POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA CIVILE,
AMBIENTALE E TERRITORIALE

Discovering Data Augmentation Best Practices for Tropical Cyclone Intensity Estimation with Deep Learning

Tesi di Laurea Magistrale in
Environmental and Land Planning Engineering -
Ingegneria per l'Ambiente e il Territorio

Author: **Giulio Palcic**

Student ID: 969054
Advisor: Prof. Andrea Castelletti
Co-advisors: Dr. Guido Ascenso
Academic Year: 2021-2022

Abstract

Tropical cyclones (TCs) are among the most devastating extreme meteorological events, causing every year dramatic losses both in terms of life and economic damages. Studying TCs can help minimising their impacts on local economies. In this work, we contribute a methodology to estimate Tropical Cyclones intensity from brightness temperature images using Convolutional Neural Networks (CNNs). In particular, we provide new techniques to alleviate the intrinsic imabalance in existing datasets, which make it difficult for learning algorithms to generalize to rare classes. This issue is particularly relevant in TC studies, since high-intensity events are more rare than weaker ones. The dataset considered here is composed of satellite images from GridSat-B1 and observed intensities from IBTraCS, and it represents the largest dataset available in the literature. Following the transfer learning method, we test ResNet, DenseNet and EfficientNet on the original dataset, freezing the deeper convolutional layers and adapting the head of the network to the particular regression problem. We use a Gini-inspired coefficient to evaluate how much the distribution of samples differs from the uniform one and, therefore, to monitor the effects of the augmented dataset on the trained network performances. Clear trends in performance are detected according to different combinations of augmentation techniques employed to build the training dataset. We adapt random erasing and use it in combination with traditional augmentation approaches, such as horizontal and vertical flipping or random rotation, demonstrating it to be effective in preventing overfitting. Label distribution smoothing is also employed here to alleviate the imbalance between contiguous target labels, coupled with the Focal-R loss function to enhance the learning of less represented targets. Relevant improvements in performances are obtained in most intense classes on the test dataset, demonstrating to be a promising technique. The performance obtained on the different trained models show that the chosen data augmentation strategy has a relevant impact on the final mean absolute error, suggesting it can enhance the generalisability of the model. Lastly, a brief discussion of Grad-CAM heatmaps is introduced as a representation of features used by the CNN to predict new images. Three classes of features are identified, in which image borders, ring-like shapes or concentrated circular gradients are highlighted, respectively. These initial observations suggest that in the first

scenario, the neural network seeks information that crosses the image boundaries. In the second scenario, it appears that the network is looking for temperature values related to wind patterns located at a certain radius. In the third scenario, the network seems to focus on identifying and analyzing particular features within the image, such as the eye of a tropical cyclone or distinctive cloud patterns, and disregarding other portions of the image. This is accomplished by applying a focused gradient to the relevant area of the image.

Abstract in lingua italiana

I cicloni tropicali (TC) sono tra gli eventi meteorologici estremi più devastanti e causano ogni anno perdite drammatiche sia in termini di vite umane che di perdite economiche. Lo studio dei cicloni tropicali può contribuire a migliorare le capacità di previsione e, di conseguenza, a minimizzare il loro impatto sulle economie locali. Il presente studio affronta il problema della stima dell'intensità dei cicloni tropicali tramite l'interpretazione delle immagini satellitari con l'utilizzo del machine learning. In particolare, per svolgere tale compito vengono utilizzate Convolutional Neural Network (CNN), che ricevono come input immagini relative alla temperatura radiante e restituiscono una stima sulla velocità massima del vento. Il dataset utilizzato è costituito da un insieme di immagini prese da GridSat (database del NOAA), rappresentanti cicloni tropicali, di varia intensità e sparsi su tutto il globo, e dai target presi da IBTraCS. Questo dataset è il più vasto utilizzato fino ad ora in letteratura. I dati relativi a questo fenomeno climatico sono caratterizzati dal problema del data imbalance, avendo molti campioni relativi alle intensità meno elevate e una scarsa rappresentatività dei fenomeni più estremi. Lo scopo principale dello studio è quello di proporre nuove tecniche che siano in grado di alleviare questo problema in modo più efficiente rispetto a quanto fatto fino ad ora. Lápproccio scelto è quello del transfer learning, grazie al quale è possibile prendere delle reti neurali che sono già state allenate su dataset notevolmente più grandi e sono state ottimizzate per l'estrapolazione di elementi dall'immagine. Per adattare la rete neurale allo scopo desiderato è necessario ricostruire parte dell'architettura, concentrandosi soprattutto sui fully connected layer. Pertanto, definiti gli ultimi layer della rete, si effettua un training per calibrare i relativi parametri. Nel frattempo, i parametri degli strati più profondi rimarranno costanti. Per affrontare questo tema, come prima cosa viene definito un coefficiente ispirato a quello proposto da Gini, adattandolo a tale scopo. Tale coefficiente è una misura della diversità della distribuzione del dataset ottenuto rispetto alla distribuzione uniforme. Questo coefficiente si è rivelato particolarmente utile al monitoraggio dell'efficacia delle diverse tecniche di data augmentation, fornendo un'indicazione più puntuale rispetto al numero di input. Confrontando varie tecniche di data augmentation, viene proposto un adattamento del random erasing al problema, riproducendo la reale difficoltà della perdita di dati

all'interno delle immagini satellitari. Tale tecnica si è rivelata efficace, soprattutto sulle classi più intense e meno rappresentate, dimostrandosi una valida alternativa alle tecniche utilizzate in letteratura. Successivamente, viene proposto un cambiamento nell'approccio da seguire. Si testa in tal senso l'efficacia delle tecniche Label Distribution Smoothing accoppiate ad una loss function che sia funzione della frequenza dei target. L'LDS, tramite un filtro gaussiano, allevia le differenze intrinseche presenti nella numerosità dei target presenti per valore di velocità del vento, trasferendo l'informazione ai label limitrofi. La loss function testata è la Focal-R, funzione di un peso inversamente proporzionale alla frequenza assoluta del target all'interno del campione. In questo modo, viene pesato maggiormente lo scarto tra valore reale e la predizione di una classe meno rappresentata, inducendo la rete ad imparare maggiormente da questa per abbassare il valore globale della loss function. Tali tecniche si sono rivelate molto efficaci sulle classi più intense e meno rappresentate. Infine, vengono brevemente introdotte le heatmaps ispirate alla tipologia Grad-CAM come metodologia per la visualizzazione di feature importanti per la rete neurale nella stima dell'intensità. Vengono individuate tre classi principali di immagini. Nella prima, la rete pone peso maggiore in prossimità del bordo dell'immagine; ciò fa sospettare che in alcuni casi l'informazione sia tagliata fuori a valle della scelta della dimensione. La seconda classe presenta delle strutture ad anello; in tali casi la stima viene fatta in base a particolari informazioni contenute ad un raggio dal centro del ciclone più o meno costante. La terza classe presenta delle strutture circolari, con un gradiente delimitante una zona molto ristretta dell'immagine. In questo caso, quando la rete riesce a distinguere un elemento importante nell'immagine, come ad esempio l'occhio del ciclone, si concentra attorno a tale figura, tralasciando il resto.

Contents

Introduction

0.1. The context

Tropical cyclones (TCs), also known as *hurricanes* in the western North Atlantic (NA) and eastern North Pacific (NP) regions and *typhoons* in the west of NP (Emanuel, 2003), are some of the most devastating natural disasters in the world in terms of both damages and fatalities. Between 2000 and 2017, TCs caused losses equating to about 946 billion USD, one-third of the total damages caused by natural disasters in that period (Re, 2018). The risk associated with a TC (as for many other natural disasters) is a function of hazards, exposure, and vulnerability (WGII, 2014), the latter being mainly related in the literature to the TC's maximum wind speed, which is therefore taken as a measure of their intensity(Schmidt et al., 2010; Murnane and Elsner, 2012; Ye et al., 2020).

Recently, deep learning (DL) algorithms, especially convolutional neural networks (CNN), have been used to determine the intensity (i.e., wind speed) of TCs, a task referred to as Tropical Cyclone Intensity Estimation (TCIE). TCIE can be framed as either a classification problem, in which a TC is assigned to a pre-determined class of intensity (according to the Saffir-Simpson scale, for example), or as a regression problem, in which the exact intensity of the TC is estimated. Nowadays, such ML models give results comparable to (and sometimes better than) state-of-the-art physically based models used in meteorological centres. Nevertheless, improvements are expected in the following years, and promising developments have already been registered. In the coming years, the scientific community aspires to the development of tools that can provide highly accurate estimations of TC intensity in almost real-time. These tools will overcome the limitations of current non-ML techniques that rely on human supervision.

0.2. Objectives of the thesis

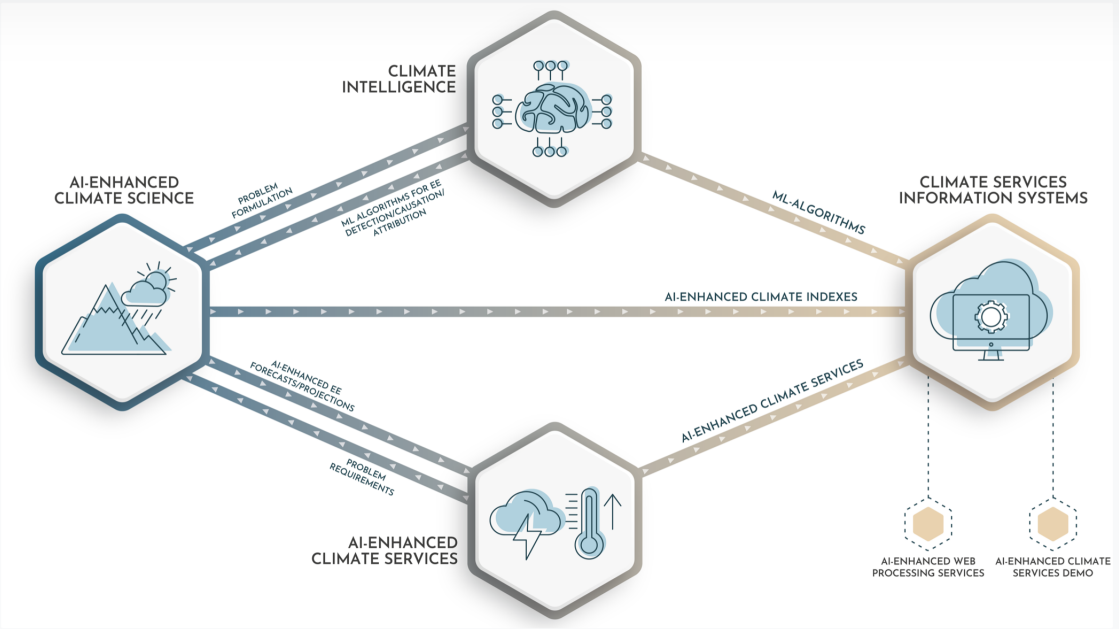
The availability of large amounts of quality data is a crucial bottleneck for the performance of deep learning algorithms. In the case of TCIE, very high-intensity events are much rarer than weaker ones, leading to a severe data imbalance. Thus, it is to be expected

that deep learning algorithms applied to this task are prone to perform well on weaker TCs (several examples of which would have been present in the training data) and poorly on the most intense (and rare) ones, which are, unfortunately, also the most dangerous. Two potential solutions to this data imbalance issue exist: obtaining more images or generating new, synthetic ones. For TCIE, the first solution would be ineffective, as the data imbalance is not due to an improper sampling of the events in climate datasets, but rather it is intrinsic to the very nature of the phenomenon; in other words, if more data were gathered (by taking images from multiple satellites, for example), the distribution of the intensity of the TCs would remain highly imbalanced. On the other hand, data augmentation could alleviate the imbalance by generating new synthetic images with a frequency that depends on the rarity of an event; in other words, more synthetic images could be developed for extreme, rare events than for weak, common ones, thus alleviating the imbalance. Data augmentation is crucial to many deep learning models (especially in Computer Vision). In TCIE, however, its use has been sparse and not guided by first principles but rather by empiricism and trial and error. Indeed, though some authors did use some form of data augmentation within their deep learning pipelines for TCIE, no guidelines exist in the literature regarding which methods of data augmentation work best *specifically for this task*, nor how the amount of augmentation done affects the ultimate performance of a given model. The dataset, composed of GridSat-B1 IR channel at 11 μm as input and IBTraCS information on Maximum Sustained Winds as a target, represents the largest one available in the literature. Taking advantage of transfer learning, State of The Art (SoTA) deep-CNNs are tested to define the backbone architecture more suitable for the purpose. Among the ones tested, EfficientNet-B0 was selected since it provided the best performance in validation. Therefore, this thesis aims to investigate the effectiveness of different data augmentation methods to address the data imbalance problem of TCIE. In particular, other than testing common data augmentation methods such as horizontal/vertical flipping and rotation, this thesis proposes three novelties:

- The use of the *Adapted Gini coefficient* as a metric to be used as a measure to find the best distribution of the augmented data;
- *Random erasing* as a novel data augmentation method for TCIE to specifically target data quality issues that arise in this field (i.e., random missing data in satellite images);
- *Label Distribution Smoothing (LDS)* and *Focal Loss (FL)* as additional data augmentation methods;

As climatologists often report concerns about the lack of interpretation of deep learning

Figure 1: Sketch of the integrated approach followed by the CLINT project.



models, following the intuition of Lee et al. (2019), the analyses performed in this thesis will be visualised through *heatmaps*, which will give insights into what parts of the images the machine learning models focus on at any given time.

0.3. The CLINT project

This thesis is developed as part of the Climate Intelligence (CLINT) project. CLINT is an EU H2020 Programme founded project. It aims to develop an Artificial Intelligence framework to study Extreme Events (EEs) such as tropical cyclones, heatwaves and warm nights, droughts, and floods, to improve the detection, causation, and attribution of EEs, but also to quantify their impacts in relation to climate conditions, considering future projections. The project's final objective is to develop AI-enhanced Climate Services to obtain operative Web Processing Services.
CLINT is organised into four macro components (1):

- *Climate Intelligence*, related to the development of new Machine Learning algorithms;
- *AI-enhanced Climate Science*, to enhance the performance of algorithms for the detection, causation, and attribution of EEs;
- *AI-enhanced Climate Services*, to develop, at continental and local scales, the water, energy, and food (WEF) Nexus;

- *Climate Services Information Systems*, to develop web processing services and commercial demos;

This paragraph was broadly taken from CLINT website [1]

0.4. Outline of the thesis

The thesis is organised as follows: Chapter 2 provides an introduction to the physics of TCs physics and their relation with climate and climate change; Chapter 3 reviews the state of the art of techniques for intensity estimation and the problems of data imbalance and data augmentation; Chapter 4 details the experiments carried out during this thesis, and Chapter 5 reports the results; and finally, Chapter 6 discusses the results and potential future research directions.

[1] https://climateintelligence.eu/

1 | Tropical Cyclones

1.1. An Introduction to Tropical Cyclones

Tropical cyclones (TCs for short) are extreme meteorological events occurring in most tropical ocean basins, threatening populations living on the coasts by wreaking havoc on infrastructures, shipping, and other offshore activities (Seneviratne et al., 2012).

A tropical cyclone is defined by (Gray, 1975) as a *"closed circulation system with a warm core of at least 5° diameter, extending through most of the troposphere with a little vertical slope and whose relative vorticity at lower levels and inner 100-200 km radius is greater than* $100x10^{-6}\,s^{-1}$*, which requires a horizontal wind speed velocity at* $100-200\,km$ *radius between* $5-10\,m/s$. Tropical cyclones are fuelled mainly by heat exchange between tropical oceans and the atmosphere.

Much work has been done trying to model these catastrophic phenomena, but still many problems persist and full understanding is missing. Predictive skills are limited, especially in terms of TC genesis, intensity, and risk forecasts (Chen et al., 2020).

The classification of TCs commonly relies upon the time-averaged maximum wind speed recorded at a 10 m height. The averaging time interval is usually 10 min. Some exceptions can be found, such as the U.S. convention which uses a time interval of 1 min. One of the most used classifications follows below (Emanuel, 2003):

1. *Tropical depression*, with maximum wind speed lower or equal to 17 $m\ s^{-1}$;
2. *Tropical storms*, when their maximum wind speed is in between 18 $m\ s^{-1}$ and 32 $m\ s^{-1}$;
3. *Severe tropical cyclones* when their maximum wind speed is above 33 $m\ s^{-1}$;

Another common classification of TCs is the so called *Saffir-Simpson Hurricane Wind Scale* (Saffir, 1973; Simpson, 1974), which divides TCs into five categories based on maximum sustained wind speed:

Table 1.1: Saffir-Simpson Hurricane Wind Scale (from NOAA site).

Category	Sustained Winds	Damages
1	64-82 kt (112-153 km/h)	Very dangerous winds will produce some damage: Well-constructed frame homes could have damage to roof, shingles, vinyl siding and gutters. Large branches of trees will snap and shallowly rooted trees may be toppled. Extensive damage to power lines and poles likely will result in power outages that could last a few to several days.
2	83-95 kt (154-177 km/h)	Extremely dangerous winds will cause extensive damage: Well-constructed frame homes could sustain major roof and siding damage. Many shallowly rooted trees will be snapped or uprooted and block numerous roads. Near-total power loss is expected with outages that could last from several days to weeks.

Continued on next page

Table 1.1 – continued from previous page

Category	Sustained Winds	Damages
3 (Major)	96-112 kt (178-208 km/h)	Devastating damage will occur: Well-built framed homes may incur major damage or removal of roof decking and gable ends. Many trees will be snapped or uprooted, blocking numerous roads. Electricity and water will be unavailable for several days to weeks after the storm passes.
4 (Major)	113-136 kt (209-251 km/h)	Catastrophic damage will occur: Well-built framed homes can sustain severe damage with loss of most of the roof structure and/or some exterior walls. Most trees will be snapped or uprooted and power poles downed. Fallen trees and power poles will isolate residential areas. Power outages will last weeks to possibly months. Most of the area will be uninhabitable for weeks or months.

Continued on next page

Table 1.1 – continued from previous page

Category	Sustained Winds	Damages
5 (Major)	137 kt or higher (252 km/h or higher)	Catastrophic damage will occur: A high percentage of framed homes will be destroyed, with total roof failure and wall collapse. Fallen trees and power poles will isolate residential areas. Power outages will last for weeks to possibly months. Most of the area will be uninhabitable for weeks or months.

Globally, approximately 80 TC develop each year, most of which form during summer and early autumn. Some difference in timing can be detected among different ocean basins; thus, the respective TC seasons can be summed up as follows (Emanuel, 2003):

- In the North Atlantic (NA): June-November (peak in September);
- In the Eastern North Pacific (ENP): peak activity in July and August;
- In the Western North Pacific (WNP): peak activity in late Summer, but their activity has been detected year round;
- In South Pacific (SP) and South Indian (SI) Oceans: October-June;

1.2. Basics of Tropical Cyclone physics

The cyclonic flow (i.e., a clockwise (in the Southern Hemisphere) or counterclockwise (in the Northern Hemisphere) circulation of wind) of a TC does not involve its entire volume, and is actually inverted near the top of the storm. Moving radially from the centre, wind speed first increases rapidly, reaching a maximum at 10-100 km from the center; it then decreases with a $r^{-1/2}$ law before a sudden decrease at further radii. At 100-1000 km from the center, wind speed is indistinguishable from that of the environment (Emanuel, 2003).

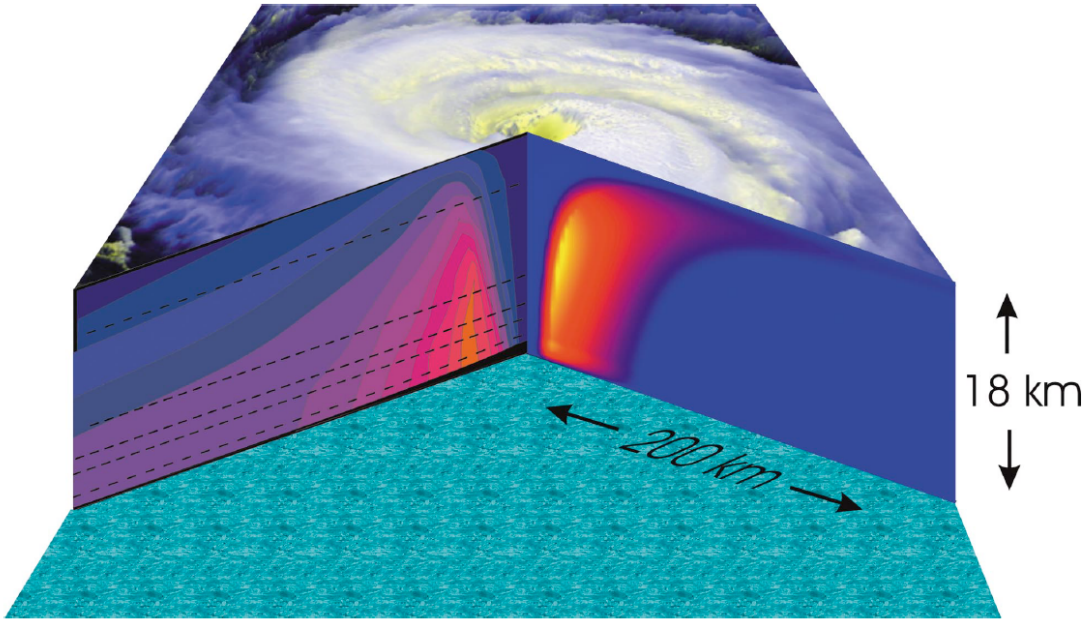


Figure 1.1: Cutaway of the structure of a tropical cyclone (from Emanuel (2003)). On the right, the results from numerical simulation of vertical component of the velocity. On the left, the tangential component measured by aircraft reconnaissance.

The structure of TCs is so that maximum wind speeds occur near the surface, decays slowly along the height of the TC, and reverses its direction at the top. The vertical velocity (5-10 ms^{-1}) has its maximum in the *eyewall*, a ring of deep convection between the outer limit of the eye and 20-50 km in external radial direction (Emanuel, 2003). The maximum value is usually located in the upper part of the troposphere (Margaret A. Jorgensen, 1985).

The *eye* of a TC is an almost cloud-free region which is usually observable in well-developed, stronger storms. Inside the eye, air is slowly subsiding, whereas outside it the slow, descending motion can be interrupted and inverted by a spiral band of cumulonimbus clouds (Emanuel, 2003).

