Towards physics-informed machine learning for wildfire simulations

A spatial temporal conditional autoregressive model

MSc. Complex Systems Engineering and Management

Gerben Bultema

Delft University of Technology

TU Delft

Towards physics-informed machine learning for wildfire simulations

A spatial temporal conditional autoregressive model

by

Gerben Bultema

In partial fulfillment of the requirements for the degree of:

Master of Science
in Complex Systems Engineering and Management

at the Delft University of Technology,
to be defended publicly on Monday August 28, 2025 at 14:30.

Supervisor:	Dr. O. Kammouh	Delft University of Technology
	Dr. Xiao Liu	Georgia Institute of Technology
Thesis committee:	Prof.dr.ir. P.H.A.J.M. van Gelder	Delft University of Technology
	Dr.ir. M. Nogal Macho	Delft University of Technology

An electronic version of this thesis is available at http:repository.tudelft.nl

TU Delft

Preface

When I first e-mailed Dr. Omar Kammouh in October 2024 inquiring if he knew of any interesting Master's thesis topics using simulation, I could have never dreamt of the truly one in a million opportunity that he would present. This Master's thesis that lies before you is the result of that opportunity and concludes my master's degree in Complex Systems Engineering and Management. It was an opportunity that allowed me to explore a different culture, meet so many amazing people and dive deeper into the aspects of my study that I enjoyed the most. It allowed me to explore how wildfires and wildfire simulations work. A very relevant topic as was evident from the wildfires in Los Angeles that occurred as the project started. From the bottom of my heart, I want to thank everyone that made this possible and that made this opportunity into the amazing time that it has been.

I want to thank my committee and especially Dr. Xiao Liu of the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology for facilitating this opportunity and for all the support throughout the countless meetings. Developing a machine learning model like we did is not something trivial and not something that is normally taught in my program. Your guidance has helped and inspired me much more than you may realize. I want to also specifically thank Dr. Omar Kammouh for providing me with many resources to improve the thesis and for providing all the advice I could ask for.

Additionally I want to thank everyone that made my stay in the United States into what it was. I want to thank Ricardo, Brandon, Nathan, Spencer and all other friends I made during my time in the United States. You all have provided me with the much needed mental support and made this a truly unforgettable time. I never would have experienced as much as I did without you. Time flew by faster for me than it ever has.

Finally, I want to thank everyone that I met throughout my studies. From the people I had group projects with to those I was a teaching assistants with. I look back on a time I truly enjoyed because of you.

Thank you.

Gerben Bultema
Delft, August 2025

Executive summary

Wildfire simulations have become increasingly important as their frequency and severity increases, posing a threat to communities and resulting in billions in damages. However, current wildfire models face a trade-off between accuracy and computational complexity. Current wildfire models can be physics-based, but computationally expensive, or based on empirical data, which allows for better computational speed, but decreases the physical basis. A model combining the computational speed of empirical models with the physics understanding of physics-based models to increase the accuracy of wildfire simulations is desired. This master thesis explores the design of such a model by answering the question: 'How can a near real-time wildfire simulation be designed using physics-informed machine learning?'.

To do this effectively with a focus on the interdisciplinary nature of wildfire simulation, the design science research methodology was used. This required identifying stakeholders, applications and their respective needs and requirements through a literature review. Operational firefighting was chosen as a main focus due to the need for accurate, interpretable, near real-time results. These formed the main requirements with other requirements covering the adaptability to different data types and use cases as well as ability to handle uncertain data. As the requirements were formulated based of off model usage and no specific requirements could be found, these remained at a high level.

After identifying the requirements, the design space was systematically outlined. First, it was investigated which parameters should be included in the simulation model. It was found that topology, vegetation, weather and human intervention were key factors to include. This data could best be found in remote sensing data sets as remote sensing data provides the most up to date data and has the best coverage leading to a model that is applicable to most regions. Several remote sensing datasets were found that provided the topological, vegetation and weather data. The Next Day Wildfirespread dataset was chosen as it also provided a proxy for human intervention, its ease of use due to providing clear documentation and the possibility to benchmark the model against existing models. After the dataset was chosen and an understanding of the types of data in the dataset were acquired, different model architectures were investigated. This revealed spatial temporal conditional auto-regressive (STCAR) models as potential wildfire simulation models. The STCAR model assumes that a cell's value is dependent on both the value in the cells around it as well as its own value in the previous time step. It provides interpretability by explicitly incorporating a spatial and temporal dependence parameter. Additionally it is space agnostic allowing it to work with different data types and shapes. These advantages provide the adaptability to different data types and the interpretability that are a requirements. The disadvantage of the STCAR model is that it is not compatible with covariates and binary data. The model was extended to work with these by assuming the covariates could be

combined to determine the odds of a cell catching on fire, which then could be used directly within the STCAR model and compared against the binary fire mask to tune the coefficients. Initially the parameters are combined linearly, but this assumption may later be replaced with a non-linear combination.

Results show that the model thus far is not able to make reasonable predictions of the wildfire spread. The model tends to not predict any fire occurring and sometimes severely overpredicts the amount of fire. Additionally the predictions lack directionality and instead provide a smoothing effect. These are a result of limitations in both the dataset and the model. Further research and model development is required to bring the model into practice. Specifically the model may be extended to work with neural networks to capture non-linear effects. This would also account for the multicollinearity issues because of which the current model has unexpected coefficients. The model can also be improved by including higher resolution data via other datasets or even mixing data of several spatial resolutions. The Next Day Wildfirespread dataset uses MODIS satellite data at a resolution of 1 kilometer. This can be replaced with for example the WildfirespreadTS dataset that provides VIIRS satellite data at a spatial resolution of 375 meters and provides multi-day observations. This will allow the model to use insights from previous days and from forecasts to understand if a fire is growing or shrinking, which is currently not possible as the Next Day Wildfirespread dataset includes independent single day observations. Finally, the model currently is not able to capture the directionality of fire spread and instead has a smoothing effect. In order to better capture the directionality, the distance metric in the weighted adjacency matrix of the STCAR model may be replaced with coefficients based on the data.

The model has multiple limitations, but can be modified, is extensible and able to run in near real-time. The current implementation thus serves as a proof of concept and foundation for modifications. Once the model provides accurate predictions, the model could be included into a data platform and be fed real-time data to bring the model into practice. The model would then be usable for a variety of wildfire simulation purposes during active wildfires and as a result may allow for improving the resilience against wildfires and saving lives.

Contents

Preface i

Summary ii

Nomenclature vi

1 Introduction 1
1.1 Societal relevance . . . 1
1.2 The link to Complex System Engineering and Management . . . 2
1.3 Prior research, academic knowledge gap and main research question . 2
1.3.1 Modeling paradigms . . . 2
1.3.2 Physics-based models . . . 2
1.3.3 Cellular automata . . . 3
1.3.4 Graph theory based simulations . . . 3
1.3.5 Machine learning simulations . . . 5
1.3.6 Conclusion and main research question . . . 5
1.4 The Design Science Research Methodology . . . 6
1.5 Sub-questions . . . 6
1.6 Structure of the thesis . . . 8

2 Systemic overview of wildfire modeling use cases 10
2.1 Wildfire simulation applications . . . 10
2.1.1 During prevention and mitigation . . . 11
2.1.2 During preparedness . . . 12
2.1.3 During response . . . 12
2.1.4 During recovery . . . 13
2.2 Requirements specification . . . 13

3 Model development 16
3.1 The conceptualization . . . 16
3.1.1 Wildfire parameters . . . 16
3.1.2 Data availability and data quality . . . 17
3.1.3 Exploratory data analysis . . . 21
3.1.4 Architectures . . . 27
3.2 Formalization . . . 28
3.2.1 Spatial temporal conditional autoregressive (STCAR) models . . 28
3.2.2 STCAR for Wildfires . . . 29
3.2.3 Code implementation and optimization . . . 32
3.3 Verification and validation . . . 33
3.3.1 Unit tests and verification of code . . . 33
3.3.2 Validation . . . 35

4 Results **36**
4.1 Experimental setup 36
4.2 Hyperparameter tuning 36
4.3 Model performance 38
4.4 Comparison to other simulation tools 38
4.5 Applicability of the model to simulation use cases 40

5 Discussion and conclusion **43**

A Structure of the initial literature review process **54**

B Coefficient values **56**

Nomenclature

Abbreviations

Abbreviation	Definition
AUC PR	Area Under the Precision-Recall Curve
BCE	(Logistic) Binary Cross Entropy
ERC	Energy Release Component
GIS	Geographic information system
GPU	Graphics Processing Unit
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
STCAR	Spatiotemporal Conditional Autoregressive model
VIIRS	Visible Infrared Imaging Radiometer Suite

Symbols

Symbol	Definition
a	Coefficient of a feature/covariate
l	Likelihood
N	Neighborhood
r	Temporal dependence
S	Spatial domain
t	Time
W	Weighted adjacency matrix
X	Features/covariates
Y	Fire mask
Z	Latent variable representing the log odds
β	Relationship between locations
μ	Mean
ϕ	Distance metric
ρ	Spatial dependence
σ	Standard deviation

List of Tables

1.1 Overview of Wildfire Simulation Models 4

2.1 Requirements of wildfire simulations 15

3.1 Comparison between different wildfire datasets 20
3.2 Next Day Wildfire Spread Dataset features 21
3.3 Descriptive statistics of dataset features 22

4.1 Wildfire simulation requirements met by the STCAR model. Fullfilment of a requirement is indicated by a plus sign. Partial fullfilment or being extendable/modifiable to achieve a requirement is indicated by a plus and minus sign. A minus sign means not achieved. 42

A.1 Overview of keywords used to create the search string 55

B.1 Parameters and their coefficient values of the trained spatial-temporal conditional autoregressive model. The values of the parameters are multiplied with their respective coefficients and then summed to derive the log odds of a cell catching fire in the initial time step. 56

List of Figures

1.1 Research flow diagram 9

3.1 Three samples of the Next Day Wildfire Dataset depicting the parameters. 21
3.2 Histograms for each of the features in the Next Day Wildfirespread dataset 24
3.3 Distribution of the unknown values in the previous fire mask and fire mask features of the training dataset. 25
3.4 Histogram of the difference in fire amount between the previous fire mask and fire mask. 25
3.5 Correlation matrix of the features in the Next Day Wildfire spread dataset. 26
3.6 Diagram depicting the flow of data in estimating the parameters of the STCAR model. The covariates $\mathbf{X}$, the firemasks Y_t and Y_{t+1} as well as the hyperparameters W, ρ, σ serve as inputs. The values for the coefficients of the covariates are tuned to minimize the binary cross entropy (BCE) loss, which consists of both the BCE loss of the first as well as later timesteps. 32
3.7 Flow chart of the complete training and evaluation process of the modified STCAR model. 34

4.1 Predicted fire mask when using the log odds to scale the weighted adjacency matrix. The fire at t=0 and t=1 are indicated. No fire is predicted. The neural network experiment provided similar results. 37
4.2 Sample of the test dataset used in Figure 4.3. The data is normalized with purple representing 0 and yellow representing 1. 38
4.3 Comparison between the previous fire mask (t=0), the actual fire mask (t=1) and the probabilistic predicted fire mask (t=1). The prediction shows the spatial smoothing effect of the STCAR model. 39
4.4 Sample of the test dataset used in Figure 4.5. The data is normalized with purple representing 0 and yellow representing 1. The precipitation is very high resulting in the overprediction in Figure 4.5 39
4.5 Comparison between the previous fire mask (t=0), the actual fire mask (t=1) and the probabilistic predicted fire mask (t=1). The predicted fire mask shows how precipitation can cause severe overestimation of the amount of fire. 39
4.6 Data sample from the test set showing artifacts in most features. Data in the sample is normalized with purple representing 0 and yellow representing 1. 40

A.1 PRISMA Flow Diagram showcasing the application of selection criteria . 55

1
Introduction

1.1. Societal relevance

Wildfire simulations have become a hot topic as wildfires increase in frequency and severity, leading to increasing damages (Bao et al., 2024). Buechi et al. (2021) have found that in California the amount of acres burned and structure damages have increased steadily over the past four decades. For all ecoregions in the western United States the rate of increase is 7 fires and 355 km^2 per year (Dennison et al., 2014). This increase in frequency and extent of fires is linked to drier conditions and longer fire seasons as a result of climate change (Halofsky et al., 2020), invasive species and past fire management (Dennison et al., 2014). The increase of wildfires is not limited to the United States. Rodrigues et al. (2022) state that wildfires are commonplace in Mediterranean countries and also found that fuel dryness is a driver in the frequency and severity of wildfires.

The increase in forests fire frequency and intensity also leads to an increase in costs. The damages from these wildfires are not limited to only loss of forests and building damages. Wang et al. (2020) found that an estimated $148.5 billion was lost due to wildfires in California in 2018, consisting of value of capital, health costs and cascading effects in supply chain. Given these economical, ecological and health costs, resilience against wildfires is very important.

Simulations of wildfires serve as a vital tool in preparing against wildfires. Given these potential damages of wildfires, containment strategies cannot be tested in the real world. Wildfire simulations provide a way to test containment strategies, investigate fire risks and to find out the effects of potential wildfires. Wildfire simulations are not limited to just wildfire containment. Wildfire models can be extended or coupled with other models to serve as a tool for testing resilience for a wide range of applications including forestry management, evacuation strategies, air pollution modeling, supply chain and critical infrastructure simulations. Thus wildfire simulations not only enhance containment strategies, but also serve as tools for preparing communities and safeguarding critical infrastructure. As a result, improvements in the accuracy and speed of simulations can have a great impact.

1.2. The link to Complex System Engineering and Management

The research is closely related to the Complex System Engineering and Management (CoSEM) program as both CoSEM and this research focus on design within complex socio-technical systems. Wildfire resilience is a multi-disciplinary topic, as shown by the wide range of applications, in which the simulations serve as the tool to test interventions. This thesis will encompass the design and creation of such a simulation tool in order to provide predictions that can be used in decision-making and can be used to test interventions. This design process further reflects the link to the CoSEM program as it will include and touch upon content from most of the coursework. It especially reflects major concepts from SEN121 Agent based modeling, SEN163 Responsible Data Analysis and SEN162 I&C Service design. All in all, the proposed research will allow for demonstrating the skills and knowledge taught during the CoSEM program.

1.3. Prior research, academic knowledge gap and main research question

In order to create an understanding of the different types of wildfire simulations, their use cases and knowledge gaps, an initial structured literature review will be conducted. A description of the process of the initial literature review can be found in A. Table 1.1 categorizes each of simulation models discussed in the papers found during the literature search. Models are classified based on their modeling paradigm: empirical, semi-empirical or physics-based. This distinction is explained in the next paragraph. Additionally the models were classified based on the methodology used to create them. These are explained in the subsequent paragraphs. In this, the specific focus will be on the advantages and disadvantages between them.

1.3.1. Modeling paradigms

In the literature it was found that fire propagation simulations can be categorized as empirical, semi-empirical and physics-based. Empirical models are models that derive the relationships from observed data. On the other end of the spectrum, physics-based models solely rely on differential equations. Semi-empirical models lie in between, usually utilizing simplified equations. The models in Table 1.1 are categorized in the most suitable category and based on their methodology. While the methodologies are presented as separate categories, the borders between categories are not as definitive as some simulation models attempt to combine different methodologies. Each of the methodologies will be discussed in the following sections.

1.3.2. Physics-based models

The models that belong to physics-based models relying on computational fluid dynamics. These models may be coupled to atmospheric models and typically integrate Navier-Stokes equations for simulating wind and may include eddies (Pimont et al., 2009). While the advantage of this is that these models resemble reality the closest, they generally are computationally expensive and require more training data. This makes them less suitable for large scale wildfire simulations and makes real-time simulation near impossible (Cheng et al., 2022).

1.3.3. Cellular automata

Another methodology for fire spread simulations are cellular automata. Cellular automata typically are semi-empirical. These simulators divide space into a raster and have functions describing when a cell changes state based on the state of neighbouring cells. Each of these cells contains data such as fuel, weather, moisture content, and topographic attributes to inform the fire spread (Pais et al., 2021). Cells initially were squares, however, some more recent models involve hexagonal cells (Xu et al., 2022). Most of these models rely on Rothermel's rate of spread model (Rothermel, 1972), the most famous being FARSITE (Finney, 1998). The drawback of this model is the amount of input data required and needing to calibrate the input parameters depending on the environment. More recently attempts were made to include machine learning withing cellular automata for fine-tuning the input parameters using genetic algorithms and for transitioning between states. Models using genetic algorithms are suitable for predicting ongoing fires, but less so for hypothetical fires. The computational demand depends a lot on how complex and detailed the cellular automaton is. While a lot of cellular automata are able to run in near real-time, others are not able to achieve this speed.

1.3.4. Graph theory based simulations

One category of semi-empirical models are models that use graphs. Graphs divide space into nodes connected via edges. This allows for irregular networks. Models compute when a node catches fire based on its connected nodes via shortest path algorithms (Hajian et al., 2016; Stepanov & Smith, 2012). The advantage over cellular automata is that the unimportant parts of the topology can be removed, solving issues in choosing resolutions and speeding up computation times (Hajian et al., 2016; Jiang et al., 2022). The drawback, however, is that parameter tuning is more difficult for irregular networks.

Table 1.1: Overview of Wildfire Simulation Models

Modeling paradigm	Methodology	Model	Source
Empirical	Machine learning	6 Unnamed models	(Bottero et al., 2020; Cheng et al., 2022; de Gennaro et al., 2017; Khanmohammadi et al., 2022; Maeda et al., 2009; Murali Mohan et al., 2021; Wood, 2021)
Semi-empirical	Graph theory	3 Unnamed models	(Hajian et al., 2016; Jiang et al., 2022; Stepanov & Smith, 2012)
	Cellular automata	FARSITE, CELL2FIRE and 10 unnamed models	(Alexandridis et al., 2011; Castrillón et al., 2011; Chi et al., 2003; Finney, 1998; Gharakhanlou & Hooshangi, 2021; Jahdi et al., 2016; Karafyllidis & Thanailakis, 1997; Mitchell et al., 2023; Ntinas et al., 2017; Pais et al., 2021; Rui et al., 2018; Subramanian & Crowley, 2018; L. Sun et al., 2021; Vahidnia et al., 2013; Wu et al., 2022; Xu et al., 2022; Yassemi et al., 2008; Yongzhong et al., 2004; Zheng et al., 2017; Zhou et al., 2020)
Physics-based	Fluid dynamics	FIRETEC, WRF-FIRE, WRF-SFIRE and 2 unnamed models	(Lopes et al., 2019; Pimont et al., 2009; Vanella et al., 2021)

1.3.5. Machine learning simulations

Machine learning models are empirical models, which derive patterns from data. There are many machine learning techniques that can be used for wildfire modeling, however the most common ones are neural networks, long short term memory and support vector machines. They are able to make predictions of the firefront by learning patterns in the provided data. However, sufficient training data needs to be available. This is not always the case. To mitigate this shortcoming, some models are trained on data generated by other simulations (Cheng et al., 2022). The advantage of machine learning methods is its speed. One model trained on data from a cellular automata ran around 1000 times faster than the cellular automata model itself (Cheng et al., 2022). Due to this increase in speed, combining machine learning with other paradigms is also widely investigated.

