lascano, H. Li, M. S. Moran, P. C. Robert, R. H. Rust, W. E. Larson, Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data., 2000. URL: <https://api.semanticscholar.org/CorpusID:128773162>.
- [42] C. HechtI, S. Hauser, A. Schmitt, M. Heurich, A. Wendleder, Kennaugh elements allow early detection of bark beetle infestation in temperate forests using sentinel-1 data, *Forests* 16 (2025). URL: <https://www.mdpi.com/1999-4907/16/8/1272>. doi:[10.3390/f16081272](https://doi.org/10.3390/f16081272).
- [43] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: *kdd*, 34, 1996, pp. 226–231.
- [44] S. Zeinali, P.-O. Olsson, T. Kronvall, M. Wiktorsson, J. Lindström, Detection of bark beetle attacks using time-aggregated satellite data with machine learning, *Remote Sensing* 18 (2026) 2234.
- [45] D. P. Kingma, M. Welling, et al., Auto-encoding variational bayes, 2013.
- [46] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation* 9 (1997) 1735–1780.
- [47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [48] X. Cai, Y. Bi, P. Nicholl, Tampered vae for improved satellite image time series classification, *arXiv preprint arXiv:2203.16149* (2022).
- [49] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).

- [50] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, Special lecture on IE 2 (2015) 1–18.
- [51] Skogsstyrelsen, Svärningsövervakning, ??? URL: <https://www.skogsstyrelsen.se/statistik/skogsskador/svarningsovervakning/>.
- [52] P. Öhrn, Carbon castles vs. beetle-fungus armies: phenology of the spruce bark beetle and Norway spruce resistance in times of climate change, Ph.D. thesis, SLU, 2024. doi:[10.54612/a.7oclp11aad](https://doi.org/10.54612/a.7oclp11aad).
- [53] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (1979) 224–227.
- [54] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.
- [55] D. J. Rezende, S. Mohamed, D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, in: International conference on machine learning, PMLR, 2014, pp. 1278–1286.
- [56] L. v. d. Maaten, G. Hinton, Visualizing data using t-sne, Journal of machine learning research 9 (2008) 2579–2605.
- [57] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, volume 4, Springer, 2006.
- [58] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, Journal of the Royal Statistical Society. Series B 39 (1977) 1–38.

List of Figure Captions:

Figure 1 : Location of the study area in Sweden, divided into two tiles. Source of the land cover data (left): [Naturvårdsverket \(2018\)](#).

Figure 2 : Data Preprocessing Workflow: 0. Images are normalized to mean zero and variance one. 1. Filtering: Sentinel-2 images with $> 60\%$ missing pixels were discarded to ensure data reliability. 2. Imputation: Remaining missing values were filled using forward-fill to maintain temporal continuity. 3. Temporal Alignment: To enable tile merging, the irregular acquisition dates were mapped onto a uniform 5-day interval calendar. This 5-day aggregation ensures a consistent time series for vegetation index calculation and seamless spatial analysis. The intervals with no observation were filled with 0; due to normalization of images, the zero-filling corresponds to mean imputation.

Figure 3 : Architecture of the CVAE used in this paper; for details on each layer, see [Appendix B](#).

Figure 4 : Distribution of the GMM-derived risk scores, illustrating the separation of observations across the classes.

Figure 5 : Performance of the proposed model versus the number of available days, for 2018. Here, days mean 5-day intervals. The swarming activity recorded in the pheromone traps at Lagan and Nässjö is represented by gray/black lines; the data are normalized so that a value of 1 equals the highest number of trapped bark beetles (6,000 at Lagan and 11,885 at Nässjö).

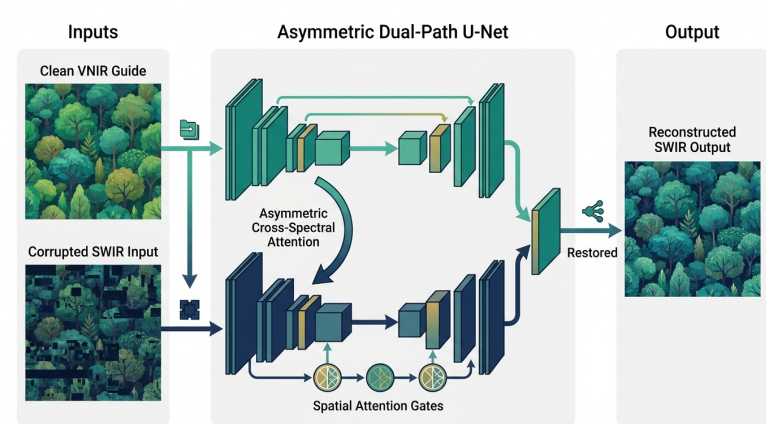
Figure 6 : Performance of the proposed model versus the number of available days, for 2019. The swarming data recorded in the pheromone traps at Lagan and Nässjö are represented by gray/black lines; the data are normalized so that a value of 1 equals the highest number of trapped bark beetles (2,600 at Lagan and 8,833 at Nässjö).

Figure 7 : Visualization of the CVAE latent space using t-SNE projections. The figure shows t-SNE projections (components 1 vs 2, 1 vs 3, and 1 vs 4).

Figure A.8 : Histogram of risk of attack in new health-only data.

Paper V





Graphical Abstract

Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction

Shokoufa Zeinali, Ted Kronvall, Langning Huo, Per-Ola Olsson, Johan Lindström

Highlights

Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction

Shokoufa Zeinali, Ted Kronvall, Langning Huo, Per-Ola Olsson, Johan Lindström

- Cross-sensor reconstruction of a corrupted SWIR band in UAV data.
- Asymmetric design of a U-net to keep the clean VNIR band intact.
- Spatial Attention uses SWIR features to identify important context-aware details in the VNIR data.

Spatial Attention Gating in Dual-Encoder Networks for Remote Sensing Reconstruction

Shokoufa Zeinali^a, Ted Kronvall^a, Langning Huo^c, Per-Ola Olsson^b, Johan Lindström^a

^a*Centre for Mathematical Sciences, Lund University, Mårtenstam
4, Lund, 22100, Sweden*

^b*Dept of Earth and Environmental Sciences, Lund University, Sölvegatan
12, 22362, Lund, Sweden*

^c*Department of Forest Resource Management, Swedish University of Agricultural
Sciences, Skogsmarksgränd, Umeå, 90183, Sweden*

Abstract

Remote sensing has received increasing attention in recent years due to the rapid development of satellite and drone imaging technologies, enabling the acquisition of large amounts of multispectral and hyperspectral imagery for environmental monitoring, agriculture, forestry, and other Earth observation applications. However, obtaining high-quality hyperspectral imagery remains challenging: Shortwave infrared (SWIR) imagery acquired by drones is often affected by noise, distortions, and sensor-related artifacts caused by platform motion, atmospheric effects, and acquisition limitations.

In this work, we propose an asymmetric dual-path U-Net for SWIR image reconstruction from corrupted SWIR observations using clean visible and near-infrared (VNIR) guidance. The network consists of separate VNIR and SWIR encoders, where asymmetric cross-spectral attention modules allow VNIR features to guide the corrupted SWIR stream without allowing corrupted SWIR information to modify the VNIR representation. The bottleneck fuses the deepest VNIR and SWIR features into a shared latent representation. During decoding, SWIR-guided spatial attention gates selectively filter VNIR skip connections, while an additional direct VNIR projection injects clean structural information at each decoder level. The model is trained with a composite loss combining robust reconstruction, gradient, edge, per-

centile, rank, SSIM, and VNIR-guided texture terms. Experimental results on two independent forest test sites show that the proposed method effectively reconstructs clean SWIR images from synthetically corrupted inputs, achieving PSNR of 28.42–30.23 dB, with high structural similarity (SSIM) 0.95–0.93 and strong pixel-wise correlation (NCC) 0.99–0.98 on an unseen forest region and the original test split, respectively, demonstrating robust generalization while preserving spatial structure, high-frequency texture, and radiometric consistency.

Keywords: Remote Sensing, Machine Learning, Drone Imagery, Band Reconstruction, Cross-Sensor

1. Introduction

Hyperspectral and multispectral remote sensing provide valuable information for environmental monitoring. The VNIR domain offers high spatial detail and strong sensitivity to vegetation structure, while the SWIR domain carries information related to moisture content, chemical composition, and stress signals [4, 23]. However, SWIR imagery collected by unmanned aerial vehicles (UAVs) is often affected by motion artefacts, scan-line distortions, illumination variability, and georeferencing errors, which degrade the acquired image when acquired by the pushbroom acquisition geometry.

We propose an asymmetric dual-encoder U-Net that restores corrupted SWIR imagery using clean VNIR guidance. The key design principle is that the two modalities are treated asymmetrically: VNIR features guide the refinement of corrupted SWIR features, but not vice versa, preserving the integrity of the clean spatial reference throughout encoding. A composite loss combining radiometric, structural, distributional, and texture terms supervises training, and performance is evaluated using both whole-image and masked-region metrics that isolate reconstruction quality within corrupted areas.

1.1. Related Work

Encoder-decoder architectures, particularly U-Net [18], have become the dominant framework for image reconstruction and fusion tasks. Their ability to combine coarse semantic features from deep layers with fine spatial detail from shallow layers via skip connections makes them well-suited for dense prediction tasks [1]. Attention mechanisms have further improved

these architectures by allowing the network to selectively emphasise informative spatial regions. Oktay et al. [17] introduce attention gates into the U-Net skip connections, suppressing irrelevant activations before they reach the decoder, a principle directly related to our spatial attention gate design. A broad survey of deep learning for remote sensing image enhancement, covering super-resolution, denoising, reconstruction, and fusion, is provided by [24], who identify cross-modal guidance and asymmetric fusion as open challenges.

Restoring missing or damaged spectral bands from complementary sensor data is a well-studied problem. Zhang et al. [28] propose a unified spatial-temporal-spectral CNN that handles MODIS band reconstruction, Landsat ETM+ gap filling, and cloud removal within one framework. Rout [20] introduces ALERT, combining residual dense blocks with adversarial learning under joint spatial and spectral priors, achieving strong cross-sensor generalisation without requiring paired ground truth. Sidorov and Hardeberg [22] extend the deep image prior framework to hyperspectral data using 3D CNNs, performing denoising, inpainting, and super-resolution without any paired training data.

For SWIR-specific synthesis, Rout et al. [21] proposed DeepSWIR, a CNN that generates high-resolution SWIR imagery from VNIR and coarse SWIR inputs, directly addressing the spatial resolution gap between commercial SWIR and VNIR sensors. Their work is the closest in motivation to ours, but treats the problem as band synthesis from a different resolution rather than reconstruction from localised corruption, and does not model the directional nature of cross-spectral guidance.