TC physics are characterized by the conservation of the angular momentum (M) per unit mass around the axis of the storm.

$$M = rV + \frac{1}{2}fr^2 \tag{1.1}$$

The first term is proportional to a component rV (radius of the storm times tangential velocity) and $1/2\, fr^2$ which represents the contribution of the rotation of Earth to the angular momentum. The angular momentum decreases upward and inward, having the greatest variation in the eyewall. The specific entropy is a variable dependent on temperature, pressure and water concentration. It increases proceeding radially to the center of the cyclone and reaching a maximum inside the eyewall. Looking at the vertical trend and proceeding from near the surface to the top layers, this variable experiences a decrease and then an increase.

Also the specific entropy, function of temperature, pressure and water concentration is conserved (Emanuel, 2003).

1.3. Basics of Tropical Cyclones Energetics

In Emanuel (1986), mature TCs are compared to an idealized Carnot engine, in which air parcels follow a path that can be summed up into isothermal expansion, adiabatic expansion, isothermal compression, and adiabatic compression.An air parcel approaching the center of the cyclone, is subjected to a pressure drop and an increase in entropy, due to both enthalpy transfer from the sea surface and dissipation of kinetic energy (Bister and Emanuel, 1998). The most relevant sink of energy is given by the angular momentum decrease provoked by frictional torque with the sea surface. Once the eyewall is reached, the movement becomes ascending and is nearly adiabatic and without frictional torque, moving from higher to lower pressure levels. At this point the movement becomes nearly horizontal and ideally continues without closing the cycle (Emanuel, 1986).

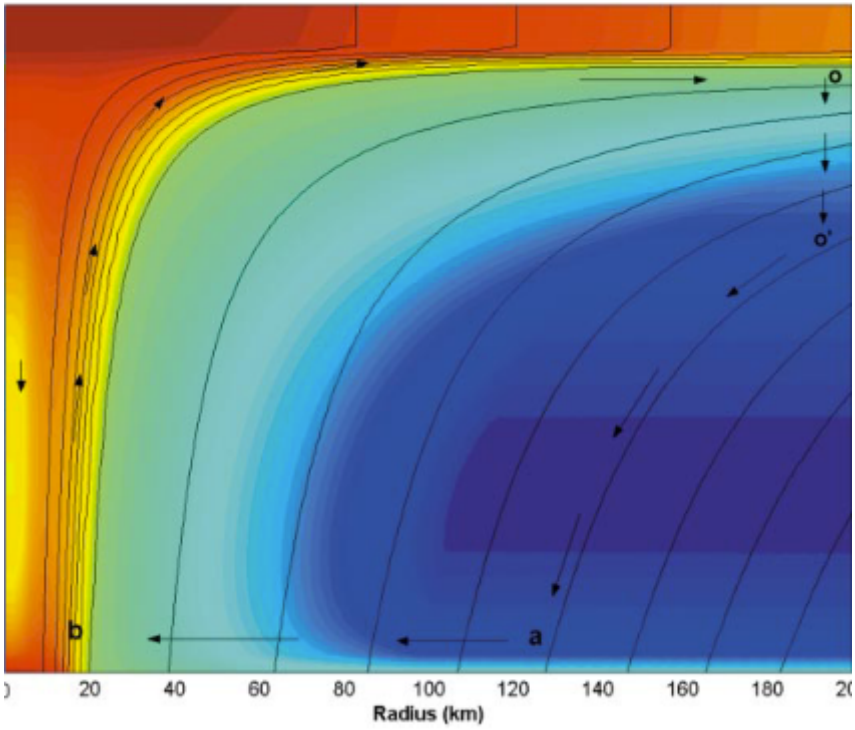


Figure 1.2: Results of a numerical models of a Tropical Cyclone seen as a Carnot engine (from Emanuel (2003). Shading of colours represents the measure of specific content of air entropy (the warmer the colour, the higher the entropy content. Black curves denote contours of constant angular momentum per unit mass (increasing with radius). Arrows qualitatively quantify air motion.

Emanuel (1986) also argues that even if in numerical modelling it is necessary to force the Carnot engine to be closed, it appears to well represent real phenomena. Thus, it is ideally closed at a point where air starts its descending movement of entropy capture thanks to electromagnetic radiation, and, because of mixing with the surroundings, angular momentum is gained. This portion of the cycle is nearly isothermal. Then, mixing with moist and dry air, the air parcel looses entropy in a path which closes the cycle and in which the angular momentum is conserved. The work produced is then transformed back into heat through turbulent dissipation in the boundary layer.

Emanuel (2003) derives in his paper an approximated formula which links the square of maximum wind speed with the deviation of the sea surface temperature and a "cold source" in the environment surrounding the TC, and the difference of enthalpy of air close to the surface and the air in contact with the ocean(assumed to be saturated with water vapour characterised by a temperature equal to the temperature of the ocean). He speculates that the energy transfer occurs because of this enthalpy difference which

endorses convection. The greenhouse effect hinders net long-wave radiative flux from the sea, thus since the squared velocity of the wind speed is directly proportional to the sea surface temperature, there is concern that climate change could provoke an increase of TC intensity (Emanuel, 2003).

1.4. Tropical cyclones genesis

Historically, modelers were focused on reproducing the growth of an already existing tropical cyclone rather than the study of its genesis, highlighting the role (among other variables) of frictional convergence, condensation heating, and ocean energy exchange. Tropical cyclone intensification is indeed quite well understood: after the formation of cyclones, lateral energy dissipation is undermined and energy builds up in the inner region. Momentum, energy, and vapour are attracted to the central part, thanks to frictionally forced convergence processes. The conditions favourable for TC genesis are not widespread throughout the whole globe (Gray, 1975, 1998).

According to historical observations, we can narrow the search for the drivers of TC genesis to within the following limits (Gray, 1975):

1. TCs form far 4-5° from the equator and are more probable in the belts of 5-15° latitude. Moreover, they are limited to below 22° and 35° latitude in the Southern Hemisphere and Northern Hemisphere respectively;
2. The majority of all TCs are found in the Eastern Hemisphere and are especially favoured for genesis in the regions centered at 90°E, 140°E and 105°W;
3. The majority of TCs are generated in summer, even if in the WNP they can occur throughout the year;
4. In the North Indian Ocean (between 5° and 15°) there are two main seasons favourable for TCs formation;
5. TCs do not form in Southeast Pacific and South Atlantic oceans;
6. In the Inter-Tropical Convergence Zone (ITCZ), TCs formation is especially likely;

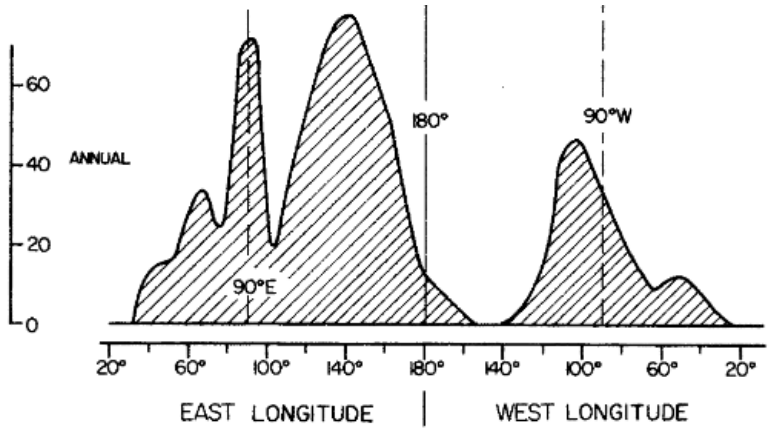


Figure 1.3: Distribution of TCs' origin by longitude. Image taken from Gray (1975).

Interestingly, about 80-85% of TCs originate near the ITCZ, (Gray, 1975), meanwhile almost all the remaining ones form in association with the trade winds and an upper atmospheric trough to their northwest (Sadler, 1967, 1976).

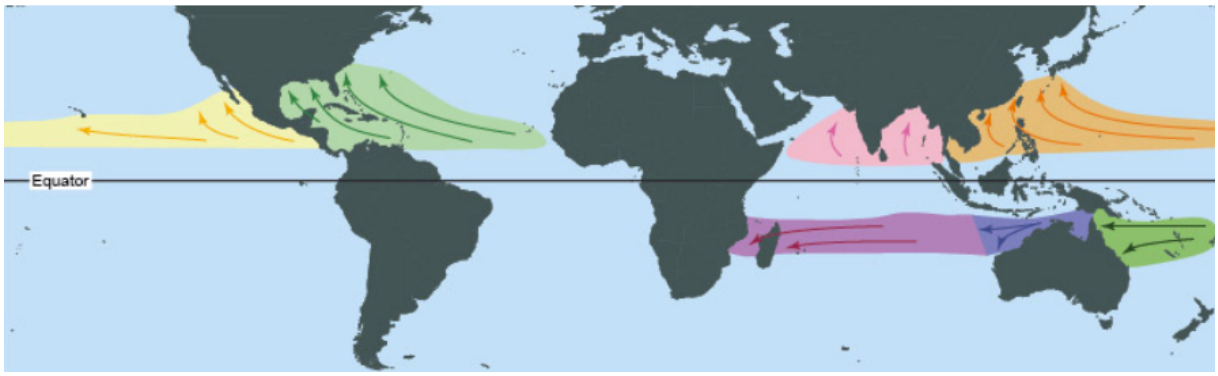


Figure 1.4: TCs' formation ocean basins. Image taken from NOAA site.

This wide share of TC percentage is caused by weak tropospheric vertical wind shear and the close similarity of 200-500 mb wind and cloud cluster velocities.Gray (1975, 1979) Moreover, favourable conditions are met if in the surroundings anticyclonic vertical wind shear assumes large values (Gray, 1968; McBride, 1981; McBride and Zehr, 1981). In addition, the large-scale low-level tropospheric winds at a certain distance (300 to 600 km) often have larger values in developing systems (Gray, 1968; McBride, 1981). This is because this is favourable to spin up of the outer growing radius tangential winds (Gray, 1998).

The genesis of TCs starts from the accumulation of enthalpy in the upper troposphere

(500-300 mb) over a small area of 200-300 km, producing a thickness increase (the top of the system reaches 100 mb) and a small surface decrease. It is clear that in order to understand cyclone genesis it is crucial to understand the reasons for this enthalpy accumulation.

As the cumulus convection proceeds, evaporating the water, the latent heat absorption induces a cooling effect in the environment. This cloud re-evaporation is especially large in the lower troposphere (Lopez, 1973; Gray, 1973; Yanai et al., 1973). The latent heat is then exported as potential energy to the surrounding environment. Sinking motions arise from the cumulus: part in the surrounding environment and removed from it, and part in the surrounding clear regions or in the distant anticyclones where the adiabatic sinking balances the atmospheric radiational losses. Furthermore, for the genesis of a TC it is fundamental that the compensation for the upward movement of the cumulonimbus (Cb) occurs in the local environment and that a sufficient amount of enthalpy is not dissipated horizontally by ventilation or divergence. It is indeed necessary that the downward return flow of the Cb induces a compressional sinking motion able to heat up the upper troposphere (Gray, 1975).

On one hand, in the cloud clusters the gradient of temperature has a linear correlation with the pressure, which causes equal amounts of enthalpy at any pressure level for equal amounts of mass subsidence. On the other, a non-linear relation links specific humidity and pressure, which means that the lower troposphere is drier than the upper layers for same magnitudes of mass subsidence (Gray, 1975).

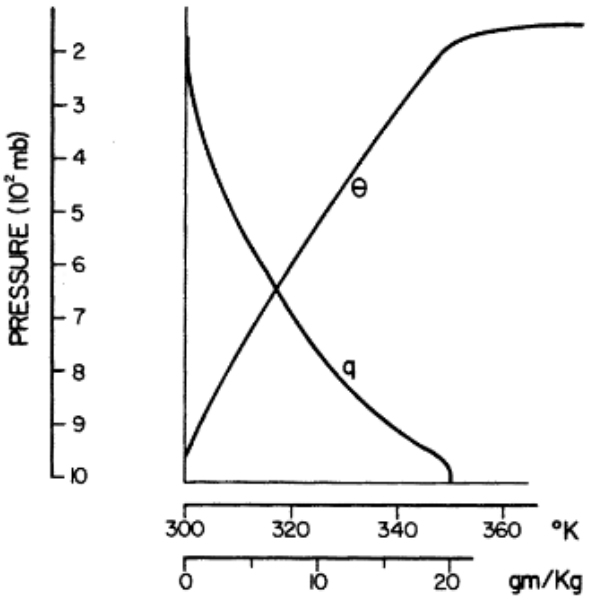


Figure 1.5: Potential temperature (θ) and specific humidity (q) in tropical cloud clusters. Image taken from Gray (1975).

There is a vertical variation in net tropospheric warming due to sinking-drying promoted vapor loss: in the lower troposphere evaporational cooling is dominant, while in the upper troposphere the sinking-drying is larger than the required evaporational cooling. As a consequence, middle and upper layer of the troposphere are the most favourable portions in which Cb induced enthalpy increase can occur (Gray, 1975).

Cloud clusters are organised in an extensive layered cloud structure. This kind of structure is responsible for another enthalpy gain, since the radiational loss is reduced. In cloud clusters, a significant portion of the upward-moving mass sinks locally, causing an up-moist and down-dry recycling circulation. This recirculation pattern is more dominant in the lower troposphere than in other layers. The sinking-drying motion promotes vapor replacement from direct transport from the cumulus (Gray, 1973).

If the accumulation of energy in the upper level of the atmosphere is inhibited by divergence or advection and ventilation, TC formation is hampered. Indeed, Gray (1975) finds that advection and ventilation are fundamental in understanding whether TCs can complete the formation process

Gray devoted part of his studies to the research of parameters correlated with seasonal

tropical cyclone frequency. This statistic is correlated to the combination of six parameters:

1. *Coriolis parameter (f)*;
2. *Low-level relative vorticity (ζ_r)*;
3. *Inverse of the tropospheric vertical wind shear ($1/S_z$)*;
4. *Ocean thermal energy (E)* ;
5. *Difference in equivalent potential temperature between the surface and 500 mb ($\Delta\theta_e$)*;
6. *Relative humidity (RH)* in the mid-troposphere;

Multiplying these parameters between each other, the seasonal frequency of TC formation (Seasonal Genesis Frequency, SGF) at any location can be estimated.

$$SGF \propto \zeta_r \cdot f \cdot \frac{1}{S_z} \cdot E \cdot \Delta\theta_e \cdot RH$$

The equation can be divided into two sub-products:a *dynamic potential*, given by the product of the first three parameters, and a *thermal potential*, given by the product of the remaining parameters listed above.

However, as highlighted by Gray, the information conveyed by his index is only an estimate of the long-term frequency of occurrence, with no further information to be used in TC forecasting which has day-to-day requirements in the formation prediction. These parameters are neither varying on a daily basis, nor assume values that allow us to distinguish between developing and non developing phenomena (Gray, 1975).

In his research, Gray noticed that *environmentally induced asymmetric lower tropospheric wind surge* plays a crucial role and that concentrated mesoscale deep convection must be present. In particular TC formation is more probable if low-level wind surges penetrate a cloud cluster or a tropical disturbance, where conditions of concentrated high vorticity are present (Gray, 1975).

TC formation is a complex process in which many factors contribute simultaneously. Even when the climatological requirements are met, some tropical disturbances evolve into TCs, but most do not. Indeed, even if favourable conditions are present, a TC may still fail to develop. Generally, three basic requirements *must* be present to make TC formation at

least a possibility:

1. Favourable climatology (in terms of region, season, SST,...);
2. Correct Synoptic Flow pattern (monsoon trough or high vorticity with small vertical wind shear, etc);
3. Active Mesoscale Convection System (MCS) within a cloud cluster system;

As previously said, these conditions are necessary but not sufficient, meaning some parameters are still missing from the equation. For example, the presence or lack of concentrated wind convergence at the center of a tropical disturbance plays a key role. Further, environmentally driven asymmetrical wind surges must occur to trigger intense convective outbreaks at locations where MCS induced previously high relative vorticity. These conditions may evolve in raising Convective Vortices (CVs) which concentrate vorticity, and if a second wind surges occur, it favours the wind spin-up.

Another key factor is the presence of bursts of low-level wind convergence in areas of previously developed relative vorticity, such as areas characterised by CVs. If this occurs, a rapid increase in vorticity is allowed; this is possible only if a strong Externally Forced Convergence (EFC) process is activated (Gray, 1998). Moreover, another important feature is related to the movement of a disturbance into a quasi-stationary wind field in a way that is able to bring a rapid increase in the convergence near the center of the disturbance (Gray, 1998). The tropical disturbance then shall intensify. This step is more efficient if the externally forced mass convergence is concentrated over a small central area of the disturbance. The rapid wind spin up can then spread outward (Gray, 1998).

Once a strong enough second surge breaks out, an upward movement arises that drives air parcels near saturation conditions. If this happens, these near-saturation conditions may cause the suppression of strong downdrafts and the influences of entrainment. This causes the establishment of Extreme Convection (EC) areas, inducing a sustained updraft motion and then initiating the Internally Forced Convergence (IFC) process. As a result, a persistent pressure drop of 5-10 mb per day begins and a rapid wind spin-up in the inner region begins. This phase of TC formation is called by Gray 'take-off' or 'ignition point', and it can occur only near the center of the disturbance, where the cumulus cloud updraft buoyancy can be sustained (Gray, 1998). Such an EC outbreak in a small region near a CV is less probable and is often missing in non-developing systems.

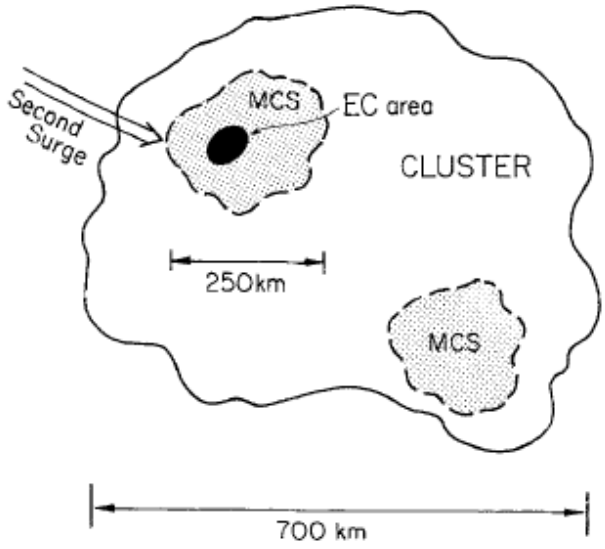


Figure 1.6: EC area formed by concentrated convection within an MCS by additional convergence caused by wind surge. Image taken from Gray (1998).

Once initiated, the IFC becomes a self-sustained process. The surface pressure drop continues over a small area beneath the cumulonimbus clouds, causing an increase in the horizontal acceleration of low-level winds. As a result, an intense upward motion is established (Gray, 1998). The combination of the vertical profiles of cloudiness, slope of pressure surfaces with temperature anomalies, and divergence profiles in MCS cause maximum convergence in the middle level of the troposphere. This motion is enhanced in later stages, promoted by evaporating and cooling downdrafts and frozen cloud melting. After 6-12 hours, MCSs weaken and die.
However, middle-level cyclonic circulations persist and can expand their influence downward to the surface, although weaker and of a smaller size. As a consequence, it is possible to notice a cold-core lower-level circulation and upper-level warm core. Since a residual MCS convective burst is present, the effect of the second EFC wind surge is different and is able to enhance with more efficacy low-level wind spin-up. Finally, this second wind surge is able to transform the disturbance's low levels from a cold-core into a warm-core system. Once this system is formed, the cyclone starts to organise into an eye-wall type convection system and the surrounding environment starts loosing his influence (Gray, 1998).

In their work, Chen and Frank (1992) highlighted how IFC initiation requires that the relative humidity be close to saturation condition (RH>95%), as in these conditions downdraft development and updraft weakening are inhibited (Chen and Frank, 1992). When

penetrating the small inner-core area, the second wind surge seems to be able to raise the water vapour content to near saturation even if a large percentage of the converging water vapour condenses and rains out. Although some downward motion can occur due to the water loading and ice melting, a general updraft buoyancy can be maintained.

The second wind surge plays a role also in mechanically forcing convergence and upward vertical motion in the EC region to the condition of near saturation. From the increase of the humidity of the upper levels, the initiation of multi-cloud sustained updraft buoyancy can take place until the start of the IFC (or CISK) intensification process (Gray, 1998). As mentioned earlier, TC formation and subsequent intensification are favoured by minimum values of vertical wind shear over the center of the convective system. Moreover, anticyclonic flow in the upper troposphere at 300-600 km radius is another enhancing factor for TC genesis.

Other studies (Dvorak, 1975) have brought evidence to how unidirectional upper-tropospheric flow over a disturbance exports negative tangential momentum, depleting the spin up of the tangential winds of the outer radius of the TC. Also, the Tropical Upper-Tropospheric Trough (TUTT), even if not being able to initiate the formation, can contribute to the development of TCs. TUTTs play a double role: they can create upper-level anticyclonic circulation, and they can also reduce the vertical tropospheric wind shear near the center of the disturbance (Gray, 1998). Gray (1998) states that equivalent potential temperature is a variable which cannot help distinguish between non-developing and developing disturbances, and that mechanical factors are indeed much more relevant than thermodynamic ones.

In conclusion, TC formation can be understood as a two-stage process separated by an intermediate 1-3 day quiescent period. Most forming storms cannot develop into TCs because they dissipate between the first and the second stage.

In the second, stage a small area of EC must arise within a MCS. Thanks to the EC area, humidity rises to near saturation conditions, weakening the downdraft and inhibiting the entrainment cooling of updrafts. The IFC is promoted by intense and concentrated cumulonimbus convection, which is caused mainly by EFC. The rapid growth of cyclonic wind is allowed only if concentrated within a small central region. Only in a second moment the area of high winds can spread outward in surrounding areas. The inner core is the first portion which spins up. Moreover, the negative tangential wind shear from the centre to outside is large while the inertial stability is small, and thus the radial movement of the air is unhindered. This causes a convergence to the central portion, enhancing an

in-up-out circulation. (Gray, 1998) The developed storms continue their path moving westward and slightly poleward with a translation speed of 2-10 ms^{-1}, after which they either quickly dissipate (after encountering land, cold water, or because of unfavourable interactions with other atmospheric wind system); or they recurve poleward and eastward and are subjected to accelerations once entrained in the strong extratropical west-to-east winds of the middle and upper troposphere (at 2-10 km height).