Another advantage of machine learning is that the patterns found can be complex and non-linear allowing for more accurate models. However, as a result it may be unknown what individual variables are represented, decreasing the interpretability of the model and allowing for potential violations of the laws of physics. Physics-informed machine learning could potentially offer an improvement by confining the machine learning's output range using differential equations. Bottero et al. (2020) have made an initial model integrating physics-informed neural networks into the physics-based WRF-SFIRE model. It shows potential as it sped up computation time and is more interpretable. However, it is still in its early stages as they mention that their model only works on relatively small domains and is unstable for larger ones.

1.3.6. Conclusion and main research question

From the literature review it can be concluded that fire propagation simulations range from being completely based on physics, such as computational fluid dynamics models, to being based on collected data, such as certain machine learning models. While physics-based simulations often better resemble reality, they also require precise input data and are computationally expensive. This makes them unable to achieve real-time simulation. Statistical models on the other hand are much faster and require less precise input data, but lack a basis in physics. As a result these machine learning models are not explainable and might even violate the laws of physics. Semi-empirical models lie somewhere in between as they use equations derived from empirical studies, most commonly Rothermel's equations. The drawback to these models is the required input data and higher computational demand than machine learning models. This reveals a need for wildfire simulations that combine the speed of machine learning with the robustness of physics-based simulations. Physics-informed machine learning could potentially fulfill this role. Therefore, this thesis will attempt to fill this gap by answering the following research question:

How can a near real-time wildfire simulation be designed using physics-informed machine learning?

1.4. The Design Science Research Methodology

As the research gap concerns the design of a model, a design oriented approach will be taken. The goal is to develop model in a structured way so that it can be implemented and used by the various stakeholders. A design oriented approach means dividing the work into requirement elicitation, prototype, evaluation and the final design.

More specifically, this master thesis will follow the design science research methodology as outlined by Peffers et al. (2007). The design science research methodology divides research into six steps that match the design approach: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication.

The design science research methodology was taken over a modeling approach as a design approach has the advantage of being much more focused on creating value by meeting the requirements of stakeholders and ensuring the relevance of the created artifact, whereas a modeling approach focuses more on explaining an observed concept without considering how the model would later be used. This is especially relevant as wildfire models are used for a multitude of applications by a variety of stakeholders. By choosing to use the design science research methodology, the focus will shift towards the relevance of the model and its use cases.

However, parts of the modeling approach will still be used during the design and development and evaluation steps of the design science research methodology. A limitation of the design science research methodology is that it does not provide methods for creating the artifacts. By including elements from the modeling cycle, a more structured way of creating the model during the design and development step is performed. In order to do this, a conceptual model and formal model will be created. Similarly for the evaluation step, where the goal is to make sure the objectives of the design are reached, there are no prescribed methods for measuring success. Peffers et al. (2007) state that ”Depending on the nature of the problem venue and the artifact, evaluation could take many forms”. To structure this, the verification and validation of the model will be completed as per the modeling cycle ensuring that the model does what it is supposed to. This way, both the strengths of the design science research methodology and the modeling approach are used.

1.5. Sub-questions

In order to further structure the research, the main research question is decomposed into three sequential sub-questions. These will lead up to the main research question via the steps of the design science research methodology. It should be noted that while the sub-questions are presented in sequential order, the design science research methodology is built on the understanding that the design process is an iterative process. The results of most of these steps will be revised multiple times throughout the design process.

In order to structure the problem identification and definition of objectives, the following sub-question will be answered:

(1) What objectives can be derived from how models are currently being used?

This will provide a deep understanding of the current models being used, by whom they are being used and for which purposes. In addition, this will provide a notion of what properties of a model are desirable. In order to get this data, a literature review can be conducted. Models used in practice will be evaluated by researchers and papers will be published about them. This serves as an opportunity to find out on the basis of what criteria they are being assessed. This directly translates to the use cases and purposes. It is expected that certain objectives and requirements will be more important than others. This is why the list of objectives will be prioritized. A literature review is chosen over other research methods such as conducting interviews due to less dispersion of information and avoiding a lengthy process. To get a near complete overview of models via interviews would require every stakeholder and application area to be interviewed, whereas a single literature review will likely provide data on all of the most important use cases.

For the design and development step, it is important to know which alternatives there are for the structure of the model. This will provide information on the design space and will facilitate organizing the trade-offs between different architectures. The ultimate goal is to narrow these options down to a single proposed design. That it why this sub-question is formulated in a way to allow for both exploring the design space and narrowing it down to one design:

(2) What could the model's architecture look like?

To get an understanding of the different model architectures and their trade-offs, a case study can be conducted on the wildfire models found during the literature review conducted for the previous sub-question. This will provide a better understanding of the way these models work and their trade-offs. One of the important trade-offs will be the data requirements. That is why this chapter will also look into the availability of this data. Examples of data include digital terrain models, weather data and fire spread data. If data is not available, this will be seen as a design constraint. Using the trade-offs and design constraints found throughout the case study, a final model architecture will be chosen. The choice for this architecture is equivalent to creating a conceptual and formal model as per the modeling cycle's steps. Based on the conceptual and formal models, the machine learning model can then be developed and trained using the available training data.

For the evaluation step, it is important to assess whether the objectives of the design are met by the created design artifact, which is the proposed wildfire spread model. This means investigating whether the requirements for a specific use case are met by the model. The requirements for a wildfire simulation for different use cases will be determined by sub-question 1. In sub-question 2 the model was developed. Now sub-question 3 will be formulated as follows to find out how applicable this model is to the use cases by checking the requirements:

(3) To what extend is the proposed model applicable to wildfire simulation use cases?

Assessing the model will consist of checking requirements that are generic for all wildfire models and checking application specific requirements. Two examples of the generic requirements are whether the model is valid and whether the model is verified. Verification means that the model matches the behavior that was set out in the formalization. For this no new data is required. Validation entails checking if the modeling relationship between the real world and the model is correct. Validation of the model can be performed as a case study by investigating the predictive capabilities of the model compared to the real life fire spread. This will require a subset of the available data from sub-question 2. Wildfire simulation use cases will also have requirements specific to those applications. By checking whether the developed model matches these requirements, the use cases of the model can be explored. In general, this sub-question will help answer whether the stakeholders would be able to use the model effectively. This will serve as an input for the discussion, conclusion and future research directions.

1.6. Structure of the thesis

Following the methodology and sub-questions, the rest of this thesis will be structured as follows. First, Chapter 2 will provide an overview of wildfire simulation use cases and their requirements to answer the sub-question ”What objectives can be derived from how models are currently being used?”. After this, Chapter 3 will use these requirements to evaluate the design space of wildfire simulations. To do this, an overview will be created of the required parameters. Then, the availability of this data will be investigated and datasets containing these parameters will be discovered and compared in order to find one dataset that will be used. After finding the dataset, different model architectures will be investigated and one will be selected and developed in order to answer the sub-question: ”What could the model’s architecture look like?”. Afterwards, the model will be compared to other models and tested on the requirements and thus on whether it provides usable results in Chapter 4. This will answer the final sub-question: ”Does the proposed model satisfy the requirements?”. Finally, a reflection on the results will be given in the discussion and conclusion in Chapter 5. An overview of the structure of the thesis, the outputs of each chapter and how they relate to each other and the design science research methodology can be found in Figure 1.1.

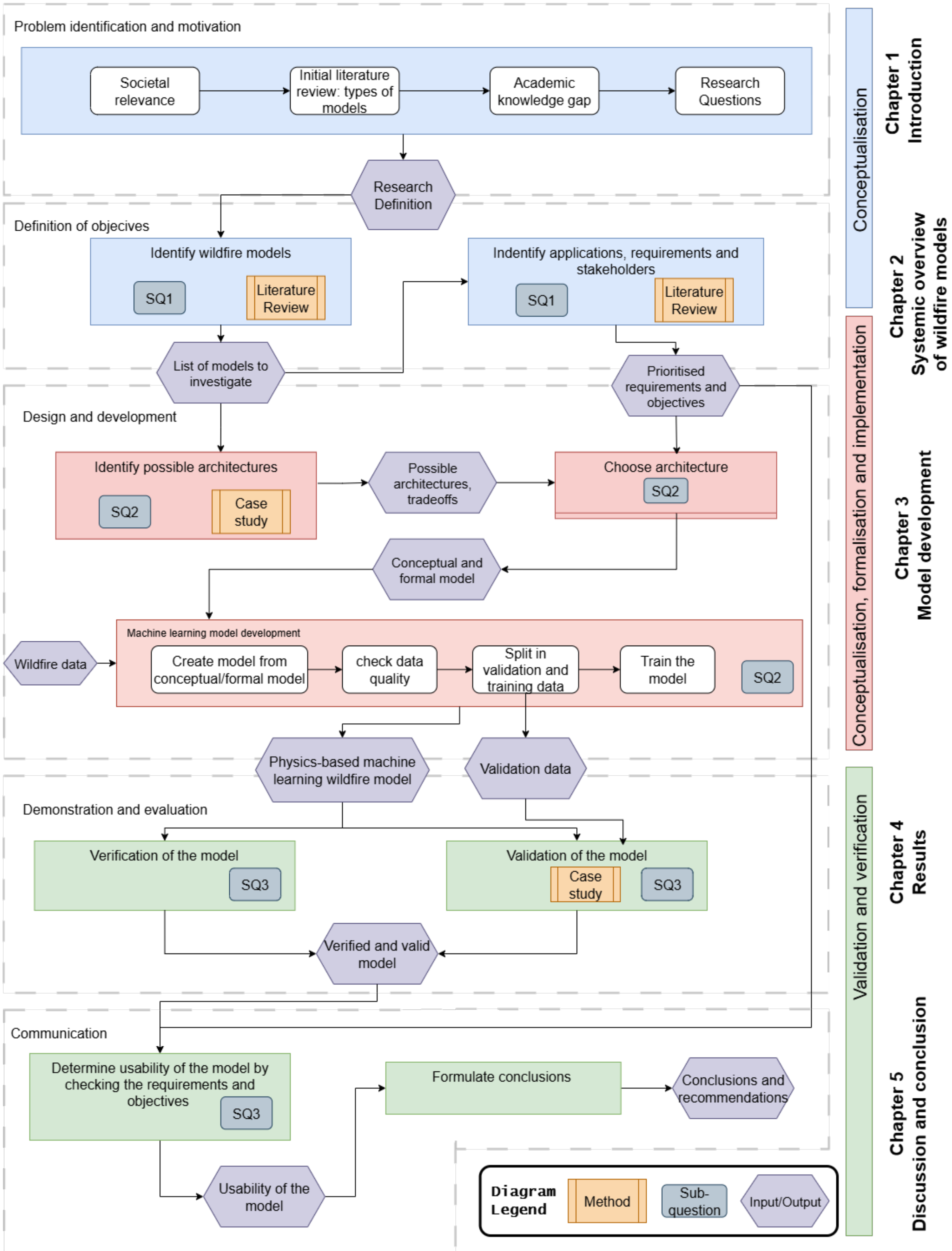

Figure 1.1: Research flow diagram

2

Systemic overview of wildfire modeling use cases

In this chapter, the first sub-question *"what objectives can be derived from how models are currently being used?"* will be answered to fulfill the 'definition of objectives' step in the Design Science Research methodology. This will create an understanding of how current wildfire simulations are used and how the system functions as a whole. In order to achieve this, a literature review will be conducted to find out which stakeholders use which wildfire models for which specific purpose. From this a list of requirements will be derived per application area. This is important as wildfire models are a means to an end, not an end in themselves. Thus, new models should strive to be useful for the application, the end, they are made for.

The search for usage of wildfire simulations will be structured using the overview of wildfire models provided by Singh et al. (2024) and Papadopoulos and Pavlidou (2011). A search will be conducted on the Google and Google Scholar search engines to find usage of the models mentioned in these papers. This data will be compiled into different application areas. The rest of this chapter will be structured as follows, first a summary of the identified application areas will be given. Then each of these application areas will be covered more in depth with a focus on their stakeholders and requirements. Finally, an overview of all the identified requirements will be given and the scope of the simulation will be determined.

2.1. Wildfire simulation applications

Wildfire simulations are applied across a broad spectrum of use cases, but are ultimately used to predict the spread of fire. The spread of fire is used for different kinds of risk management surrounding the wildfires. Wildfire risk management incorporates the four phases of emergency management: prevention and mitigation, preparedness, response and recovery (Tymstra et al., 2020). As such the applications of wildfire simulations can generally be matched to these phases. The prevention and mitigation phase is all about limiting the amount of potential damage and consists of fuel management, home preparedness and land management use cases. The preparedness phase entails planning for if a wildfire happens and contains evacuation simulations and danger rating applications. The response phase encompasses all actions once a fire does occur and consists of decision support during active wildfires. Finally, the recovery phase includes all actions after a wildfire has occurred and consists of disaster response and ecology applications. Each of the wildfire risk management phases and how simulations are applied in them will be covered in more depth below.

2.1.1. During prevention and mitigation

The prevention and mitigation phase focuses on reducing the negative effects of potential wildfires. This is done through fuel and land management and home preparedness. Fuel management means removing potential fuel sources from the forest. This could be achieved by thinning trees and shrubbery via mechanical treatment, by utilizing prescribed burns as well as by converting the fuel type to fire resistant vegetation (Moreira et al., 2011). In a prescribed burn a small portion of the forest will be set on fire to burn away the shrubbery when conditions are controllable. This decreases the amount of fuel that is available if the forest were to catch fire later on, effectively decreasing the size, intensity and damage of future wildfires when conditions may not be favorable. The effectiveness of prescribed burns and fire breaks is evaluated through experiments, fieldwork and simulations, such as those using FIRETEC and FARSITE (Moreira et al., 2011). Due to the risky nature of experiments and the limited data that can be derived from fieldwork, simulations are preferred over experiments and fieldwork (Moreira et al., 2011). In addition to evaluating the effectiveness, simulations are also used for planning the prescribed burns. The amount of land that can be subjected to fuel treatment is limited by operational, ecological and social pressures and thus it is important to carefully select treatment locations (Fernandes & Botelho, 2003). There are several tools that help with this selection process. In the United States the Interagency Fuel Treatment Decision Support System is used for fuel treatment planning by using the FlamMap, minimum travel time and other simulation models (Drury et al., 2016). In Canada the Prometheus simulator may be used for fuel treatment planning.

Additionally burn probability maps are created to aid in the decision making using FSIM. FSIM is a stochastic simulation tool that predicts the fire spread of many different ignition points over a given area at a resolution of 270 meters (Finney et al., 2011). The fire scars left by the fire spreading are determined for various weather scenarios. Afterwards, these are aggregated to find the likelihood of a given place catching on fire if a wildfire occurs in the area.

These model results may shape policies and educational programs for home preparedness. For example, the American government advises to create a fire resistant zone, making an emergency plan, and using fire-resistant materials (Federal Emergency Management Agency, 2025). It has been shown that a homeowners' perceived risk and thus their willingness to act is influenced by this information from official sources, as well as information from unofficial sources (Brenkert-Smith et al., 2013). Additionally these results may be used by insurance providers.

2.1.2. During preparedness

The preparedness phase includes creating plans for when a wildfire occurs. Wildfire simulations are used for this especially in planning evacuations. Evacuation simulations are used to understand human behavior during evacuations and to determine the optimal thresholds for when to notify and evacuate people as a wildfire approaches. Disaster response and evacuations planning has to account for significant logistical and operational challenges. Simulation studies have been conducted that accounted for road conditions (Ma & Lee, 2025) and failing communication networks (Grajdura et al., 2022). The FARSITE model was coupled with evacuation simulations to find evacuation triggers (Mitchell et al., 2023). Mitchell et al. (2023) used the rate of spread output in raster format together with Dijkstra's shortest path algorithm to find the minimum travel time to communities.

Typically these evacuation cues are linked to fire danger ratings. In America the Fire Weather Index is used as a fire danger rating. In Canada the Canadian Fire Weather Index is used. In Australia McArthur's MK 5 forest fire danger meter is used. And in Europe the Copernicus Fire Danger Firecast system may be used. These systems provide a numerical output based on the likelihood of a fire occurring and the expected rate of spread once a fire does occur. The rating these systems provide are also used to determine the manpower and resources that need to be ready at a given time and thus determine the readiness level of firefighters (Alexander, 2000).

In addition to facilitating the planning of readiness levels, wildfire simulations may be used to train first responders themselves. Emergency services may simulate a fire and determine the best course of action. This is not limited to firefighters, but is also for making plans for medical services (Hertelendy et al., 2024).

2.1.3. During response

The response phase includes all actions when an emergency occurs. For wildfires this includes the operational planning during an active wildfire. Simulations are used as decision support tools. Specifically, the wildfire simulations are used to determine if and how to attack the fire through fire spread modeling. Fire spread modeling entails figuring out where the fire will go and at what time it will get there given a scenario. Typically the focus will be on a fire perimeter that is advancing, rather than showing the entire fire scar. This can then be used by forest fire managers to determine the appropriate firefighting resource allocation. More specifically, in the United States as a fire increases in size, more resources and agencies are included in fighting the fire and the National Interagency Fire Center (NIFC) will coordinate the resource allocation across these agencies, relying on simulation-based forecasts to do so. As such it is important for a simulation to accurately provide the location of a fire and its arrival time to different points. There have been efforts to facilitate this interagency collaboration and make data consistent between agencies such as the LANDFIRE platform in the United States (Rollins, 2009).

An example of a simulator used in planning the initial attack is the California Fire Economics Simulator (CFES2). It includes the ability to model extreme fire spread rates and multiple fire starts. Models that are used in planning the initial attack typically model the fire rate of spread and the fire dispatch level and include various firefighting strategies to answer 'what if' questions (Gilless & Fried, 1999). In Australia the PHOENIX, managed by the Fire Predictive Services, is used to support investment decisions, positioning of fire resources during a fire and evacuation decision making (FLARE Wildfire Research, 2021).

2.1.4. During recovery

The recovery phase includes all actions taken after a wildfire has occurred in order to recover from the damages. In this phase wildfire simulations are used for various purposes. One application is forensics. Wildfire simulations such as Prometheus may be used for forensic purposes by running the simulation in reverse to find the origin of the fire (Tymstra et al., 2010). Typically physics-based simulations are used for this purpose due to their high accuracy and computational demand being less important.