Encoder-decoder designs have also been extended for spectral band reconstruction from multispectral inputs. Neagoe et al. [16] modify U-Net with additional skip connections, residual blocks, and spectral attention for Sentinel-2 band reconstruction, achieving 15–20% reduction in mean absolute error. Deng et al. [7] integrate residual learning with multi-scale convolutional kernels to jointly model spectral correlations and spatial structure. More recently, transformer-based architectures have been explored for long-range spectral-spatial dependencies, including the Hierarchical Spectral-Spatial Transformer [29] and the Progressive Spatial-Spectral Joint Network [14].

Cross-spectral fusion methods exploit the complementary properties of different spectral domains to enhance a degraded or low-resolution modality.

For comprehensive reviews of hyperspectral pansharpening and fusion, covering both classical and learning-based approaches, see e.g. [15, 26]. Chen et al. [3] reviewed deep learning methods for multispectral image fusion, identifying attention mechanisms as a key component for effective cross-modal interaction.

A common limitation of existing fusion methods is that they treat the two input modalities symmetrically, so that both streams can influence each other equally. This is appropriate when both inputs are of comparable quality, but is problematic when one modality is corrupted. Allowing corrupted features to modify the clean guidance signal degrades the spatial reference that the reconstruction depends on. Our model addresses this directly through an asymmetric cross-spectral attention module in which only the SWIR stream is updated, while the VNIR stream remains unchanged throughout encoding.

Existing methods for SWIR reconstruction or cross-spectral fusion either target sensor-specific degradation patterns with classical signal processing [28, 20], treat the two modalities symmetrically without accounting for the different reliability of VNIR and SWIR [16, 7], or focus on spectral synthesis rather than spatial reconstruction from localised corruption [21]. Methods based on U-Net with attention gates [17] provide selective use of skip connections but operate on a single modality and do not model cross-spectral guidance. To our knowledge, no existing method combines asymmetric cross-spectral attention, SWIR-guided spatial gating of VNIR skip connections, and a VNIR-guided texture loss that is applied specifically within estimated corruption regions. Our method is designed for UAV-acquired SWIR hyperspectral data where a clean VNIR image of the same scene is available, a setting that is increasingly common in precision forestry and environmental monitoring applications.

2. Data

In this section, we go through the details of the data used and the preprocessing steps taken to prepare the data for the model.

2.1. SWIR and VNIR

The multispectral (VNIR) and hyperspectral (SWIR) datasets used in this study were both acquired in 2023, in the same week, at Remningstorp Experimental Forest, a managed forest landscape in southern Sweden primarily

composed of Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*). The study area is relatively flat, with an average elevation of approximately 140 m above sea level. Both datasets were collected over the same experimental sites using UAV-based remote sensing systems.

VNIR images were acquired using a MAIA S2 sensor mounted on a DJI Matrice 210 RTK UAV, shown in Fig. 1, recording nine spectral bands between 443 and 865 nm. Further details regarding the multispectral dataset are provided in [10, 11].



Figure 1: From left to right: The MAIA camera used to collect VNIR images, the MAIA ILS, the DJI Matrice 210 RTK drone, and the radiometric reference targets. Images from [11]

SWIR images were collected using a SPECIM AFX17 pushbroom sensor mounted on a Freefly Alta X UAV equipped with a Gremsy T7 gimbal, illustrated in Fig. 2, capturing 112 contiguous bands spanning 1000–1700 nm. Additional acquisition and preprocessing details are described in [5].

Table 1 summarizes the main acquisition parameters of both sensors.

The VNIR and SWIR datasets differ in imaging geometry and acquisition methods. The MAIA S2 system operates as a frame-based multispectral camera, capturing two-dimensional image frames across broad VNIR bands and generating spatially consistent orthomosaics with relatively stable geometric properties. The SPECIM AFX17 is a pushbroom hyperspectral sensor¹ that records imagery line by line along the UAV flight direction, producing contiguous SWIR hyperspectral cubes. Although the pushbroom configuration provides substantially richer spectral information, it is also more sensitive to UAV movement, gimbal instability, scan-line misalignment, and illumination variability.

¹A pushbroom scanner is a line-scanning sensor that captures images one narrow line at a time as the device moves forward.



Figure 2: From left to right: The hyperspectral drone system used for acquisition of SWIR data. Ground calibration targets used for post-processing reflectance estimation. Images from [5].

Table 1: Specifications of the MAIA S2 (VNIR) and SPECIM AFX17 (SWIR) sensors.

Parameter	MAIA S2 (VNIR)	AFX17 (SWIR)
Spectral range (nm)	443–865	1000–1700
Bands	9	112
Spectral sampling (nm)	–	6.8
Ground sampling distance (cm)	6.0	10.8
Flight speed (m s^{-1})	–	7.0
Frames per second (FPS)	–	64.8
Flight height (m)	120.0	100.0
Swath width (m)	–	69.0
Overlap	–	71%

2.2. Data Preprocessing, Normalization and Augmentation

The preprocessing steps are shown in Fig. 3. To enable pixel-wise comparison between VNIR and SWIR, the two datasets were first geometrically aligned. The SWIR imagery was co-registered to the VNIR reference using manual ground control points in QGIS [6], and a validity mask was generated from the SWIR data to remove background and noisy regions. Since the two

modalities had different spatial resolutions, VNIR was resampled to match the SWIR grid using bilinear interpolation, which preserves spatial continuity and avoids abrupt intensity transitions. A few centimeters of misalignment remained after these steps, which were handled by the model itself.

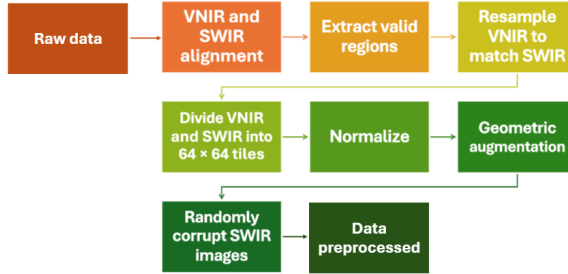


Figure 3: Diagram illustrating the preprocessing workflow.

The aligned drone images were then divided into fixed-size 64×64 tiles, the input size required by the network. Images whose height or width was not divisible by 64 were cropped to the largest region that was, so that tiles could be extracted without padding.

Before training, VNIR and SWIR images were normalized using percentile-based min-max scaling, clipping to the 2nd and 98th percentiles and scaling to $[0, 1]$. This reduces the influence of outlier pixels while keeping the useful dynamic range. A single global range was computed for SWIR, and a separate range for each VNIR band. All percentiles were computed from the training set only and reused for validation, test, and inference data, preventing data leakage and allowing the same values to be used later for denormalizing predictions.

Since training data were limited, geometric augmentations were applied: random rotation by $k \times 90^\circ$ ($k \sim \{0, 1, 2, 3\}$), random horizontal flip, and random vertical flip. These augmentations preserve spectral values, which is important since intensity-based augmentations could distort the VNIR–SWIR spectral relationship.

From the SWIR-related bands, one band was selected as the clean SWIR reference. The proposed method reconstructs one band per run, although

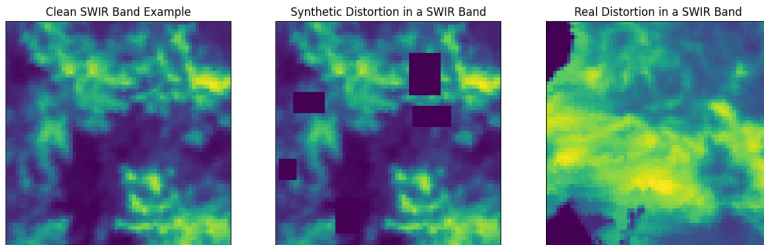


Figure 4: Examples of SWIR band corruption used in the proposed reconstruction framework. From left to right: a clean SWIR band, a synthetically corrupted SWIR band generated using random block-based distortion, and a real SWIR band affected by localized distortion.

this can be modified. Each training sample thus consists of a VNIR image, a clean SWIR target, and a synthetically corrupted SWIR input generated during training. Corruption followed one of two random block-based patterns, chosen with equal probability: a single larger rectangular region or several smaller rectangular regions, mimicking the localized block artifacts and radiometric distortions seen in real SWIR imagery. Each corrupted region was filled with one of three values $\{0, \text{mean}(S), u\}$, chosen at random, where:

$$u \sim \mathcal{U}(0.2, 0.8).$$

Figure 4 illustrates representative examples of SWIR band data used in this study, including a clean, a synthetically corrupted, and a real SWIR band affected by localized distortion.

3. Methodology

In this section, we explain the proposed method. More details of this model can be found in Appendix A.

3.1. Model Architecture

We propose an asymmetric dual-path U-Net [19] for SWIR reconstruction using two spectral inputs: clean VNIR imagery with 8 channels and synthetically corrupted SWIR imagery with 1 channel. The network is designed to

reconstruct the clean SWIR image by exploiting the complementary properties of the two modalities. VNIR provides rich spatial structure, texture, and boundary information, while SWIR contains the target spectral response but may be corrupted or locally degraded. The overall model therefore learns a mapping from the clean VNIR image and corrupted SWIR image to the corresponding clean SWIR reconstruction. Fig. 5 shows the overall architecture of this model. The main blocks are:

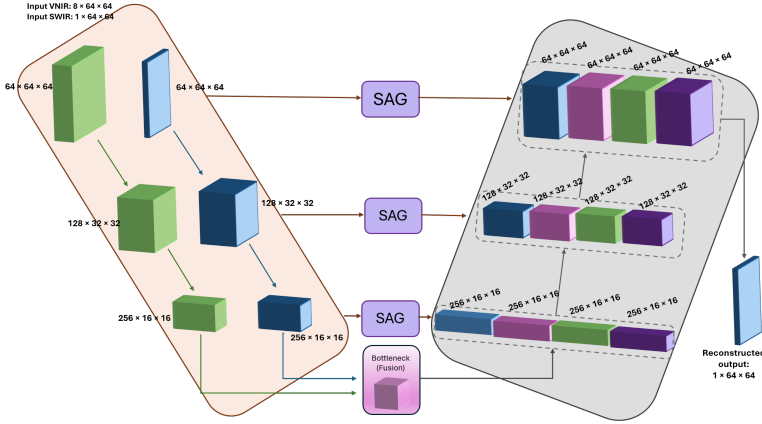


Figure 5: Overall architecture of the proposed Asymmetric Dual U-Net for SWIR reflectance reconstruction from VNIR imagery. The network consists of two parallel encoder branches that process VNIR (8 bands) and SWIR (1 band) inputs separately. The bottleneck fuses high-level features from both modalities. During decoding, Spatial Attention Gates (SAGs) use SWIR features to selectively emphasize relevant VNIR skip connections, while direct VNIR projection paths inject clean structural information into each decoder stage. Decoder blocks combine the upsampled decoder features (pink blocks), gated VNIR features (purple blocks), SWIR features (blue blocks), and direct VNIR projections (green blocks) to progressively reconstruct the target SWIR image. Tensor dimensions are shown as *channels* $\times$ *height* $\times$ *width*, with encoder outputs of $64 \times 64 \times 64$, $128 \times 32 \times 32$, and $256 \times 16 \times 16$, and decoder outputs of $256 \times 16 \times 16$, $128 \times 32 \times 32$, and $64 \times 64 \times 64$. The final 1×1 convolution followed by a sigmoid activation produces the reconstructed SWIR reflectance map.

