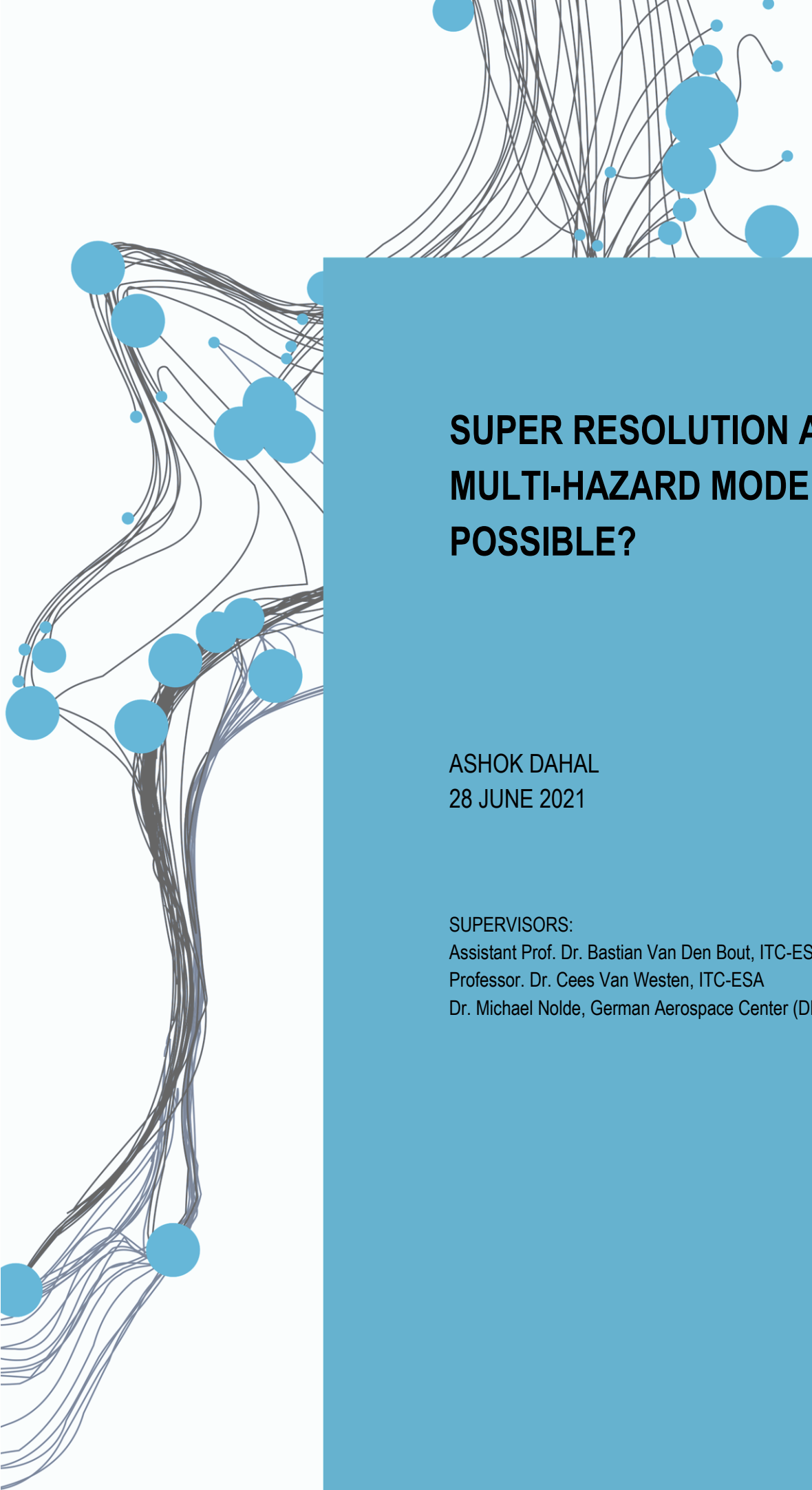
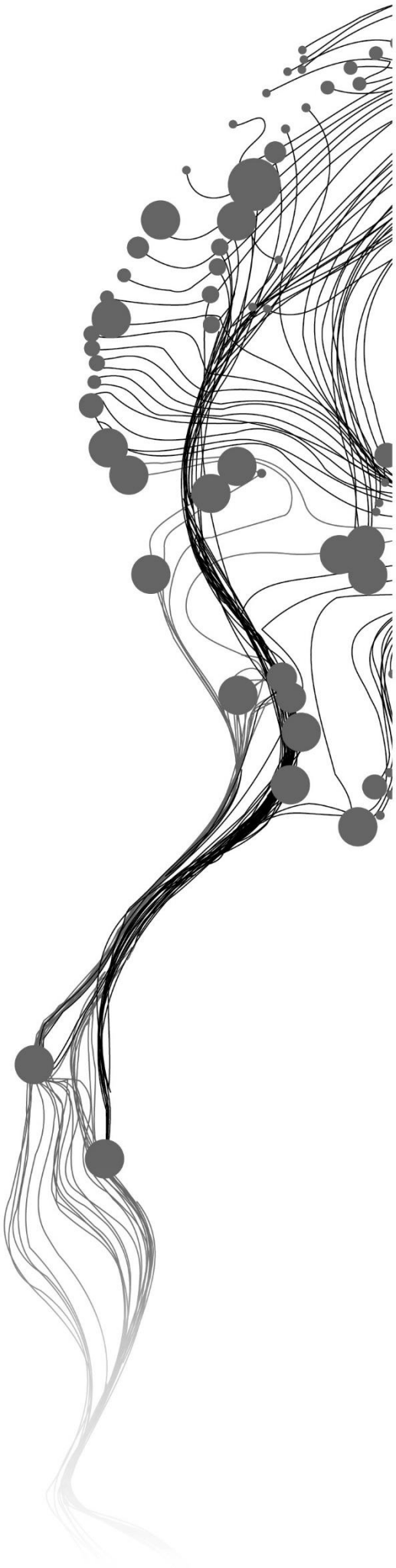


SUPER RESOLUTION AIDED MULTI-HAZARD MODELING: IS IT POSSIBLE?

ASHOK DAHAL

28 JUNE 2021

SUPERVISORS:
Assistant Prof. Dr. Bastian Van Den Bout, ITC-ESA
Professor. Dr. Cees Van Westen, ITC-ESA
Dr. Michael Nolde, German Aerospace Center (DLR)



SUPER RESOLUTION AIDED MULTI HAZARD MODELING: IS IT POSSIBLE?

ASHOK DAHAL
ENSCHEDE, THE NETHERLANDS, JUNE 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialisation: Natural Hazards and Disaster Risk Reduction

SUPERVISORS:
Assistant Prof. Dr Bastian Van Den Bout, ITC-ESA
Professor. Dr. Cees Van Westen, ITC-ESA
Dr Michael Nolde, German Aerospace Center (DLR)

THESIS ASSESSMENT BOARD:
Professor Dr Victor Jetten (Chair)
Dr. A.C. Seijmonsbergen (External Examiner, Universiteit van Amsterdam)

ABSTRACT

The losses due to natural hazards are very high and show an increasing trend due to climate change; human and economic growth; and unplanned development. The risk due to those hazards can be reduced using multi-hazard risk assessment using hazard, the element at risk and vulnerability data. However, due to the lack of good quality and high-resolution data in developing nations, modelling hazards at infrastructure level is difficult. Deep learning-based Super-Resolution can be a solution to increase the spatial resolution of freely available global datasets. However, no studies exist that produced high-resolution output from globally available DEM data using Super-Resolution to improve the quality of physically based hazard modelling. Furthermore, due to differences in data collection sources and value ranges in DEMs, they cannot be compared in absolute values, and there is a lack of techniques to evaluate the improvement done with Super-Resolution in geoscientific data. Moreover, none of the existing research has trained the Super-Resolution models in one region and applied them to another region.

To address these problems, our research aimed to increase the applicability of physically based models in data-poor regions by improving the spatial resolution of globally available datasets by using deep learning-based Super-Resolution. To fulfil our objectives, we selected to work on Digital Elevation Models as the target variable due to its importance in hazard modelling and global availability. We used the two of the most advanced Super-Resolution models (EBRN and ESRGAN), each from different groups of deep learning architecture. These models were trained extensively using high-resolution LiDAR DEM data from Austria. After proving that they perform better than most used interpolation techniques such as bicubic interpolation in the study areas in Austria, they were applied in globally available free datasets in Colombia and Dominica. Furthermore, novel loss function and evaluation metrics were developed to train and evaluate the results focusing on improving DEM data. Furthermore, physically based modelling was used to evaluate the impact of Super-Resolution in multi-hazard modelling. We used 21 different scenarios to test the applicability of Super-Resolution compared to existing interpolation techniques and global commercial data. Each scenario was calibrated for 20 iterations (total 420 iterations, ~5460 CPU hours) in Microsoft Azure, which is the first time that OpenLISEM was used in a cloud computing environment. The results were evaluated in terms of the modelled extent of hazardous processes, the height of flow, and the time of solid and fluid flow to prove the applicability of the Super-Resolution approach.

The analysis shows that the use of global DEM data with Super-Resolution processing was able to increase the accuracy of hazard modelling output as compared to DEMs made with existing interpolation techniques. Furthermore, when evaluating derivative DEM products through visual analysis, it is observed that the Super-Resolution has increased the crispness of valley lines and ridgelines in the DEM datasets. However, the specific topographic features that are not present in low-resolution data could not be reconstructed using the Super-Resolution, limiting its use in geomorphological mapping. The applicability of Super-Resolution was tested in multiple locations, and it could prove that the technique resulted in 8-25% improvement in all of the study sites. The results also show that the capacity of both models (EBRN and ESRGAN) is generally very similar. There are a few challenges in calibration, such as the use of Gradient Descent requiring more iterations and the lack of datasets and metrics to compare our results with existing Super-Resolution models. Furthermore, we could not compare our results on multi-hazard modelling to other research because there is no published work using the Super-Resolution in multi-hazard modelling.

Keywords: Super-Resolution, Deep Learning, Multi-Hazard Modelling, Digital Elevation Models

ACKNOWLEDGEMENTS

The more we learn, the more we discover how much we do not know – Yoda, Star Wars

Acknowledgement is a very difficult section to write because the butterfly effect suggests that any tiny change in the condition, a million years ago, I would not have been doing this research or even existed. So, I am thankful to all of those living and non-living things that had indirectly contributed to my thesis.

As of direct contribution, I would like to extend my sincere gratitude to Dr Bastian van den Bout for his initial suggestion and exciting information about this research topic and his guidance throughout the research process. His understanding of broad aspects of mathematics, programming and physically-based modelling has inspired me to conduct this thesis and my future work. Furthermore, his guidance and comments on my research were extremely helpful for me to understand the concepts of Super-Resolution and multi-hazard modelling.

I am extremely thankful to Prof.Dr. Cees van Westen, for his support and guidance throughout the research and his contribution to my study at ITC. His suggestion to join ITC instead of other University and suggestions throughout my MSc research has changed my career and research in a better direction. Throughout the research, Professor Cees always taught me to think critically rather than just reporting my results, and his politeness and encouragement have added energy to my research. Without his guidance and encouragement, I could not have completed this research. Furthermore, I would also like to thank Dr Michael Nolde from German Aerospace Center for his genuine suggestions and support during my research as an advisor. His comments on my research and support to obtain the globally available commercial DEM for research purpose have added more value to my research. As an advisor, Dr Michael has supported my research in every way he could, and it is my pleasure to have him in my research as an advisor.

I want to express my sincere thankfulness to Prof. Cees van Westen and Drs. Nanette Kingma, for their geomorphological evaluation, which added more perspective in my thesis in terms of the geomorphological mapping using Super-Resolution. Furthermore, their evaluations and comments were extremely helpful to understand the limitations of this research. Furthermore, I would like to thank Ms Maria Isabel Arango Carmona for her support in providing the dataset for Mocoa, Colombia. Her support on data and local information about Colombia has helped extremely to apply Super-Resolution in Colombia.

I would like to show my appreciation to Dr Manzul Kumar Hazarika and Ms Sudchai Nikaset (Khun Lek) from the Asian Institute of Technology for their financial and moral support for me to join ITC, their support and encouragement made it possible for me to join the ITC and constant guidance from Dr Manzul in ethical research and use of science to create a better world has encouraged me to work on the field of natural hazards. If it were not for Dr Manzul, being a Geomatics Engineer and software developer, I would have never thought of using my knowledge to solve real-world problems.

Last but not the least, I would like to thank my parents and sisters for their encouragement and support. Without their encouragement, I would not have been able to conduct the research or even study at the master's level. Furthermore, I would like to appreciate the efforts Ms Pratichhya Sharma had put into my thesis through encouragement and support when I was procrastinating. Moreover, I would not have completed the thesis without her moral support and guidance during the setbacks. Finally, I would like to show my genuine gratitude to my International and all of the Nepalese friends at ITC for their moral support during the thesis research, and their moral support has encouraged me to achieve better in my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

CNN	Convolution Neural Network
DEM	Digital Elevation Models
D-SRGAN	DEM- Super-Resolution Generative Adversarial Networks
ESA	European Space Agency
ESRGAN	Enhanced Super-Resolution Generative Adversarial Networks
GAN	Generative Adversarial Networks
GD	Gradient Descent
HR	High-Resolution
LR	Low Resolution
NASA	National Aeronautics and Space Administration
PSNR	Peak Signal to Noise Ratio
SR	Super-Resolution
SR. HR.	Super-Resolution. High-Resolution
SRGAN	Super-Resolution Generative Adversarial Networks
SRTM	Shuttle Radar Topography Mission
SISR	Single Image Super-Resolution
SSIM	Structural Similarity
USGS	United States Geological Survey

1. INTRODUCTION

1.1. Background

The earth lost almost three trillion US dollars in damage by natural disasters to infrastructure from 1980 to 2020, of which 75% is evoked by hydrometeorological surface hazards such as floods, storm surge, landslides, debris flow, etc. (The World Bank, 2020a). The loss due to natural hazards has increased from 1.63 to 2.97 trillion dollars (adjusted to inflation and economic factors) from the last two decades of the 20th century to the first to decades of the 21st century and is expected to increase more because of climate change, unplanned infrastructure, economic and population growth (CRED & UNDRR, 2020). Jongman et al. (2012) estimate that by 2050 the exposure due 1/100 year flood events will affect more than 1.3 billion people and 158 trillion US Dollars in infrastructure compared to 46 trillion US Dollars in 2010.

To reduce such impact of natural hazards, it is essential to understand and quantify these impacts on the infrastructure and population (Sanders, 2017). However, there is a gap in data, understanding, and technology in many developing countries to perform risk analysis at the infrastructure level and plan the best alternative for disaster risk reduction due to the lack of funds, capacity, and resources (Rahman & Fang, 2019). The impact of natural hazards can only be reduced if we understand their characteristics, such as their location, intensity, frequency, and their impact on the element at risk. Mapping, monitoring, and modelling are the most important aspects to understand the key spatial and temporal characteristics such as location, intensity, and frequency of natural hazards. Furthermore, mapping and monitoring helps to understand the past and current situation, but for understanding the future scenario and planning disaster risk reduction alternatives, modelling natural hazards is necessary.

Numerical modelling of Multi-Hazards and risk is the process of numerically simulating the different characteristics of natural hazards either on a physical or data-driven basis to obtain information such as the probability of occurrence, intensity, and potential consequences of natural hazards in terms of hazard interaction as well as its impact (Komendantova et al., 2014; Thompson & Warmink, 2016). Modelling of natural hazards is a data-intense and computationally complex task. Physics-based simulations are best suited for the local and infrastructure planning level because of their relatively lower dependency on past event data and capacity to deterministically simulate real-world scenarios (Briggs, 2016, p. 153). For this research, we primarily consider physically based multi-hazard modelling and do not discuss single hazard modelling approaches because multi-hazard modelling combines multiple physical processes and their interaction and better represents the real-world scenarios. In the physically based multi-hazard analysis, the interaction between various hazards is simulated using the governing physical equations such as the two-phase flow equations, which work with the principle of conservation of mass, energy, and momentum (van den Bout et al., 2018). Liu et al. (2016) categorise those interactions of hazards as an independent, mutex, parallel, and series. Physically based multi-hazard modelling is a data-intensive process and requires ample data to run the model. Most of the resources involved in multi-hazard modelling are related to the data collection model calibration and validation process.

In many developing countries, high-resolution data required for multi-hazard modelling are not available due to the lack of financial and skilled human resources (Hawker et al., 2018). However, there are global datasets that are freely available. However, their use in multi-hazard modelling is limited because of (I) low

spatial resolution (typically 30 – 10000 meters), (II) high level of uncertainty in the quality because of varying noise and error depending on the local geographical characteristics, (III) often not available in continuous time series (IV) mostly generated using proxy measurements instead of direct measurement (Li et al., 2019; Osuteye et al., 2017). A coarse-resolution dataset can be used for modelling processes on regional scales, but to analyse the impact of hazard at the local or infrastructure level, higher resolution data is required (Lindersson et al., 2020). Some processes such as gully formation or landslides that have a higher impact on agriculture happen on a large cartographic scale and are not possible to model physically using a global dataset. In those cases, only susceptibility assessment is possible (Dewitte et al., 2015).

To put things into more perspective, a hillshade image of the Caribbean Island Dominica is shown in Figure 1. In Figure 1, we can observe that even though both DEMs are resampled to 10 meters resolution, the hillshade image from a low-resolution dataset is very blurred, and, during the modelling, uncertainties will increase due to the coarse representation of the terrain with such low-quality data. For small size infrastructure, it can only provide coarse level information, leading to overestimated or underestimated risk in many cases because of the uncertainties in hazard intensity. Therefore, for high-resolution multi-hazard modelling for infrastructure level risk assessment, the high-resolution data must be collected at a local scale, or the quality and resolution of the freely available global/local dataset must be increased using some techniques.

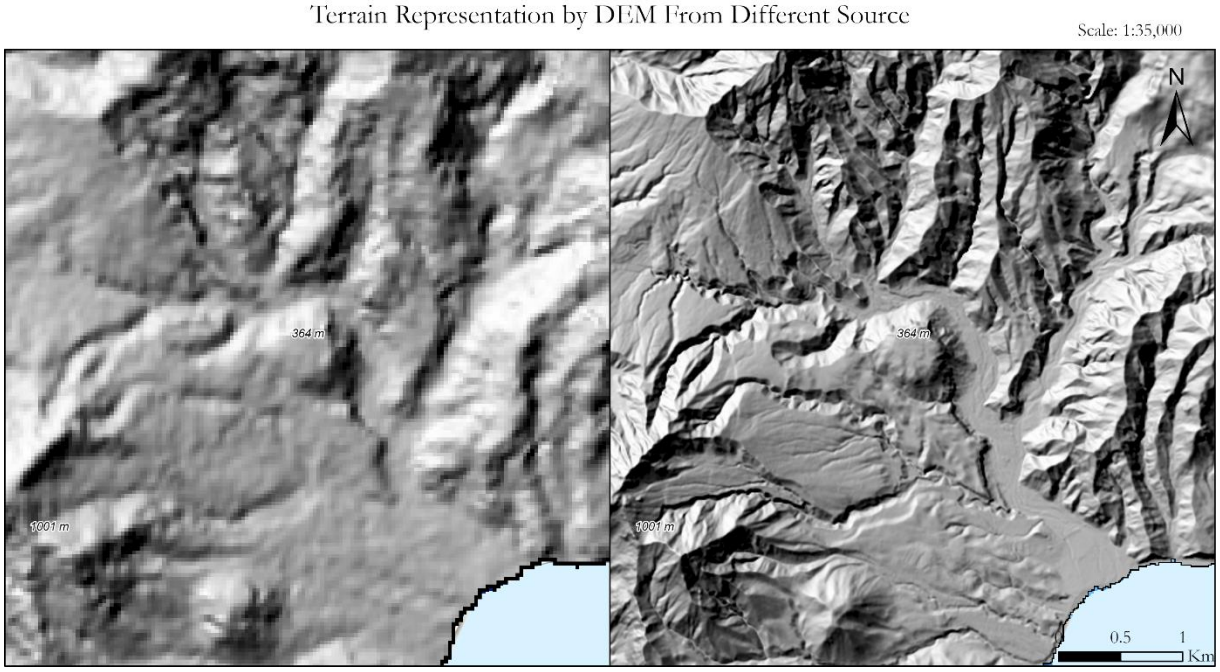


Figure 1: Terrain representation with data from low-resolution SRTM DEM (left) and high-resolution LiDAR DEM (right), both are resampled to 10 meters. Basemap sources: ESRI, HERE

High-resolution data collection is an expensive procedure, and often the costs cannot be covered by the local authorities without external funding. Another approach of local-level multi-hazard modelling with a cost-efficient approach can be by improving the resolution of the existing global data (GEOSENSENL, 2018). In the current stage, most of the resolution improvement is done using a technique called *Super-Resolution*. Super-resolution is the process in which higher resolution data (mostly raster images) are generated based on lower resolution data inputs using a specialised algorithm (Yang & Huang, 2017). Application of Super-Resolution in an ideal scenario where all technological constraints, implementation constraints, and scientific constraints are perfectly in favour of Super-Resolution can reduce the cost of

high-resolution data collection leading to cheaper disaster risk mitigation. However, those constraints might lead to a reduction of the applicability of the Super-Resolution.

Before the advancement of deep learning in the 2010s, Super-Resolution was mostly achieved using statistical and mathematical techniques such as Interpolation, Maximum Likelihood, Maximum a Posteriori, Joint MAP restoration, Bayesian Treatments (Yang & Huang, 2017). For the Super-Resolution of the geospatial data, Interpolation and geostatistical approaches such as kriging, co-kriging, regression kriging were used widely (Bhunia et al., 2018; Rata et al., 2020). Recent advancement in computer vision and deep learning has worked towards a new concept of Super-Resolution based on deep learning. Most of the emerging algorithms and approaches use convolution neural networks and Generative Adversarial Networks (GAN) for Super-Resolution (Al-falluji et al., 2017; Ji et al., 2020; Ledig et al., 2017; Luo et al., 2017; Xu et al., 2019). The geostatistical approaches such as the kriging can produce smooth data which might be accurate but cannot generate spatial variability, and deeper and faster CNN and GAN-based approaches are trying to overcome such problems (Al-falluji et al., 2017).

1.2. Knowledge Gap

Most of the deep learning-based Super-Resolution models focus predominantly on improving the resolution of multimedia images and videos and do not focus on the Super-Resolution of the geospatial data (Al-falluji et al., 2017; Ledig et al., 2017; X. Wang et al., 2018). In the case of multimedia Super-Resolution, perception is more important, and if the produced image looks better and more real than the input image, the results are considered good (Chu et al., 2020). However, in geospatial data, the Super-Resolution should represent the actual values and spatial variability of specific measurements on the earth's surface. Only a few research efforts have focused on using Super-Resolution in the geospatial field. One of the interesting examples of Super-Resolution is to create a higher resolution Sentinel-2 image from the original resolution (up to 6 times upscaling) (Lanaras et al., 2018). Even though the application with Sentinel-2 image was with satellite image data, the improvement was done in the reflectance values rather than digital numbers, which makes it challenging than image Super-Resolution (Lanaras et al., 2018). Unlike image fusion technology, where images from many sensors are fused to create a new image, the single image Super-Resolution is used to improve the resolution of an image without the addition of other data after the deep learning model is trained with high-resolution data and its low-resolution counterpart (Zhong et al., 2016).

Most of the Super-Resolution research is done by smoothing original high-resolution data using interpolation and then downscaling it using the Super-Resolution. However, none of the works available to the author have used the approach to improve the quality of existing low-resolution data. Theoretically, training on an available high-resolution and its counterpart low-resolution should generalise the model so that it can perform well in the case of Super-Resolution of another set of low-resolution data. However, there are no existing evaluation matrices and use cases for the use of Super-Resolution models trained in high-resolution data and inference in low-resolution data. Moreover, in our case, if the high-resolution dataset is already available for specific locations, there is no need to convert global free datasets to a higher resolution. Furthermore, the current state-of-the-art Super-Resolution models are developed for photo and video processing, but there are very few studies on using them for downscaling geospatial datasets. The deep learning architectures designed for the multimedia Super-Resolution may not perform well in the case of the geospatial data and might need modification at the architecture level.

Even though studies on Super-Resolution of DEMs, satellite images, landcover maps, etc. have implemented the Super-Resolution for geospatial data (Demiray et al., 2020; Jia et al., 2019; Lanaras et al.,

2018; Leong & Horgan, 2020), there has not been any study in which the downscaled data were used for physically based multi-hazard modelling. Furthermore, those research only focus on one location to observe the improvement, but they have not studied whether a model trained at one specific location can be used in another location to improve the spatial resolution or not. In the climate model downscaling, Super-Resolution is used for downscaling the climate variables (Vandal et al., 2018). However, most of the downscaling is done on the product of the model and not on the input data, and it is also difficult to claim that the downscaling of climate models have enough spatial variability because those models mostly cover larger areas and do not depict the high spatial variability of earth surface properties such as elevation and soil depth. So, in the current stage, it is unknown whether the Super-Resolution models are suitable for downscaling global datasets as better input data for physically based multi-hazard modelling. Demiray et al. (2020) suggest that the output DEM from their algorithm works better in flatter terrain where it adds spatial variability but performs worse in the mountainous regions. Furthermore, Demiray et al. (2020) also suggested that the improvement in the model architecture and training dataset might make it possible to add spatial variability in the mountainous terrain as well. On the other hand, this problem might be solved by adding functions to minimise the slope error on the training process of GAN models. Further research is required to understand if the output from the Super-Resolution model can be further improved by calculating and minimising the derivative during the training process or not.

1.3. Objective

This study aims to increase the applicability of physically based multi-hazard modelling of hydrometeorological hazards in data-poor regions by increasing the spatial resolution of input datasets using deep learning-based Super-Resolution techniques. Based on this major objective, the following sub-objectives are identified:

1. To design and develop (or adopt) Super-Resolution models and their components using the deep learning technology for downscaling global datasets to a higher resolution.
2. To train, calibrate, and test the Super-Resolution model with a high-resolution dataset and its low-resolution counterpart.
3. To define evaluation metrics and evaluate the quality of Super-Resolution output from the global dataset relative to the measured high-resolution dataset.
4. To analyse the applicability of the Super-Resolution in multi-hazard modelling by comparing the results produced by the OpenLISEM hazard model using the downscaled DEM data with results obtained from lower resolution data.

Based on the above research objectives and sub-objectives, the following research questions are defined:

1. To what extent are the Super-Resolution models capable of downscaling geospatial data with higher accuracy and spatial variability than the interpolation approach? (Objective: 1)
2. What architecture of the Super-Resolution models is best for the data downscaling, and which modifications will create a more robust model for optimal downscaling of the datasets in question? (Objective: 1)
3. How to optimally minimise the training loss functions, validate, and generalise the Super-Resolution models for downscaling the global dataset for multi-hazard modelling? (Objective: 2)
4. Is the Super-Resolution model capable of generating a better dataset than existing interpolation techniques in terms of peak signal to noise ratio and structural similarity? (Objective: 2)
5. How the different architectures and complexity levels of deep learning models perform in the Super-Resolution of geoscientific data with different sources? (Objective: 2)
6. How to compare the quality of Super-Resolution output from the global dataset and their derivatives without any available high-resolution counterpart of the data from the same sensor? (Objective: 3)

7. Is the output from the Super-Resolution model for the global dataset comparable to that of the actual higher resolution data in terms of the evaluation metrics defined above? (Objective: 3)
8. Does the OpenLISEM hazard model produce similar output from the downscaled data as compared to outputs based on higher resolution input data? If not, to what extent the downscaled dataset influences the model accuracy? (Objective: 4)
9. Based on the quantitative assessment of the outputs from multi-hazard modelling, is it possible to use the Super-Resolution techniques in data scare regions? (Objective: 4)

1.4. Research Overview and Thesis Structure

Figure 2 shows the overall steps followed in the research and the research questions answered by those steps. Each step is crucial towards answering the research questions, and some of those steps are qualitative such as literature review. Steps 2,3 and 4 are the Super-Resolution model development and training works; steps 5 and 6 are quantitative analysis steps.

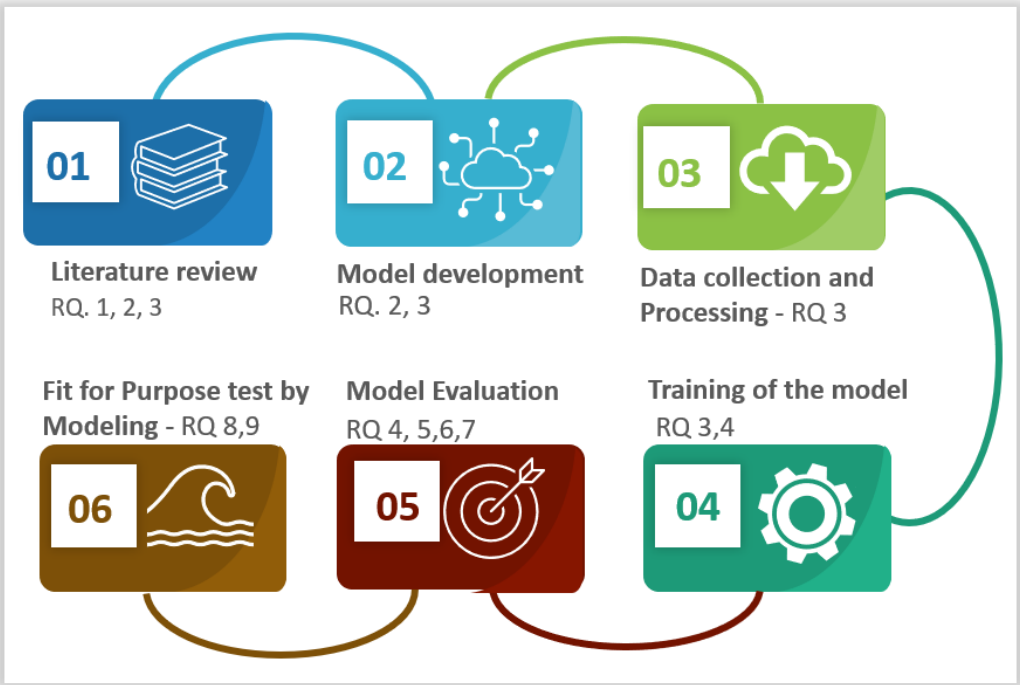


Figure 2: Research Design shows the overall steps that are followed in the research.

The overall thesis is designed in a conventional way and consists of six chapters. The first chapters provide a brief background on the research and its social aspects, followed by a literature review chapter which provides a basic understanding of the literature and published work in the field of Super-Resolution and rationale to select specific models and data. The third chapter elaborates on data collected, detailed design of the models and their input, and modelling and calibration details. The results and analysis chapter (chapter 4) reports results obtained from the research and its analysis compared with the existing literature where possible, followed by a discussion chapter explaining the reason behind such results. The last chapter is the conclusion and discussion chapter, which will further conclude the whole research relating to the original research questions.

2. LITERATURE REVIEW

This chapter will provide background information on the major concepts required to understand the thesis, a literature review on the latest developments in the area, and a rationale for selecting specific data, algorithms, and basic background on multi-hazard modelling.

2.1. Super-Resolution

Super-Resolution emerged from the medical and nuclear imaging field to define an image enhancement process where the spatial or spectral resolution of a low-resolution image is synthetically increased to higher resolution using any mathematical, statistical, or deep learning approaches. The term was first used by Inouye (1964) to define the process of increasing the resolution gamma-ray spectrum using Fourier transformation. Bannore (2009, p. 1) defines Super-Resolution as "an image reconstruction process in which the higher resolution image (in terms of spatial resolution) is obtained by the fusion of partial information contained within the low-resolution images". This is a more generic and broad definition; in the case of single image Super-Resolution, the partial information is obtained from a single image, not multiple images making the problem more challenging to solve. The Super-Resolution aims to increase the spatial resolution of a lower resolution image, caused by different image degradation factors such as camera motion, camera optics, atmospheric interference, and insufficient sampling (Bannore, 2009, p. 2). In terms of research done in the field of Super-Resolution, the Scopus (2021) database shows the following number of documents published with the keyword "Super-Resolution" from 1964 to April 2021, as represented by Figure 3. It can be observed from Figure 3 that after the advancement of computation capacity and availability at around 2000 A.D, the research in Super-Resolution has increased exponentially. With the implementation of deep learning techniques in Super-Resolution from 2015, the number of published works in Super-Resolution increased with very high rate.

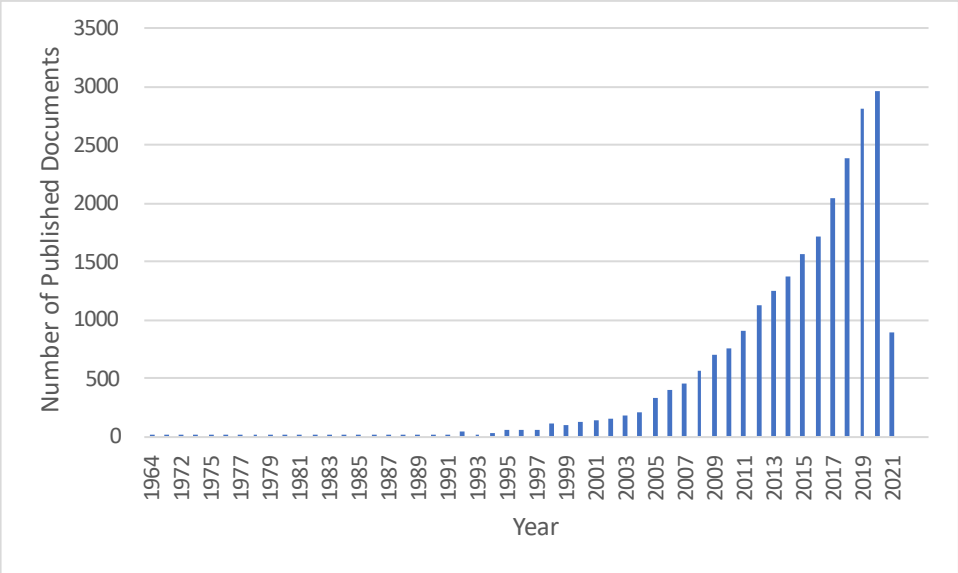


Figure 3: Number of published documents with the keyword Super-Resolution from 1964 to April 2021.

Even though Super-Resolution can be applied to different datasets such as image, signal, video, etc., our major focus is on the Single Image Super-Resolution (SISR). The SISR is the image restoration problem where a higher resolution (HR) image is recovered from a degraded lower resolution image (LR) using some restoration process (Soh et al., 2019). In this branch of Super-Resolution, multiple images of the same features/scene are not available to fuse, and only a single is used to generate a new HR image. This makes

the task more challenging because there can be too many possible solutions to HR images from a single LR image (Soh et al., 2019). Our major focus is on using the single image Super-Resolution for geoscientific data, so we will not discuss algorithms which are focused on image fusion, video, and signal Super-Resolution but only discuss the techniques developed for SISR. The SISR Super-Resolution techniques can be further classified as mathematical, regularisation, and learning-based approaches which are further explained in the following sections.

2.1.1. Mathematical Approaches

In the mathematical approach, Super-Resolution (SR) images are obtained using a standard function of low-resolution images, which can be represented as SR=f(LR), where function f can be any mathematical function that can estimate the data with higher spatial resolution. The mathematical approach can be further classified as a Non-uniform interpolation approach and frequency domain approach (Park et al., 2003). The non-uniform interpolation approach directly reconstructs the SR image from the LR image by interpolating the grid values near the desired location. Major algorithms for non-uniform interpolation approaches are listed as below:

1. *Nearest Neighbour Interpolation*: the desired location takes the value of the nearest available data in the Euclidian domain even though it is a fast approach, generated data have very low accuracy (Han, 2013).
2. *Bilinear interpolation*: the four nearest pixels from LR in Euclidian distance are averaged to estimate the value for the desired location in SR. It provides better results than NN but has poor visual quality (Han, 2013).
3. *Bicubic interpolation*: the bicubic interpolation considers 16 nearest pixels in a 4x4 grid from the LR data and calculates their weighted average to estimate the value at SR image. This method is most wildly used and creates relatively smooth results (van Ouwerkerk, 2006).
4. *Bicubic Spline Interpolation*: similar to bicubic interpolation, this approach considers 16 nearest pixels but instead of taking the weighted average of those pixels, it uses a Cubic Hermite spline to generate a continuous function which is then used to estimate the required value at SR. This approach generates very smooth results, which might not be useful for many approaches (van Ouwerkerk, 2006).
5. *Inverse Distance Weighted Interpolation (IDW)*: It estimates the SR values based on the neighbouring locations using the average distance weighted LR values. This approach does not have any limitations on the number of LR data points to consider and can be used in different scenarios with different data points based on the requirements(Shepard, 1968). Furthermore, this approach also does not require the data to be in a grid, making it more applicable in an irregularly spaced dataset.

In the case of frequency domain approaches, most of the algorithms are based on the Fourier transform where the algorithms such as discrete and continuous Fourier transform are used to relate the high and low-resolution images by the aliasing relationship, and SR images are reconstructed from LR images using Fourier transformation (Bannore, 2009, p. 10). This approach is more complex and not used much in geospatial data, so it is not further discussed here, please see Bannore (2009, pp. 9–11) and (Tian & Ma, 2011) for further details.

2.1.2. Regularisation-based Approaches

These approaches mostly use statistical or specific methods to reconstruct the SR from LR (Tian & Ma, 2011). The major algorithms that are used in this approach are Maximum Likelihood, Maximum a Posteriori, Joint MAP restoration, Bayesian Treatments (Yang & Huang, 2017, pp. 11–17), and Kriging (Panagiotopoulou & Anastassopoulos, 2007; Q. Zhang & Wu, 2015). In general, all those algorithms derive

the statistical likelihood from the LR image and inference the values based on statistical equations in the grid of SR images. These methods reconstruct the SR image stochastically from the LR where LR images are considered stochastic variables, and then based on those LR images, a statistical (mostly Bayesian) model is fitted and which is then further used to generate the SR image (Yang & Huang, 2017, pp. 11–13). For example, in ordinary kriging, a semi-variogram can be estimated from the low-resolution data, and then that pattern of the data can be further used to estimate SR values. Kriging is also considered the best local predictor because it tries to preserve the local variance in the data. However, the kriging approach produces the image with smooth results, and the output will lack spatial variability.

2.1.3. Learning-Based Approaches

Learning-based algorithms work in such a way that algorithms learn from one set of LR and HR images about the relationship between LR and HR image and then use that information to inference in newer LR image to generate SR image. With the advancements in machine learning and deep learning, these methods have been extensively researched and implemented. There are many algorithms developed using the deep learning-based approaches for Super-Resolution, and in general, all of them use convolution neural networks (CNN) as layers of the neural network. Based on the network design and its characteristics, Anwar et al. (2020) classified the deep learning-based Super-Resolution techniques as shown in Figure 4. The study of Anwar et al. (2020) is very sophisticated and has outlined and explained all major algorithms at the time of publication. However, Figure 4 can be further simplified by classifying the existing deep learning algorithms as Generative Adversarial Network GAN based models and Non-GAN based models. The reason to classify them as GAN-based and Non-GAN based is that the focus of GAN-based models is to generate images with better visual representation, and that of Non-GAN based is to generate higher PSNR. Furthermore, they are trained in a very different way, and due to the complexity of training and output results from GAN based approach are not usually compared with Non-GAN based approach:

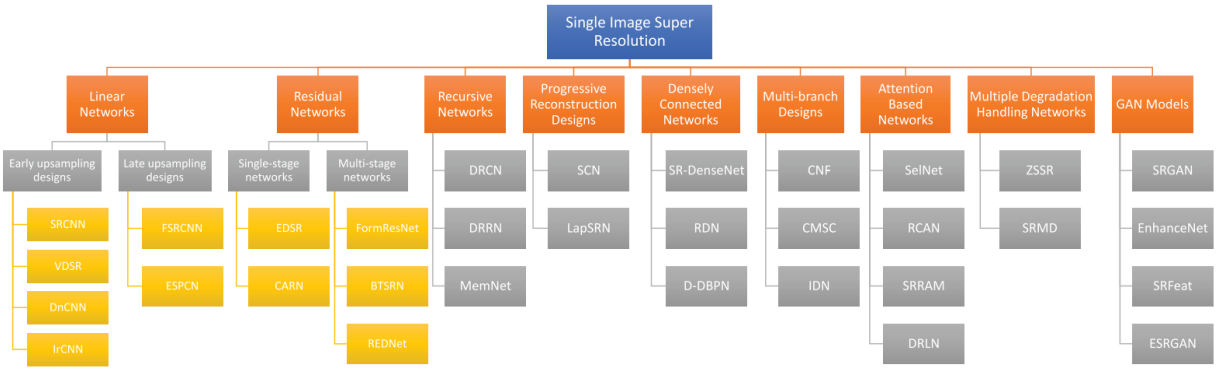


Figure 4: Classification of deep learning-based Super-Resolution algorithms. Source: (Anwar et al., 2020)

2.1.3.1. Generative Adversarial Network (GAN) Based Models

The GAN based model was first developed by Goodfellow et al. (2014); this class of networks features a generator network, which tries to create fake data based on noise (or low-resolution data), and the discriminator network, which tries to identify if the prediction is artificial or real. Furthermore, if the discriminator network identifies the fake data, it calculates the adversarial loss and sends the feedback to both the generator and discriminator. During the training process, the generator network tries to minimise the adversarial and content loss, whereas the discriminator tries to increase the adversarial loss (Demiray et

al., 2020). The generator uses both adversarial and content loss to learn from past errors and generate better results in the next epoch, whereas the discriminator uses the adversarial loss to detect the fake high-resolution data (Ledig et al., 2017). The competition between the generator and the discriminator algorithm continues until the stage at which the discriminator network can no longer identify the fake data (Goodfellow et al., 2014). GANs can be understood as a money counterfeit problem; the Generator model can be considered a counterfeiter who tries to create fake banknotes and discriminator as a bank detective. Initially, the counterfeiter creates fake banknotes that the detective can identify, and then the counterfeiter learns how to create better and better banknotes whereas, the detective learns how to detect them until a convergence point where the counterfeiter produces the banknote so perfectly, the detective can no longer identify the fake notes. This process of learning for both counterfeiter and the detective is the overall idea of GAN training.

The GAN-based approach is infamous for its difficulty in training and famous for its capacity to generate likely images. The first use of Generative Adversarial Network in single-image Super-Resolution application was done by Ledig et al. (2017) with their famous Super-Resolution using GAN (SRGAN) model. The main highlight of the SRGAN model is that (i) multiple loss functions which provide feedback to the network about their performance, were used to generate visually pleasing as well as quantitatively better (high PSNR) results, and (ii) use of generator and discriminator model in the adversarial min-max game to improve the performance of generator model (Ledig et al., 2017). This research is considered as the benchmark in the GAN-based SISR task, and it was further developed by Wang et al. (2018) in their ESRGAN model. The ESRGAN modified the generator network with Residual-In-Residual blocks without any batch normalisation layers and further modified the loss functions to add relativistic loss functions (X. Wang et al., 2018). ESRGAN is the best GAN-based model available in the literature, and it has been used in a multitude of the fields such as medical imaging, natural image processing, etc. (Anwar et al., 2020). However, the GAN-based approach sometimes generates artefacts that seem to appear from nowhere, ESRGAN has tried to reduce those problems by averaging the model, but the problem persists in many GAN-based SR applications (Anwar et al., 2020).