Additionally wildfire simulations can be coupled with smoke simulations to provide an understand of where contaminants may have traveled to and the resultant impacts on the air quality. This typically uses coupled fire and atmosphere models. These typically are physics-based simulations such as FIRETEC and WRF-SFIRE. Both of those models used fire grid spacings of centimeters up to meters and could only model domains up to roughly a kilometer (Liu et al., 2019). Due to this these models are typically used for retrospective fire cases at relatively limited scales (Hung et al., 2025). Alternative models that were made for more operational use cases include Daysmoke and PB-P and CMAQ-Bluesky which respectively have a grid spacing of about 100 meters and 4 to 12 kilometers and domains of respectively 5 km and 1000s of kilometers (Liu et al., 2019). Other models that are used operationally for this purpose include Prometheus, Simple Smoke Screening tool and VSMOKE-GIS as these execute fast enough and the data they require is readily available (Liu et al., 2019). In order for a wildfire model to be connected to a smoke and emission simulation, the model requires fuel condition, fuel consumption, fire spreading and atmospheric condition properties as described by Liu et al. (2019). Fuel condition properties include the amount of fuel, type, distribution and moisture contents. The fuel consumption includes the amount consumed and stage of the flame. Information about the fire spread includes the fireline location, shape and evolution over time. The atmospheric conditions include the temperature, winds, moisture, pressure and precipitation. Air quality forecasting models often use the fire intensity and fire size, which are derived from these properties, to determine the smoke emissions (Hung et al., 2025)

Additionally, wildfire simulations may be used for ecological monitoring. FlamMap has been used in hydrology studies as a wildfire changes the runoff and erosion properties of soil (Srivastava et al., 2018).

2.2. Requirements specification

In order to make a new model useful and implementable, it should offer an improvement over the models that are currently being used. As such, this section provides an overview of the requirements that would need to be met in order for the model to see adoption within the application domains. All of these requirements are listed in Table 2.1. The reasoning behind these requirements is linked to the use cases found in the previous section, but will be briefly iterated and explained in the following paragraphs. Afterwards the requirements will be prioritized using the MoSCoW prioritization method. This method breaks the requirements down into must-have, should-have, could-have and won't have. This will show the relative importance and value of each of the requirements and provide an understanding of which requirements will be integrated into the model and which will not.

In general, the most important performance indicators will be the computational efficiency, accuracy and interpretability of the models. As was found during the initial literature review, empirical simulations run quickly, but may not offer interpretability and physical rigor, whereas physics-based simulations do provide this, but require a lot of data and computational resources. As such the interpretability and ability to run many simulations quickly are major requirements in table 2.1. As a result, the accuracy, the computational efficiency and interpretability will be key performance indicators when discussing the model's applicability to wildfire simulations. There are several accuracy measures that could be used to compare the performance of different models, however the most common way to compare the results of different wildfire spread simulations is by comparing the fire scars. This shows the estimated burned area. The amount of burned area and shape of the fire scar should match reality. As such, while some applications do not outright require a fire scar, such as smoke simulations, they do require accurate information on where a fire is at a given time. The fire scar provides a way to measure this and is thus considered an important output of the model. For computational efficiency the inference time is the key performance indicator. Inference time is the time that the model takes to run a given scenario. It is more important than the training time, which is the time it takes for the model to learn its parameters from a given dataset, as the training can take place before a model is required for a given application.

For simulations that are used for risk assessments the ability to perform many simulations quickly is vital. As such the computational complexity should be low. Additionally the model must be able to run in parallel. To then provide an output of the risk of a fire reaching a certain point, the area burned of each simulation will be aggregated. As such the burned area is a necessary output for the model for risk assessments.

For simulations during active wildfires it is most important that the models are able to assist in planning how to attack the fire. This means it needs to be able to ingest data in near real-time. For operational tools this means that the data the models require needs to be readily available. Additionally the simulation models needs to be able to be executed quickly. While no specific requirements were listed for the spatial and temporal scales, most models that are currently used for operational firefighting provide outputs at an hourly time scale and a spatial resolution of between 30 and 270 meters.

For other use cases the simulations provide outputs that are compatible with Geographic Information Systems. This allows the wildfire spreading outputs to be connected to for example evacuation simulations, smoke simulations, hydrology studies and other applications. As such providing outputs in a georeferenced raster or vector format is essential for making these applications possible. This allows it to be used as a Geo Information System layer and coupled to other types of simulations and data. For smoke simulations, hydrology studies, ecology studies and similar applications the computational efficiency on the model is much less important than the accuracy of the model as the results are not needed in near real-time.

In order to prioritize the requirements the MoSCoW prioritization method will be used. The requirements that are essential to wildfire spread modeling will be categorized as must. These include requirements 1, 1.1, 1.2 in Table 2.1. These requirements specify the basic behavior of wildfire simulations by modeling the rate of spread and location of a wildfire. The information processing capabilities are also considered a must as all operational use cases rely on the model having the ability to ingest data in near real-time.

Table 2.1: Requirements of wildfire simulations

Number	Requirement	Prioritization
Goal	To create an interpretable wildfire simulation to support fire-fighters	
1	Provide information on wildfire spread	Must
1.1	Provide an accurate fire scar	Must
1.2	Show when the fire arrives at different locations	Must
2	Provide interpretable outputs explaining fire behavior drivers	Should
3	Test the effectiveness of interventions	Should
3.1	Have an adjustable fuel load	Could
3.2	Show risk profiles	Could
3.2.1	Aggregate information from several simulations	Could
3.2.2	Be able to run in parallel	Should
4	Process information	Must
4.1	Be able to integrate near real-time data	Must
4.2	Validate information	Should
4.3	Handle missing or uncertain input data	Must
5	Be extendable to different use cases	Should
5.1	Be usable as a GIS layer	Should
5.2	Adaptable to different data types	Should

3
Model development

This chapter will answer the sub-question 'What could the model's architecture look like?". This chapter will be split into two parts. First, the model's design space and design constraints will discussed. These will be used to create a conceptual model in Chapter 3.1. After this, the chosen model will be formalized and denoted in mathematical equations in Chapter 3.2.

3.1. The conceptualization

The conceptualization will go through the following process: first it will be denoted which types of parameters should be included. Then, potential data sources will be investigated. The availability of these datasets will act as a constraint on the design space. Once the data sources are found and a single dataset has been chosen, different types of models will be investigated with the goal of identifying the one that matches the requirements the best.

3.1.1. Wildfire parameters

First it is important to consider the parameters that are included in the model. There are several factors that influence where a given wildfire will spread to. According to Alexander and Cruz (2013) "the difficulty in predicting wildland fire behaviour boils down to the fact that there are numerous, interacting variables involved". These are captured in several conceptual models readily available in literature.

In Rothermel's model (Rothermel, 1972), the rate of spread is dependent on the reaction intensity, the bulk density, moisture dampening coefficients, propagating flux ratio, wind, slope and heat of preignition. The model was designed to work with fuel, moisture and terrain data (Andrews, 2018). According to Cawson et al. (2020), who researched conceptual models of fire spread in eucalyptus forests by eliciting fire experts, there is a broad agreement that drought, dead fine fuel moisture, weather and topology were main drivers of flammability in these forests. Additionally, they state fire management may limit the spread of wildfires.

The conceptual models found in literature can be summarized by four key factors:

topology, vegetation, weather and human intervention. Each of these factors may be further decomposed into several sub-factors that influence the fire spreading behavior.

The topology, or terrain features, can be decomposed into the aspect, slope and elevation. It is widely understood that fire travels quicker uphill than downhill due to the tops of the flames heating the fuel load that is uphill. Topology data is typically included by using a digital terrain model that was created using satellites.

Vegetation plays a major role in the spread of fire as it provides biomass as fuel for the fire to spread. The amount of fuel that is available for any given fire in an area, also called the fuel load, is largely determined by the type of vegetation, its moisture contents and its distribution over the terrain. Different types of vegetation exhibit different behaviors. Popular simulation models account for this by including several fuel models to distinguish between these behaviors. Some types of vegetation may be prone to spotting, which is embers flying from the fire that may start a new fire elsewhere. Additionally a fire may be a surface or brush fire, which is low to the ground, or a crown or canopy fire in which entire trees catch fire. The FARSITE model accounts for this by including layers of the canopy and crown heights (Finney, 1998).

The weather and climate also determine how a fire spreads by affecting the fuels. If the fuel is warmer or contains less moisture due to droughts, less heat is required for it to catch fire. Thus the temperature is an important factor to consider. Additionally, precipitation in the previous days will also add to the moisture content of the fuels and increase the humidity making it harder for it to catch fire. Lastly the wind may influence the spread of the fire through several ways. J. Sun et al. (2023) state that the wind may accelerate the spread of wildfires by drying fuel, by carrying embers resulting in spot fires and by providing oxygen for the fire to burn.

Additionally human intervention is an important factor to account for. The suppression efforts of the firefighters in the locations they are deployed to will hinder the spread of fire. Historically, the focus of wildfire firefighting has been on extinguishing fires as quickly as possible regardless of the values at risk (Smith, 2017). These strategies lead to increased fuel loading as fires were not able to burn away fuels and thus more uncontrollable fires. Current strategies consider which values are at risk. These include manmade structures and critical infrastructure at the wildlife-urban interface. This makes it more likely for firefighters to be deployed near these locations.

While these parameters represent the fire behavior in wildfires at a large scale, there are several factors that influence the fire behavior at smaller spatial and temporal scales, such as precise fluid dynamics. While it is unrealistic to model these due to the computational demand involved, this does leave a level of uncertainty and behavior that will not be able to be captured by current models. However, it is believed that the topology, weather, vegetation and human intervention factors will capture the fire behavior well enough to be useful in the applications mentioned in the previous chapter.

3.1.2. Data availability and data quality

Specific data is required to represent the parameters in the conceptual model. There are several data sources available that are used in wildfire simulations which can be classified into three categories: remote sensing data, land survey data and simulated data.

Remote sensing data is data gathered by satellites. This has the advantage of providing up to near real time data. However, this data may be noisy due to for example cloud cover or solar flares. Additionally, remote sensing data has the advantage of having worldwide coverage, whereas other data types may only be collected in certain locations. Thus, using remote sensing data leads to more generally applicable models.

Land survey data on the other hand is measured by people going out into the field. Historically this would be people in watchtowers, but more recently drones have been used to gather this data. Gathering the data is very labor intensive and making it relatively expensive to acquire. As such it is only gathered infrequently. As it is based off of visual assessment at a distance, it may be less accurate for wildfire perimeters. It does, however, usually have a better spatial resolution, but only at a given location. It is not available everywhere. As Chen et al. (2024) puts it ”While these methods provided valuable in situ data in specific locations, they were inherently limited in scope, unable to provide the reliable and continuous monitoring required at a continental or global scale.”

The advantage of simulated data is that it is able to provide extra training data at different spatial resolutions, however this data may contain errors that would compound into the newer simulations trained on the data. Drury et al. (2016) mention errors compounding as an important limitation using chained models within the Interagency Fuels Treatment Decision Support System and infer that the biggest uncertainty of the modeling framework occurs in the fuel model data input step.

As accuracy and near real time data is important to meet the requirements, remote sensing data was chosen as a main focus throughout the search for datasets. The search for datasets was conducted on the Google search engine and Google Scholar with the keywords ”remote sensing”, ”dataset” and ”Fire spread”. Only datasets that contain vegetation, fire, weather and topology data were considered in order to match the required parameter types found in Chapter 3.1.1. Four remote sensing wildfire datasets were discovered, each created with data from different sources and thus each with their own characteristics. These will be listed in Table 3.1. In the following sections these characteristics and their advantages and disadvantages will be explained in more detail, after which a final dataset is chosen.

Remote sensing wildfire data comes from two sources: the Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) satellite sensors. MODIS provides wildfire data at a resolution of 1 square kilometer, whereas VIIRS is more detailed at 375 meters. Similar differences occur in the vegetation data where the WildfireDB uses a LANDFIRE dataset that has vegetation at a resolution of 30 meters. This is very detailed, but comes at the cost that the data is only updated every two years as it is derived from field studies. As was mentioned in chapter 3.1.1, outdated data can lead to significant under- and over-predictions of the fire spread. While less detailed, the always near up to date nature of the remote sensing data of the Next Day Wildfire Spread dataset and the WildfirespreadTS dataset make them more suitable for near real time prediction.

In addition to the datasets listed in table 3.1, there are also is the possibility to create a new dataset using some of the previously mentioned data sources. The source code that is available for the Next Day Wildfire and WildfireTS datasets could be a start for this. The advantage of this would be that the dataset could be completely customized taking advantage of the upsides of the specific data sources. The downside, however, is a lot of extra work in preprocessing. Given the datasets provide appropriate data given the parameters found in chapter 3.1.1, creating a custom dataset is unlikely to be worthwhile.

Given the advantages and disadvantages of each dataset, it was chosen to use the Next Day Wildfire Dataset for creating the model as it offers data on all categories of parameters identified in chapter 3.1.1 and it is the easiest to use. In future iterations of the model it can be applied to the WildfireTS dataset to facilitate a better spatial resolution and to incorporate multi-day data.

Table 3.1: Comparison between different wildfire datasets

	Next day wildfire spread	WildfirespreadTS	Fire-Image-DenseNet dataset	WildfireDB
Source	(Huot et al., 2022)	(Gerard et al., 2023)	(Pang et al., 2025)	(Singla et al., 2020)
Resolution (space)	1 km (64x64)	375m	1km	375m
Resolution (time)	1 day	1 day	1 day	1 day
Vegetation data	NVDI, ERC	VIIRS vegetation index,	ORNL DACC biomass (above, below ground), PROBA-V densities (grass, tree, bare, snow, water)	LANDFIRE (vegetation, fuel type) at 30m resolution, updated every 2 years
Weather data	Wind, Temperature, precipitation, humidity, drought index	Wind, Temperature, precipitation, humidity, drought index and forecasts (precipitation, wind, Temperature, humidity)	ERA5 Wind, precipitation	Meteostat (part of NOAA)
Fire data	MODIS	VIIRS in GlobFire (MODIS) (607 fires, USA, 2018-2021)	MODIS (2012-2019, 304 events, western USA)	VIIRS
Other data	Elevation, population density,	Elevation, slope, aspect	slope	Data on neighbouring polygons, elevation, slope
Filetype	TF records	GeoTIFF, recommended conversion to HDF5	npy (numpy array in binary format)	CSV (of polygons)
Advantages	Extensive documentation, ease of use, easy benchmarking	Multi-day data, extensive documentation	More detailed vegetation	very detailed vegetation
Disadvantages	relatively low resolution, single observations, Specific to the United States	Harder to use, Specific to the United States	Less detailed weather. Minimal documentation	Specific to United States (LANDFIRE data), harder to process polygons

3.1.3. Exploratory data analysis

In order to better understand the Next Day Wildfire Spread dataset, a exploratory data analysis was conducted. The dataset contains 18,545 samples collected between 2012 and 2020, with each sample representing a 64×64 kilometer area divided into cells of one kilometer each. Each cell has data on 12 environmental covariates as shown in Table 3.2. The Fire Mask, which is the last feature shown in Table 3.2, is what has to be predicted by the model. In order to easily discern between the fire mask and the previous fire mask, the fire mask will be refered to as the actual fire mask. The model will be evaluated by comparing its predictions to this fire mask. Three samples of the dataset are shown in Figure 3.1.

Table 3.2: Next Day Wildfire Spread Dataset features

Feature name	Description	Unit/Range
Elevation	Average height of cell above sea level	m
Wind direction	Azimuth clockwise from north	degrees
Wind velocity	Average wind speed of that day	m/s
Mini temp	Minimum daily temperature	K
Max temp	Maximum daily temperature	K
Humidity	Concentration of water vapor in the air	kg/kg
Precip	Average precipitation per m²	mm
Drought	10-day Palmer Drought Severity Index	[-10, 10]
Vegetation	NDVI (Normalized Difference Vegetation Index). Ratio between red and near infra-red light reflectance as a measure of photosynthesis	[-1, 1]
Population density	Amount of people living in that cell	people/km²
Energy Release Component	releasable energy according to fuel model G (conifer forest)	unitless
Previous Fire mask	Previous day fire presence (1), no fire (0) or unknown (-1)	0, 1, -1
Fire Mask	Fire presence in the current day.	0, 1, -1

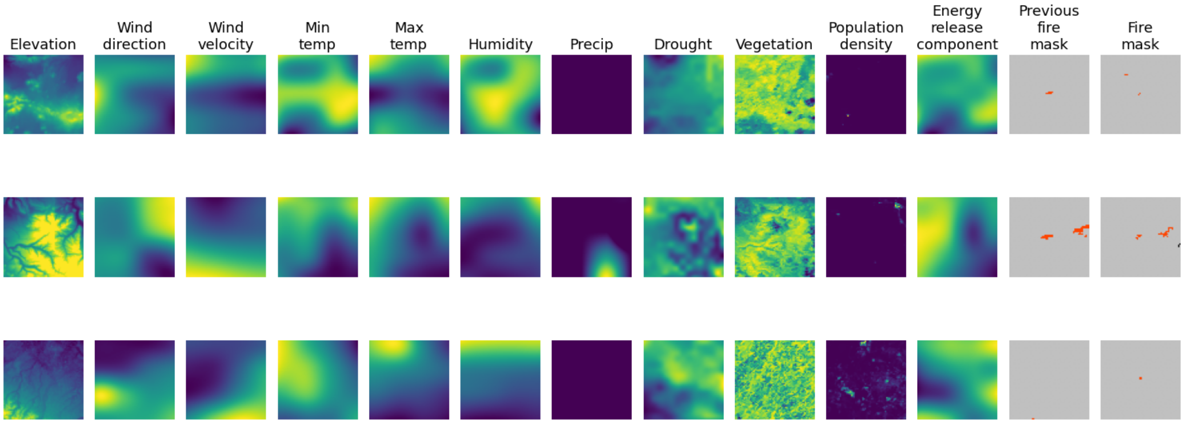

Figure 3.1: Three samples of the Next Day Wildfire Dataset depicting the parameters.

The data was split in a ratio of roughly 8:1:1 between training, validation and testing data by the publishers of the dataset by dividing all the weeks in the dataset randomly among these three categories (Huot et al., 2022). This resulted in a training set of 14979 samples, a validation dataset of 1877 samples and a testing dataset of 1689 samples. The exploratory data analysis is performed on the training dataset in order to not gain knowledge of the patterns of the validation and test dataset. This way tuning hyperparameters on can be performed using the validation set and evaluating the model on an evaluation set. This guarantees the model generalizes to unseen data if it passes the evaluation. For each of the features the descriptive statistics were gathered of the training dataset. As can be seen in Table 3.3, most features have extreme minimum and maximum values. Some of these values are unreasonable and cannot be possible. Including these values in the model would lead to exploding gradients when using machine learning. Fortunately Huot et al. (2022) provide a data reader with tools to clamp these features to realistic values. All features will be clamped using the provided functions to minimize the effect of measurement errors. Samples including these unrealistic values will not be deleted in order to avoid sampling bias. If the model is used in practice, data of similar quality is expected and thus the preprocessing steps need to match.