Dual-Encoder Design: The model contains two separate encoders: one for the VNIR input and one for the corrupted SWIR input. Each encoder extracts hierarchical features at multiple spatial scales using residual

convolutional blocks with Group Normalization and GELU activations. The VNIR encoder processes the 8-channel VNIR image, while the SWIR encoder processes the 1-channel corrupted SWIR image. This dual-encoder design allows the network to preserve modality-specific representations before cross-spectral fusion. Fig. 6 illustrates the architecture of a single layer of the dual-encoder.

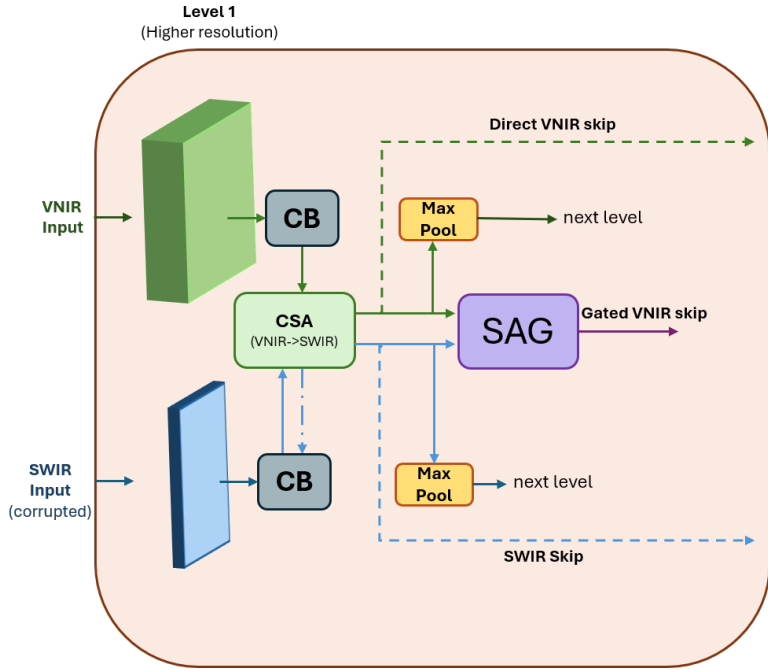


Figure 6: One-level illustration of the dual-encoder feature extraction and skip preparation. VNIR and corrupted SWIR inputs are first processed by separate convolutional blocks. The resulting features are passed to the asymmetric cross-spectral attention module, where VNIR features guide the refinement of the SWIR stream while the VNIR stream remains unchanged. The post-attention features are used to form the direct VNIR skip, the SWIR skip, and the SWIR-guided spatial attention gate output. The same post-attention features are then max-pooled and passed to the next encoder level.

At each encoder level, an asymmetric cross-spectral attention module is applied. In this module, VNIR features provide the query representation, while SWIR features provide the key and value representations (for details and equations, see Appendix A). The resulting attention output is used only to update the SWIR feature stream, while the VNIR stream remains unchanged. This design is motivated by the fact that the SWIR input is corrupted, whereas VNIR provides a cleaner structural reference. Therefore, preventing corrupted SWIR features from modifying VNIR features helps preserve the integrity of the spatial guidance signal. In this way, cross-spectral attention enables VNIR-guided refinement of SWIR features without symmetric contamination between modalities.

Bottleneck Fusion: After the final encoder level, the deepest VNIR and SWIR features are downsampled and concatenated along the channel dimension. This concatenated representation is passed through the bottleneck convolutional blocks to form a shared latent representation. Unlike element-wise addition, channel-wise concatenation preserves the full information from both modalities before the joint processing. The bottleneck therefore integrates high-level VNIR structural cues and SWIR spectral information at the coarsest spatial resolution.

Spatial Attention Gate: During decoding, the network progressively reconstructs the restored SWIR image using bilinear upsampling followed by convolution. At each decoder stage, encoder skip connections are incorporated to recover spatial details. However, instead of passing all VNIR skip features directly to the decoder, we introduce a SWIR-guided spatial attention gate.

Specifically, the VNIR skip features at a given scale are filtered using the corresponding SWIR feature as the gating signal. The gate first projects both the SWIR gating feature and the VNIR skip feature into an intermediate feature space using 1×1 convolutions and Group Normalization. These projected features are added, passed through a ReLU activation, and then mapped to a single-channel spatial attention mask using another 1×1 convolution followed by a sigmoid function. The resulting attention map assigns an importance weight between 0 and 1 to each spatial location. This map modulates the VNIR skip feature before it is passed to the decoder.

This mechanism has two important effects:

- It suppresses VNIR structures that are not relevant for SWIR reconstruction.
- It makes the use of VNIR spatial information context-aware, because the attention mask is guided by the SWIR feature representation at the same scale.

As a result, the decoder receives VNIR spatial details selectively, focusing on structures that are useful for restoring missing or degraded SWIR regions.

Direct VNIR Projection: In addition to the gated VNIR skip connection, the model includes a direct VNIR projection at each decoder stage. The corresponding VNIR encoder feature is passed through a 1×1 convolution to match the decoder channel dimension and is then concatenated with the other decoder inputs. This direct path provides an independent residual route for clean VNIR structural information. It ensures that useful VNIR details remain available to the decoder even when the spatial attention gate suppresses part of the VNIR skip feature.

Decoder: At each decoder stage, four components are fused:

- the upsampled decoder feature,
- the SWIR-guided gated VNIR skip feature,
- the corresponding SWIR skip feature,
- the direct VNIR projection.

These features are concatenated along the channel dimension and refined using residual convolutional blocks. This decoder design combines spectral information from the corrupted SWIR stream with both selective and direct structural guidance from the clean VNIR stream. Thus, attention in the encoder provides global VNIR-to-SWIR feature refinement, while attention in the decoder performs localized, task-driven spatial selection.

Finally, a 1×1 convolution maps the decoder output to a single-channel SWIR image. A sigmoid activation is applied to constrain the reconstructed output to the normalized intensity range. The final prediction

represents the restored clean SWIR image.

3.2. Loss Function

The proposed model is trained to reconstruct the clean SWIR image S from a synthetically corrupted SWIR observation S_c using VNIR guidance V . Given the prediction

$$\hat{S} = f_{\theta}(V, S_c),$$

the training objective combines radiometric fidelity, structural consistency, distribution preservation, and VNIR-guided texture recovery. The full mathematical definition of each loss component is provided in the Appendix A.9. In this section, we summarize the overall objective.

The total loss is defined as a weighted sum of eight complementary terms:

$$\begin{aligned} \mathcal{L}_{\text{total}} = & \lambda_{\text{recon}}\mathcal{L}_{\text{recon}} + \lambda_{\text{high}}\mathcal{L}_{\text{high}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}} + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}} \\ & + \lambda_{\text{pct}}\mathcal{L}_{\text{pct}} + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{tex}}\mathcal{L}_{\text{VNIR-tex}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{SSIM}}. \end{aligned}$$

The reconstruction term $\mathcal{L}_{\text{recon}}$ is a Charbonnier loss that provides robust pixel-wise supervision between the restored and clean SWIR images. The high-intensity term $\mathcal{L}_{\text{high}}$ gives additional weight to bright SWIR responses, reducing the tendency of the network to underestimate high-value regions. The gradient loss $\mathcal{L}_{\text{grad}}$ compares Sobel gradients of the prediction and clean target to preserve edges and local transitions, while the VNIR edge-weighted term $\mathcal{L}_{\text{edge}}$ increases the penalty near structural boundaries detected from the VNIR image.

To preserve the global radiometric distribution, the percentile loss $\mathcal{L}_{\text{pct}}$ matches selected intensity percentiles between the prediction and target, and the rank-based patch loss $\mathcal{L}_{\text{rank}}$ maintains the relative ordering of bright and dark regions across local patches. The VNIR-guided texture term $\mathcal{L}_{\text{VNIR-tex}}$ is applied inside regions estimated to be corrupted in S_c . This term encourages the restored SWIR image to recover plausible high-frequency texture by aligning its gradient and Laplacian structure with VNIR-derived spatial information only in degraded areas. Finally, the SSIM term $\mathcal{L}_{\text{SSIM}}$ encourages local structural similarity between the restored and clean SWIR images.

During the initial training stage, the loss weights, found by grid search, across $[0, 3]$, are set as:

$$\lambda_{\text{recon}} = 1.0, \quad \lambda_{\text{high}} = 1.0, \quad \lambda_{\text{grad}} = 0.5, \quad \lambda_{\text{edge}} = 0.8,$$

$$\lambda_{\text{pct}} = 0.8, \quad \lambda_{\text{rank}} = 0.5, \quad \lambda_{\text{tex}} = 2.0, \quad \lambda_{\text{ssim}} = 1.0.$$

After initial training, the model is fine-tuned using a focused corruption strategy. The fine-tuning loss reduces the weights of the high-intensity, edge, percentile, rank, and VNIR-texture terms to avoid disrupting the pixel-accurate reconstruction learned during the first stage:

$$\lambda_{\text{recon}} = 1.0, \quad \lambda_{\text{high}} = 0.5, \quad \lambda_{\text{grad}} = 0.5, \quad \lambda_{\text{edge}} = 0.5,$$

$$\lambda_{\text{pct}} = 0.3, \quad \lambda_{\text{rank}} = 0.3, \quad \lambda_{\text{tex}} = 0.5, \quad \lambda_{\text{ssim}} = 1.0.$$

Sensitivity of the model to each of these weights are evaluated in Appendix A.9.

3.3. Evaluation metrics

The reconstruction performance was evaluated using several complementary metrics that measure radiometric accuracy, structural similarity, edge preservation, and overall agreement between the reconstructed SWIR image $\hat{S}$ and the clean reference SWIR image S . Table 2 briefly defines these metrics. More details with formulas are given in Appendix B.

4. Results

In this section, the results of the proposed method are presented.

4.1. Evaluation on test data

To assess reconstruction performance, full evaluations were conducted on two independent test regions: the original test split (224 tiles) and an additional forest region not seen during training or validation (272 tiles). For each tile, five independent synthetic corruptions were generated, with approximately 20% of SWIR pixels corrupted per realization, giving 1120 and 1360 reconstruction trials, respectively. Table 3 reports the average performance across all tiles and corruption realizations for both sites.

On both sites, the model substantially improves PSNR over the corrupted input and achieves high SSIM, NCC, Sobel edge correlation, and R^2 , with a low SAM, indicating strong structural, radiometric, and directional agreement with the clean SWIR targets. The close agreement between the two sites shows that the model generalizes well to unseen forest imagery rather than overfitting to the training region.

Table 2: Summary of evaluation metrics used for SWIR reconstruction assessment.