2.1.3.2. Non-GAN Based Models

As shown in Figure 4, there are many deep learning-based SR models with different width, depth, and network architecture, but all of them have one thing in common, they all try to increase the PSNR with their loss function. The early CNN-based SR models used to be sequential models with stacked convolution layers as in the Super-Resolution using CNN (SRCNN) model (Dong et al., 2014). After the development of SRCNN, there have been many models with different aims towards solving the SISR problem in different scales such as 2x, 3x, 4x. Currently, the 2x, 3x, and 4x Super-Resolution have started to reach a bottleneck, but according to one of the most detailed review by Anwar et al. (2020), the Embedded Block Residual Network (EBRN) of Qiu et al. (2019) has performed best with the highest PSNR with fewer layers and parameters. Dong et al. (2014) stated that the major problem in this type of model is that once the model reaches a very high number of parameters, it is very difficult to stabilise the model and train it properly. However, EDSR (Lim et al., 2017)and VDSR (Y. Zhang et al., 2018) models have shown that it is possible to increase the PSNR with very deep networks and with regularisation, it is possible to train such a deep network. However, the major problem is that there is always a trade-off between the computational complexity and improvement in accuracy with such a high number of parameters. Another advancement in SISR is done by the RCAN model of Y. Zhang et al. (2018), where they treat each band separately with "Channel Attention", which proved to have significant improvement in the overall quality of the image, but

in our case, most of the geoscientific data are of single band, and it might not improve the quality as such. EBRN (Qiu et al., 2019) model uses the block residual network that processes different frequencies of data through different complexity levels and concatenates that information at the end of the model. This approach has shown the highest PSNR at the time of publication and also at the time of review by Anwar et al.(2020). This method could be of good use for geoscientific data because it investigates the data frequency and different cases that are very important for geoscientific data, such as Digital Elevation Models (DEM), where different slopes could be treated as different frequency and which might help to produce better SR images. There are many types of networks in this class of networks but only relevant models are reported, for further information please see Anwar et al. (2020), Bashir et al.(2021), and Ooi & Ibrahim (2021).

2.2. Super-Resolution in Geoscientific Data

Since the research aim to select a single dataset and perform the Super-Resolution on that dataset to check its impact on the accuracy of physically based multi-hazard modelling, the dataset was searched to which the physically based modelling is most sensitive. In physically based modelling, Franchini et al. (1996) has concluded that their model in hand (TOPOMODEL) was much sensitive to the grid size (spatial resolution) of the digital elevation models, and Kelleher et al. (2015) has shown that the physically based hydrological modelling was most sensitive to climate factors and then to the topography. In our case, the future climate change scenario will not be considered, and we will only run the standard event, so, based on those observations, DEM is the best data to check the model performance. Furthermore, the research of (2020) has shown that the use of higher resolution data in physically based modelling can provide better representation, which is contradicted by the research of H. Zhang et al. (2016), where they suggested the impact of resolution largely depends on the quality of data and requirement of the model. Considering the availability of DEM data with high- and low-resolution pairs in some countries and the availability of multiple global datasets and importance in different kinds of hazard simulations, we decided to perform our research with the Digital Elevation Models. However, in the case of statistical modelling and susceptibility mapping, the spatial resolution of the terrain dataset does not have much influence, and also, having higher resolution might not always generate better quality (Chang et al., 2019).

In the case of the digital elevation models, NASA used a Super-Resolution technique to improve the spatial resolution of Flash LIDAR in 2011, and the results were published in 2014 (Bulyshev et al., 2014, 2011). NASA used the patented approach for the Super-Resolution purely based on multi-frame matching and mathematical projection (Bulyshev et al., 2011). Liu et al. (2018) used Super-Resolution for the lunar surface reconstruction using the improved sparse representation. Some works on DEM Super-Resolution use the convolution neural network (Moon & Choi, 2016; Xu et al., 2019) and GAN approaches (Demiray et al., 2020; Leong & Horgan, 2020; Shin & Spittle, 2019). For the soil moisture data, there are two published studies on the Super-Resolution of SMAP data (Nguyen et al., 2019), but there is no published work on Super-Resolution for other data such as soil depth.

A limited number of published research in the Super-Resolution of Digital Elevation Models using the Deep Learning techniques is available; most of those research focuses on incrementing image quality. However, none of the studies has tested their applicability in any modelling. Furthermore, most of the results are compared via the raw elevation values than that of the derivatives, but raw elevations are not much used in most of the modelling but having relatively better-quality derivatives is very important. This leads to the need of having evaluation metrics that consider the derivatives and improvement on the derivatives using SR. Furthermore, due to the lack of a standard dataset for comparing the models, the published works are

not comparable without any bias. Table 1 lists the available models and their limitations and advancements in DEM Super-Resolution in chronological order.

Model	**Data Used**	**Advancements**	**Limitations**
Convolutional neural network-based DEM Super-Resolution (Z. Chen et al., 2016)	DEM in mountainous regions, source unknown.	The first model to use CNN-based SR in DEM with a very basic CNN-based approach.	Very small model with very few parameters which caused very small improvement compared to bicubic interpolation.
Deep gradient prior network for DEM Super-Resolution: Transfer learning from image to DEM (Xu et al., 2019)	Satellite images and the LiDAR DEM	Transfer learning-based approach for faster convergence of the model and a First detailed comparison on DEM and its derivatives.	The use of transfer learning is an interesting approach, but it might be difficult to obtain a very high-resolution image each time to perform Super-Resolution. The stitching of output patches is still a problem because it shows small lines in between each patch.
LoGSRN: Deep Super-Resolution network for digital elevation model (Shin & Spittle, 2019)	DEM from multiple sensors with varying spatial resolution from 0.5-1 meters	They have tested the model with multiple combinations of functions and used the Laplacian pyramid to reconstruct the data.	The model is trained and tested in very high-resolution data (0.5-1 meters) in that scale improvement might seem bigger, but using it on globally available data might not improve its quality.
D-SRGAN: DEM Super-Resolution with Generative Adversarial Networks (Demiray et al., 2020)	LiDAR DEM	The First GAN-based approach on Super-Resolution, even though the model is a bit old, is trained without any additional datasets and only with DEM data.	The researchers mistakenly designed their model to improve 16x even though they have written 4x improvement, the model does not perform well in the mountainous terrain and performs better in the flatter terrain. Furthermore, it is also not compared with other models rather than Bicubic.
Super-resolution reconstruction of a digital elevation model based on a deep residual network (Jiao et al., 2020)	SRTM DEM	This model shows a very high improvement in the PSNR by using the residual layers in the model.	The reconstructed HR image is around 30 meters, and it is not easy to have any visible features at that scale. While improving from 120 meters to 30 meters, the PSNR can improve with a large amount because generating the LR images due to very high variation in the data and bicubic interpolation will overly smooth the dataset. Furthermore, it also consists of usability problems because of using SRTM as HR data which is already available globally.

Model	Data Used	Advancements	Limitations
ESRGAN-based DEM Super-Resolution for enhanced slope deformation monitoring in Lantau island of Hong Kong (Wu & Ma, 2020)	SRTM Digital Elevation Models	First ESRGAN based approach to use in Super-Resolution of DEM. This method used transfer learning from greyscale images.	The amount of DEM for training is very low with just 3601x3601 pixel size image, which might overfit the model quite fast. The training was done in the SRTM data, which dilutes the need for developing the Super-Resolution model because SRTM data is available globally in 30 meters resolution, so the Super-Resolution model generates 30 meters DEM is not much useful. Furthermore, research also does not compare its results with existing research.
Feedback Neural Network-Based Super-Resolution of DEM for Generating High Fidelity Features (Kubade et al., 2020)	Aerial and LiDAR DEM	Use of feedback networks with an extensive amount of data for training and testing.	The model is trained and tested in the same data, but evaluation of performance in other data such as SRTM could have been better. The improvement is lower compared to FCN based method. Furthermore, the authors removed the boundary in output with overlapping, increasing the PSNR, but it does not mean it has good performance.
AFN: Attentional Feedback Network Based 3D Terrain Super-Resolution (Kubade et al., 2021)	Use of Aerial Images and Digital Elevation Models	It is the most recent and first model trained with enough data combined with images and has significant improvement in the output.	The major problem is the availability of very high-resolution images, and also, the model is trained and used in very high-resolution images only, limiting its uses in real-world scenarios.

Table 1: List of available deep learning-based Super-Resolution techniques for DEM data

2.3. Multi-Hazard Modelling

To understand the risk imposed by hydrometeorological or geological hazards in infrastructure and reduce the impact of such hazards, understanding hazard characteristics is important. To understand the impact, modelling the hazards is generally done using statistical or physically based numerical simulation. The physically based models are better suited to provide more detailed information on the underlying process of the hazard because they can simulate the information based on the empirical or physical relationships between different factors that play a crucial role in hazard characteristics (van den Bout, 2020, p. 4). In general definition, multi-hazard modelling is a process of simulating real-world hazard scenarios using different input parameters and triggering events with the combined interaction of multiple hazards. Because most of the naturally occurring hazards interact with each other and cause intensification in each other's frequency, density, distribution, and density(Bell & Glade, 2011; Finlay & Fell, 1997; Yousefi et al., 2020). The different interaction between multiple hazards is categorised by van Westen & Greiving (2017, pp. 39–42) as the following types:

1. Independent: where each hazard happens without relation to one another, such as earthquake and fluvial flood
2. Coupled: Both hazards are strongly coupled together and occur most often due to the same triggering event as drought and heat waves. The temporal probability of the events is the same because of the same triggering events.
3. Conditional: One hazard occurs due to the impact of another hazard, but with a long-term relationship, such as flood plain created by landslides can be flooded after a certain amount of time. In this case, one hazard is not a triggering event of another hazard.
4. Domino or Cascading: One hazard occurs due to the triggering of another hazard that happens in a very short duration, such as earthquake and earthquake-induced landslide. This adds more damage and extremely difficult to model.

The above-stated relations in multiple hazards are extremely difficult to simulate because multiple earth surface processes simultaneously occur (van Westen & Greiving, 2017, p. 39). Each of them has a different relationship that can alter the intensity and propagation of the hazards. The relationship between such hazards contains many uncertainties in the quantification process because of the multiple interactions and dynamic situations, and small change in one process can alter another hazard to a large extent. For example, the flood propagation directions and their velocity can be extremely deviated by the landslide occurrence, which blocks or formats the dam. Modelling of such events requires multiple physical process simulation at the same time. There are very few tools available to model such events, and some of the recent developments have increased the computation capacity as well as the theoretical understanding to model such events. Even though there are multiple models for physically based single hazard modelling, there is a lack of models capable of considering multiple hazards and their interaction at a detailed spatial level(van den Bout et al., 2018). The OpenLISEM model developed by van den Bout et al. (2018) is a good example of multi-hazard modelling, which has implemented such a model which considers shallow slope failure, flash flood, and debris flow. Some models also developed using multi-criteria evaluation and machine learning methods (Skilodimou et al., 2019; Yousefi et al., 2020). However, even though those models can work on a larger scale than physically based models, those methods cannot simulate the events based on the physical relationships and largely depend on data limiting the usability in data-poor scenarios.

To further explain the types of quantification of the multi-hazard interactions, Tilloy et al. (2019) has classified the multi-hazard quantification (modelling) process as **Stochastic**, **Empirical**, and **Mechanistic** models. The stochastic models are designed to assume that all processes show random behaviour and can be considered stochastic in nature (Cox & Miller, 2017). These models generally generate data from statistical distributions and usually use extreme value and multivariate statistics as processing methods (Tilloy et al., 2019). The types of stochastic multi-hazard models are copulas and multivariate models, and they are mostly used to model compounding hazards; some examples of such models are extreme value copulas models, joint tail regions models, and compounding flood models (Gudendorf & Segers, 2010; Ledford & Tawn, 1997; Z. Liu et al., 2018). The empirical models are mostly based on the observations and measurements, and their modelling is done using the data-based empirical relations observed on the data and such model have no capacity to extrapolate beyond the range of the observed data making it difficult to use in future scenarios (Tilloy et al., 2019; Zou et al., 2003). Some examples of empirical models are dependence measure and regression used in different scenarios (Guzzetti et al., 2007; Svensson & Jones, 2004; van den Hurk et al., 2015). The most detailed models are mostly mechanistic models that use physics-based simulations, and they are ideal mathematical representations of real phenomena (Devia et al., 2015). These models can be further classified as conceptual and physical models based on their calculation strategies. The conceptual

models usually describe all the components of a hazardous process using conceptual relationship and require a large amount of data and calibration (Tilloy et al., 2019). Some examples of such conceptual models are HBV, TOPMODEL and MORDOR (Devia et al., 2015). The physical models usually use standard physical equations to simulate such phenomenon using standard equations; the best example of such is the two-phase flow OpenLISEM model, which considers the relationship between different hazards and models such processes using physics-based equations (van den Bout et al., 2018).

The availability of such physically based models is very limited, and the review of Tilloy et al. (2019) have presented such availability in Figure 5. We can observe that the multi-hazard models using mechanistic approach are largely limited in hydrological context. The models such as linear regression, extreme value and hydrodynamic models are most widely used for multi-hazard modelling. The reasons to have such a limited number of multi-hazard models might be due to limited computational capacity in large regions and understanding interrelationships between such hazards. The need to develop models that can consider multi-hazard interactions is growing day by day.

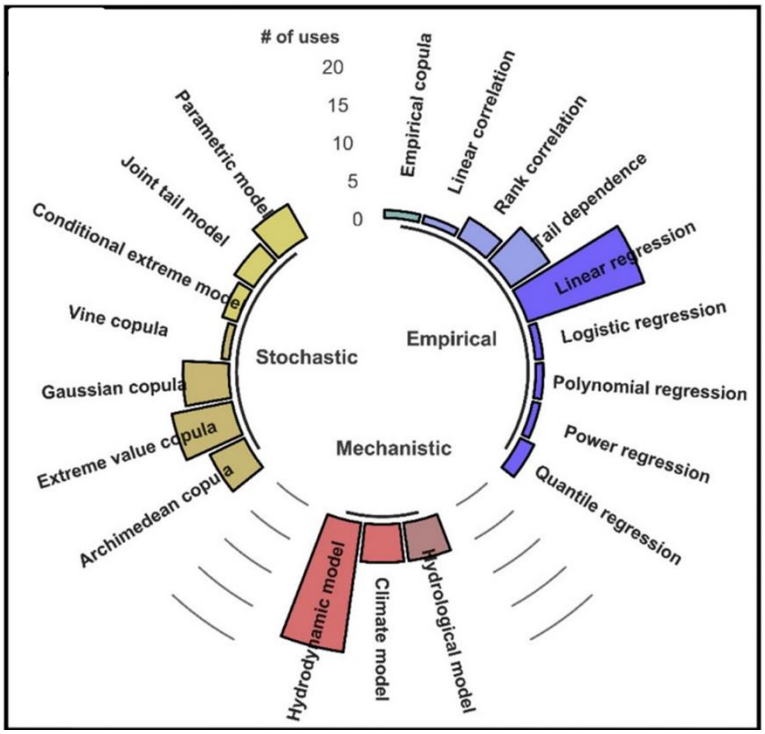


Figure 5: Multi-Hazard models using different modelling techniques. Source: Tilloy et al. (2019)

In multi-hazard modelling, there are many geospatial datasets used, such as Digital Elevation Model, soil maps, vegetation maps, etc., to name a few. Such hazard models usually take those maps as input and simulate the real-world scenario with different triggering events to estimate the propagation of multi-hazard event in different spatial locations. Those simulations and their quality are mostly dependent on the quality of the input data, and their performance is limited by the quality of elevation and soil data (van den Bout et al., 2018). Furthermore, simplifying modelling equations and assumptions on the data can also increase the uncertainties on the output to a large extent (Tilloy et al., 2019). In most of the simulations, DEM is the most important input parameter because most of the surface processes (even the effects of geological processes on the surface, such as an earthquake) are greatly affected by the terrain parameters such as slope steepness, aspect, and elevation. Based on the modelling approach, such as a data-driven or physically based approach, such input parameters are fitted or simulated through modelling equations to mimic the real

event. In our case, the research mostly focuses on physically based multi-hazard modelling because of its intense use of spatial data at pixel levels rather than regional/aggregated level, representing the impact of Super-Resolution in a more elaborative manner.

There are many mechanistic simulations in hazard modelling, such as for earthquakes, landslides, floods, etc. In our case, hydrometeorological triggered scenarios are chosen as our main hazard modelling approach and selected a combined model to process the hydrology and slope failure process in combination. Table 2 shows the examples of multi-hazard models (only accessible models) and information about the underlying equations as well as modelled hazards and their interaction by those models. Since multi-hazard modelling has recently started getting advanced due to the improvement in computation capacity, there are a limited number of models and software/ models available as of now.

Physically based Multi-Hazard Model	**Modelling Equations**	**Modelled hazards and interactions**
ICRESTRIGRS(K. Zhang et al., 2016)	Integrated CREST and TRIGRS model equations.	Flood and landslide, cascading hazard interactions.
LHT Model (Fan et al., 2017)	Integration of Rickenmann, Perla, and RAMMS models	landslide-debris flow model and neglects hydrological process. The link between hazards is one way.
EDDA 2.0(Shen et al., 2018)	Mass and momentum conservation equations	Integrates the surface failure and the erosion of debris flow.
STEP-TRAMM (von Ruette et al., 2017)	Hydro-mechanical triggering model and simple debris flow runout model.	Integrates the shallow landslide and debris flow induced by rainfall.
OpenLISEM (van den Bout et al., 2018)	Saint-Venant equations, factor of safety equations, and debris flow equation from Pudasaini (2012)	Integrates the flash floods, debris flow, and slope failures for catchment-wide modelling.

Table 2:Overview of different Multi-Hazard Models.

3. METHOD AND DATA

This section provides an overview of the methods used in this research with the explanation of the models, methods used to prepare the data, details on training the models, multi-hazard modelling, and evaluation approaches. The major workflow of the research is given by the first section, followed by the different steps used in research which explains in detail the steps followed in each section.

3.1. Research Overview

The research methodology is developed, as shown in Figure 6. The whole research is divided into two major phases; the first is the development of a deep learning-based Super-Resolution model, and the second is to test the applicability of Super-Resolution in multi-hazard modelling. The first stage is also further divided into two tasks: training of the model (green part) and evaluation of the models (yellow part). For this research, we have selected the best GAN-based model (ESRGAN) and Non-GAN based (EBRN) model for single image Super-Resolution tasks from the review of Anwar et al. (2020). Those models are further discussed in their respective sections in this chapter. The training of the model is done using the freely available DEM from Austria and its synthetic low-resolution data. After training the models, their quality and generalisation are tested at the test sites in Austria using standard computer vision test approaches such as PSNR, MSE, and SSIM. The training process of the model is improvised many times to obtain the best results, and the best model is used to further evaluate using multi-hazard modelling.

Once the Super-Resolution model reached sufficient accuracy with the test dataset, further two more test sites were selected in different locations with completely different terrain characteristics as described in the data section and tested the applicability of Super-Resolution. Those test sites are located in Colombia and Dominica, where high-resolution DEM, globally available DEM and global commercial DEM are available. Furthermore, to test the capacity of the model to reconstruct relative to its trained area, these evaluations were also performed in Austria so that we can understand if the model is well generalised or not. Globally available low-resolution Digital Elevation Models from those selected test sites are super-sampled and interpolated using bicubic interpolation. The capacity of Super-Resolution to reconstruct the DEM derivatives and geomorphological features are analysed and compared relative to high-resolution DEM. Furthermore, to compare how globally available commercial DEM can perform compared to that of the super-sampled DEM, the same test was done with TanDEM-X data in Colombia and Dominica Test sites as well. This section of research is represented by the blue box in the methodology diagram. The derivative test and geomorphological test methodology are newly developed for this research which provides the relative information between derivatives of high-resolution DEM and low-resolution DEM. The geomorphological test includes visual evaluation by geomorphological experts, and derivative analysis includes quantitative analysis using mathematical and statistical functions.

Once that process was completed, one of the two catchments was modelled using the OpenLISEM hazard model to physically simulate a multi-hazard scenario, and its results were compared using different methods as explained in relative sections. The simulation is done for the specific events to reconstruct those events and understand the change of model accuracy using Super-Resolution methods. Furthermore, all simulations are conducted in a multi-hazard environment which consists of landslides, debris flow, and flooding. To make a fair comparison, firstly, in Dominica, we selected multiple combinations of DEMs and created eight scenarios with channel flow and eight scenarios with the non-channel flow to understand how SR-based methods improve local drainage overall terrain. Furthermore, we added five more scenarios in Colombia

without channel flow to observe Super-Resolution applicability in multiple locations. The OpenLISEM model was run and calibrated in the Microsoft Azure cloud computation environment for faster computation. The next sections will further discuss the methodology in detail with the reason to use the SR models, their explanation, and specific details on their use.

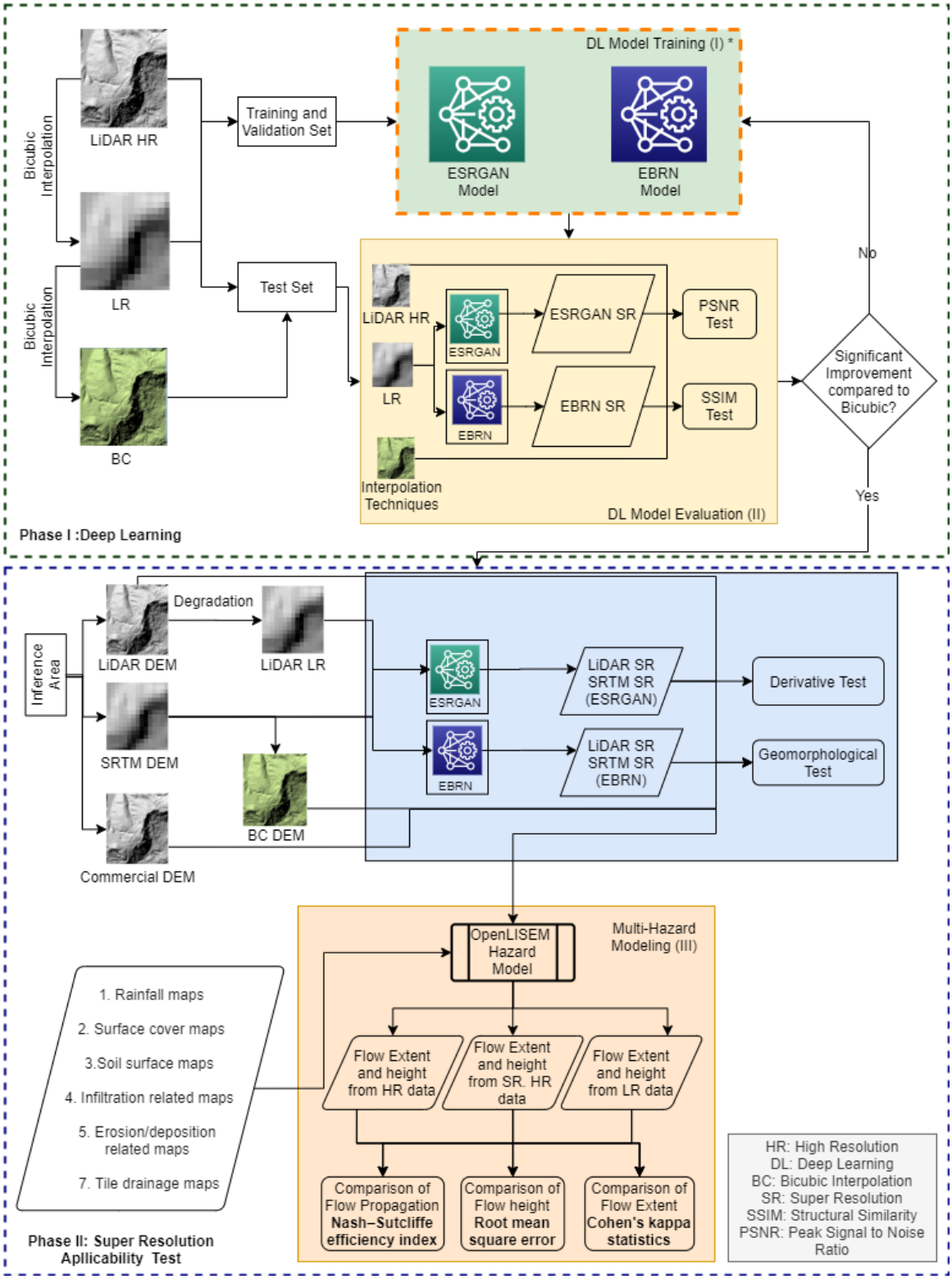


Figure 6: Research Methodology. The figure shows three different phases of research and how each phase was conducted; see the specific section for further details.

3.2. Dataset and Study Area

3.2.1. Dataset for Deep Learning (Phase: I)

To conduct this research, we required two types of datasets, first, a high-resolution dataset with sufficient quality to train the model, and second, a low-resolution dataset to inference using the Super-Resolution model and compare its quality to its high-resolution counterpart. Since DEMs generated from different observation systems over different periods do not have comparable elevation values due to datum shift, noise in the data, central location of the grid, and change in the terrain itself, we cannot directly compare the elevation values from different sensors. However, it is more relevant to compare their derivatives such as slope steepness, aspect, drainage direction, and therefore, a high-resolution DEM was also required for the inference area.

For training the deep learning network, we need high-resolution data that is freely available in locations with sufficient terrain variability where enough steep slopes and flat terrain are available. For this purpose, we found the freely available DEM from the four states of Austria most useful. We collected the 5-meter DEM dataset from Austria (Salzburg, Tyrol and Carinthia) and selected six catchments with ranges of 0-30, 30-50, and 50-70 degrees slope (Land Carinthia, 2015; Land Salzburg, 2016; Land Tirol, 2018). Each range of slope consisted of 2 catchments in random locations to create variable information. Furthermore, we randomly selected a square region with sufficient terrain variability to test the unbiased quality of the model performance; in this test area, we did not calculate the slope, but it consists of major terrain features such as mountains, valleys, and some flatter terrain. Those selected DEMs were then degraded to lower resolution data using the Bicubic Interpolation function of MATLAB and then further converted to patches of 128x128 pixels for HR samples and 32x32 for LR samples. Wang et al. (2018) have shown that using higher size patches is better for training bigger networks because it can provide more information about the local geographical characteristics of the terrain, which enables the model to learn about geographic relationships and use that information in the reconstruction of other images. The regions selected for training the deep learning model are shown in Figure 7.

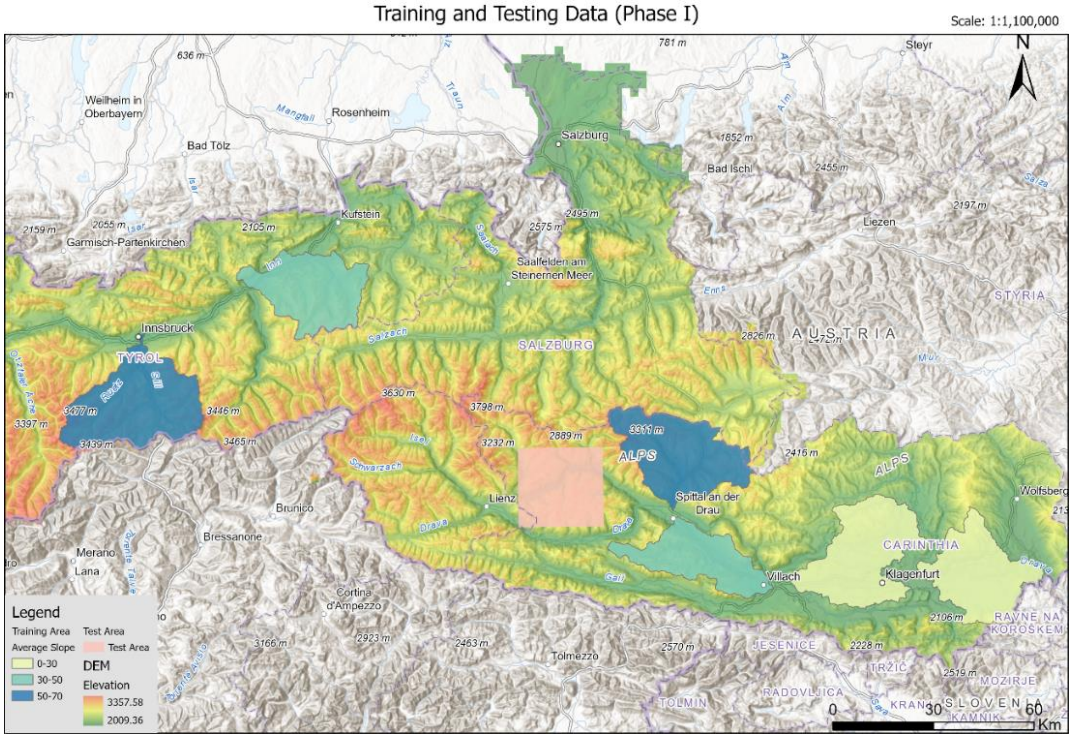


Figure 7: Training and test data samples from Austria LiDAR Data. Basemap sources: ESRI, HERE, OpenStreetMap

3.2.2. Dataset for Applicability Test (Phase: II)

For quality evaluation of the Super-Resolution globally available free DEM dataset and the performance evaluation of SR dataset in multi-hazard modelling, we have selected two more sites. The first area that we selected is Dominica, located in the Caribbean, and another is Mocoa of Colombia, the description about the modelling catchment and the study sites are present in hazard modelling specific details about the data (NASA, 2009; Stott, 2018). The global SRTM DEM and TanDEM-X provided by the German Aerospace Center (DLR) with Licence was used in both Dominica and Colombia (DLR, 2010). Because of the Non-disclosure agreement, the data cannot be shared with third parties, but the details about the licence can be further provided upon request. However, to test how the Super-Resolution performs in the global dataset in Austria compared to those areas, we also performed Super-Resolution with SRTM data in the test site of Phase I. For quality evaluation using evaluation metrics such as derivative analysis and geomorphological evaluation, we selected the places where high-resolution DEMs were available in both Dominica and Colombia; for Austria, the applicability test site was kept the same as the test site from phase I. The regions selected for those areas are shown in Figure 8. In Dominica, because the high-resolution DEM was available only on the seashore sides and was not available for the whole country (due to problems in data collection), we also extracted the SRTM and TanDEM-X DEM in the region where high-resolution DEM was available and made the comparison on those regions only. For Colombia, we also clipped the SRTM DEM to catchment shape where data was available, and all places where data was not available are set to NoData values (not to be confused with zero).

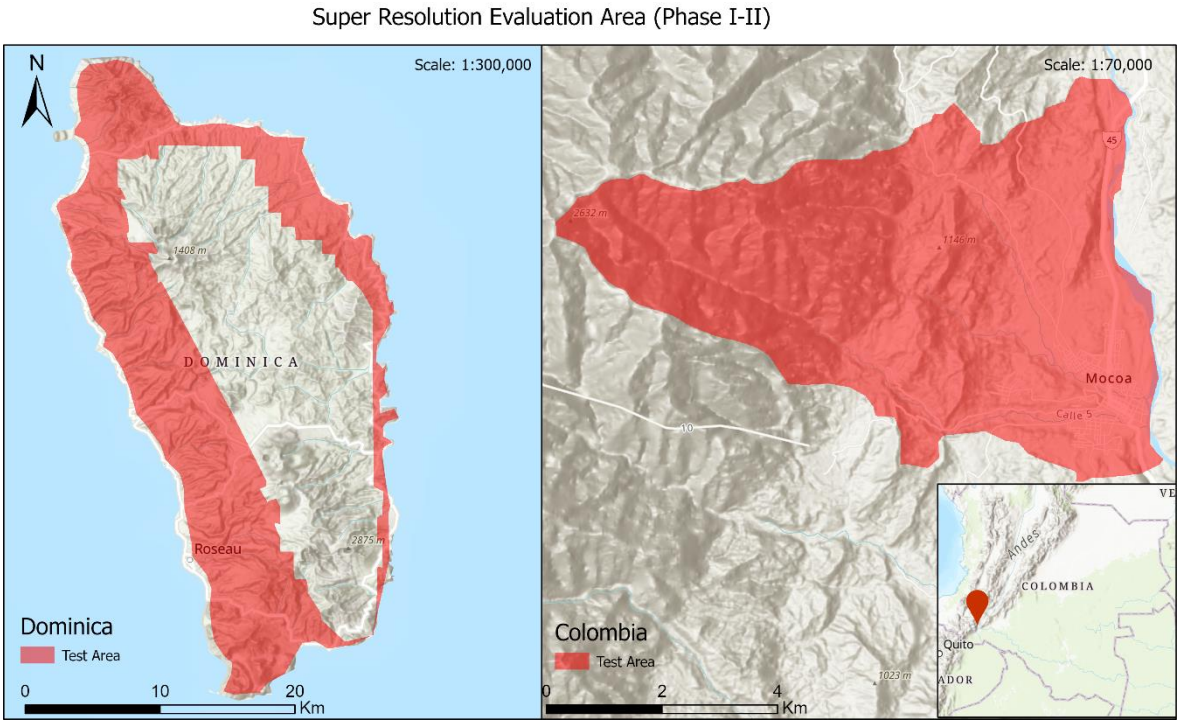


Figure 8: Inference site for global digital elevation models in Dominica and Colombia. The DEMs are clipped to match the available pixels of High-Resolution data. Basemap Sources: ESRI, HERE, OpenStreetMap.

To understand the SR-based approach's capacity to improve the quality of multi-hazard modelling, we selected two specific catchments as shown in Figure 9 and their events in Dominica and Colombia. The catchment in Dominica was modelled extensively to understand how SR-based methods perform in different types of data, such as SRTM data and LiDAR DEM data. The high-resolution DEM we had in all the Dominica study sites was degraded using MATLAB bicubic interpolation function to create low-resolution DEM. The reason for using both globally available free, commercial, and high-resolution DEM and its degraded low-resolution is that the deep learning algorithms mostly perform better in the dataset

they are trained and the performance changes with the data from a different source different amount of noise. Furthermore, the perfect recreation of the training scenario can show the capacity of the model to reconstruct HR data from LR data. However, in the case of a globally available dataset, its capacity might be degraded, and to evaluate how the performance degrades, we created those combinations of the dataset. To evaluate the performance of SR methods, we used multiple DEMs in Dominica, keeping all other parameters constant and only changing the DEM and its derivatives. Those DEMs are:

(1) SRTM bicubic interpolation
(2) SRTM Super-Resolution using ESRGAN
(3) SRTM Super-Resolution using EBRN
(4) LiDAR bicubic Interpolation from degraded DEM
(5) LiDAR high-resolution without degradation
(6) LiDAR Super-Resolution from degraded DEM using ESRGAN
(7) LiDAR Super-Resolution from degraded DEM using EBRN
(8) Commercial DEM, TanDEM-X

To further understand the applicability of SR in multi-hazard modelling at multiple locations, we further ran the multi-hazard model of the events in Colombia only for the SRTM and TanDEM-X datasets. The dataset from Dominica was evaluated in more scenarios compared to Colombia to prove that whether SR-based methods improve performance in modelling or not. However, when we already have results from Dominica, which proves the applicability of SR in multi-hazard modelling, only five scenarios of Colombia are required to check its applicability in multiple locations. The DEMs used in Colombia to evaluate applicability in multiple locations are listed as below:

(1) SRTM bicubic interpolation
(2) SRTM Super-Resolution using ESRGAN
(3) SRTM Super-Resolution using EBRN
(4) Commercial DEM, TanDEM-X
(5) High-resolution DEM

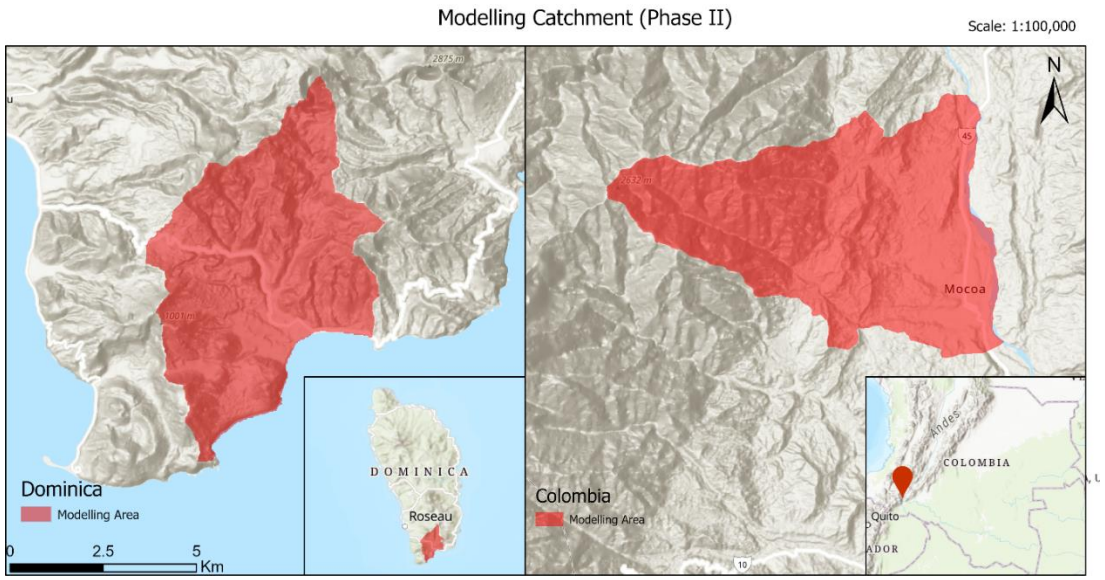


Figure 9: Modelling catchment area, Grand Bay in Dominica and Mocoa in Colombia. Basemap Sources: ESRI, HERE, OpenStreetMap.

Dominica is a small island nation with an approximate population of 71,000 people and a GDP of 551 million USD in 2018 (The World Bank, 2018). The overall terrain in Dominica is mountainous in the central part of the country, and the whole country is surrounded by the ocean making it prone to hurricanes. There

are many damaging major hurricane events in Dominica, such as Hurricane David (1972), Erika (2015), and Hurricane Maria (2017). For modelling the event, we used the hurricane Maria event, which occurred in 2017, and the physically based multi-hazard modelling was done in the Grand bay catchment of Dominica. The Maria event caused estimated damage of 930.9 million USD, and most of the damage was in the housing sector (Government of Commonwealth of Dominica, 2017). The event also caused many landslides and flashflood (Government of Commonwealth of Dominica, 2017); their interaction has changed the event's overall impact and modelling both hazards simultaneously is very important. The rainfall event that was measured in the Canefield Airport station Dominica due to the Maria hurricane is shown in Figure 10. The amount of peak rainfall that occurred due to the hurricane reached more than 120 mm/hr, and total cumulative rainfall was more than 400 mm, which caused a large amount of landslide and flood around the country. The input dataset except for digital elevation models for the Maria event was obtained from the existing model of van den Bout et al. (2020, p. 174). To simulate the event, we used canfield airport rainfall data as the rainfall scenario, and other data used for simulating the events are listed in Table 3:

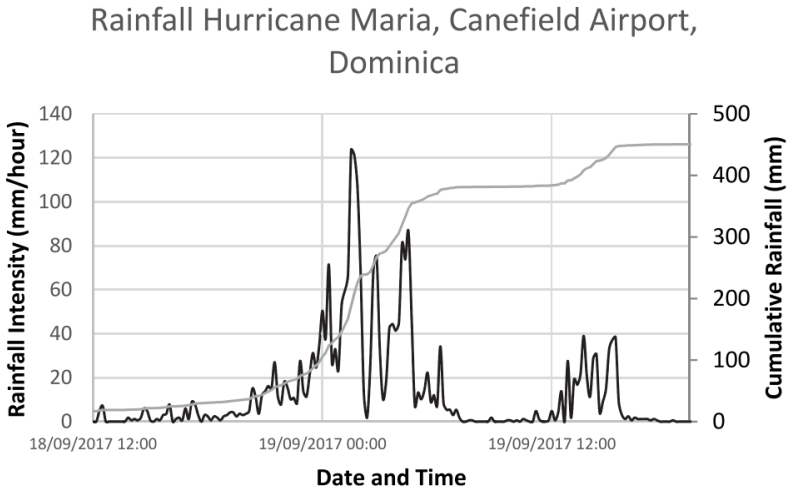


Figure 10: Rainfall in Canefield airport due to hurricane Maria. Source: van den Bout (2020, p. 175)

Input Parameter	**Method Used/Source**
Elevation	LiDAR DEM, SRTM DEM and TanDEM-X DEM
Channel Network	Created from the DEM source
Channel Properties	Channel width and depth obtained from van den Bout et al. (2020)
Land Use	Obtained from the work of van den Bout et al. (2020)
Vegetation Density	Obtained from the work of van den Bout et al. (2020)
Soil Depth	Obtained from the work of van den Bout et al. (2020), derived from empirical relations.
Urban Elements	Vector data from OpenStreetMaps, later rasterized to model grid size of 10x10 meters
Soil Physical Parameters	Obtained from van den Bout et al. (2020), the calculation was done by the pedo-transfer function from Saxton et al. (2006)
Soil Strength Parameters	Obtained from van den Bout et al. (2020), the obtained internal friction angle was later calibrated.
Precipitation	Measured rainfall data obtained from the work of van den Bout et al. (2020)
Inventory	Mass movements and flood inventory maps from the work of van Westen et al. (2020)

Table 3: Input dataset and their sources for Grand Bay, Dominica

The second catchment is selected in the Mocoa region of Colombia, a developing South American nation with an estimated population of 50 million and a GDP of 323 million USD in 2019 (The World Bank,

2020b). Being a mountainous country, Colombia has a frequent occurrence of landslide and debris flow causing frequent damage, and the World Risk Report 2020 shows that Colombia ranks 88th country in terms of risk due to hazards in between India(89) and Pakistan (87) (Behlert et al., 2020). The report also shows that Colombia's high exposure value, compared to India and Pakistan, increases the risk index, even though vulnerability and adaptive capacity are better than India and Pakistan.

To simulate the event, we used landslides and flash flood events in 2017 due to extreme rainfall. AON Benfield reports that the event caused 336 deaths with 400 injuries and missing (AON, 2017). It is the third deadliest hydrometeorological disaster in Colombia and the deadliest disaster in Mocoa in the recorded history (Masters & Grenci, 2017; Rojas, 2017). The economic damage of the event was more than 10s of millions USD (AON, 2017). The event was triggered by more than 130 mm rainfall during a very short period of 23:00 Friday, 31 March and 01:00 Saturday, 01 April 2017, whereas the region usually gets around 400 mm of rain in a month during that season (Davis, 2017). The overall rainfall graph recorded in the Acueducto rain gauge is shown in Figure 11, the total cumulative rainfall for our scenario was 630 mm, and peak rainfall was 73.8 mm/hr. This intense rainfall at short duration caused large amount of landslide and flooding leading to larger slope failures and loss.