Table 3.3: Descriptive statistics of dataset features

Feature	Mean	Std. dev.	Minimum	25%	50%	75%	Maximum
Elevation	896.572	842.610	-45.000	119.000	611.000	1595.000	4193.000
Wind direction	146.647	3435.084	-505870.062	149.760	208.086	254.313	37735.629
Wind velocity	3.628	1.309	-82.653	2.716	3.427	4.332	103.220
Min temp	281.851	18.497	-444.693	277.875	283.079	287.881	716.628
Max temp	297.716	19.458	0.000	293.976	299.818	304.222	1229.849
Humidity	0.007	0.004	-0.129	0.004	0.006	0.008	0.086
Precipitation	0.323	1.534	-167.448	0.000	0.000	0.000	56.215
Drought	-0.773	2.441	-125.711	-2.609	-1.355	1.189	52.269
Vegetation	5350	2185	-9567	3735	5520	7123	9966
Population density	30.460	214.200	0.000	0.000	0.166	3.538	27103.605
Energy release component	53.469	25.098	-1196.089	30.999	49.440	75.067	2470.882
Previous fire mask	-0.003	0.138	-1.000	0.000	0.000	0.000	1.000

To get a better understanding of the features after this transformation, histograms were created as shown in Figure 3.2. Clamping the features has left some artifacts in the data as can be seen by the spikes near the minima, maxima and the origin of the features. These artifacts should not lead to significant performance issues of the model as their quantity is relatively low. The distribution of wind direction is centered around 240 degrees clockwise from north corresponding with a south-west wind. This corresponds with winds originating from the north pacific ocean. This reveals a bias in the data for fires in the western side of the United States and may limit the model's applicability to other regions. This bias is a result of most fires occurring in this region and thus being included in the dataset. The effects of the western side of the United States being most prevalent in the dataset will also be present in the other features. The distribution of drought and energy release component are bimodal suggesting either two distinct climatic regimes. Vegetation as measured by the NDVI ranges from -10000 to 10000. This is because the data is stored as a scaled integer. -10000 thus represents -1 and 10000 represents 1. It includes mostly positive values suggesting green vegetation with few negative values indicating water bodies. The imbalance in the distribution could make the model weaker at identifying areas that cannot burn due to fuel shortages. The precipitation and population density also exhibit a tailed distribution. They contain many values around the origin with relatively few higher values. This may lead to lower performance when there is a lot of rain or when there is a high population density as there are few samples including these cases.

Additionally special attention should be given to the previous fire mask distribution and its relationship to the distribution of the fire mask that is to be predicted. The previous fire mask consists of 0.8% cells on fire, 98.1% cells not on fire and 1.1% unknown values. These unknown values are a result of for example cloud cover obscuring the satellite's measurements. The fire mask distribution shows similar patterns with 1.1% cells on fire, 96.6% not on fire and 2.3% unknown values. In 44% of all samples an unknown value occurs in either the previous fire mask or the fire mask. The distribution of these values is right tailed with many samples containing almost no unknown values and relatively few samples including many unknown values as can be seen in Figure 3.3.

Due to the imbalance in both the previous fire mask and the fire mask, models could output that there is no fire reach up to 98% accuracy. the area under the precision recall curve (AUC-PR) is a better metric to evaluate the performance of binary classification for imbalanced datasets as it balances the precision and recall, which are respectively the true positive rate and the proportion of positives that are actually positive.

In addition to the amount of cells on fire in the entire training dataset, the amount of cells on fire within independent samples should be inspected as well. 54.5% of fires increase in size, 38.7% shrink in size while 6.8% remains the same size. The amount the size of the fire changes in each sample is shown in the histogram in Figure 3.4. There are relatively few samples where the amount of fire changes a lot. This may make it so the model is not able to learn patterns of extreme wildfire events.

In addition to clamping the values of the features to reasonable extents, the values will be min-max scaled to decrease the differences between different features, while keeping the distribution within the features themselves the same. This will lead to better performance for machine learning models. Finally the correlations of the features were inspected. As shown in Figure 3.5, the minimum and maximum temperature are highly correlated ($r \approx 0.99$). The elevation shows a negative correlation with humidity ($r \approx -0.43$) and a positive correlation with the energy release component ($r \approx 0.58$), consistent with the expectation that higher elevations are drier and more fire-prone.

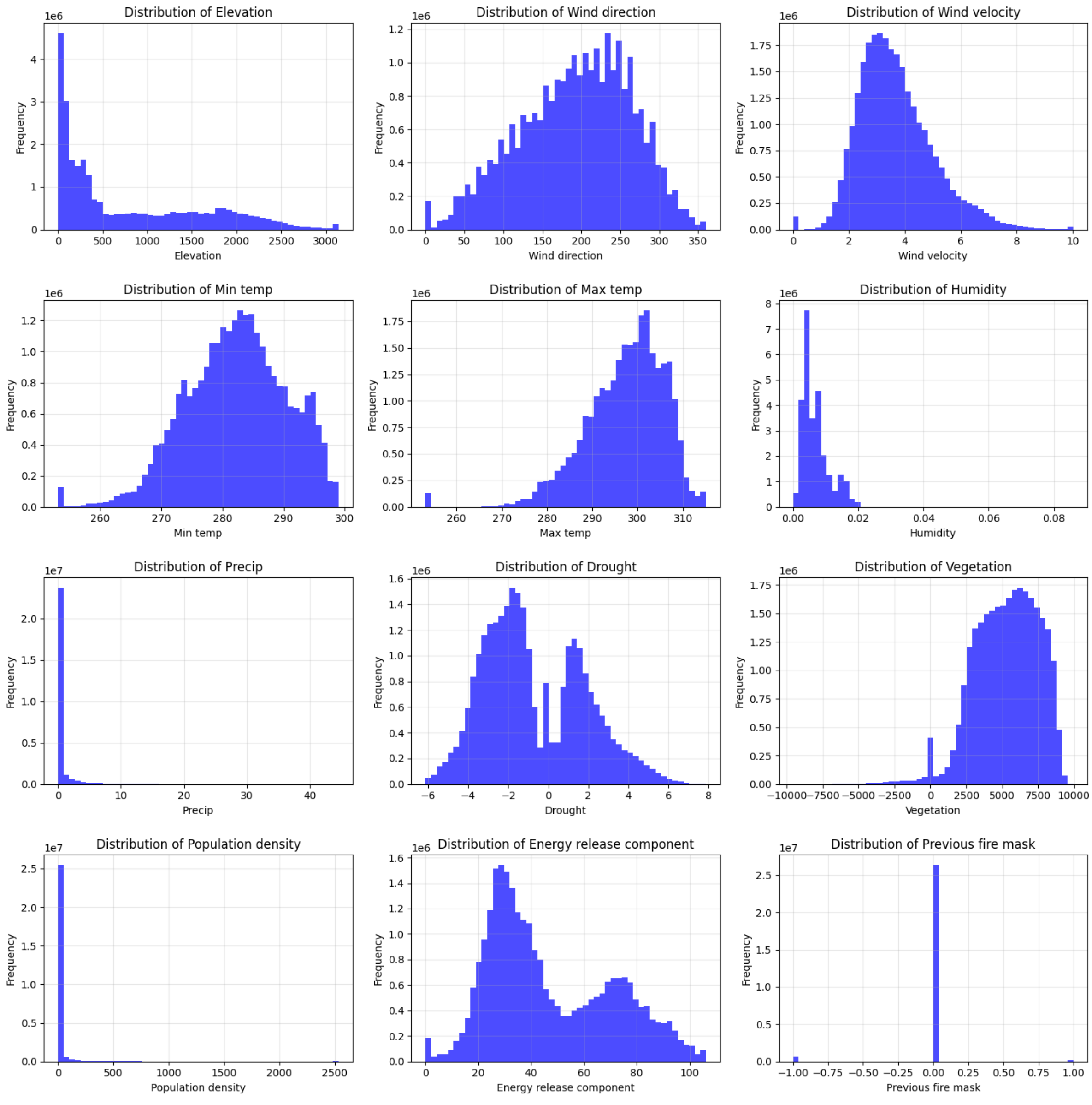

Figure 3.2: Histograms for each of the features in the Next Day Wildfirespread dataset

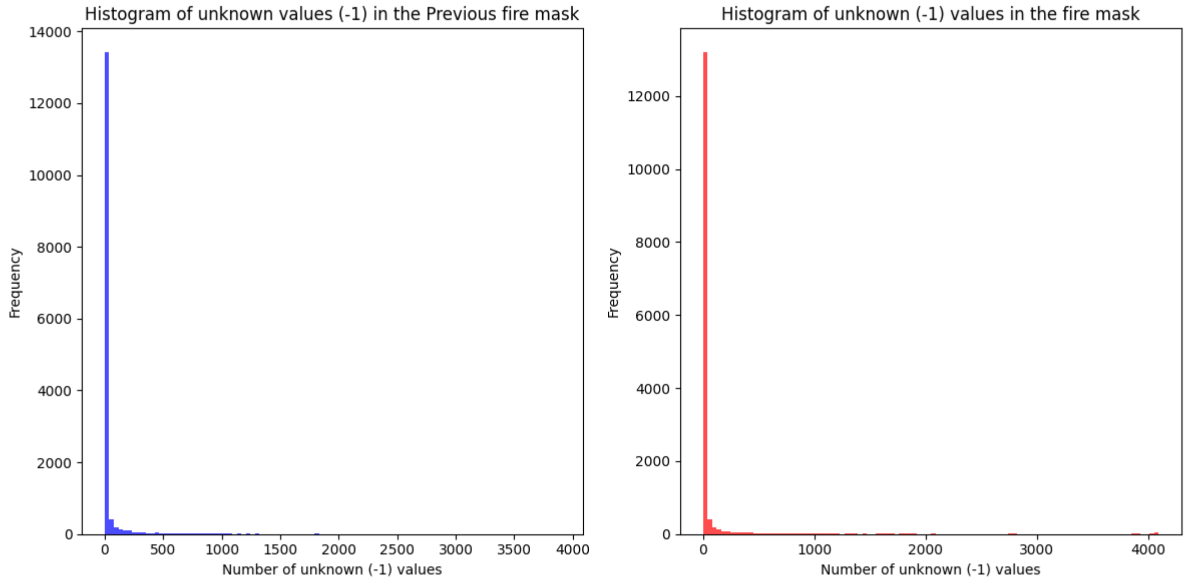

Figure 3.3: Distribution of the unknown values in the previous fire mask and fire mask features of the training dataset.

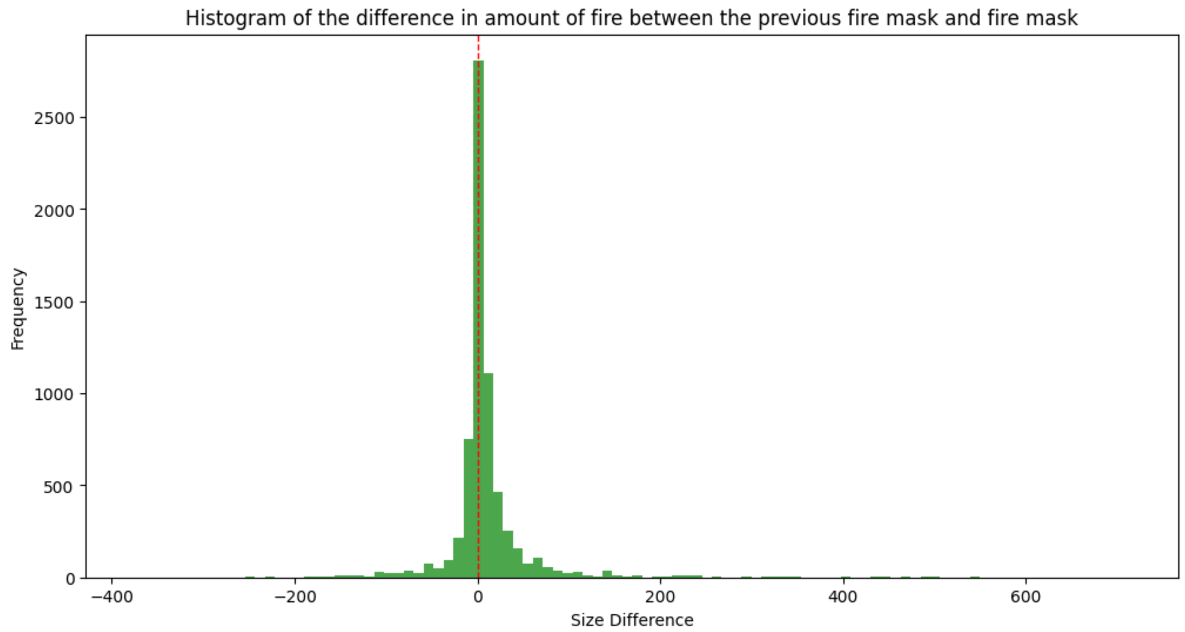

Figure 3.4: Histogram of the difference in fire amount between the previous fire mask and fire mask.

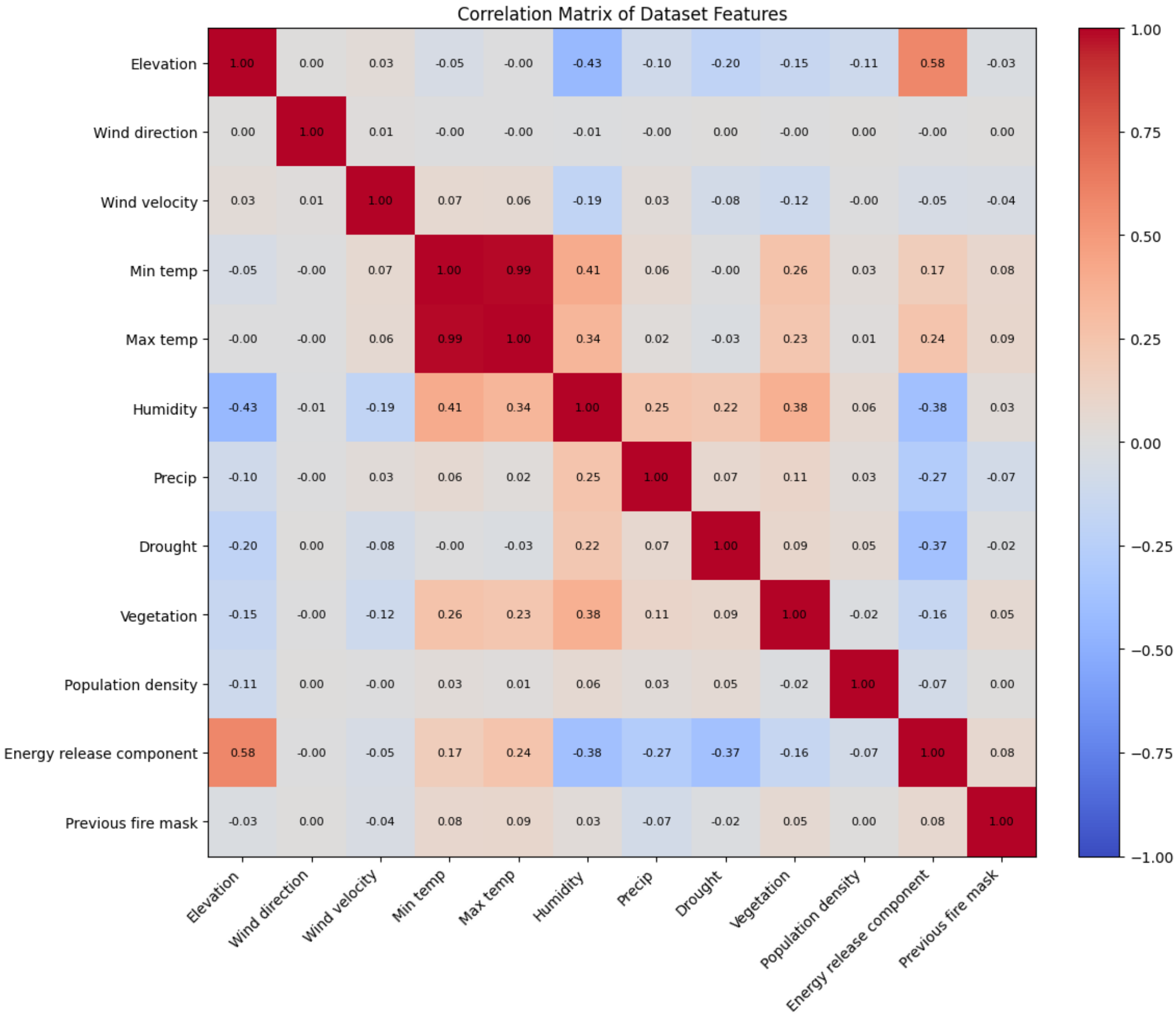

Figure 3.5: Correlation matrix of the features in the Next Day Wildfire spread dataset.

3.1.4. Architectures

Another important consideration is the type of model architecture that is used. The model architecture will be chosen in this section using the requirements from Chapter 2.2, data that was found in Chapter 3.1.2 and explored in Chapter 3.1.3 and given the overarching goal of creating an interpretable, near real-time simulation using physics-informed machine learning.

As previously identified in chapter 1, there are several kinds of wildfire models with unique views on the wildfire spread modeling problem. Given the binary data, the problem can be approached as an image recognition task. Machine learning models, including convolutional neural networks and physics-informed machine learning models, can be used directly for this. While their advantage is that they are able to capture non-linear spatial patterns, their interpretability is poor as they act as black-boxes.

The problem can also be approached as a binary classification problem to which logistic regression models can be applied. While this model provides interpretable outputs as the coefficients show the effect of each parameter, the model is not able to capture non-linear spatial patterns. Additionally the logistic regression model has the assumption that observations are independent. This assumption is violated as there is spatial correlation.

The fire spreading can also be seen as a infection model. In this view the location that is on fire may attempt to set other locations on fire to infect them. In terms of the compartmental models in epidemiology, this process can be represented as a SIR model where a location may be susceptible to wildfires, infected or recovered. In this case the transition function between these states would still need to be estimated using another model type. Additionally the compartmental models typically do not model spatial relationships, which would have to be added.

An additional view that could be used is that of areal data analysis. Within the disease mapping and geostatistics field, the conditional autoregressive (CAR) model is used to model spatial relationships. The model assumes that the state of the cell in question is influenced by the state of the neighboring cells. This allows the model to explicitly capture the spatial effects. The CAR model itself only includes spatial dependence and does not change over time. This model was expanded to include a temporal dimension by Mariella and Tarantino (2010), making it viable for simulation purposes. The advantage of this spatial temporal conditional autoregressive (STCAR) model is that it provides interpretable results: parameters have a clear meaning and the spatial and temporal dependence of a cell have a specific, understandable value. Additionally the method is space agnostic, which allows it to potentially include different types and resolutions of data. While this is not needed for the selected dataset as it is a raster, simulations used in operational firefighting may be able to ingest live data from drones as it becomes available. Other models would not be able to do this. The disadvantage is that the spatial temporal conditional autoregressive model currently only works with continuous data and does not include covariates. However, it does have the potential to be modified and expanded to work with this data. Due to these advantages matching the interpretability requirements and adaptability requirements, this architecture was chosen and will be used in the model as will be described in the next sections.