Metric	Description
Peak Signal-to-Noise Ratio (PSNR; dB)	Measures pixel-wise reconstruction quality based on the mean squared error between the prediction and clean SWIR target. Higher PSNR indicates lower reconstruction error [9].
Structural Similarity Index(SSIM)	Measures structural similarity between the prediction and target by comparing local luminance, contrast, and structure. Values closer to 1 indicate stronger structural agreement [25].
Normalized Cross-Correlation (NCC)	Measures normalized correlation between the predicted and clean SWIR images. Higher values indicate stronger pixel-wise agreement independent of scale [13].
Sobel Gradient Correlation (Sobel-Corr)	Computes the correlation between Sobel gradient magnitude maps of the prediction and target. This evaluates how well edges and spatial transitions are preserved [8].
Spectral Angle Mapper (SAM; rad)	Measures the angular difference between the predicted and target image vectors. Lower values indicate better directional agreement between the reconstructed and clean SWIR patterns [27].
Coefficient of Determination (R^2)	Measures the proportion of variance in the clean SWIR target explained by the prediction. Values closer to 1 indicate stronger overall agreement.

Quantitative evaluation was also performed exclusively within the corrupted regions. Within corrupted regions only, both the original test split and the independent test site showed substantial recovery from the synthetically corrupted input, as summarized in Table 4.

For qualitative inspection, we visualize individual samples from each site by comparing clean SWIR target, corrupted SWIR input, reconstructed output, an absolute difference map, an intensity profile along the most corrupted row, and a scatter plot of predicted versus ground-truth SWIR values. This

Table 3: Full-image quantitative evaluation on the original test split (224 tiles) and an independent test site (272 tiles), each with five independent corruptions per tile. Values are reported as mean $\pm$ standard deviation. For PSNR, larger values indicate better performance, whereas for NCC, SSIM, Sobel-cor, and R^2 , values nearer to 1 are preferable; for SAM, smaller values reflect better evaluation.

Metric	Test split	Independent site
PSNR (dB)	30.23 ± 3.09	28.42 ± 4.34
SSIM	0.95 ± 0.02	0.93 ± 0.04
NCC	0.99 ± 0.007	0.98 ± 0.02
Sobel-Corr	0.95 ± 0.01	0.93 ± 0.03
SAM (rad)	0.05 ± 0.02	0.06 ± 0.06
R^2	0.96 ± 0.18	0.87 ± 0.68

Table 4: Masked evaluation within corrupted regions only. Values are reported as mean $\pm$ standard deviation.

Metric	Test split	Independent site
Input PSNR (dB)	9.99 ± 4.27	9.24 ± 5.09
Model PSNR (dB)	19.91 ± 4.02	19.85 ± 5.23
Δ PSNR (dB)	$+9.92 \pm 5.13$	$+10.61 \pm 6.00$
NCC	0.774 ± 0.181	0.773 ± 0.198

combination provides an interpretable view of both structural reconstruction and radiometric accuracy. Fig. 7 and Fig. 8 show representative tiles from the test split and the independent site, respectively.

4.2. Comparison with other methods

To provide a simple comparison with non-learning reconstruction approaches, we evaluated several baseline methods. The first baseline is the corrupted SWIR input itself, which represents the degraded image before reconstruction and provides a lower-bound reference. Moment matching adjusts the corrupted image by matching its global mean and standard deviation to those of the reference distribution, correcting global brightness and contrast but not local spatial artifacts. FFT notch filtering operates in the frequency domain by suppressing selected frequency components associated with periodic or stripe-like noise; however, it is less effective for localized block corruptions. Median filtering replaces each pixel with the median value of its local neighborhood and is commonly used to remove impulsive noise and small local

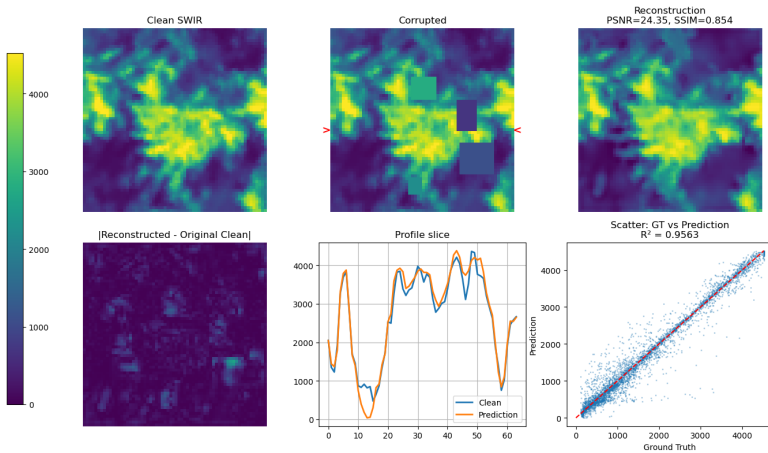


Figure 7: SWIR reconstruction example on a tile from the original test split. Top row: clean SWIR target, corrupted SWIR input, and the reconstructed SWIR target. Bottom row: difference between the reconstructed and original SWIR image, a profile slice along the most corrupted row (marked with red chevrons at the edge of the corrupted image), and a scatter plot of ground-truth versus predicted SWIR values. Larger values for PSNR values nearer to 1 are for SSIM show better performance.

artifacts, but it can blur fine spatial structures. These baselines are compared with the proposed VNIR-guided neural reconstruction model, which uses both SWIR context and VNIR structural information to recover missing or corrupted regions. The results are shown in Table 5.

5. Conclusion

In this work, we proposed an asymmetric dual-path U-Net for SWIR image reconstruction using clean VNIR guidance and synthetically corrupted SWIR observations. The model is designed around the complementary properties of the two spectral domains: VNIR provides reliable spatial structure and texture, while SWIR contains the target spectral information but may be locally corrupted or degraded. To exploit this relationship, the architecture uses separate VNIR and SWIR encoders, asymmetric cross-spectral attention, SWIR -guided spatial attention gates, and direct VNIR projections in

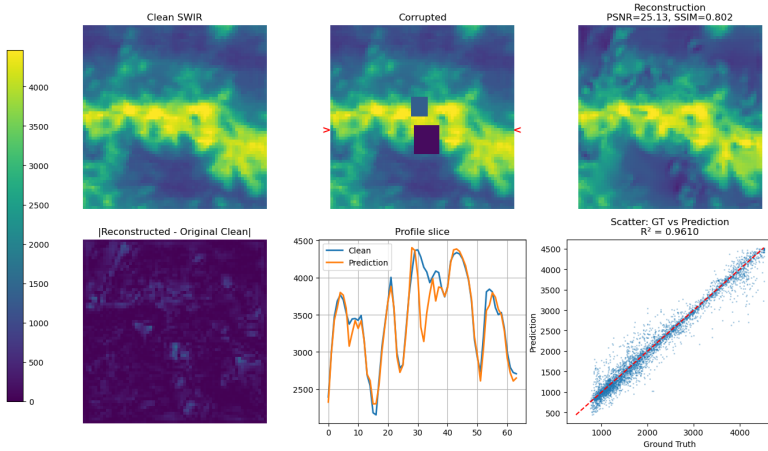


Figure 8: SWIR reconstruction example on a tile from the independent test site. Panels are the same as in Fig. 7. Larger values for PSNR values nearer to 1 are for SSIM show better performance.

the decoder.

A key feature of the proposed design is that cross-spectral attention is applied asymmetrically. VNIR features guide the refinement of SWIR features, while the VNIR stream remains unchanged. This prevents corrupted SWIR information from contaminating the cleaner VNIR structural representation. During decoding, SWIR-guided spatial attention gates selectively filter VNIR skip features, allowing the decoder to use only the VNIR structures most relevant for SWIR reconstruction. The additional direct VNIR projection provides a stable residual pathway for clean structural information at each decoder level.

The model was trained using a composite reconstruction loss that combines robust pixel reconstruction, high-intensity preservation, gradient consistency, VNIR edge weighting, percentile and rank-based distributional constraints, SSIM, and VNIR-guided texture recovery. A two-stage training strategy was used, where the initial training stage learns general SWIR reconstruction and the fine-tuning stage focuses on localized inpainting of block-like corruptions.

Table 5: Comparison with baseline reconstruction methods. Values are reported as mean $\pm$ 95% confidence interval. Larger values For PSNR, and values close to 1 for NCC and SSIM reflect better evaluation.

Method	PSNR (dB)	SSIM	NCC
Corrupted input	23.11 ± 0.60	0.91 ± 0.005	0.93 ± 0.008
Moment matching	17.05 ± 0.36	0.77 ± 0.013	0.82 ± 0.01
FFT notch	18.17 ± 0.26	0.66 ± 0.009	0.85 ± 0.01
Median filter	22.17 ± 0.45	0.88 ± 0.005	0.93 ± 0.009
Proposed model	30.23 ± 3.09	0.95 ± 0.02	0.99 ± 0.007

Experimental results demonstrate that the proposed method effectively reconstructs corrupted SWIR images. On the main test set, the model reconstructed the corrupted input with PSNR 30.23. The reconstructed images also achieved a high SSIM of 0.95, strong normalized correlation of 0.99, and accurate edge preservation of 0.95. Evaluation on an independent forest region further showed that the method generalizes to unseen spatial areas, achieving a PSNR of 28.42 dB and an SSIM of 0.93.

Overall, the proposed VNIR-guided asymmetric reconstruction framework provides an effective approach for recovering degraded SWIR imagery by combining spectral reconstruction with spatial guidance. Future work will focus on extending the method to multi-band SWIR or full hyperspectral cube reconstruction, evaluating the framework on real sensor artifacts beyond synthetic corruptions, and improving robustness under stronger geometric misalignment between VNIR and SWIR imagery.

5.1. Limitations

Although the proposed model shows strong reconstruction performance, several limitations remain. First, the corruption process is synthetic and may not fully capture all real SWIR sensor artifacts. Second, the VNIR-guided texture loss assumes that VNIR and SWIR structures are spatially related, which may not always hold for materials with different spectral behavior. Third, the method currently restores a single SWIR band; extending the framework to multiple SWIR bands or full hyperspectral cube reconstruction is an important direction for future work.

No publicly available deep learning method has been trained or evaluated on the same sensor configuration (SPECIM AFX17 SWIR and MAIA S2

VNIR over forest canopy), and retraining existing methods would require paired corrupted-clean data that is not available for this sensor pair. The comparison is therefore restricted to classical signal-processing baselines and the proposed method.

An ablation study on the architecture of the proposed model and on the selected loss function is presented in Section Appendix C.

5.2. Future Work

In future work, the proposed framework can be extended to reconstruct the complete set of SWIR bands from multispectral drone imagery. While this study focuses on reconstructing a single SWIR band, extending the model to predict multiple SWIR wavelengths would provide a more comprehensive representation of the spectral information. Future investigations can also explore the integration of additional spectral bands, larger and more diverse datasets, and improved cross-sensor adaptation strategies to enhance the generalization capability of the model across different platforms and acquisition conditions.