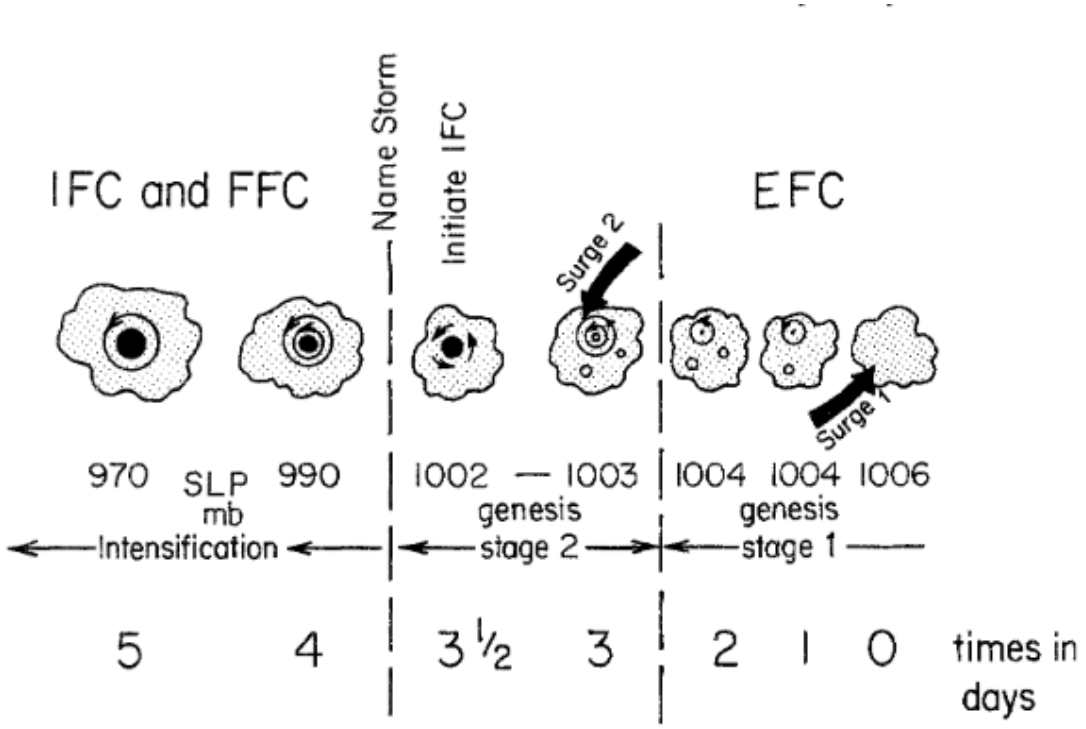


Figure 1.7: Westward propagating sequence of TC formation. Image taken from Gray (1998).

1.5. Tropical cyclones and global warming

1.5.1. Tropical cyclones and climate

Tropical cyclones are strongly related to climate conditions. Strong correlations can be detected between inter-annual fluctuations of Atlantic hurricane activity and changes in the general circulation (Namias, 1955) and climatic parameters such as the phase of the quasi-biennial oscillation and Sea Surface Temperature (SST) in west Africa(Shapiro, 1982a,b). El Nino was shown to be related to Atlantic hurricane activity, as wind shear is greater in this basin during El Nino years, preventing the formation of TCs (Gray, 1984;

Goldenberg and Shapiro, 1996). Landsea and Gray (1992) found a relevant correlation between hurricane activity and sub-Saharan Africa rainfalls. The abnormally high hurricane activity in 1995 was detected by Saunders and Harris (1997) to be related to very high SSTs. A year later, Shapiro and Goldenberg (1998) noted that the real cause of the 1995 TC activity was more related to the SST gradient and vertical wind shear. Thanks to the work of Goldenberg et al. (2001), it was also discovered how Atlantic hurricane activity is linked to the North Atlantic Oscillation. Even the Madden-Julian Oscillation has an influence on both TC genesis at a global scale and their rapid intensification (Klotzbach, 2014).

1.5.2. Tropical cyclones and climate change

TCs are characterized by a natural variability that hampers the detection of any ongoing trend. Although Henderson-Sellers et al. (1998) were unable to detect any changing trend in the number of cyclones or differences in locations with respect to past years because of limitations of Climate Models, there are results pointing towards the possibility of an increase in Maximum Potential Intensities (MPI) of TCs proportional to increases in CO_2 concentrations. These results were also confirmed in the Third Assessment Report of the IPCC, where it was noted that a change in numbers could be related to changes in the behaviour of the El-Niño Southern Oscillation (ENSO). At the time it was accepted that with a doubling of carbon dioxide concentration the MPI of TCs was likely to increase by 5-10% followed by an increase in peak precipitation rates of 20-30% (YDJG Houghton et al., 2001). These result were based on the estimation of the MPI of TCs based on average climate conditions (Emanuel, 1988; Holland, 1997; Tonkin et al., 2000) and a few modelling studies (Knutson and Tuleya, 1999; Walsh and Ryan, 2000; Knutson et al., 2001). Even if the result were compliant with modelling studies, the resolution of these models is too coarse to make these results conclusive.

Earlier studies which were questioning whether there would be more TCs in a warmer climate did not converge to a unique answer, as they were unable to reproduce TC formation without large discrepancy from the reality in some basins (Walsh, 2004).
Neither global (Webster et al., 2005) nor regional trends in specific ocean basins (Chan and Xu, 2008; Kubota and Chan, 2009; Callaghan and Power, 2011) have been detected. Some trends have been discovered in the North Atlantic, but still their reliability is under discussion (Holland and Webster, 2007; Landsea, 2007; Mann et al., 2007).

One of the main issues in determining the influences of climate change on tropical cyclones is that they are characterised by a strong inter-annual and decadal variability, making

trends hard to detect. Since accurate TC intensity estimation needs several measurements over the lifetime of a TC, data about it are particularly sensitive to the technology and methodology applied (Seneviratne et al., 2012).

While Kossin et al. (2007) were doing global reanalyses of TC intensities using homogeneous satellite records, they found that using different technologies through the years introduced a non-stationary bias, inflating trends of intensities. However, an increasing trend in TC intensities can be detected even if this bias is removed. Further studies Elsner et al. (2008) have indeed revealed significant trends for higher quantiles, especially for the 90th percentile, indicating that the strongest TCs are getting stronger due to increasing SSTs, which is qualitatively consistent with the heat-engine theory of TC intensities (Elsner et al., 2008). This suggests a link between SST and climate change, even if this is not proven because of the limited data available (Seneviratne et al., 2012).

Since the greenhouse effect is able to enhance the imbalance between atmosphere and ocean energy content, hindering the net long-wave radiative flux, Emanuel (1987) worried about the effects an increase in GHGs content induced by human emissions could have on this equilibrium and on TC intensities. His estimates on TC intensities for increasing SST of tropical oceans were based on a simple equation relating SST, mean temperature of a cold source at a certain distance from the center of TC, specific enthalpy near the surface, enthalpy of air parcels near the sea surface, and dimensionless transfer coefficients of momentum and enthalpy. According to his estimates, an increase of 3°C of SST at the tropics would increase the wind speed by 15-20%.

Another line of research was carried out by Emanuel (2007), who investigated time series of the previous 25 years of power dissipation, which is an indicator that aggregates cyclone frequency, duration, and intensity and is a measure of TC power consumption. This indicator showed upward trends in the North Atlantic and less relevant upward trends in the North Pacific. Emanuel also found that these trends could be related to SST, tropopause temperature, and vertical wind shear, even though other scientists questioned whether the local SST or the difference with local SST and mean SST are the more relevant variable (Swanson, 2008). In the Summary for Policymakers of the AR4 of IPCC it was reported that there *likely* has been the increase in intense TC activity in some regions since the 1970s (Alley et al., 2007). Afterwards, Kunkel et al. (2008) found that also frequencies in the North Atlantic *likely* have increased in the last 100 years, noting that in the same period the average SST has increased as well. In the same report, also an increase in power dissipation is noted as *likely* substantial since the 1950s. However, Knutson et al. (2010) asserted that the confidence in these detected trends are not trusted completely since it is not demonstrated (due to lack of data) that they exceed the natural

variability that has been detected through the past millennia.

There is also a branch of climatology which studies the link between the increase of tropical SST and greenhouse gases increase. Some studies reported that there is a direct correlation between these two variables (Karoly and Wu, 2005; Knutson et al., 2006; Santer et al., 2006; Gillett et al., 2008). Moreover, Kunkel et al. (2008) stated that the anthropogenic influence on greenhouse gases increase has *very likely* impacted SST increase in the North Atlantic and on TC genesis in North West Pacific. The potential intensity theory by Bister and Emanuel (1998) endorses the link between TC variability and the thermodynamic state of the tropics, since with an increase in the ambient potential intensity a shift toward greater intensities is expected to follow (Emanuel, 2000; Wing et al., 2007). Thus, understanding completely the relationship between SST and potential intensity appears to be a research milestone to be achieved (Seneviratne et al., 2012). More recent studies suggest that the correlation between potential intensities and differences between local SST and spatially averaged SST in the tropics is more important than the SST itself (Vecchi and Soden, 2007b; Xie et al., 2010; Ramsay and Sobel, 2011). In this perspective, the increasing SST gradients are not expected to continuously grow together with global warming. This suggests that the physical link has not yet been fully understood (Seneviratne et al., 2012).

TC intensity is, among the variety of metrics of tropical cyclone activity, the most physically linked to climate variability in the potential intensity theory framework (Kossin and Vimont, 2007). A reason of concern is whether the detected statistical relationship between ambient environmental conditions (reported in the above paragraphs) and TCs genesis (DeMaria et al., 2001) remains valid in future conditions of increased GHGs concentrations (Seneviratne et al., 2012). As an example, the increase of SST above the threshold of 26°C could either lead to an increase of TCs frequencies or to an increase of the SST threshold itself (Ryan et al., 1992; Dutton et al., 2000; Yoshimura et al., 2006; Bengtsson et al., 2007; Knutson et al., 2008; Johnson and Xie, 2010).

Moreover, uncertainties in reanalysis data related to areas of development and evolution of TCs introduce another variable in the complex task of cause identification of cyclogenesis changes (Bister and Emanuel, 2002; Emanuel, 2010). Knutson and Tuleya (2004) concluded that detectable increases in of TC intensity may not occur before decades even if an SST increase due to global warming is expected in long-term evolution model simulations. Furthermore Bender et al. (2010) confirmed such results in a study involving high-resolution dynamical down-scaling of the SRES A1B scenario and concluded that

the strongest Atlantic storms could not show a statistically significant trend before 2050. Another contribution of Bengtsson et al. (2007) further explored the possible evolution of TCs in a changing climate using coupled climate models at different resolutions using the IPCC SRES A1B scenario and evaluated for the end of 19th, 20th, and 21th century. The results showed that the number of TCs is expected to decrease, while the number of intense storms showed a significant increase in the 21st century with respect to the previous one. In particular, the number of storms with maximum wind speeds greater than $50ms^{-1}$ is expected to increase by 30%.

Another issue is related to the availability and reliability of data records. Ocean data records are limited to the period from the second half of the 20th century to the present. The reliability of TC data is also not constant throughout this period, since technologies and methods of measurement have changed. Also, TCs are characterized by intrinsic random variability and variability related to climate modes such as El-Niño, making it difficult to detect trends. IPCC AR4 asserts that it is *more likely than not* that human activities had some influence on the increase in more intense TCs. However, further studies downgraded this statement to *low confidence*, since the understanding of the process is hampered by the entities of the above explained uncertainties (Knutson et al., 2010). Facing questions about the interaction between TCs and climate change, the considerations are usually focused on whether climate changes have already affected TCs and how can these impact the future (Walsh et al., 2016). This field of the research uses information coming from different approaches. The data are obtained from *Paleotempestology*, whose purpose is to detect (by using proxy variables) over different geological eras if there have been variations in TC number or intensity. This could be helpful to have a longer dataset which can be used to understand the dependency of TCs on climate (Walsh et al., 2016). Paleotempestology is limited since it is based on data related to single basins. Thus, it is unable to discern basin-wide trends or variability from systematic changes in TC tracks (Simon Banholzer, 2014). Analyzing stalagmites' oxygen isotopes, Haig et al. (2014) discovered that TC activities in the Australian region are at their minimum since 1,500 years ago. Brandon et al. (2013) analyzed over-wash deposits, discovering that the decrease in intense hurricane frequency from 2,500 years ago up to nowadays was interrupted from only from 1,500 to 600 years ago. These variations in TC climatology remain unexplained. Korty et al. (2012a) and Korty et al. (2012b) were able to show how in past geological eras, and especially during the Last Glacial Maximum and Holocene, even if the climate was characterized by lower temperatures, TC activity did not seem to be less relevant with a clear trend.

Sugi et al. (2015), running a General Circulation Model (GCM) with a climate cooler by 4

K, obtained significantly increased TC frequency. In his simulations, TC formation could occur even where SSTs were well below 26°C (most of them form over oceans with SST $>$ 24°C and some of them even with 22 °C $<$ SST $<$ 24°C. Since the 19th century, large amounts of data have been recorded from ship observations. These data are constantly under review and allow us to have information relative to TC positions and intensities. The IBTrACS dataset (described in detail in 3) represents in this sense a helpful tool which can facilitate global climate trend analysis. Creating a homogeneous dataset is a challenging task, since the estimates of TC intensities especially are continuously improving (Walsh et al., 2016). More recently, some studies detected trends over the past few decades thanks to satellite-based temporally homogenized datasets. Significant results are obtained at the basin level. The most significant trends are detected in the North Atlantic and Western North Pacific basins, increasing and decreasing, respectively. In the South Pacific and South Indian basins, the trends are less significant, while in the East North Pacific no trend was detected. Last, in the North Indian basin (NI) there was insufficient data to draw conclusions (Walsh et al., 2016). Globally, a significant movement poleward of the path of TCs was discovered between 1982-2012 (Kossin et al., 2014). Also, an increase in the proportion of intense hurricanes was detected both globally and at the basin level in almost all locations, due to a link between a global warming indicator and the proportion of intense hurricanes (Holland and Bruyère, 2014). ENSO was shown to be strongly connected with TCs (Camargo et al., 2010), but the projected changes of ENSO are still characterized by a low confidence and so the same holds for the projected effects on TC activity (Christensen et al., 2013).

In summary, here are reported the conclusions of the IPCC SREX report on the projected changes and uncertainties in TC activity:

1. Meehl et al. (2007) states that a number of models indicate a *likely* increase in the intensity of winds and precipitations related to TCs; also (with a lower confidence) a decrease (increase) in the frequency of weak (intense) storms;

2. Knutson et al. (2010) confirmed the above conclusions and their statistical confidence. The increase of maximum wind speed and rainfall rates are found to be *likely*, the frequency increase of most intense storms in some basin are *more likely than not* and the overall decrease (6 - 34%) or stability of overall TC frequency is found to be *likely*. Greater uncertainties are detected at the individual basin scale. Rainfalls in TC areas are projected to increase by 3-37%;

3. An increase in TC intensity in a warming climate is endorsed by both theoretical considerations of Emanuel (1987) and results from idealized dynamical models of

(Knutson and Tuleya, 2004); with an increase in mean maximum wind speed of 2-11% globally (Seneviratne et al., 2012).

4. The same studies showed little to no change in the frequency of TC genesis at a global scale, Gualdi et al. (2008); Sugi et al. (2009); Murakami et al. (2011) together with an increase of frequencies of stronger events;

5. Randall et al. (2007) state that spatial resolution of current ocean-atmosphere models should not be considered high enough to capture TC intensities;

6. Some other studies reported that higher-resolution global model are able to reproduce only coarser characteristics of global models (Chauvin et al., 2006; Oouchi et al., 2006; Zhao et al., 2009);

7. Downscaling techniques using boundary conditions provided by either reanalysis datasets or coarser resolution models can provide some advances in this research field (Knutson et al., 2007; Emanuel et al., 2008; Knutson et al., 2008; Emanuel, 2010);

8. Bender et al. (2010) obtained promising results applying a cascading technique involving models at global and regional scales whose results were fed to a very high-resolution hurricane model. Among the results, the simulation projected a 28% overall reduction in frequencies and an 80% increase in frequency of strongest storms;

9. Major limitations in TC projections are represented by the inability of capturing natural climate patterns which are known to be influential on them, such as ENSO or MJO (Seneviratne et al., 2012);

10. The decrease in frequency is to trace back to an increasing vertical wind shear (Vecchi and Soden, 2007a; Zhao et al., 2009; Bender et al., 2010), a weakening tropical circulation (Sugi et al., 2002; Bengtsson et al., 2007) or an increasing middle troposphere saturation deficit (Emanuel et al., 2008);

11. The increase in tropics water vapour is expected to evolve into an increase in rainfalls related to TCs (Trenberth et al., 2005; Held and Soden, 2006);

12. Caution on the validity of statistical relationships detected in present and past observation needs to be taken, since they could probably be no longer valid in a warming climate (Emanuel, 2007; Vecchi et al., 2008; Knutson et al., 2010; Vecchi and Soden, 2007b; Vecchi et al., 2006; Ramsay and Sobel, 2011);

"The proportion of tropical cyclones that are intense is expected to increase (high confi-

dence), but the total global number of tropical cyclones is expected to decrease or remain unchanged (medium confidence)" (IPCC, 2022). In the Technical Summary of IPCC it is also reported that intense tropical cyclones are expected to increase by 10%, 13%, 30% as consequence of an increase of global average temperature of 1.5°C, 2°C and 4°C respectively.

1.6. How TC data are collected

Before World War II, the records regarding TC data were compiled locally by coastal stations, islands, and ships at sea. Therefore, many events escaped entirely or were observed only once, towards the end of their life (as TCs dissipate upon making landfall). Furthermore, the availability of records was not homogeneously widespread in different ocean basins, since shipping was more focused in certain ocean basins. Also, many storms which never made landfall have never been detected.

An increasing effort arose during the war because of the interest in naval operation security. Aircraft reconnaissance began at that time. While the first operations did not penetrate storms, on the 27th of July 1943 an AT-6 trainer was able to enter a hurricane in the Gulf of Mexico. In following years, TC reconnaissance via airplane became routine in the WP and NA basins.

Another step forward was achieved in the late 1940s, when Wexler (1947) analysed the very first radar image of a TC, from which it was possible to obtain information about the structure of TCs. This allowed the detection of TCs located far from coasts, up to several hundred kilometers offshore.

Doppler radar in late 1950s allowed trustworthy estimates of ground speed of aircraft. After that, wind speed estimates quality increased.

A great innovation occurred in 1960, when from a polar orbiting satellite a TC was captured for the first time. From the 1970s, virtually all TCs have been captured thusly. Moreover, techniques were developed to evaluate their intensities by recognising patterns or by performing measurement of infrared radiance. Although most of our knowledge about the structure of TCs derives from reconnaissance aircraft measurements (and the dropwindsondes which they deploy) and from ground-based or airborne radar, satellites and simulation models have greatly enriched the information available.

The historical records of TCs have issues such as the continuously changing technology and reporting protocols within the agencies reporting this type of data (Seneviratne et al., 2012). Recently, there have been efforts to obtain more homogeneous records of TC

intensity globally. This is achieved through satellite imagery—which, however, are quite limited, as they started only in 1970s (Knapp and Kossin, 2007; Kossin and Vimont, 2007).

1.7. Tropical Cyclones Damages Modelling

It is well established how the extreme winds of TCs can cause devastating damage. However, most of the damages and deaths caused by TCs are due primarily to the flooding associated with TC-induced storm surge and extreme rainfall. Before comparing historical losses, it is necessary to normalize them to account for changes in population, wealth, and inflation (Neumayer and Barthel, 2011; Pielke et al., 2008; Pielke and Landsea, 1998). Pielke et al. (2008) made an effort in establishing whether over the past century there has been an increase in damages of TCs in terms of economic and life loss in the US. After normalization of the data available considering inflation, wealth of the nation, and changes in urbanization of the coasts, no trends were noticed. These results are supported by the lack of a trend of increasing TC landfall frequency or intensity in the same period (Seneviratne et al., 2012).

The damages caused by TCs are extremely relevant and globally they represent a leading cause of damage in economic terms (Klotzbach et al., 2018). For example, Hurricane Katrina in 2005 costed approximately $108 billions NOAA, while in Bangladesh, in 1970 a storm killed almost 500 thousands people (Emanuel, 2003).

Throughout the years, some indexes have been developed to correlate TC-induced damages and their intensities measured as maximum wind speed, especially linking damages to some exponential of the maximum wind speed of a TC. For example, damages were hypothesized to be proportional to the square (Pielke and Landsea, 1999), the third power (Emanuel, 2005), or the ninth power of the maximum wind speed (Nordhaus, 2010). To model these dependencies into a unique formula, the damage of a TC (L) can be expressed as (Murnane and Elsner, 2012):

$$L = \alpha V^{\beta} \tag{1.2}$$

Where $\beta = 3 - 9$.

Other relationships of the loss with some other important measures such as TC size, its speed of motion, the induced precipitation, or factors such as building construction, roughness of the surface, or some other effects such as storm surge are possible but less easily adopted by decision makers and stakeholders (Murnane and Elsner, 2012). Murnane

and Elsner (2012) proposed an exponential relationship between aggregate normalized economic losses and wind speed at landfall, providing better results for the U.S, and found an increment of 5% for each ms^{-1} increase in wind speed. There are a number of indexes which attempt to estimate the potential damage caused by TCs. Among others, the Saffir-Simpson Scale (Simpson and Riehl, 1981), the already cited PDI (Emanuel, 2005), the Hurricane Intensity Index (HII) and Hurricane Hazard Index (Kantha, 2006), the Hurricane Severity Index (HSevI) (Hebert et al., 2010), the Carvill Hurricane Index (CHI) (Smith, 2010), the Integrated Kinetic Energy (IKE) (Powell and Reinhold, 2007; Kozar and Misra, 2014), and the Cyclone Damage Potential (CDP) (Holland et al., 2019). The dependence on storm intensity is different among the indices. Some of them are connected to maximum wind speed, others to its square or cube. Also, they can include shape and translation speed (Holland et al., 2019).