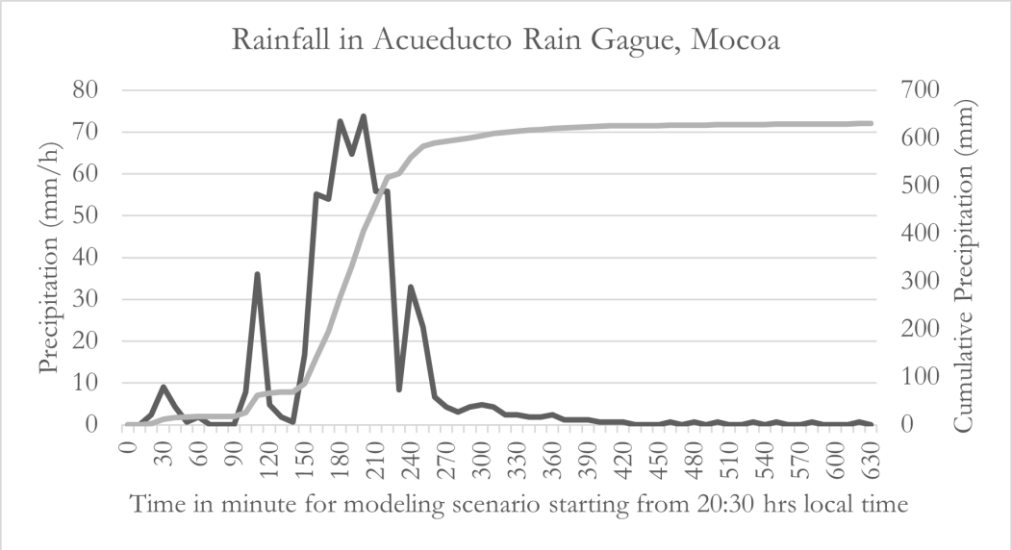


Figure 11: Rainfall in Acueducto rain gauge for Mocoa simulation. The time started from 20:30.

To simulate the event and observe the effect of Super-Resolution in modelling such extreme event, we created four different scenarios in the Colombia Mocoa region and used them to further understand the possibility of using a Super-Resolution to generate better quality modelling results in Colombia.

Urban Elements	Openstreet Maps, rasterized to map grid
Soil Physical Parameters	UNGRD and Pontificia Universidad Javeriana (2018)
Soil Strength Parameters	UNGRD and Pontificia Universidad Javeriana (2018)
Precipitation	Obtained from SGC (2017)
Inventory	Created by aerial photographs obtained after the event SGC (2017)

Table 4 shows the available input dataset for Colombia that we used and their sources.

Input Parameter	Method Used and Source
Elevation	SRTM DEM, ALOS PALSAR (HR) and TanDEM-X DEM
Channel Network	Created from the DEM source
Channel Properties	Created from the DEM source
Land Use	National Cover Map of Colombia scale 1:100.000 (IDEAM, 2010) used to create Random Roughness and Manning's roughness coefficient (n) using methods suggested by (Floors et al., 2018; Papaioannou et al., 2018; van den Bout, 2020)
Vegetation Density	Created using sentinel-2 images as suggested by van den Bout (2020)
Soil Depth	UNGRD and Pontificia Universidad Javeriana (2018)
Urban Elements	Openstreet Maps, rasterized to map grid
Soil Physical Parameters	UNGRD and Pontificia Universidad Javeriana (2018)
Soil Strength Parameters	UNGRD and Pontificia Universidad Javeriana (2018)
Precipitation	Obtained from SGC (2017)
Inventory	Created by aerial photographs obtained after the event SGC (2017)

Table 4: Input dataset for Mocoa region of Colombia

3.3. Phase I: Deep Learning Super-Resolution

We had selected two major models (GAN Based and Non-GAN Based) with the highest performance and lower computational complexity as compared by Anwar et al. (2020). The specific models and their training and evaluation strategies are explained in the following sub-sections.

3.3.1. EBRN

The Embedded Block Residual Network (EBRN) was developed by Qiu et al. (2019). The model has specific features compared to other methods, making it the best model for PSNR oriented approaches in the review of Anwar et al. (2020). Unlike other existing models, EBRN does not process data with all frequency (such as elevation difference or slope steepness) through a single network (no matter the number of layers), but it has different blocks through which the different frequency domain data get processed (Qiu et al., 2019). In our case, the patches with a very high slope and very low amount of slope usually get processed by a single network in other models, but in the case of EBRN, such processing is done through a different network depth. Having such a network structure is theoretically very beneficial for DEM data because, with different elevation changes (slope), different processing levels are required to generate better representation. If the terrain has a small slope, a smooth reconstruction might be useful, but a higher amount of processing and reconstruction is crucial to generate better SR images for a higher slope. Furthermore, it also developed a novel approach for block residual and its embedding through concatenation rather than stacking, proving to be better at reconstruction than the existing methods (Qiu et al., 2019).

We have modified the model to work on our case because it was developed for use in optical images, firstly we changed the input size of the model from 3 channel input to 1 channel input. Secondly, we added a normalization layer at the top, which converts input data from the range 0-8000 to 0-1. We kept the maximum value as 8000 because most elevations are in that range, and if we keep the range too big, the normalized values will get too small, causing the network to ignore such differences making smoother terrain. However, it will be possible to use the model in sub-zero and above 8000 meters by vertical scalings, such as in case of more than 8000 meters, we can subtract certain values before processing and adding the same amount after processing. This normalization was necessary to limit the data range from which the model can learn; otherwise, the model can generate elevation values in very high ranges, which might not

be realistic. Thirdly, we added a de-normalization layer at the last of the model, which converts the 0-1 range back to 0-8000, giving the elevation values in a similar range to that of the input.

The overall architecture of the model after modification is shown in Figure 12, and In the diagram, we can observe that the input LR image is passed through different levels of Block Residual Modules (BRM) based on their frequency (in our case, slope). Each BRM, as shown in Figure 13, aims to reconstruct the parts of higher resolution images that are in a specific frequency domain and pass the remaining signals to the next module, which again reconstructs some level of frequency, the high to low-frequency output are denoted by colour lines in the diagram. Once the data enters the BRM module, it first creates a bigger layer of specific Super-Resolution scale, such as in our case when we provide input of 32x32 pixel image, it creates 128x128 pixel image and this image then goes to two places. First, it passes through multiple convolution layers shown in Figure 14, which uses several filters of 3x3 kernel size to extract learnable features. The weights of convolution layers are adjusted during the training process to better extract terrain features, which helps the model better reconstruct the image. Second, the upsampled image gets combined with the input image using the subtraction function, which again passes through several convolution filters until it reaches the next BRM module. The image that went through the first process does not further pass through another BRM module, but the data (same essentially) went through the second process goes through another BRM module following the same process. After the next BRM module again generates the 128x128 image, the image processed through the first step is added to the recent output, and they are passed through another number of convolution layers.

In general, the first BRM module creates the SR image that has gone through a few processing steps and can reconstruct low-frequency data with better quality, and it is further processed by another module that performs similar operations in the image so on and so forth, and at last, all of them get concatenated. Once all the output from BRM is added and passed through convolution layers, it gets concatenated and further goes through the convolution layers, which estimates the weights to reconstruct the SR DEM based on all that processed information. As we can see in the image, for the last BRM, convolution operation is not done in its data which is not added to the previous BRM; this is because when the data is not added, there is no need to perform the convolution. After all, it has been done already inside BRM. In simpler terms, each BRM module process the information from an input up to its capacity and then passes it to the next BRM, which will further process that information to generate better elevation. However, when a higher amount of processing is not required, the more complex process can produce unwanted artefacts and to reduce such problems, all the outputs are first concatenated and passed through the convolution layer. So, the more BRM a data has gone through, the better it is at the reconstruction of higher frequency data, but at last, convolution layers will estimate the weights which generate the final elevation value, which should have the lowest error value compared to ground truth data. In the training process of this model, the weight of the model gets repeatedly adjusted based on some function which calculates the error of super-sampled data compared to ground truth. The detailed design of the EBRN model we used is added in Annex 1 to reconstruct the model.

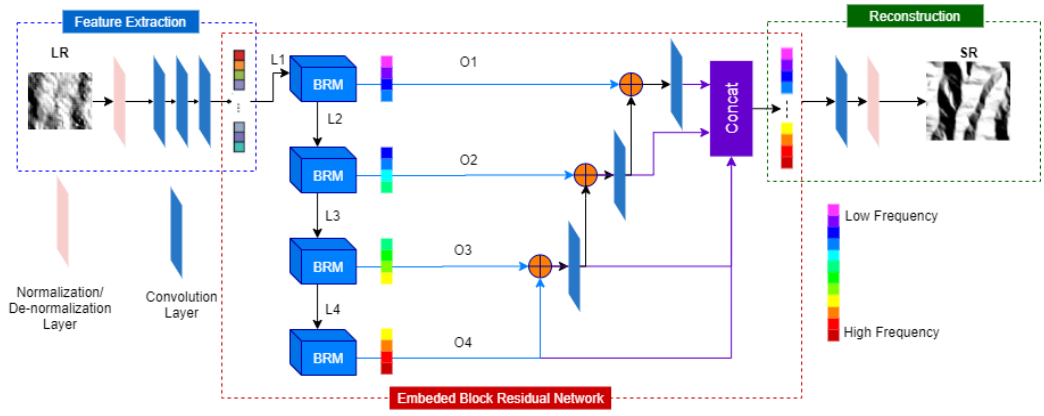


Figure 12: Embedded Block Residual Network for the Digital Elevation Model.

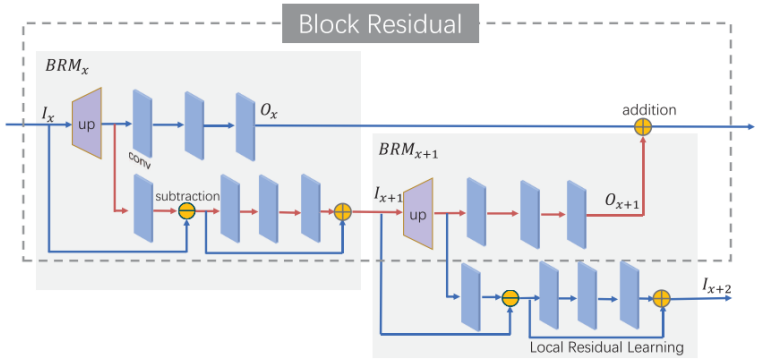


Figure 13: Block Residual Module, which shows the combination architecture of BRMS. Source: Qiu et al. (2019)

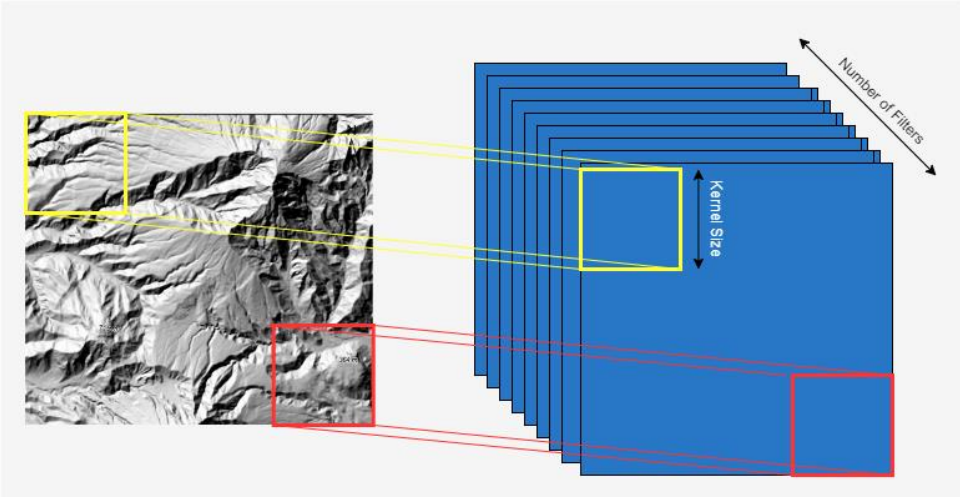


Figure 14: Simplified process of CNN where filter kernel of arbitrary size performs the convolution operation, each filter weights are estimated per number of filters and adjusted during the training process.

To train this model, we have provided the combination of HR and LR training data as described in the data section. The training is the process of adjusting each convolution layer's bias and weights in the model to reduce the loss in the model. As we briefly discussed before, the loss function is a simple function that calculates the error between ground truth data and the generated images, and an optimizer function updates the weights and bias on the model to generate the data that consists of lower loss values. In the original paper, they used L1 loss (Mean Absolute Error) as the target function and then fine-tuned it using the L2 (Mean Squared Error) as shown in Equation 1. However, for our purpose of multi-hazard modelling, only having elevation values close to high-resolution is not sufficient because most of the geospatial analysis is

not affected by absolute elevation but relative elevation between neighbouring pixels has a very high impact on the derivatives and the overall quality of model output. To minimize the relative error between the neighbouring pixels, we introduced a novel loss function called TopoLoss function, an abbreviation for Topographic Loss as shown in Equation 6, which is derived in the next paragraph. After introducing TopoLoss, we created a final loss function with the weighted average of L1 Loss and TopoLoss as in Equation 7. However, the model could not converge initially because of complexity, so we first trained the model with L1 Loss and then followed by a mixture of L1 Loss and TopoLoss. To train the model, we used Adam Optimizer to optimize the training process because of its robustness in such work (Qiu et al., 2019).

$$l1 = \mathbb{E}_n\left|y - \bar{y}\right|$$

$$l2 = \mathbb{E}_{xi}|(y - \bar{y})^2| \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$ (1) L1 and L2 Loss functions

To derive the topographic loss function, let us consider a convolution window as shown in Figure 15, where derivative values will be estimated for the pixel "e". Then, to calculate the slope and aspect as defined by Burrough & McDonell (1998, p. 190), we calculated derivative in X and Y direction as shown by Equation 2

a	b	c
d	e	f
g	h	i

Figure 15: Kernel defined to TopoLoss

$$dY_x^{\rightarrow} = \frac{(c+2f+i)-(a+2d+g)}{8}$$

$$dY_y^{\rightarrow} = \frac{(g+2h+i)-(a+2b+c)}{8}$$ …….. (2) Derivatives in X and Y directions

With the help of derivatives in x and y directions, we further calculated the slope for ground truth data Y and generated high-resolution data ($\bar{Y}$) as in equation 3.

Slope $Y = \sqrt{\left(\frac{(c+2f+i)-(a+2d+g)}{8}\right)^2 + \left(\frac{(g+2h+i)-(a+2b+c)}{8}\right)^2}$

Slope $\bar{Y} = \sqrt{\left(\frac{(c+2f+i)-(a+2d+g)}{8}\right)^2 + \left(\frac{(g+2h+i)-(a+2b+c)}{8}\right)^2}$ …….. (3) Slope for ground truth and generated data.

Similarly, we also computed the Aspect values for ground truth data Y and generated high-resolution data ($\bar{Y}$) as shown in equation 4.

$$Aspect\ Y = atan2\left(dY_y^{\rightarrow}, -dY_x^{\rightarrow}\right) \times \frac{180}{\pi}$$

$$= atan2\left(\frac{(g+2h+i)-(a+2b+c)}{8}, -\frac{(c+2f+i)-(a+2d+g)}{8}\right) \times \frac{180}{\pi}$$

$$Aspect\ \bar{Y} = atan2\left(d\bar{Y}_y^{\rightarrow}, -d\bar{Y}_x^{\rightarrow}\right) \times \frac{180}{\pi}$$

$$= atan2\left(\frac{(g+2h+i)-(a+2b+c)}{8}, -\frac{(c+2f+i)-(a+2d+g)}{8}\right) \times \frac{180}{\pi}$$ …….(4) Aspect for ground truth and generated data

Now that we have both aspect and slope, the aspect and slope loss are calculated by subtracting generated data from ground truth data.

$$Slope\ Loss = \mathbb{E}_n\left|\left(\left(\sqrt{\left(\frac{(c+2f+i)-(a+2d+g)}{8}\right)^2 + \left(\frac{(g+2h+i)-(a+2b+c)}{8}\right)^2}\right)_Y - \left(\sqrt{\left(\frac{(c+2f+i)-(a+2d+g)}{8}\right)^2 + \left(\frac{(g+2h+i)-(a+2b+c)}{8}\right)^2}\right)_{\bar{Y}}\right)^2\right|$$

$$Aspect\ Loss = \mathbb{E}_n \left| \begin{pmatrix} \left(atan2\left(\frac{(g+2h+i)-(a+2b+c)}{8}, -\frac{(c+2f+i)-(a+2d+g)}{8}\right) \times \frac{180}{\pi}\right)_Y \\ -\left(atan2\left(\frac{(g+2h+i)-(a+2b+c)}{8}, -\frac{(c+2f+i)-(a+2d+g)}{8}\right) \times \frac{180}{\pi}\right)_{\bar{Y}} \end{pmatrix}^2 \right| \quad \ldots (5)$$ Slope and Aspect loss.

After we have slope and Aspect loss, we combined them to create the TopoLoss as represented by equation 6. Where alpha and beta are regularization parameters

$$TopoLoss = \alpha \cdot Slope\ Loss + \beta \cdot Aspect\ Loss \quad \ldots\ldots\ldots (6)$$ Combined topographic loss.

After combining TopoLoss with the initial EBRN model loss, the final loss function we developed for the EBRN model is shown by equation (7), where gamma and delta are regularization parameters.

$$Final\ EBRN\ Loss = \gamma \cdot L1\ Loss + \delta\ TopoLoss \quad \ldots\ldots\ldots (7)$$ Final loss function for EBRN model.

3.3.2. ESRGAN

ESRGAN model is one of the most used models in GAN-based Super-Resolution approaches. The model is developed with two parts: a generator and a discriminator. The generator model creates the Super-Resolution image given the low-resolution image, and the discriminator model tries to identify whether the generated image looks more like a higher resolution image or not and based on that, adversarial feedback is provided to the generator network as a loss function(X. Wang et al., 2018). The generator model of the ESRGAN is composed of residual in residual blocks without any batch normalization layers to make it easy to converge.

The existing ESRGAN model was modified to add the data normalization and denormalization layer at the start and end of the model as discussed in the previous model, and the number of blocks was fixed to 20. The normalization and denormalization were done in the range from 0-8000 to 0-1 and vice versa. The residual in residual block, unlike the residual block of Ledig et al. (2017), each layer is connected to another layer inside the residual block, and the learning is done at different levels (X. Wang et al., 2018). The modified network architecture of the Generator model is shown in
Figure 16. When an input map is normalized and enters the basic block, it first passes through one convolution layer with process same as described in the EBRN model, and then it passes through an activation function called Leaky Rectified Linear Unit (LReLU). The LReLU function is a regularization function that keeps the non-zero positive values as it is, but the values below zero are converted to very small values. This function regularizes the information obtained in the convolution layers and decides which information is to be passed to the next convolution layer. The important use of the LReLU function is to minimize the extreme negative values in the convolution layers, which can limit the capacity of the model. The graphical representation of LReLU function is shown in Figure 17, where the LReLU function minimizes the negative values by a factor of 0.01 and keeps the positive values as it is. Once the data passes through the convolution layer and LReLU layer, it again gets mixed with the previous data within that block and goes to the next convolution layer. At the last of each block, there is only a convolution layer that creates the final output of the block and passes into the next block, which again processes in the same way.

Each residual in the residual block (Basic Block) on the model is densely connected to each other sequentially to increase the model's capacity. When data passes through all the processing steps, it gets enlarged based

on the information obtained from learning and then passes through few convolution layers, regenerating the elevation data. Unlike the EBRN model, where a different frequency level is processed through different layers, all the information is processed through the same network. After modifying the ESRGAN model with the addition of a normalization layer in the start and end and reduction of input channels from 3 to 1, the generator model looks like as shown in
Figure 16. Furthermore, the discriminator model of the ESRGAN model was used as is, without any further modification from Wang et al. (2018). The detailed design of the EBRN model we used is added in annex 2.

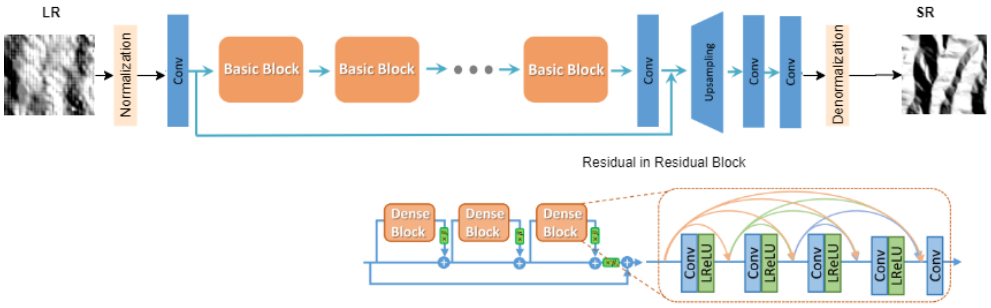


Figure 16: ESRGAN Generator model architecture. Modified from Wang et al. (2018)

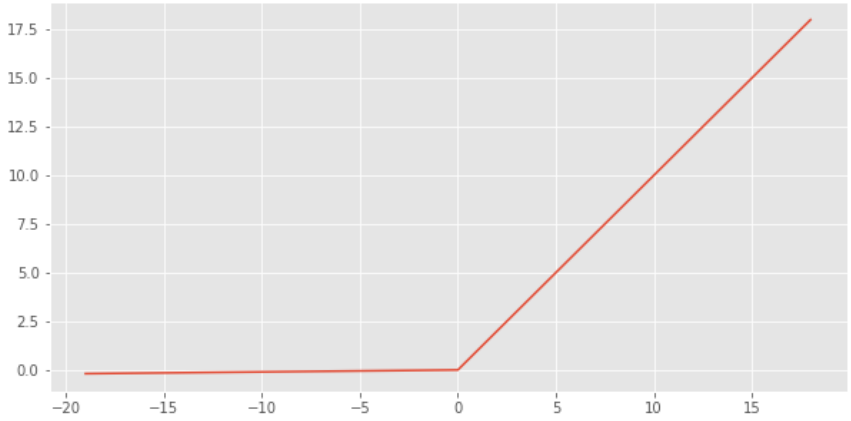


Figure 17: Leaky ReLU activation function where the negative values are minimized by factor 0.01 and positive values are kept as it is. Source: Mujtaba (2020).

In the ESRGAN paper, the model was trained to minimize the loss functions; initially, we used the L1 loss to train and L2 as shown in Equation `1 to fine-tune the generator model before GAN training to avoid collapse mode as suggested by Wang et al. (2018). However, to improve the quality of the model to minimize the relative elevation values, we introduced TopoLoss as shown in Equation 6 together with the L2 loss, to avoid collapse mode; first, we trained the model with L1 loss, and after that, the model was trained with L2 loss followed by the combination of L2 Loss and TopoLoss. Once the generator model was sufficiently trained, the GAN training was started. The GAN training is represented by the following Figure 18. For the Generator and discriminator, the Relativistic loss as defined by Wang et al. (2018) was used, which provides the likelihood of the SR image looking like an HR image. Since the Generator and Discriminator are competing against each other in a zero-sum game, the relativistic loss for the generator is shown in Equation 8. Component of 1-D_{ra} (x_r, x_f) provides the how HR data is not more realistic than SR data, and $D_{Ra}(x_f, x_r)$ provides how SR is less realistic than HR data. After combining the content loss and TopoLoss with the generator model, the final generator loss is represented by Equation (9). We did not include the perceptual loss from the original paper because, in our case, there are no distinguishable features such which could be

used for perception and using perceptual loss only increased the complexity of the model without much impact.

Disc Relativistic Loss =-$\mathbb{E}_{xr}$ $[\log(\mathrm{D_{ra}}(\mathrm{x_r}, \mathrm{x_f})]$ -$\mathbb{E}_{xf}$ $[\log(1 - \mathrm{D_{ra}}(\mathrm{x_f}, \mathrm{x_r})]$
Gen Relativistic Loss=-$\mathbb{E}_{xr}$ $\left[\log\left(1 - D_{ra}(x_r, x_f)\right]\right.$-$\mathbb{E}_{xf}$ $\left[\log\left(D_{ra}(x_f, x_r)\right]\right.$
$D_{ra}(\mathrm{x_r}, \mathrm{x_f}) = \sigma\big(C(x_r) - \mathbb{E}\big[C(x_f)\big]\big)$
$D_{ra}(\mathrm{x_f}, \mathrm{x_r}) = \sigma\big(C\big(x_f\big) - \mathbb{E}[C(x_r)]\big)$ …… (8) Relativistic Loss for Generator and Discriminator. Where D is the relativistic discriminator function. Source: Wang et al. (2018)

$GenLoss = \alpha L_g^{Ra} + \eta L1 + \beta\, TopoLoss$
$L1 = \mathbb{E}_{xi}\big||G(xi) - y|\big|$ ……. (9) Final Generator Loss function where TopoLoss is represented by topographic loss above and function G is generator model. α, β, η are the regularization weights.

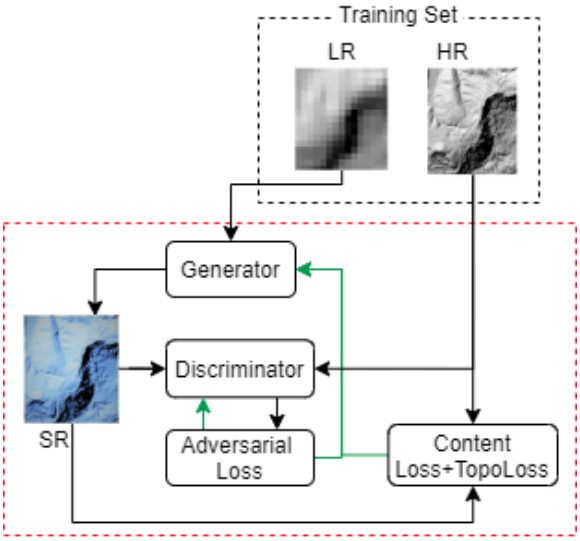


Figure 18: GAN training strategy.

3.3.3. Experimentation Details

The EBRN model was trained with a 20% randomly selected validation set of the data from the training dataset and was trained for 1000 epochs with checkpoints to avoid overfitting, where the best results on validation will be saved. The model was trained with Adam Optimizer, and the learning rate was set to 1e-04 in the beginning and then reduced by a factor of 0.5 in every 100 epochs until it reaches 1e-06. The batch size for the model was 10, and there was a total of 500 steps in each epoch. Due to the very complex combined loss functions model had problems in converging initially, and to avoid that, we used a similar concept as curriculum learning (Bengio et al., 2009) but with increasingly complex loss functions instead of increasingly complex datasets. Initially, we used simple Mean Absolute Error, and after the model converged, we used Mean Squared Error for the better generalization, and finally, we used a combination of L1 Loss and TopoLoss as suggested by Equation 7. The major reason behind this was that when we used the combined loss function without a gradual increase of complexity, it took too long to converge the model. To our knowledge, this is the first research to use loss functions in such a way for optimization.

For the ESRGAN Model, the Generator model was trained first without the presence of a discriminator to avoid collapse mode because of too weak generator followed by GAN training as suggested by Wang et al.

(2018). The training was done using the Adam optimizer with a learning rate initially at 1e-04 and then decayed by factor two until it reaches 1e-06 for the generator. The Generator training was done in similar ways as the EBRN model with a curriculum-like training, where instead of gradually increasing the data complexity, we increased the complexity in the loss function. Firstly, the model was trained with an L1 loss followed by L2 Loss and finally, the combination of TopoLoss and L1 Loss. Once the generator was trained for 1000 epochs, with L1 and TopoLoss, we started the GAN learning process where both generator and discriminator performed against each other for another 1000 epochs. During the GAN training, the discriminator learning rate was set to 1e-03 for faster learning in the beginning as compensation to that of the pretrained generator, and the generator learning rate was set to 1e-05. After both models converged and have reliable and acceptable results, we stopped the training process, and model averaging was done. The model averaging identifies the best suitable network weights that have a lower amount of artefacts, higher PSNR, and higher visual quality. This approach was suggested by Wang et al. (2018). The averaging was done between the pretrained generator network and the GAN-based trained network as in Equation 10, where we could decide the factor for each model based on our requirement of higher PSNR or visual quality images.

$Final\ Model = \mathrm{A} \cdot G_{Pretrained} + (1 - A) G_{GAN-trained}$ … (10) Model Averaging function where A is the weight ranging from 0-1.

After the model training, during the inference process, we had to create smaller patches according to the model shape and again stitch them after prediction. We observed that the model produces noisy data in the boundaries of patches introducing artefacts in the inference boundaries. To solve that, we applied a basic photogrammetric overlap approach and developed an inferencing program in such a way that the output parts from the SR model during inference are overlapped, as shown in Figure 19, to remove noise in bordering pixels. A similar approach has been used by Kubade et al. (2021) during the training and evaluation process, but such application would mean over-amplification of accuracy/PSNR and only falsely represent that the model has better capacity when in reality it does not have. To avoid making such mistakes to get a false impression of the model capacity, we implement this approach in inference only, so the quality of the model during training and evaluation remained

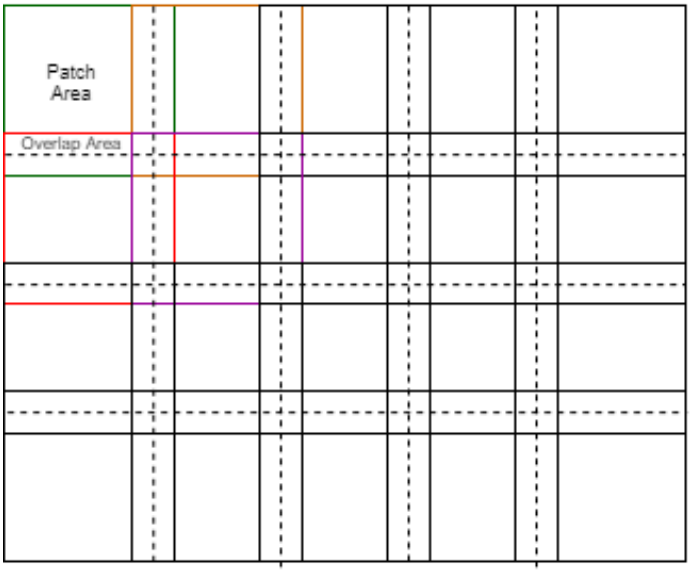


Figure 19: Overlapping of the Inference Patches. Here, the dotted lines separate the pixels used from each patch, and the bold line shows the boundary of the patches. All patches are overlapping with each other, and only certain portions of the overlapping area are used.

3.3.4. Super-Resolution Evaluation Methods

After training the Super-Resolution model, the evaluation of the Super-Resolution output must be conducted. For that, we have selected four different approaches. The first two (PSNR and SSIM) were applied on the dataset from Austria, Dominica, and Colombia, where we used high-resolution and low-resolution data from the same DEM source. The low-resolution data is created using bicubic interpolation degradation from the high-resolution data. The remaining two (Derivatives and Geomorphological testing) are used in all sites where we do not have a high-resolution counterpart from the same source for low-resolution data such as SRTM. Even though LiDAR HR images are available in Dominica and Austria, it is not possible to compare them using the PSNR, and SSIM approaches because the elevation values in each pixel are different from both sources due to their measurement bias and noise, grid structure, quality of measurement, and dates when they are obtained. Therefore, to compare such data from different sources, we developed two novel comparison methods: Derivative Evaluation and Geomorphological Evaluation to suit our application.

3.3.4.1. Peak Signal to Noise Ratio (PSNR)

The PSNR is emerged from electrical engineering to measure the ratio between the signal's maximum power and the power of the noise (MATLAB, 2020). In computer vision and machine learning, it has been used frequently for quality checking of the output from different classification algorithms, and it is also a common method to check the quality of the Super-Resolution algorithms (Ledig et al., 2017). The PSNR in this context will measure the quality of the generated pixels compared to original high-resolution data and will not consider the existing errors in the measured high-resolution data. It is the logarithmic version of mean squared error, where it is computed based on the maximum possible values. We also evaluated MSE to compare the error reduction in each method and relative comparison between improvement by both EBRN and ESRGAN. The equation for measuring PSNR is shown in Equation 11.

$PSNR = 10log_{10}\left(\frac{R^2}{MSE}\right)$ ……(11) Where R is the maximum fluctuation in the image data and depends on the image's bit depth, such as an 8-bit image will have 256 value of R and MSE is the Mean Square Error of the difference between generated and real image (MATLAB, 2020).

3.3.4.2. Structural Similarity Index Measure (SSIM)

The PSNR measures the quality of the measurement based on the mean square error and does not consider the human perception and spatial variability of the images (Z. Wang & Bovik, 2009). To measure the perceived quality, we planned to use the SSIM method, which compares the image quality of the generated high-resolution image with the measured high-resolution image. The SSIM method considers the luminance (brightness), contrast, and structural information while comparing the data (Leong & Horgan, 2020). In our case, the luminance and contrast are represented by the actual ground measurement instead of the digital number of the reflectance so, the SSIM is expected to show very high similarity compared to the photo Super-Resolution. To better represent the similarity between generated DEM, instead of using the DEM directly, we used SSIM because with very small changes in elevation values, the SSIM did not change much, and it was already saturated up to 1e-5 range. This method has been further improved for DEM analysis and used to show improvement using derivatives such as hillshade instead of raw elevation. It is computed using a complex set of equations using covariance and variance.

3.4. Phase II: Super-Resolution Applicability Test

In this section, as shown in the methodology diagram, we will test the applicability of Super-Resolution in the global dataset and its capacity to improve the modelling output. We will further explain Super-Resolution evaluation techniques for the global dataset, physically based modelling, and evaluation methods.

3.4.1. Super-Resolution Evaluation in Global Dataset

To analyse how the Global datasets are changed using Super-Resolution methods compared to that of the interpolation techniques, we defined two evaluation methods, the first is the derivative analysis, and the second is visual and geomorphological analysis. This section will discuss their needs and the approach that they use to evaluate such data.

3.4.1.1. Analysis of DEM Derivatives

Our research aims to use global freely available datasets with models trained with the high-resolution dataset and apply them when there is no high-resolution counterpart available. In those case, we cannot measure the improvement using PSNR and SSIM methods because they always need a high-resolution dataset which has exactly aligned pixels. In our case, the SRTM DEM and HR DEM are available to compare, but their pixels are not aligned, and it is more important for us to have better derivatives than absolute values. To overcome that problem, we developed and used this DEM derivative analysis method to evaluate to what extent the bicubic and SR methods can reconstruct the topographic properties compared to high-resolution data. This evaluation can provide meaningful information because, in many geoscientific applications, it is more important to have the correct derivatives of the DEM than that of the correct and aligned elevation itself. Most of the hazard simulation models work on the relative elevation differences of the neighbouring pixels rather than that of the absolute elevation.

The major information that can be inferred from this analysis is to understand how derivatives obtained from each DEM are more likely linked to that of the higher resolution data. Xu et al. (2019) compared the DEM derivatives using the mean absolute error in each DEM compared to a high-resolution DEM, but the major problem with that approach is that it can be prone to outliers present in the data and especially in SR based techniques, the presence of artefacts can overshadow the overall improvement of the data. Another problem with the error-based approach is that if the high-resolution pixel has a certain shift in terrain representation, then all the derived derivatives will be shifted by a few pixels, which cause a large amount of error even though the actual error might be lower. Furthermore, it also does not show where the SR-based methods have improved compared to the interpolation-based methods.

To overcome those problems, we used the Kernel Density Estimation Function. This function can analyse the data in a more elaborative manner where we could observe the overall performance of the models in all the pixels without being affected by bias. In the derivative analysis, we calculated the relevant derivatives (namely Slope steepness, Aspect, and Topographic Wetness Index) and plotted the kernel density estimate for each derivative in each study site represented in green colour by Table 5. The kernel density estimate function is calculated using the Seaborn library in Python. The kernel density estimation function estimates the probability density of the slope, aspect, and TWI in different elevation datasets using equation 12 (Rosenblatt, 1956). The Kernel Density Estimation function in Equation 12 shows the function to estimate kernel density, where K is a kernel function, x and x_i are the points where the kernel density is estimated and point used to calculate KDE, respectively. The h parameter is the smoothing parameter that decides how smooth the histogram we want to generate, and n is the number of points to consider. We used this function together with a histogram plot to understand how the derivatives are distributed and how they are more likely to be similar to the high-resolution data. To put more perspective, if we compare the histogram usually, due to the bars, it is difficult to visualize which data is more likely to be similar to high-resolution data (because we have eight different datasets, its difficult to understand the difference) when we use KDE function, it smooths the histogram. When plotted, we can visualize the relation between dataset more clearly. One example of a histogram plot and its KDE function plot is shown in Figure 20.

$$f = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \quad \text{..... (12) Kernel Density Estimation Function}$$

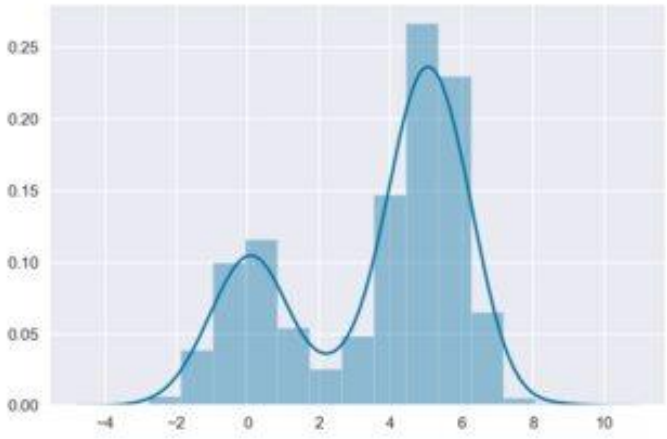


Figure 20: Input distribution and its KDE estimation sample. Graph generated from random data using Python.

To understand how SR-based approaches can perform compared to commercial data such as TanDEM-X, we also did the derivative analysis with TanDEM-X data. The combination matrix is shown, where all comparison is done with the HR data, and all green combinations have a respective evaluation of slope, aspect, and TWI, and due to the unavailability of TanDEM-X in Austria, we have only made the comparison with SRTM data. There are 33 different combinations of derivative analysis to check the quality in different locations with different derivatives.

Test Site	SRTM Low Resolution	SRTM Bicubic Interpolation	SRTM Super Resolution EBRN	SRTM Super Resolution ESRGAN	TanDEMX
Dominica					
Colombia					
Austria					NA

Table 5: Evaluation matrix for Derivative analysis.

3.4.1.2. Visual and Geomorphological Analysis

To better understand how the improvement is done in the SR-based methods, we did an extensive visual and geomorphological evaluation where first we visually compared the results to observe the improvement, and after that, we asked geomorphological experts to score them. The PSNR, SSIM, and Derivative evaluation were quantitative, but we also considered a qualitative visual analysis to conclude that the Super-Resolution is working better than the existing method to generate crisp images. Because the quantitative analysis can provide information on how the improvement is there numerically, but that might not always be useful for geoscience application. For example, even though the mean squared error is low if the terrain features are not well reconstructed or visible in the DEM data, it might not improve the quality of model or map products. To understand if the geomorphological features such as landslide scars and other features are recognizable and clearer in the SR image, we asked the expert opinion from two Geomorphologists. We provided the landslide scar (vector point data) and Low resolution, Bicubic interpolated, Super Sampled, and High-Resolution hillshading images for the three study locations to two different Geomorphologists and asked them to score them from 0-10, being HR data as highest in a relative scoring and fill the following Table 6. Then the scores are averaged, and the conclusions are drawn based on those independent evaluations in the next chapters.

Test Site	SRTM Low Resolution	SRTM Bicubic Interpolation	SRTM Super Resolution EBRN	SRTM Super Resolution ESRGAN	TanDEMX	LiDAR HR DEM	Remarks per test site
Dominica						10	
Colombia						10	
Austria					NA	10	

Table 6: Geomorphological evaluation sheet

3.4.2. Multi-Hazard Modelling

In our case, we have used the OpenLISEM model as our model of choice because of its flexibility to work with different geospatial data and efficient and intuitive graphical user interface. For the study area of Dominica, we already have an existing model available from van den Bout (2020), and because of the Covid-19 pandemic situation, the collection of new data in the field was not possible. The OpenLISEM hazard model is based on the different sets of equations for different processes such as flood, landslide, and debris flow as shown in Equation 13, 14, 15; the equations are taken from van den Bout et al. (2018); please see the source for further explanations and derivation of the flow equations. The model takes different data as input as suggested by (van den Bout & Jetten, 2018); there are the following major categories of the data to be used in the OpenLISEM multi-hazard model: (i) Topography related maps (ii) Rainfall maps (iii) Surface cover maps (iv) Soil surface maps (v) Infiltration maps (vii) Channel maps (viii) Tile drainage maps.

$$\frac{\partial h}{\partial t} + \frac{\partial (hu_x)}{\partial x} + \frac{\partial (hu_y)}{\partial y} = R - I$$

$$\frac{\partial hu_x}{\partial t} + \frac{\partial (hu_x^2)}{\partial x} + \frac{\partial (hu_x u_y)}{\partial y} = gh\left(S_x - S_{f,x}\right)$$

$$\frac{\partial hu_y}{\partial t} + \frac{\partial \left(hu_y^2\right)}{\partial y} + \frac{\partial (hu_x u_y)}{\partial x} = gh\left(S_y - S_{f,y}\right)$$ …….. (13) Saint Venant Equation for shallow flow modelling where h is height, R is rainfall, I Is infiltration and s is friction, and sf is momentum(van den Bout et al., 2018).

$$SF = \frac{c' + c + ((\gamma - m\gamma_w)z + m\gamma_w z)cos(\beta)^2 \tan(\phi')}{((\gamma - m\gamma_w)z)sin(\beta)\cos(\beta)}$$ ……(14) Factor of Safety Equation for Shallow Slope Failures, Where SF is the safety factor, b is the slope, c is cohesion, c' is apparent cohesion, and f is the frictional angle (van den Bout et al., 2018).

$$S_{x,s} = \alpha_s\left(g\left(\frac{\partial b}{\partial x}\right) - \frac{u_s}{|\vec{u}_s|}\tan(\partial P_{b_s}) - \varepsilon P_{b_s}\left(\frac{\partial b}{\partial x}\right)\right) - \varepsilon\alpha_s\gamma P_{b_f}\left(\frac{\partial h}{\partial x} + \frac{\partial b}{\partial x}\right) + C_{DG}\left(u_f - u_s\right)\left|\vec{u}_f - \vec{u}_s\right|^{j-1}$$

$$S_{y,s} = \alpha_s\left(g\left(\frac{\partial b}{\partial y}\right) - \frac{v_s}{|\vec{u}_s|}\tan(\partial P_{b_s}) - \varepsilon P_{b_s}\left(\frac{\partial b}{\partial y}\right)\right) - \varepsilon\alpha_s\gamma P_{b_f}\left(\frac{\partial h}{\partial y} + \frac{\partial b}{\partial y}\right) + C_{DG}\left(v_f - v_s\right)\left|\vec{u}_f - \vec{u}_s\right|^{j-1}$$

$$S_{x,f} = \alpha_f\left\{g\left(\frac{\partial b}{\partial x}\right) - \varepsilon\left[\frac{1}{h}\frac{\partial}{\partial x}\left(\frac{h^2}{2}P_{b_f}\right) + P_{b_f}\frac{\partial b}{\partial x} - \frac{1}{\alpha_f N_R}\left(2\frac{\partial^2 u_f}{\partial x^2} + \frac{\partial^2 v_f}{\partial y \partial x} + \frac{\partial^2 u_f}{\partial y^2} - \frac{\chi u_f}{\varepsilon^2 h^2}\right) + \frac{1}{\alpha_f N_R}\left(2\frac{\partial}{\partial x}\left(\frac{\partial \alpha_s}{\partial x}\left(u_f - u_s\right)\right) + \frac{\partial}{\partial y}\left(\frac{\partial \alpha_s}{\partial x}\left(v_f - v_s\right) + \frac{\partial \alpha_s}{\partial y}\left(u_f - u_s\right)\right)\right) - \frac{\xi\alpha_s\left(v_f - v_s\right)}{\varepsilon^2\alpha_f N_{R_y} h^2}\right]\right\} - \frac{1}{\gamma}C_{DG}\left(u_f - u_s\right)\left|\vec{u}_f - \vec{u}_s\right|^{j-1}$$

..(15)

Debris flow Equation from Pudasaini (2012). Please see the article for further explanation.