Acknowledgment

The data used in the study were funded by the Forest Damage Center at the Swedish University of Agricultural Sciences for the data collection [project grants 260FDC05, 260FDC57, 260FDC44, 260FDC12 to LH] and the Kempe Foundation on the hyperspectral drone system [project grant JCSMK22-0126 to LH].

SZ and JL have received financial support from the Swedish Strategic Research Area (SRA) MERGE (Modelling the Regional and Global Earth System). The study is also a contribution to the SRA BECC (Biodiversity and Ecosystem Services in a Changing Climate) and to the Lund University profile area Nature-based future solutions.

Appendix A. Model details

In this section, we go through the details of the proposed model.

Appendix A.1. Detailed Description of the Proposed Model

The proposed model is an asymmetric dual-path U-Net designed for SWIR restoration from synthetically corrupted SWIR observations using clean VNIR guidance. The model receives two inputs:

- **VNIR input** $V \in \mathbb{R}^{B \times 8 \times H \times W}$, processed by the VNIR encoder.
- **Corrupted SWIR input** $S_c \in \mathbb{R}^{B \times 1 \times H \times W}$, processed by the SWIR encoder.

Here, the image size is $H = W = 64$ with 8 channels of VNIR data and 1 SWIR channel, and B is the batch size. The output is a reconstructed clean SWIR image:

$$\hat{S} \in \mathbb{R}^{B \times 1 \times H \times W}.$$

The central idea is that VNIR contains reliable spatial structure, texture, and boundary information, whereas the SWIR input contains the target spectral information but is corrupted. Therefore, the model uses VNIR as a structural guide while reconstructing the clean SWIR image from the degraded SWIR observation.

Let a feature tensor at a given layer be written as:

$$X \in \mathbb{R}^{B \times C \times H_l \times W_l},$$

where B is the batch size, C is the number of channels, and $H_l \times W_l$ is the spatial resolution at level l . We also define the number of pixels as:

$$N = H_l W_l.$$

Appendix A.2. Residual Convolutional Block

The basic convolutional unit of the network is a residual convolutional block. Each block contains two 3×3 convolutional layers, followed by Group Normalization and GELU activation, as shown in Fig. A.9.

For an input feature map x , the main branch is:

$$F(x) = \phi(\text{GN}(W_2 * \phi(\text{GN}(W_1 * x)))),$$

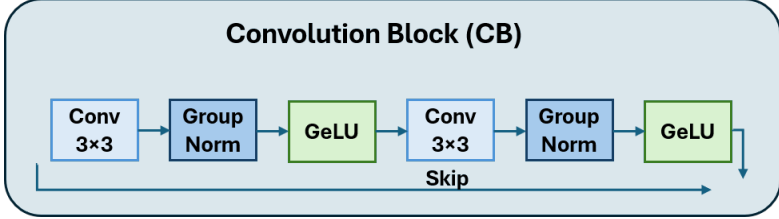


Figure A.9: A schematic figure of the convolution block.

where W_1 and W_2 are 3×3 convolution kernels, $*$ denotes convolution, GN denotes Group Normalization, and $\phi(\cdot)$ denotes the GELU activation function.

The output of the block is:

$$y = F(x) + R(x),$$

where $R(x)$ is the residual shortcut. This residual design improves gradient flow and allows the network to preserve low-level information while increasing the feature dimension across encoder levels. Group Normalization, which normalizes features within each individual sample by dividing channels into groups (μ and σ^2 are computed within channel groups of one sample), is used instead of Batch Normalization because it is stable for small batch sizes and does not mix statistics across samples, which is important for dense SWIR regression.

Appendix A.3. Dual-Encoder Design

The architecture uses two modality-specific encoders. The VNIR encoder processes the clean 8-channel VNIR input, while the SWIR encoder processes the corrupted 1-channel SWIR input. For encoder level l , the feature extraction is written as:

$$V_l = E_l^V(V_{l-1}), \quad S_l = E_l^S(S_{l-1}),$$

where: $V_0 = V$, and $S_0 = S_c$.

With base channel size $ch = 64$, the encoder produces three feature scales:

$$\begin{aligned} V_1, S_1 &\in \mathbb{R}^{B \times 64 \times H \times W}, \\ V_2, S_2 &\in \mathbb{R}^{B \times 128 \times \frac{H}{2} \times \frac{W}{2}}, \\ V_3, S_3 &\in \mathbb{R}^{B \times 256 \times \frac{H}{4} \times \frac{W}{4}}. \end{aligned}$$

Max-pooling is applied between encoder levels, except after the final encoder level. These multi-scale features are stored as skip connections for the decoder.

Appendix A.4. Asymmetric Cross-Spectral Attention

At each encoder level, the VNIR and SWIR features are passed through an asymmetric cross-spectral attention module. Unlike symmetric fusion approaches, this module updates only the SWIR feature stream using VNIR guidance, so it works primarily in feature space. The VNIR feature stream remains unchanged.

Given VNIR and SWIR features of the same shape:

$$V_l, S_l \in \mathbb{R}^{B \times C \times H_l \times W_l},$$

the module computes a query from VNIR and keys and values from SWIR:

$$Q = W_q * V_l, \quad K = W_k * S_l, \quad U = W_v * S_l,$$

where W_q , W_k , and W_v are 1×1 convolutions. The query and key have reduced channel dimension:

$$Q, K \in \mathbb{R}^{B \times d \times H_l \times W_l},$$

where:

$$d = \max\left(\frac{C}{4}, 8\right),$$

and the value tensor has:

$$U \in \mathbb{R}^{B \times C \times H_l \times W_l}.$$

The tensors are reshaped into sequence form:

$$Q \rightarrow \tilde{Q} \in \mathbb{R}^{B \times d \times N}, \quad K \rightarrow \tilde{K} \in \mathbb{R}^{B \times d \times N}, \quad U \rightarrow \tilde{U} \in \mathbb{R}^{B \times C \times N},$$

where $N = H_l W_l$.

To avoid the quadratic memory cost of full spatial attention, a linear attention formulation is used. The query is normalized over spatial locations [12]:

$$\tilde{Q}' = \text{softmax}(\tilde{Q}, \dim = N),$$

and the key is normalized over the feature dimension:

$$\tilde{K}' = \text{softmax}(\tilde{K}, \dim = d).$$

A compact global context matrix is computed as:

$$C_{\text{ctx}} = \tilde{K}' \tilde{U}^T,$$

where:

$$C_{\text{ctx}} \in \mathbb{R}^{B \times d \times C}.$$

The attention output is then:

$$A = C_{\text{ctx}}^T \tilde{Q}',$$

where:

$$A \in \mathbb{R}^{B \times C \times N}.$$

which is reshaped to:

$$A \in \mathbb{R}^{B \times C \times H_l \times W_l}.$$

The SWIR feature stream is updated using a residual attention connection:

$$S'_l = \text{GN}(S_l + \gamma A),$$

where γ is a learnable scalar initialized to zero. The VNIR stream remains unchanged:

$$V'_l = V_l.$$

Thus, the attention module implements:

$$(V_l, S_l) \longrightarrow (V_l, S'_l).$$

This asymmetric design, also shown in Fig. A.10, is important because the SWIR input is corrupted. Allowing corrupted SWIR features to modify VNIR features could degrade the clean spatial guidance signal. Therefore, VNIR acts as a stable structural reference, while only the SWIR representation is refined.

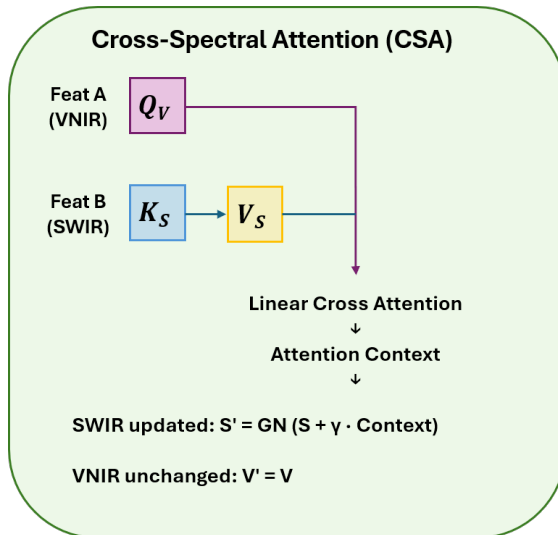


Figure A.10: Asymmetric cross-spectral attention: clean VNIR features guide the corrupted SWIR stream. VNIR generates queries, SWIR generates keys and values, and only the SWIR feature map is updated. The VNIR stream remains unchanged to avoid contamination by corrupted SWIR information.

Appendix A.5. Bottleneck Fusion

After the final encoder level, the deepest VNIR and SWIR features are down-sampled and concatenated along the channel dimension:

$$B_0 = [P(V_3), P(S_3)],$$

where $P(\cdot)$ denotes max-pooling and $[\cdot, \cdot]$ denotes channel-wise concatenation.

For base channel size $ch = 64$, V_3 and S_3 each have 256 channels. After pooling and concatenation:

$$B_0 \in \mathbb{R}^{B \times 512 \times \frac{H}{8} \times \frac{W}{8}}.$$

The bottleneck then processes this fused representation using residual convolutional blocks:

$$B = F_{\text{bottleneck}}(B_0).$$

This bottleneck representation combines high-level VNIR structural information with SWIR spectral information at the coarsest scale.

Appendix A.6. SWIR-Guided Spatial Attention Gate

During decoding, VNIR skip features are filtered using a SWIR-guided spatial attention gate. The goal is to decide which VNIR spatial structures are useful for reconstructing the clean SWIR image, so it works primarily in spatial space.

At a decoder level, let: $x \in \mathbb{R}^{B \times F_l \times H_l \times W_l}$ be the VNIR skip feature and: $g \in \mathbb{R}^{B \times F_g \times H_l \times W_l}$ be the corresponding SWIR feature used as the gating signal.

Both inputs are projected to an intermediate feature space:

$$g_1 = \text{GN}(W_g * g), \quad x_1 = \text{GN}(W_x * x),$$

where W_g and W_x are 1×1 convolutions.

The projected features are combined additively:

$$z = \text{ReLU}(g_1 + x_1).$$

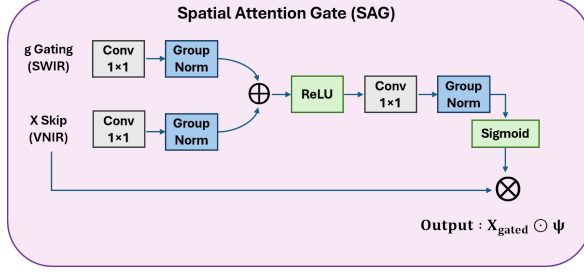


Figure A.11: Spatial Attention Gate used in the decoder. The SWIR feature g acts as a gating signal for the VNIR skip feature x . Both features are projected with 1×1 convolutions and Group Normalization, combined additively, and passed through ReLU, 1×1 convolution, Group Normalization, and *sigmoid* activation to generate a spatial attention map Ψ . The VNIR skip feature is then filtered by element-wise multiplication, producing the gated VNIR feature $x_{gated} = x \odot \Psi$, which is passed to the decoder for SWIR reconstruction.