2 | Tropical Cyclones Intensity Estimation: State of the Art

2.1. Dvorak Technique

This section introduces the Dvorak Technique (DT), one of the most used method for tropical cyclone intensity estimation.
The following paragraph is broadly taken from Dvorak (1984), which represents the last update of the technique developed by Vernon Dvorak. Thus, for readability purpose it is cited here once.
Before Dvorak introduced his technique, others used satellite images for TC monitoring. For example, Fett (1964) and Timchalk et al. (1965) provided tools for intensity estimation based on patterns of cyclones such as the appearance of the eye, its banding, and size of cloud patterns. However, these first methods, used to have poor performance on cyclones with unclear features or sudden changes. The availability of more data and the emergence of infrared (IR) sensors have created new opportunities for analyzing tropical cyclones (TCs) using satellite imagery. In particular, it became possible to detect clear, subsequent steps of TC development, and to model their variability to aid recognition. By the late 1960s, the first models were developed to forecast the intensity of a TC 12-36 hours in advance based on its current and past features.

The Dvorak technique consists of a set of methods and rules which allows for TC intensity estimation and forecasting by using visible (VIS), enhanced infrared (EIR), and digital infrared images captured from satellites. These images are used to build a model of the development of the cyclones which leads to estimation of current and future intensity.The primary advantage of IR technology is its ability to facilitate continuous monitoring. Unlike VIS images, IR images can be captured at night, allowing for uninterrupted surveillance. The model contains both cloud pattern descriptions and information regarding detected changes during subsequent steps.

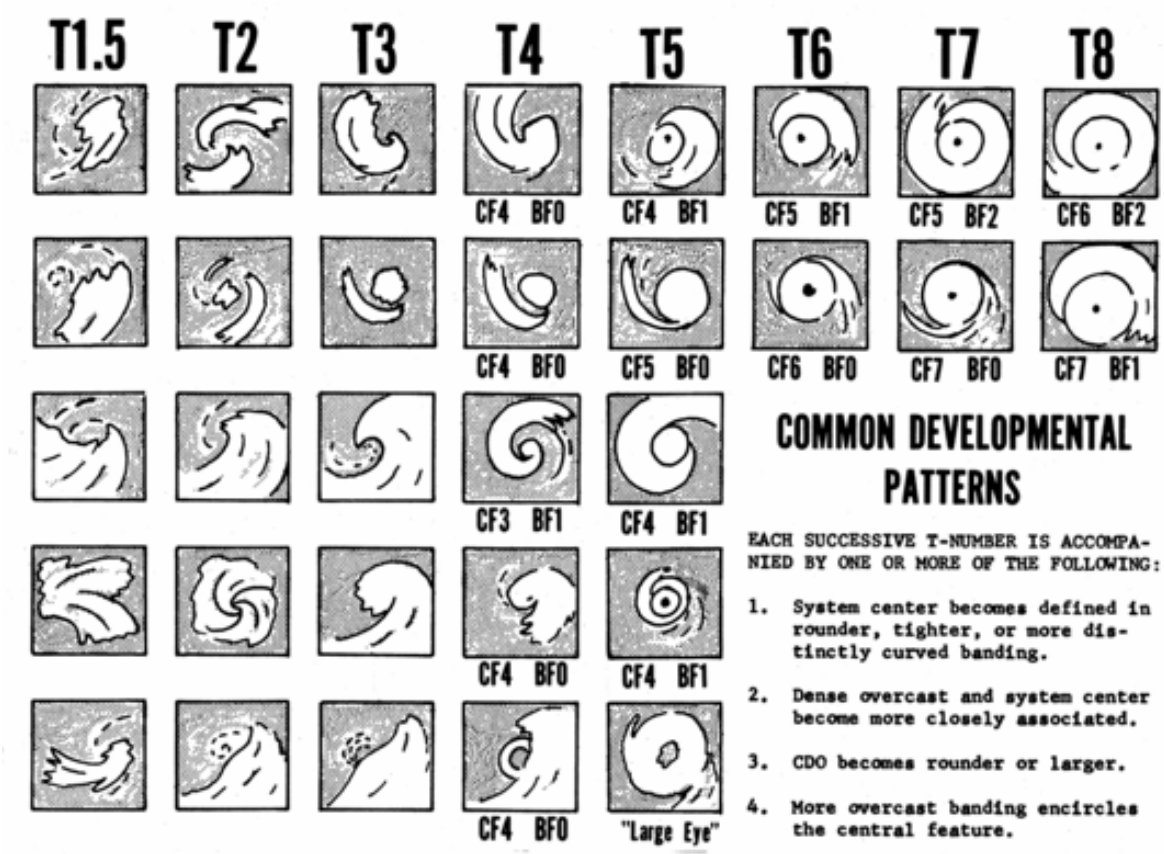


Figure 2.1: Patterns searched in image processing for T-number attribution.

The Dvorak technique consists of ten steps (Dvorak, 1984).

1. *Locate the Cloud System Center* (CSC), which is determined as the point through which all curved cloud lines or bands are converging to;

2. *T-number estimation from cloud feature measurement.* Through these measurement on cloud features, which could be performed when can be clearly recognised, a Tropical cyclone number (T-num) is identified. This number is related to intensity through given tables, as explained later on. T-numbers series, obtained from the analysis of successive images, define the model of TC development, through which it is possible to define the rate of growth of its intensity;

3. *T-number estimate from pattern comparison with model.* When a Central Cold Cover (CCC) appears, the TC arrests its development. In such case, the analysis considers the past T-num and, if T-num<=3.0, it maintains the model trend for 12 hours, while if T-num>=3.5, it maintains T-num. The T-num determined is used for the final estimate and the analyst can proceed to step nine;

If step two was followed, a second estimate of the T-num is performed.

4. *Determine past 24-hour trend.*It is determined whether the TC is developing or weakening;

5. *Determine Model Expected T-num.* This estimate is provided after a comparison between the current and the previous pictures, determining whether TC is following its development path. This comparison is done on selected cloud features such as band curved. Considering these, a second intensity estimation is provided, according to an extrapolation of the intensity change curve of the development model;

6. *Determine the pattern T-num.* A refinement of previous estimate is obtained by comparing current picture with patterns representing the stage of development reported in Fig.2.1.The adjustment of the intensity estimate can be performed through the analysis of differences;

   From Step 7 to 9, the T-number estimate is chosen, constraints are applied and the Current Intensity is determined.

7. *T-number determination.* According to the different situations, chosen T-num determined in Step 2 or MET determined in step 5-6;

8. *Final T-number Constraints.* Several rules are applied, determining the maximum T-number that could be defined according to the number of days passed from its formation and to the maximum change that can be obtained in two successive images;

9. *Current Intensity (CI) number rules.* When the TC is intensifying, one should select the T-num. On the other hand, if the TC is starting to weaken, one should choose the previous CI. For successive stages, a CI that is either equal to 1/2 or 1 higher than the T-num should be chosen;

   The last step is related to forecasting.

10. *24-hours forecast.* From past trends a 24-h forecast is obtained, unless it is clear that environmental conditions are changing;

Once determined the CI through tables of conversion, estimates of MSW or MSLP are obtained. A detail from Velden et al. (2006) of these conversions is reported in the following table.

Table 2.1: Summary of Dvorak (1984) Atlantic and West Pac wind pressure relationships (Velden et al., 2006).

CI	MSW (kt)	Atlantic MSLP (hPa)	West-Pac MSLP (hPa)
1.0	25	-	-
1.5	25	-	-
2.0	30	1009	1000
2.5	35	1005	997
3.0	45	1000	991
3.5	55	994	984
4.0	65	987	976
4.5	77	979	966
5.0	90	970	954
5.5	102	960	941
6.0	115	948	927
6.5	127	935	914
7.0	140	921	898
7.5	155	906	879
8.0	170	890	858

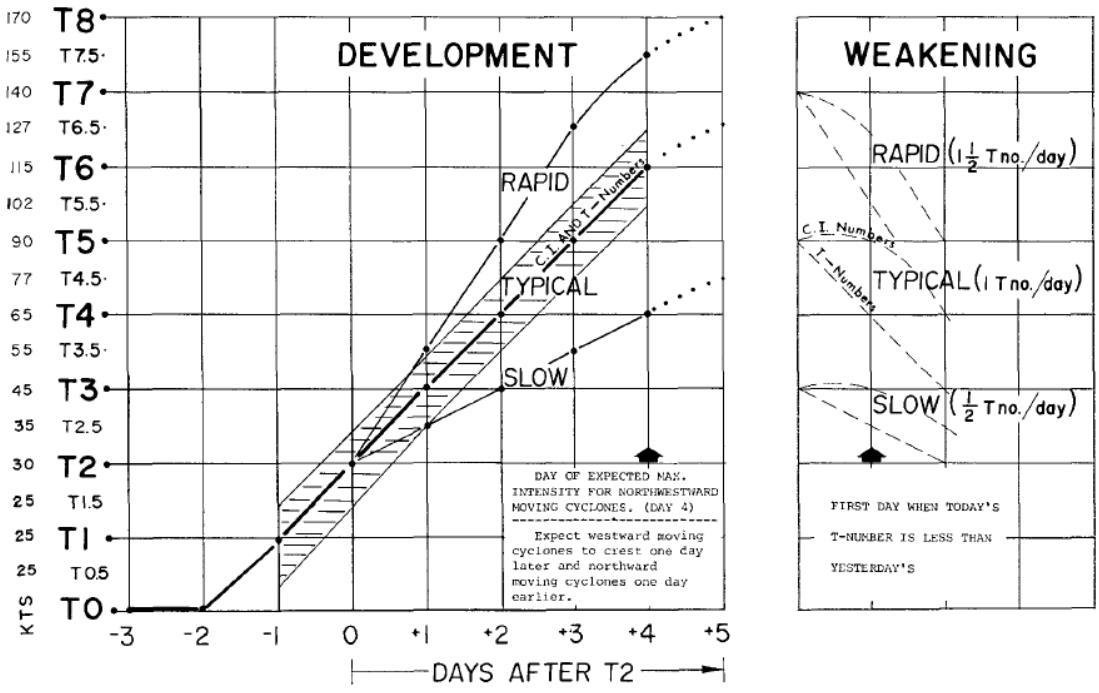


Figure 2.2: Graph representing a model of TC development; used by the analyst to monitor if TC is developing or decaying.

Latest version of the Dvorak technique introduced innovations which enhanced the relia-

bility of cloud feature measurements (Velden et al., 2006):

1. Introduction of patterns such as *curved band* (CB) and *shear patterns* enabled the analysis of eye-free or central dense overcast (CDO) TCs;
2. The use of IR imagery enabled the use of *embedded center* (EMBC) pattern recognition in place of CDO;
3. Enhanced Infrared (EIR) is used for eye pattern analysis, using cloud-top IR temperatures difference between the center of the cyclone and its surroundings. This is especially relevant, being a strictly objective measurement which lays the foundations for automating the analyses (Zehr, 1989; Velden et al., 1998);

Although the Dvorak technique once was a groundbreaking innovation in this field and it is still used by many meteorological agencies, limitations have been identified throughout the years (Velden et al., 2006):

1. It suffers from two intrinsic sources of errors. First, measures of intensity such as wind and pressure are not direct, but rather inferred from cloud patterns. Second, the analyst must make subjective interpretations, limiting the objectivity of the whole method Zehr (1989),Velden et al. (1998),Olander et al. (2004),Olander and Velden (2007);
2. The use of IR images can be misleading in presence of cirrus, since they obscure cloud patterns beneath them. Central Dense Overcasts (CDOs) captured by an IR could be concealing weak eyes or developing eye-walls. The analysis of Embedded Centers (EMBCs) tries to solve this issue but it is difficult and not reliable at present. Moreover, IR cannot capture concentric eyes and eye-wall replacement cycles, which are found to be indicative of intensity changes (Willoughby et al., 1982);
3. When using geostationary satellites, large scan angles underestimate TCs with small eyes since they are hidden by the eye-wall;
4. The technique shows limited skill for rapidly weakening TCs moving over strong SST gradients;
5. The technique cannot be applied either in case of particularly small TCs, since cloud pattern measurement is not possible, or in the case of large monsoon depression, because of the lack of central deep convection;
6. It does not account for TC motion;

2.2. Dvorak-based Techniques

The Dvorak technique inspired through the years other methods, aiming at the reduction of the influence of the analyst on the estimation. The first attempt came from Dvorak (1984), who proposed to provide an estimate of intensity by measuring the gradient between the warmest IR pixel within the eye and the warmest at 55-km radius. In particular, it was found that the T-number increases with decreasing surrounding temperature and increasing eye temperature. Some improvements were proposed by Zehr (1989) and Dvorak (1995):

1. the reference radius value was substituted with a range (25-125 km);
2. it was preferred to average computations over time intervals to reproduce more realistic trends;
3. some modifications were introduced to include also cases with colder eyes;

Zehr (1989) proposed the Digital Dvorak technique (DDT), an objective method on which modern versions of the Dvorak technique are based. It is "objective" in the sense that it removes the subjectivity introduced by an analyst. This technique was found to be satisfactorily accurate for well-organized TCs with low intensities (Velden et al., 1998).
Velden et al. (1998) introduced the Objective Dvorak Technique (ODT), which involves Fourier transform analysis of the center and cloud-top regions in IR imagery. In the ODT, T-numbers are replaced by their time average, to smooth out their variability. The accuracy of ODT is on par with that of the subjective EIR technique Velden et al. (2006), but it has two crucial limitations. First, it can be applied only to storms above an intensity threshold of hurricane strength (i.e. T-4.0, 65 kt). Second, it still requires human intervention in the location of the TC center. The Advanced Objective Dvorak Technique (AODT) by Olander et al. (2004) made further progress by automating the localisation of the TC center. The idea behind it is to start from a first guess using short-term track forecasts, which are taken as an objective estimation of the centre from which to search for curvature patterns and strong gradients in the brightness temperature (BT) field. The performance of this technique is only slightly worse than that of the manual Dvorak technique, but with the added benefit of skipping the subjective identification of the storm centre (Olander et al., 2004).
Olander and Velden (2007) further improved on the AODT by introducing new rules to govern the time-averaging scheme of T-numbers, intended to hamper or allow intensity changes in specific time periods. Also, they introduced a bias adjustment for MSLP noticed by Kossin and Velden (2004). Furthermore, they introduced a regression-based

equation that uses a set of large-scale environmental variables to determine the intensity from eye and CDO imagery. Still, even this more advanced technique struggles to estimate the intensity of TCs with small eyes, as the IR images have too low resolution to capture them fully (Olander and Velden, 2007).

2.3. Other traditional techniques

Another popular technique for TC intensity estimation is the Deviation Angle Variance technique (DAV-T), developed by Pineros et al. (2008),Pineros et al. (2010), and Pineros et al. (2011). Originally (Pineros et al., 2008), the technique consisted of a set of operations:

1. Computation of the gradient of IR image and definition of the IR gradient field in vector form;
2. Choice of reference pixel and computation of the deviation angle between the vector representing the gradient and the radial line. Deviation angles are determined for all pixels up to a certain distance from central ones;
3. Definition of the distribution of angles and variance determination (DAV). This is a measure of cloud organization: the more they are organised in an axisymmetric structure, the lower the variance;

Pineros et al. (2010) improved the technique by adding a map of DAVs, reporting variance values in reference to the centre of the storm, to identify forming cyclones. In Pineros et al. (2011), this map was used to estimate the intensity of TCs by means of a parametric curve related to DAV values. The relationship was formalised as a sigmoid function with two parameters (α and β) fitted to input data and one (σ^2) representing the filtered DAV value to smooth out the signal. In this technique, DAV values are computed at each pixel, and the pixel corresponding to minimum variance is used for intensity estimation. Ritchie et al. (2014) further improved the DAV-T technique by quantifying the level of organization of clouds and relating this to the intensity of the TC.

2.4. Machine learning and Tropical Cyclone Intensity Estimation

Machine learning (ML) techniques are used more and more often in the field of climatology, partly due to their widespread success in most scientific fields, and partly because they appear to be able to solve some of the intrinsic limitations of traditional methods such

as the Dvorak and DAV techniques. In particular, Pradhan et al. (2018) highlighted four major limitations of traditional techniques that ML could alleviate:

1. Inconsistency of RMSE in different regions of an image;
2. Significant effort/resources needed for pre-processing, such as detection and measurement of specific patterns and their changes over time;
3. Complexity of their applications (requiring analyst expertise);
4. Poor generalisation, as traditional techniques are usually tuned to specific basins and latitudes;

One of the first attempts to use ML for TCIE were by Bankert and Tag (2002), who used the k-nearest neighbor (k-NN) algorithm to analyse data from the Special Sensor Microwave Imager (SSM/I); and by Kulkarni et al., who trained two neural networks (on the same data as Bankert and Tag (2002)): a two-layer perceptron used for classification, and a three-layer perceptron used for intensity estimation (i.e. solving a regression problem). Similarly, Sakuragi et al. used multi-channel Tropical Rainfall Measurement Mission Microwave Imager (TRMM/TMI) Brightness Temperature for classification with the k-means algorithm.

The advent of more powerful architectures, and especially the transition from the CPU to the GPU as the main processing unit, produced increased interest in Deep Learning (DL) architectures for image processing tasks (He et al., 2016). Pradhan et al. (2018) were the first to use *deep convolutional neural networks* (deep-CNNs) for the problem of TC intensity classification. Their model was not only orders of magnitude faster than previously available ones, but also beat them in terms of RMSE (Pradhan et al., 2018). For training, their model used hurricane images from the TC repository of the Marine Meteorology Division of the U.S. Naval Research Laboratory [1] labelled through HURDAT2 data [2]; for testing it used a recon-only dataset [3] composed of 2,646 images. Since their training dataset was unbalanced, they performed over-sampling on hurricanes of Saffir-Simpson categories H3 or higher. The over-sampling was done by interpolating data from 6h to 2h. Other data augmentation techniques were applied, such as horizontal and vertical flips and image rotation (90°, 180°, and 270°). The dataset obtained consisted of 48,828 images. They also speculated that two other data augmentation techniques, brightness and contrast enhancing, could be beneficial for this task. However, since the network learns to predict intensity from values of BT represented in the input pixels, modifying

[1] http://www.nrlmry.navy.mil
[2] http://www.nhc.noaa.gov/data/#hurdat
[3] http://www.nhc.noaa.gov/recon.php

the relationships between pixels is not believed to be an effective augmentation technique. Thus, these techniques are not explored in this thesis.

Combinido et al. (2018) provided an interesting example of how *transfer learning* could help researchers. They used the Visual Geometry Group 19-layer CNN (VGG19) (Simonyan and Zisserman, 2014) pretrained on ImageNet to regress TC intensity from IR images, and they obtained results comparable to methods requiring human interventions for feature extraction. They used IR images with a time resolution of 6 hours from the Weather Home archives (University, 2015) and best track data from RSMC Tokyo-Typhoon Center of the Japan Meteorological Agency (JMA). The dataset consisted of 10,015 images from 493 TCs. The authors maintained the pre-trained weights of the VGG19 except for those of the last few fully connected layers, which they re-trained on their TC data. Interestingly, the trained network was found to extract features considered important in manual TC intensity estimation, such as organised cloud circulation and the eye of the TC.

Another interesting contribution for regressing TC intensity with DL was provided by Chen et al. (2018). First, they released a benchmark dataset (named *TCIR*) to make it easier for researchers to compare the performance of their algorithms [4]. Observational data were taken from the GridSat (channel IR1, WV, and VIS) satellite mission for brightness temperature (Knapp et al., 2011), the CMORPH dataset for information on global precipitation (Joyce et al., 2004), and the Joint typhoon Warning Center (JTWC) and revised Atlantic Hurricane database (HURDAT2) for TC best tracks. Nevertheless, in their studies they used only the IR1 channel of the GridSat data, as they found WV added little value (and potentially could cause overfitting, being so similar to the IR1 channel) and the VIS channel was unstable, being dependent on daylight. The CNNs they used were AlexNet (Krizhevsky et al., 2012) and VGG19 (Simonyan and Zisserman, 2014). Interesting findings of the experiments are reported below:

1. The difference between the discrete output searched by classification models and the continuous one of regression results in a need for different loss functions. For classification, a cross-entropy loss function is commonly used, while for regression MSE is more appropriate;

2. They argued that max-pooling layers, commonly used inside CNNs for classification, could be deleterious for TCIE regression. For this reason, they decided to remove

[4] https://www.csie.ntu.edu.tw/~htlin/program/TCIR

them from their models;

3. Although random cropping is frequently used in classification tasks, they chose not to use this data augmentation technique, arguing that it may cause some key features of the TC to be missed. Instead, they chose to perform data augmentation using random rotations;

4. They removed all dropout layers from their networks, showing increased performance compared to models that used them;

With these considerations, they modified AlexNet into a network they called they termed *CNN-TC*, which at the time was state of the art compared to other traditional techniques. Lee et al. (2019) developed a two-dimensional and a three-dimensional CNN. They used grid search to optimize the filter size and the depth of the convolutional layers. Also, they tested different evaluation metrics, and in the end used MAE and RMSE to select the most optimised version of CNN-based model, since more significant for the identification of their overall accuracy. They were also the first authors to use heat maps (Zeiler and Fergus, 2014) to visualize the patterns learned by a CNN trained for TCIE. Interestingly, it was noted that the architecture was searching for patterns similar to those identified in the Dvorak technique.

2.5. Data imbalance and image augmentation

Data imbalance is problematic for deep neural networks, in terms of both convergence in training and generalization (Buda et al., 2017), He and Garcia (2009). Having classes of data with wildly different samples sizes hampers network performances (Buda et al., 2017), as the network will tend to learn to predict the most represented class regardless of the input. If the data imbalance is severe enough, it also affects regression algorithms, as they would be biased towards the skew in the distribution of the samples. In TCIE, the relative scarcity of data due to the temporal resolution and the methodologies for maximum wind speed estimation cause strong heterogeneity in the data. For example, a dataset could have a high number of samples characterised by wind speed velocity of x, but zero samples with a velocity of x+-1. In this case, even if we split the dataset into classes of one unit, we still would have strong data imbalance.