In the multi-hazard modelling phase of the research, we only selected two test sites, Dominica and Colombia, for our test purpose. We did an extensive amount of modelling with 16 different scenarios in Dominica and only 4 Scenarios in Colombia because modelling extensively in each study site will be redundant because of the similar characteristics of the model and simulation. Firstly, we modelled 16 different scenarios in Dominica to test the applicability of Super-Resolution in physically based modelling. We modelled and calibrated both channel and non-channel flow simulations in Dominica to understand how the flow from the 1d channel flow and non-channel flow behaves with super-sampled Digital Elevation Models. The channel flow is the case where all the fluid is flowing through specific channels of a certain depth, and non-channel flow is where the flow is directed based on the elevation data rather than channels of specific width and depth; this will provide more information on the performance of Super-Resolution on improving the channels as well as overall terrain.

To understand how the noise and uncertainties present on the global dataset are addressed by Super-Resolution methods, we also created a synthetic LiDAR low-resolution dataset by degradation function, which was also super-sampled and modelled. This will provide a fair comparison of how the Super-Resolution performs in a global dataset compared to the dataset with no or less noise. All the combinations are run in 10-meter spatial resolution where all DEMs are either super-sampled or interpolated using bicubic interpolation. To check the model's applicability in data scare regions, we created a bicubic interpolated and super-sampled model dataset from SRTM data and ran and calibrated the model with those datasets. Lastly, to compare the quality of super-sampled DEM compared to commercially available global DEMs such as TanDEM-X, we also created one scenario with TanDEM-X data resampled from ~12 meters to 10 meters. For TanDEM-X, we did not use supersampling because the Super-Resolution model that we developed can scale in a factor of 4 and to convert from 12 meters to 10 meters bicubic interpolation suits better than using the Super-Resolution because of the flexibility of bicubic interpolation in scaling. Furthermore, to understand which genre of super-sampled model performs better (either PSNR oriented- Non-GANs or SSIM oriented GANs), we created scenarios with both EBRN and ESRGAN. The combinations of all scenarios are shown in Table 7where we can observe that the total number of physically based model combinations where scenarios in blue are channel flow and scenarios in green are non-channel flow.

To calibrate all those scenarios 20 times, we needed at least 400 simulations, and each simulation takes around 8 to 12 hours. With a single computer or computation platform, it was impossible to run it for almost 4800 hours, so to solve that problem, we used cloud computing and implemented the model in Microsoft Azure cloud. These models are run using the OpenLISEM software version 2.0 developed using an equation from van den Bout et al. (2018). The software had the capacity to run in the graphical processing unit (GPU), so, to simulate multiple scenarios in a shorter time, we used 4 AMD powered GPU in Microsoft Azure and one computer with GPU processing capability. To run the OpenLiSEM in Azure cloud, we needed GPU with an AMD GPU processor; NVIDIA processors were more expensive and did not support OpenLiSEM. Furthermore, the group policy and environmental settings were also tweaked in the Azure servers to run OpenLiSEM without any issues. This analysis will simulate the hazards in Colombia and Dominica, which will eventually show how the Super-Resolution affects the quality of the model output. As a by-product of this thesis, we have also identified ways to utilize high-performance GPUs in Microsoft Azure to run the OpenLisem model. A modified and ready-to-use Windows operating system installation disk is also available for future use to run OpenLiSEM in Microsoft Azure.

Scenario	DEM Data	Resolution Improvement Technique	Study Site	Channel flow
1	SRTM	Bicubic Interpolation	Dominica	Yes
2	SRTM	Super Resolution ESRGAN		Yes
3	LiDAR	Original High Resolution. N/A		Yes
4	LiDAR	Super Resolution ESRGAN		Yes
5	TanDEMX	Original High Resolution. N/A		Yes
6	LiDAR	Bicubic Interpolation		Yes
7	SRTM	Super Resolution EBRN		Yes
8	LiDAR	Super Resolution EBRN		Yes
9	SRTM	Bicubic Interpolation		No
10	SRTM	Super Resolution ESRGAN		No
11	LiDAR	Original High Resolution. N/A		No
12	LiDAR	Super Resolution ESRGAN		No
13	TanDEMX	Original High Resolution. N/A		No
14	LiDAR	Bicubic Interpolation		No
15	SRTM	Super Resolution EBRN		No
16	LiDAR	Super Resolution EBRN		No
17	SRTM	Bicubic Interpolation	Colombia	Yes
18	SRTM	Super Resolution ESRGAN		Yes
19	SRTM	Super Resolution EBRN		Yes
20	Aerial Survey	Bicubic Interpolation		Yes
21	TanDEMX	Original High Resolution. N/A		Yes

Table 7: Combinations for Physically based multi-hazard modelling.

The models in Table 7 had different input datasets as described in the data section. Most of the physical and soil parameters were kept constant for a fair comparison, but the DEM-derived products, such as slope steepness, channel maps, etc., are changed to create each scenario as a completely new dataset from that DEM. Table 8 shows the input dataset that was used in the modelling for both cases and their brief description; for the source of those data and detailed preparation methods, please see their relevant sources. In the model configuration, we have included the hydrology, slope stability, slope failure, and initial stability to mimic the simulated events as explained in the study area and data section. Furthermore, to estimate the precise parameters to fit the model, all models were calibrated at least 20 times using a gradient descent algorithm (Leon Bottou, 1991).

Input map	Colombia	Dominica
Triggering		
Rainfall	Measured data and all catchments have the same values.	
Topographic		
DEM (meters)	Obtained from various sources, TanDEMx and SRTM are hydrologically corrected during the modelling. The catchment has different values.	
Mask (binary)	Processed from DEM for each scenario and used post-processing to create the mask file	
channel (meters)	Processed from DEM for each scenario and used post-processing to create the channel file, standard lddcreate function is used in PcRaster to obtain the channel. Spatially varying dataset.	
channel width (meters)	N/A	Varying in most of the locations based on the terrain parameter, inferred from ground observation
channel depth (meters)	N/A	

Input map	Colombia	Dominica
Surface		
Manning's N (s/[m^1/3])	Less varying than Dominica and ranges from 0.013-0.1	Spatially varying with a range from 0.05 - 0.203
Building Cover (binary)	It was created from vector data mostly varying in spatial locations based on the availability of Buildings.	
Road Width (meters)		
Vegetation Height (meters)	N/A	
Vegetation Cover (fractional)	Spatially varying, computed from optical imagery	
Smax Canopy (meters)	N/A	Spatially varying with the different range depending on the vegetation
Smax Surface (meters)	N/A	Spatially varying with the different range depending on the surface and landuse
Subsurface		
Ksat (mm/h)	Very high range of the difference, from 0.003 to 157, in most places it has 0.003 except few which is most likely a lake.	N/A
Clay (fraction)	N/A	Similar to Manning's N with a value range from 0.28-0.33
Sand (fraction)	N/A	Spatially varying from 0.2 to 0.4
Gravel (fraction)	N/A	Constant value of 0.001
Organic Material (fraction)	N/A	Spatially varying, ranging from 0-130
Density (kg/m^3)	Spatially varying, ranging from 1300-1800	Spatially varying, ranging from 1000-1270
Soil Depth (meters)	Spatially varying, ranging from 2.0-4.252	Spatially varying, ranging from 0-2.7286
Ground Water Height (meters)	N/A	N/A
Internal Friction Angle (radians)	Spatially varying, ranging from 0.17 to 0.57	Mostly same except few pixels ranges from 0.52-0.54
Theta Initial (fraction)	Spatially varying, ranging from 0.30 to 0.54	Spatially varying, ranging from 0.012-0.71995
Cohesion Top (Pa)	N/A	Similar to Manning's N with a value range from 25-30
Cohesion Bottom (Pa)	Spatially varying, ranging from 7 to 100	Similar to Manning's N with a value range from 25-31
Density Bottom (kg/m^3)	N/A	Fixed value of 2700
Rock Size (meters)	Spatially varying, ranging from 0.0002 to 2	The constant value of 0.05

Table 8: Brief description and properties of the dataset used.

The gradient descent algorithm (GD) is a widely used algorithm in deep learning as well as in mathematical optimization problems. It first simulates the OpenLISEM model with initial values that we provided, and then results of the OpenLISEM model are evaluated using the continuous Cohen's kappa statistics, and it returns an error (1-kappa) value to the GD algorithm. The goal of the GD algorithm is to update the input parameters in such a way that it minimizes the error (Leon Bottou, 1991). As shown in abstract Figure 21, the goal of the GD algorithm is to reach the blue point of the error space by updating the parameters. Once the error values are obtained from the initial run, it again runs with updated parameters and calculates the first derivative. The first derivative provides the slope, making the GD algorithm understand how to update the next parameters to minimise the error. After GD knows the slope, it updates the parameters by using Equation 16, where learning rate is used to control how much effect the slope has in updating the parameters, and momentum is a decay factor that controls the weights from the previous slope, with the use of momentum, the GD can remember what the slope in n-1 was run, and better estimate the next parameters which lead to a better solution. This approach learns from the data and updates the calibration parameters in each step based on the previous step's results, but learning rates and momentum parameters are crucial and should be chosen with great care.

Because we had to automate the process due to 420 different simulations, using the GD algorithm was easier where we do not have to calibrate the parameters manually. The Internal friction angle, Manning's N, and Cohesion at the bottom of the channel were calibrated 20 times in each scenario using the GD algorithm because the model is highly sensitive to those parameters and requires significant adjustment with the new Digital Elevation Models. So in our case, in each step of the modelling, the gradient descent calculates the first derivative and then multiplies it with the learning rate provided by us to obtain the step size of the change used in computing new parameter values (Internal friction angle, Manning's N and Cohesion at bottom) by subtracting the step size from the old parameter value.

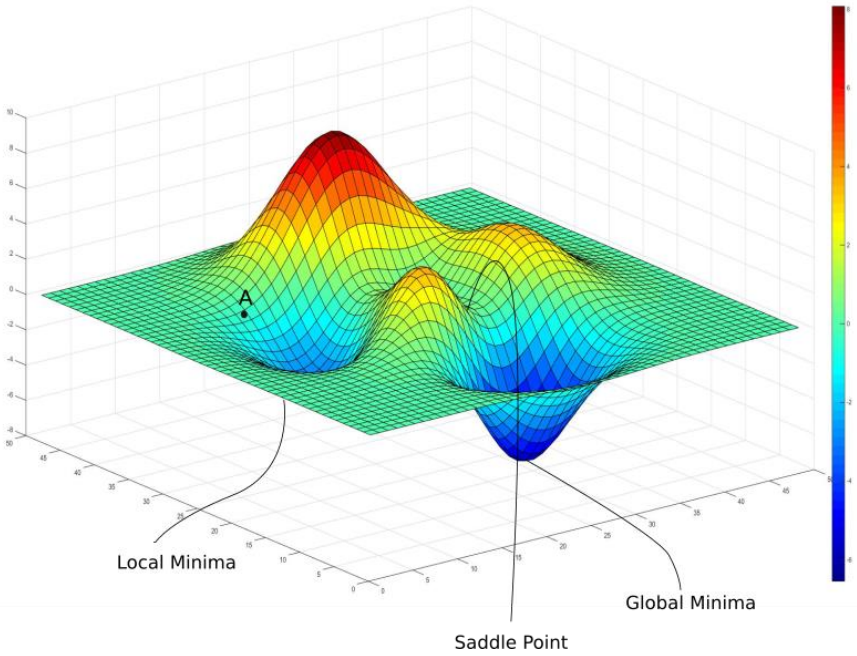


Figure 21: Error space of GD algorithm where the goal of the GD is to reach the global minima by updating calibration parameters

New coefficient = coefficient – (learning rate * slope) + momentum* previous slope…….... (16) Formula to update the coefficient in GD algorithm

3.4.3. Multi-Hazard Model Evaluation

This section will further explain the methods used to evaluate the results obtained from multi-hazard modelling and will show if and how Super-Resolution can improve the quality of multi-hazard modelling. All the methods used in this section are considering all hazards simulated, such as the Cohen's Kappa will use combined slope failure and flood to calculate the Kappa statistics. The root mean square error will use the combined height of solid and fluid to calculate root mean square, and in the case of NS-Index, the fluid at outpoint is mostly affected by slope failures in the catchment, so NS-Index will account for that effect. Therefore, even though it is written as fluid height, it means solid and fluid combined height.

3.4.3.1. Cohen's Kappa

To evaluate the accuracy of the modelled flow (of solid and fluid) extent in a certain region, we need to compare it with the observed flow extent. One of the most used methods to evaluate the reliability of the output from the model is Cohen's kappa statistics (Blackman & Koval, 2000). This measure considers not only the correct predictions (pixels true positive and true negative in this case) but also the incorrect predictions making it a robust measure of agreement between datasets. The Cohen's kappa will be calculated using the pixels with the maximum combined height of solid and fluid above 0.25 meters and the extent of (solid and fluid) flow observed from the satellite right after the modelled event. This threshold of height is defined in the Continuous Cohen's Kappa function developed on OpenLISEM software which can take different threshold value and calculate the Cohen's Kappa for modelled values above those thresholds. This functionality is specifically designed for evaluating the results from OpenLISEM. The kappa statistics were computed using equation 17, where we can observe that the observed agreement and probability of random agreement are used to calculate the Cohen's Kappa. The probability of observed agreement is calculated by the total correct classification divided by the total number of pixels. In our case, it was calculated using the sum of true positive and true negative divided by the total pixels. The total number of pixels having a maximum combined solid and fluid height above 25 cm that have a hazard in the inventory are true-positive, and the pixels that had the maximum combined height less than 25 cm and did not have any hazard in observed data are true negative. The probability of random agreement indicates the model predicting the affected pixels by random chance. To calculate that, we sum the probability of correct classification and probability of incorrect classification as shown in Equation 18. The probability of correct classification and incorrect classification is calculated using the formula in Equation 18 for confusion matrix as shown in Figure 22, where the green arrow shows true positive and true negative. Followed by kappa statistics, a critical evaluation will be done to evaluate whether the Super-Resolution model significantly impacts the model's output.

$\kappa = \frac{P_o - P_e}{1 - P_e}$ ..... (17) Cohen's Kappa statistics for inter-rater reliability measurement, P_o denotes the observed agreement between the modelled and observed data, and P_e is the probability of random agreement

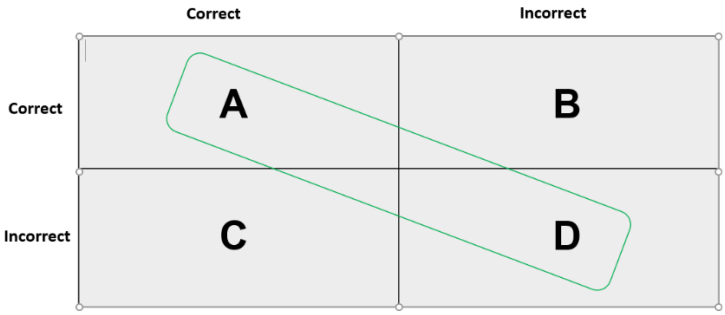


Figure 22: Confusion matrix for correct and incorrect classification of impacted pixels. Source: Pykes (2020).

$$P_o = (A + D)/(A + B + C + D)$$
$$P_{correct} = \left(\frac{A+B}{A+B+C+D}\right) * \left(\frac{A+C}{A+B+C+D}\right)$$
$$P_{incorrect} = \left(\frac{C+D}{A+B+C+D}\right) * \left(\frac{B+D}{A+B+C+D}\right)$$
$P_e = P_{correct} + P_{incorrect}$ ……...(18)Equations to calculate the probability of observed agreement and random agreement. Source Blackman & Koval (2000).

3.4.3.2. Root Mean Squared Error

For comparing the (combined solid and fluid) flow height outputs generated using different datasets, we use the Root Mean Square Error of the difference between the observed and modelled one. Unfortunately, there are no observed solid and liquid heights during this event, so the modelling results with the HR DEM are used to represent the "correct" values. Hypothetically, the OpenLISEM should generate the same flow height for a certain multi-hazard event with the same triggering event, but it will not produce the same due to the different resolution and quality of the data. To compare and evaluate whether downscaled data significantly improves flow height estimation than that of the low-resolution data, we will calculate RMSE for each output using high-resolution data as baseline data. The high-resolution data is considered baseline data because usually, we do not have the measured/observed flood height maps in many cases. So, in this case, we will compare the quality of output using downscaled data to that of the low-resolution data.

3.4.3.3. Nash–Sutcliffe Efficiency Index

The Nash-Sutcliffe Efficiency Index is a popular statistical method of accessing the goodness of fit of the hydrological models (McCuen et al., 2006). This statistical method calculates the goodness of fit of the hydrological model by comparing it with the ground truth data to get the magnitude of residual variance (Nash & Sutcliffe, 1970). In our case, we will compare the goodness of fit of the discharge calculated by the OpenLISEM hazard model at the channel outlet for different DEM input data. Since the past event data is not available, a relative comparison will be done with the model output from high-resolution data. The Nash-Sutcliffe efficiency index is computed using Equation 19 defined by Nash & Sutcliffe (1970). We will compare the performance of flow propagation quantitatively by comparing the discharge output from the model outlet for different scenarios and conclude the usability of Super-Resolution for multi-hazard modelling.

$NSE = \left[\frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{mean})^2}\right]$ …….. ……. (19) Nash-Sutcliffe Efficiency Index equation. the indication Y^{obs} represent the observed discharge, and Y^{sim} represent the simulated discharge (Moriasi et al., 2007)

After all the fit-for-purpose tests and the quantitative evaluations, we will conclude either the Super-Resolution models are useful in multi-hazard modelling or not.

4. RESULTS AND ANALYSIS

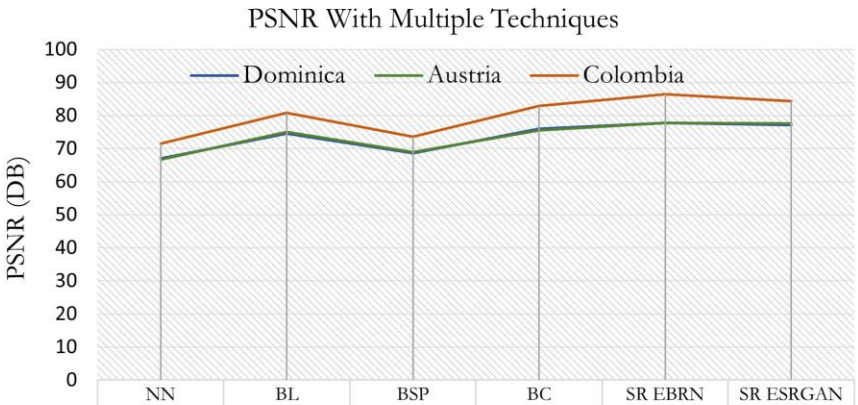
After comprehensive Super-Resolution training and Multi-Hazard modelling, the results were obtained for further analysis. The results obtained from the study and their analysis are presented in this chapter, categorized by each phase of the research as defined in the methodology section.

4.1. Phase I: Deep Learning Super-Resolution Results

The results from SR models are shown and analysed in this section. The testing represents the use of high-resolution DEM, and the low-resolution DEM obtained using bicubic interpolation from high-resolution DEM. Results shows the capacity of the model to generate SR images. To check the model's applicability with the global free LR dataset, we then further evaluate SRTM data without having the HR counterpart available from the same sensor; this result will be explained in the next section. Explanation of those evaluation metrics are already explained in the methodology chapter.

4.1.1. PSNR and MSE Analysis

The PSNR obtained for the test area in Austria and the inference area in Colombia and Dominica where High-Resolution DEM (mostly LiDAR) were available are represented in Figure 23. As we can observe, both EBRN and ESRGAN based methods have improved the PSNR values compared to Bicubic Interpolation in all the study sites. We did this comparison based on the practice used in most of the Super-Resolution studies, where the low-resolution counterpart of HR data is obtained through bicubic interpolation. Then a quality assessment is done compared to those HR data with SR data obtained from different algorithms using LR data. As we can observe in Figure 23, in all study sites, we can observe that the Super-Resolution with DL techniques has superior results compared to other methods.

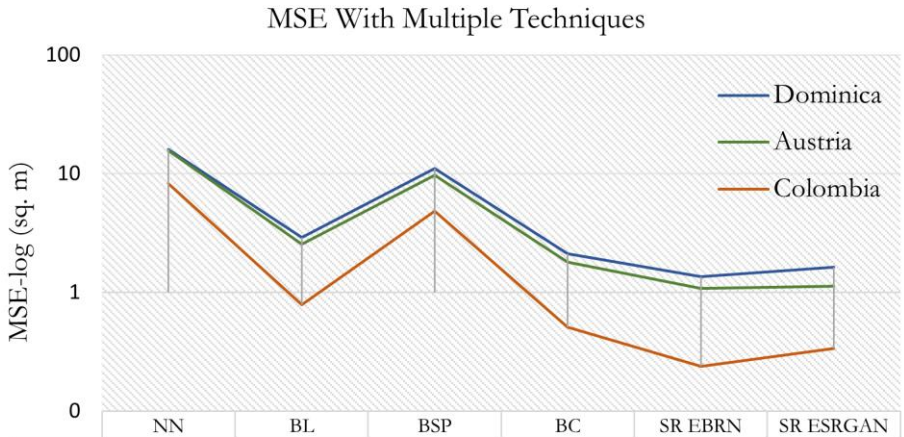


	NN	BL	BSP	BC	SR EBRN	SR ESRGAN
Dominica	67.027	74.506	68.617	75.943	77.789	77.084
Austria	66.752	75.051	68.848	75.500	77.734	77.531
Colombia	71.428	80.802	73.661	82.889	86.443	84.336

Figure 23: PSNR with different interpolation and Super-Resolution techniques.

PSNR being a logarithmic scale, the amount of improvement in such scale is difficult to perceive; the increase of 1DB of PSNR in the lower range of the data is not similar to that of 1DB increase in the higher range. So, to make an absolute comparison mean squared or mean absolute error and improvement in reducing such error are a better way of comparison. So, we present the findings as mean squared error in Figure 24. As we can observe in Figure 24, the MSE is decreased by a significant amount in both EBRN and ESRGAN at all study area with Super-Resolution techniques; also, we can observe that the reduction

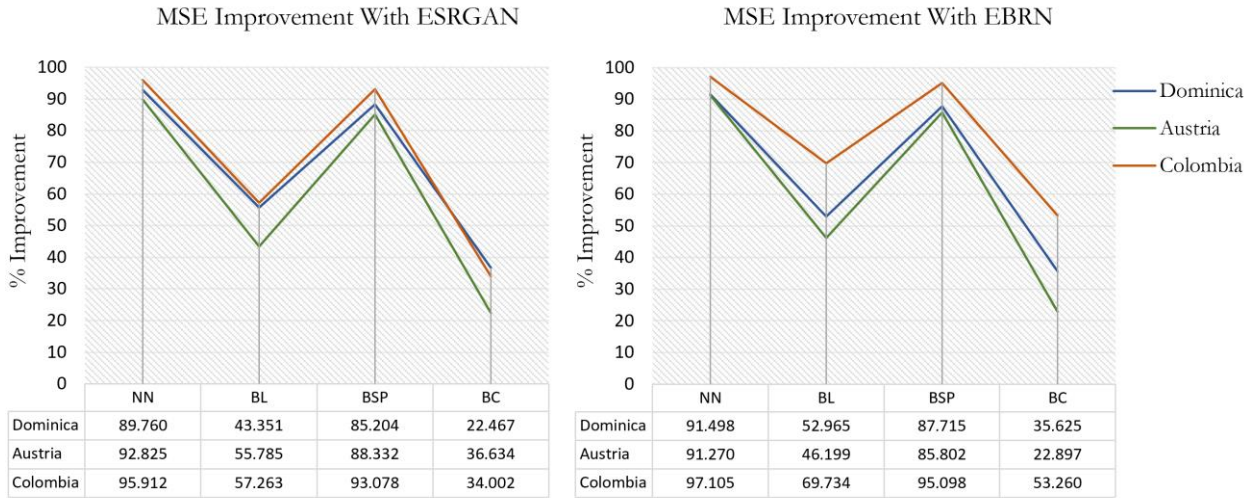
of MSE from NN to BL is very high and from BL to BC is lower, and BC to SR methods are lowest, this is because once the accuracy is high, it is more difficult to improve the quality of data due to the fact that there is much less room available for further improvement. Furthermore, all the curves have shown the very same pattern on the reduction of MSE, showing that the model can perform similarly with different amounts of noise present in the LR data. However, MSE and PSNR techniques only provide the overall alignment of the elevation with high resolution. However, these quantitative methods cannot provide information on whether there is enough spatial variability increased or how the important features such as ridgelines and valleys are conserved/improved or not, and those properties of the improvement will be looked at in the following sections.



	NN	BL	BSP	BC	SR EBRN	SR ESRGAN
Dominica	15.993	2.891	11.069	2.112	1.360	1.638
Austria	15.750	2.556	9.684	1.8036933	1.0784966	1.130
Colombia	8.227	0.787	4.859	0.510	0.238	0.336

Figure 24: Mean Square Error in the different study areas with different interpolation and SR techniques. Y-axis in log scale for better representation.

To see relative improvement of the SR techniques compared to different interpolation techniques, we have computed the relative percentage improvement, which provides how much the MSE is improved compared to that method, and the results are represented in Figure 25. As we can observe in Figure 25, the improvement of SR techniques compared to all interpolation techniques, including bicubic, has more than 20% increase in all the cases.



	NN	BL	BSP	BC
Dominica	89.760	43.351	85.204	22.467
Austria	92.825	55.785	88.332	36.634
Colombia	95.912	57.263	93.078	34.002

	NN	BL	BSP	BC
Dominica	91.498	52.965	87.715	35.625
Austria	91.270	46.199	85.802	22.897
Colombia	97.105	69.734	95.098	53.260

Figure 25: MSE reduction with ESRGAN and EBRN model compared to different interpolation techniques.

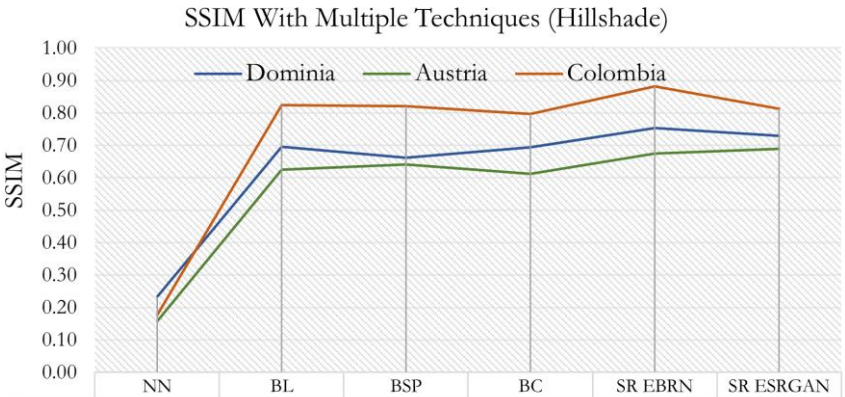
4.1.2. SSIM

Another evaluation metric most extensively used by computer vision scientists in Super-Resolution evaluation is the SSIM. In our research, to understand how the model can reconstruct the quality of the image, we computed the SSIM for all study areas and models. The results from the SSIM study are shown in the following Table 9. As we can observe, the EBRN and ESRGAN model have improved the image's visual quality, but the SSIM value has been already saturated with BC interpolation, which shows an almost perfect SSIM value. There is an improvement with SR methods, but because of already saturated SSIM, it is difficult to observe the improvement. This issue is further discussed in the discussion section for a reason behind the saturated performance of SSIM in DEM data. Maybe due to that, most of the DEM SR research has not included SSIM in their research (Jiao et al., 2020; Kubade et al., 2021; Shin & Spittle, 2019; Xu et al., 2019). Wu & Ma (2020) have represented the SSIM results of their work, and they also have faced a similar problem; the bicubic interpolation already has SSIM of more than 0.9999, making the improvement visible in the 1e-5 range. While comparing the increase in SSIM from Bicubic to SR-based approach, we have higher improvement, but we cannot make any quantitative comparisons because the testing data are from different locations.

		Study Area		
		Austria	Colombia	Dominica
Improvement Technique	NN	0.999734	0.999860	0.999725
	BL	0.999974	0.999995	0.999959
	BSP	0.999971	0.999991	0.999858
	BC	0.999976	0.999995	0.999959
	SR EBRN	0.999984	0.999997	0.999979
	SR ESRGAN	0.999983	0.999996	0.999975

Table 9: SSIM with different interpolation and SR techniques in different case study sites

To overcome this analysis bottleneck that occurred due to the SSIM method not being able to differentiate much in the case of the DEM, we further developed an approach where we converted our elevation data to hill shade with the constant azimuth of 315 degrees and altitude of 45 degrees. This allowed us to convert the range of data from an unlimited range to 0-255, which makes it more relevant for SSIM and also it can perform in similar ways as human comparing the different Hillshade images, similar to SSIM in natural and medical images(Renieblas et al., 2017; Z. Wang et al., 2004). The results of SSIM in Hillshade images are shown in Figure 26, where we can observe that the SR-based approaches, specifically EBRN, have significantly improved the similarity of images to that of the HR images. Furthermore, more interesting is to see that, in the case of hillshade Bilinear and Bicubic Spline interpolation methods perform better than bicubic interpolation, which represents that for better visualization, Bilinear Interpolation techniques are better and for more accurate values, Bicubic Interpolation techniques are better. More interestingly, the EBRN model has shown similar characteristics in all regions and study area, but ESRGAN has performed better than that of EBRN in Austria, where it was trained but performed less in other study areas; this behaviour is due to the architecture of the model and its quality, and it will be further discussed in the discussion chapter. Since this is a new approach we used to compare the results, and previous papers have not done such SSIM evaluation with derivatives, we cannot compare our model with others in terms of visual quality improvement.



	NN	BL	BSP	BC	SR EBRN	SR ESRGAN
Dominia	0.23250	0.69599	0.66187	0.69365	0.75274	0.72911
Austria	0.15710	0.62513	0.64078	0.61172	0.67433	0.68893
Colombia	0.17689	0.82406	0.82036	0.79619	0.88110	0.81250

Figure 26: SSIM in hillshade Images obtained from different techniques.

To put things into more perspective, we have included some visual samples of the ESRGAN and EBRN quality improvement with LiDAR data in Dominica in Figure 27. As we can observe, both EBRN and ESRGAN models have increased the visual quality of DEM while improving degraded low-resolution LiDAR DEM (by bicubic interpolation) to high-resolution counterpart significantly over the bicubic method, which fails to generate a crisp looking image with clear ridge and drainage lines. The ESRGAN and EBRN model have very similar results even though in terms of parameters, EBRN is big and complex while in terms of training complexity, ESRGAN is more complex; this phenomenon can also be observed in SSIM values that they have very similar SSIM, ESRGAN being slightly better, but with the human eye, that difference is negligible.

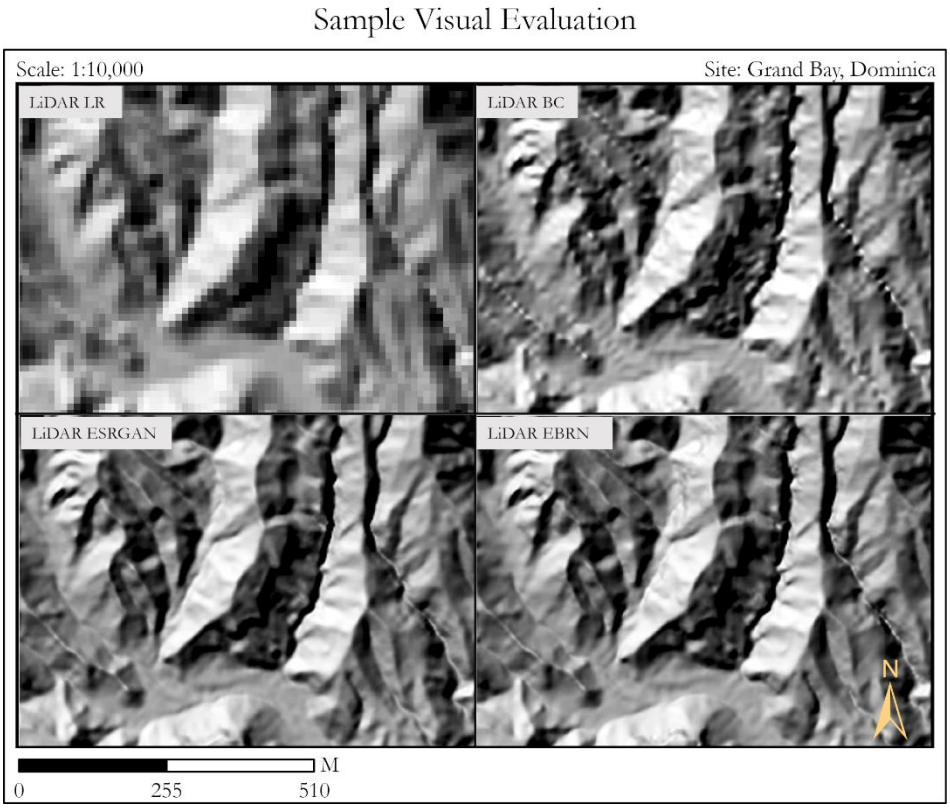


Figure 27: Sample visual evaluation of LiDAR DEM Super-Resolution in Dominica test area.

4.2. Phase II: Deep Learning Super-Resolution Results

This section will first test the applicability of Super-Resolution in generating better derivatives and geomorphological mapping followed by multi-hazard modelling results analysis.

4.2.1. Super-Resolution Evaluation in Global Dataset

In this section, we evaluate the capacity of Super-Resolution to reconstruct the global dataset and its derivatives, as well as visual and geomorphological evaluation. This evaluation section focuses more on the applicability in modelling and mapping than technological improvement, whereas the previous section was more focused on technological development. This part will deal with the applicability of Super-Resolution in the domain of geoscience and natural hazards modelling in specific.

4.2.1.1. Derivative Analysis

In the derivative analysis, to understand how the global free DEM has been improved compared to that of the high-resolution DEMs in terms of the DEM derivatives, we estimated the KDE function for all the available datasets and plotted against high-resolution DEMs in all three study areas in Figure 28. As we can observe in subplot [1,1] (to explain better their position row and column represent the sub-images), the slope in Dominica, the bicubic interpolation (red line) has its peak in a bit below the peak of the EBRN, and ESRGAN methods which are performing very similar, we can also observe that the TanDEM-X have its slope distribution closer to that of the high-resolution DEM. Furthermore, we can observe that the improvement in ESRGAN and EBRN model is low, but we can observe in all three study sites that the distribution of ESRGAN and EBRN are more like HR than that of the Bicubic Interpolation method. If we observe the case of Austria, the bicubic interpolation have a higher amount of pixels in 1.0 to 1.2 radians, but a lower amount of pixels is present in the 1.2-1.5 radians range. In contrast, HR DEM has a higher number of pixels in those regions showing that Bicubic Interpolation have more smooth data, and ESRGAN and EBRN have tried to improve that to generate more pixels with the higher slope as shown by the upper peak.

In the case of Aspect, we can observe that all the aspects except the TanDEM-X are aligned; this can occur due to the quality and sensor of the data. What makes it more interesting is that the aspect has not much impact on any interpolation techniques, and it is mostly in alignment with the high-resolution DEM, especially in Austria and Colombia. This can be due to the fact that any interpolation techniques are barely going to change the overall direction of the terrain so that we can see very little change in aspect in all SRTM DEM. As we can observe in [2,2] subplot, the SR-based techniques have very slightly lower curves near the peak, making it more towards the high-resolution data, but that improvement is not that significant.

For TWI, we can see that in sub-image [3,3], the TWI from EBRN has almost perfectly aligned with the HR data, whereas the bicubic interpolation is less aligned, followed by ESRGAN TanDEM-X. The improvement, in this case, is very significant. However, in the case of Dominica, the TWI are mostly clustered for each dataset, and in Colombia, even though all the values are clustered together, we can observe that the ESRGAN and EBRN models are nearer to HR data than that of the BC data. The case in Dominica is more interesting because the HR-DEM that we have available in Dominica do not cover the major mountainous part and mostly present in the seashore and relatively flatter areas which might have caused such peak and clustering. The further reasons for this behaviour in Dominica will be discussed in the discussion section.

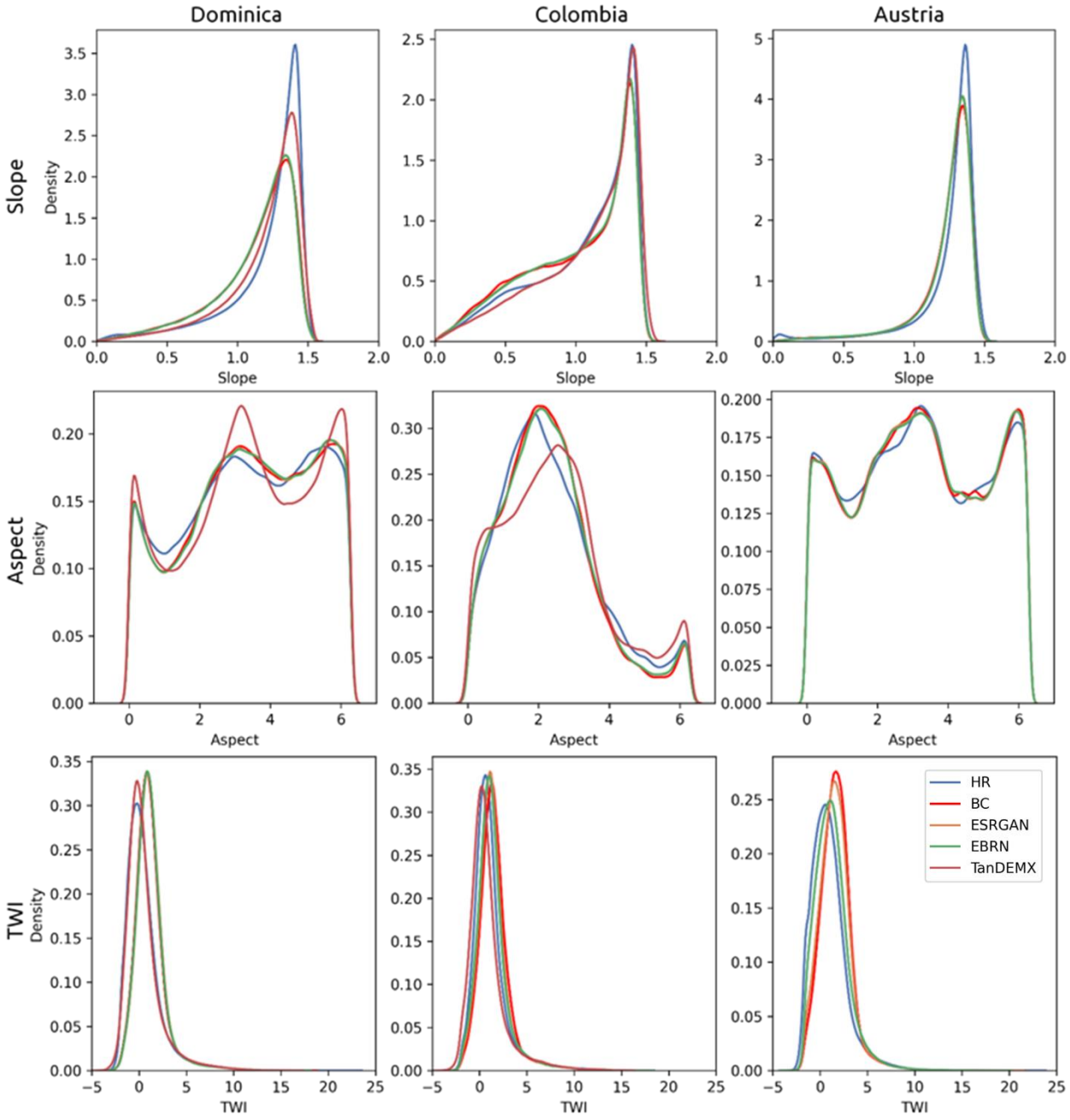


Figure 28: KDE estimation of multiple DEM and their derivatives

4.2.2. Visual and Geomorphological Analysis

The PSNR and SSIM approaches can be used when we have the HR data available, but the goal of this research is to use the model trained with HR data in one part of the world to use it in another part of the world where the model can be used with low-resolution data. To observe the improvement while using the different source data, we used the visual analysis method, where we visually analysed the quality of SR images compared to Bicubic Interpolation. Figure 29 and Figure 30shows the sample cases from Austria, Dominica, and Colombia with different terrain characteristics in different map scales. As we can observe in the upper image of Figure 29 (Colombia), the EBRN has improved the ridge and valley lines in those areas compared to SRTM low-resolution and Bicubic Interpolation techniques making the streams and valleys more visible and crisper looking. However, the ESRGAN based approach has also increased the number of artefacts in the image, making it more unpleasant and has a better and crisp-looking image than bicubic interpolation.

In the case of the Austria DEM, where we trained all our models with High-Resolution DEM, we can see significant improvement in the quality of hillshade in EBRN and slightly less improvement in ESRGAN compared to that of the EBRN. If we observe the SSIM evaluation in Austria in the previous section, we can see that ESRGAN has better performance than that of the EBRN in the case of HR DEM, but here by visual observation, we can see that when used with SRTM, EBRN has produced better-looking image than that of the ESRGAN. This might be due to the difference in model characteristics and capacity of the model, but with global freely available data in Austria, the EBRN model shows better visually pleasing results.

In the case of Dominica, we have included a steeper slope and less steep slope as two examples, and we can observe that for Dominica, both models have very similar performance in both steep and non-steep slopes. The EBRN model has a crisper-looking hillshade image in this case, but it is not a significant improvement over the other image. We can also observe that the lower image that we selected intentionally, a noisier DEM, has reduced the amount of noise in the SR-based methods compared to that of the LR and BC. When there are simple features and ridgelines and valleys, the model has performed better reconstructing them and generating more crisp images, but when there is a lot of sudden change in features and elevation, the model generates smoother images, as we can see in the lower image, but it is still better than that of the BC and LR DEM.

Overall, in other areas as well, we could observe that the model has improved the visual quality of the DEM, but due to the lack of a high-resolution counterpart from the same source, it is not possible to check the improvement in terms of PSNR. However, with visual assessment, the DEM produced from Super-Resolution techniques have better visual quality and better reconstruct the features. Especially, the EBRN model has performed similarly in all locations and datasets in the reconstruction of the images, but the ESRGAN model has produced better results with HR DEM but has lesser performance than that of EBRN with SRTM data. Due to the very low resolution of Input data (SRTM DEM), the features such as road are not generated and are not recognizable because the SR based methods are good at reconstructing the features they have learned during the training. However, it requires some low-resolution/noisy input, and if the feature is not visible in input SRTM DEM, it cannot be available in the SR images as well. This phenomenon satisfies the theory of data processing inequality, which instructs that any post-processing operations in data cannot increase the information content (Beaudry & Renner, 2012). So, with Super-Resolution techniques, we can make images more visibly plausible and pleasing and increase the sharpness of the image but cannot add new features that are not existing in the original image.