A spatial attention map is produced by:

$$\Psi = \sigma(\text{GN}(W_{\Psi} * z)),$$

where: $\Psi \in \mathbb{R}^{B \times 1 \times H_l \times W_l}$, and $\sigma(\cdot)$ denotes the sigmoid function.

The VNIR skip feature is modulated by this attention map:

$$x_{att} = x \odot \Psi,$$

where $\odot$ denotes broadcasted element-wise multiplication. For each spatial location (i, j) :

$$x_{att}(i, j) = \Psi(i, j)x(i, j), \quad \Psi(i, j) \in [0, 1].$$

Thus, $\Psi(i, j)$ acts as a learned importance coefficient. Locations with values close to 1 are preserved, while locations with values close to 0 are suppressed. Since the attention mask is computed using SWIR features as the gating signal, the decoder selectively receives VNIR structures that are relevant to SWIR restoration. Fig. A.11 shows the spatial attention gate explained.

Appendix A.7. Direct VNIR Projection

In addition to the gated VNIR skip connection, the model includes a direct VNIR projection at each decoder level. The VNIR encoder feature is passed through a 1×1 convolution:

$$V_l^{\text{dir}} = W_l^{\text{dir}} * V_l.$$

This direct projection matches the channel dimension required by the decoder and provides an independent residual path for clean VNIR structural information. This is useful because the spatial attention gate may suppress some VNIR features. The direct path ensures that the decoder still has access to clean VNIR structure even when the gated skip is conservative.

Appendix A.8. Decoder and Upsampling

The decoder reconstructs the restored SWIR image progressively from the bottleneck representation. Upsampling is performed using bilinear interpolation followed by a convolution:

$$D_l = W_l^{\text{up}} * \text{Interp}(D_{l+1}),$$

where $\text{Interp}(\cdot)$ denotes bilinear upsampling by a factor of 2.

At each decoder stage, four sources of information are fused:

- the upsampled decoder feature,
- the SWIR-guided gated VNIR skip feature,
- the corresponding SWIR skip feature,
- the direct VNIR projection.

For decoder level l , the fused feature is:

$$F_l = [D_l, V_l^{\text{gate}}, S_l, V_l^{\text{dir}}],$$

where:

$$V_l^{\text{gate}} = \text{AG}(g = S_l, x = V_l).$$

The decoder output at that level is then:

$$D'_l = \text{Dec}_l(F_l),$$

where Dec_l is a residual ConvBlock.

This means that each decoder stage receives both spectral information from the SWIR stream and structural information from the VNIR stream. The VNIR information enters in two forms: a selective gated skip and a direct projected skip.

Appendix A.9. Loss Function

Let $\hat{S}$, S , and S_c denote the model prediction, the clean SWIR target, and the corrupted SWIR input, respectively, and let V represent the VNIR input, all having the same dimensions as in Section Appendix A.

The loss combines robust reconstruction, high-intensity preservation, gradient consistency, VNIR edge weighting, distribution matching, rank preservation, VNIR-guided texture recovery, and SSIM. The full objective is designed to preserve both the radiometric accuracy of the SWIR band and the spatial structure guided by VNIR.

Gradient loss. We use Sobel kernels for horizontal and vertical gradients [8]:

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad K_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

The loss is:

$$\mathcal{L}_{\text{grad}} = \|K_x * \hat{S} - K_x * S\|_1 + \|K_y * \hat{S} - K_y * S\|_1.$$

This term, which compares horizontal and vertical gradients between prediction and clean SWIR, encourages the restored SWIR image to preserve sharp local transitions and reduces over-smoothing.

Charbonnier reconstruction loss. The main pixel-wise reconstruction term is the Charbonnier loss:

$$\mathcal{L}_{\text{recon}} = \frac{1}{BHW} \sum_{b,i,j} \sqrt{(\hat{S}_{b,i,j} - S_{b,i,j})^2 + \epsilon^2}.$$

Here, ϵ is a small positive constant (we use $\epsilon = 10^{-3}$) introduced for numerical stability rather than any physical interpretation. Without it, the loss reduces to the standard L_1 norm, $|\hat{S}_{b,i,j} - S_{b,i,j}|$, whose gradient is discontinuous at

zero error and can destabilize optimization as predictions converge toward the target. The addition of ϵ^2 inside the square root smooths this discontinuity: for small residuals the loss behaves approximately quadratically (as in L_2), yielding a gradient that vanishes smoothly near zero, while for larger residuals it retains the near-linear, outlier-robust behavior of L_1 . This combination makes the Charbonnier loss both numerically stable and robust to outlier pixels, which is desirable for dense SWIR regression.

High-intensity weighted loss. A high-value mask is defined as:

$$M_{\text{high}}(b, i, j) = \mathbb{I}(S(b, i, j) > 0.6).$$

The high-intensity loss is:

$$\mathcal{L}_{\text{high}} = \frac{\sum_{b,i,j} M_{\text{high}}(b, i, j) S(b, i, j) |\hat{S}(b, i, j) - S(b, i, j)|}{\sum_{b,i,j} M_{\text{high}}(b, i, j) S(b, i, j) + \epsilon}.$$

The high-intensity loss prevents bright SWIR structures from being suppressed or regressed toward the mean, as neural networks often predict conservative mean values.

VNIR edge-weighted loss. The last VNIR channel (closest to the SWIR) is used to compute an edge-weighting map:

$$V_{\text{nir}} = V_{:,8,:},$$

The VNIR edge magnitude is:

$$E_V = \sqrt{(K_x * V_{\text{nir}})^2 + (K_y * V_{\text{nir}})^2 + \epsilon}.$$

The normalized edge weight is:

$$\tilde{E}_V = \frac{E_V}{\text{mean}(E_V) + \epsilon}.$$

The edge-weighted reconstruction loss is:

$$\mathcal{L}_{\text{edge}} = \frac{1}{BHW} \sum_{b,i,j} \tilde{E}_V(b, i, j) |\hat{S}(b, i, j) - S(b, i, j)|.$$

This term places stronger emphasis on reconstruction errors near VNIR-derived spatial boundaries.

Percentile loss. To preserve the global radiometric distribution, selected percentiles of the predicted and clean SWIR images are compared. For

$$\mathcal{P} = \{10, 25, 50, 75, 90, 95, 99\},$$

The percentile loss is:

$$\mathcal{L}_{\text{pct}} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} |Q_p(\hat{S}) - Q_p(S)|,$$

where $Q_p(\cdot)$ denotes the p -th percentile. This checks whether the predicted SWIR image has a similar intensity distribution to the clean SWIR image.

Rank-based patch loss. The prediction and target are divided into non-overlapping patches. Let $m_i^{\hat{S}}$ and m_i^S denote the mean intensities of patch i in the prediction and target. Pairwise differences are:

$$d_{ij}^{\hat{S}} = m_i^{\hat{S}} - m_j^{\hat{S}}, \quad d_{ij}^S = m_i^S - m_j^S.$$

The rank loss is:

$$\mathcal{L}_{\text{rank}} = \left\| \tanh(5d_{ij}^{\hat{S}}) - \tanh(5d_{ij}^S) \right\|_1.$$

$\tanh$ is just a practical choice. Without it, a single pair of patches with a very large brightness difference could dominate the whole loss and make training unstable. $\tanh$ squashes large differences into the range $(-1, 1)$, so all pairs contribute roughly equally regardless of how extreme the difference is. The scale factor of 5 is used so that the $\tanh$ function responds nonlinearly to typical patch-mean differences in normalized SWIR images. This limits the effect of very large differences while still keeping the method sensitive to small differences. When the network reconstructs a corrupted region, the easiest thing to do is predict a value close to the global mean, resulting in low reconstruction error but spatially flat. However, flat prediction, where all patches have the same mean, will get every pairwise comparison wrong, producing a large penalty. This forces the network to recover the spatial contrast structure of the image, not just the average intensity. It is complementary to the percentile loss, which checks the global intensity distribution, whereas the rank loss checks local relative ordering across space. The rank-based patch loss is inspired by pairwise ranking objectives [2] and encourages the predicted image to preserve the relative spatial contrast structure of the clean SWIR target across local regions.

SSIM loss. The SSIM loss is [27]:

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\hat{S}, S).$$

A 7×7 local window is used, which is suitable for small image tiles. This term encourages local structural similarity between the restored and clean SWIR images.

VNIR-guided texture loss. When a region of the SWIR image is corrupted, the network receives no useful signal from those pixels and must reconstruct them from context and cross-spectral guidance. To support this, we introduce a texture loss that encourages the reconstructed SWIR image to recover spatial structure that is consistent with the clean VNIR image, but only inside the corrupted regions. The VNIR image is always clean and provides reliable information about canopy boundaries, tree crown structure, and local texture patterns that are spatially correlated with the SWIR response in forest environments.

Step 1: Detecting corrupted regions. Rather than using the ground-truth corruption mask, the loss estimates which regions are corrupted directly from the corrupted SWIR input S_c . The key observation is that block corruption fills regions with a constant or near-constant value, producing locally flat areas with very low variance. We compute the local variance of S_c using average pooling:

$$\mu_c = \text{AvgPool}(S_c), \quad \sigma_c^2 = \text{AvgPool}(S_c^2) - \mu_c^2.$$

The variance is then normalised to $[0, 1]$ and inverted, so that flat (likely corrupted) regions receive high scores:

$$M_c = 1 - \frac{\sigma_c^2}{\max(\sigma_c^2) + \epsilon}.$$

A binary mask is produced by thresholding at the median of M_c :

$$\bar{M}_c = \mathbb{I}(M_c > \text{median}(M_c)).$$

Locations where $\bar{M}_c = 1$ are treated as corrupted. The median threshold ensures that a fixed proportion of spatial locations is always selected, providing a stable signal regardless of corruption severity.

Step 2: VNIR structural reference. The VNIR input V contains eight spectral bands. To obtain a single-channel spatial reference that can be compared with the single-channel SWIR prediction, we average all eight bands:

$$V_{\text{ref}}(b, 1, i, j) = \frac{1}{8} \sum_{c=1}^8 V(b, c, i, j).$$

This produces a greyscale image that captures the spatial structure of the scene as seen in the VNIR domain.