3.2. Formalization

The spatial temporal conditional autoregressive (STCAR) model that was chosen as the model architecture in the conceptualisation will be formalized in this chapter. First, a brief overview of the formulation of the STCAR model by Mariella and Tarantino (2010) will be given. Then, this model will be extended to be compatible with covariates and binary response data. This will allow it to be applied to the wildfire use case. Finally, the implementation and an optimization using spatial partitioning will be discussed.

3.2.1. Spatial temporal conditional autoregressive (STCAR) models

The spatial temporal conditional autoregressive (STCAR) model as formulated by Mariella and Tarantino (2010) works under the presumption that the variable of interest Z is a continuous, normally distributed random variable that is dependent the value of Z in neighboring cells and its own past values. The STCAR model essentially connects several conditional autoregressive (CAR) models in order to simulate change over time. CAR models are only dependent on their own timestep and thus do not have the temporal dependence. As the model connects several CAR models, the CAR model will first be formally defined. In order to do so, Equations 3.1 to 3.3 were adapted from Mariella and Tarantino (2010). After formally defining the CAR model, the STCAR model will be introduced and then modified to include covariates and work with binary data for the wildfire simulation use case.

Given the spatial domain $S = \{1, .., n\}$ with neighborhoods $\mathcal{N}_i$ for any location i where i and its neighbors, shown as $i*$, are part of S. A CAR model assumes the value of a given location is dependent on its neighbors and is thus characterized by a conditional probability density function as shown in Equation 3.1 and a joint probability density function as will be shown later in Equation 3.2.

$$Z_i \mid Z_{\mathcal{N}_i} \sim \mathcal{N}\left(\mu_i + \rho \sum_{i^* \in \mathcal{N}_i} \beta_{(ii^*)}(Z_{i^*} - \mu_{i^*}), \sigma_i^2\right) \tag{3.1}$$

Here ρ is the spatial dependence. The spatial dependence is a measure for how dependent the value of a cell is on its neighbors. It scales the total effect of all neighboring locations. The effect of a single neighboring location is given by the the deviation from its expected mean scaled by the strength of the relationship $\beta_{(ii*)}$ between the two locations. This is to say that any given location is conditional on all of its neighbors' values based on the strength of their relationship scaled by the amount of spatial dependence.

In addition to the conditional probability density equation 3.1, Mariella and Tarantino (2010) has shown that the joint probability density function can be written as Equation 3.2. This uses a symmetric weighted adjacency matrix W containing the strength of the relationship between site i and its neighbors and a diagonal matrix W_D for normalization with the sum of the weights on the diagonal. The weights are determined by a distance metric φ. The distance metric is closely related to the relationship $\beta_{(ii*)}$ between two locations shown in Equation 3.1 as the normalized weights i.e. $W_D^{-1}W$ are equal to the matrix of interactions containing the relationship $\beta_{(ii*)}$ between two locations for all locations.

$$\mathbf{W} = (w_{ii^*}) \quad \text{with} \quad w_{ii^*} = \begin{cases} 0 & \text{if } i^* = i, \\ \varphi(i, i^*) & \text{if } i^* \in \mathcal{N}_i, \\ 0 & \text{otherwise}, \end{cases}$$

$$\mathbf{W}_D = \operatorname{diag}\left(w_{(1+)}, w_{(2+)}, \ldots, w_{(n+)}\right), \quad \text{where } w_{(i+)} = \sum_{i^* \in \mathcal{N}_i} w_{(ii^*)}, \quad i, i^* \in S.$$

$$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, [\frac{1}{\sigma^2}(W_D - \rho W)]^{-1}) \tag{3.2}$$

Mariella and Tarantino (2010) expand the CAR model by including a temporal domain T. In their model, they chain together several CAR models with expected value 0 linearly with a temporal dependence r as can be seen in Equation 3.3. They do this in such a way that if the temporal dependence is excluded, the model reverts to a CAR model. If the spatial dependence is excluded, the model reverts to a linear regression model. The STCAR model of order p can be written as:

$$\mathbf{B}_t \mathbf{Z}_t = r_1 \mathbf{B}_{(t-1)} \mathbf{Z}_{(t-1)} + r_2 \mathbf{B}_{(t-2)} \mathbf{Z}_{(t-2)} + \cdots + r_p \mathbf{B}_{(t-p)} \mathbf{Z}_{(t-p)} + \boldsymbol{\epsilon}_t, \tag{3.3}$$

where $\boldsymbol{\epsilon}_t$ is the vector of pseudo errors, i.e.

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}\left(0,\, \sigma_t^2 \mathbf{W}_D^{-1} \left(\mathbf{I} - \rho_t \mathbf{W}_D^{-1} \mathbf{W}\right)^\top\right),$$

supposed $\mathbf{B}_t = \mathbf{I} - \rho_t \mathbf{W}_D^{-1} \mathbf{W}$.

The resulting model now provides a continuous value that is dependent upon neighbors in the same timestep, its own value in the previous timesteps as well as the values in the neighborhood in the previous timestep. Similarly to the CAR model, the spatial dependence is given by ρ_t and the relationship between any two locations is contained in W for each location. Unlike the CAR model, the STCAR model also includes the temporal dependence r_t.

3.2.2. STCAR for Wildfires

In the case of wildfire simulation, the goal will be to predict whether any given cell in the fire mask Y in the next timestep is on fire (Y=1) or not (Y=0). This is a classification problem, where the fire mask Y, denoting whether a cell i is on fire or not, is given by a Bernoulli distribution.

$$Y_i \sim \text{Bernoulli}(p_i) \tag{3.4}$$

Here the chance of catching fire p_i for any location is assumed to be dependent on the vegetation, weather and terrain and fire being present in the previous timestep as well as these same variables in the neighboring cells as per the conceptualization in Chapter 3.1.

Unfortunately the STCAR model was developed to work with continuous data and does not support binary data. In order to make the binary data work with the STCAR model, the logit link function as shown in (3.5) can be used. This transforms $P(Y_i = 1)$'s range from [0,1] to $(-\infty, \infty)$, which is suitable for the STCAR model.

$$z_i = \log\left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right) \tag{3.5}$$

Where $P(Y_i = 1)$ is the chance of a given cell i being on fire and z_i is a latent variable. This transformation can later be undone using the inverse logit function (3.6) to retrieve the odds of a given cell catching fire.

$$P(Y_i = 1) = \frac{1}{1 + \exp(-z_i)} \tag{3.6}$$

As z_i is a continuous variable, it can directly be used in a CAR and STCAR model and serve as the expected mean μ_i. As mentioned previously, the chance of catching fire p_i for any location is assumed to be dependent on the vegetation, weather and terrain and fire being present in the previous timestep as well as these same variables in the neighboring cells as per the conceptualization in Chapter 3.1. Since z_i now represents the chance of catching fire, the assumption can be made z_i is dependent on these covariates. This will allow for the model to include these covariates. To start simply, the assumption will be made that the expected mean is equal to a linear combination of the covariates similar to a logistic regression model: $\mu_i = a_0 + \sum a_n * X_{i_n}$. In future iterations of the model, the assumption that it is a linear relationships may be dropped by including a different, non-linear relationship here. Implementing this into the CAR model given in (3.2), looks like this:

$$\mathbf{Z} \sim \mathcal{N}\left(a_0 + \sum_n \boldsymbol{a_n}\mathbf{X_n}, \left[\frac{1}{\sigma^2}(W_D - \rho W)\right]^{-1}\right) \tag{3.7}$$

where $a_0 ... a_n$ are the coefficients for each of the respective covariates $X_1, ..., X_n$ of each location.

This so far is a CAR model, equation (3.3) is used to connect two CAR models together and create an STCAR model. As the data of the Next Day Wildfire Spread dataset (Huot et al., 2022) that was chosen in Chapter 3.1.2 contains only data about the current day and the fire mask of the next day, the wildfire STCAR model will be of order 1. As such there will only be a single spatial dependence parameter r that needs to be estimated, instead of potentially multiple that are included in the original STCAR formulation. In addition, the weights matrix W, σ and ρ are treated as a hyperparameter. This makes it so that only μ, and thus $a_0..a_n$, and the the spatial dependence r are unknown.

To estimate the parameters the likelihood should be maximized. This is equal to minimizing the loss. Since there are two consecutive CAR models, the likelihood of each of the values of Z should be maximized and thus the loss consists of a combination of both of these given a set of parameters that need to be estimated. To evaluate this a build in optimizer can be used that uses the binary cross-entropy (BCE) as loss function as given by (3.9).

$$\begin{aligned} l(Z_1, Z_2 \mid a, \rho, \sigma, r) &= l(Z_1 \mid a, \rho, \sigma) \cdot l(Z_2 \mid a, \rho, \sigma, r) \\ &= \mathrm{BCE}(Z_1) + \mathrm{BCE}(Z_2) \end{aligned} \tag{3.8}$$

$$BCE(Z_i) = -\frac{1}{N}\sum_{i=1}^{N}\left[Y_i \log(Z_i) + (1 - Y_i)\log(1 - Z_i)\right] \tag{3.9}$$

However, we unfortunately do not work directly with the continuous variable's observations, but instead make an estimation of its distribution. In order to then estimate the STCAR model with binary data and covariates, the distribution needs to be estimated using:

$$l(a) = P(Y|a) = \int P(Y|Z)P(Z|a)dZ \tag{3.10}$$

To do this, Monte Carlo simulation can be used. Monte Carlo simulation means that for N runs $a_0..a_n$ are chosen, Z is simulated, the probability is computed using the inverse logit (3.6), determine the likelihood and get the average likelihood for those values of $a_0..a_n$. This can later be used to find the optimal parameters for $a_0..a_n$.

This, however, just assumes that there is only one Z. In the case of the wildfire dataset, the goal is to find the optimal value for Z_{t+1}. In order to do so, Z_t can be used to find a value for Z_{t+1} using one step of Equation (3.3). The loss can then be a combination of both the loss for the estimation of Z_t and Z_{t+1}.

The data flow within a training step is visualized in Figure 3.6. The vector containing the covariates $\mathbf{X}$ is combined with the current best estimation for the coefficients $\mathbf{a}$ to form the mean, resembling the log odds, which is used in the conditional autoregressive model. The log odds is first corrected for the values in the neighboring cells to acquire the log odds in the first time step Z_1. Afterwards, the temporal dependence is used to find the value of the log odds in the next step Z_2. As the prediction based on the assumed log odds should be correct for both the initial and next time step, both are used in the loss function. To do this, the probability they represent is compared to the actual fire masks in both time steps, from which the binary cross-entropy loss is derived. This loss is the value that gets minimized by the built in solver, finding the optimal values for all the parameters.

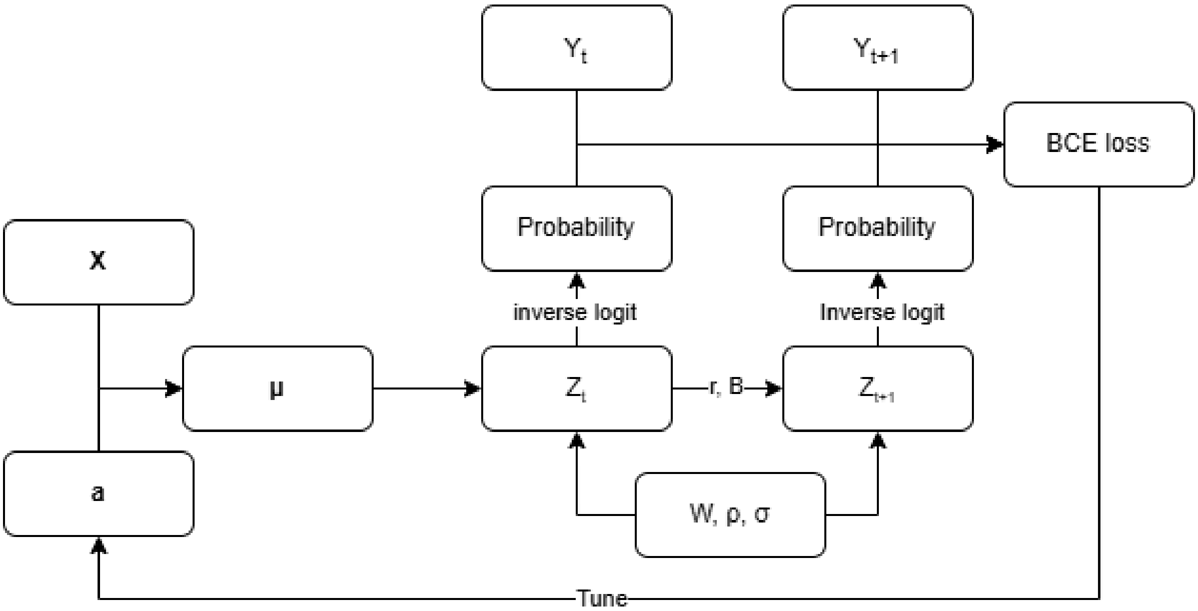

Figure 3.6: Diagram depicting the flow of data in estimating the parameters of the STCAR model. The covariates $\mathbf{X}$, the firemasks Y_t and Y_{t+1} as well as the hyperparameters W, ρ, σ serve as inputs. The values for the coefficients of the covariates are tuned to minimize the binary cross entropy (BCE) loss, which consists of both the BCE loss of the first as well as later timesteps.

3.2.3. Code implementation and optimization

The modified STCAR model as outlined in the previous section was implemented using Tensorflow 2.12. Unfortunately, it turned out that training the STCAR model can be computationally expensive, which makes parameter estimation more challenging. In order to investigate which parts of the STCAR model caused this, a timing function was created that registers the time it took for each part of the STCAR model to run.

Using the timing information the implementation of the weight matrix generation algorithm was improved. Additionally it was found that the time complexity of the STCAR model is $O(N^2)$ where N is the region size or amount of cells to take into account. This was predominantly caused by the multivariate normal distribution used in the STCAR model. Given that the amount of cells increases the time it takes to compute results to this extend, it was investigated whether the region sizes could be decreased. Using insights from the data generating process being a fire that spreads, the included cells can be limited to only those to which the fire should be able to get to within the timestep. This will be referred to as spatial partitioning or spatial splitting. Performing a spatial split comes with the downside that the weighted adjacency matrix has to be recalculated for each sample as the dimensions of the sample are no longer constant. This is seen as an acceptable trade-off as timing the model revealed that recalculating the weighted adjacency matrix takes less time than not performing a spatial split. Using the spatial split with a distance of 10 cells Manhattan distance around a fire decreased the time it takes to assess a single sample using the full 64 by 64 grid from 9 seconds to less than a second for the average sample afterwards. In the current implementation of the model the neighborhood of the model is considered a hyperparameter. In a future iteration of the model the neighborhood and thus the distance metric $\varphi(i, i^*)$ of the weighted adjacency matrix may be determined based on the feature data to more accurate represent which cells are neighbors. If this is the case, the weighted adjacency matrix would already need to be recalculated for each sample, further solidifying the trade-off.

Similarly for spatial partitioning, the algorithm will be implemented using the cell's distance from a fire as the inclusion criteria. In a future iteration the effectiveness of spatially splitting the data may be improved by using the feature's data to determine which cells should be out of bounds, similar to the adjacency matrix mentioned previously. This will allow the area that is taken into account to be limited as much as possible, resulting in a model that would train as fast as possible.

In order to provide a complete overview of the implementation of the modified STCAR model, a flow chart was created that depicts the entire training and evaluation process. This can be seen in Figure 3.7. The training steps indicated in blue match Figure 3.6. The inference process, colored green, uses the trained model to make predictions on the validation or test sets. Here running the model is equivalent to performing the training step on these datasets without tuning the coefficients. The performance statistics that serve as an output of the model here are the area under the precision-recall curve and the time it took to perform the inference.

3.3. Verification and validation

In the previous sections the spatial temporal conditional auto-regressive model was developed. This section will show which steps were taken to ensure that the model is verified and the outputs of the model are valid.

3.3.1. Unit tests and verification of code

In order to verify that the code written for each formula of the model works as expected, the code will be broken up into pieces to look at intermediary results and whether they are correct.

For the weighted adjacency matrix W this meant checking if it was constructed correctly. This meant asserting that it is positive definitive, symmetric, that the distance decay for weights functioned properly and asserting that the connectivity was correctly performed for mock data by investigating the connections for a smaller 3 by 3 area. These all gave expected results.

For the spatial division, verification included inspecting all the outputs of intermediary steps including identifying separate fires, finding the area of interest per fire by dilation, combining areas where relevant and providing the resulting areas. Each of these steps worked and gave expected results. For the parameter estimation and cross entropy loss calculation built in functions were used. These were assumed to perform as expected.