From the review of Geomorphological experts, after they evaluate the DEM produced with our approach and the quality of commercial DEMs, we present the average score in Table 10. As we can observe in the table, The SRTM low-resolution has the lowest score, followed by bicubic interpolation and EBRN. The evaluation was more targeted for detection of landslides and landforms, so, obviously, HR high-resolution DEM has the highest score, and as we stated before, the data processing inequality also holds in this evaluation because the data collected by TanDEM-X have higher information content available, the recognizable features in the TanDEM-X data is also higher (Beaudry & Renner, 2012).

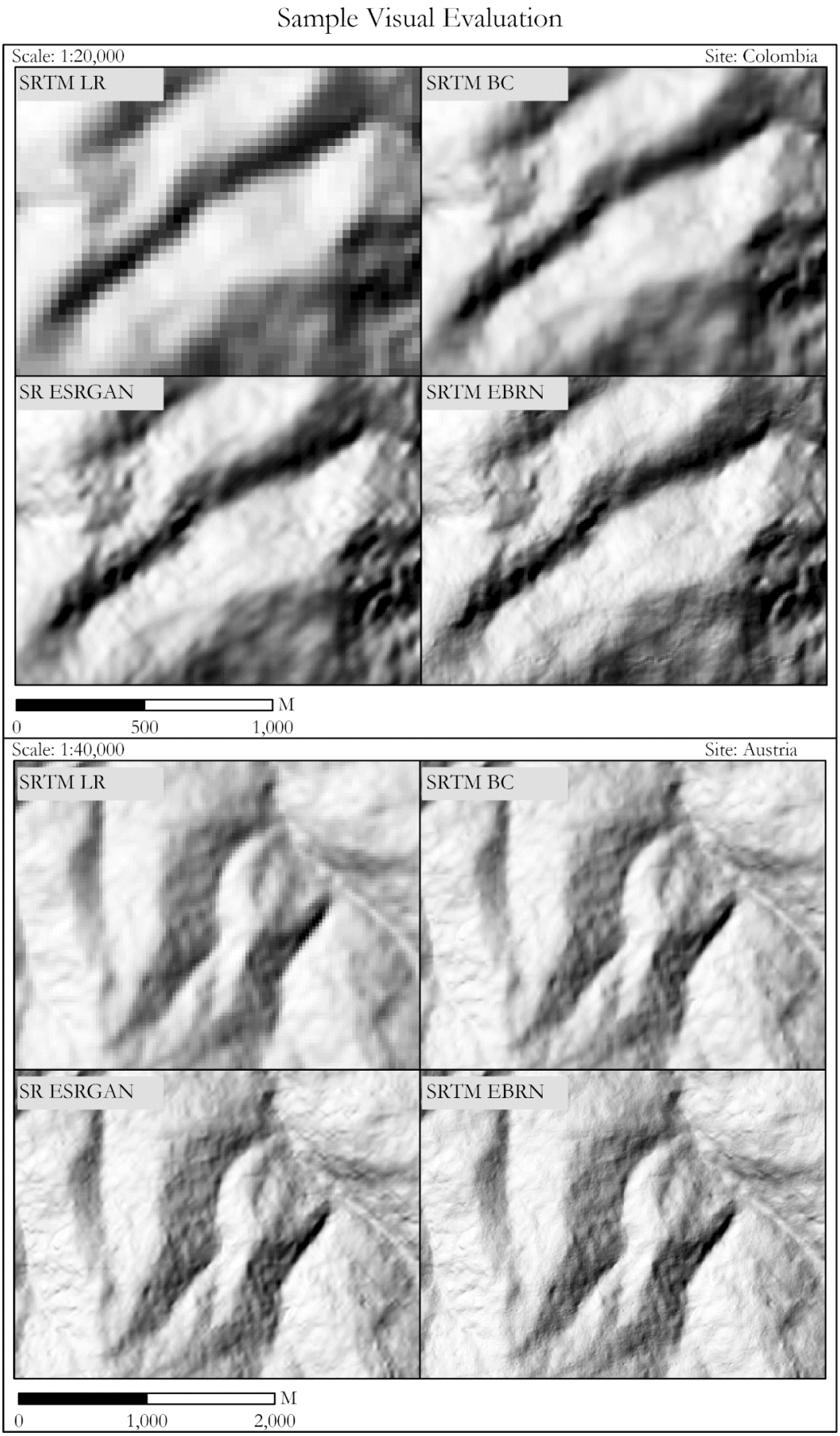


Figure 29: Visual Evaluation of SR and BC methods in Colombia and Austria

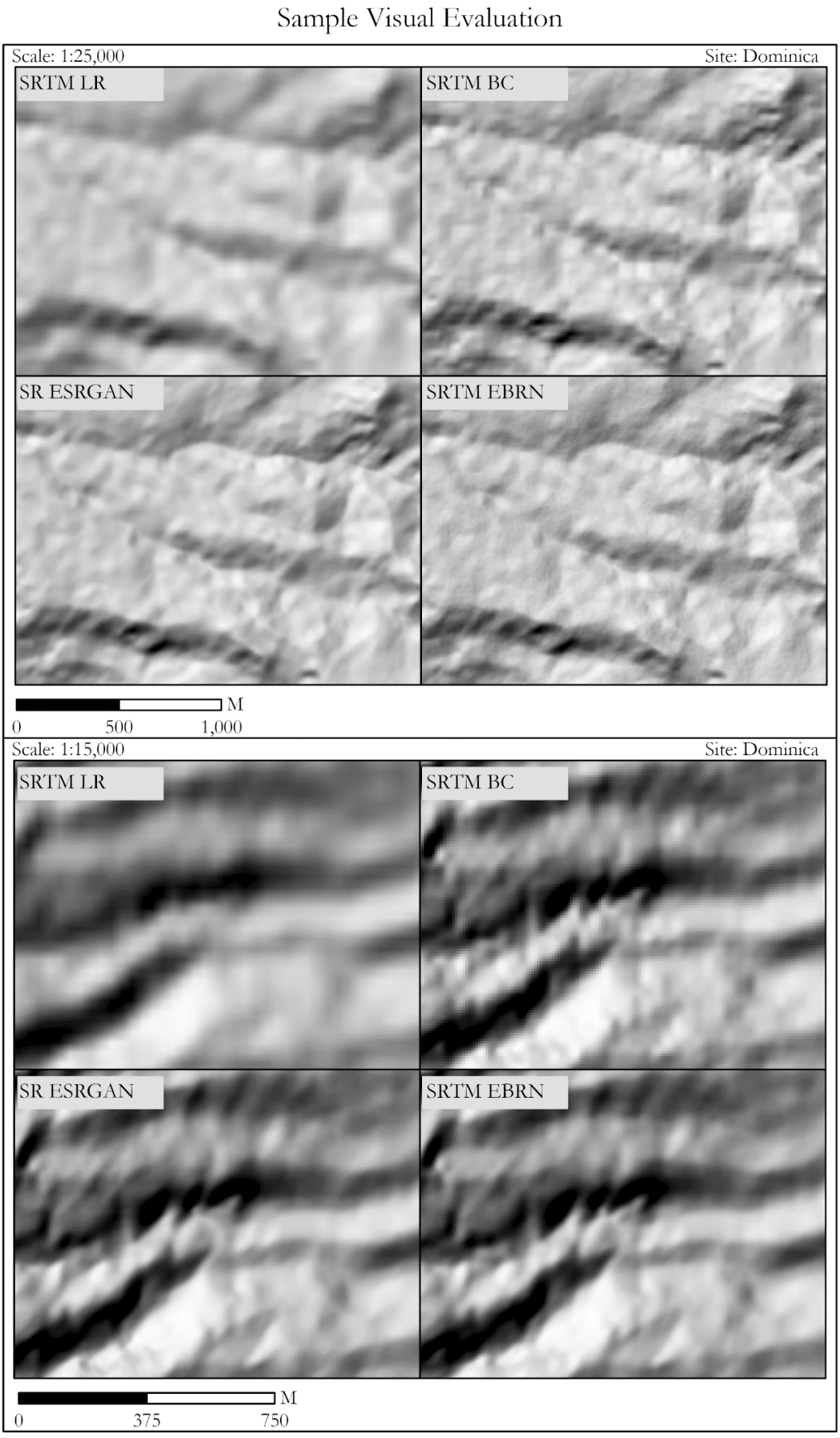


Figure 30: Visual Evaluation of SR and BC methods in Dominica

Interestingly, on average, the ESRGAN method has better performance in Dominica, whereas the EBRN method has better performance in both other regions. However, one of the reviewers scored EBRN as the top performer in all three sites and ESRGAN as equal in Dominica, but another reviewer scored ESRGAN is the top performer in all the regions making the evaluation more subjective. In general, the ESRGAN has better performance in the image's visual quality and should be valid because the GAN-based approach is better known for its visually pleasing images (X. Wang et al., 2018). Based on the training data, the EBRN has performed best in Austria, where it was trained, but ESRGAN has shown comparatively better results in Dominica, but because this is relative scoring per site, we cannot conclusively claim that it has performed better in inference data than in test data.

		DEM Generation Methods					
		SRTM Low Resolution	SRTM Bicubic Interpolation	SRTM SR EBRN	SRTM SR ESRGAN	TanDEM-X	HR DEM
Test Site	Dominica	0.75	2.5	3.5	4	4	8.5
	Colombia	0.75	2.5	3.5	3.5	5	8.25
	Austria	0.75	3	4	3.75	NA	9.5
Average Score		**0.750**	**2.667**	**3.667**	**3.750**	**4.500**	**8.750**

Table 10: Geomorphological score of different methods for DEM resolution improvement.

Because all the review was done with the geomorphological point of view and the resolution of DEM were fixed to 7.5 meters (by resampling all images), the reviewers found that In Dominica, LiDAR has the best performance even though due to resampling, many features were not visible furthermore they also commented that the TanDEM-X has highly mottled/speckled structure. Furthermore, the SRTM low-resolution was very coarse and had rounded terrain forms, and both EBRN and ESRGAN had smoother surfaces, but landslide detection was still not possible. In the case of Colombia, they had reviewed that, similar to Dominica, landform recognition is very difficult at that scale, but SRTM EBRN has a more natural-looking hillshade. However, they have commented that it is completely difficult to recognize the landslides in the image. In the case of Austria, the LiDAR DEM had a very good image where all geomorphological features were available, but in the case of other DEM, it was very difficult to recognize those features and interpret them.

In conclusion, the reviewers have suggested that SR images were not better in terms of recognizing the geomorphological features, and also the pixel size of 7.5 meters are too large for any good quality recognition of the landslide features, making it difficult to use in geomorphological mapping of medium and small-scale landslides from global freely available DEMs such as SRTM.

4.2.3. Multi-Hazard Modelling Evaluation

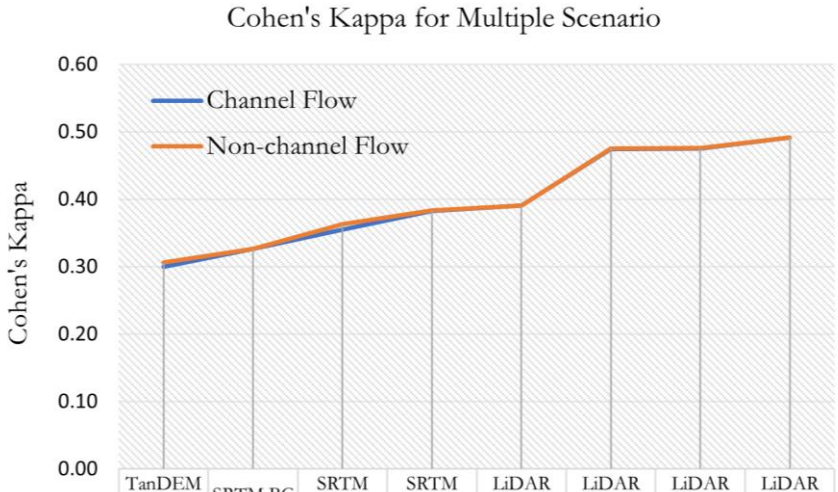
In this section, we evaluated the results from multi-hazard modelling. Results are presented for each of the case study sites Dominica and Colombia. Each case study site will explain the results obtained in that study area. Dominica has more extensive analysis to evaluate whether Super-Resolution is useful in multi-hazard modelling or not, and Colombia has fewer scenarios that only test the applicability in multiple scenarios locations.

4.2.3.1. Case Study Site: Dominica

After simulating all the scenarios as defined in the methodology section, we further analysed the results obtained from the simulation. Firstly, we computed the continuous Cohen's Kappa statistics which we made a comparison with the inventory dataset available, and this evaluation section is the only absolute comparison we can make due to the lack of other data such as discharge and flood height in Dominica. The continuous Cohen's Kappa values for different scenarios (1-16), including and excluding channel flow, are shown in Figure 31. Before calculating the Cohen's Kappa value, every scenario was calibrated for 20 iterations using a gradient descent algorithm and based on the best values with the calibration; we obtained the modelling results.

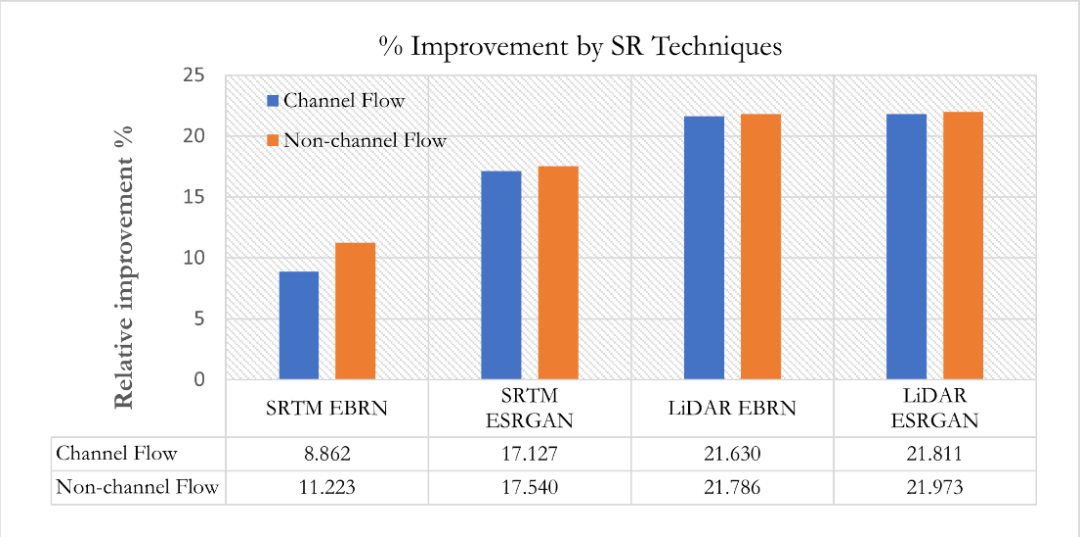
As we can observe in the image, the least performing Cohen's Kappa value is from TanDEMx data, followed by SRTM BC then Super-Resolution techniques. Furthermore, in both the channel and non-channel flow cases, the kappa value has been increased by using Super-Resolution techniques. To clarify the increased accuracy, the percentage improvements are calculated relative to bicubic interpolation. The relative improvement using ESRGAN and EBRN methods in modelling compared to bicubic interpolation is shown in Figure 32. As we can observe in Figure 31and Figure 32, the amount of improvement in Cohen's kappa is very similar in all cases; SR-based methods have slightly higher Cohen's Kappa in Non-Channel flow that difference is insignificant.

The improvement by both ESRGAN and EBRN is mostly the same except for the case of SRTM data, where ESRGAN has significantly higher performance. This shows that ESRGAN has more improvement in reducing noise and generating better surface in the case of SRTM data as well, but in the case of EBRN, the improvement is limited in low-resolution data because of the non-generative nature of the model. In conclusion, both ESRGAN and EBRN have improved the modelling accuracy by 11%-21% in the case of non-channel flow and by 8-21% in the case of non-channel flow, which is a significant improvement compared to bicubic interpolation. Moreover, the improvement of the modelling output is largely affected by the presence of noise in the low-resolution data.



	TanDEM X -HR	SRTM BC	SRTM EBRN	SRTM ESRGAN	LiDAR BC	LiDAR EBRN	LiDAR ESRGAN	LiDAR HR
Channel Flow	0.300	0.326	0.355	0.382	0.390	0.475	0.475	0.492
Non-channel Flow	0.306	0.326	0.363	0.384	0.390	0.475	0.476	0.492

Figure 31: Cohen's Kappa for different modelling scenario



	SRTM EBRN	SRTM ESRGAN	LiDAR EBRN	LiDAR ESRGAN
Channel Flow	8.862	17.127	21.630	21.811
Non-channel Flow	11.223	17.540	21.786	21.973

Figure 32: Relative improvement compared to bicubic interpolation techniques by using SR methods

To observe the results visually, the map of modelling output from ESRGAN, bicubic, high-resolution and TanDEM-X are shown in Figure 33; we did not include the EBRN case because it has very similar improvement and results compared to ESRGAN in most of the cases. The results are shown in the extent where most of the hazardous process was occurring. Furthermore, the results are also shown only for Non-Channel flow because output results were very similar, and they were visually indifferentiable. We can observe that SR-based methods have improved the amount of flow spread in multiple pixels compared to bicubic methods, making the model more realistic. However, because the output data is largely dependent on the input in SR based methods, the flow direction is not much improved, making it slightly non-aligned with the inventory data. Furthermore, the amount of slope failures is very low due to the automated calibration, and the reasons for that will be further discussed in the discussion section for multi-hazard model calibration. If we observe the TanDEM-X results due to the presence of the noise in DEM, there are many smaller pixels filled with water; even though the sinks were filled before the modelling process, the overall improvement is not much with sink filling. The overall extent of hazard (combined flood and solid) is not much aligned, also represented by the Cohen's Kappa values. In LiDAR, even though the improvement in the upper part of the catchment is difficult to visualize, In the flatter area, we can observe that the SR-based method has reduced the amount of false positive by a significant amount very similar to that of the high-resolution data.

To have a more general understanding of the overall performance of the dataset, we also compared the total channel outflow for the modelling scenarios 1-8, which had channel flow included, as shown in Table 11. This evaluation is not an absolute evaluation because we do not have the ground truth data from the event, but to have a general idea, we considered LiDAR HR DEM as the best performing DEM and calculated the relative difference in total channel flow. Compared to LiDAR HR, the absolute error is maximum in the case of the TanDEM-X, the reason behind that might be due to the noise and speckle presented in the TanDEM-X data, but further discussion will be done in the next chapter. Interestingly, in LiDAR ESRGAN, even though Cohen's kappa value was maximum, the difference in overall channel flow is higher than that of LiDAR BC, but the difference between them is very small, which might have also occurred due to other factors. Furthermore, in the case of SRTM, the total channel inflow has increased using SR DEM, making it more equivalent to that of the LiDAR HR DEM, and the improvement is in line with Cohen's kappa values.

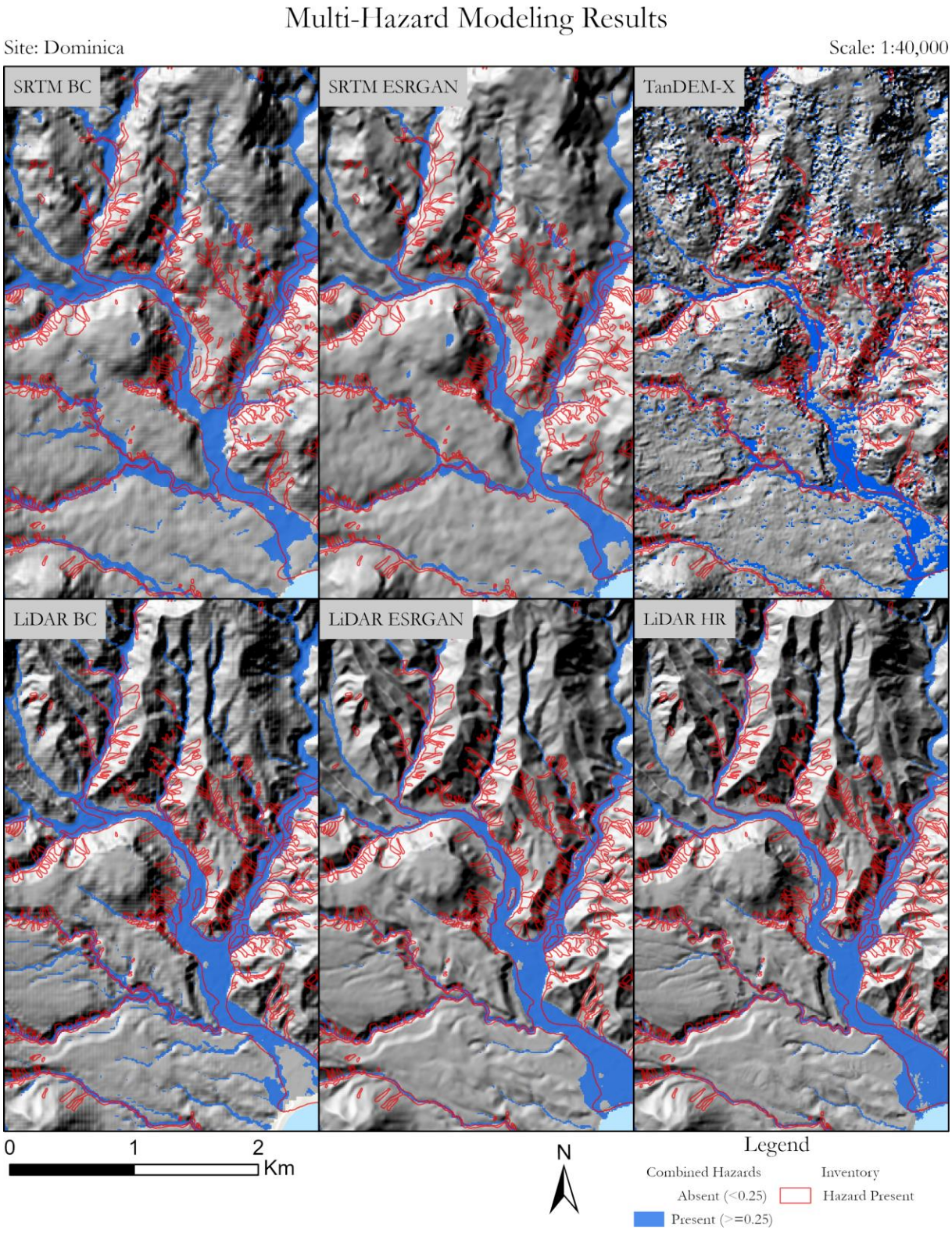


Figure 33: Multi-Hazard modelling results in Dominica; most relevant cases are shown in the map.

Dataset	Total Channel Flow (mm)	Absolute Error relative to LiDAR HR (mm)
LiDAR EBRN	296.977	7.675
LiDAR BC	312.494	7.842

Dataset	Total Channel Flow (mm)	Absolute Error relative to LiDAR HR (mm)
LiDAR ESRGAN	296.620	8.031
SRTM EBRN	281.190	23.461
SRTM ESRGAN	278.879	25.773
SRTM BC	274.360	30.291
TanDEM-X	357.503	52.852
LiDAR HR	304.651	

Table 11: Relative comparison of total channel flow with different DEM dataset with LiDAR HR DEM

To further understand how different approaches have produced the different flood height and compare them, we tried to compare the flood heights using RMSE. However, before that, we found in data from different sensors that the exact location of pixels might have different features showing a shift in elevation data and relatively comparing them pixel-by-pixel without any ground truth data can cause serious problem. To further clarify the situation, Figure 34shows one red point, which is the same point in both DEM, but as we can observe in both DEMs, the same point lies on the left side of the hill in one LiDAR DEM and on the right side of the SRTM DEM. In this case, if we compare the values on a pixel-by-pixel basis without having the ground truth, the comparison will produce irrelevant data. Since we did not have any high-resolution SRTM data so to check whether the SR based methods can improve the quality of height and outflow from the channel compared to bicubic interpolation methods, we computed the Nash Sutcliffe index at a certain outlet point and also the RMSE in each pixel for the whole study area.

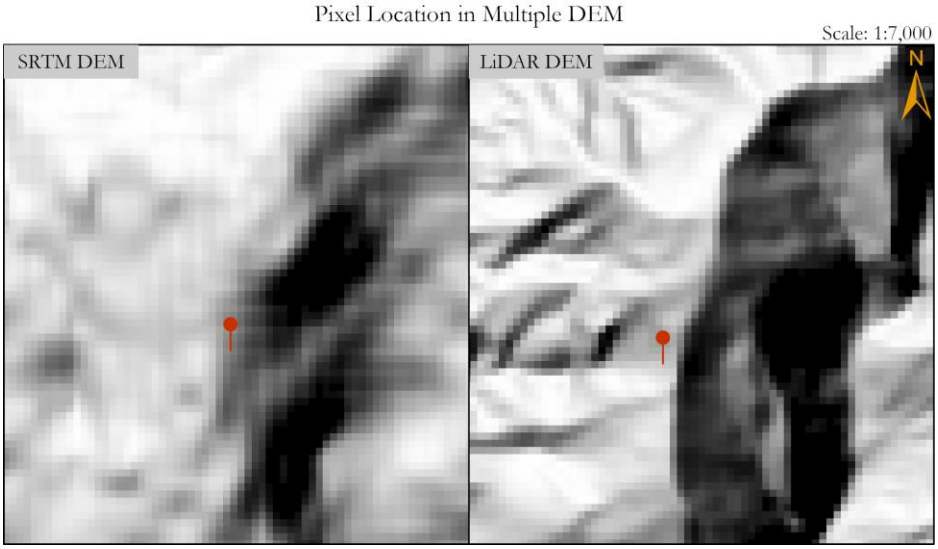


Figure 34: Evaluation of how the same pixel might represent different terrain in DEM.

In the case of the Nash Sutcliffe (NS) Index (Nash & Sutcliffe, 1970), we observed how each hydrograph at the same points (end of the channel) has a different level of flows using the Hydroeval python library (Hallouin, 2021). The model versus observation plot of the NS index for different DEM is shown by Figure 35, where we can observe that LiDAR bicubic interpolation has a significantly higher amount of flow compared to LiDAR HR in ranges up to 18 cusecs but after that, it has decreased the amount of flow. This phenomenon might be because the LiDAR bicubic interpolation data is smoother than that of the other DEMs, making the flow of water and solid less obstructed, leading to higher water flow in BC than HR DEM. Furthermore, if we observe the curves, the ESRGAN model has closer relation in terms of hydrographs to LiDAR HR followed by the EBRN model. However, discharge from ESRGAN is more fluctuating. To observe and compare numerically, we further checked the NS Index as shown in Table 12,

where we can observe that the use of SR techniques has improved the NS Index value. From this observation, we can also infer that the SR techniques may also improve the flow to match more realistic scenarios in the global dataset. This observation was made at the channel location so that for non-channel flow, the observation is not available; the reason to include non-channel flow was to observe how overall terrain improvement is affecting the model, but in general, channels are included in flow models, so this evaluation more suits for channel included flows.

DEM	**N-S Index**
LiDAR BC	0.9764
LiDAR EBRN	0.9928
LiDAR ESRGAN	0.9934
LiDAR HR	NA

Table 12: NS Index for multiple DEMs compared to LiDAR HR DEM.

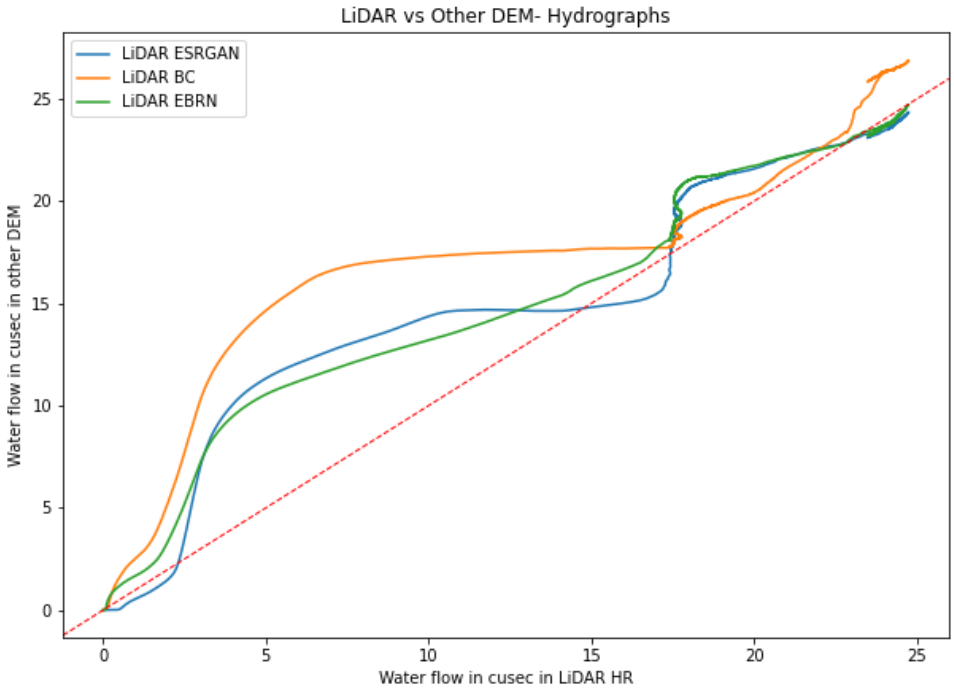
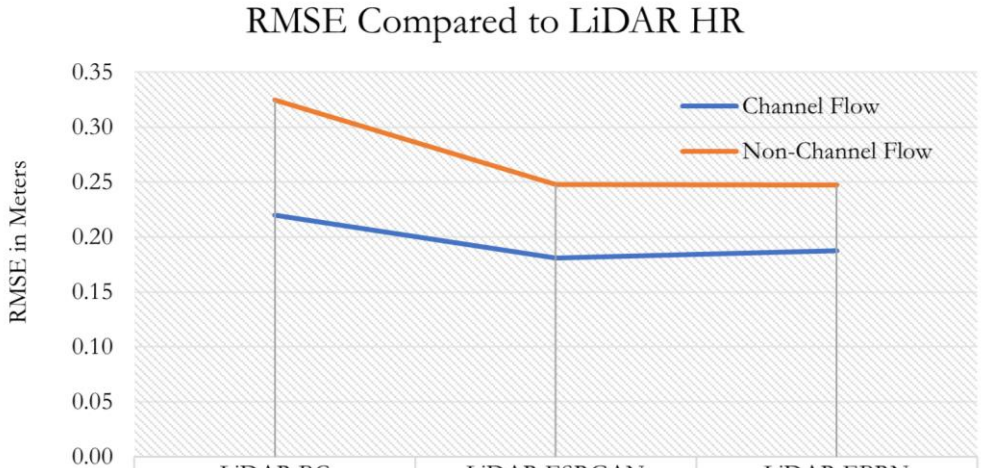


Figure 35: LiDAR HR vs other DEM flow in hydrograph, units in mm.

The NS Index provided an improvement estimation for a single point with relation to flow time and discharge, to further understand how the maximum height of fluid and to differ in SR and Interpolation techniques, we computed the root mean squared error, in this case also we could not compare the SRTM data with LiDAR DEM due to the above-stated reasons but unlike NS-Index we could compute this for non-channel flow as well. The RMSE with different DEM compared to LiDAR HR is shown in Figure 36, where we can see that the SR-based methods have improved the RMSE values in both channel and non-channel flow. Particularly, in the case of Non-channel flow, we can see that the RMSE has been improved by more this is because in non-channel flow, most of the fluid flows through the surface, and with better DEM, the error in the height of fluid can be minimized. With RMSE we can Hence, from these observations, we can conclude that the SR-based approaches have significantly improved the quality of the Physically based multi-hazard modelling.



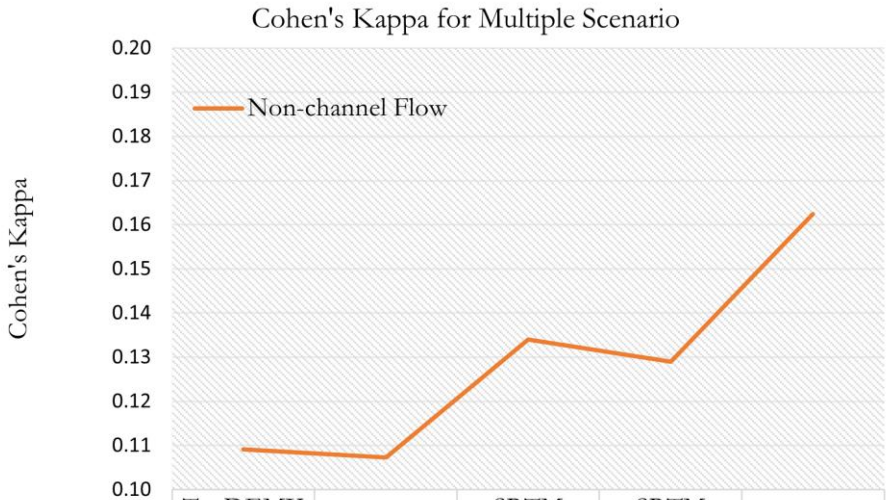
	LiDAR BC	LiDAR ESRGAN	LiDAR EBRN
Channel Flow	0.2200	0.1810	0.1875
Non-Channel Flow	0.3246	0.2480	0.2469

Figure 36: RMSE compared to LiDAR HR.

4.2.3.2. Case Study Site: Colombia

To further understand the applicability of Super-Resolution in multiple locations, multi-hazard modelling was further done in Mocoa of Colombia. We performed the analysis in non-channel flow to test the overall performance of SR methods in multi-hazard modelling. This section can only test the performance using Cohen's kappa and total discharge from the catchment; the evaluation using the RMSE and NS-Index is not possible because we did not run the model with high, low, and SR combinations from the same data source. Because we do not have the ground truth data of the solid and fluid height and the discharge records, and by the logic of the reasons explained in the previous section about the non-overlapping of the pixels, we cannot make the comparison using RMSE and NS-Index with HR datasets.

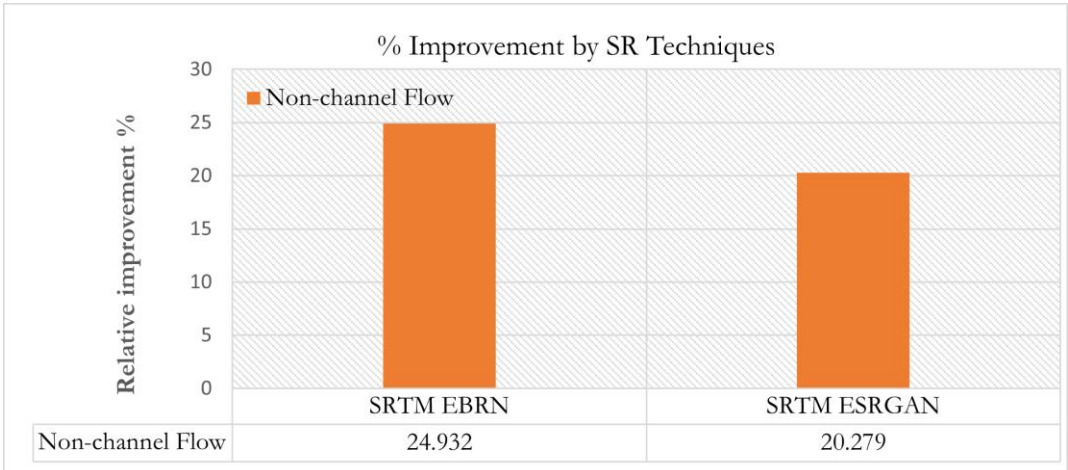
The results obtained by Cohen's kappa for combined solid and fluid flow is shown in Figure 37. We can observe that the minimum Kappa value is for the SRTM-BC followed by the TanDEM-X and SRTM based approaches. Unlike Dominica, TanDEM-X has a higher amount of kappa value than SRTM bicubic because the TanDEM-X in Colombia was less noisy than that of Dominica. We can further observe that the EBRN have more Kappa value compared to ESRGAN based methods, which is also contrary to that of the results for Dominica, which shows the model's performance largely depends on the study site than in general. The results also show that the SR based methods have surpassed Cohen's kappa value of the TanDEM-X, showing that the SR based methods could obtain better results. However, the HR data is obtained from a very similar sensor and spatial resolution as TanDEM-X; it has shown better performance due to post-processing on the DEM, as explained in the data section. In SRTM and TanDEM-X data, no post-processing except filling sink was done to reduce the amount of bias on the comparison. Even though the SR based methods have improved Cohen's Kappa, the overall values of Cohen's Kappa are very low, meaning that in this case, the requirement of High-Resolution data is required; the reasons for such low kappa values will be discussed in the discussion section.



	TanDEMX-HR	SRTM BC	SRTM EBRN	SRTM ESRGAN	HR
Non-channel Flow	0.109	0.107	0.134	0.129	0.162

Figure 37: Cohen's Kappa values for multi-hazard modelling in Colombia

To further observe the quantitative improvement, we computed and plotted the percentage improvement using the SR-based methods compared to bicubic interpolation, as shown in Figure 38. We can observe that the EBRN model has around 24.9% improvement, whereas the ESRGAN model has 20.2%. The results are surprisingly different compared to that of Dominica; in the case of Non-Channel flow in SRTM DEM, the improvement was 8-17 % for EBRN and ESRGAN. The reasons to have more percentage improvement, in this case, is also because the overall Cohen's kappa value is a very low and very small improvement in Cohen's Kappa value can be relatively higher but in the case of Dominica, even though the absolute change was large relative change was lower. If we compare absolute change both case study site has similar (in the range of 0.3-0.5) of improvement.



	SRTM EBRN	SRTM ESRGAN
Non-channel Flow	24.932	20.279

Figure 38: Improvement in Cohen's Kappa using SR based methods in Colombia

To understand the results better, we have plotted the results in maps for Colombia, focusing on the extent to which most of the hazard processes were occurring. In Figure 39, We can observe that the bicubic interpolation have mostly a very low amount of combined height, and also it misses the flow in the upper right corner, whereas the other all DEMs have flow occurring and almost aligned with the ground truth.

Most of the scenarios have a huge flow in the middle part of the area, not recorded in the inventory. This might have occurred because of the not-proper representation of terrain, and some channels or flow path might not have been visible in the DEMs. Such a high number of false positives is also validated by such a high number of false positives. In this case, also many slopes failure process is not well represented. This will be further clarified in the discussion section.

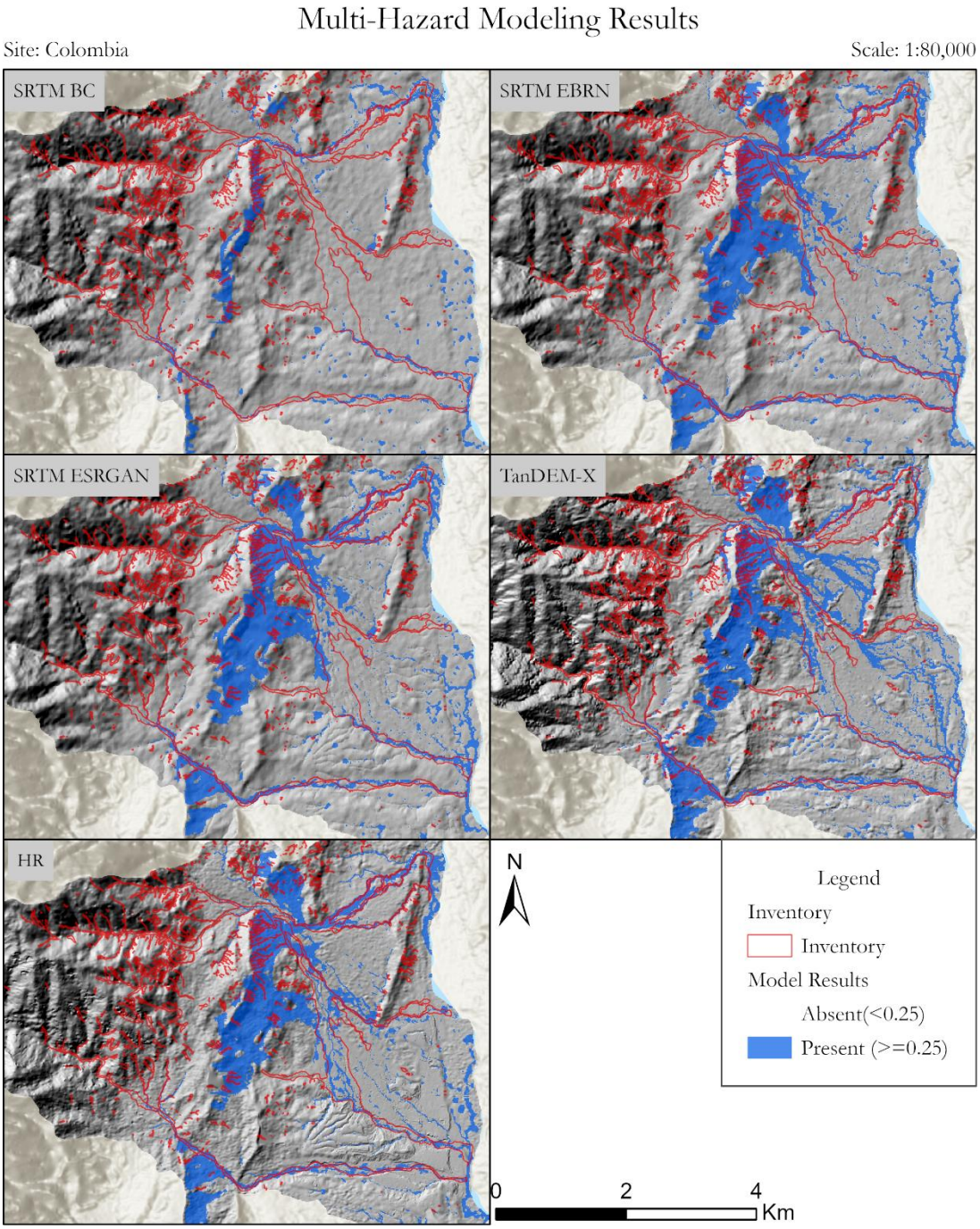


Figure 39: Multi-Hazard modelling results in Colombia, compared to inventory datasets. The hillshade images are from the respective modelling scenario.

To further understand the overall performance of the models in creating overall discharge as better as high-resolution data, we compared the overall discharge in multiple scenarios that we modelled. This observation is biased because it depends on high-resolution data. However, due to lack of ground truth on such discharge values, we simply compare for the understanding using mean error, and there are still many uncertainties

available such as if the high-resolution model is wrong in terms of overall discharge, the comparison will be more biased and show complete wrong values. The author suggests understanding this observation critically and not as an absolute comparison such as Cohen's Kappa.

The total discharge for each simulation and their absolute error compared to HR data is shown in Table 13. We can observe that the total discharge by SRTM bicubic interpolation is largely less compared to HR data, but we can not claim it is bad because of a lack of ground truth. However, if we take HR as standard and compare the absolute error with SRTM, BC is very high, and TanDEM-X follows that. SR based methods seem to have improved the total discharge amount, and both models show very similar performance on total discharge. The difference in total discharge is very less in the case of SR based methods compared to other approaches. This result, even though biased, shows that the SR-based methods can generate better discharge compared to that of the bicubic interpolation methods because of the capacity of the model to reconstruct the terrain better.

Dataset	Total discharge (mm)	Absolute Error relative to HR (mm)
SRTM BC	13.915	24.875
TanDEM-X	32.245	6.545
SRTM ESRGAN	35.597	3.193
SRTM EBRN	35.606	3.184
HR	38.790	N/A

Table 13: Absolute error in total discharge compared to high-resolution data.