Step 3: Gradient alignment inside corrupted regions. We compute the spatial gradient magnitude of both the predicted SWIR image $\hat{S}$ and the VNIR reference V_{ref} using Sobel operators K_x and K_y :

$$G(I) = \sqrt{(K_x * I)^2 + (K_y * I)^2 + \epsilon}.$$

The gradient maps are then masked to retain only the corrupted regions:

$$G_S^m = \bar{M}_c \odot G(\hat{S}), \quad G_V^m = \bar{M}_c \odot G(V_{\text{ref}}).$$

The loss encourages the gradient pattern of the reconstruction to match the gradient pattern of the VNIR image inside corrupted areas. Rather than penalising absolute differences in gradient magnitude (which would force the SWIR to copy VNIR values exactly), we use normalised cross-correlation (NCC), which measures the similarity of spatial patterns independently of their overall scale:

$$\text{NCC}(A, B) = \frac{\sum(A - \bar{A})(B - \bar{B})}{\sqrt{\sum(A - \bar{A})^2} \sqrt{\sum(B - \bar{B})^2 + \epsilon}},$$

where $\bar{A}$ and $\bar{B}$ are spatial means. The gradient alignment loss is:

$$\mathcal{L}_{\text{grad-NCC}} = 1 - \text{NCC}(G_S^m, G_V^m).$$

This loss equals zero when the two gradient maps are perfectly correlated and increases as they diverge. In practice, it encourages the reconstructed SWIR image to place edges and boundaries at the same spatial locations as the VNIR image inside corrupted regions, without requiring the edge magnitudes to match exactly.

Step 4: Curvature alignment inside corrupted regions. Gradient-based losses capture where intensity changes occur, but not the finer texture structure within those regions. To address this, we add a second component based on the Laplacian operator K_Δ , which measures the local curvature of the image detecting blob-like structures, fine texture, and points where intensity changes direction [8]:

$$K_\Delta = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The masked Laplacian loss directly penalises differences in curvature between the prediction and the VNIR reference, averaged over the corrupted region:

$$\mathcal{L}_{\text{lap}} = \frac{\sum \bar{M}_c |K_\Delta * \hat{S} - K_\Delta * V_{\text{ref}}|}{\sum \bar{M}_c + \epsilon}.$$

This term is given half the weight of the gradient term because the Laplacian is more sensitive to small geometric misalignments between the VNIR and SWIR sensors, which always exist at the centimetre level after co-registration.

Combined loss. The full VNIR-guided texture loss combines both components:

$$\mathcal{L}_{\text{VNIR-tex}} = \mathcal{L}_{\text{grad-NCC}} + 0.5 \mathcal{L}_{\text{lap}}.$$

This loss is only active inside the estimated corruption mask $\bar{M}_c$ and has no effect on uncorrupted regions, where the reconstruction is already supervised by the direct SWIR target. The assumption underlying this term is that VNIR and SWIR spatial structures are correlated in forest environments, where both modalities respond to the physical structure of the vegetation canopy. This assumption may not hold for all surface types, which is discussed further in Section 5.1.

Although the spatial attention gate produces a mask Ψ that implicitly identifies unreliable SWIR regions, we use this variance-based method here to define the corruption mask for the texture loss for two reasons. First, Ψ is only available after the full forward pass, creating a circular dependency if used to define the loss that trains the same weights. Second, the attention mechanism is initialised to contribute nothing ($\gamma = 0$), so Ψ is uninformative during the early training epochs when the texture supervision is most needed.

Total loss. The final training objective is:

$$\begin{aligned}\mathcal{L}_{\text{total}} = & \lambda_{\text{recon}}\mathcal{L}_{\text{recon}} + \lambda_{\text{high}}\mathcal{L}_{\text{high}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}} + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}} + \lambda_{\text{pct}}\mathcal{L}_{\text{pct}} \\ & + \lambda_{\text{rank}}\mathcal{L}_{\text{rank}} + \lambda_{\text{tex}}\mathcal{L}_{\text{VNIR-tex}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{SSIM}}.\end{aligned}$$

For the initial training stage, the weights are:

$$\begin{aligned}\lambda_{\text{recon}} = 1.0, \quad \lambda_{\text{high}} = 1.0, \quad \lambda_{\text{grad}} = 0.5, \quad \lambda_{\text{edge}} = 0.8, \\ \lambda_{\text{pct}} = 0.8, \quad \lambda_{\text{rank}} = 0.5, \quad \lambda_{\text{tex}} = 2.0, \quad \lambda_{\text{ssim}} = 1.0.\end{aligned}$$

For fine-tuning, the weights are:

$$\begin{aligned}\lambda_{\text{recon}} = 1.0, \quad \lambda_{\text{high}} = 0.5, \quad \lambda_{\text{grad}} = 0.5, \quad \lambda_{\text{edge}} = 0.5, \\ \lambda_{\text{pct}} = 0.3, \quad \lambda_{\text{rank}} = 0.3, \quad \lambda_{\text{tex}} = 0.5, \quad \lambda_{\text{ssim}} = 1.0.\end{aligned}$$

The initial loss places stronger emphasis on VNIR-guided texture recovery, while the fine-tuning loss reduces this weight to preserve the accurate reconstruction already learned during initial training and to focus on localized inpainting improvements.

Appendix B. Evaluation metrics

The reconstructed SWIR image $\hat{S}$ was evaluated against the clean reference SWIR image S using PSNR, SSIM, NCC, Sobel-Correlation, SAM, and R^2 .

The peak signal-to-noise ratio is defined as [9]:

$$\text{PSNR}(S, \hat{S}) = 10 \log_{10} \left(\frac{L^2}{\text{MSE}(S, \hat{S})} \right),$$

where

$$\text{MSE}(S, \hat{S}) = \frac{1}{HW} \sum_{i,j} \left(S(i, j) - \hat{S}(i, j) \right)^2,$$

and L denotes the data range of the clean SWIR image. Higher PSNR indicates lower reconstruction error.

The structural similarity index is computed as [25]:

$$\text{SSIM}(S, \hat{S}) = \frac{(2\mu_S\mu_{\hat{S}} + C_1)(2\sigma_{S\hat{S}} + C_2)}{(\mu_S^2 + \mu_{\hat{S}}^2 + C_1)(\sigma_S^2 + \sigma_{\hat{S}}^2 + C_2)}.$$

SSIM evaluates local luminance, contrast, and structural similarity, with values closer to 1 indicating better reconstruction.

Normalized cross-correlation is defined as [13]:

$$\text{NCC}(S, \hat{S}) = \frac{\sum_{i,j} (S(i,j) - \bar{S}) (\hat{S}(i,j) - \bar{\hat{S}})}{\sqrt{\sum_{i,j} (S(i,j) - \bar{S})^2} \sqrt{\sum_{i,j} (\hat{S}(i,j) - \bar{\hat{S}})^2} + \epsilon}.$$

This metric measures pixel-wise correlation between the prediction and reference.

To evaluate edge preservation, Sobel-Correlation is computed from gradient magnitude maps:

$$G(I) = \sqrt{(\text{Sobel}_x(I))^2 + (\text{Sobel}_y(I))^2},$$

$$\text{Sobel-Corr}(S, \hat{S}) = \text{corr}(G(S), G(\hat{S})).$$

A higher Sobel-Correlation indicates better preservation of edges and local spatial transitions.

The spectral angle mapper is computed by treating the image as a vector [27]:

$$\text{SAM}(S, \hat{S}) = \arccos \left(\frac{\langle \text{vec}(S), \text{vec}(\hat{S}) \rangle}{\|\text{vec}(S)\|_2 \|\text{vec}(\hat{S})\|_2 + \epsilon} \right).$$

Lower SAM values indicate stronger directional agreement between the reconstructed and clean SWIR patterns.

Finally, the coefficient of determination is:

$$R^2 = 1 - \frac{\sum_{i,j} (S(i,j) - \hat{S}(i,j))^2}{\sum_{i,j} (S(i,j) - \bar{S})^2 + \epsilon}.$$

Values closer to 1 indicate that the prediction explains a larger proportion of the variance in the clean SWIR target.

Appendix C. Ablation study

In this section, we present results from different model architectures and loss compositions.

Appendix C.1. Model Architecture

To evaluate the contribution of the two key architectural components, we trained two ablation variants under identical conditions: the same training schedule, loss function, and data splits as the full model.

The first variant, *symmetric attention*, replaces the asymmetric cross-spectral attention with a bidirectional version in which SWIR features are also allowed to update the VNIR stream. This removes the directional constraint that keeps the VNIR representation clean throughout encoding. The second variant, *no spatial attention gate* (No-SAG), replaces the SWIR-guided spatial attention gate in the decoder with a plain 1×1 projection, so VNIR skip features are passed to the decoder without any context-dependent filtering.

Table C.6 reports full-image metrics for both variants alongside the full model. Removing the asymmetric constraint and allowing corrupted SWIR features to modify the VNIR stream reduced PSNR and SSIM, confirming that VNIR stream contamination during encoding measurably degrades restoration quality. Removing the spatial attention gate produced a smaller but consistent drop in PSNR and in SSIM, indicating that selective filtering of VNIR skip features contributes to reconstruction accuracy, though the decoder partially compensates through the direct VNIR projection path that remains in place. Together, these results support both design choices: the asymmetric attention constraint has the larger effect, while the spatial gate provides an additional and independent improvement.

Appendix C.2. Loss Function Ablation

To evaluate the contribution of each term in the composite restoration loss, we trained seven variants of the full model, each with one term removed by setting its weight to zero. All other training settings, including architecture, optimiser, schedule, and two-stage training procedure, were kept identical. Table C.7 reports full-image metrics for each variant, and Fig. C.12 shows reconstructions on the same corrupted tile across all variants.

The most critical term is the high-intensity loss. Removing it caused complete model collapse: PSNR dropped to 6.05 dB, SSIM to 0.02, and NCC

Table C.6: Ablation study on 224 test tiles with five corruptions per tile. All variants use VNIR guidance and identical training configurations. The full model includes both asymmetric cross-spectral attention and SWIR-guided spatial attention gating. For PSNR, larger values indicate better performance, whereas for NCC, SSIM, Sobel-cor, and R^2 , values nearer to 1 are preferable; for SAM, smaller values reflect better evaluation.

Variant	PSNR (dB)	SSIM	NCC	Sobel-Corr	SAM (rad)
Symmetric attention	27.83	0.93	0.98	0.94	0.07
No spatial gate	28.22	0.94	0.99	0.95	0.06
Full model	30.23	0.95	0.99	0.95	0.05

became negative, indicating the model predicted near-constant near-zero output. This is visible in Fig. C.12, where the *no_high* reconstruction is uniformly dark. Without explicit upweighting of bright SWIR responses, the network regresses toward the mean of the training distribution and suppresses the high-reflectance regions that carry the most diagnostic information in SWIR imagery. This confirms that the high-intensity loss is not a refinement but a necessary condition for stable training.

Removing the SSIM loss produced the second-largest degradation, reducing PSNR by 2.5 dB and Sobel correlation from 0.95 to 0.91. The reconstructions in Figure C.12 show visibly softer spatial structure and reduced contrast in the *no_ssim* row, consistent with the role of SSIM in enforcing local luminance and contrast consistency.