In addition to performing unit tests, each section was first tested with synthetic data to ensure formulas were implemented correctly before real data was introduced. For the covariates values between 0 and 1 were chosen randomly. The fire mask is an

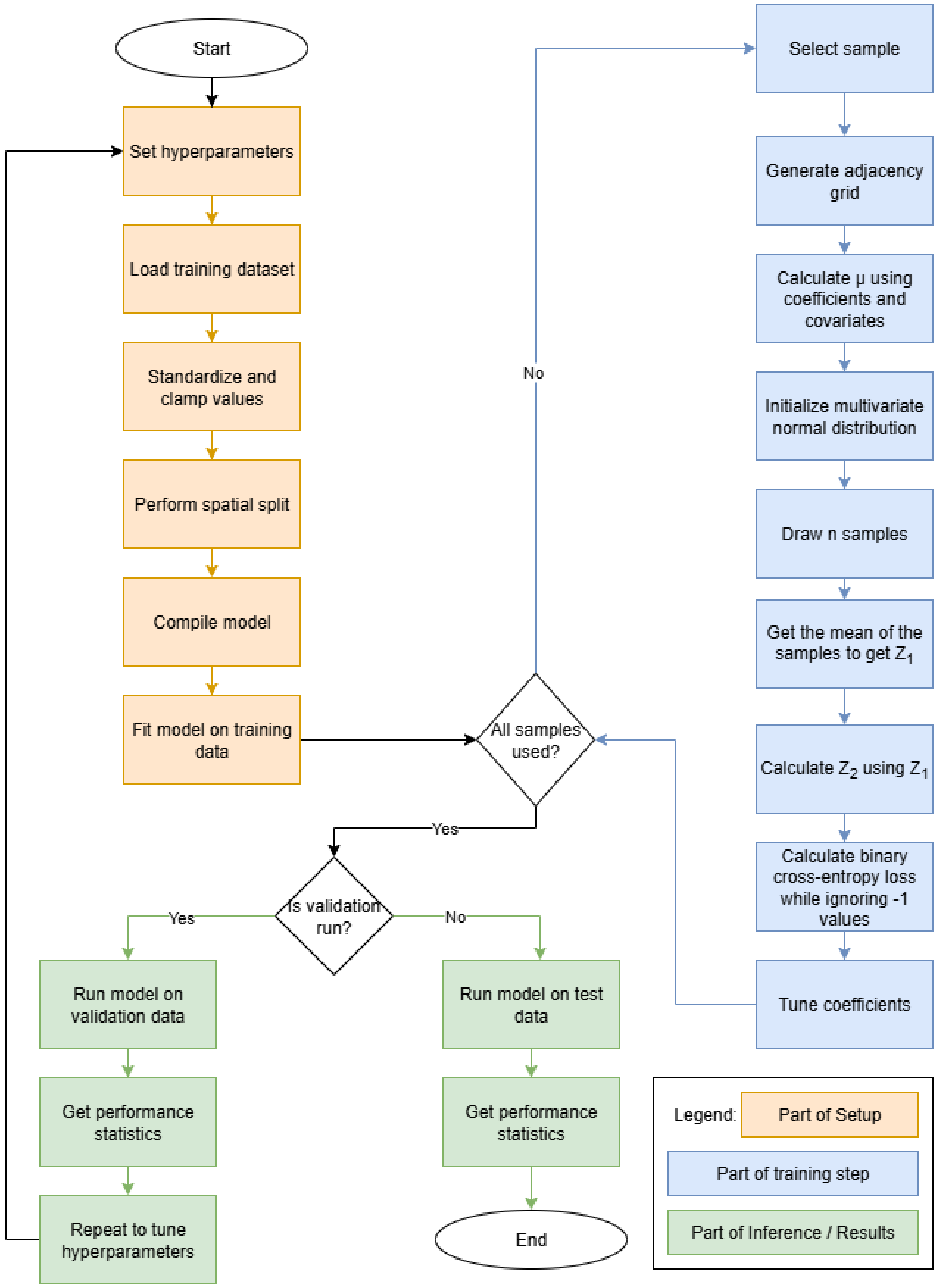

Figure 3.7: Flow chart of the complete training and evaluation process of the modified STCAR model.

exception to this rule as it contains binary values and is imbalanced. Initially, the synthetic fire masks was created by choosing between 0 and 1 randomly. However, as the code to generate the adjacency matrix relies on the distribution of cells on fire, it was chosen to represent this imbalance by having only 5% of the cells on fire in the synthetic data. This approach helped in debugging the functions of the model and ensuring the functions were correctly implemented.

3.3.2. Validation

In order to facilitate the validation of the model, the training data was divided into training, validation and testing data. As previously mentioned, the data was split in a ratio of 8:1:1 between training, validation and testing data by dividing all the weeks in the dataset randomly among these three categories (Huot et al., 2022). This split was already performed by the creators of the dataset to allow models trained on this data to be benchmarked against each other. The validation data were used to make sure the tuned parameters generalize well to unknown data during hyperparameter tuning. Afterwards the model is tested using the test set. Both the results of the hyperparameter tuning and the performance on the test set will be described in Chapter 4.

4 Results

In the previous chapters, the design science research methodology was followed to gather requirements and objectives for wildfire simulations, to understand the design space of wildfire simulations and to develop a spatial temporal autoregressive (STCAR) wildfire simulation model. This chapter will outline the results and assess the performance of the model by comparing it to other simulation tools and the requirements that were established in Chapter 2.2.

4.1. Experimental setup

To evaluate the performance and applicability of the proposed spatial temporal conditional autoregressive (STCAR) model, the model was developed in python using Tensorflow 2.12 using the Adam optimizer. This allows the model to be trained in parallel on the graphical processing unit (GPU). Unfortunately the system used for training did not have a GPU that could be used. The model was trained on an Intel(R) Core(TM) i7-9750H CPU with 16GB DDR4 RAM. All features were min-max scaled for training. In order to be able to compare the performance of the model, the area under the precision recall curve (AUC-PR) is used as a key performance indicator and compared to existing models. The AUC-PR is used over other metrics as it effectively considers the model's precision and recall. These are respectively the ability to find true positives and penalizing false positives. This is important as the dataset is imbalanced. The comparison between the STCAR model and other models will be described in more detail in chapter 4.4.

4.2. Hyperparameter tuning

Before running on the test set to determine the AUC-PR, the model's parameters were inspected and performance was visually inspected using the validation set. In order to improve the model, several models with different hyperparameters and slight modifications were developed and tested on the validation set. While the initial goal was to complete a grid search of the hyperparameters, training each model turned out to be relatively expensive and time consuming as it took up to four hours to train a model. As such only specific hyperparameters were changed. Each of these models and their performance will briefly be described.

First the regular STCAR model was tested. Its parameters revealed that it completely prioritizes the fire mask and the precipitation, generally disregarding the other parameters. The parameter values of the fire mask and precipitation were respectively 18 and 11 where other parameters generally were between -1 and +1. By pausing the model training halfway through the training data, it was revealed that this gap was only growing over time. This is likely due to not accounting for the multicollinearity of the features. The model can't uniquely identify the contribution of each correlated feature, leading to unstable and large coefficient estimates.

The assumption that the model uses linear parameters was changed to use a very basic neural network. The neural network that was used consists of three layers of respectively 64, 32 and 1 node to map the covariates to the expected mean. This was done to both proof that modifying the model to work with a physics-informed neural network would be possible, as well as to test whether the linear assumption hindered the model performance as significantly as is expected. The result of this experiment was that, while the implementation in Tensorflow made it easy to make this modification, the predictions did not improve significantly. The model predicted for nearly all tested samples that the log odds were 0.5, just below predicting there is fire anywhere.

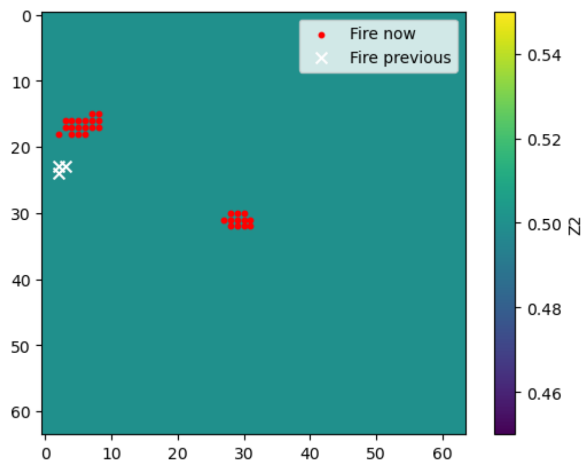

Figure 4.1: Predicted fire mask when using the log odds to scale the weighted adjacency matrix. The fire at t=0 and t=1 are indicated. No fire is predicted. The neural network experiment provided similar results.

As the reason for the neural network model not providing predictions was not fully understood, it was decided to attempt to modify the model to address another shortcoming of the STCAR model's current implementation. Investigating the outputs of the model visually revealed that rather than providing a directional flow, the current weight matrix being symmetrical and using a distance metric resulted in spatial smoothing without directionality. To account for this another model was developed in which the log odds was added to the weighted adjacency matrix. The idea behind this modification is that by including the log odds in the adjacency matrix, the covariates influence the directionality of the spread directly. This ended up giving similar predictions to the neural network experiment. One prediction is shown in figure 4.1. Surprisingly it turned out that this model completely disregards the temporal dependence and spatial dependence with $r = 0$ and $\rho = 0$ respectively. This is likely a result of the loss for predicting the first mask weighing as much as the loss in the second timestep, while it can be completely derived from the previous fire mask.

4.3. Model performance

A final model was created using the linear assumption, but with a bias in the loss of the second timestep. This meant that errors in the predictions of the second timestep would weigh thrice as much as errors in the first timestep. Overall this STCAR model still performed poorly with a AUC-PR of 0.1264. This model also still primarily relied on the precipitation and fire mask, being 18 and 39 times as important as the other parameters. The exact coefficients for each parameter can be found in Appendix B. Inspecting these coefficients revealed strange values. The minimum and maximum temperature both have negative signs while a higher temperature is expected to increase the likelihood of wildfires. Additionally the precipitation has a very strong positive coefficient which is unexpected as higher precipitation is often associated with a higher moisture content, which decreases fire spread. It is assumed that these unexpected values are a result of the multicollinearity not being accounted for. This may flip signs in order to account for the contribution of other covariates. Furthermore most samples do not contain any precipitation as shown by the right tailed distribution in the exploratory data analysis. The model may not have enough samples to properly learn the effects of precipitation.

Most sample's predictions do not have any fire predicted. The predictions where fire is predicted, such as shown in figure 4.3, show the spatial smoothing effects of the STCAR model. These also show that the STCAR model is not able to capture the directionality in the fire spread. There are also some samples where the precipitation is relatively high leading to a significant overprediction of the fire mask such as can be seen in Figure 4.5. For both predictions the sample's covariates were included in respectively Figure 4.2 and Figure 4.4 to provide context to the prediction.

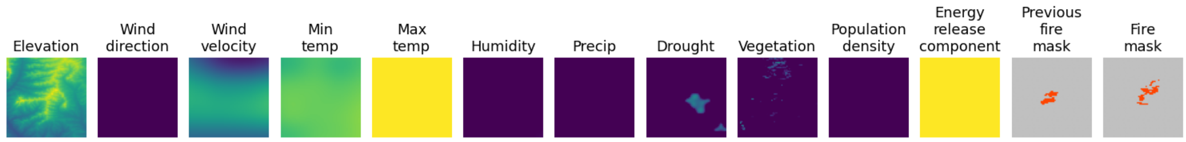

Figure 4.2: Sample of the test dataset used in Figure 4.3. The data is normalized with purple representing 0 and yellow representing 1.

4.4. Comparison to other simulation tools

In order to properly estimate the success of the STCAR model, its performance will be compared to other simulation tools using the same dataset. In the dataset's paper, Huot et al. (2022) demonstrate the usability of their data by training a logistic regression model, a random forest model and a deep learning model. The logistic and random forest models were created by using a 3 by 3 kernel. The deep learn-

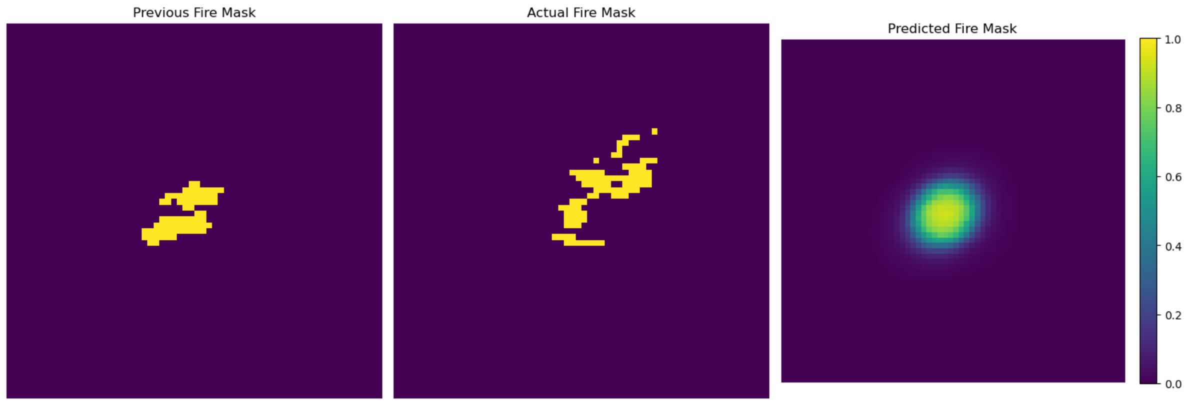

Figure 4.3: Comparison between the previous fire mask (t=0), the actual fire mask (t=1) and the probabilistic predicted fire mask (t=1). The prediction shows the spatial smoothing effect of the STCAR model.

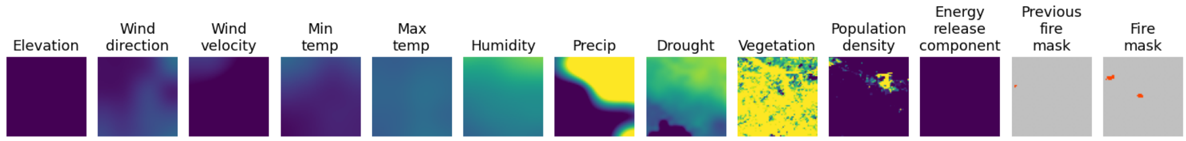

Figure 4.4: Sample of the test dataset used in Figure 4.5. The data is normalized with purple representing 0 and yellow representing 1. The precipitation is very high resulting in the overprediction in Figure 4.5

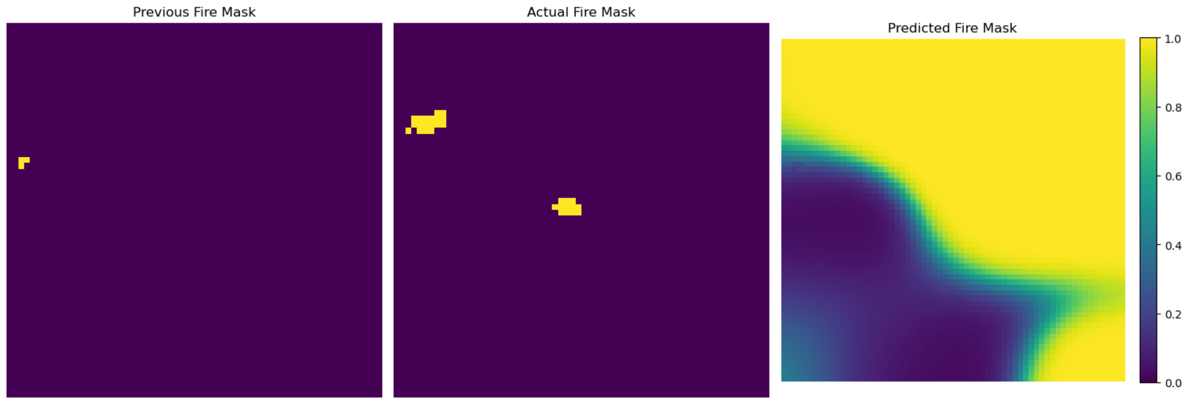

Figure 4.5: Comparison between the previous fire mask (t=0), the actual fire mask (t=1) and the probabilistic predicted fire mask (t=1). The predicted fire mask shows how precipitation can cause severe overestimation of the amount of fire.

ing model consists of several residual blocks. For full implementation details of these models, the paper describing the dataset and developed models by Huot et al. (2022) may be consulted.

In order to provide a benchmark for models such as the STCAR model, Huot et al. (2022) assess their models using the area under the precision curve (AUC PR) as a key performance indicator. Their logistic regression model has a AUC PR of 0.198. The random forest has a AUC PR of 0.225 and the deep learning model has a AUC PR

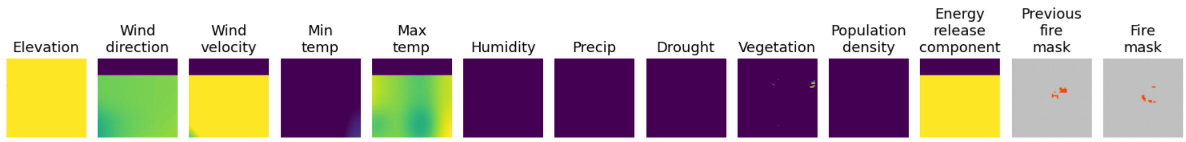

Figure 4.6: Data sample from the test set showing artifacts in most features. Data in the sample is normalized with purple representing 0 and yellow representing 1.

of 0.284. While these models also have a relatively low performance, each of these performs better than the STCAR model, which has an AUC PR of 0.126. Despite providing fire scars in their paper, Huot et al. (2022) does not provide the index of these fire scars in the dataset and as such no direct comparison can be made. Similar to the STCAR model, they report that their models have a strong dependence on the previous fire mask and provide smoother boundaries and can merge separate segments. These behaviors were also present in the STCAR model as is evident from Figure 4.3. In addition to the limitations mentioned, the relatively poor performance of both the STCAR model and the models presented by Huot et al. (2022) may also be a result of artifacts in the dataset. During the visual inspection of the model on test data it was discovered that some samples still contain illogical values that likely are a result of noise, such as shown in Figure 4.6.

4.5. Applicability of the model to simulation use cases

Identifying the use cases, stakeholders and their requirements led to finding a list of requirements presented in table 2.1. This table will be used to find which requirements were and were not satisfied. The result of this can be found in Table 4.1. Requirements were classified as either being achieved by the current implementation (+), not achieved by the current implementation, but should reasonably be achievable using the STCAR model (-+) and not achievable (-). The most important requirements and results will now be discussed.

As is evident from the results in Chapter 4.3, the performance of the current implementation of the STCAR wildfire model is not suitable for any application. As the model cannot provide an accurate fire scar, requirement 1.1 is not fulfilled. However, the model does provide the fire spread as output of the model. This achieves requirement 1 with a spatial resolution of 1 km and a temporal resolution of 1 day. Requirement 1.2 is only partially achieved as the model only predicts one day ahead. While the model could theoretically be used to predict multiple days ahead, the current dataset contains no data to test these predictions against. This resolution could be improved in the future by using a dataset with a better spatial or temporal resolution.

Requirement 2 of providing interpretable outputs is fulfilled as the model's outputs can be interpreted directly. It should be noted that the current model results provide illogical coefficients as a result of the linearity multicollinearity that is not accounted for, however the parameters are interpretable as the spatial dependence and temporal dependence parameters have clear meanings. Similarly the coefficients for the covariates can be interpreted as it denotes the effect it has on the cell catching fire, before being corrected for spatial and temporal effects. The current implementation of the model works with the assumption that the parameters are combined in a linear fashion. If this assumption is dropped in favor of using a neural network or other model that can capture non-linear effects, the model may perform better at the cost of the interpretability.

The requirements for testing interventions were partially fulfilled. Requirement 3.1 was partially achieved as the model can use custom data for the energy release component and as a vegetation index to simulate changing the fuel load. However these modifications are limited to the spatial resolution of 1 km of the model. The model provides probabilistic outputs for the fire scar, as such requirement 3.2 is fulfilled. This however assumes that a fire is given. The output could be improved by running multiple simulations with different ignition points to improve the result. In order to facilitate this effectively, the model also needs to be able to be run in parallel. While the current model was not run in parallel as the setup did not have a GPU to achieve requirement 3.2.2, this is possible within the current implementation.

The requirements surrounding the processing of data were partially achieved, but also could be improved. The data validation of requirement 4.2 is fulfilled as extreme data is clipped to reasonable numbers. Nevertheless, the remote sensing data that is used in the model is noisy by nature and further validation of input data could be added to find and limit the noise. Similarly requirement 4.3 on handling missing and uncertain data is not fulfilled as uncertain data currently is ignored in training.