There are three main approaches to tackle data imbalance (Krawczyk, 2016):

1. Data-level methods, which imply the modification of the distribution of the data and/or the removal of difficult/extremely undersampled classes;

2. Algorithm-level methods, which modify the learning algorithms themselves (for example at the level of the loss function), alleviating the intrinsic bias towards majority classes;

3. Hybrid methods, which combine the other two;

Data-level methods, also referred to as *data augmentation techniques*, generally involve oversampling of the underrepresented classes or undersampling most represented classes, or both. In particular, oversampling is robust in most of the cases, while undersampling is typically used only in cases of extreme data imbalance (expressed as the ratio between the most and the least numerous class) and when only a few classes are heavily over-represented (Buda et al., 2017). Data-level methods are usually random, thus introducing risks of deleting important data or adding redundant information (Krawczyk, 2016), (Seiffert et al., 2010). Data-level methods can be subdivided into two categories of transformation: geometric transformations and pixel-level transformations. Examples of traditional geometric data augmentation techniques are as follows:

- Horizontal and vertical flipping: the image is mirrored along horizontal or vertical axis;
- Random rotation: the image is rotated according to an angle of random amplitude within a certain interval;
- Random shifting: the center of the image is shifted by a random number of pixels, in order to obtain an image of the same dimension of the input one, a zero padding can be applied;
- Random cropping: the image is cropped around a random point and within a rectangle of random size;

Examples of pixel-level data augmentation techniques are as follows:

- Image noising: the values in random pixels are substituted with random values;
- Random brightness change: pixel values are substituted with higher ones, increasing the overall brightness;
- Image blurring: filters of different shapes are applied to the image, averaging the pixel values with the neighbour ones;

Algorithm-level techniques, on the other hand, try to deal with skewed distributions by modifying the learning algorithm directly (Krawczyk, 2016). An example would be to use a loss function with weights that are proportional to the sample size of a class: the lower

the number of samples in the class, the higher the weight used in the loss function (Zhou and Liu, 2010). In doing so, the importance of different samples is regulated increasing the skill in learning rarer data. Other algorithm-level methods, such as one-class learning (Japkowicz et al., 1995) exist in literature, but they are less relevant for the purposes of the thesis and therefore will not be discussed further.

The issue of data imbalance is seldom discussed or addressed directly in papers dealing with TCIE. When authors do mention the use of data augmentation techniques, they typically follow guidelines for their use that come from different areas of computer vision—and, more often than not, that are designed for classification algorithms. However, not all data augmentation techniques described above are appropriate for TCIE regression. For example, altering the brightness of satellite images or shifting them could alter the contents of the images in an unwanted way. There currently are no guidelines in the literature for which data augmentation methods are most appropriate for TCIE, which is a gap this thesis attempts to fill.

3 | Data and Methods

The flowchart in 3.1 provides an overview of the experiment sequence. First, the optimal combination of backbone architecture and image crop dimension is searched. By doing so, it is possible to minimise the validation loss measured during training on the dataset which is chosen as input. Then, entering the core of the experiments, the same dataset is augmented, testing different combinations of the augmentation techniques (horizontal and vertical flipping, random rotation and random erasing), monitoring the input distribution with a Gini-inspired coefficient which is described in the following paragraphs. The trained models are further evaluated in terms of mean absolute error on test dataset. Once the augmented dataset providing best performances is defined, a novel approach is proposed, introducing Label Distribution Smoothing coupled with a loss function named Focal-R, devoted to alleviate the imbalance of input dataset, by weighting more the losses on the represented classes samples. Lastly, Grad-CAM heatmaps are obtained to visualise patterns of the image used for predictions. This chapter begins by describing the data used and the choice of a backbone DL architecture. As transfer learning has been shown to be effective for TCIE, some state-of-the-art deep-CNNs are tested in training and validation with different resolutions of input images. Once the model and resolution of the image are chosen, the selected augmentation techniques are tested. A Gini-inspired coefficient is introduced, as a metric to measure the effects of the augmentation on the original distribution of the data.

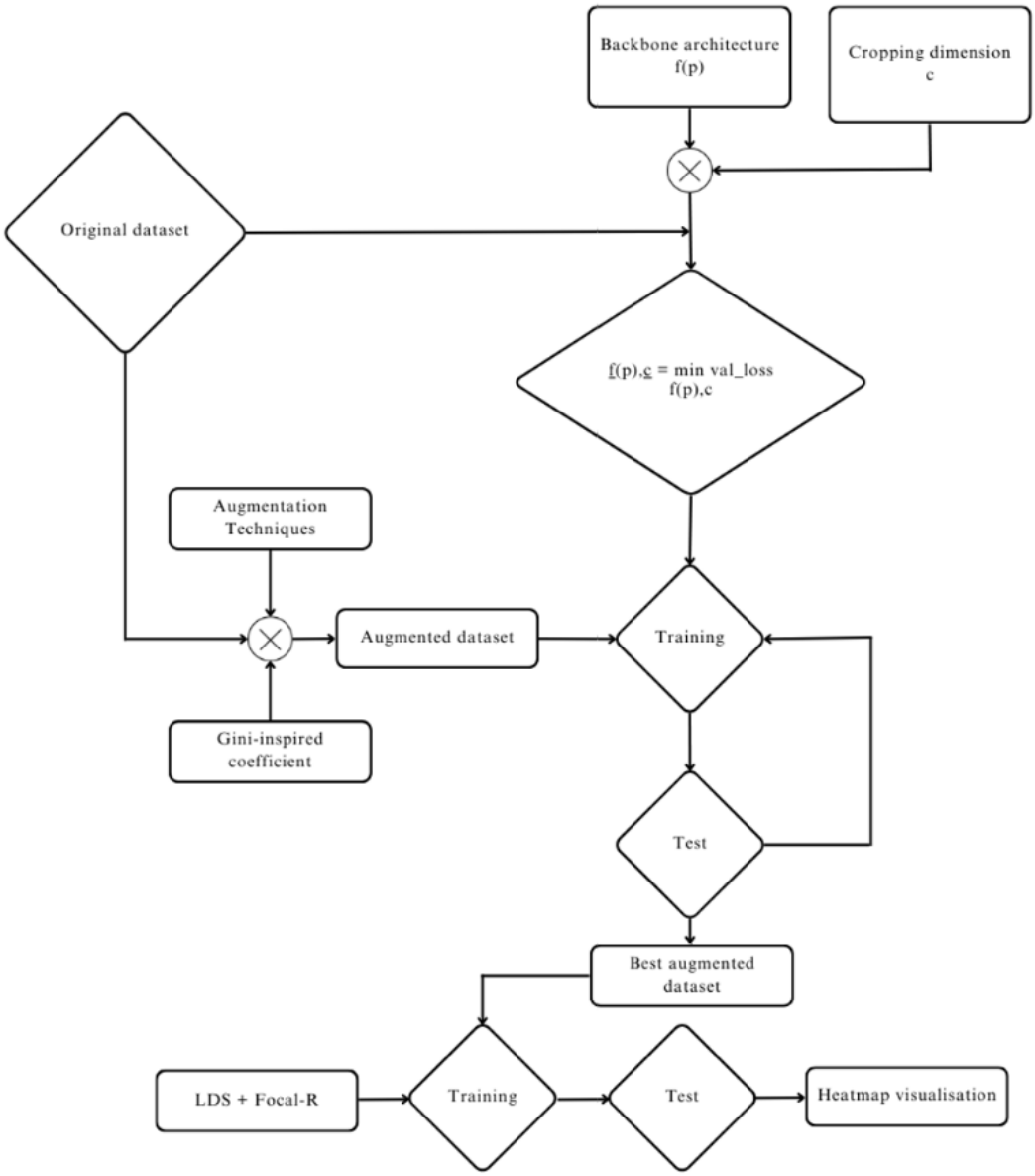


Figure 3.1: Methods flowchart.

Next, the random erasing augmentation technique is described, motivating the changes applied to the original algorithm by Zhong et al. (2017). Afterwards, two techniques for resolving data imbalance are described that have never been applied to TCIE before, namely Label Distribution Smoothing (LDS), to obtain the effective distribution of target dataset Yang et al. (2021), and Focal-R, a loss function that enables neural networks to learn better under-represented samples Yang et al. (2021). Finally, GradCAM heatmaps are proposed as a tool to "explain" the final neural network topology, gaining information on how they use inputs to predict the output.

3.1. Data

3.1.1. GridSat

The Gridded satellite-B1 (GridSat-B1) data (Knapp et al., 2011) is a calibrated and mapped geostationary dataset built to overcome difficulties characterising previous data records. GridSat was built upon ISCCP B1 data (Knapp, 2008), being processed to enable their use by a wider part of the research community. The resulting product has three channels: infrared (IR) at 11 μm (IR); visible (VIS) at 0.6 μm; and water vapour (WV) at 6.7 μm. In this thesis the IR channel at 11 μm is used as the only input, following the recommendations by (Chen et al., 2018), who found that WV channel provides information similar to IR, and that the VIS channel was very unstable due to the daylight variation. The spatial resolution of the dataset is 0.07 ° latitude (8 km at the equator) and the temporal resolution is 3 h. The dataset begins from 1980 and reaches the present day, with some lag not being real-time, and covers the globe from 70°S to 70°N. Throughout the years, the coverage has not been constant, with the Indian Ocean being the last gap to be filled (Fig. 3.2).

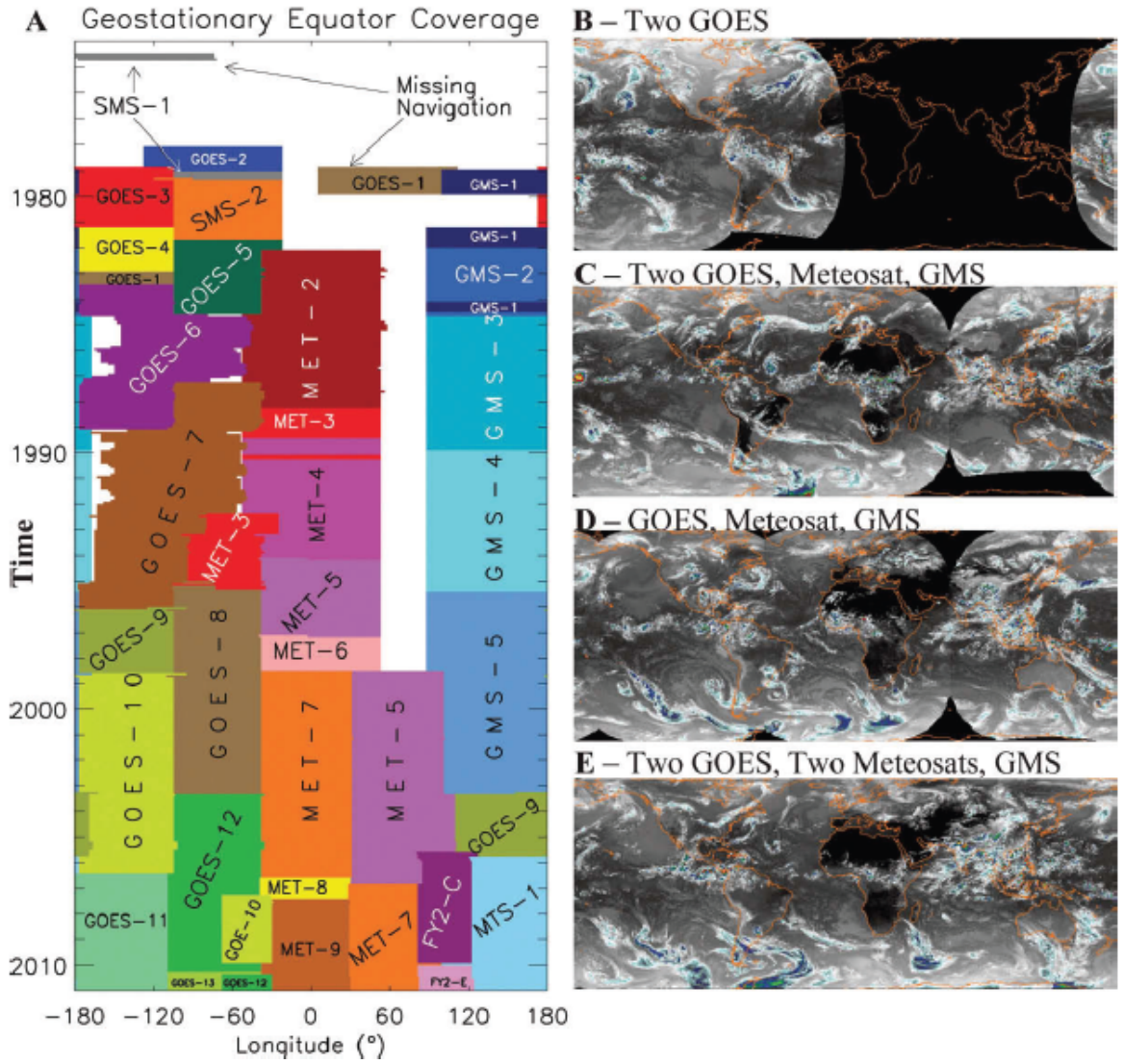


Figure 3.2: GridSat coverage (from Knapp et al. (2011). On the left: ISCCP B1 coverage. On the right, details on typical coverage of different periods and a different number of satellites. B: 1980; C: 1982-1998 with four satellites; D: 1985-87 or 1989-92 with three satellites; E: five-satellites coverage, 1998-present.

3.1.2. IBTraCS

The International Best Track archive for Climate Stewardship IBTraCS (Knapp et al., 2010) dataset represents a unification of best-track data coming from agencies throughout the globe. "Best-track" data are obtained by the postseason reanalysis of TC information coming from ships, surface stations, and satellites. These data, which are heterogeneous in the way they are collected and stored, are then converted into a common data format (netCDF) and merged in a way that compensates their different original formats and

makes them uniform. To do this, several operations are performed, from the elimination of duplicate or spurious data to the homogenization of differences in storm localisation to the conversion of maximum sustained wind (MSW) measurements to a common unit of measure. IBTraCS contains storm tracks at 6 h temporal resolution across all ocean basins. For each track, it reports the position (in terms of latitude and longitude coordinates) of the TC every 6 hours, its maximum sustained wind speed (MSW) and mean sea level pressure, and measures of its size and translation speed. For this thesis, we use only data relative to the position and MSW of each cyclone. In particular, we use the latitude/longitude coordinates to localize the TCs in GridSat images and the MSWs as the targets.

3.2. Choosing a backbone architecture

The first step was the choice of a pre-trained backbone architecture, leveraging the benefits of transfer learning discussed earlier. Due to the limited computational power available for this thesis, the performance of the architectures tested was evaluated on the non-augmented dataset, in terms of training and validation loss (which was the Mean Absolute Error). The following architectures were chosen for testing, as they represent the state of the art of many computer vision benchmarks, including ImageNet (Deng et al., 2009), MS COCO (Lin et al., 2014), and CIFAR (Krizhevsky, 2009):

- ResNet (He et al., 2015): a deep-CNN introducing the use of *residual blocks*, which are shortcut connections linking non-adjacent layers. Introducing these "skip connections", it is possible to overcome the problem of "vanishing gradient", where gradients become so small they cannot update weights. ResNet is widely used as backbone architecture in a variety of research fields. Its first version, ResNet50, achieved state-of-the-art performance in many tasks. Later on, other two versions of ResNet, with increasing depth, have been released: ResNet101 and ResNet 152;
- DenseNet (Huang et al., 2016): designed for image classification, this architecture introduces the use of *dense blocks*. These blocks are built to be connected to all feature maps from deeper layers and to provide input to all shallower ones. Reusing all features already extracted, each layer benefits from a reduction of the number of parameters to be tuned. The architecture can be scaled in the number of layers and parameters. The most common variants are DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264;
- EfficientNet (Tan and Le, 2019): a family of state-of-the-art deep-CNNs in image recognition designed to be more efficient than some of their counterparts. The focus

is on a scaling approach which enables balanced tuning of the width, depth, and resolution of the network. By doing so, it is possible to achieve better performance in image recognition, reducing computational costs and the risk of overfitting. There are six variants of EfficientNet (from B0 to B6);

All these architectures are designed to perform classification tasks. Therefore, their top layers (the *head*), which are the ones specialized for classification, must be rebuilt so that they output a single number (needed for regression). To allow for a fair comparison, the same structure is used for fine-tuning the new head:

- The pre-trained model is loaded with weights derived from training on ImageNet;
- A 2-D Global Average Pooling layer is added after the last convolutional layer of the pre-trained model, obtaining a 1D vector from the feature map;
- A Batch Normalization layer is added to obtain a normalization of the activation function;
- Top dropout rate at 40% of the feature to prevent overfitting;
- A dense layer with a single output is added to produce the final prediction in regression mode.

With the above modifications, ResNet50, DenseNet121, EfficientNet-B0, and EfficientNet B2 were trained using different cropping of the images (250 km, 500 km, and 750 km from the centre). EfficientNet-B0 reached the lowest loss on the validation data (Fig.3.4), and therefore was selected for all further analyses. Furthermore, it was found that the best performance was reached when images were cropped within a 750 km box from the centre (Fig.3.3); therefore, all subsequent experiments used this crop. Interestingly, validation loss consistently decreases with the increase of the cropping dimension, highlighting that useful information is contained at a higher distance from the centre.

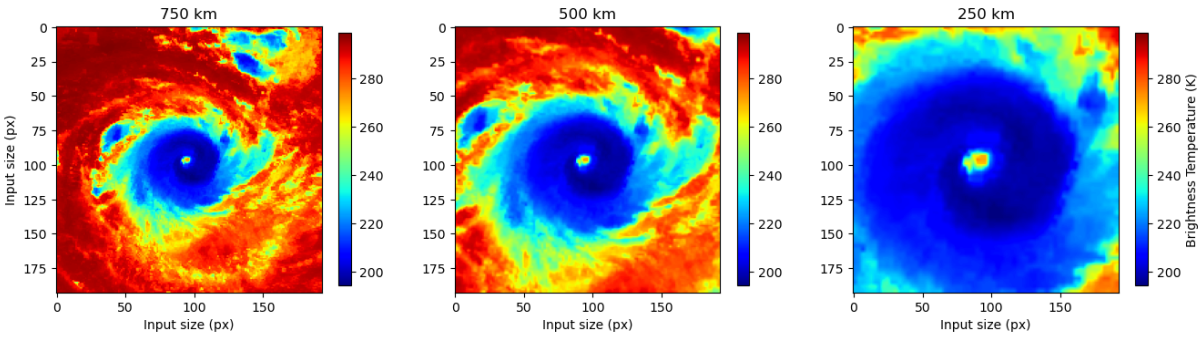


Figure 3.3: Different crops applied in the choice of the deep-CNN architecture.

It should be noted that there is still room for improvement on the chosen architecture, for example by tweaking its hyper-parameters or increasing/decreasing the depth of its layers. However, due to resource limitations in terms of computational power, it was decided to proceed with this layout, especially as it has been validated extensively in the literature.

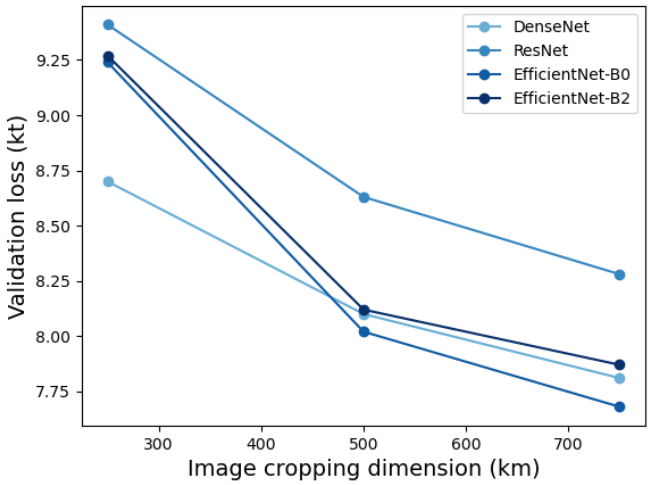


Figure 3.4: Results of preliminary tests on backbone structure.

3.3. Data augmentation

Having chosen the backbone model, several experiments were performed on different augmentation scenarios.

3.3.1. A new Gini-inspired coefficient

The coefficient here introduced is an economic coefficient. However, it can be a powerful instrument to measure the distribution under study is similar to the uniform one. To the authors' knowledge, this application of the Gini coefficient has never been developed so far. Comparing this metric for the distribution of the original dataset and the distribution of the augmented dataset, one has the feeling of whether the resulting sample is closer to a uniform distribution than it started. Using this coefficient it is possible to find an optimal combination of augmentation techniques applicable to the dataset.

The Gini coefficient, proposed by Corrado Gini, is defined as a measure of statistical dispersion of a distribution of a social group, measuring income inequality or wealth inequality within a nation or a social group. The Gini coefficient provides a consistent

measure of how well distributed a certain resource is (Farris, 2010). This coefficient was derived from speculations made on the Lorenz curve (Lorenz, 1905), which was intended to be a visual representation of a certain quantity in a population. Following the paper of Farris (2010), the principles of these two statistical subjects are described below. Defining the quantity as Q, ordering the population in ascending order of the share of Q given a certain class width the Lorentz curve $y = L(p)$ we have that the fraction of population characterized by low values of p is the poorest, with associated low values of $L(p)$, which are low values of the shares of the total Q. For the computation of the Gini coefficient, it is necessary to compute the line of perfect equality, which is simply given by $L(p) = p$, which represents a uniform distribution in which each member of the population accounts for the same share of quantity to the total amount. At this point the Gini index is computed as an integral which quantifies the distance between the Lorentz curve and the curve of perfect equality:

$$G := 2 \int_0^1 [p - L(p)]\, dp \tag{3.1}$$

The index is scaled by a factor of 2 so that in case of perfect inequality (where one person has 100% of the share of Q) G = 1. On the other hand, in case all the individuals share the same amount of the total, G = 0. In Fig.3.5, the Lorenz curve for a population divided into two classes is represented.

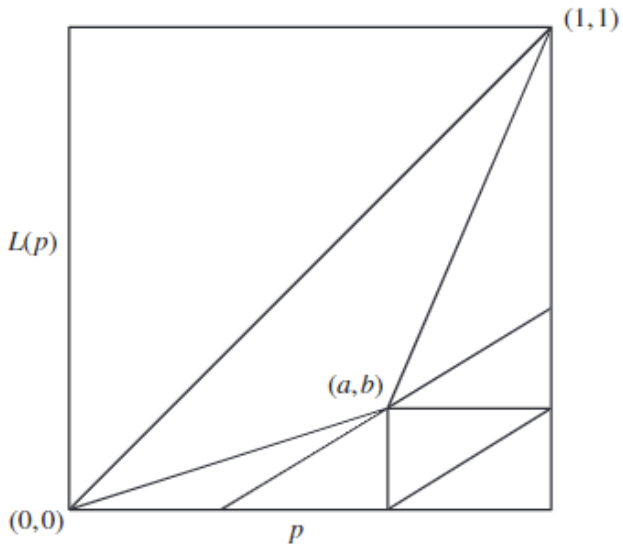


Figure 3.5: Lorenz curve for a population divided into two classes. Image taken from Farris (2010).