Removing the VNIR-guided texture loss reduced PSNR by 1.0 dB and slightly increased SAM from 0.05 to 0.07 rad. Visually, the *no_vnir_tex* reconstruction recovers the broad spatial structure but shows weaker high-frequency detail inside corrupted regions, which is the specific behaviour this term was designed to address. Since the texture loss is applied only within detected corruption masks and encourages gradient and Laplacian alignment with the VNIR reference, its removal disproportionately affects restoration quality in degraded areas rather than the uncorrupted background.

Removing the rank loss produced a reduction of 0.63 dB in PSNR with NCC and Sobel correlation decreasing by 0.001 and SAM increasing to 0.071 rad. The rank loss penalises violations of relative intensity ordering across local patches, so its removal allows the model to predict globally plausible pixel values while losing fidelity to the spatial contrast structure of the original

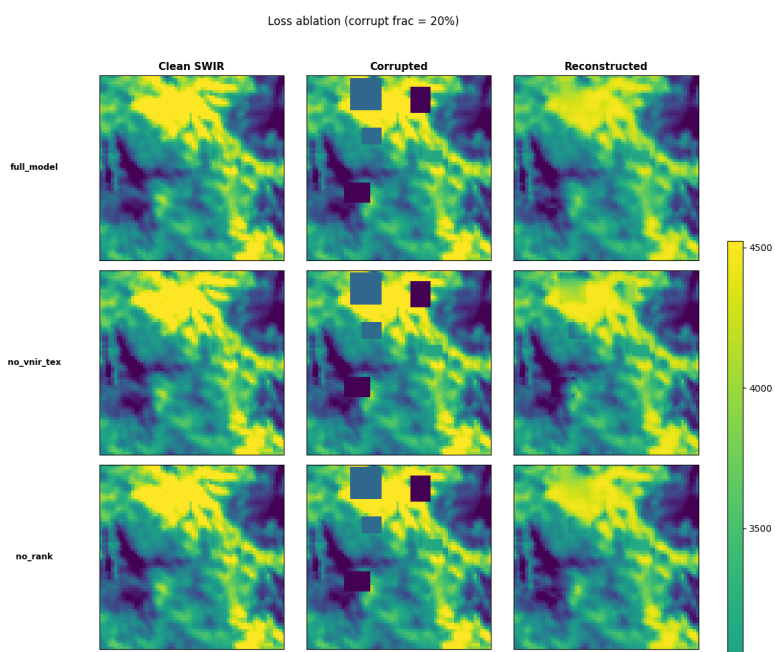
SWIR image. This effect is subtle in single-tile visualisations but consistent across the full test set. Removing the percentile loss produced a similarly modest drop of 0.27 dB; together, the rank and percentile terms act as complementary distribution-matching constraints, the former preserving local relative ordering and the latter aligning the global intensity distribution, and their individual contributions are moderate precisely because each partially compensates for the absence of the other.

Removing the edge loss reduced PSNR by 1.9 dB and Sobel correlation from 0.95 to 0.93, indicating that VNIR-guided boundary weighting contributes meaningfully to spatial transition accuracy.

Overall, the ablation demonstrates that the composite loss is not over-engineered: each term addresses a distinct failure mode, with the high-intensity, SSIM, edge, and VNIR texture terms having the largest individual effects and the rank and percentile terms providing consistent but moderate distributional regularisation.

Table C.7: Loss function ablation on 224 test tiles with five corruptions per tile. Each variant removes one loss term by setting its weight to zero. All other settings are identical to the full model. For PSNR, larger values indicate better performance, whereas for NCC, SSIM, Sobel-cor, and R^2 , values nearer to 1 are preferable; for SAM, smaller values reflect better evaluation.

Variant	PSNR (dB)	SSIM	NCC	Sobel Corr	SAM (rad)
No percentile loss	28.36	0.94	0.99	0.95	0.06
No rank loss	27.97	0.94	0.99	0.95	0.07
No VNIR texture	27.63	0.94	0.98	0.95	0.07
No edge loss	26.71	0.93	0.98	0.93	0.07
No SSIM loss	26.11	0.88	0.98	0.91	0.08
No high-intensity loss	6.05	0.02	-0.06	0.00	1.57
Full model	30.23	0.95	0.99	0.95	0.05



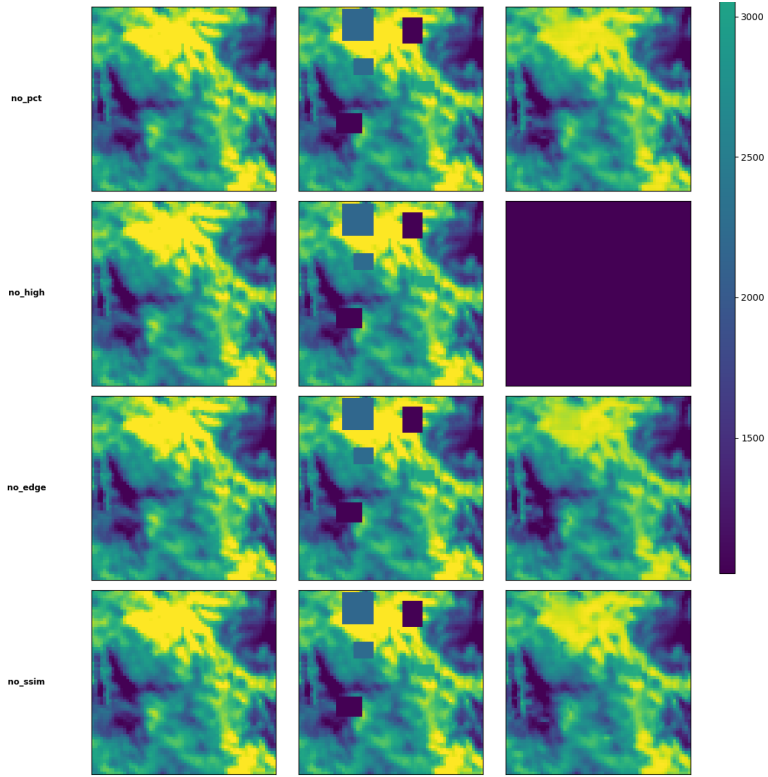


Figure C.12: Continued. Qualitative comparison of loss ablation variants on a single test tile with 20% synthetic corruption. Each row shows the clean SWIR target, the corrupted input, and the reconstruction produced by the corresponding variant. All images share the same colour scale. Removing the high-intensity loss causes complete output collapse (*no_high*, row 5). Removing the SSIM or VNIR texture loss produces visibly softer reconstructions with reduced spatial detail compared to the full model (top row).

References

- [1] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.
- [2] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G. (2005). Learning to rank using gradient descent. In *ICML*, pages 89–96.
- [3] Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., and Feng, J. (2021). Multi-spectral image fusion using deep learning: A review. *Information Fusion*, 76:325–342.
- [4] Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of Remote Sensing*, 3:3–58.
- [5] Cosimo, L. H. E., Lopez, J. G., Lindberg, E., and Huo, L. (2026). Linking hyperspectral remote sensing and tree hydraulic functioning for pre-emergence detection of bark beetle infestations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 237:628–651.
- [6] Dawson, N., Fischer, J., Kuhn, M., Pasotti, A., mhugent, Rouzaud, D., Bruy, A., Sutton, T., Dobias, M., Pellerin, M., Rouault, E., Olaya, V., Blottiere, P., Macho, W., Blazek, R., Sherman, G., Sant-anna, H., Cabieces, J., Woodrow, N., Bartoletti, L., signedav, rldhont, Shaffer, L., Natsis, S., Belgacem, N., Santilli, S., Larosa, S., Mani, S., Cloarec, V., and Jurgiel, B. (2025). qgis/qgis: 3.44.0.
- [7] Deng, L., Sun, J., Chen, Y., Lu, H., Duan, F., Zhu, L., and Fan, T. (2021). M2h-net: A reconstruction method for hyperspectral remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:323–348.
- [8] Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall, 3rd edition.
- [9] Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.

- [10] Huo, L., Cosimo, L. H. E., Bozzini, A., Bijou, S., de Oliveira, R. A., Suomalainen, J., Vennervirta, E., Koivumäki, N., Näsi, R., Kupková, L., Faccoli, M., and Honkavaara, E. (2026). A cross-european assessment on the pre-emergence detection of trees attacked by spruce bark beetle using uav imagery. *Remote Sensing of Environment*, 342:115445.
- [11] Huo, L., Lindberg, E., Bohlin, J., and Persson, H. J. (2023). Assessing the detectability of european spruce bark beetle green attack in multispectral drone images with high spatial- and temporal resolutions. *Remote Sensing of Environment*, 287:113484.
- [12] Katharopoulos, A., Vyas, A., Pappas, N., and Fleuret, F. (2020). Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pages 5156–5165. PMLR.
- [13] Kirch, W., editor (2008). *Pearson’s Correlation Coefficient*, pages 1090–1091. Springer Netherlands, Dordrecht.
- [14] Li, T. and Gu, Y. (2021). Progressive spatial–spectral joint network for hyperspectral image reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14.
- [15] Loncan, L., de Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simões, M., et al. (2015). Hyperspectral pansharpening: A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46.
- [16] Neagoe, I. C., Faur, D., Vaduva, C., and Datcu, M. (2023). Band reconstruction using a modified unet for sentinel-2 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:6739–6757.
- [17] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. In *arXiv preprint arXiv:1804.03999*.
- [18] Ronneberger, O., Fischer, P., and Brox, T. (2015a). U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer.

- [19] Ronneberger, O., Fischer, P., and Brox, T. (2015b). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- [20] Rout, L. (2020). Alert: Adversarial learning with expert regularization using tikhonov operator for missing band reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4395–4405.
- [21] Rout, L., Bhateja, Y., Garg, A., Mishra, I., Moorthi, S. M., and Dhar, D. (2019). Deepswir: A deep learning based approach for the synthesis of short-wave infrared band using multi-sensor concurrent datasets. *arXiv preprint arXiv:1905.02749*.
- [22] Sidorov, O. and Hardeberg, J. Y. (2019). Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3844–3851.
- [23] Thenkabail, P. S., Lyon, J. G., and Huete, A. (2012). *Hyperspectral Remote Sensing of Vegetation*. CRC Press.
- [24] Tsagkatakis, G., Aidini, A., Fotiadou, K., Giannopoulos, M., Pentari, A., and Tsakalides, P. (2019). Survey of deep-learning approaches for remote sensing observation enhancement. *Sensors*, 19(18).
- [25] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.
- [26] Yokoya, N., Grohnfeldt, C., and Chanussot, J. (2017). Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56.
- [27] Yuhas, R. H., Goetz, A. F., and Boardman, J. W. (1992). Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, summaries of the third annual JPL airborne geoscience workshop. volume 1: AVIRIS workshop*.

- [28] Zhang, Q., Yuan, Q., Zeng, C., Li, X., and Wei, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4274–4288.
- [29] Zhu, T., Liu, Q., and Zhang, L. (2024). Hierarchical spectral–spatial transformer for hyperspectral and multispectral image fusion. *Remote Sensing*, 16(22):4127.