5. DISCUSSION

The major results and problems that we had during the research are discussed in this chapter. There are five major sections in this chapter that will discuss the Super-Resolution results, multi-hazard modelling results, comparison of the models, applicability in multi-hazard modelling, and the advancement and limitations of the study. The discussion also follows the same sequence as the methodology section

5.1. Phase I: Super-Resolution

This section will discuss the deep learning-based Super-Resolution and its details on the training and test results.

5.1.1. Training and Fine-tuning the Models

Initially, the model was trained with the combined loss functions involving many parameters. This model did not converge well after 1000 epochs, and in both cases, the accuracy of bicubic interpolation was not exceeded. To overcome this issue, we tried to remodify the concept of curriculum learning (Avrahami et al., 1997; Bengio et al., 2009), which has already proved the better convergence with having gradually increased complexity of the data. Since complexity in our data was fixed, we gradually increased the complexity of loss functions, meaning that if a neural network was a student, instead of gradually increasing the complexity of the curriculum, we increased the complexity of examination, which proved to help us to use the novel TopoLoss to reconstruct better ridgelines and valley lines. So, instead of training a model with the more complex loss for longer epochs, we first trained the model with a less complex loss function for some epochs and then gradually increased the complexity of the loss functions. Furthermore, using the gradual increment of the complexity of loss functions also helped reduce the artefacts and gradual errors in the output.

When we trained the model with combined loss, it produced a horizontal stripe causing spiky noise in data as shown in Figure 40 because the model could not fit the weight in such a way that it can reduce all combined loss function. However, when we used the gradual increment of loss functions, it solved those problems because the model when training with complex loss functions was already initialized with weights obtained from simpler functions. This type of problems is also prominent in computer vision, where the model performance is very sensitive to initialization in multimodal loss functions. Without proper initialization, the model does not converge well (Candès & Fernandez-Granda, 2013), so with our approach, it solved three major problems (1), the problem with proper convergence in the model (2), reduction of noise in the output with complex loss functions and (3) better initializers for complex loss function solving the requirement to have proper initialization.

The training and testing datasets were obtained from the LiDAR DEM, but we also observed that the input data degradation function from LiDAR DEM HR to LR also impacted the model performance. In most of the SR literature in computer vision, they have used the MATLAB bicubic interpolation function for the image degradation for a fair comparison (Agustsson & Timofte, 2017; Dong et al., 2016; Park et al., 2003; X. Wang et al., 2018; Yang & Huang, 2017; Y. Zhang et al., 2018). However, we observed that if we use a different library such as OpenCV or ArcGIS, the training performance will be different; also, when we perform the testing, the LR created from other sources performs differently than that of LR prepared from the same methods. The reason behind that is in the interpolation function even though all the libraries and software claim it to be a bicubic interpolation; the weighting strategy is different; for example, the OpenCV

and MATLAB use the same function for bicubic interpolation as shown in Equation 20; however, the weight value a is different for each software (0.5 for MATLAB and 0.75 for OpenCV) (MATLAB, 2021; OpenCV, 2021). Furthermore, MATLAB has implemented an anti-aliasing function that completely changes the input dataset to the model, and the model performance will not be the same in the case of data prepared through different libraries. So, as a standard library used in most of the research, MATLAB is considered the standard, and we used it as a basis for our research.

$p(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}x^i y^j$ …. (20) Bicubic interpolation function implemented in multiple libraries.

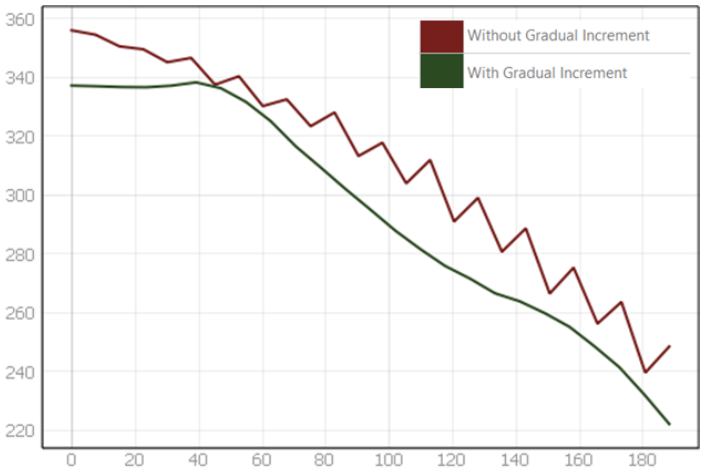


Figure 40: Comparison of reconstructed DEM with and without gradual increment of Loss function complexity

Another major challenge we faced during the training of the EBRN and ESRGAN model was vanishing and exploding gradients, the major reasons for vanishing gradients and exploding gradients were caused by extremely low and extremely high gradients or norms during the training process, which makes the updated weights and bias range to infinity making the loss values to Nan (Pascanu et al., 2012). To solve this problem, we used gradient clipping and also a lower learning rate to reduce the fluctuating of the model, as suggested by Goodfellow et al. (2016, p. 289). Furthermore, in the case of GAN training, we faced more problem of collapse mode, which was caused by either the discriminator model was too powerful and its loss becomes nearly zero, or both models keep oscillating during the training as observed by Durall et al. (2021). To solve that, we used the pretrained Generator model and stopped updating the weights of the discriminator model after certain epochs to avoid collapse mode, as suggested by Ham, Jun, & Kim (2020) and X. Wang et al. (2018). Furthermore, the hyperparameters were tuned with the different range using the Keras tuner, which optimized the learning rate, decay factor for learning rate, and with using model checkpoint with validation, we could train the model for 1000 epochs without overfitting the model.

5.1.2. Test on High-Resolution Data

The HR DEM inference was used to understand how the Super-Resolution model improves the quality of DEM in multiple locations before using it for a global free dataset. The PSNR scale being an MSE dependent scale, as we can observe in Figure 23, the Colombian DEM had higher PSNR in all scenarios compared to Dominica and Austria, even though it had more noise in the data and the quality of DEM was not better as compared to Dominica and Austria. The reason to have higher PSNR was that the MSE value for Colombia was less because the noise present in the DEM is smaller. When we degraded the HR image to make it LR,

those noises were also smoothed, but when we calculate MSE, which is an average measurement, the mean value is very low for Colombia compared to Austria and Dominica. In Austria and Dominica, even though the DEM was good quality because of the smoothing function and few large noises, the MSE was higher. So, this shows that using PSNR and MSE can show the improvement in noise and error reduction but cannot show the quality of the data that should be considered during the training and evaluation.

As we observed in the results section, we could obtain the model accuracy with more than 20% improvement in all the locations, but the improvement in EBRN was quite interesting, with the highest improvement in Colombia followed by Dominica and Austria. Even though the model was trained in Austria, it had more improvement in Colombia and Dominica; this is because the EBRN model has more capacity to reduce the noise and improve the data based on its frequency details (Qiu et al., 2019). The LiDAR DEM available in Austria was very high quality, and it did not have much noise in its LR counterpart, making it very difficult to improve the quality. However, in the case of Colombia, the available HR DEM was very noisy and consisted of artefacts that were also preserved in its LR counterpart, and the EBRN model being capable of reducing such noise had removed much of the noise from the input dataset, making it better in terms of MSE improvement. Furthermore, the quality of DEM in Dominica was also relatively better, showing that the MSE improvement is more than Austria but far less than Colombia. This observation shows that, in the EBRN model, using more noisy data has more room for improvement, but the amount of improvement it can do is less than that of the ESRGAN model. ESRGAN model, on the other hand, shows performance based on the input data is received(X. Wang et al., 2018); the model is trained in Austria heavily, have more capacity to improve in Austrian terrain. Whereas it also could improve better in Colombia than Dominica due to the fact that terrain characteristics in Colombian DEM are very similar to that of Austria whereas, being an Island Nation, the Dominican terrain was quite different where ESRGAN model suffered to improve as much as EBRN.

Most of the available research has used their algorithms in different locations and with different peak values for PSNR without relative comparisons, and we cannot compare those existing works (Argudo et al., 2018; Z. Chen et al., 2016; Demiray et al., 2020; Jiao et al., 2020; Kubade et al., 2021, 2020; Shin & Spittle, 2019; Wu & Ma, 2020). Furthermore, as opposed to computer vision Super-Resolution, where existing test datasets are available for a fair comparison between algorithms, in the geoscience community, especially for DEM data, there is no standard dataset for comparison, making it impossible to compare the performance of our approach relative to others. Even though the works of Argudo, Chica, & Andujar (2018) (FCND Model) and Kubade et al. (2021, 2020) (DSRFB-2020 and AFND-2021)have used data from Austria, their scale factor is other than 4 (7.5), and also the resolution of the data they used is 2 meters compared to ours five meters making it incomparable. Furthermore, the versions of research by Kubade et al. (2021, 2020), which includes RGB images and overlapping in the model, are also not comparable because we do not have included any auxiliary information, as well as overlapping, can cause false overrepresentation of accuracy as explained in section 3.3.3. The only way to compare our results is by checking the relative improvement compared to the existing studies in a similar location with a similar scale (because the model performs differently in different scales). The research by OpenDEM (2021) in Austria has provided the relative improvement compared to Bicubic Interpolation with RMSE: 7%, whereas our results have improvement (In Austria) of 20.397% for ESRGAN and 12.191% EBRN, respectively, which shows our results have significantly improved the results than their studies.

On the 4x scale of Super-resolution, Xu et al. (2019) have provided some interesting results. However, because the data in the study is not published, we cannot do the absolute comparison and for the relative

comparison also, as we can observe in the study of Kubade et al. (2021), relative improvement also largely depends upon the study locations making it difficult to make comparisons in the unpublished and non-standard datasets. One possible way to compare is to compare the model's capacity to reduce the squared root of standard deviation in error (MSE), which shows the capacity of the model to reduce the noise and random error compared to bicubic interpolation (because Bicubic is the most wildly used among interpolation techniques due to its fast nature with good quality outputs). This approach is less biased to the dataset location than the reduction of mean error, but not completely without any bias and should be treated accordingly only for relative information. The comparison of our models with the work of Xu et al. (2019) is shown in Table 14. As we can see in the table, the standard deviation in RMSE has been significantly decreased by our both models compared to BP (Xu et al., 2019), Sun11(Sun et al., 2011), D-SRCNN(Z. Chen et al., 2016), DGPN (SRCNN) (Xu et al., 2019)and DGPN (EDSR) (Xu et al., 2019).

Method	**AVG RMSE**	**RSSE**	**Avg St. Dev Improvement in RMSE (%)**	**Data**
Bicubic	13.635	1.952		Different than ours
BP	11.136	1.882	3.586	
Sun11	10.782	1.885	3.432	
D-SRCNN	10.962	1.919	1.691	
DGPN(SRCNN)	10.130	1.798	7.889	
DGPN(EDSR)	9.785	1.805	7.531	
Bicubic	1.170	2.127		Ours
EBRN (Ours)	**0.898**	**1.681**	**20.957**	
ESRGAN (Ours)	**0.974**	**1.813**	**14.780**	

Table 14: Comparison of standard deviation reduction by different methods for the different datasets.

While comparing the quality of our model to that of other published work, we have realized that the requirement of a standard dataset for modelling and comparing the quality of the model is very important and required for geoscience data. In the case of computer vision, there are many such data available such as DIV2K (Agustsson & Timofte, 2017), General100(Dong et al., 2016), and MANGA109 (Aizawa et al., 2020), which are wildly used in training and testing the model. However, as we mentioned earlier, due to the lack of such a dataset in the geoscience community, it is very difficult to compare the quality of the model and its output. To make a reasonable comparison, we must consider the terrain the model is trained, the degradation function used to train the model, the quality of the input dataset, and the test data itself, but there is no other matching dataset that we can make a reasonable comparison with our models. The one way to compare how much percentage of MSE, but the presence of noise in multiple datasets are very different it is not possible to compare them. To have a relative comparison, we proposed to use percentage reduction in the standard deviation of MSE, which represents the capacity of the model to reduce the noise can be one approach, instead of directly comparing the MSE, which is largely dependent on data comparing the capacity of the model to reduce noise makes more fair comparison. However, this approach is temporary and is biased because the complexity of data is not considered, and with different complexity of the dataset, it might be difficult to make an absolute comparison making it a temporary solution. Hence, there is a strong need for the standard dataset in geoscience for Super-Resolution modelling and comparison.

When we started comparing the SSIM values, which is a standard tool to measure the visual quality in computer vision Super-Resolution tasks, it considers the visual representation of the data and compares their similarity (Z. Wang et al., 2004). However, In the case of DEM, the major problem is that even though the DEM is very different, it looks very similar for the observer. The values of SSIM are quite similar in both cases, as shown in Table 10, where we can observe that for all DEMs, the SSIM value was extremely high, making it difficult to compare and find the difference, this is because SSIM was designed for the optical images which have ranged from 0-255 in most of the cases(Z. Wang et al., 2004). All the features in optical images are visible, and small changes can be recognized, but in the case of DEM data, the changes are very difficult to visualize without converting them to derivatives. To overcome this issue, we used SSIM with hillshade images that range from 0-255, and it is easier to compare the difference for the human observer with hillshade, which also makes it more reasonable to use SSIM with hillshade images. As per our knowledge, we are the first to use SSIM with DEM derivatives to make a reasonable comparison. Most DEM SR research has not considered SSIM as evaluation metrics and only focus on PSNR, which lacks the part of research to consider improvement in visual quality (Argudo et al., 2018; Z. Chen et al., 2016; Demiray et al., 2020; Jiao et al., 2020; Kubade et al., 2021, 2020; Shin & Spittle, 2019; Xu et al., 2019) because the major need of SR technique is to generate better-looking images with high quality not only images with high PSNR. Furthermore, as we explained before, having higher PSNR might not mean better and noise-free images in case of the presence of small frequency noise in the data. Wu & Ma (2020) used the SSIM Index in their research, but the problem changed in SSIM with DEM SR was in the scale of 1e-5, which makes it very difficult to understand if there is an improvement or not, so, our approach of using SSIM with derivatives has proven to be a better metric to measure the improvement in the visual quality of the image.

5.2. Phase II: Super-Resolution Applicability test

This section will first discuss the applicability of Super-Resolution in increasing the quality of the global digital elevation model and followed by applicability in multi-hazard modelling.

5.2.1. Super-Resolution Evaluation in Global Dataset

This section discusses Super-Resolution in the global dataset and its evaluation using the derivative and geomorphological evaluations. The quantitative comparison using methods such as RMSE, MSE and other methods described by Polidori & Hage (2020) was not possible due to the lack of ground truth data, we could have used the HR DEM as ground truth, but the pixels with HR DEM were not aligned to SRTM DEM making it impossible to compare pixel-by-pixel. Furthermore, the range of elevation in each pixel are also randomly related, and there is no systematic error which made the comparison with HR data more biased. For a simple understanding, please observe Figure 41, where SRTM and LIDAR HR dem cross-section are plotted in a 1600 meters long cross-section; both DEM have no specific error patterns of the elevation and SRTM have more similar data in some places and completely different values at other places. To overcome those, we used the derivative, visual and geomorphological analysis.

5.2.1.1. Derivative analysis

The derivative analysis shows how the SR-based methods can improve the quality of the derivatives compared to bicubic interpolation. The major discussion points on using KDE function, are, the pixel-by-pixel comparison with HR DEM from another source is not possible due to the fact that elevation obtained from different sensor have a small shift in the dataset due to their grid structure as well as the spatial resolution in which the sampling was done. Such phenomenon of the horizontal shift was observed by many authors such as Alganci et al. (2018) and suggested that co-registration was required to solve the issue. However, in our case, because the shift was not along a single direction and there was a vast difference in

sampling resolution between HR DEM and SRTM DEM, it was difficult to co-register them. Furthermore, the research aimed to develop a post-processing algorithm to improve global data in case of lack of high-resolution data, so we did not perform such post-processing because it might change the model's overall performance and make it more biased.

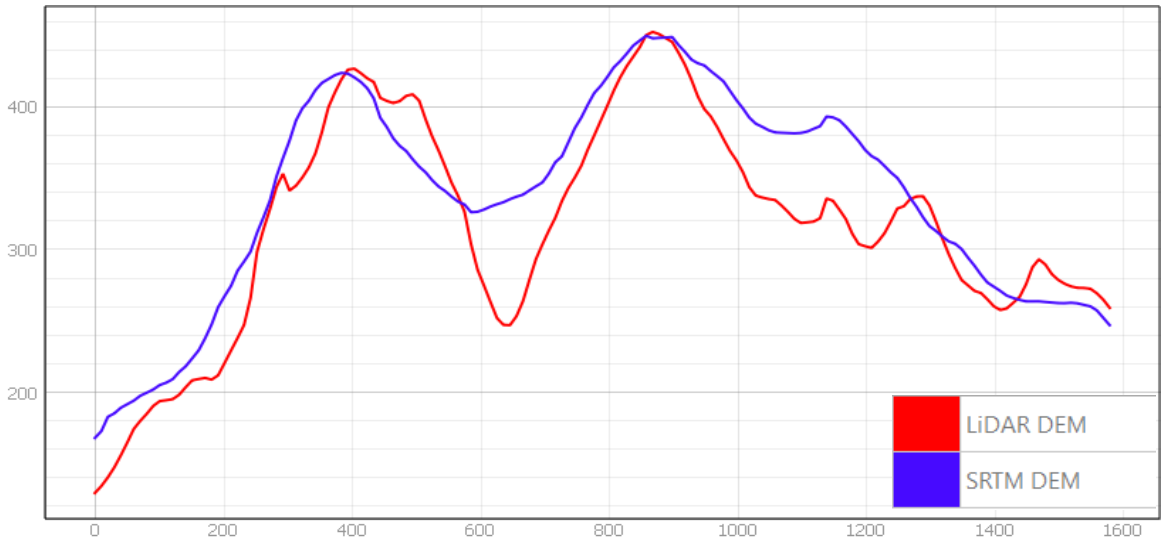


Figure 41: Demonstrating the difference in elevation values and its randomness in SRTM and LiDAR DEM. X-Axis represents distance, and Y-Axis represents elevation.

The KDE function estimated the kernel density for all DEM, and when we plotted them with each other, we could see that the SR-based methods have KDE more near to HR than the BC interpolation. However, the differences were comparatively small; in the case of MSE and PSNR, we could observe the differences in numerical values, which could show that in absolute terms, but here the derivatives have a very smaller range making it difficult to observe the difference. For example, for a 10-meter horizontal distance, if the elevation difference changes by 0.5 (let us say from 3 to 3.5), the slope will change by ~0.04 radians which is not significant in terms of derivatives but which is very significant in terms of elevation, so drastic change in the derivative is very difficult to obtain. However, those smaller changes have a very big impact on the physically based modelling because when the slope changes by a very small amount, the stability and flow of the pixels can change exponentially. For example, in the case of multiplication by sin(a), if the slope value changes by ~0.04, the multiplication factor will change by ~0.039, and such deviation can cause a significant change in physically based simulations. As we observe in all three cases, the slope has been improved (even though a small amount), making it more like that of the HR; the impact of this improvement can be observed in the results of multi-hazard modelling. The TanDEM-X have a very similar curve to that of the HR DEM in most of the case. We can see the impact of initial data collection resolution in this case; as long as the high amount of content is available in the original data, the derivatives are more likely to be similar to high-resolution DEM but due to such feature and terrain variability are not available in the SRTM LR DEM the improvement in such cases are very difficult to obtain. This problem is also explained and backed by the theory of data processing inequality (Beaudry & Renner, 2012).

In the case of Aspect, the improvement is very low and negligible, but the HR and SRTM DEM have very similar KDE functions; this can be further explained by the fact that small changes in elevations and small noises cannot change the direction to which overall terrain is facing to. Even though the small values might change on aspect angle, but in a larger context, the overall face of the slope unit will not change by noise.

However, in the case of Dominica, TanDEM-X has a more variable aspect compared to the other two datasets, with amplified peak and crest. The reason behind such an effect in TanDEM-X is due to very large noise and spikes in the dataset. It has created amplified aspect angles due to sudden change in the direction of the terrain; other than that, in the case of noise-free data, all kinds of DEM have very similar Aspect with a small difference. However, in this case, SR-based methods have also improved the quality of Aspect by making their distribution closer to HR DEM.

For TWI, in the case of both Colombia and Austria, we have seen very similar results compared to that of the Slope. However, in the case of Dominica, the TWI has shifted a bit far from the HR DEM for all BC and SR methods while keeping the similar structure of the curve. This observation represents that more pixels had a higher value of wetness compared to HR DEM. The TWI is calculated logarithmically by dividing the flow accumulation by slope, and there are two possible reasons to have higher TWI, the SRTM DEM is relatively flatter (with lower slope) than that of the HR and TanDEM-X, causing more TWI values or due to the smoother DEM more pixel's flow into the lower region contributing higher TWI. From slope KDE, we can safely conclude that the lower amount of slope has caused the higher values to TWI in this case, and the value of TWI has not much changed by using SR based methods which is because, in logarithmic scale, the change of slope in Dominica is insignificant which makes not much impact on TWI. For further reference on explanation, please see how the TWI is calculated and elevation affects the TWI in Agren et al. (2014), Mattivi et al. (2019), Qin et al. (2011), and Radula et al. (2018). Furthermore, the majority of terrain covered by HR had a relatively lower slope, and in SR based approach, those slopes are not changed by a higher amount, which causes the TWI improvement insignificant.

Even though it was the best possible approach to check the distribution on the whole dataset to estimate the improvement, the KDE function smooths the data, and usually, the kernel functions remove the required variability in the dataset (Hwang et al., 1994). Furthermore, the KDE functions are largely dependent on the smoothing functions as well as the kernel width, which is kept constant in our case, but this can largely affect the results (Portugués, 2021)

5.2.1.2. Visual and Geomorphological Analysis

In visual analysis, the results are very self-explanatory, but few things are worth discussing. The model has better preserved the ridge, and the valley lines than bicubic interpolation, whereas the middle slopes, are crisper, but we cannot find much difference on them. The reason is also related to the theory of data processing inequality (Beaudry & Renner, 2012), here in SRTM LR, because we have a coarse representation of the ridgelines and the valley lines, the SR based methods could create better ridge and valley lines. However, in most of the terrain, the features on mountain slope are not visible in the SRTM DEM making it impossible to generate via SR methods. If SR-based methods generate such features, that would be completely wrong and considered as artefacts because without any indication of features in LR data, if SR models start to create artefacts, they would increase the noise in the model instead of decreasing. However, in the case of SRTM DEM, where smaller feature representation is available and ridge and valley lines are visible, the SR-based methods have created a very good representation of them in SR output. This has a limitation of SR-based methods in mapping the geomorphological features, which are not available in the LR dataset, and this will be discussed in the next section.

Another important point to discuss in the visual analysis is that the ESRGAN based method has smoother terrain in most of the cases of SRTM input, the reason lies in the model architecture, the EBRN model process different frequency of information with different level and complexity. In contrast, ESRGAN passes

all of them through the same network. When the model was trained, we had high-resolution DEM and its degraded counterpart, which was better in terms of visual quality compared to SRTM LR. The ESRGAN Generator model learned to generate HR data from those perfectly created bicubic samples, but it had suffered to provide better quality images when we provided real-world data. However, the ESRGAN model has better and comparable performance in the HR DEM and their degraded LR DEM; it suffered heavily in SRTM DEM. On the other hand, the EBRN model has different blocks of processing complexity, and when it was trained, all those blocks have their weights, and when we provided more noisy data, it could pass it through more complex blocks and process to generate the relatively better-looking dataset. The model architecture and its development strategy have played a major role in producing such discrepancies in similar performing models.

In the visual analysis, we can also observe that the ESRGAN based approach has created some artefacts in Colombia DEM and few in Austria DEM, especially in the lower hill of Colombia. Being a Generative Adversarial Network, the artefacts are very common on the ESRGAN, and the model averaging has tried to reduce the artefacts, but this phenomenon is inevitable, and we can expect some sort of artefacts in ESRGAN based networks (X. Wang et al., 2018). However, there are few ways to reduce such artefacts, and one of them was to average the pretrained generator with the GAN trained generator network, even though we did the model averaging with the factor of 0.5 and taken the mean of the model to ensure that we obtain the best visual and PSNR quality, we could still find some artefacts on Colombia case. In contrast, Austria and Dominica do not have many artefacts. Therefore, both EBRN and ESRGAN model have performed similarly in the case of derivative analysis. However, in the case of visual analysis, the ESRGAN model has lower performance than that of the EBRN model in terms of visual clarity and fewer artefacts, and most of those effects are due to the architecture of the model and their design for specific tasks.

Multiple experts did the geomorphological evaluation in DEM, and their score and remarks about the geomorphological data were analysed. The major comment was, even though the SR-based approach creates a crisp-looking image, features such as Landslide detection in those DEMs are not possible. This phenomenon is expected, and it is due to the data processing inequality (Beaudry & Renner, 2012), which states that information that are not present in the data while collection cannot be obtained by post-processing. As we can also observe in the visual analysis section, the images produced by SR methods have more crisp looking ridgelines and valley lines, and the features that are present have become crisper, but if the terrain has less amount of slope, it could not produce any features but just created clearer and crisper looking surface. This phenomenon has a very strong impact on the geomorphological mapping from the SR methods; in terms of mapping large features such as mountain slope, flood plain, etc., the SR-based methods can provide good visualization. However, in terms of mapping features like landslides and faults are very difficult with SR-based methods because to generate a better quality hillshade with more features, the features in the LR image should be present even though in low-resolution. When there is no feature present in Low resolution, it cannot be generated by SR methods and hence cannot be mapped. For better understanding, Figure 42 shows that in the upper image where ridge and valley lines were present in LR DEM, the SR-based method has created realistic terrain with crisp terrain and crisp ridgelines. However, in the lower image, because there was very few features present, even to a bigger extent, there was not much improvement in visual characteristics because there were not many features present in the LR image. This phenomenon can be both interpreted as boon and curse because the capacity to limit fake artefacts helps to generate better terrain, and not being able to create better features by prediction without the availability of visible features limits the possibility of using SR methods in mapping small scale geomorphological features.

Interestingly, if we have large landslides visible with low-resolution images, SR-based methods can generate more crisp-looking landslides making it possible to map them easily.

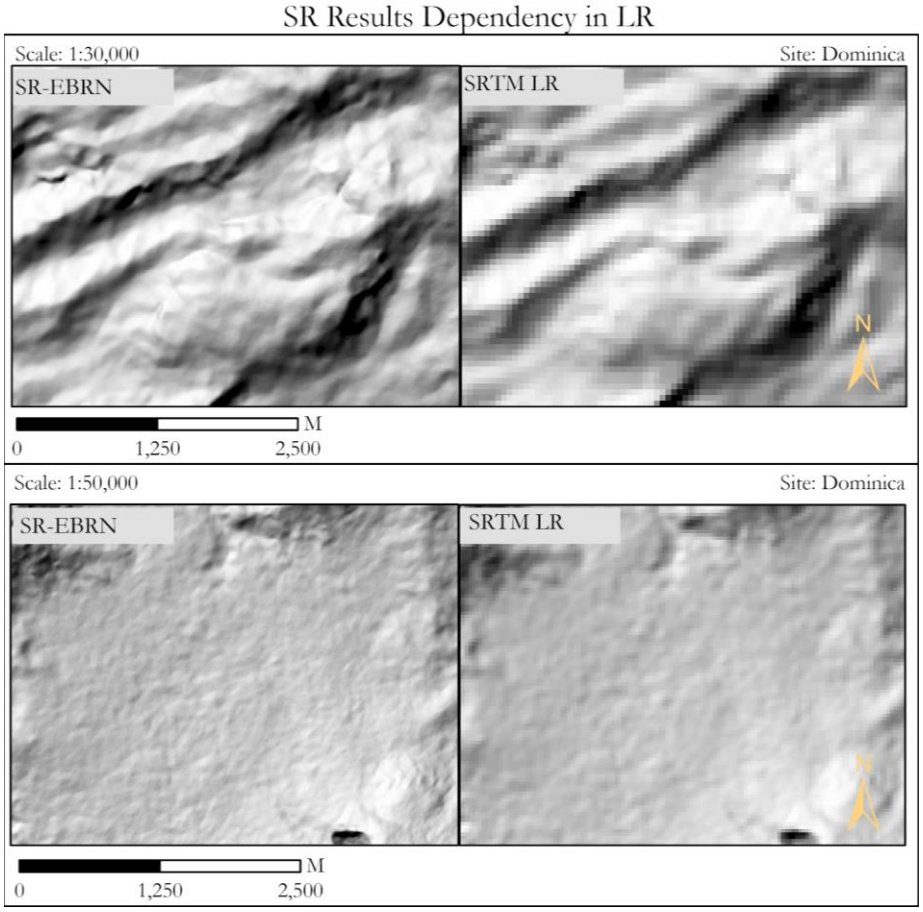


Figure 42: Dependency of SR products on the information content available in LR image.

5.2.2. Multi-Hazard Modelling

The multi-hazard modelling was done with different possible scenarios to understand the impact of SR in flow dynamics and flow characteristics. The observation that we had was interesting, with significant improvement in the quality of output and data. The major discussion points on multi-hazard modelling and applicability of SR in multi-hazard modelling are calibration of the model, brief discussion on results we obtained, and applicability of SR methods in data-poor regions.

5.2.2.1. Calibration of the Models

All the scenarios were calibrated with the gradient descent algorithm; the major benefit of using GD over that of the manual values is, it is adaptable to the current accuracy of the model and adapts quickly than that of the manual method (Léon Bottou, 1999). Furthermore, because we had 420 different scenarios to simulate, manually calibrating all those was virtually impossible in the given time frame of the research. The GD method has proven to be a good estimator, and we could also get better results with some improvement over the initial parameters, which were calibrated in Dominica in the existing model. The calibration was easier for us because we already had the calibrated parameters from the model run in HR LIDAR DEM, and with the use of SR methods, we could obtain better results. The results obtained from the calibration are reported as Cohen's Kappa coefficient in the previous section.

Some of the issues we faced with calibration using GD was, sometimes it created the multiplication factors unrealistic such as 400 for manning's N, and such a high number of multiplication factor had caused very small accuracy; we monitored the calibration process via GD and did not consider the accuracy obtained by such unrealistic multiplication factors in our analysis. Such unrealistic values are because the GD algorithm does not understand the physical relationship between values and their impact on modelling; it just tries to minimize the error at any cost. Furthermore, to reach the best minimum error, it might take way more simulations because GD is very slow on converging, and if there are multiple local minima, it is possible that the model gets stuck in some values and never improves (Chi et al., 2009, p. 472). So, In our context with 20 simulations, an expert modeller with a higher understanding of the relation of hazards with input parameters might generate better results. This has been observed by the situation in Dominica, where van den Bout (2020, p. 187) could obtain Cohen's Kappa (without filtering below 25 cm) value of 0.562, whereas our approach could only obtain 0.55, this shows that expert modeller like van den Bout can generate better results than gradient descent methods in few simulations. However, with the addition of more simulations, GD might be able to get better results. The loss function (error function) we used is not a convex optimization problem, and reducing any values, in this case, might not show a large effect making it more prone to lack of convergence. To optimize better and faster, the implementation of algorithms such as Adam optimizer would help to generate better results (Kingma & Ba, 2015). As we observed in the visual evaluation of both Colombia and Dominica, many slope failures were not present in the results. This is because when we calibrated the models, the major goal was to reduce (1-Cohen's kappa) value.

Furthermore, because flood event is spread largely and true prediction of such large event increases with more flood and Cohen's kappa increases reducing the error, compared to when it creates more landslides, so, the model calibration output was more biased towards flood. We could not solve this problem because we had around 420 simulations to run, and manual calibration of such a huge number of simulations was impossible. However, this can be solved in the future by adding multiple or weighted error functions that consider slope failure and flooding events separately and add their values with the required weight.

The parameter of GD, learning rate, and momentum are very sensitive to its performance, and setting those parameters right is very important; otherwise, the multiplication factors will be changed unrealistically, as suggested by Chen et al. (2019). To set those parameters right, we started calibration from 0.1 for learning rate and 0.5 for momentum and calibrated the sample model with a bigger timestep for obtaining good parameters; then, we decreased the learning rate multiple times until the model became stable and gave more realistic results. The final parameters we used were 0.001,0.9 for learning rate and momentum in both sites, which did not let the multiplication factor grow absurd. The major drawback of using GD for calibration is that it is difficult to store intermediate results. It is only possible to get the final parameters that are best to re-run the model, and visually inspecting intermediate maps is more difficult in GD compared to the manual calibration method. However, in this case, our concern was with the results but not the intermediate results, and we gave all the scenarios a fair chance of equally calibrating for 20 epochs so that intermediate results were not much relevant.

5.2.2.2. Results With SR Methods: Dominica

The Cohen's kappa was used to understand the overall performance of the model instead of the Accuracy metrics because the accuracy is a non-normalized scale, and it does not account for the random chance of having such classification, but the Cohen's Kappa statistics compensates for the random chance (Ben-David, 2008). When we compared the Kappa values, the improvement is higher in the LiDAR DEM, whereas the SRTM DEM had a lower amount of improvement. The LiDAR low-resolution DEM is a smoothened DEM

from the high-resolution by bicubic interpolation, but no artificial noise is present on the data. When we used SR-based methods, our models are better at creating crisper and better-looking ridgelines, valley lines, and channels, showing that in the case of LiDAR data, there is significantly higher improvement in the modelling output because of the reconstruction. However, in the case of SRTM DEM, the noise present on the low-resolution data had limited the capacity of the model to reconstruct the images better. This phenomenon is well observed and recorded in computer vision by the works of Guan et al. (2019). Moreover, if we observe in the results section, we can observe that the bicubic interpolation has more extended flow compared to SR based methods, this is because the bicubic interpolation smoothed the channel or rivers, which lead the lower depth of the channel, making the fluid flow more in horizontal extent. Figure 43 shows a random cross-section in a channel where both ESRGAN and bicubic interpolation is shown, and we can observe that ESRGAN have significantly higher channel depth.

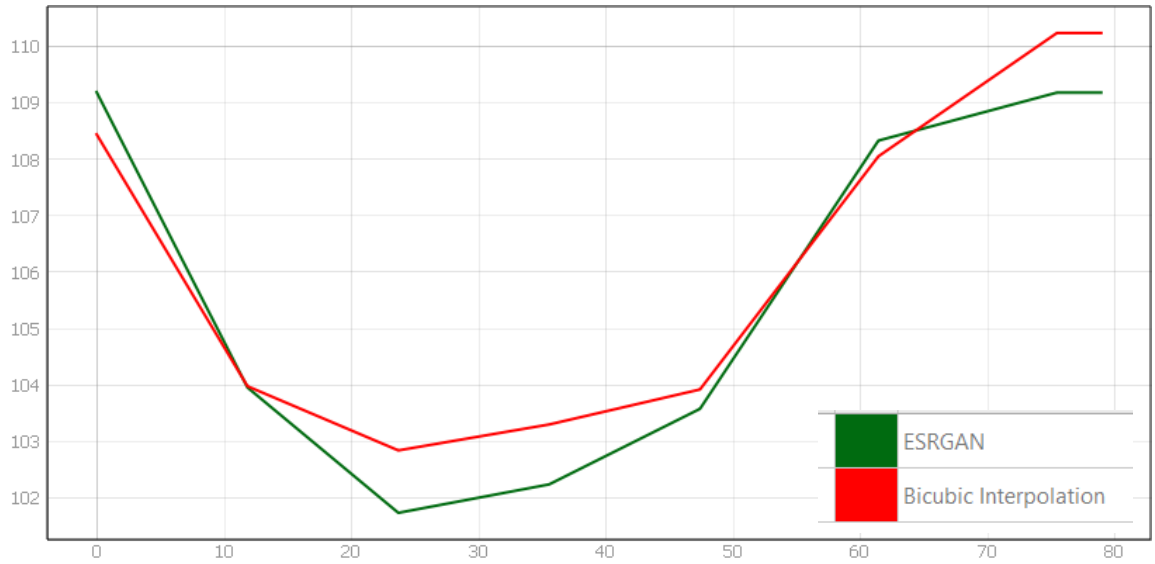


Figure 43: Cross-section in the channel, which shows how bicubic interpolation have smoothened the channel height.

Furthermore, with the change in input low-resolution data from the different sensor with different noise levels, the quality of the output can change significantly (Guan et al., 2019); this might also have had an impact on improvement using EBRN and ESRGAN models. However, both models have shown the capacity to improve the DEM and modelling output using such data. Comparing our high-resolution model with van den Bout (2020, p. 187), the Cohen's kappa value is lower in our case when we used 0.25 as threshold the Cohen's kappa value is ~0.47, whereas if we do not use the threshold as van den Bout (2020, p. 187), our Cohen's kappa value is 0.55. This shows that the use of threshold reduces Cohen's kappa value significantly, but that is very important because, in most cases, the impact of the hazards is very low, below 0.25 meters. However, in non-threshold results, our calibration only provided results up to 0.55 due to our automated calibration process limitations as described in the calibration section. To reduce such problems, either increasing the number of simulations or use of multiple and better error functions such as cross-entropy will help because of their capacity to check correct classification in a more holistic manner with different hazards as different classes (Haussler, 1992).

On the other hand, there is a higher probability of improvement in LiDAR DEM in both channel and non-channel flow due to less noise in the input dataset and because the model was trained in the LiDAR DEM rather than SRTM DEM. This phenomenon is called transferability of Super-Resolution in natural images where non-synthetic images suffer to generate better results than the ones which were used to train the model, this is a new research field in deep learning-based Super-Resolution, and very active research is going

on (Anwar et al., 2020; Guan et al., 2019; Mousavi et al., 2018). Furthermore, in both LiDAR and SRTM DEM, we can observe that the improvement in Channel flow is a bit higher, which is in line with our visual observation. The channel flow improvement in LiDAR DEM is very higher than that of SRTM data, which is because the channels were visible in LR LiDAR data because it was captured in the HR LiDAR DEM. However, due to the smaller ground sampling distance, SRTM had not captured such features making it impossible to reconstruct. Non-Channel flow results are very similar to channel flow, showing that the improvement in the non-channel area is very low compared to channels. The EBRN model has improved less in the Channel and Non-Channel flow of SRTM; this is due to the model architecture, where ESRGAN being a generative model, it can produce better images than the EBRN, which works as a noise reduction method (Qiu et al., 2019; X. Wang et al., 2018). If we compare the output from our SR data to commercial global DEM, we can observe that it has higher improvement, this because in our modelling site, the TanDEM-X have very noisy data compared to other places in Dominica shown in Figure 44. This noise in DEM compared to SRTM reduced its accuracy to very low in the case of physically based modelling, but if we observe the derivative analysis, it has better performance than SR-based methods.

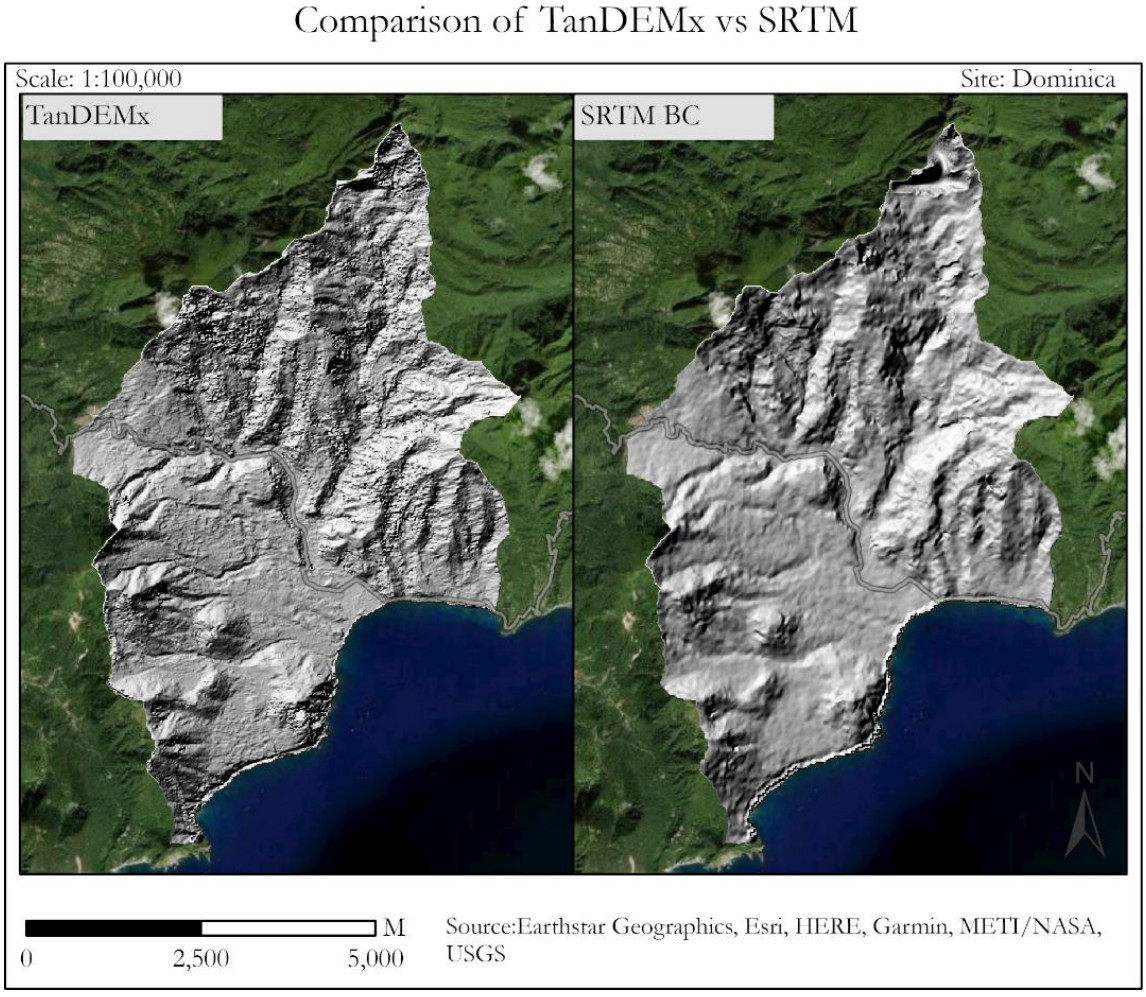


Figure 44: Comparison of TanDEM-X vs SRTM in physically based multi-hazard modelling site

In the overall channel flow of solid and fluid, when we considered LiDAR HR as standard and observed computed the absolute error, we observed that absolute error was significantly reduced while using SR-based methods compared to that of the BC-based method in SRTM data. However, in the case of LiDAR data, the total channel flow in ESRGAN is more erroneous than that of the bicubic interpolation, but when we compare kappa values, it shows more improvement. This discrepancy between kappa statistics and total

water flow in the channel can be due to 2 major reasons, (1) the ESRGAN method increased the artefacts in LiDAR DEM, causing less water to flow towards the channel, (2) when the crisper image was obtained using ESRGAN, the amount of flow reduced than that of the smoother data. Furthermore, if we compare in the case of LiDAR DEM, the improvement in total flow in the channel is not that significant. With bicubic interpolation, the amount of flow was higher than HR, but in SR methods, it is lower, showing that the smoother surface cause more water to flow into the channels, and SR based methods creating more crisp images cause the lower amount of flow in the channel which can be due to more artefacts in DEM, not connected channel flows, etc. The SR-based methods also had the drawback that when creating crisp images, they sometimes create hydrologically incorrect DEM causing non-connected flow, making the overall channel flow smaller. In the case of SRTM DEM, the results show that the SR-based methods in SRTM DEM have improved the quality making more water flow through the channel and reducing absolute error compared to LiDAR HR. In this observation, we are confident that the SR-based methods improve the performance of the model. In this case, if we compare the TanDEM-X, it has a very high amount of total channel flow compared to other methods; this is also caused by noise present in the DEM, which created small ponds making a higher amount of water in the channel.