For a discrete distribution, the Gini coefficient is equal to half the relative absolute difference divided by the mean of values, and it is obtained according to the following formula:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{i=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu} \tag{3.2}$$

Where x_i is the frequency of class i.

The Gini coefficient corresponds to a geometric interpretation of the Lorenz curve. It is equal to the ratio between the area below the line of perfect equality limited by the axes, and the area delimited by the line of perfect equality and the Lorenz curve.

Considering a *population* of images of tropical cyclones, we can view them as belonging to different categories of intensity, where weaker TCs are detected with the higher absolute frequency with respect to stronger ones. The concepts of Gini and Lorentz can therefore be applied in the following terms. The earning of the "social classes" can be substituted with the intensity classes in terms of discretized maximum wind velocities (e.g. $0-10, 10-20, \dots kt$) and the quantity Q as the percentage of images depicting a cyclone classified with a certain velocity belonging to the class. In so doing, the Gini index becomes 0 in the case of perfect equality and 1 in the case of perfect inequality, and the Lorentz curve can be used as a graphical representation of the inequalities of shares in the population.

In the case of right-hand-tailed distributions (which is the case of the dataset used here), the Lorenz curve lies above the line of perfect equality. As a consequence, the mathematical meaning of the Lorenz curve itself is lost, and the graph obtained is thus a common P-P plot. Nevertheless, the Gini-inspired coefficient described above preserves the usefulness of the original coefficient, still varying between 0 (perfect equality) and 1 (perfect inequality). In other words, the more G tends towards 0, the more the distribution is similar to a uniform. Also, the area between the curve of the distribution and the line of perfect equality is still a graphic representation of the distances between the actual distribution and the uniform distribution. Fig.3.7 shows a graphical representation of how the Gini-inspired coefficient and the P-P plot change after data augmentation.

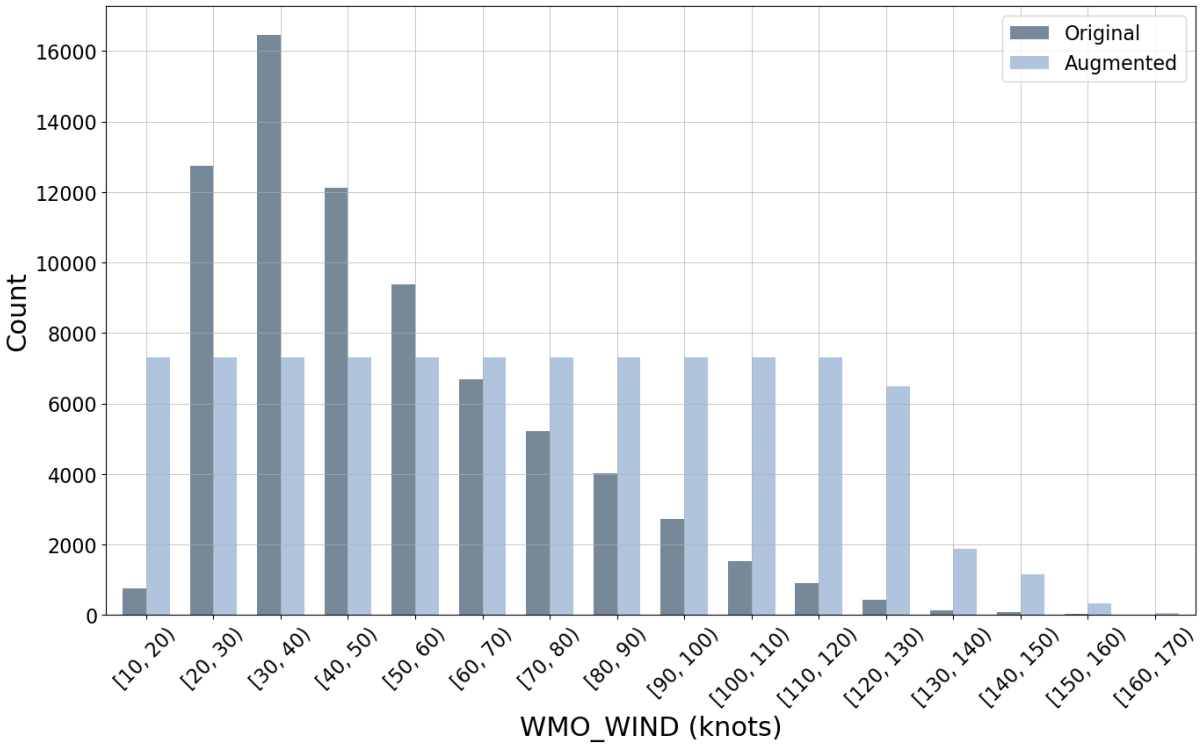


Figure 3.6: Original dataset (Gini coefficient = 0.68) and augmented dataset (Gini = 0.22).

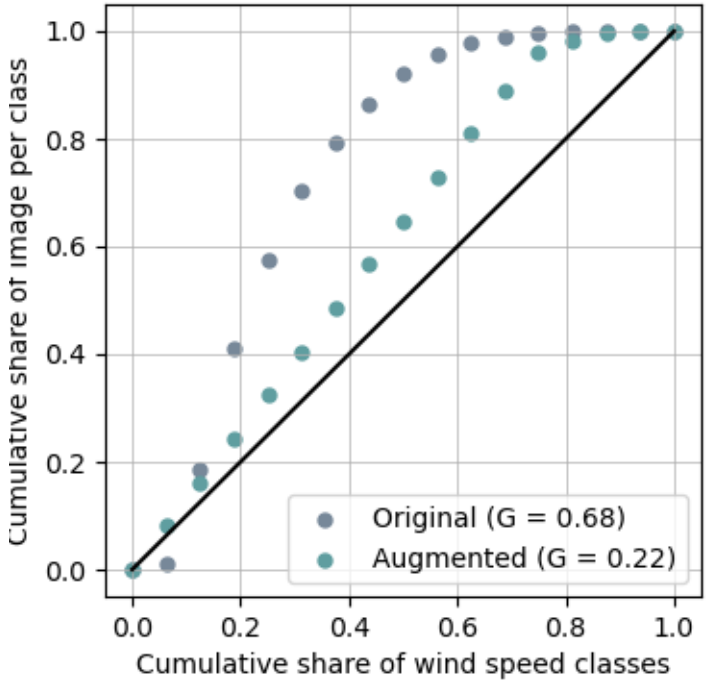


Figure 3.7: P-P plot of the original and augmented datasets.

The Gini-inspired coefficient is thus a quantitative measure of the distribution obtained, giving more useful information than just the number of images in each class. In particular, it makes it easy to compare the efficacy of different augmentation scenarios:

- Fixing the distribution (monitored by the Gini-inspired coefficient) and changing the augmentation techniques used, it is possible to determine the efficacy of different augmentation techniques with respect to the decided target distribution;
- Fixing the augmentation techniques to be tested and changing the distribution, it is possible to control whether increasing the degree of augmentation, performance improvement is achieved.

These considerations open the possibility to choose from the start the most suited augmentation scenario depending on the problem. It may be the case that an augmentation technique reveals to improve model performance when applied to obtain a certain distribution (characterised by a certain G), but worsen performance when G decreases further. Or, it could happen that improvements are achieved only when G gets lower. As a consequence, it could be better to use the first technique in case of a lower degree of data imbalance, the second one in case of deep imbalance.

3.3.2. Data augmentation techniques

By using the Gini-inspired coefficient as a guide, several augmentation scenarios were tested to improve the generalization of EfficientNet-B0, with the goal of understanding which augmentation techniques worked best. The techniques employed were random rotation, horizontal and vertical flipping, horizontal and vertical random erasing. For random rotation (Fig. 3.8), the images were rotated by an angle varying randomly between +270° and -270°, with a step of 5°. Horizontal (Fig. 3.9) and vertical (Fig. 3.10) flipping, as previously mentioned, apply a mirroring with respect to the vertical or horizontal axis respectively. These techniques have been used before for TCIE, but their usefulness has not been quantified.

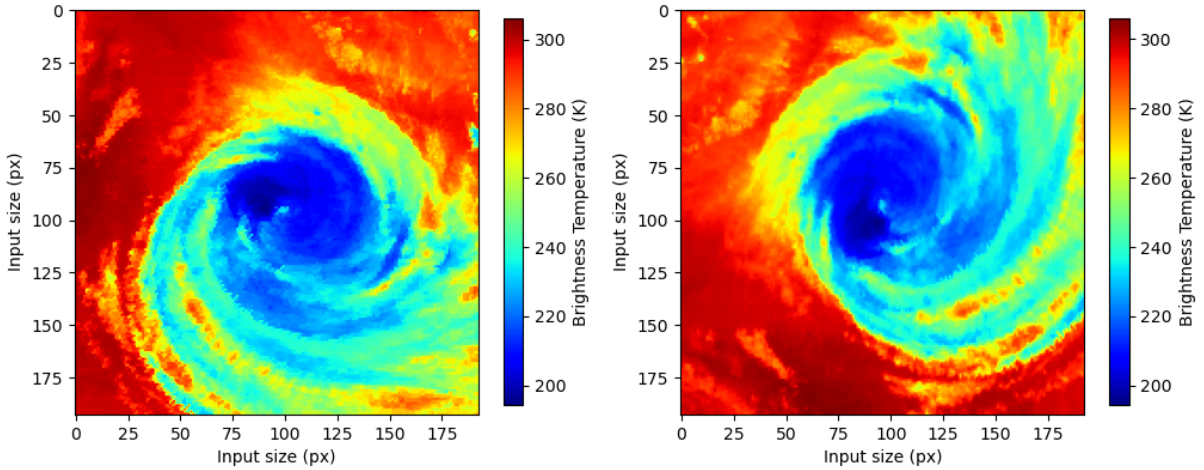


Figure 3.8: Example of random rotation applied to an image.

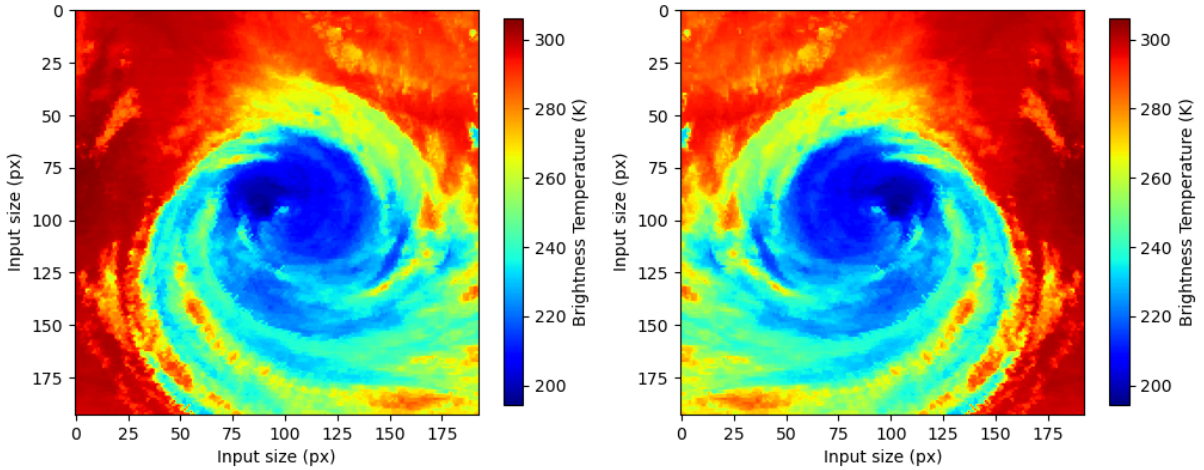


Figure 3.9: Example of horizontal flipping applied to an image.

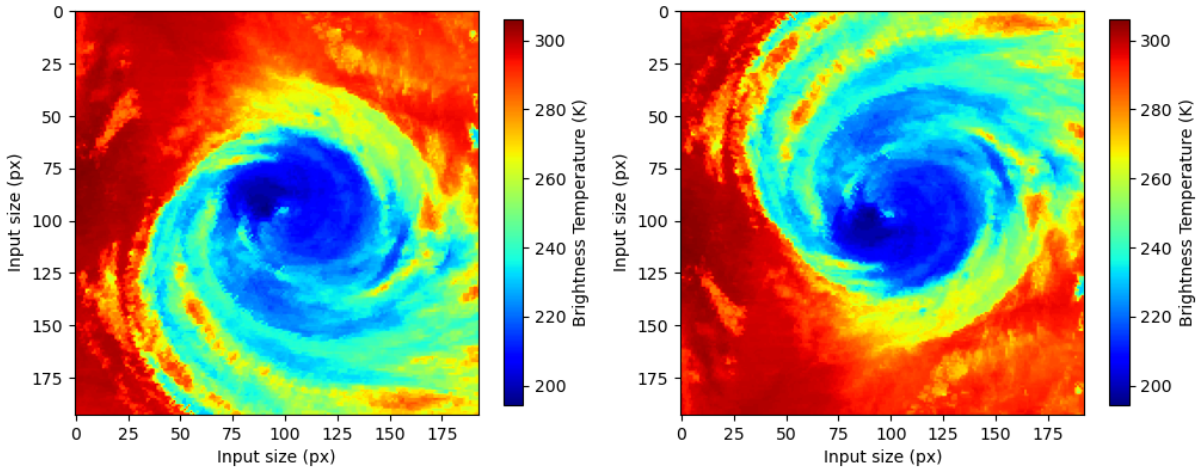


Figure 3.10: Example of vertical flipping applied to an image.

The use of random erasing, on the other hand, is a novelty for TCIE. In its original formulation, random erasing was conceived as an augmentation technique that superimposes the image squares of random dimensions in random positions, filled with random values Zhong et al. (2017). Its effect is to encourage the network to not overly rely on any single feature in images, and to learn to adapt to occlusions. Moreover, it does not introduce any new parameter to be tuned and it does not require additional memory consumption, as it can be performed before training the network. As it is known that satellite images can present missing data (which appear as stripes of missing values) due to sensor errors, it is reasonable to assume that random erasing could make the network invariant to this source of error. Therefore, an adaptation of the random erasing algorithm is proposed here (Fig. 3.11), as follows:

- Before training, each image is subjected to a random erasing transformation with a prefixed probability;
- For images randomly selected for random erasing, a random point along the edge of the image is selected;
- A rectangle of random width (maximum 0.1 x image width) is drawn from that point to the opposing edge of the image and filled with random values.
- Optionally, this is repeated in the opposing direction (i.e., if a horizontal bar of random values is created, a vertical one can be created afterwards, as shown in Fig. 3.12).

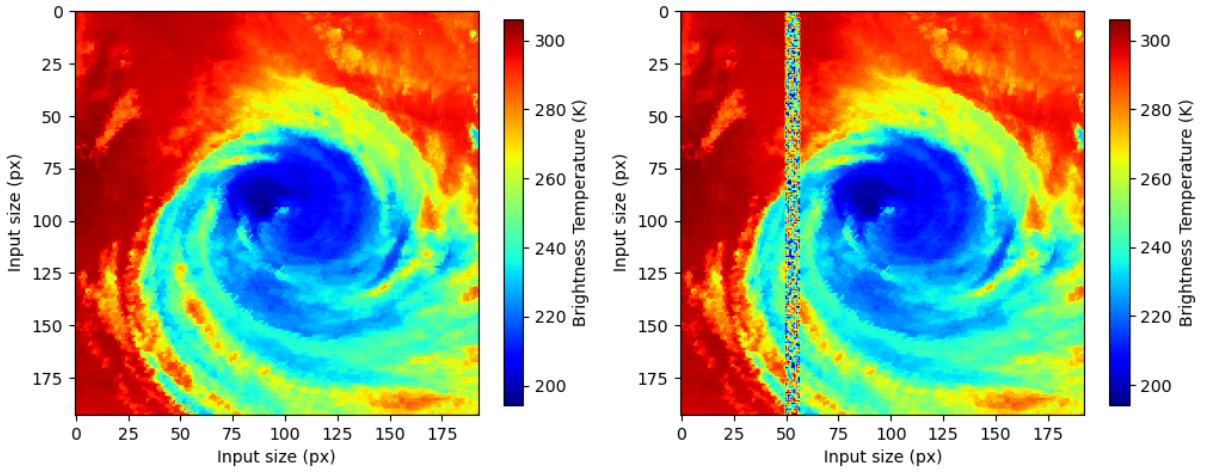


Figure 3.11: Example of vertical random erasing.

Random rotation, horizontal and vertical flipping, and random erasing were applied to images according to a pre-determined probability. In other words, each image in the dataset had a random chance of receiving one or more of the transformations.

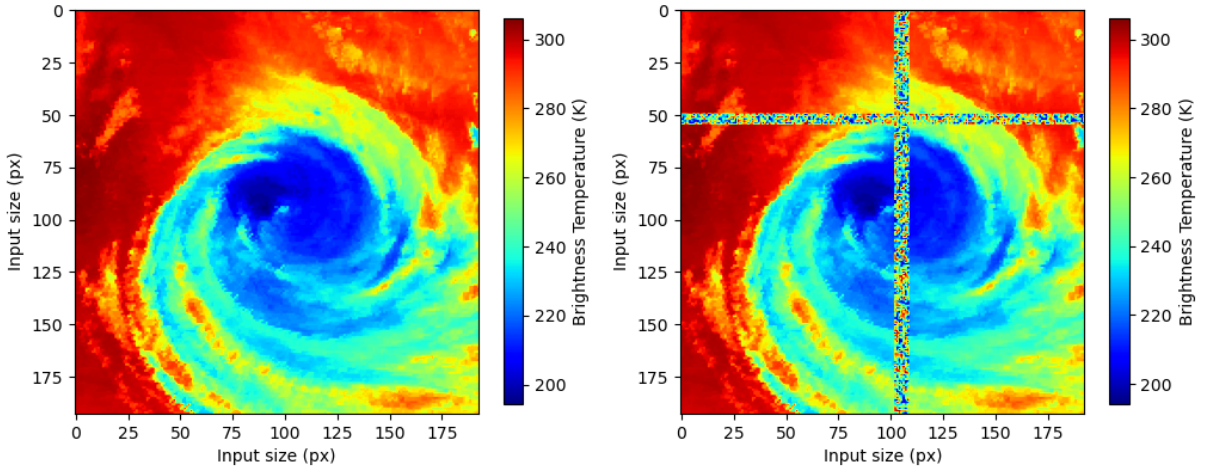


Figure 3.12: Example of horizontal and vertical random erasing.

Having established which data augmentation techniques to use and how to implement them, different augmentation scenarios (i.e., combinations of possible augmentation techniques applied to images) were chosen:

- *Business as usual*: random rotation, horizontal and vertical flipping;
- *All augmentations*: random rotation, horizontal and vertical flipping, vertical random erasing;

- *Flipping and erasing*: horizontal and vertical flipping, vertical random erasing;
- *Rotating and erasing*: random rotation, vertical random erasing;
- *Rotating and double erasing*: random rotation, horizontal and vertical erasing;

Similarly, different distribution scenarios were chosen, and monitored by the Gini-inspired coefficient:

- G = 0.28, with random undersampling of classes representing more than 15% of the total dataset, and augmentation of images of underrepresented classes up to a maximum of 10 times;
- G = 0.24, with random undersampling of classes representing more than 10% of the total dataset, and augmentation of images of underrepresented classes up to a maximum of 10 times;
- G = 0.23, with random undersampling of classes representing more than 10% of the total dataset, and augmentation of images of underrepresented classes up to a maximum of 12 times;
- G = 0.22, with random undersampling of classes representing more than 10% of the total dataset, and augmentation of images of underrepresented classes up to a maximum of 14 times;

The regulation on the augmentation was so gradual because the most intense class is very little represented. In the original dataset (G = 0.60), only four images out of 91,472 belong to this class. Below, the absolute frequency of images for the highest represented classes is reported:

3.3.3. Label Distribution Smoothing (LDS) and Focal Loss for Regression

In a recent paper, Yang et al. (2021) highlighted the need for a radical change of view when dealing with deep imbalanced regression, questioning the applicability of classification loss functions to regression tasks. Their main concern was related to the difference between categorical and continuous label spaces. In Deep Imbalanced Regression (DIR), the aim is to learn continuous targets from an imbalanced dataset. Thus, the division into classes loses meaning. Moreover, the imbalance is not constant throughout the frequency distribution since two samples could have the same associated frequency but be in two very differently represented neighbourhoods. Finally, values belonging to a class could be not represented at all. It is thus necessary to implement interpolation or extrapolation

Table 3.1: Absolute frequency of images by class with varying Gini-inspired coefficient.

Intensity class	**G = 0.60**	**G = 0.28**	**G = 0.24**	**G = 0.23**	**G = 0.22**
[10, 20)	753	8283	7317	7317	7317
[20, 30)	12758	10976	7317	7317	7317
[30, 40)	16452	10976	7317	7317	7317
[40, 50)	12117	10976	7317	7317	7317
[50, 60)	9386	10976	7317	7317	7317
[60, 70)	6683	10976	7317	7317	7317
[70, 80)	5215	10976	7317	7317	7317
[80, 90)	4019	10976	7317	7317	7317
[90, 100)	2714	10976	7317	7317	7317
[100, 110)	1519	10976	7317	7317	7317
[110, 120)	903	9933	7317	7317	7317
[120, 130)	432	4752	4752	5616	6480
[130, 140)	125	1375	1375	1625	1875
[140, 150)	77	847	847	1001	1155
[150, 160)	22	242	242	286	330
[160, 170)	4	44	44	52	60

of values. For this, they proposed two main techniques: Label Distribution Smoothing (LDS) and Feature Distribution Smoothing (FDS).

In LDS, the basic idea is to smooth out differences in the original density distribution of labels among close labels. This is done by applying a kernel density function, such as a Gaussian kernel.

The basic idea of FDS is to smooth out outliers and noise in the feature values. This is done again with a kernel function such as a Gaussian kernel. Unfortunately, for a matter of time, FDS was not explored in this thesis.