The model architecture was chosen specifically to achieve the adaptability requirements. Requirement 5.2 highlights the strength of the model as the model is space-agnostic and thus different spatial resolutions can be integrated together in future iterations. This can allow data from drones to work together with satellite data to create an operational model with the most up-to-date data. As for requirement 5.1, the current dataset is not georeferenced and thus the outputs cannot be used as a GIS layer. However, the code used to retrieve and compile the dataset from the Google Earth Engine platform is open source. Thus the correct coordinates and coordinate reference systems can be added to the dataset in future iterations of the model.

Overall the STCAR model falls short of being applicable to any use case due to its poor accuracy. The STCAR methodology does have the potential to achieve many of the requirements with slight modifications to its implementation or using different training data. These modifications were not implemented due time restrictions and the chosen scope of the project. The main goal of the thesis was to provide the basic model for physics-informed wildfire simulation to be built upon. The STCAR methodology itself provides a good basis for these extensions.

Table 4.1: Wildfire simulation requirements met by the STCAR model. Fullfilment of a requirement is indicated by a plus sign. Partial fullfilment or being extendable/modifiable to achieve a requirement is indicated by a plus and minus sign. A minus sign means not achieved.

Number	Requirement	Prioritization	Achieved?
Goal	To create an interpretable wildfire simulation to support firefighters		
1	Provide information on wildfire spread	Must	+
1.1	Provide an accurate fire scar	Must	-
1.2	Show when the fire arrives at different locations	Must	-+
2	Provide interpretable outputs explaining fire behavior drivers	Should	+
3	Test the effectiveness of interventions	Should	-+
3.1	Adjust the fuel load	Could	-+
3.2	Show risk profiles	Could	+
3.2.1	Aggregate information from several simulations	Could	-+
3.2.2	Be able to run in parallel	Should	+
4	Process information	Must	+
4.1	Be able to integrate near real-time data	Must	+
4.2	Validate information	Should	+
4.3	Handle missing or uncertain input data	Must	-+
5	Be extendable to different use cases	Should	+
5.1	Be usable as a GIS layer	Should	-+
5.2	Adaptable to different data types	Should	+

5
Discussion and conclusion

This study examined existing wildfire modeling approaches and identified a fundamental trade-off in the literature: current real-time models are typically purely data-driven and act as black boxes, whereas physics-based models, although interpretable and grounded in physics, often fail operational constraints due to computational and data demands. This gap motivates the need for hybrid approaches that are both interpretable and computationally efficient. As such the main research question was posed as: *"How can a near real-time wildfire simulation be designed using physics-informed machine learning?"*

In order to fill this gap, the requirements for wildfire simulations were investigated. This included a brief review of wildfire simulation use cases, stakeholders and requirements guided by the first sub-question: *"What objectives can be derived from how models are currently being used?"*. This resulted in identifying multiple use cases for wildfire simulations including decision support during active wildfires, and during planning for prescribed burns, risk assessments applications, disaster response and ecological monitoring. Requirements for each of the applications were gathered by trying to find usage of the models, however model usage is not always reported and if it is reported the requirements for the models may not be clearly stated. The current requirements were thus estimated based on the way the models are used and are mostly generic, but do provide guidance in the design process as well be outlined in the next paragraphs. Improvements to the list of requirements could be made by directly inquiring the stakeholder involved. Given the overarching goal of creating a model that is interpretable and computationally efficient, operational firefighting was chosen as the core use case to focus on. This use case was chosen as it necessitates accurate results with low inference times. The requirements for operational firefighting were prioritized using the MoSCoW method. Providing accurate results and being able to integrate near real-time data were found to be the most important requirements.

After finding the requirements for the simulation, the design space of wildfire simulations was explored in order to answer the second sub-question: *"What could the model's architecture look like?"*. First the data requirements and available data were investigated. This resulted in identifying topology, vegetation, weather and human intervention as desired parameters. Additionally it resulted in identifying remote sensing data as preferred data source over simulated and field data. Remote sensing data was preferred as it has a much greater availability and coverage. This allows the model to be applicable world-wide and significantly decreases the risk of outdated data, which could lead to inaccurate results. Four remote sensing datasets were found and compared. The Next Day Wildfire Dataset was chosen out of these four for its completeness, as it also includes a proxy for human interaction, because of its ease of use, as it has complete documentation and because the code used to create it is open source and easily modifiable for future adaptation.

Knowing the available data and requirements, different types of models were investigated and several possible architectures were identified that could meet part of the requirements. Ultimately spatial temporal conditional autoregressive (STCAR) models were chosen for having the most potential. This model is based on the assumption that the value of a given location is based on its surroundings in the current time step and itself in the previous time steps. The advantage of this model is that it is space-agnostic, is interpretable and relatively computationally efficient. The downside to this kind of model is that it is not specifically designed to work with the types of data in the Next Day Wildfire Dataset: it does not include covariates and does not work with binary data and thus the model had to be modified. In order for the model to work with binary data a logit link function was used. In order to include covariates, the model was modified by making the log odds a combination of the covariates, which then can be used in the STCAR model. this effectively makes it so that the STCAR model predicts the spread of the odds of a cell catching on fire.

Testing showed that the model did not perform well compared to existing models trained on the same dataset. The modified STCAR model has an area under the precision-recall curve (AUC-PR) of just 0.126. This is significantly less than the neural network, random forest and logistic regression models trained by Huot et al. (2022) which achieved 0.284, 0.225, and 0.198 respectively. The modified STCAR model relied primarily on the previous fire mask and precipitation for its predictions. Relying on the precipitation caused severe overpredictions in some samples.

The poor performance of the modified spatial temporal conditional auto-regressive model is a result of several limitations in both the dataset used as well as the model architecture itself. First the limitations in the dataset will be addressed. Afterwards the limitations in the model architecture will be discussed. For each of these future research directions will be listed.

The first limitation of the dataset is that it uses single observations of daily data from the MODIS satellite at a resolution of 1 kilometer. Independent daily observations are a severe limitation to the predictive potential of a model as no knowledge of the previous days can be integrated into the model. This means that based purely on the data, it is impossible to see if a fire is growing or shrinking. Additionally the trajectory of the fire in previous days may be used in the prediction of future days as the available fuel decreases in those areas. In addition to using insights from previous days, models may be able to use weather forecasts in their predictions. The spatial resolution of the MODIS data also limit the predictive performance of the model. Due to this resolution, certain features that may act as natural barriers such as rivers might be missed. In their research comparing the predictive performance of the MODIS and VIIRS satellite data, Karlsson et al. (2025) conclude that MODIS data is unsuitable for next day wildfire spread prediction due to exhibiting a highly stochastic fire mask and that the VIIRS product is better suited for this task due to its improved spatial resolution. The WildfirespreadTS dataset addressed both of these limitations by providing multi-day observations, providing weather forecast data, using VIIRS wildfire data and providing data at a spatial resolution of 375 meters (Gerard et al., 2023). Using this dataset may offer an improvement over the current dataset.

In addition to using a different remote sensing dataset, other data could be integrated into the model. Remote sensing data has a lower level of detail than is available through field data. Integrating data of a higher resolution into the model may cause it to not miss natural barriers that could still be missed with an improved spatial resolution. This could thus improve the model's performance. Since the STCAR method is space-agnostic, there is potential for combining higher resolution data that is known to be up to date and accurate, such as from unmanned aerial vehicles, with the lower resolution data that is currently used. Similarly, if real time data is available from geostationary satellites at a greater spatial resolution, these could be used to check the inputs given by the other sensors and fill in uncertainty as the Next Day Wildfirespread dataset contained 1% uncertain labels.

Another limitation of the dataset is that extreme wildfire behavior also likely will not be represented well by the model as these events and associate fire behavior are rare and thus does not occur often in the dataset as was revealed in the exploratory data analysis in Chapter 3.1.3.

Additionally the Next Day Wildfirespread dataset is not georeferenced as exact locations are missing in the training data and metadata. This is a limitation as the importance of factors may differ per ecosystem (Alexander & Cruz, 2013). While the Normalized Difference Vegetation Index (NDVI) does act as a proxy for the eco-region, it is unlikely that this effect is fully covered. As the code to generate the dataset of both the Next Day Wildfirespread dataset and the WildfirespreadTS dataset is open source and Google Earth Engine, which the Next Day Wildfire dataset was derived from, is able to provide the locations, this data can be added to the dataset that is used in future iterations. Adding this data will also allow the model to work as a layer in geographic information system (GIS) software. This makes the model interoperable with other models expanding the use cases the model may be used for and meeting the interoperability requirement.

There are also limitations present in the implementation of the spatial temporal conditional auto-regressive (STCAR) model. The most important limitation is that the STCAR model in its current implementation is not able to capture the directionality of fire spreading. The weighted adjacency matrix used in the STCAR model consists of a distance metric to show the relationship between cells. This causes the STCAR model to show a smoothing effect rather than directionality. This could be replaced by a data driven metric. Additionally, directionality may be promoted by transforming the features in the dataset. The wind direction and elevation could be transformed into vectors. The features could be corrected for the similarity of these vectors to the vector in the direction of the fire. This was attempted but not successfully implemented due to time constraints. Another method that was considered to promote directionality as well as improve the computational efficiency is by dividing the prediction of one day into several predictions at a smaller temporal resolution, which also facilitates using smaller spatial resolutions multiple times. It is assumed this decreases the computation time as the computation time has a quadratic relationship with the amount of cells it needs to account for. Improvements in computation time will also allow for further hyperparameter tuning as the current hyperparameter tuning was limited due to the complete training taking up to four hours.

Finally the model may be improved by not assuming the log odds is a result of a linear combination of the covariates. The current model assumes a linear relationship between the features and the odds of a cell catching on fire, corrected by the spatial on the cells around it and temporal dependence on itself in the previous timestep. It is widely understood that wildfire spreading is not a linear process. As such the linearity assumption was meant as a temporary assumption that would later be replaced. The model was created in such a way that replacing this assumption with, for example, a physics-informed neural network is as easy as possible. This should make it so that the nonlinearities in the data generating process of wildfires are captured to a much greater extend. This does, however, decrease the interpretability of the model and decreases the computational efficiency and thus forms a trade-off. The attempt to implement a simple neural network resulted in no fire being predicted. The reason for this is not fully understood, as such the linearity assumption was kept. Even if the linear assumption is kept the model can be improved by selecting specific features according to the correlations between them as the current model does not account for the multicollinearity. This also hindered the interpretation of the parameters and may have caused the flipped signs in the coefficients of the model.

While the STCAR model in its current implementation suffers from these limitations and as a result does not provide accurate predictions and thus cannot be applied to any use case, it does serve as an initial proof of concept. The model architecture has interpretable parameters, is able to run in parallel, adaptable to different data types and set up to be able to integrate neural networks. Once these limitations are addressed, the model may form the foundation for extensions that integrate physics-informed machine learning to provide the interpretable near real-time wildfire spread predictions that this thesis set out to create.

In conclusion, near real-time physics-informed wildfire simulations can be designed by carefully considering the stakeholders, requirements, available data and model architectures. As a result of this process, this thesis contributes to the field of wildfire modeling by introducing the spatial temporal conditional autoregressive model that can be modified with physics-informed and data-driven extensions. While the current implementation suffers from multiple limitations and does not outright use physics-informed machine learning yet, it does provide a foundation for future research in scalable, transparent wildfire simulation systems, which may hopefully make more accurate predictions possible that can be used in practice to make informed decisions that save lives.

Data and code availability

The data of the Next Day Wildfire Spread dataset is available at https://www.kaggle.com/datasets/fantineh/next-day-wildfire-spread. Data is available upon request to the author in the event that the dataset is no longer available on the Kaggle website. The dataset and the preprocessing of the dataset is described by Huot et al. (2022).

Cleaned code of the exploratory data analysis of the Next Day Wildfirespread dataset and the cleaned code of the spatio-temporal conditional autoregressive model implementation for wildfire simulations are available at https://github.com/GerbenBultema/STCAR-wildfire-model. Uncleaned code may be made available upon request. The model was originally developed on a private repository on a private account and later moved to a public repository, as such the complete commit history will not be available.

Bibliography

Alexander, M. E. (2000). *Fire behaviour as a factor in forest and rural fire suppression* (Forest Research Bulletin No. 197, Forest and Rural Fire Scientific and Technical Series Report No. 5). Forest Research, Rotorua, in association with the New Zealand Fire Service Commission and the National Rural Fire Authority, Wellington.

Alexander, M. E., & Cruz, M. G. (2013). Limitations on the accuracy of model predictions of wildland fire behaviour: A state-of-the-knowledge overview. *The forestry chronicle*, *89*(3), 372–383. https://doi.org/10.5558/tfc2013-067

Alexandridis, A., Russo, L., Vakalis, D., & Siettos, C. (2011). Simulation of wildland fires in large-scale eterogeneous environements. *Chemical Engineering Transactions*, *24*, 433–438. https://doi.org/10.3303/CET1124073

Andrews, P. L. (2018). The rothermel surface fire spread model and associated developments: A comprehensive explanation. *Gen. Tech. Rep. RMRS-GTR-371. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 121 p.*, *371*. https://doi.org/10.2737/RMRS-GTR-371

Bao, M., Liu, J., Ren, H., Liu, S., Ren, C., Chen, C., & Liu, J. (2024). Research trends in wildland fire prediction amidst climate change: A comprehensive bibliometric analysis. *Forests*, *15*(7). https://doi.org/10.3390/f15071197

Bottero, L., Calisto, F., Graziano, G., Pagliarino, V., Scauda, M., Tiengo, S., & Azeglio, S. (2020, December). *Physics-informed machine learning simulator for wildfire propagation*. https://doi.org/10.48550/arXiv.2012.06825

Brenkert-Smith, H., Dickinson, K. L., Champ, P. A., & Flores, N. (2013). Social amplification of wildfire risk: The role of social interactions and information sources. *Risk analysis*, *33*(5), 800–817. https://doi.org/10.1111/j.1539-6924.2012.01917.x

Buechi, H., Weber, P., Heard, S., Cameron, D., & Plantinga, A. J. (2021). Long-term trends in wildfire damages in California. *International Journal of Wildland Fire*, *30*(10), 757–762. https://doi.org/10.1071/wf21024

Castrillón, M., Jorge, P., López, I., Macías, A., Martín, D., Nebot, R., Sabbagh, I., Quintana, F., Sánchez, J., Sánchez, A., Suárez, J., & Trujillo, A. (2011). Forecasting and visualization of wildfires in a 3d geographical information system. *Computers and Geosciences*, *37*(3), 390–396. https://doi.org/10.1016/j.cageo.2010.04.011

Cawson, J. G., Hemming, V., Ackland, A., Anderson, W., Bowman, D., Bradstock, R., Brown, T. P., Burton, J., Cary, G. J., Duff, T. J., et al. (2020). Exploring the key drivers of forest flammability in wet eucalypt forests using expert-derived conceptual models. *Landscape Ecology*, *35*, 1775–1798. https://doi.org/10.1007/s10980-020-01055-z

Chen, Y., Morton, D. C., & Randerson, J. T. (2024). Remote sensing for wildfire monitoring: Insights into burned area, emissions, and fire dynamics. *One Earth*, *7*(6), 1022–1028. https://doi.org/10.1016/j.oneear.2024.05.014

Cheng, S., Prentice, I., Huang, Y., Jin, Y., Guo, Y.-K., & Arcucci, R. (2022). Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. *Journal of Computational Physics*, *464*. https://doi.org/10.1016/j.jcp.2022.111302

Chi, S.-D., Lim, Y.-H., Lee, J.-K., Lee, J.-S., Hwang, S.-C., & Song, B.-H. (2003). A simulation-based decision support system for forest fire fighting. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2829*, 487–499. https://doi.org/10.1007/978-3-540-39853-0_40

de Gennaro, M., Billaud, Y., Pizzo, Y., Garivait, S., Loraud, J.-C., El Hajj, M., & Porterie, B. (2017). Real-time wildland fire spread modeling using tabulated flame properties. *Fire Safety Journal*, *91*, 872–881. https://doi.org/10.1016/j.firesaf.2017.03.006

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophysical Research Letters*, *41*(8), 2928–2933. https://doi.org/10.1002/2014gl059576

Drury, S. A., Rauscher, H. M., Banwell, E. M., Huang, S., & Lavezzo, T. L. (2016). The interagency fuels treatment decision support system: Functionality for fuels treatment planning. *Fire Ecology*, *12*(1), 103–123. https://doi.org/10.4996/fireecology.1201103

Federal Emergency Management Agency. (2025). *Wildfires*. https://www.ready.gov/wildfires

Fernandes, P. M., & Botelho, H. S. (2003). A review of prescribed burning effectiveness in fire hazard reduction. *International Journal of wildland fire*, *12*(2), 117–128. https://doi.org/10.1071/WF02042

Finney, M. A. (1998). *Farsite: Fire area simulator-model development and evaluation*. U.S. Department of Agriculture, Forest Service. https://doi.org/10.2737/rmrs-rp-4

Finney, M. A., McHugh, C. W., Grenfell, I. C., Riley, K. L., & Short, K. C. (2011). A simulation of probabilistic wildfire risk components for the continental united states. *Stochastic Environmental Research and Risk Assessment*, *25*(7), 973–1000.