When comparing the NS index and RMSE, the EBRN and ESRGAN model have provided better discharge than the bicubic interpolation technique, showing that in multi-hazard modelling of hydrometeorological hazards, it is better to use SR based methods for modelling. The NS index is a widely used method for hydrographs comparisons, and it has shown that our approach has improved the quality of discharge at a random location. However, if we investigate Figure 25, up to 18mm of flow, the SR-based methods have way better similarity with HR data compared to BC methods, but after that, SR-based methods show more divergence with a higher amount of flow. The reason behind this lies in the DEM itself; when the DEM is smoother, the water flows faster, reaching the discharge point due to that the amount of flow in the lower range is higher for BC DEM. However, in SR based approach, the flow gets slower because it has to fill the terrain undulations and reaches later than that of the BC DEM, making a higher peak in the range of 18-21mm. The BC also has higher values after 22mm during the peak, showing a higher amount of flow where SR-based methods almost perfectly match the LiDAR HR. This might not be due to the time taken to flow to that point but might be due to some other processes such as slope failure and debris because the peak of rainfall caused that flow, and which has created a huge amount of surface failures in the study area. The RMSE also has shown the model output in line with the kappa coefficient; in channel flow, ESRGAN has lower RMSE and higher improvement in Kappa, which can be explained with the same reasons as in kappa coefficient and NS Index.

5.2.2.3. Results With SR Methods: Colombia

After running the models in Colombia, the overall results show that the SR based methods have improved the quality of the model at least in two dimensions of flow extent and total discharge. We could not analyse the root mean squared error and discharge due to the lack of ground truth data. The use of SR with HR data was not used in this case because it was for testing the capacity of SR based methods, not an applicability test. We observed that the SR-based methods could improve the quality of the outputs, but the model's overall performance is subpar. This model could not obtain good results with our 20 simulations for each scenario, and more extensive simulations with manual calibration might improve the results. The problem with modelling was that it generated a very high number of false-positive with high-resolution and SR data decreasing the kappa value. If we just compare it using the accuracy metric model can be reliable, but with Cohen's kappa values, we can observe that the model is purely unrealistic and cannot show better values. Furthermore, because of the automated calibration using gradient descent, we could not observe the

intermediate results, and the model mostly relied on the Kappa values, which might have limited the cases of slope failures. To further analyse this, Figure 45 shows that, even though in most of the inventory areas the flow is near the inventory data, there is a large amount of solid and liquid flow in the middle of the study area. Furthermore, when calculating Cohen's kappa, the false negatives are taken very seriously, and that created the low values of Cohen's Kappa. The DEM used as HR, in this case, was from ALOS PALSAR, and it had some post-processing done, but due to the lack of features, such errors occurred. A very high-resolution DEM such as LiDAR is essential for this area to have better modelling result.

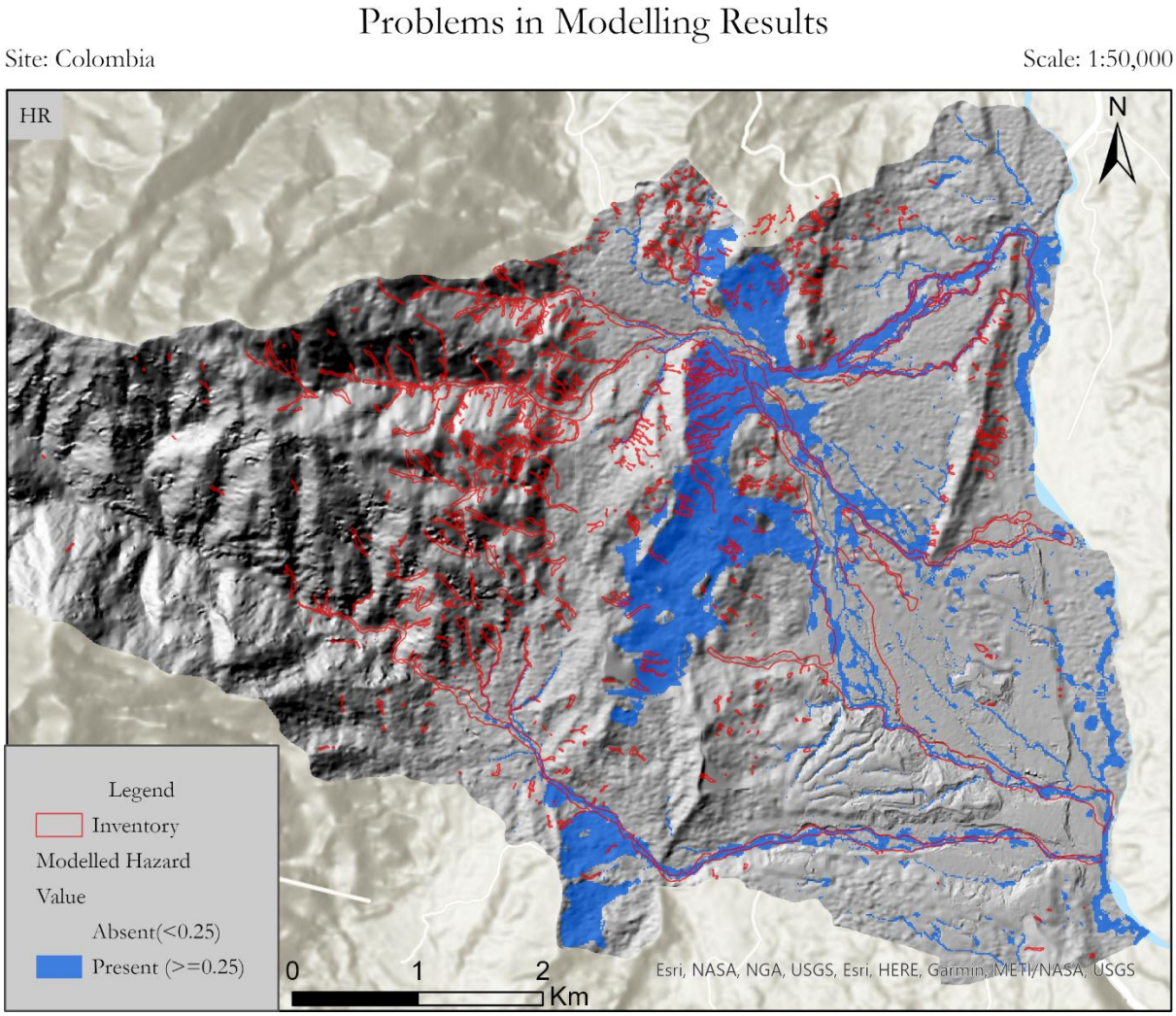


Figure 45: Error and false negative in the modelling output with high-resolution data at Mocoa region of Colombia. Basemap Sources: ESRI, HERE, NASA and USGS

The TanDEMx data being observed at 12.5 meters, similar to ALOS PALSAR data, have lower performance than that of the SRTM SR data because firstly, the ALOS DEM was post-processed to generate better terrain represented, but in the case of TanDEM-X, we did not perform much post-processing. Furthermore, the TanDEM-X had a higher amount of noise in the left parts (mountains) of the study area and middle parts where mountain or slope was available, which has reduced its capability. The flatter terrain was better in the case of TanDEM-X compared to SRTM DEM, but a major impact on the physically based modelling was due to the flow from mountains and slope failures. To further improve the results, hydrologically corrected TanDEM-X might be better to use.

If we observe the overall results on visual analysis of these outputs, the bicubic interpolation has very limited flow, limiting the model's improvement. This is because of the low resolution of the model; the failures are limited, and also the flow in the northern part is very low. However, with all other DEMs, the false-positive hazard is too much, making it more problematic. Even though those had more Cohen's kappa values, they seem to have failed to generate more realistic data than the inventory. The major reasons for such cases are (1) due to lack of high-resolution terrain representation, major streams or flow networks were not visible in DEMs, limiting the generation of flow direction, and fluid and solid usually flow in different directions than actual. (2) The bicubic interpolation techniques smoothened the terrain to a larger extent, and we hypothesize that it can limit the number of slope failures because of more stability in smooth terrain than crisper terrain. However, no published work that evaluates different interpolation techniques has what kind of impact on slope stability modelling, making it an open question for research. The closest research we could found was of Fuchs et al. (2014), which shows that with good input parameters and calibration, the higher resolution data can better generate slope failures, which can also be similar for different interpolation techniques research is required in this area. (3) in SR and other data, the slope failures in the wrong locations might have deviated water flow in different directions, generating more false positives.

The overall improvement in total discharge is very impressive; the reason to have such high error in SRTM bicubic interpolation is because all the models have very high false-positive flow in the middle part, which essentially increase the total discharge. However, in the case of bicubic interpolation, the amount of flow is limited, contributing to the lower amount of total discharge. This shows the overall improvement in SR-based methods, but we cannot still be sure because of the uncertainties present in the HR data and the lack of the ground truth on total discharge. The observed discharge is required for an absolute comparison, and similar to the case of Dominica SR based methods have improved the amount of discharge. Furthermore, the TanDEMx data had more error than that of the SR-based methods and that DEM also had similar false positives in the central part of the study area, indicating that SR-based methods are performing better.

5.3. Comparison of the Models

It is very difficult to decide which model works better in the case of SR, but by comparing the number of parameters to train, we can make a reasonable choice. To further evaluate our work, we have plotted the % improvement in MSE vs the number of model parameters in Figure 46. The model improvement in the EBRN model is higher in terms of MSE improvement compared to the ESRGAN model, which is similar to the observation of Anwar et al. (2020). EBRN is a PSNR oriented model and processes the different frequency of data with different model depth; the model's performance is better in the inference area. However, very small changes in MSE after certain values do not have an impact on modelling as we observed that both models have a very similar improvement in Cohen's kappa values for LiDAR DEM in Dominica. The number of parameters for the EBRN model is higher (3x) than the ESRGAN Generator model, but being Non-GAN based model, it is easier to train the EBRN model even though the training process is slower compared to training the ESRGAN model (Kodali et al., 2017). Since the ESRGAN model is more focused on visually better images, the ESRGAN output for Austria has higher SSIM as shown in the results section, but EBRN has shown better performance in terms of inference. In terms of improving the quality of DEM, the EBRN model is better. However, even though the overall SSIM and PSNR are better with the EBRN model, we have observed that they both have a similar impact on multi-hazard modelling; this will be further explained in the next section.

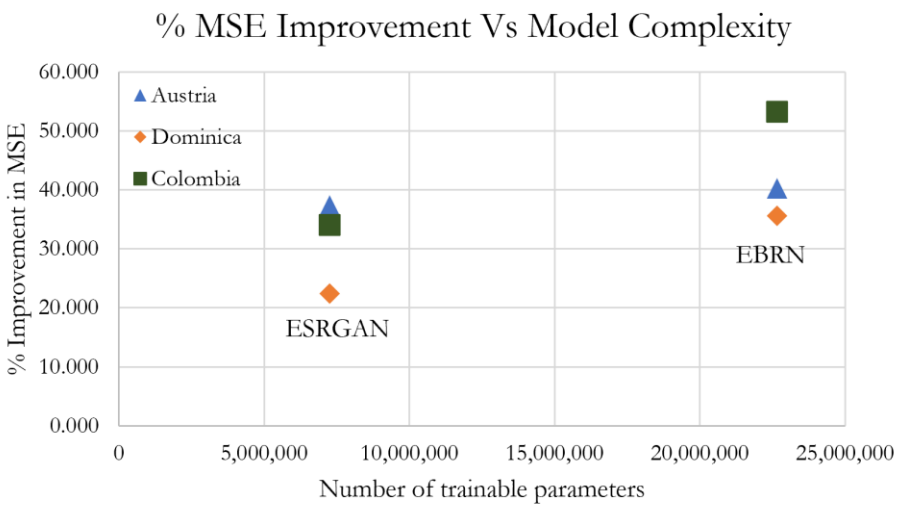


Figure 46: Improvement in MSE by model vs the complexity of the model

5.4. Applicability of SR in Multi-Hazard Modelling at Data-Poor Regions

After all the observation and modelling, we have seen that the SR models can improve the DEM quality, making it applicable to use in data-poor regions. In mapping geomorphological features, it is impossible to expect SR-based methods to predict and create features, but if the features are present with coarse resolution in low-resolution data, it is possible to visualize them with a crisp boundary. In the case of modelling, the results are very similar for channel and non-channel flow. In practice, most of the modelling works are done with including the channel flow in case of hydro-meteorological hazards, and our approach showed that the improvement in the global dataset by using Super-Resolution than Bicubic Interpolation can improve the results by 8-17% depending on the modelling scenario in Dominica and 20-24% In Colombia. Considering the model complexity, the ESRGAN model is suitable for complex areas with more noise, and EBRN is more suitable for less complex and flatter terrain with higher noise.

There is still, few non-technical challenges to use the models limited by the capacity and time of modelling organization to develop such complex models; in the real-world project, the funds and time are extremely limited, and such projects might not have enough time to generate such models to perform deep learning training and analysis to create SR models. Furthermore, the current model we developed can be used in some of the cases, but its performance largely depends on the terrain in consideration, and it might not work globally with the same capacity, which will be further discussed in the limitation section. Another problem that we see with the approach is, Deep Learning based algorithms need strong programming background and knowledge to run them, and inference from them and small changes in the input datasets might create a large discrepancy in the results. Such technical capacity might not be available with multiple local organizations in developing countries. Furthermore, a lack of knowledge about the existence of such models and processes can also limit the applicability of the SR in multi-hazard modelling.

5.5. Advancement and Limitation of the study

The research that we have performed has first tested the applicability of Super-Resolution in physically based modelling and proved that SR-based methods can improve the quality of physically based modelling in data-poor regions. Furthermore, we also have advanced the knowledge on Super-Resolution in Geoscientific data than other research as per our knowledge such as, (1) we used novel TopoLoss function

which considers differentiation in the layers itself while the majority of deep learning models perform differentiation in the different time epochs only, and such novel TopoLoss has shown the advancement in creating better SR outputs compared to that of L1 and L2 loss. (2) The modification of curriculum learning to curriculum examination also has proven to be a good way to train the DL models with complex loss functions. (3) we also developed an overlapping inference concept inspired by photogrammetric observations; even though a similar approach was used by Kubade et al. (2021, 2020), their approach was to use it on the model, which overamplified the model accuracy, but we used it during the inference which did not have any influence in training and testing but created, even more, better results in inference. (4) Definition of multiple approaches for SR-DEM evaluation was more of necessity than novelty because we were first to test the applicability of SR in geoscientific modelling. We needed to rethink the existing methods such as SSIM to evaluate our work in better ways. That led us to develop a novel approach of evaluating using derivatives of SR than that of the DEM itself and derivative evaluation with KDE functions. (5) This research also trained the model in data from one source (measurement techniques) and used it in data from another source successfully, unlike any existing SR methods in both computer vision and SR-DEM based approach where they train and test the model with HR and its degraded counterpart rather than testing on another real LR data. (6) Finally, this research has shown higher accuracy than that of the existing SR models for DEM analysis and proven to improve the quality in a similar dataset by more than 10 % than existing methods of OpenDEM (2021).

Even though there is advancement in the research, there are also few limitations of the study; being heavily trained in Austria DEM, the DEM is biased towards the mountainous terrain, which was also a necessity for physically based modelling because most of the flow and slope failure processes occur in the mountainous terrain. However, when applied to flatter terrain or terrain with a very gentle slope, the improvement might not be as much as expected. However, in the case of Dominica, in flatter terrain also the results were satisfactory. The model is also inapplicable in terrain above 8000 meters and below sea levels because of the limitations during the training, as a solution, the below-sea-level elevation could be shifted to zero, and for above 8000 meters, it could be shifted to lower elevations. Another limitation is to use it together with elevations that are null values; if any null values exist in the model, we had to fill it before using SR-based methods; otherwise, the DL model cannot generate the results making all values null. This might be difficult to have in global DEM because SRTM is not filled in most of the world. Another limitation with using the SR-based method is that it is not much usable in geomorphological mapping because the model's output mostly depends on the input, making it not much applicable in mapping features. The calibration of the physically based models was done automatically using the GD methods. However, due to the inability of the model to understand the relationship between multiple input maps and a limited number of calibration simulations, the model could not find the global minima with the lowest error. Furthermore, evaluating or manual calibration of such a large number of simulations was impossible, and calibrating using only Cohen's Kappa largely reduced the number of slope failures. In the Dominica study site, the TanDEM-X data was very noisy, followed by lesser noise in Colombia, leading to its lower accuracy than SRTM SR methods.

6. CONCLUSION AND RECOMMENDATION

To conclude, with two different types of Super-Resolution models and physically based modelling at two different sites, we have revealed that the Super-Resolution can increase the applicability of physically based multi-hazard modelling in data scare regions. To further explain, Super-Resolution models could improve the Cohen's Kappa values by 8-25% (for EBRN) in different locations, whereas the ESRGAN model had improvement ranging from 17-21%. This improvement was largely dependent on the noise available in the low-resolution data and terrain characteristics. Furthermore, with the use of multiple test matrices, we could prove that the results in all four aspects of the model results, such as flow extent, height, and duration, are improved with the application of SR-based methods than traditional methods. In both Super-Resolution models, the quality of the output from physically based multi-hazard modelling is increased significantly. This shows the possibility of using free and open data in the data-scarce region for modelling, monitoring, and mitigating the hydrometeorological natural hazards. However, the applicability of Super-Resolution is also dependent on uncertainties in the dataset, consistency of the global datasets, technological understanding, and willingness to use from different stakeholders.

The research was mostly focused on the DEM data resolution improvement because of its major role in most of the numerical modelling and its global availability and complexity. Models such as single hazard and multi-hazard models simulated through statistical to deep learning-based DEM are extensively used in natural hazard modelling. With DEM data, we could prove that SR-based methods can be suitable for resolution improvement and which implies that these methods could also be suitable for other datasets where global free data and its high-resolution counterpart are available. The SR-based models also have shown an excellent capacity to increase the spatial variability and crispness in the images compared to traditional techniques such as bicubic interpolation. Furthermore, in the case of application on a global dataset where many uncertainties are present, the ESRGAN based model is more suitable than the EBRN model because of its generative nature. EBRN can process different level of frequency information with different depth of the network, which can be more usable in the case where less error is required in the output. In contrast, in the case of higher visual quality and derivative reconstruction, ESRGAN based model is suited to increase the crispness and generate better-looking images. We have also demonstrated that even though computationally complex, the ESRGAN model is more flexible to a different type of noise present in the input data due to its generative nature and can perform similarly to EBRN in data with lesser noise.

In the technological advancement part of the work, to reconstruct better slope and aspect, we developed a novel loss function that penalizes the model for not generating better slopes, improving the overall quality of our results. Furthermore, to train with the more complex loss functions, we modified the existing curriculum learning approach to the curriculum examination approach where we use gradually more complex loss function than data itself. Additionally, because there is not much published research to evaluate Super-Resolution in data without a high-resolution counterpart, we had to develop novel evaluation methods using derivatives of the SR Images rather than SR images themselves. This also inspired us to check quantitative SSIM improvement using derivatives rather than that of the DEM itself. Due to the lack of standard comparison data, we also had to evaluate based on relative improvement than absolute values, which can be detrimental because of the model's bias on specific locations and noise in the input data. With such evaluations, it has been shown in both test and inference data that our approaches have proven to work better than published works and shown better performance in the overall improvement of modelling quality. However, SR methods are largely dependent on the quality of input data; the mapping hazards using

this approach are limited to only the features present in the low-resolution dataset. Furthermore, we have also observed with available commercial DEM that the SR-based methods can only improve quality compared to traditional processing methods but cannot reach the level of data that are collected in higher resolution, making this a temporary solution when data is not available.

The overall improvement in physically based modelling with global dataset using SR-based methods is better compared to interpolation techniques. However, such improvement is nowhere near the actual high-resolution DEM, making the SR-based solution only useful in case of an absolute lack of data and capacity to capture the higher-resolution dataset. This phenomenon of incapacity to increase the information content is further supported by the theory of data processing inequality, meaning that we can only improve the information quality but not improve the information content using SR-based methods. Furthermore, in case of the absolute impossibility to capture a new high-resolution dataset with cost, technology, and time limitation, SR-based methods can improve the quality of models, making it a viable solution. However, the real-world applicability of the SR-based methods is also limited due to factors such as the availability of training data, technological capacity, and time.

For further improvement of the work, (1), we would like to recommend that the more advanced model that is developed in 2021 for natural image Super-Resolution might also further improve the generalization of the model with higher capacity to generate SR images, and those models could be tested for geoscience data as well. (2), the loss function that we developed is still in the immature phase. It could be further improved by including other topographic characteristics in loss function such as TWI error function, error in channel location, function to estimate the error in drainage density etc. might improve the quality of physically-based models. (3), the data that we used in training the current model is from Austria and which might not work well in cases of very different terrains. To improve such quality, the addition of more data from different types of terrain would help to generate a global model which can generate better global free data. (4), there is still a need to develop public training and testing data as well as standard evaluation methods for geoscientific Super-Resolution, which will make it possible to compare different models and their quality without bias. (5), change of the error functions and faster optimization algorithms or more simulations are recommended to generate better results from multi-hazard modelling; the current calibration approach, due to the lack of available time, could not perform better than that of the manual calibration. (6), the physics-aware SR-based methods can be excellent in climate data downscaling to downscale the climate data than existing mathematical, statistical, and machine learning methods. Further research on that area can also open better possibilities to understand and reduce the global impact of climate change.

LIST OF REFERENCES

Agren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J., & Arp, P. A. (2014). Evaluating digital terrain indices for soil wetness mapping-a Swedish case study. *Hydrology and Earth System Sciences*, *18*(9), 3623–3634. https://doi.org/10.5194/hess-18-3623-2014

Agustsson, E., & Timofte, R. (2017). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2017-July*, 1122–1131. https://doi.org/10.1109/CVPRW.2017.150

Aizawa, K., Fujimoto, A., Otsubo, A., Ogawa, T., Matsui, Y., Tsubota, K., & Ikuta, H. (2020). Building a Manga Dataset "Manga109" with Annotations for Multimedia Applications. *IEEE Multimedia*, *27*(2), 8–18. https://doi.org/10.1109/MMUL.2020.2987895

Al-falluji, R., Guirguis, S., & Youssif, A. (2017). Single Image Super Resolution Algorithms: A Survey and Evaluation. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *6*(9), 2278–1323. https://www.researchgate.net/publication/320286761

Alganci, U., Besol, B., & Sertel, E. (2018). Accuracy assessment of different digital surface models. *ISPRS International Journal of Geo-Information*, *7*(3), 1–16. https://doi.org/10.3390/ijgi7030114

Anwar, S., Khan, S., & Barnes, N. (2020). A Deep Journey into Super-resolution: A Survey. In *ACM Computing Surveys* (Vol. 53, Issue 3). https://doi.org/10.1145/3390462

AON. (2017). *Global Catastrophe Recap* (Issue April). http://thoughtleadership.aon.com/Documents/20191009_analytics-if-september-global-recap.pdf

Argudo, O., Chica, A., & Andujar, C. (2018). Terrain super-resolution through aerial imagery and fully convolutional networks. *Computer Graphics Forum*, *37*(2), 101–110. https://doi.org/10.1111/cgf.13345

Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997). Teaching by Examples: Implications for the Process of Category Acquisition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 586–606. https://doi.org/10.1080/713755719

Bannore, V. (2009). *Iterative-Interpolation Super-Resolution Image Reconstruction* (Vol. 195). Springer. https://doi.org/10.1007/978-3-642-00385-1

Bashir, S. M. A., Wang, Y., & Khan, M. (2021). A Comprehensive Review of Deep Learning-based Single Image Super-resolution. *CoRR*, *abs/2102.0*. http://arxiv.org/abs/2102.09351

Beaudry, N. J., & Renner, R. (2012). An intuitive proof of the data processing inequality. *Quantum Information and Computation*, *12*(5–6), 432–441. https://doi.org/10.26421/qic12.5-6-4

Behlert, B., Diekjobst, R., Felgentreff, D. C., Manandhar, T., Mucke, P., Pries, P. D. L., Radtke, D. K., & Weller, D. (2020). *World Risk Report 2020: Forced Displacement and Migration* (L. Kirch & R. Prütz (eds.)). Bündnis Entwicklung Hilft and Ruhr University Bochum – Institute for International Law of Peace and Armed Conflict (IFHV). https://reliefweb.int/sites/reliefweb.int/files/resources/WorldRiskReport-2020.pdf

Bell, R., & Glade, T. (2011). Multi-hazard analysis in natural risk assessments. *WIT Transactions on State of the Art in Science and Engineering*, *53*, 1755–8336. https://doi.org/10.2495/978-1-84564-650-9/01

Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications*, *34*(2), 825–832. https://doi.org/10.1016/j.eswa.2006.10.022

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *ACM International Conference Proceeding Series*, *382*. https://doi.org/10.1145/1553374.1553380

Bhunia, G. S., Shit, P. K., & Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, *17*(2), 114–126. https://doi.org/10.1016/j.jssas.2016.02.001

Blackman, N. J. M., & Koval, J. J. (2000). Interval estimation for Cohen's kappa as a measure of agreement. *Statistics in Medicine*, *19*(5), 723–741. https://doi.org/10.1002/(SICI)1097-0258(20000315)19:5<723::AID-SIM379>3.0.CO;2-A

Bottou, Leon. (1991). Stochastic Gradient Learning. *In Proceedings of Neuro-Nîmes. EC2*.

Bottou, Léon. (1999). On-line Learning and Stochastic Approximations. In *On-Line Learning in Neural Networks* (pp. 9–42). Cambridge University Press. https://doi.org/10.1017/cbo9780511569920.003

Briggs, W. (2016). Statistical and Physical Models. In *Uncertainty: The Soul of Modeling, Probability {\&} Statistics* (pp. 153–202). Springer International Publishing. https://doi.org/10.1007/978-3-319-39756-6_9

Bulyshev, A., Amzajerdian, F., Roback, E., & Reisse, R. (2014). A super-resolution algorithm for enhancement of flash lidar data: flight test results. *Computational Imaging XII*, *9020*, 90200B. https://doi.org/10.1117/12.2035251

Bulyshev, A., Vanek, M., Amzajerdian, F., Pierrottet, D., Hines, G., & Reisse, R. (2011). A super-resolution algorithm for enhancement of FLASH LIDAR data. *Computational Imaging IX*, *7873*, 78730F. https://doi.org/10.1117/12.876283

Burrough, P. A., & McDonell, R. A. (1998). *Principles of Geographical Information Systems* (3rd ed.). Oxford University Press.

Candès, E. J., & Fernandez-Granda, C. (2013). Super-Resolution from Noisy Data. *Journal of Fourier Analysis and Applications*, *19*(6), 1229–1254. https://doi.org/10.1007/s00041-013-9292-3

Chang, K. T., Merghadi, A., Yunus, A. P., Pham, B. T., & Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, *9*(1), 1603–1604. https://doi.org/10.1038/s41598-019-48773-2

Chen, B. C., Chen, A. X., Chai, X., & Bian, R. (2019). A Statistical Test of the Effect of Learning Rate and Momentum Coefficient of Sgd and Its Interaction on Neural Network Performance. *Proceedings - 2019 3rd International Conference on Data Science and Business Analytics, ICDSBA 2019*, 276–281. https://doi.org/10.1109/ICDSBA48748.2019.00065

Chen, Z., Wang, X., Xu, Z., & Hou, W. (2016). CONVOLUTIONAL NEURAL NETWORK BASED DEM SUPER RESOLUTION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B3*, 247–250. https://doi.org/10.5194/isprs-archives-xli-b3-247-2016

Chi, C.-Y., Li, W.-C., & Lin, C.-H. (2009). Convex Optimization Problems. In *Convex Optimization for Signal Processing and Communications* (2nd ed.). Cambridge University Press. https://doi.org/10.1201/9781315366920-5

Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L., & Thuerey, N. (2020). Learning temporal coherence via self-supervision for GAN-based video generation. *ACM Transactions on Graphics*, *39*(4). https://doi.org/10.1145/3386569.3392457

Cox, D. R., & Miller, H. D. (2017). The theory of stochastic processes. In *CRC Press* (1st ed.). https://doi.org/10.1201/9780203719152

CRED, & UNDRR. (2020). *An overview of the last 20 years*. https://reliefweb.int/report/world/human-cost-disasters-overview-last-20-years-2000-2019

Davis, R. (2017, April 2). *Colombia – Over 200 Dead After Massive Landslide and Floods in Mocoa – FloodList*. Floodlist. http://floodlist.com/america/colombia-flood-landslide-mocoa-april-2017

Demiray, B. Z., Sit, M., & Demir, I. (2020). D-SRGAN: DEM Super-Resolution with Generative Adversarial Networks. *SN Ccomputer Science*, *2*, 48. https://doi.org/10.31223/osf.io/frd8x

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Dewitte, O., Daoudi, M., Bosco, C., & Van Den Eeckhaut, M. (2015). Predicting the susceptibility to gully initiation in data-poor regions. *Geomorphology*, *228*, 101–115. https://doi.org/10.1016/j.geomorph.2014.08.010

DLR. (2010). *TanDEM-X - the Earth in three dimensions* . German Aerospace Center. https://www.dlr.de/content/en/missions/tandem-x.html

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9906 LNCS*, 391–407. http://arxiv.org/abs/1608.00367

Durall, R., Chatzimichailidis, A., Labus, P., & Keuper, J. (2021). Combating mode collapse in GAN Training: An empirical analysis using hessian eigenvalues. *VISIGRAPP 2021 - Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, *4*, 211–218. https://doi.org/10.5220/0010167902110218

Fan, L., Lehmann, P., McArdell, B., & Or, D. (2017). Linking rainfall-induced landslides with debris flows runout patterns towards catchment scale hazard assessment. *Geomorphology*, *280*, 1–15. https://doi.org/10.1016/j.geomorph.2016.10.007

Finlay, P. J., & Fell, R. (1997). Landslides: Risk perception and acceptance. *Canadian Geotechnical Journal*, *34*(2), 169–188. https://doi.org/10.1139/t96-108

Floors, R., Enevoldsen, P., Davis, N., Arnqvist, J., & Dellwik, E. (2018). From lidar scans to roughness maps for wind resource modelling in forested areas. *Wind Energy Science*, *3*(1), 353–370. https://doi.org/10.5194/wes-3-353-2018

Franchini, M., Wendling, J., Obled, C., & Todini, E. (1996). Physical interpretation and sensitivity analysis

of the TOPMODEL. *Journal of Hydrology*, *175*(1–4), 293–338. https://doi.org/10.1016/S0022-1694(96)80015-1

Fuchs, M., Torizin, J., & Kühn, F. (2014). The effect of DEM resolution on the computation of the factor of safety using an infinite slope model. *Geomorphology*, *224*, 16–26. https://doi.org/10.1016/j.geomorph.2014.07.015

GEOSENSENL. (2018, February 27). *The importance of Spatial Resolution in Elevation Data – geosense.nl*. Geosensenl. https://www.geosense.nl/importance-of-spatial-resolution-elevation-data/

Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). *Deep Learning* (1st ed.). MIT Press. http://files.sig2d.org/sig2d14.pdf#page=5

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *3*(January), 2672–2680. http://www.github.com/goodfeli/adversarial

Government of Commonwealth of Dominica. (2017). *Post-Disaster Needs Assessment Hurricane Maria*. https://reliefweb.int/sites/reliefweb.int/files/resources/dominica-pdna-maria.pdf

Guan, J., Pan, C., Li, S., & Yu, D. (2019). SRDGAN: learning the noise prior for Super Resolution with Dual Generative Adversarial Networks. *Computer Science*, *abs/1903.1*. http://arxiv.org/abs/1903.11821

Gudendorf, G., & Segers, J. (2010). *Extreme-Value Copulas* (pp. 127–145). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12465-5_6

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, *98*(3–4), 239–267. https://doi.org/10.1007/s00703-007-0262-7

Hallouin, T. (2021). *hydroeval: an evaluator for streamflow time series in Python*. https://doi.org/10.5281/ZENODO.4709652

Ham, H., Jun, T. J., & Kim, D. (2020). Unbalanced GANs: Pre-training the Generator of Generative Adversarial Network using Variational Autoencoder. *Computer Science, Mathematics*, *abs/2002.0*. http://arxiv.org/abs/2002.02112

Han, D. (2013). Comparison of Commonly Used Image Interpolation Methods. *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, 1556–1559. https://doi.org/10.2991/iccsee.2013.391

Haussler, D. (1992). Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, *100*(1), 78–150. https://doi.org/10.1016/0890-5401(92)90010-D

Hawker, L., Bates, P., Neal, J., & Rougier, J. (2018). Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM. *Frontiers in Earth Science*, *6*, 233. https://doi.org/10.3389/feart.2018.00233

Hussain Mujtaba -. (2020, August 29). *An Introduction to Rectified Linear Unit (ReLU) | What is RelU?* Grea Learning. https://www.mygreatlearning.com/blog/relu-activation-function/

Hwang, J. N., Lay, S. R., & Lippman, A. (1994). Nonparametric Multivariate Density Estimation: A Comparative Study. *IEEE Transactions on Signal Processing*, *42*(10), 2795–2810. https://doi.org/10.1109/78.324744

IDEAM. (2010). *Leyenda nacional de coberturas para la tierra. Metodología Corine Land Cover adaptada para Colombia. Escala 1:100.000*.

Inouye, T. (1964). The super resolution of gamma-ray spectrum. *Nuclear Instruments and Methods*, *30*(2), 224–228. https://doi.org/10.1016/0029-554X(64)90275-7

Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., & Huang, F. (2020). Real-world super-resolution via kernel estimation and noise injection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 1914–1923. https://doi.org/10.1109/CVPRW50498.2020.00241

Jia, Y., Ge, Y., Chen, Y., Li, S., Heuvelink, G. B. M., & Ling, F. (2019). Super-resolution land cover mapping based on the convolutional neural network. *Remote Sensing*, *11*(15), 1815. https://doi.org/10.3390/rs11151815

Jiao, D., Wang, D., Lv, H., & Peng, Y. (2020). Super-resolution reconstruction of a digital elevation model based on a deep residual network. *Open Geosciences*, *12*(1), 1369–1382. https://doi.org/10.1515/geo-2020-0207

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004

Kelleher, C., Wagener, T., & McGlynn, B. (2015). Model-based analysis of the influence of catchment

properties on hydrologic partitioning across five mountain headwater subcatchments. *Water Resources Research*, *51*(6), 4109–4136. https://doi.org/10.1002/2014WR016147

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kodali, N., Abernethy, J., Hays, J., & Kira, Z. (2017). On Convergence and Stability of GANs. *ArXiv: Artificial Intelligence*, *1705.07215*, 1–18. http://arxiv.org/abs/1705.07215

Komendantova, N., Mrzyglocki, R., Mignan, A., Khazai, B., Wenzel, F., Patt, A., & Fleming, K. (2014). Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: Feedback from civil protection stakeholders. *International Journal of Disaster Risk Reduction*, *8*, 50–67. https://doi.org/10.1016/j.ijdrr.2013.12.006

Kubade, A., Patel, D., Sharma, A., & Rajan, K. S. (2021). AFN: Attentional Feedback Network Based 3D Terrain Super-Resolution. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12622 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-69525-5_12

Kubade, A., Sharma, A., & Rajan, K. S. (2020). Feedback Neural Network Based Super-Resolution of DEM for Generating High Fidelity Features. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 1671–1674. https://doi.org/10.1109/IGARSS39084.2020.9323310

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., & Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018

Land Carinthia. (2015, December 31). *Digital Terrian Model Carinthia*. Open Data Austria. https://www.data.gv.at/katalog/dataset/a188992b-4071-45c3-99ce-65662395ebe6

Land Salzburg. (2016, September 15). *Digital Terrian Model 5 m (DTM) Land Salzburg*. Open Data Austria. https://www.data.gv.at/katalog/dataset/d585e816-1a36-4c76-b2dc-6db487d22ca3

Land Tirol. (2018, February 22). *Digital Terrian Model Tyrol*. Open Data Austria. https://www.data.gv.at/katalog/dataset/land-tirol_tirolgelnde

Ledford, A. W., & Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *59*(2), 475–499. https://doi.org/10.1111/1467-9868.00080

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*(12), 105–114. https://doi.org/10.1109/CVPR.2017.19

Leong, W. J., & Horgan, H. J. (2020). DeepBedMap: Using a deep neural network to better resolve the bed topography of Antarctica. *The Cryosphere Discussions*, 1–27. https://doi.org/10.5194/tc-2020-74

Li, G., Zhao, J., Murray, V., Song, C., & Zhang, L. (2019). Gap analysis on open data interconnectivity for disaster risk research. *Geo-Spatial Information Science*, *22*(1), 45–58. https://doi.org/10.1080/10095020.2018.1560056

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2017-July*, 1132–1140. http://arxiv.org/abs/1707.02921

Lindersson, S., Brandimarte, L., Mård, J., & Di Baldassarre, G. (2020). A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies. *WIREs Water*, *7*(3), 1–20. https://doi.org/10.1002/wat2.1424

Liu, B., Siu, Y. L., & Mitchell, G. (2016). Hazard interaction analysis for multi-hazard risk assessment: A systematic classification based on hazard-forming environment. *Natural Hazards and Earth System Sciences*, *16*(2), 629–642. https://doi.org/10.5194/nhess-16-629-2016

Liu, C., Du, W., & Tian, X. (2018). Lunar DEM Super-resolution reconstruction via sparse representation. *Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017*, *2018-Janua*, 1–5. https://doi.org/10.1109/CISP-BMEI.2017.8301904

Liu, Z., Cheng, L., Hao, Z., Li, J., Thorstensen, A., & Gao, H. (2018). A Framework for Exploring Joint Effects of Conditional Factors on Compound Floods. *Water Resources Research*, *54*(4), 2681–2696. https://doi.org/10.1002/2017WR021662

Luo, Y., Zhou, L., Wang, S., & Wang, Z. (2017). Video Satellite Imagery Super Resolution via Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2398–2402. https://doi.org/10.1109/LGRS.2017.2766204

Masters, J., & Grenci, L. (2017, April 3). *Meteorology of Saturday's Colombian Flood Disaster That Killed 254*.