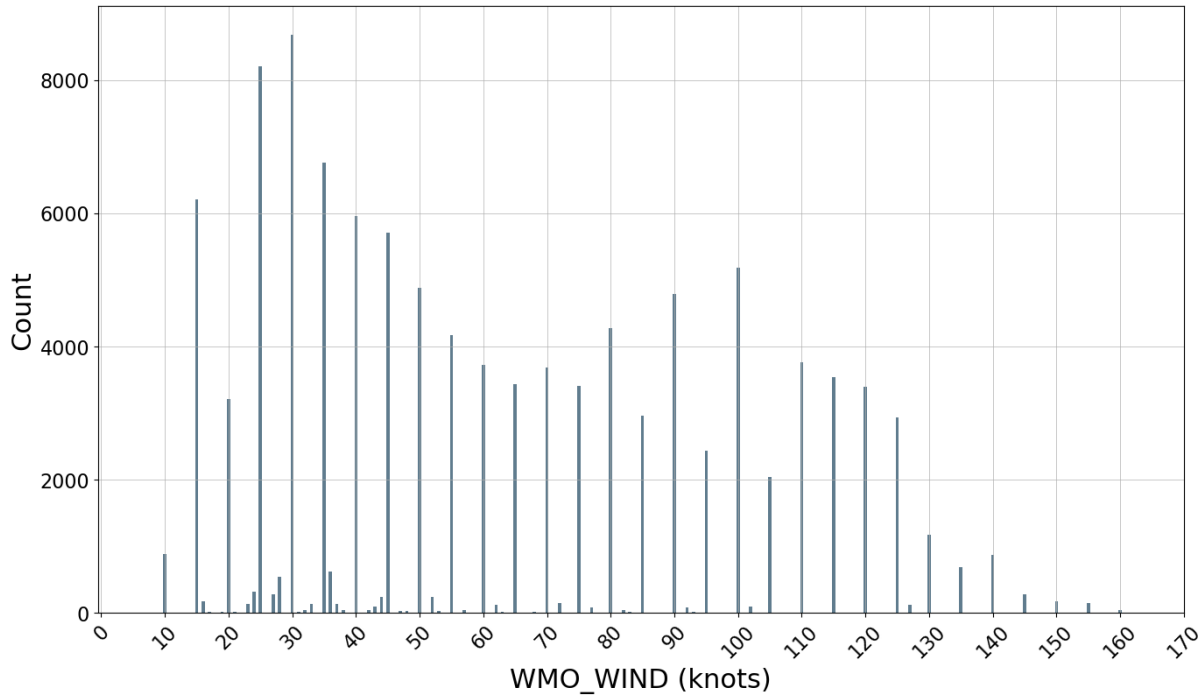


Figure 3.13: Empirical label distribution of targets.

Detailing LDS, a symmetric kernel is applied to the distribution of the labels, to smooth it and thus extend the information of more represented labels to nearby, potentially poorly represented, ones. The operation performed is the following:

$$\tilde{p}(y') = \int k(y, y')p(y)dy$$

Where:

- *p(y)* is the empirical number of occurrences of a label;
- $\tilde{p}$*(y')* is the label density, which is called effective because a representative of the real phenomena;

In the experiment done in this thesis, a Gaussian kernel with kernel size of 5 and a standard deviation of 3 was applied. The resulting effective distribution is reported below.

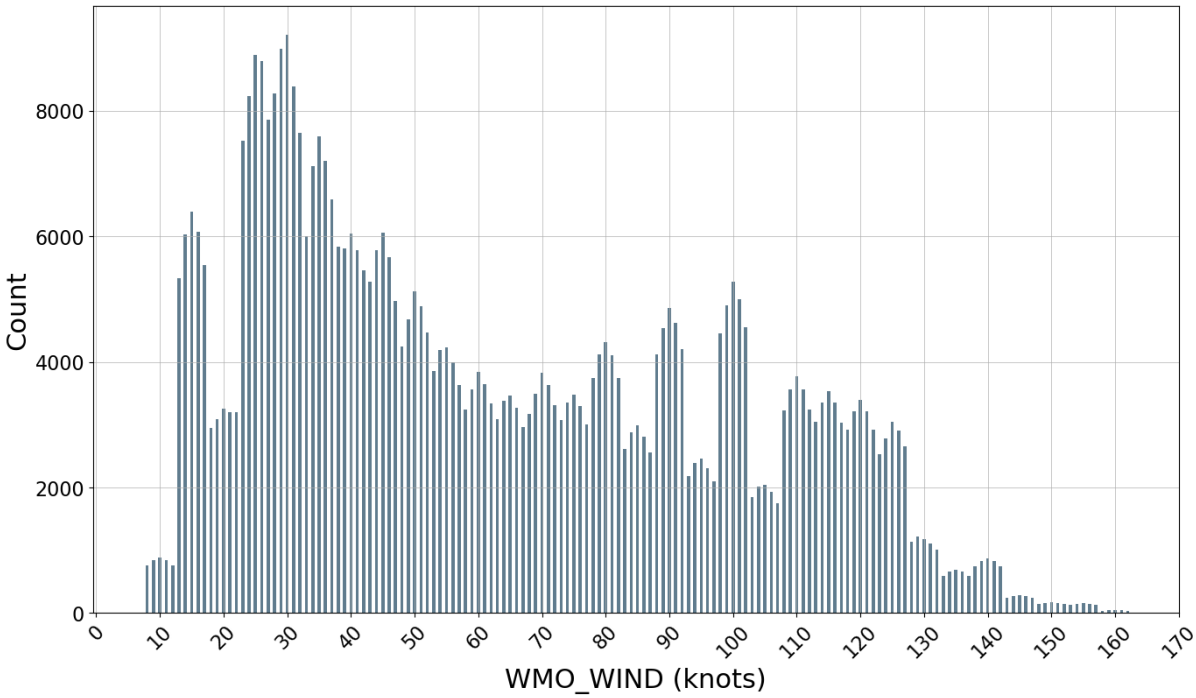


Figure 3.14: Effective label distribution of targets.

Yang et al. (2021) noted how the effective density function is much more correlated with the distribution of errors than the empirical one, meaning that the resulting distribution better represents the real imbalance.

The use of LDS is particularly appropriate for TCIE, as the empirical distribution of TC intensities in IBTrACS is little representative of the underlying phenomenon: the sparse distribution is more related to measurement (and potentially post-processing, as performed by IBTrACS to standardise data coming from different agencies) than of reality. This is immediately evident when looking at the empirical distribution (Fig. 3.13, in which the highest occurrences are for classes that are multiples of five. This is clearly related to approximations decided during the estimation of the maximum wind speed velocity. Thus, smoothing out the distribution using LDS could help the deep-CNN implemented to better capture the real (supposedly continuous) distribution of the labels.

All this reasoning is preparatory to the implementation of a Cost-Sensitive Loss Function (C-SLF). C-SLF, commonly used in ML classification tasks, are used to alleviate class imbalance problems, assigning different misclassification costs to different classes, based on their rarity. Yang et al. (2021), took inspiration from the loss function introduced by Lin et al. (2017): Focal Loss (FL). FL is a dynamically scaled cross-entropy loss, governed by a scaling factor decaying to zero and thus down-weighting the contribution to the loss

for well-classified samples, while increasing the weight for hard ones. Yang et al. (2021) adapted the Focal Loss for regression tasks, introducing Focal-R. They substituted the scaling factor by a continuous function which normalises L1 distance into an interval between 0 and 1. The Focal-R is constructed as follows:

$$Focal - R = \frac{1}{n}\sum_{i=1}^{n} w_i \sigma(|\beta e_i|)^\gamma = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{p_i}\sigma(|\beta e_i|)^\gamma$$

Where:

- e_i is the L1 distance for the i-th sample;
- $\sigma(\cdot)$ is the sigmoid function;
- β,γ are hyper-parameters;
- w_i are the weights, inversely proportional to target value occurrences;
- p_i is the smoothed label distribution value;

Focal-R can handle dataset imbalance more effectively, improving the performance of the model on the minority class. The use of this loss function is received as an effective tool which helps to overcome the intrinsic limits of loss functions such as MAE which has no power to contrast deep data imbalance. In the experiments, Focal-R hyper-parameters β,γ were set to 0.2 and 1 respectively, as proposed by Yang et al. (2021).

Focusing on the case study, here the weights are set as the inverse of the frequencies in effective label distribution. The distribution of w_i for the GridSat-B1 dataset is shown in Fig. 3.15.

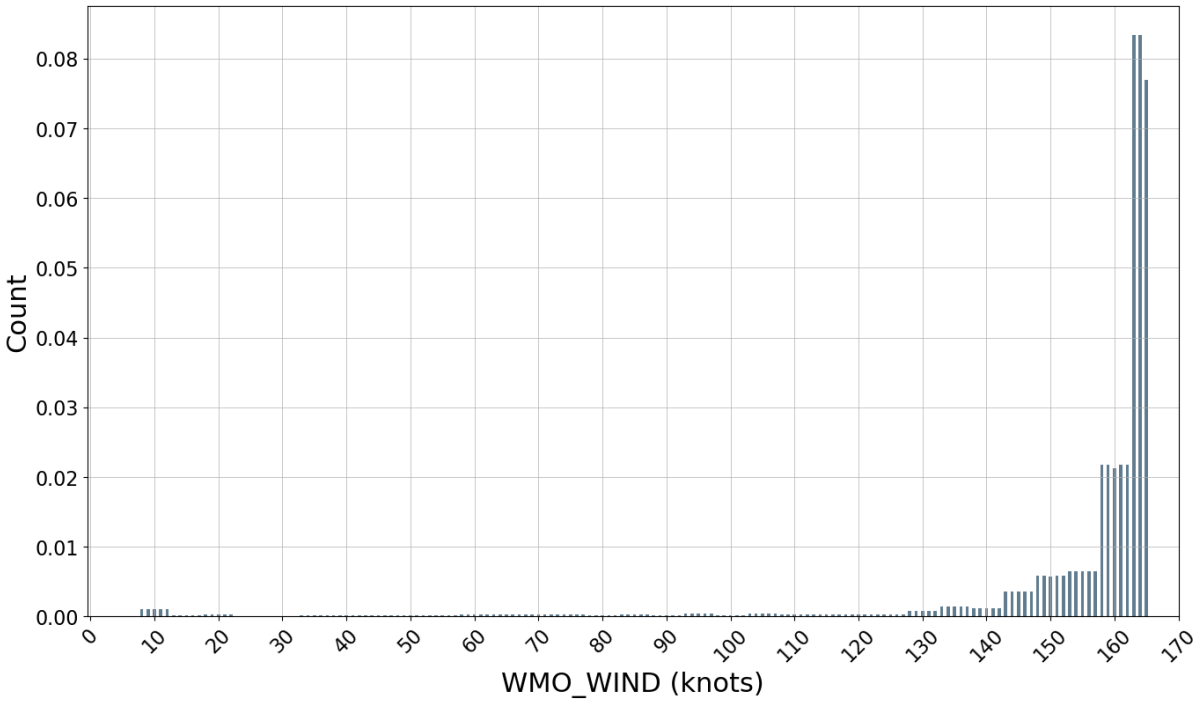


Figure 3.15: Weights (w_i) of the Focal-R.

3.4. Unboxing black-box models: heatmaps

Although CNNs have been shown to outperform traditional algorithms in most computer vision tasks, they are essentially "Black-box" (i.e., not interpretable) models, meaning their adoption by the scientific community at large has been met with some reluctance. Some authors proposed to use heatmaps in this field, as Lee et al. (2019). For TCIE, Zeiler and Fergus (2014) proposed to use heatmaps based on the sum of the activation maps in the last convolutional layer of a CNN. With resizing, they adapted heatmaps dimensions to match the size of the original image, to visualize what features the network focused on before making the final prediction. They observed how the eye region becomes more relevant as the intensity of the cyclones increases, reflecting the structure of weaker TCs which do not have well-defined eyes. A different approach to visualize heatmaps was proposed by Selvaraju et al. (2016), who developed the popular Gradient-based localisation Class Activation Mapping (Grad-CAM). The rationale behind it is to compute the gradient of the predicted class score with respect to the feature map corresponding to the last convolutional layer. Gradients are then weighted by the global average pooling of the same feature maps.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\delta y^c}{\delta A_{ij}^k} \tag{3.3}$$

The result obtained is then summed and a ReLU activation function is applied, obtaining the heatmap.

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k) \tag{3.4}$$

The algorithm used here is inspired by Grad-CAM, with small modifications:

1. Gradients of output are computed with respect to input for each channel of the output of the last feature map;
2. Gradient tensors pass through global average pooling;
3. The obtained pooled gradients are used to weigh the feature maps;
4. The final heatmap is obtained by summing the weighted feature maps.

In the following chapter, a few heatmaps are analyzed from the best model found. A preliminary classification of features is identified, highlighting the need for further analysis.

3.5. Implementation details

In this section, further details are provided, ensuring that the experiments can be reproduced subsequently. For the sake of readability, all details are provided as bullet points.

- Experiments leaned on a TensorFlow Framework, which supports CUDA;
- Networks were trained on NVIDIA GeForce GTX 1650 (4 GB) GPU;
- The batch size was set at 16;
- Top layers of networks were rebuilt, adding a Global Average Pooling (2D) to reduce spatial dimension to a single vector, Batch Normalization layer to normalise the activation functions, Dropout with a dropout rate fixed at 0.4 to regularise the model, and a final single output unit to make regression;
- The loss function chosen was Mean Absolute Error (except for the experiment involving Focal-R) and the optimizer was RMSprop;

- The starting learning rate was 0.0001, setting the decay at 0.1 every time the validation loss did not improve by at least 10^{-4} in 5 consecutive epochs;
- An early stopping was set to prevent overfitting and save time in training. The parameter monitored was validation loss, stopping the training after 6 epochs not reaching a minimum improvement of 10^{-4};
- The training/validation split of the input dataset was set respectively at 0.8/0.2;
- A normalisation of inputs was applied using batch mean and standard deviation;

4 | Results

4.1. Data augmentation tests

The following graphs report the results achieved by the different augmentation scenarios, in terms of mean absolute error on the test set, averaged over all predictions and grouped by the value of the labels. To improve their visual quality, the displayed curves were smoothed using a moving average to filter out outliers and reduce the noise.

Figure4.1 shows how different augmentation scenarios impact model performance differently. Several augmentation scenarios, among which the combination of horizontal and vertical flipping, random rotation and vertical random erasing, decreased the performance on intensities higher than 125 kt. Furthermore, all combinations decreased the performance for the most represented classes (i.e., weak TCs), which is an inevitable trade-off that happens when re-balancing the data (especially when undersampling the most represented classes).

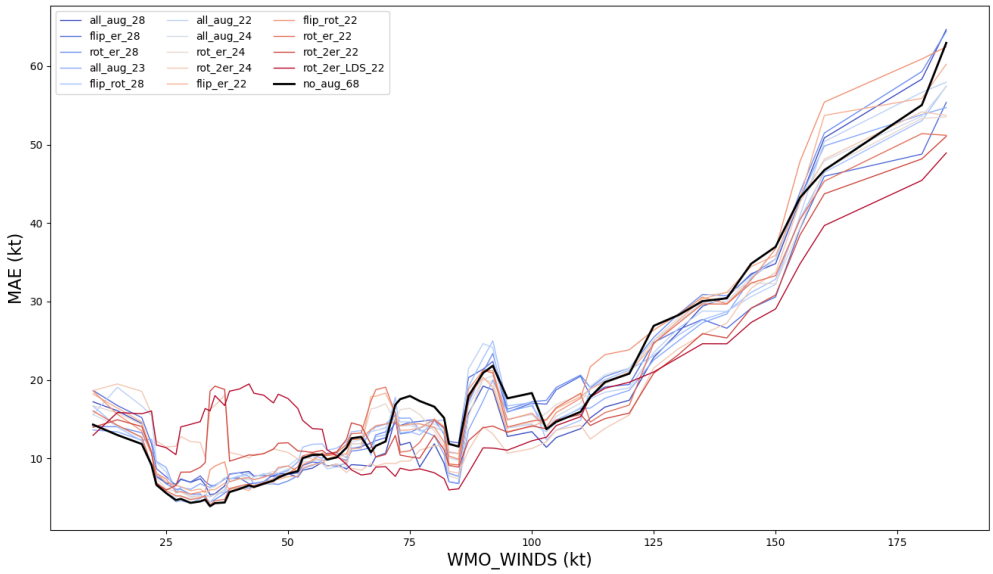


Figure 4.1: Comparison of MAE obtained in all the experiments.

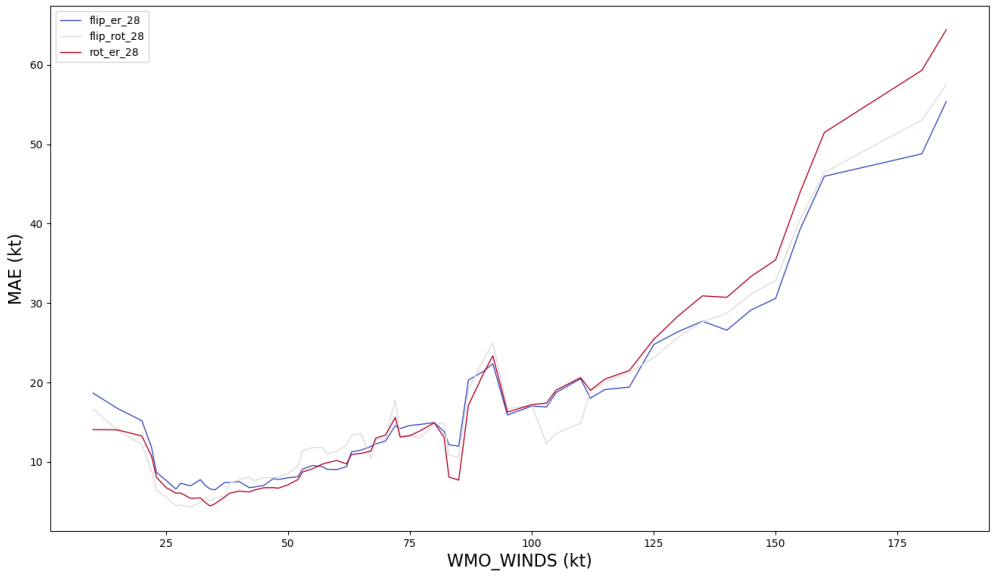


Figure 4.2: MAE obtained on test datasets in models trained on datasets with G = 0.28.

By isolating some of the augmentation scenarios (Fig.4.2) and fixing the Gini coefficient at G = 0.28, a comparison of MAE on most intense classes is made possible. In particular, flipping + random erasing and random rotation + random erasing scenarios show

divergent trends starting from intensities around 125 kt. The MAE measured in intensity classes higher than 100 kt increases by 4 kt between the first and the second scenario. Here, random rotation + random erasing are clearly disturbing the network in training, worsening its performance systematically in most intense classes. On the other hand, the combination of random erasing and flipping is slightly improving the MAE in prediction with respect to the business as usual.

The combination of flipping and random erasing was further explored, considering a distribution with a Gini-inspired coefficient of 0.22. Results of MAE obtained are reported in Fig.4.3. As can be seen, performances worsen systematically from intensity equal to 125 kt. A possible explanation could be linked to the fact that an image can be transformed three times by flipping without repetition: horizontal, vertical and combined horizontal and vertical transformations. As a consequence to a number of images, the same flipping transformation will be applied, and the only difference will be the width and location of erasing window. These decreasing performances can be linked to the fact that there is a lower amount of information introduced in the training by augmentation-derived images and the network is overfitting the dataset.

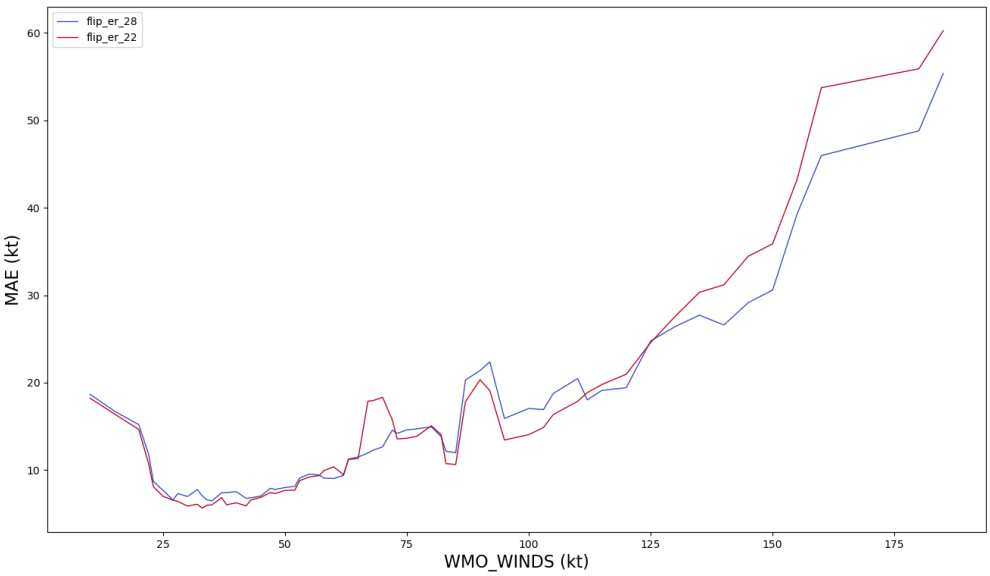


Figure 4.3: Comparison of MAE obtained by models trained on dataset augmented with random erasing and flipping with different Gini-inspired coefficient values.

Once noticed the different trends obtained varying the Gini-inspired coefficient, it was decided to explore deeply the effect of different distributions on the scenario random eras-

ing coupled with random rotation. First, the distribution with G = 0.22 was tested. The outcomes of this are reported in Fig.4.4. Surprisingly, the new dataset fed during training provided a model with increased performances measured on the test dataset. Performance in the most intense classes appears to improve consistently as the Gini coefficient decreases. This increase in performance is even more evident in classes with labels above 160 kt, which are not present in the training and validation sets. To conclude the experiment on the combination random erasing and random rotation, another training and testing was performed on an augmented dataset characterised by a G = 0.24. Performances obtained are overlapping G = 0.22 scenario for most of the intensity values, with a relevant deterioration between 110 and 125 kt, and a smaller deterioration for intensities higher than 160 kt. The decrease in performance in classes between 110 kt and 125 kt is hardly explained because it corresponds to classes characterised by the same number of images. The decrease in performance in intensity classes higher than 160 kt is better explained because the difference between G = 0.24 and G = 0.22 distributions are linked to classes with intensities higher than 120 kt, as reported in Tab. 4.1.