FLARE Wildfire Research. (2021, December). *Phoenix rapidfire*. https://www.flarewildfire.com/software/phoenix-rapidfire/

Gerard, S., Zhao, Y., & Sullivan, J. (2023). Wildfirespreadts: A dataset of multi-modal time series for wildfire spread prediction. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in neural information processing systems* (pp. 74515–74529, Vol. 36). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2023/file/ebd545176bdaa9cd5d45954947bd74b7-Paper-Datasets_and_Benchmarks.pdf

Gharakhanlou, N., & Hooshangi, N. (2021). Dynamic simulation of fire propagation in forests and rangelands using a gis-based cellular automata model. *Inter-*

national Journal of Wildland Fire, *30*(9), 652–663. https://doi.org/10.1071/WF20098

Gilless, J. K., & Fried, J. S. (1999). Stochastic representation of fire behavior in a wildland fire protection planning model for california. *Forest Science*, *45*(4), 492–499. https://doi.org/10.1093/forestscience/45.4.492

Grajdura, S., Borjigin, S., & Niemeier, D. (2022). Fast-moving dire wildfire evacuation simulation. *Transportation Research Part D: Transport and Environment*, *104*, 103190. https://doi.org/https://doi.org/10.1016/j.trd.2022.103190

Hajian, M., Melachrinoudis, E., & Kubat, P. (2016). Modeling wildfire propagation with the stochastic shortest path: A fast simulation approach. *Environmental Modelling and Software*, *82*, 73–88. https://doi.org/10.1016/j.envsoft.2016.03.012

Halofsky, J. E., Peterson, D. L., & Harvey, B. J. (2020). Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. *Fire Ecology*, *16*(1). https://doi.org/10.1186/s42408-019-0062-8

Hertelendy, A. J., Howard, C., Sorensen, C., Ranse, J., Eboreime, E., Henderson, S., Tochkin, J., & Ciottone, G. (2024). Seasons of smoke and fire: Preparing health systems for improved performance before, during, and after wildfires. *The Lancet Planetary Health*, *8*(8), e588–e602. https://doi.org/10.1016/S2542-5196(24)00144-X

Hung, W.-T., Baker, B., Campbell, P. C., Tang, Y., Ahmadov, R., Romero-Alvarez, J., Li, H., & Schnell, J. (2025). Fire intensity and spread forecast (fira): A machine learning based fire spread prediction model for air quality forecasting application. *GeoHealth*, *9*(3), e2024GH001253. https://doi.org/10.1029/2024GH001253

Huot, F., Hu, R. L., Goyal, N., Sankar, T., Ihme, M., & Chen, Y.-F. (2022). Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–13. https://doi.org/10.48550/arXiv.2112.02447

Jahdi, R., Salis, M., Darvishsefat, A., Alcasena, F., Mostafavi, M., Etemad, V., Lozano, O., & Spano, D. (2016). Evaluating fire modelling systems in recent wildfires of the golestan national park, iran. *Forestry*, *89*(2), 136–149. https://doi.org/10.1093/forestry/cpv045

Jiang, W., Wang, F., Su, G., Li, X., Wang, G., Zheng, X., Wang, T., & Meng, Q. (2022). Modeling wildfire spread with an irregular graph network. *Fire*, *5*(6). https://doi.org/10.3390/fire5060185

Karafyllidis, I., & Thanailakis, A. (1997). A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, *99*(1), 87–97. https://doi.org/10.1016/S0304-3800(96)01942-4

Karlsson, J., Xu, Y., Berg, A., & Haglund, L. (2025). Comparing next-day wildfire predictability of modis and viirs satellite data. https://arxiv.org/abs/2503.08580

Khanmohammadi, S., Arashpour, M., Golafshani, E., Cruz, M., Rajabifard, A., & Bai, Y. (2022). Prediction of wildfire rate of spread in grasslands using machine learning methods. *Environmental Modelling and Software*, *156*. https://doi.org/10.1016/j.envsoft.2022.105507

Liu, Y., Kochanski, A., Baker, K. R., Mell, W., Linn, R., Paugam, R., Mandel, J., Fournier, A., Jenkins, M. A., Goodrick, S., et al. (2019). Fire behaviour and smoke modelling: Model improvement and measurement needs for next-generation smoke research and forecasting systems. *International journal of wildland fire*, *28*(8), 570–588. https://doi.org/10.1071/WF18204

Lopes, A., Ribeiro, L., Viegas, D., & Raposo, J. (2019). Simulation of forest fire spread using a two-way coupling algorithm and its application to a real wildfire. *Journal of Wind Engineering and Industrial Aerodynamics*, *193*. https://doi.org/10.1016/j.jweia.2019.103967

Ma, F., & Lee, J. Y. (2025). Analyzing wildfire evacuation dynamics with agent-based modeling in damaged road networks. *Safety Science*, *187*, 106835. https://doi.org/https://doi.org/10.1016/j.ssci.2025.106835

Maeda, E., Formaggio, A., Shimabukuro, Y., Arcoverde, G., & Hansen, M. (2009). Predicting forest fire in the brazilian amazon using modis imagery and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, *11*(4), 265–272. https://doi.org/10.1016/j.jag.2009.03.003

Mariella, L., & Tarantino, M. (2010). Spatial temporal conditional auto-regressive model: A new autoregressive matrix. *Austrian Journal of Statistics*, *39*(3), 223–244. https://doi.org/10.17713/ajs.v39i3.246

Mitchell, H., Gwynne, S., Ronchi, E., Kalogeropoulos, N., & Rein, G. (2023). Integrating wildfire spread and evacuation times to design safe triggers: Application to two rural communities using peril model. *Safety Science*, *157*. https://doi.org/10.1016/j.ssci.2022.105914

Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., et al. (2011). Landscape–wildfire interactions in southern europe: Implications for landscape management. *Journal of environmental management*, *92*(10), 2389–2402. https://doi.org/10.1016/j.jenvman.2011.06.028

Murali Mohan, K., Satish, A., Mallikharjuna Rao, K., Yarava, R., & Babu, G. Leveraging machine learning to predict wild fires. In: 2021, 1393–1400. https://doi.org/10.1109/ICOSEC51865.2021.9591952

Ntinas, V., Moutafis, B., Trunfio, G., & Sirakoulis, G. (2017). Parallel fuzzy cellular automata for data-driven simulation of wildfire spreading. *Journal of Computational Science*, *21*, 469–485. https://doi.org/10.1016/j.jocs.2016.08.003

Pais, C., Carrasco, J., Martell, D., Weintraub, A., & Woodruff, D. (2021). Cell2fire: A cell-based forest fire growth model to support strategic landscape management planning. *Frontiers in Forests and Global Change*, *4*. https://doi.org/10.3389/ffgc.2021.692706

Pang, B., Cheng, S., Huang, Y., Jin, Y., Guo, Y., Prentice, I. C., Harrison, S. P., & Arcucci, R. (2025). Fire-image-densenet (fidn) for predicting wildfire burnt area using remote sensing data. *Computers & Geosciences*, *195*, 105783. https://doi.org/https://doi.org/10.1016/j.cageo.2024.105783

Papadopoulos, G. D., & Pavlidou, F.-N. (2011). A comparative review on wildfire simulators. *IEEE systems Journal*, *5*(2), 233–243. https://doi.org/10.1109/JSYST.2011.2125230

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Pimont, F., Dupuy, J.-L., Linn, R., & Dupont, S. (2009). Validation of firetec wind-flows over a canopy and a fuel-break. *International Journal of Wildland Fire*, *18*(7), 775–790. https://doi.org/10.1071/WF07130

Rodrigues, M., Camprubí, À. C., Balaguer-Romano, R., Megía, C. J. C., Castañares, F., Ruffault, J., Fernandes, P. M., & De Dios, V. R. (2022). Drivers and implications of the extreme 2022 wildfire season in Southwest Europe. *The Science of The Total Environment*, *859*, 160320. https://doi.org/10.1016/j.scitotenv.2022.160320

Rollins, M. G. (2009). Landfire: A nationally consistent vegetation, wildland fire, and fuel assessment. *International Journal of Wildland Fire*, *18*(3), 235–249. https://doi.org/10.1071/WF08088

Rothermel, R. C. (1972). *A mathematical model for predicting fire spread in wildland fuels* (tech. rep. No. INT-115). U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. Ogden, UT.

Rui, X., Hui, S., Yu, X., Wu, B., & Zhang, G. (2018). Forest fire spread simulation algorithm based on cellular automata. *Natural Hazards*, *91*(1), 309–319. https://doi.org/10.1007/s11069-017-3127-5

Singh, H., Ang, L.-M., Lewis, T., Paudyal, D., Acuna, M., Srivastava, P. K., & Srivastava, S. K. (2024). Trending and emerging prospects of physics-based and ML-based wildfire spread models: A comprehensive review. *Journal of Forestry Research*, *35*(1), 135. https://doi.org/10.1007/s11676-024-01783-x

Singla, S., Diao, T., Mukhopadhyay, A., Eldawy, A., Shachter, R., & Kochenderfer, M. (2020). WildfireDB: A Spatio-Temporal Dataset Combining Wildfire Occurrence with Relevant Covariates. *AI for Earth Sciences Workshop at NeurIPS 2020*.

Smith, D. M. (2017). Sustainability and wildland fire: The origins of forest service wildland fire research. *FS-1085. Washington, DC: US Department of Agriculture, Forest Service. 120 p.*

Srivastava, A., Wu, J. Q., Elliot, W. J., Brooks, E. S., & Flanagan, D. C. (2018). A simulation study to estimate effects of wildfire and forest management on hydrology and sediment in a forested watershed, northwestern us. *Transactions of the ASABE*, *61*(5), 1579–1601. https://doi.org/10.13031/trans.12326

Stepanov, A., & Smith, J. (2012). Modeling wildfire propagation with delaunay triangulation and shortest path algorithms. *European Journal of Operational Research*, *218*(3), 775–788. https://doi.org/10.1016/j.ejor.2011.11.031

Subramanian, S., & Crowley, M. (2018). Using spatial reinforcement learning to build forest wildfire dynamics models from satellite images. *Frontiers in ICT*, *5*(APR). https://doi.org/10.3389/fict.2018.00006

Sun, J., Qi, W., Huang, Y., Xu, C., & Yang, W. (2023). Facing the wildfire spread risk challenge: Where are we now and where are we going? *Fire*, *6*(6), 228. https://doi.org/10.3390/fire6060228

Sun, L., Xu, C., He, Y., Zhao, Y., Xu, Y., Rui, X., & Xu, H. (2021). Adaptive forest fire spread simulation algorithm based on cellular automata. *Forests*, *12*(11). https://doi.org/10.3390/f12111431

Tymstra, C., Bryce, R., Wotton, B., Taylor, S., Armitage, O., et al. (2010). Development and structure of prometheus: The canadian wildland fire growth simulation model. *Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Information Report NOR-X-417.(Edmonton, AB)*.

Tymstra, C., Stocks, B. J., Cai, X., & Flannigan, M. D. (2020). Wildfire management in canada: Review, challenges and opportunities. *Progress in Disaster Science*, *5*, 100045. https://doi.org/10.1016/j.pdisas.2019.100045

Vahidnia, M., Alesheikh, A., Behzadi, S., & Salehi, S. (2013). Modeling the spread of spatio-temporal phenomena through the incorporation of anfis and genetically controlled cellular automata: A case study on forest fire. *International Journal of Digital Earth*, *6*(1), 51–75. https://doi.org/10.1080/17538947.2011.603366

Vanella, M., McGrattan, K., McDermott, R., Forney, G., Mell, W., Gissi, E., & Fiorucci, P. (2021). A multi-fidelity framework for wildland fire behavior simulations over complex terrain. *Atmosphere*, *12*(2). https://doi.org/10.3390/atmos12020273

Wang, D., Guan, D., Zhu, S., Mac Kinnon, M., Geng, G., Zhang, Q., Zheng, H., Lei, T., Shao, S., Gong, P., & Davis, S. J. (2020). Economic footprint of California wildfires in 2018. *Nature Sustainability*, *4*(3), 252–260. https://doi.org/10.1038/s41893-020-00646-7

Wood, D. (2021). Prediction and data mining of burned areas of forest fires: Optimized data matching and mining algorithm provides valuable insight. *Artificial Intelligence in Agriculture*, *5*, 24–42. https://doi.org/10.1016/j.aiia.2021.01.004

Wu, Z., Wang, B., Li, M., Tian, Y., Quan, Y., & Liu, J. (2022). Simulation of forest fire spread based on artificial intelligence. *Ecological Indicators*, *136*. https://doi.org/10.1016/j.ecolind.2022.108653

Xu, Y., Li, D., Ma, H., Lin, R., & Zhang, F. (2022). Modeling forest fire spread using machine learning-based cellular automata in a gis environment. *Forests*, *13*(12). https://doi.org/10.3390/f13121974

Yassemi, S., Dragićević, S., & Schmidt, M. (2008). Design and implementation of an integrated gis-based cellular automata model to characterize forest fire behaviour. *Ecological Modelling*, *210*(1-2), 71–84. https://doi.org/10.1016/j.ecolmodel.2007.07.020

Yongzhong, Z., Feng, Z.-., Tao, H., Liyu, W., Kegong, L., & Xin, D. Simulating wildfire spreading processes in a spatially heterogeneous landscapes using an improved cellular automaton model. In: *5*. 2004, 3371–3374. https://www.scopus.com/inward/record.uri?eid=2-s2.0-15944364403&partnerID=40&md5=8e782a41c960f34958d422d76825b2a6

Yoo, M., & Wikle, C. (2022, October). *A bayesian spatio-temporal level set dynamic model and application to fire front propagation*. https://doi.org/10.48550/arXiv.2210.14978

Zheng, Z., Huang, W., Li, S., & Zeng, Y. (2017). Forest fire spread simulating model using cellular automaton with extreme learning machine. *Ecological Modelling*, *348*, 33–43. https://doi.org/10.1016/j.ecolmodel.2016.12.022

Zhou, T., Ding, L., Ji, J., Yu, L., & Wang, Z. (2020). Combined estimation of fire perimeters and fuel adjustment factors in farsite for forecasting wildland fire propagation. *Fire Safety Journal*, *116*. https://doi.org/10.1016/j.firesaf.2020.103167

A

Structure of the initial literature review process

The literature review will be structured as follows. First, an overview of relevant key words will be created inspired by three papers provided as recommended readings (Bottero et al., 2020; Finney, 1998; Yoo & Wikle, 2022). This overview of the keywords can be found in Table A.1. Secondly, search strings will be derived from these key words. These will be created by using ”AND” between categories and ”OR” within categories shown in the table. Afterwards Scopus will be searched using this search string. Scopus was chosen as it has a broad and interdisciplinary focus. This is suitable for exploring wildfire propagation simulation models as wildfire modeling is used in many fields. As the goal of the literature review is to give an overview of wildfire simulation models and wildfires have a high societal relevance, it is expected that the search strings will return many papers. To keep the literature review to a manageable scope, several criteria will be used to filter the results. This is visualized in a PRISMA flow diagram in Figure A.1. First, the search will be limited to English articles and conference papers. Secondly, only papers with more than 10 citations are included. These papers are expected to be more relevant as it is expected that designs of wildfire models that are used in practice will be cited frequently. While this does create a bias towards older papers and newer papers might be missed, it is assumed that the newer paradigms will be discussed sufficiently within the threshold to be included. Thirdly, the papers that are found will be subjected to a relevance check. Papers discussing the creation of a wildfire spreading simulation will be included. Papers discussing wildfire risk assessments, gathering data for simulations and forest fire detection will not be included. Finally, the three initial papers were added as core literature.

The results of each of the steps of the literature review process can be found in Figure A.1. The literature review was conducted between the first and third of December 2024. In total 37 relevant papers were included in the literature review.

Fire domain	Simulation	Model type	Design
wildfire	modeling	spread*	design
forest fire	simulation	propagation	creation
bushfire	prediction		algorithm
	digital twin		framework
	Statistical model		model development

Table A.1: Overview of keywords used to create the search string

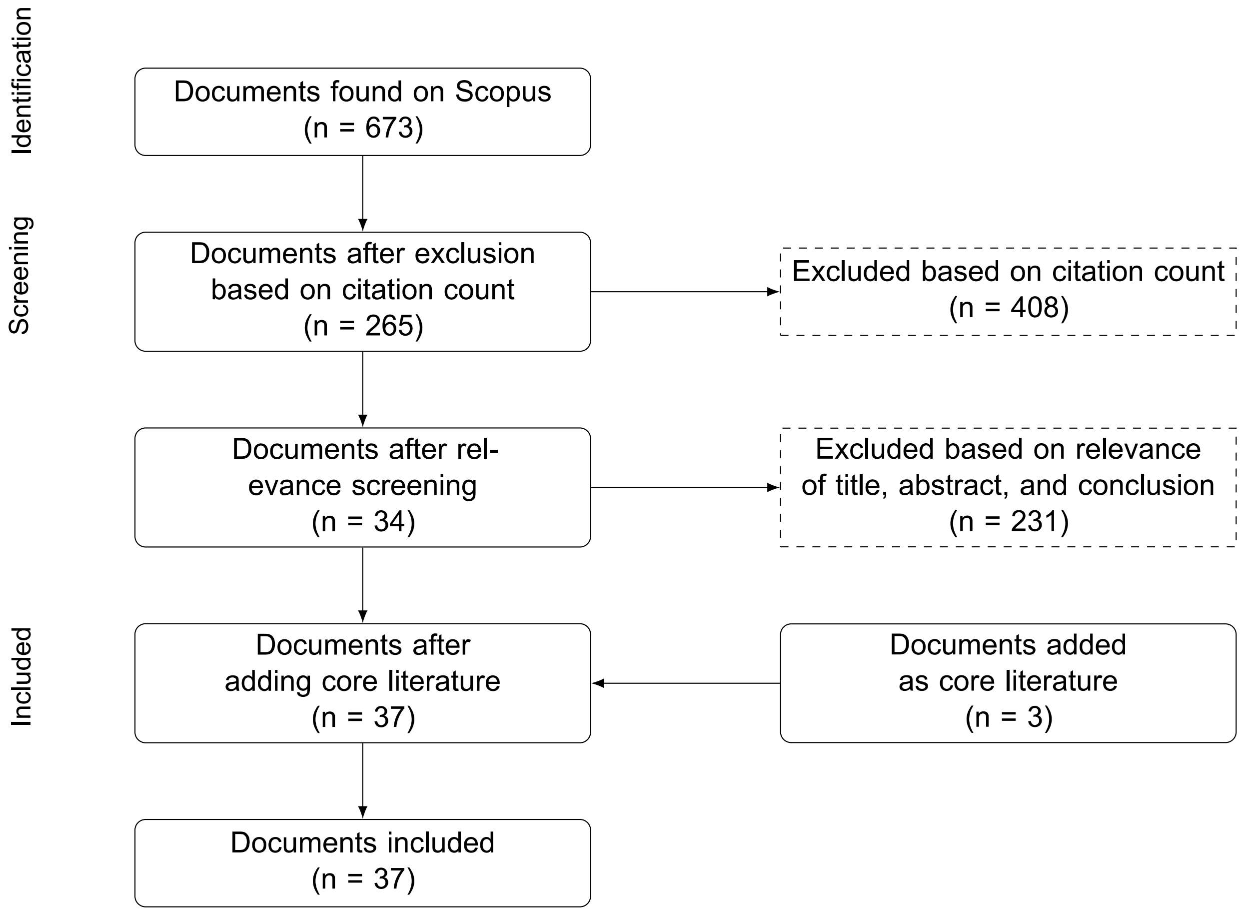

Figure A.1: PRISMA Flow Diagram showcasing the application of selection criteria

B

Coefficient values

Table B.1: Parameters and their coefficient values of the trained spatial-temporal conditional autoregressive model. The values of the parameters are multiplied with their respective coefficients and then summed to derive the log odds of a cell catching fire in the initial time step.

Parameter	**Coefficient value**
Spatial dependence ρ	0.748
Temporal dependence r	0.751
Elevation	-0.023
Wind direction	0.175
Wind velocity	0.401
Minimum temperature	-0.535
Maximum temperature	-0.520
Humidity	0.411
Precipitation	18.120
Drought	0.486
Vegetation	0.299
Population density	-1.088
Energy release component	-0.201
Previous fire mask	39.168