Weather Underground. https://www.wunderground.com/blog/JeffMasters/meteorology-of-saturdays-colombian-flood-disaster-that-killed-254.html
MATLAB. (2020). *Compute peak signal-to-noise ratio (PSNR) between images - Simulink - MathWorks Benelux*. MathWorks. https://nl.mathworks.com/help/vision/ref/psnr.html
MATLAB. (2021). Resize image - MATLAB imresize. In *Mathworks help center* (No. 2021a). Mathworks. https://www.mathworks.com/help/images/ref/imresize.html
Mattivi, P., Franci, F., Lambertini, A., & Bitelli, G. (2019). TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, *4*(1). https://doi.org/10.1186/s40965-019-0066-y
McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash–Sutcliffe Efficiency Index. *Journal of Hydrologic Engineering*, *11*(6), 597–602. https://doi.org/10.1061/(asce)1084-0699(2006)11:6(597)
Moon, S. H., & Choi, H. L. (2016). Super-resolution based on deep learning technique for constructing digital elevation model. *AIAA Space and Astronautics Forum and Exposition, SPACE 2016*. https://doi.org/10.2514/6.2016-5608
Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900.
Mousavi, H. S., Guo, T., & Monga, V. (2018). Deep Image Super Resolution via Natural Image Priors. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2018-April*, 1483–1487. http://arxiv.org/abs/1802.02721
NASA. (2009, June). *ASTER Global Digital Elevation Map*. Jet Propulsion Laboratory. https://asterweb.jpl.nasa.gov/gdem.asp
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Nguyen, P., Bhaskar, A. V., Shivadekar, S., Yesha, Y., & Halem, M. (2019, December 13). A Super Resolution Convolutional Neural Network approach for simulating NASA's SMAP Radar observations from Radiometer Data. *AGU Fall Meeting 2019*. https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/550204
Ooi, Y. K., & Ibrahim, H. (2021). Deep Learning Algorithms for Single Image Super-Resolution: A Systematic Review. *Electronics*, *10*(7), 867. https://doi.org/10.3390/electronics10070867
OpenCV. (2021). *OpenCV: Geometric Image Transformations* (3.4.14). Open Source Computer Vision. https://docs.opencv.org/3.4/da/d54/group__imgproc__transform.html
OpenDEM. (2021). *Super-resolution*. OpenDEM. https://www.opendem.info/superResolution.html
Osuteye, E., Johnson, C., & Brown, D. (2017). The data gap: An analysis of data availability on disaster losses in sub-Saharan African cities. *International Journal of Disaster Risk Reduction*, *26*(September), 24–33. https://doi.org/10.1016/j.ijdrr.2017.09.026
Panagiotopoulou, A., & Anastassopoulos, V. (2007). Super-resolution image reconstruction employing Kriging interpolation technique. *2007 IWSSIP and EC-SIPMCS - Proc. 2007 14th Int. Workshop on Systems, Signals and Image Processing, and 6th EURASIP Conf. Focused on Speech and Image Processing, Multimedia Communications and Services*, 144–147. https://doi.org/10.1109/IWSSIP.2007.4381174
Papaioannou, G., Efstratiadis, A., Vasiliades, L., Loukas, A., Papalexiou, S. M., Koukouvinos, A., Tsoukalas, I., & Kossieris, P. (2018). An operational method for Flood Directive implementation in ungauged urban areas. *Hydrology*. https://doi.org/10.3390/hydrology5020024
Park, S. C., Park, M. K., & Kang, M. G. (2003). Super-resolution image reconstruction: A technical overview. In *IEEE Signal Processing Magazine* (Vol. 20, Issue 3, pp. 21–36). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MSP.2003.1203207
Pascanu, R., Mikolov, T., & Bengio, Y. (2012). On the difficulty of training Recurrent Neural Networks. *30th International Conference on Machine Learning, ICML 2013*, *PART 3*, 2347–2355. http://arxiv.org/abs/1211.5063
Polidori, L., & Hage, M. El. (2020). Digital elevation model quality assessment methods: A critical review. In *Remote Sensing* (Vol. 12, Issue 21, pp. 1–36). MDPI AG. https://doi.org/10.3390/rs12213522
Portugués, E. G. (2021). *Notes for Nonparametric Statistics* (6.4.4). Bookdown. https://bookdown.org/egarpor/NP-UC3M/#citation
Pudasaini, S. P. (2012). A general two-phase debris flow model. *Journal of Geophysical Research: Earth Surface*, *117*(3). https://doi.org/10.1029/2011JF002186
Pykes, K. (2020, February 27). *Cohen's Kappa. Understanding Cohen's Kappa coefficient | by Kurtis Pykes | Towards*

Data Science. Towards Data Science. https://towardsdatascience.com/cohens-kappa-9786ceceab58
Qin, C. Z., Zhu, A. X., Pei, T., Li, B. L., Scholten, T., Behrens, T., & Zhou, C. H. (2011). An approach to computing topographic wetness index based on maximum downslope gradient. *Precision Agriculture*, *12*(1), 32–43. https://doi.org/10.1007/s11119-009-9152-y
Qiu, Y., Wang, R., Tao, D., & Cheng, J. (2019). Embedded block residual network: A recursive restoration model for single-image super-resolution. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 4179–4188. https://doi.org/10.1109/ICCV.2019.00428
Raduła, M. W., Szymura, T. H., & Szymura, M. (2018). Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. *Ecological Indicators*, *85*(March 2017), 172–179. https://doi.org/10.1016/j.ecolind.2017.10.011
Rahman, A. ur, & Fang, C. (2019). Appraisal of gaps and challenges in Sendai Framework for Disaster Risk Reduction priority 1 through the lens of science, technology and innovation. *Progress in Disaster Science*, *1*, 100006. https://doi.org/10.1016/j.pdisas.2019.100006
Rata, M., Douaoui, A., Larid, M., & Douaik, A. (2020). Comparison of geostatistical interpolation methods to map annual rainfall in the Chéliff watershed, Algeria. *Theoretical and Applied Climatology*, *141*(3–4), 1009–1024. https://doi.org/10.1007/s00704-020-03218-z
Renieblas, G. P., Nogués, A. T., González, A. M., Gómez-Leon, N., & del Castillo, E. G. (2017). Structural similarity index family for image quality assessment in radiological images. *Journal of Medical Imaging*, *4*(3), 035501. https://doi.org/10.1117/1.jmi.4.3.035501
Rocha, J., Duarte, A., Silva, M., Fabres, S., Vasques, J., Revilla-Romero, B., & Quintela, A. (2020). The importance of high resolution digital elevation models for improved hydrological simulations of a mediterranean forested catchment. *Remote Sensing*, *12*(20), 1–17. https://doi.org/10.3390/rs12203287
Rojas, J. C. (2017, April 3). La avalancha que provoco la peor catastrofe en la historia de Mocoa. *EL TIEMPO*. https://www.eltiempo.com/colombia/otras-ciudades/avalancha-en-mocoa-cifras-de-la-tragedia-y-causas-73984
Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, *27*(3), 832–837. https://doi.org/10.1214/aoms/1177728190
Sanders, B. F. (2017). Hydrodynamic Modeling of Urban Flood Flows and Disaster Risk Reduction. In *Oxford Research Encyclopedia of Natural Hazard Science*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199389407.013.127
Scopus. (2021, April 19). *Scopus - Document Database*. Elsevier. https://www.scopus.com/results/results.uri?sort=plf-f&src=s&sid=85952c8ad696359306a871b89c310caf&sot=a&sdt=a&sl=33&s=TITLE-ABS-KEY%28%22Super+resolution%22%29&origin=patentresults&editSaveSearch=&txGid=f33eff2267e5fcef9079de81f29286b7&zone=queryBar
Servicio Geológico Colombiano (SGC). (2017). *Caracterización del Movimiento en Masa Tipo Flujo del 31 de marzo de 2017 en Mocoa – Putumayo*.
Shen, P., Zhang, L., Chen, H., & Fan, R. (2018). EDDA 2.0: Integrated simulation of debris flow initiation and dynamics considering two initiation mechanisms. *Geoscientific Model Development*, *11*(7), 2841–2856. https://doi.org/10.5194/gmd-11-2841-2018
Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference, ACM 1968*, 517–524. https://doi.org/10.1145/800186.810616
Shin, D., & Spittle, S. (2019). LoGSRN: Deep super resolution network for digital elevation model. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2019-Octob*, 3060–3065. https://doi.org/10.1109/SMC.2019.8914037
Skilodimou, H. D., Bathrellos, G. D., Chousianitis, K., Youssef, A. M., & Pradhan, B. (2019). Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. *Environmental Earth Sciences*, *78*(2), 47. https://doi.org/10.1007/s12665-018-8003-4
Soh, J. W., Park, G. Y., Jo, J., & Cho, N. I. (2019). Natural and realistic single image super-resolution with explicit natural manifold discrimination. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8122–8131.
Stott, R. (2018). The World Bank. In *The World Bank*. https://doi.org/10.1136/bmj.318.7187.822
Sun, J., Xu, Z., & Shum, H. Y. (2011). Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Transactions on Image Processing*, *20*(6), 1529–1542. https://doi.org/10.1109/TIP.2010.2095871
Svensson, C., & Jones, D. A. (2004). Dependence between sea surge, river flow and precipitation in south and west Britain. *Hydrology and Earth System Sciences*, *8*(5), 973–992. https://doi.org/10.5194/hess-8-

973-2004
The World Bank. (2018, July 24). *Dominica | Data*. The World Bank Data. https://data.worldbank.org/country/dominica
The World Bank. (2020a, April 14). *Disaster Risk Management Overview*. The World Bank. https://www.worldbank.org/en/topic/disasterriskmanagement/overview
The World Bank. (2020b, July 8). *Colombia | Data*. The World Bank Data. https://datos.bancomundial.org/pais/colombia
Thompson, M., & Warmink, J. J. (2016). Natural Hazard Modeling and Uncertainty Analysis. In G. E. P. Box (Ed.), *Natural Hazard Uncertainty Assessment: Modeling and Decision Support, Geophysical Monograph* (1st ed., pp. 11–19). American Geophysical Union. https://doi.org/10.1002/9781119028116.ch2
Tian, J., & Ma, K. K. (2011). A survey on super-resolution imaging. In *Signal, Image and Video Processing* (Vol. 5, Issue 3, pp. 329–342). Springer. https://doi.org/10.1007/s11760-010-0204-6
Tilloy, A., Malamud, B. D., Winter, H., & Joly-Laugel, A. (2019). A review of quantification methodologies for multi-hazard interrelationships. In *Earth-Science Reviews* (Vol. 196, p. 102881). Elsevier B.V. https://doi.org/10.1016/j.earscirev.2019.102881
Unidad Nacional de Gestión del Riesgo de Desastres (UNGRD), & Pontificia Universidad Javeriana (PUJ). (2018). *Informe General Del Ejercicio Con Las Consideraciones Supuestas Y Restricciones De Los Resultados Generados*.
van den Bout, B. (2020). *Integrated Physically-Based Multi-Hazard Modelling*. University of Twente.
van den Bout, B., & Jetten, V. G. G. (2018). *OpenLISEM- Multi-Hazard Land Surface Process Model*. https://blog.utwente.nl/lisem/
van den Bout, B., Lombardo, L., van Westen, C. J., & Jetten, V. G. G. (2018). Integration of two-phase solid fluid equations in a catchment model for flashfloods, debris flows and shallow slope failures. *Environmental Modelling & Software*, *105*, 1–16. https://doi.org/10.1016/j.envsoft.2018.03.017
van den Hurk, B., Van Meijgaard, E., De Valk, P., Van Heeringen, K. J., & Gooijer, J. (2015). Analysis of a compounding surge and precipitation event in the Netherlands. *Environmental Research Letters*, *10*(3), 035001. https://doi.org/10.1088/1748-9326/10/3/035001
van Ouwerkerk, J. D. (2006). Image super-resolution survey. In *Image and Vision Computing* (Vol. 24, Issue 10, pp. 1039–1052). Elsevier Ltd. https://doi.org/10.1016/j.imavis.2006.02.026
van Westen, C. J., & Greiving, S. (2017). Multi-hazard risk assessment and decision making. In *Environmental Hazards Methodologies for Risk Assessment and Management* (pp. 31–94). International Water Association. https://doi.org/10.2166/9781780407135_0031
van Westen, C. J., van den Bout, B., Lombardo, L., & Jetten, V. G. G. (2020). *Reducing Debris and Water Flow Risks from landslides for Infrastructure Resilience of the Commonwealth of Dominica*.
Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2018). Generating high resolution climate change projections through single image super-resolution: An abridged version. *IJCAI International Joint Conference on Artificial Intelligence*, *2018-July*, 5389–5393. https://doi.org/10.24963/ijcai.2018/759
von Ruette, J., Lehmann, P., Fan, L., Bickel, S., Or, D., & Anonymous. (2017). STEP-TRAMM; a modeling interface for simulating localized rainfall induced shallow landslides and debris flow runout pathways. *Geophysical Research Abstracts*, *19*, Abstract EGU2017-11408. https://ui.adsabs.harvard.edu/abs/2017AGUFMNH41D..03O/abstract
Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Loy, C. C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11133 LNCS*, 63–79. https://doi.org/10.1007/978-3-030-11021-5_5
Wang, Z., & Bovik, A. C. (2009). Mean squared error: Lot it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, *26*(1), 98–117. https://doi.org/10.1109/MSP.2008.930649
Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861
Wu, Z., & Ma, P. (2020). ESRGAN-based DEM super-resolution for enhanced slope deformation monitoring in lantau island of Hong Kong. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B3), 351–356. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-351-2020
Xu, Z., Chen, Z., Yi, W., Gui, Q., Hou, W., & Ding, M. (2019). Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM. *ISPRS Journal of Photogrammetry and Remote*

Sensing, *150*(August 2018), 80–90. https://doi.org/10.1016/j.isprsjprs.2019.02.008

Yang, J., & Huang, T. (2017). Image super-resolution: Historical overview and future challenges. In *Super-Resolution Imaging* (1st ed., pp. 1–33). CRC Press. https://doi.org/10.1201/9781439819319

Yousefi, S., Pourghasemi, H. R., Emami, S. N., Pouyan, S., Eskandari, S., & Tiefenbacher, J. P. (2020). A machine learning framework for multi-hazards modeling and mapping in a mountainous area. *Scientific Reports*, *10*(1), 12144. https://doi.org/10.1038/s41598-020-69233-2

Zhang, H., Li, Z., Saifullah, M., Li, Q., & Li, X. (2016). Impact of DEM resolution and spatial scale: Analysis of influence factors and parameters on physically based distributed model. *Advances in Meteorology*, *2016*. https://doi.org/10.1155/2016/8582041

Zhang, K., Xue, X., Hong, Y., Gourley, J. J., Lu, N., Wan, Z., Hong, Z., & Wooten, R. (2016). ICRESTRIGRS: A coupled modeling system for cascading flood-landslide disaster forecasting. *Hydrology and Earth System Sciences*, *20*(12), 5035–5048. https://doi.org/10.5194/hess-20-5035-2016

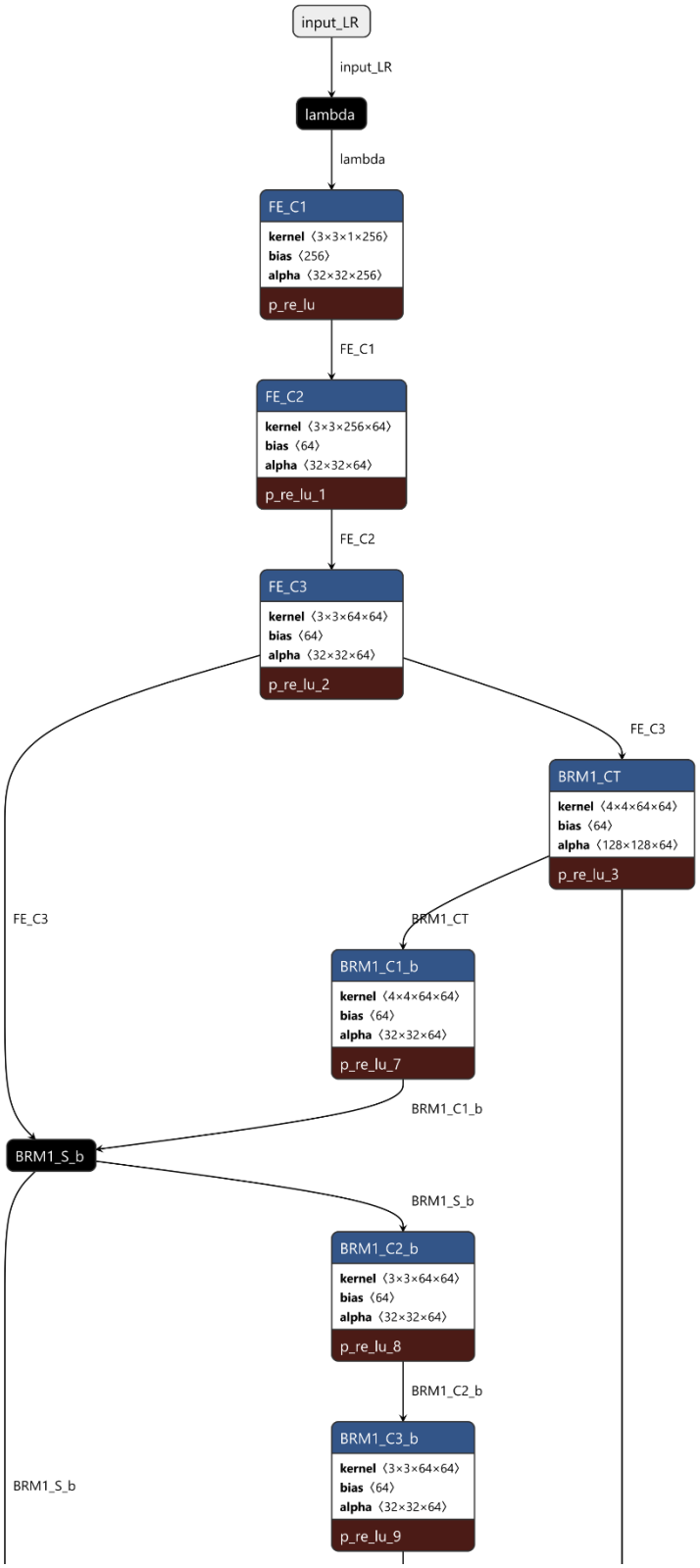
Zhang, Q., & Wu, J. (2015). Image super-resolution using windowed ordinary Kriging interpolation. *Optics Communications*, *336*, 140–145. https://doi.org/10.1016/j.optcom.2014.09.060

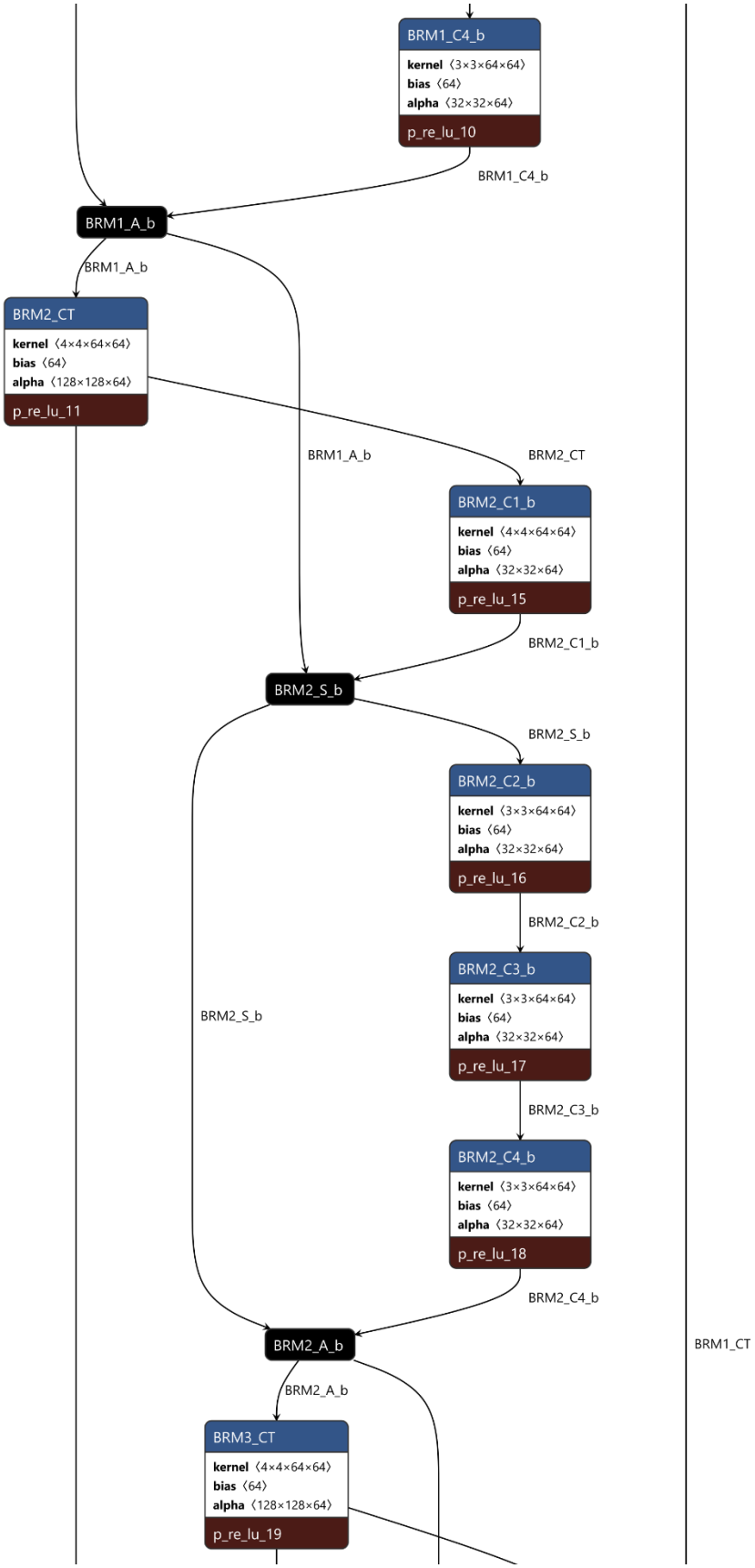
Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 294–310. http://arxiv.org/abs/1807.02758

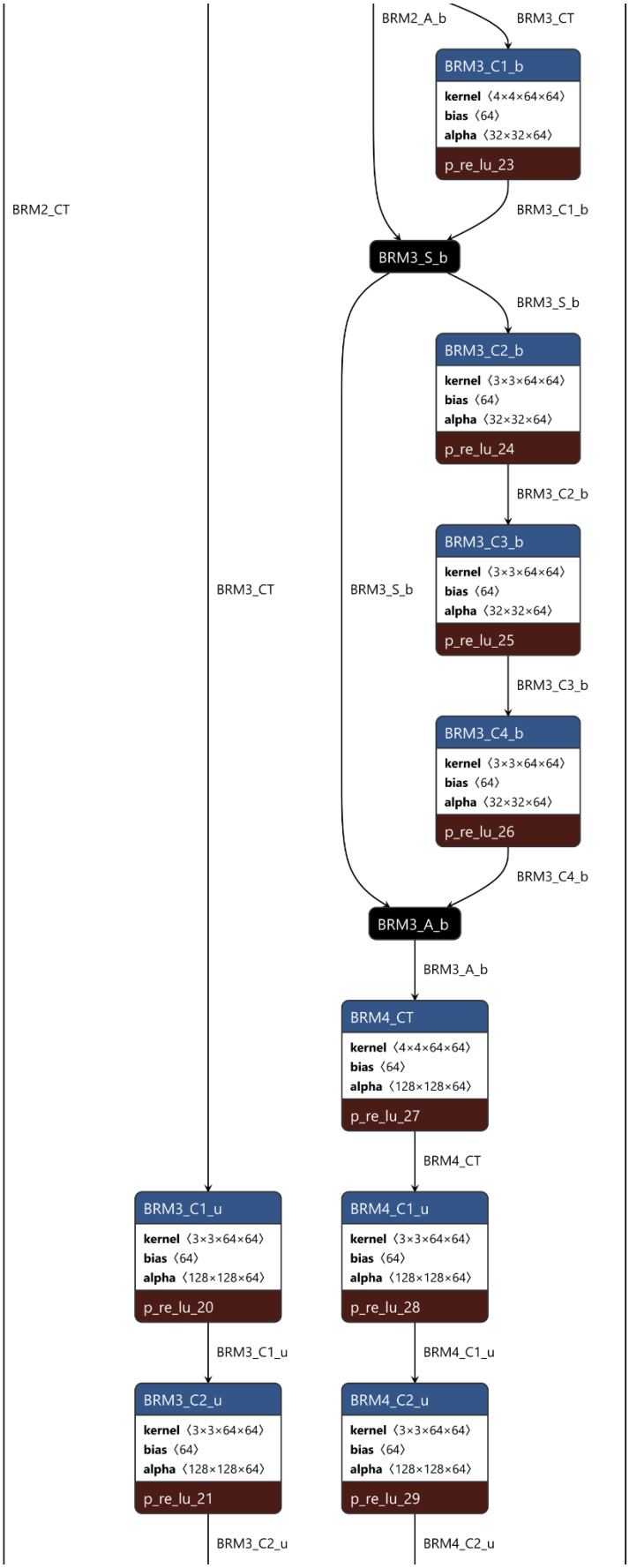
Zhong, J., Yang, B., Li, Y., Zhong, F., & Chen, Z. (2016). Image fusion and super-resolution with convolutional neural network. *Communications in Computer and Information Science*, *663*, 78–88. https://doi.org/10.1007/978-981-10-3005-5_7

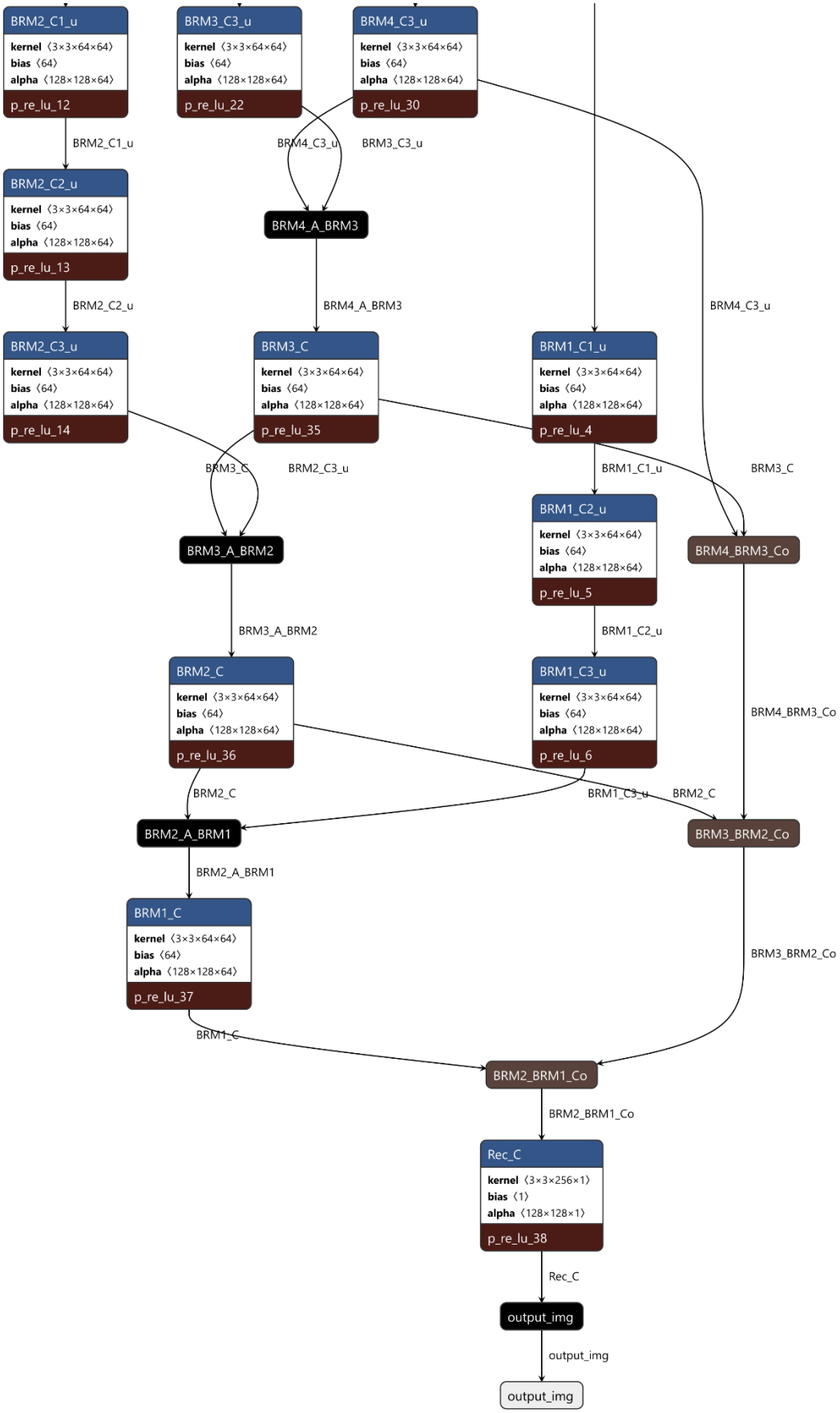
Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. In *Radiology* (Vol. 227, Issue 3, pp. 617–622). Radiological Society of North America . https://doi.org/10.1148/radiol.2273011499

ANNEX 1: EBRN MODEL DETAILED DESIGN

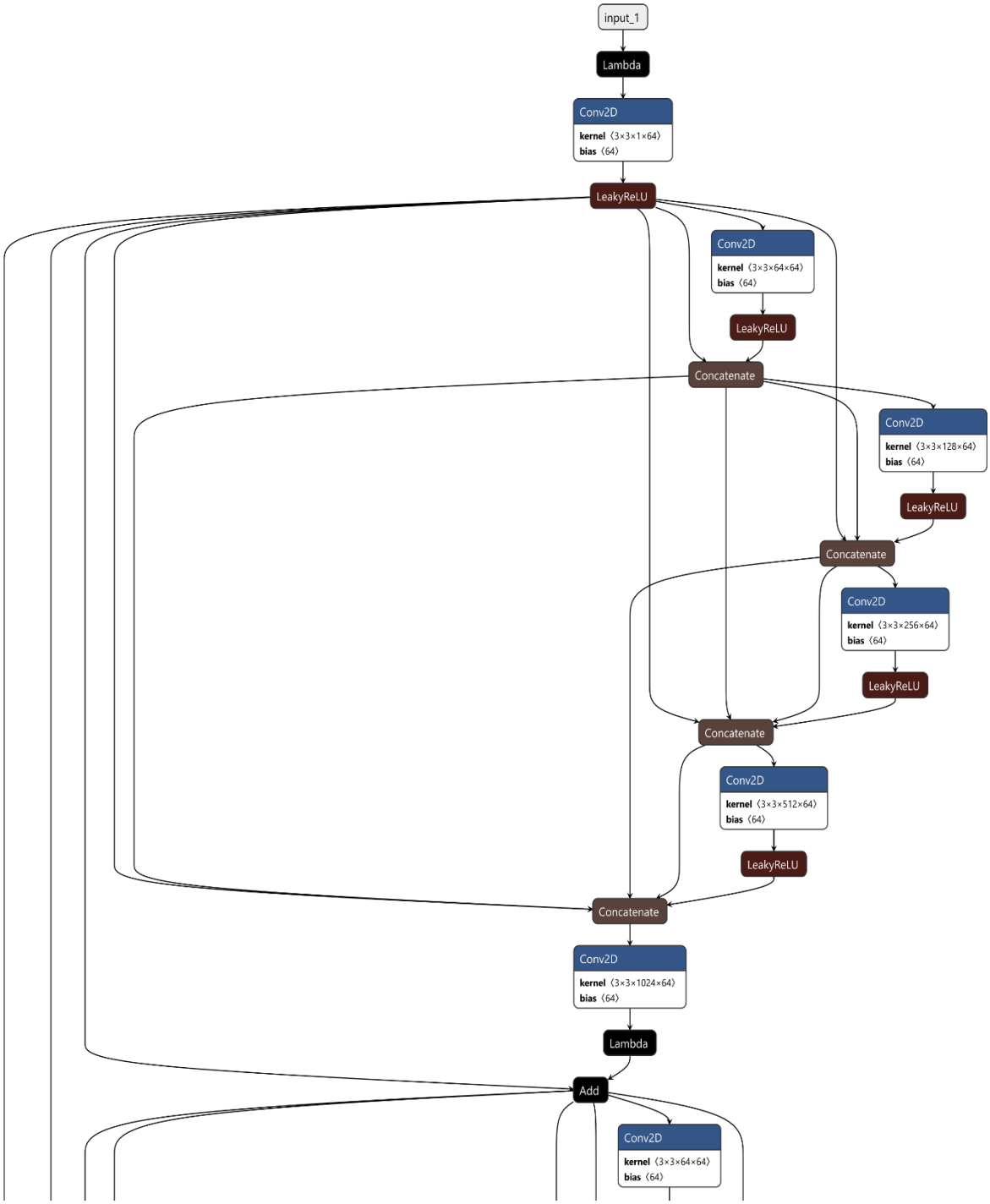


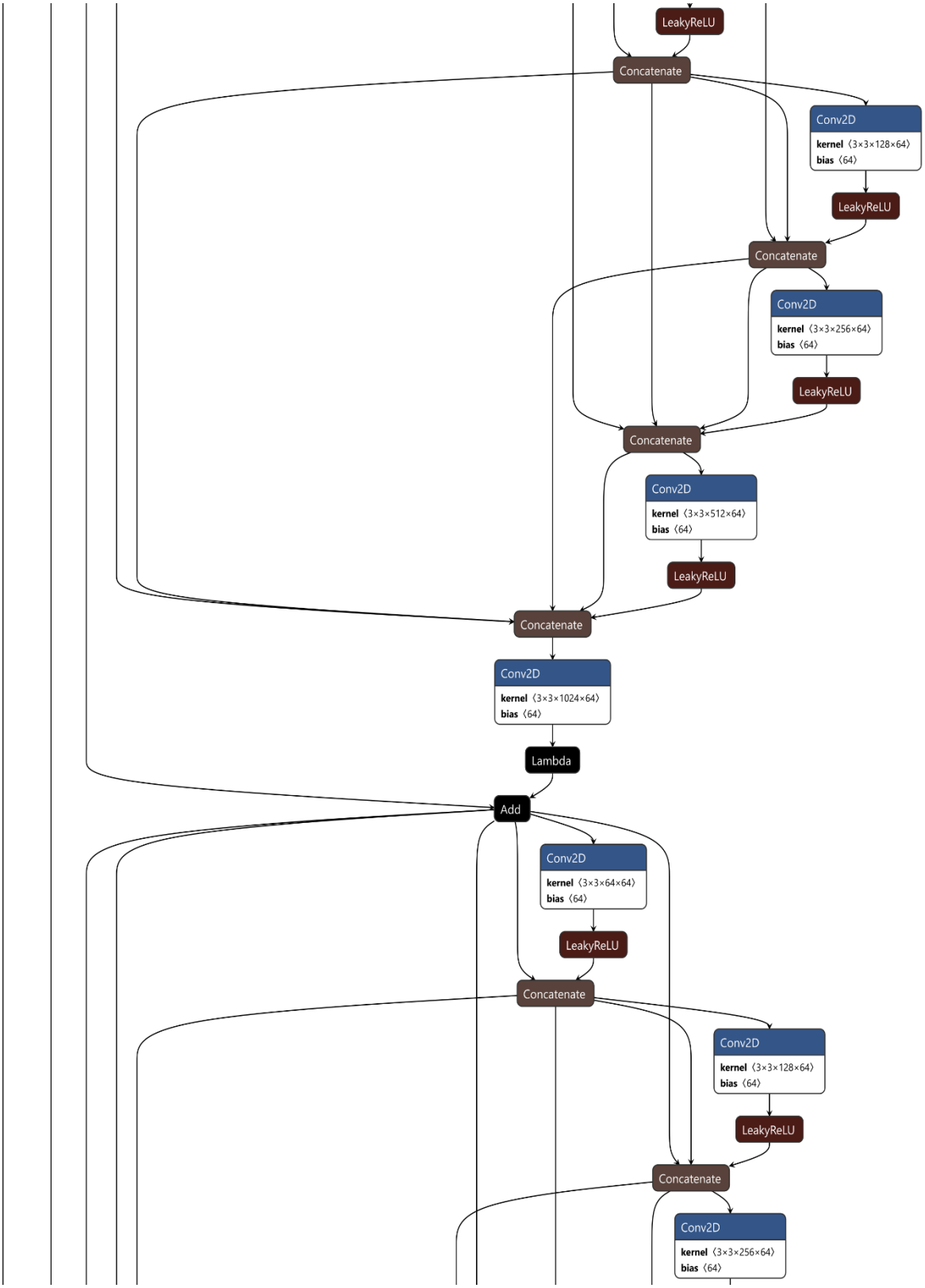

BRM1_C4_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_10
BRM1_C4_b
BRM1_A_b
BRM1_A_b
BRM2_CT
kernel 〈4×4×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_11
BRM1_A_b
BRM2_CT
BRM2_C1_b
kernel 〈4×4×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_15
BRM2_C1_b
BRM2_S_b
BRM2_S_b
BRM2_C2_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_16
BRM2_C2_b
BRM2_C3_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_17
BRM2_S_b
BRM2_C3_b
BRM2_C4_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_18
BRM2_C4_b
BRM2_A_b
BRM1_CT
BRM2_A_b
BRM3_CT
kernel 〈4×4×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_19

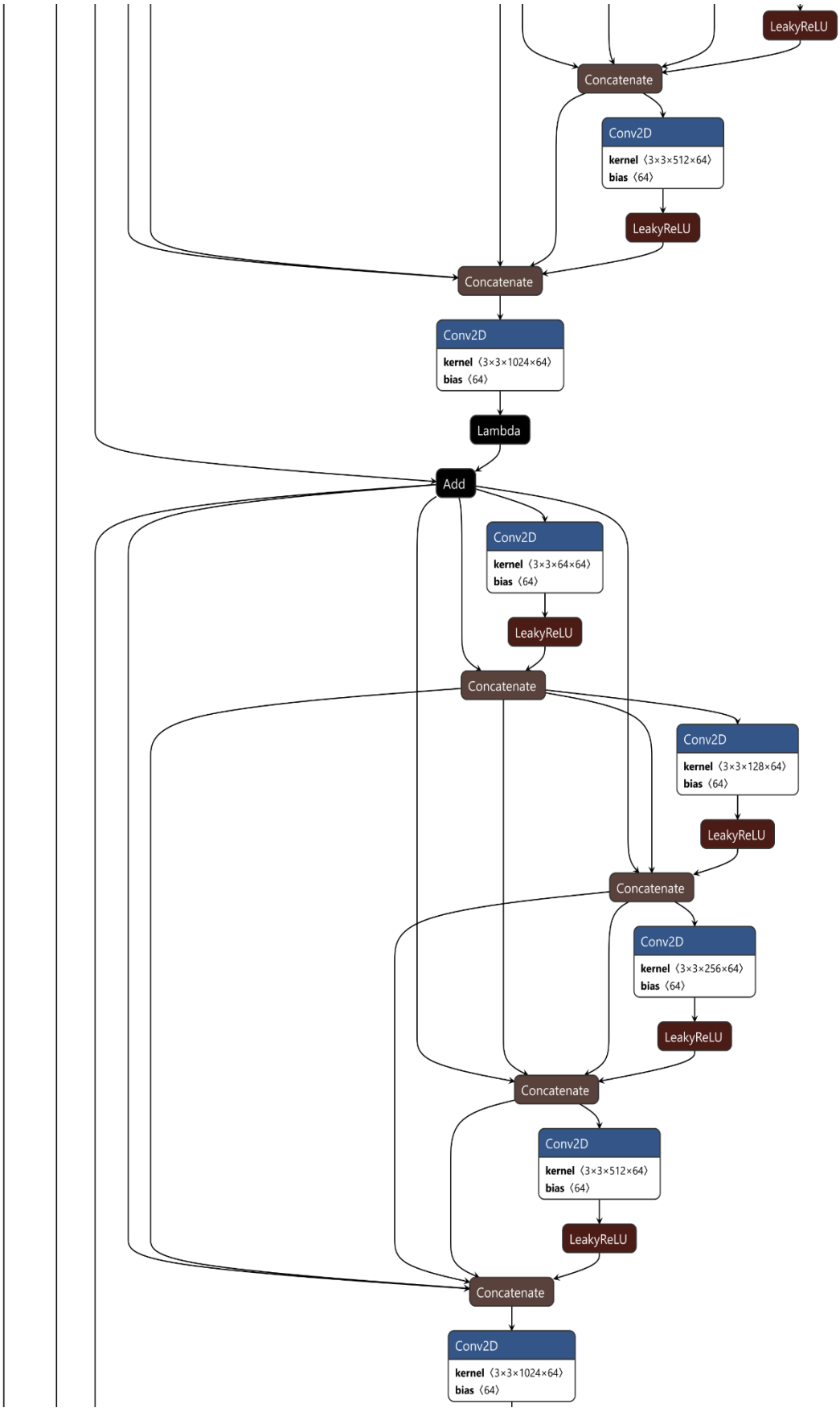

BRM2_A_b
BRM3_CT
BRM3_C1_b
kernel 〈4×4×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_23
BRM2_CT
BRM3_C1_b
BRM3_S_b
BRM3_S_b
BRM3_C2_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_24
BRM3_C2_b
BRM3_C3_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_25
BRM3_CT
BRM3_S_b
BRM3_C3_b
BRM3_C4_b
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈32×32×64〉
p_re_lu_26
BRM3_C4_b
BRM3_A_b
BRM3_A_b
BRM4_CT
kernel 〈4×4×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_27
BRM4_CT
BRM3_C1_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_20
BRM4_C1_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_28
BRM3_C1_u
BRM4_C1_u
BRM3_C2_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_21
BRM4_C2_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_29
BRM3_C2_u
BRM4_C2_u

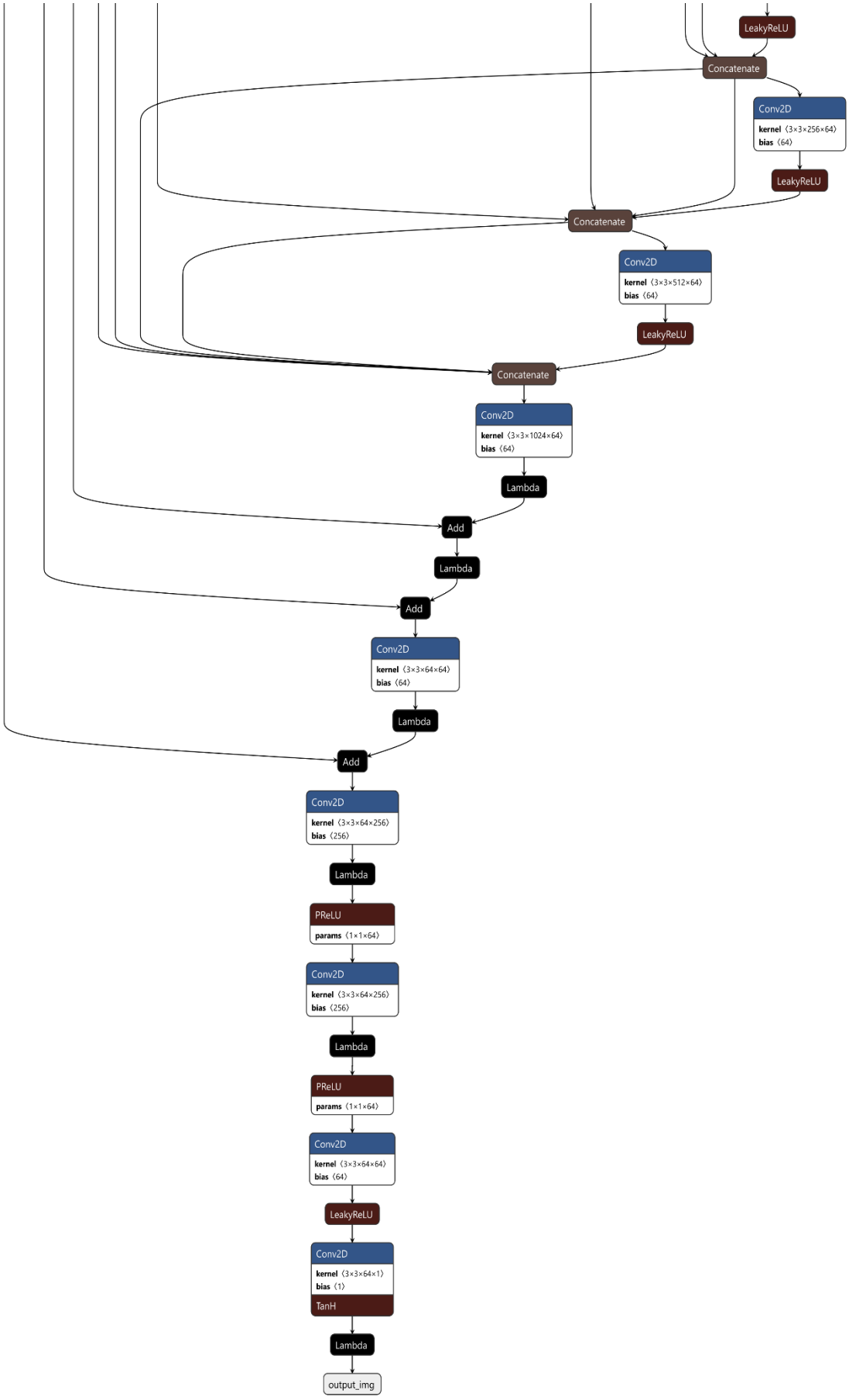

BRM2_C1_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_12
BRM3_C3_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_22
BRM4_C3_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_30
BRM2_C1_u
BRM4_C3_u
BRM3_C3_u
BRM2_C2_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_13
BRM4_A_BRM3
BRM2_C2_u
BRM4_A_BRM3
BRM4_C3_u
BRM2_C3_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_14
BRM3_C
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_35
BRM1_C1_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_4
BRM3_C
BRM2_C3_u
BRM1_C1_u
BRM3_C
BRM1_C2_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_5
BRM3_A_BRM2
BRM4_BRM3_Co
BRM3_A_BRM2
BRM1_C2_u
BRM2_C
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_36
BRM1_C3_u
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_6
BRM4_BRM3_Co
BRM2_C
BRM1_C3_u
BRM2_C
BRM2_A_BRM1
BRM3_BRM2_Co
BRM2_A_BRM1
BRM1_C
kernel 〈3×3×64×64〉
bias 〈64〉
alpha 〈128×128×64〉
p_re_lu_37
BRM3_BRM2_Co
BRM1_C
BRM2_BRM1_Co
BRM2_BRM1_Co
Rec_C
kernel 〈3×3×256×1〉
bias 〈1〉
alpha 〈128×128×1〉
p_re_lu_38
Rec_C
output_img
output_img
output_img

ANNEX 2: ESRGAN MODEL DETAILED DESIGN




LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×128×64⟩
bias ⟨64⟩
LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×256×64⟩
bias ⟨64⟩
LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×512×64⟩
bias ⟨64⟩
LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×1024×64⟩
bias ⟨64⟩
Lambda
Add
Conv2D
kernel ⟨3×3×64×64⟩
bias ⟨64⟩
LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×128×64⟩
bias ⟨64⟩
LeakyReLU
Concatenate
Conv2D
kernel ⟨3×3×256×64⟩
bias ⟨64⟩


LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×512×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×1024×64〉
bias 〈64〉
Lambda
Add
Conv2D
kernel 〈3×3×64×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×128×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×256×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×512×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×1024×64〉
bias 〈64〉


LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×256×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×512×64〉
bias 〈64〉
LeakyReLU
Concatenate
Conv2D
kernel 〈3×3×1024×64〉
bias 〈64〉
Lambda
Add
Lambda
Add
Conv2D
kernel 〈3×3×64×64〉
bias 〈64〉
Lambda
Add
Conv2D
kernel 〈3×3×64×256〉
bias 〈256〉
Lambda
PReLU
params 〈1×1×64〉
Conv2D
kernel 〈3×3×64×256〉
bias 〈256〉
Lambda
PReLU
params 〈1×1×64〉
Conv2D
kernel 〈3×3×64×64〉
bias 〈64〉
LeakyReLU
Conv2D
kernel 〈3×3×64×1〉
bias 〈1〉
TanH
Lambda
output_img