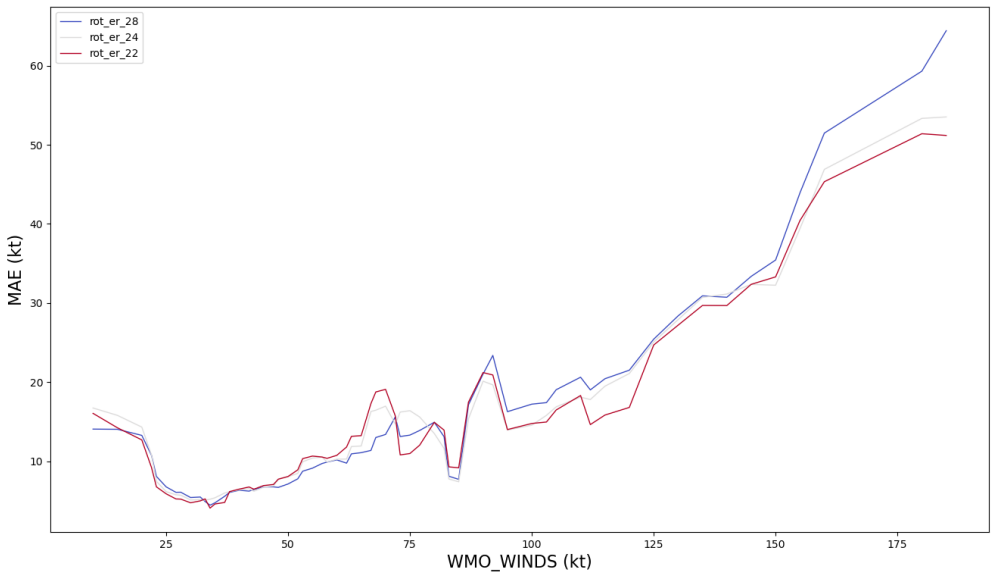


Figure 4.4: Comparison of MAE obtained by targets obtained by models trained on datasets augmented with random erasing and rotation with different Gini-inspired coefficient values.

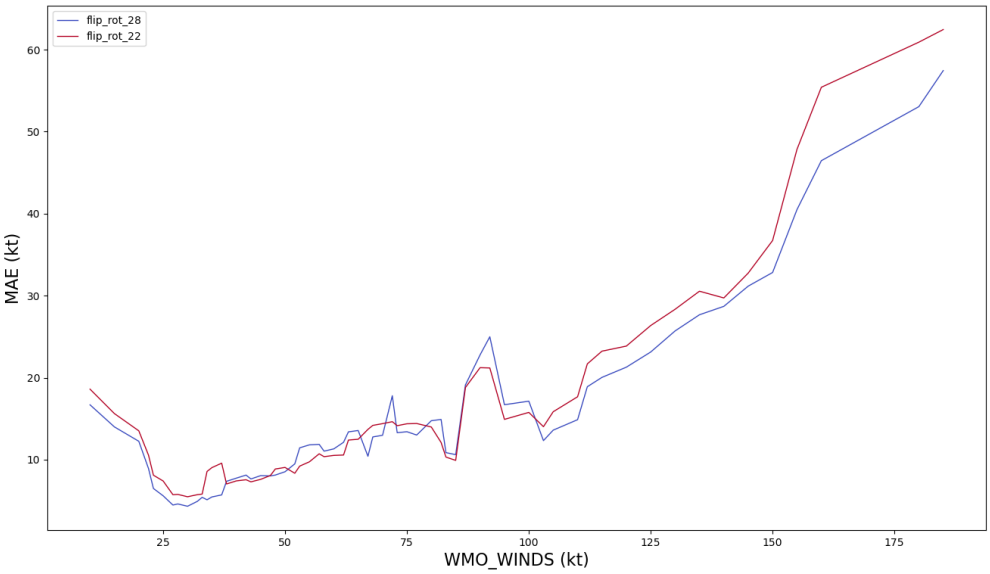


Figure 4.5: Comparison of MAE obtained by a model trained on datasets augmented with horizontal and vertical flipping and random rotation with different G values.

Focusing on the effects of different G values for augmented dataset in the business-as-usual scenario, the Gini coefficient was decreased to a value of 0.22. In the graph in Fig.4.5 it is clear that the performance deteriorated. In particular, a decrease in MAE equal to 4.03 kt is detected for intensities higher than 100 kt, with the highest loss in performance around 155 kt (the delta between the MAE of G = 0.22 scenario with respect to G = 0.28 is equal to -9 kt).

Comparing the results obtained for the scenario in which random rotation and random erasing (G = 0.22) with the business as usual in its best-performing distribution (G = 0.28), there is no clear improvement in performance, as can be seen in Fig.4.6, except for classes with intensities higher than 150 kt.

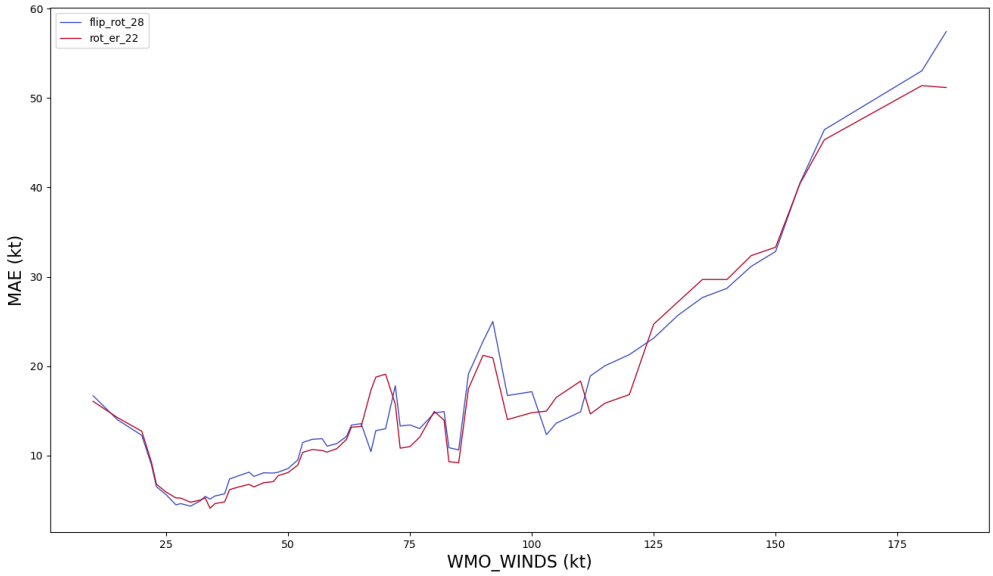


Figure 4.6: Comparison of MAE by models trained on datasets augmented with traditional techniques (flip_rot_28) and with rotation and vertical random erasing (rot_er_22).

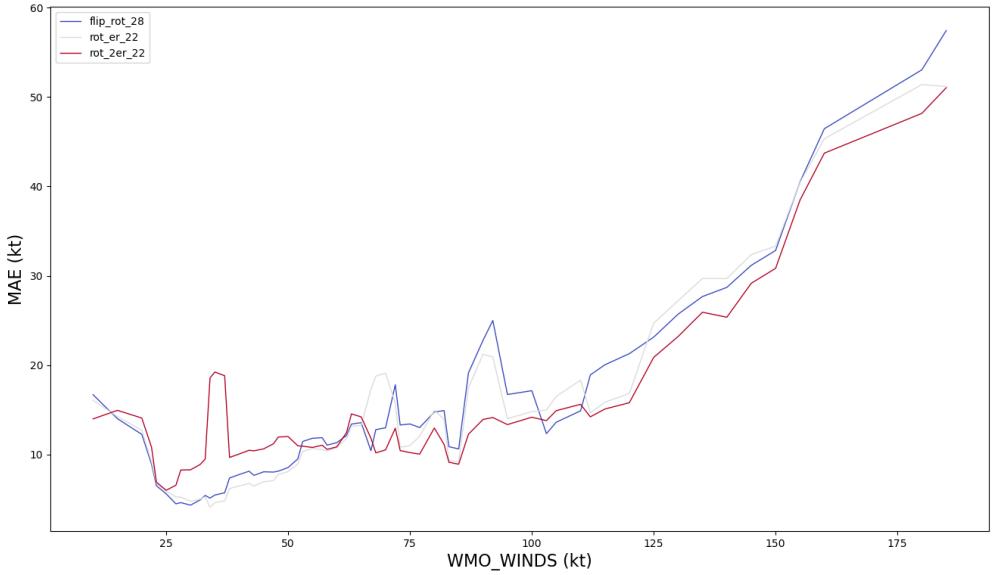


Figure 4.7: Comparison of MAE obtained in business as usual scenario (flip_rot_28); random rotation and random erasing scenario (rot_er_22); random rotation, and double random erasing scenario (rot_2er_22).

Having noticed the impact on TCs with maximum sustained winds higher than 150 kt, also horizontal random erasing was introduced. Unexpectedly, as reported in Fig.4.7, performances were systematically improved with respect to both business as usual and random rotation + random erasing. Having established this last scenario as the best-performing one, the effect of LDS and the new loss function were examined. In the graph in Fig.4.8 it is evident that the performance on classes greater than or equal to 125 kt is systematically better. Also, apart from a small interval around 110 kt, from 55 kt to the end of the dataset, the performance is consistently better than the business-as-usual scenario, with a considerable improvement, especially for the most intense classes. Entering the details, this last scenario outperforms business as usual, with relevant effects on MAE of classes with intensities higher than 100-110 kt (-3.99 kt) and the most relevant effect on class 150-160 kts (-9.25 kt).

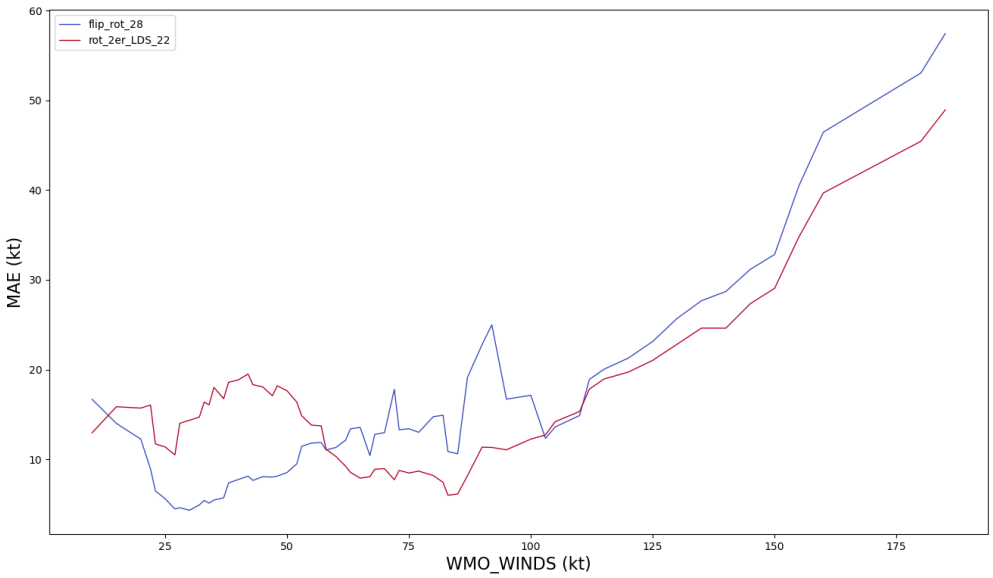


Figure 4.8: Comparison of MAE obtained models trained on datasets augmented with business as usual scenario (flip_rot_28); random rotation and double erasing and Focal-R as loss function (rot_2er_LDS_22).

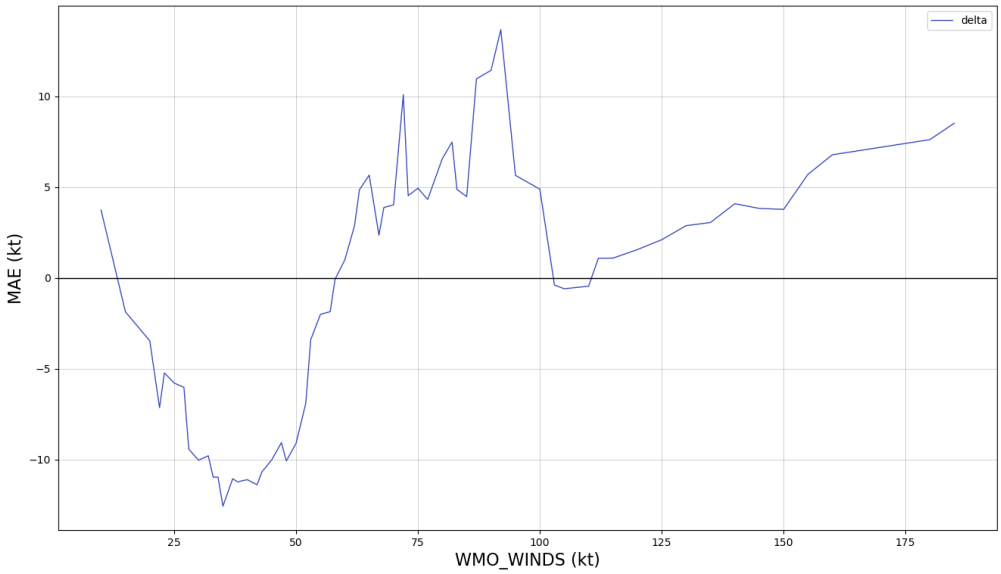


Figure 4.9: MAE improvement obtained in the scenario with random rotation, double random erasing and Focal-R as loss function with respect to business as usual scenario.

To conclude, Fig.4.9 shows a graph highlighting the MAE differences between business as usual scenario (random rotation and flipping with G = 0.28) and the best model obtained (trained on dataset augmented with random rotation, double random erasing with G = 0.22 and Focal-R as loss function). When the curve is above y = 0, there is a gain in performance. It can be seen how model performance improves for winds with MSW bigger than 55 kt, with a limited exception for TCs with intensities between 100 and 115 kt, where performances are comparable. Obtaining better performances in TC with intensities of MSW bigger than 160 kt is one of the most promising results since these TCs are not present in the training and validation data.

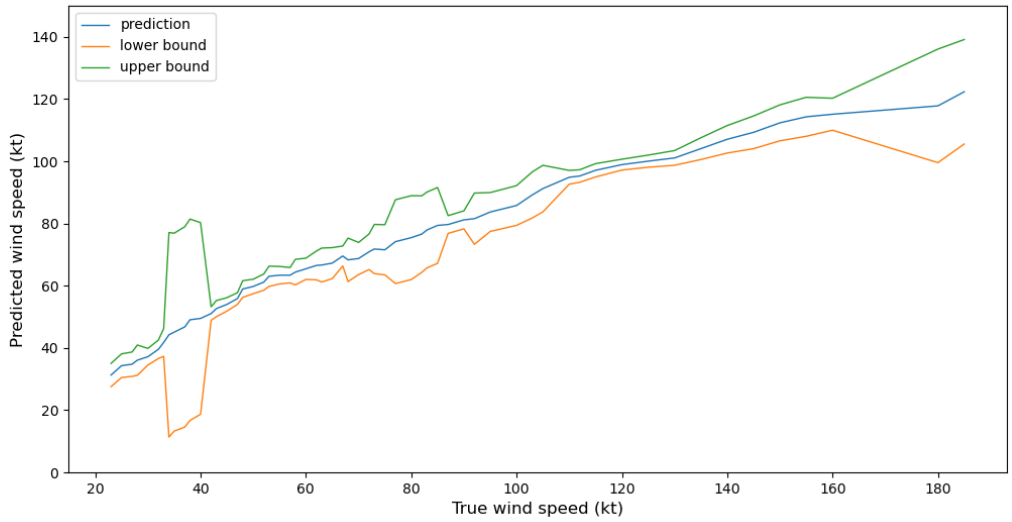


Figure 4.10: Prediction values and respective confidence interval for best-performing model (training dataset augmented of G = 0.22 by random rotation and double random erasing, processing targets with LDS and Focal-R as loss function).

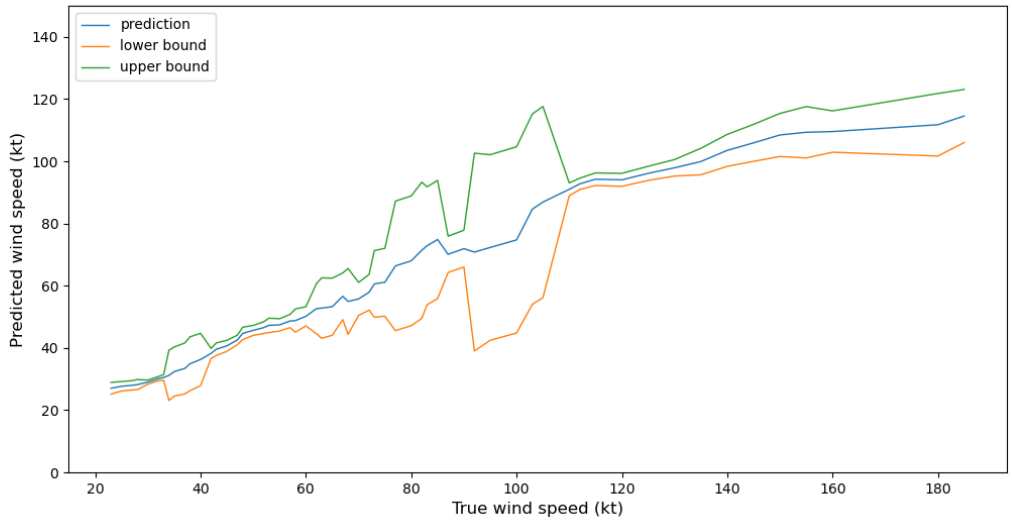


Figure 4.11: Prediction values and respective confidence interval for *traditional* model (training dataset augmented of G = 0.28 by random rotation and horizontal, vertical flipping, and MAE as the loss function).

Improving performance in these classes implies that the proposed augmentation strategy can make the model generalise even for previously unseen class intensities. The relevant

impact of the new approach proposed is captured when looking at the confidence intervals. The trend of predictions and the associated confidence interval at 95% are represented in Fig.4.10 and Fig.4.11. From this plot, one can see the effects of the two different augmentation scenarios on the final uncertainties associated with predictions. Especially in the range between 60 kt and 110 kt, there is a decrease in the range of values predicted by the network, making the network more reliable in estimation. Even wind speeds up to 160 kt show a detectable effect, regardless of their higher intensity values.

4.2. Preliminary results on heatmaps

Finally, a brief analysis of the heatmaps is reported below. Three pictures (Fig.4.12, Fig.4.13, and Fig.4.14) are represented, among the ones obtained from a subsample of predictions characterised by an absolute error smaller than 1 kt with respect to the target. Please note that the visible shapes in these heatmaps do not represent the TC intensity class. Indeed, these shapes are mainly found across all predicted intensities, indicating how the network routinely looks for the same types of patterns to make its estimation: border features (Fig.4.12) along the outskirts of the TC; spiral bands features (Fig.4.13); and cyclone eye features (Fig.4.14). The features selected along the border of the image bring into question the choice of crop size of the images; further experiments should investigate whether, by using a larger crop size, these border features assume more prominence, indicating relevant information at the edge of a TC. This is a reasonable assumption, as TC size is strongly linked to TC intensity.

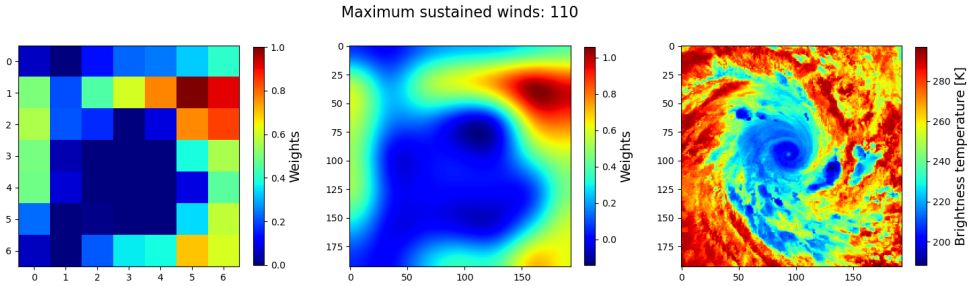


Figure 4.12: Example of heatmap highlighting the picture border.

After analyzing additional examples (not presented here) of similar heatmaps, it seems that the ring-shaped features emphasized in Fig.4.13 are more common when the cyclone's eye is not clearly defined. This can happen when the TC eye is occluded (as in Fig. 4.13),

but also when it is not fully formed, which occurs for weaker TC. As a result, in the absence of a distinct eye that would facilitate intensity estimation, the network adapts by searching for alternative features located further away from the center.

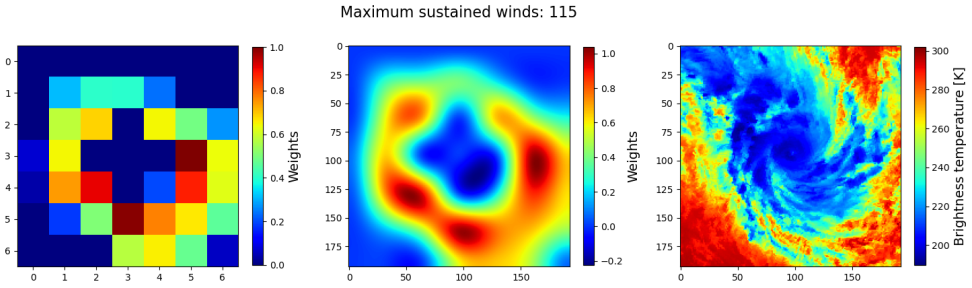


Figure 4.13: Example of heatmap highlighting a ring-like shape.

Finally, if an eye is visible (Fig. 4.14), the network identifies it as a key source of information to make its estimation. In a somewhat peculiar behaviour, the type of heatmap shown in Fig. 4.14 is occasionally found (not shown here) to not coincide with the eye but rather with some particular feature in the image. The flowchart of how the network interprets images, therefore, seems to be the following:

1. Look for a well-defined eye;
2. Look for well-defined features near the core of the TC;
3. Look for well-defined spiral bands within the main vortex;
4. Look for border features, possibly indicating TC size;

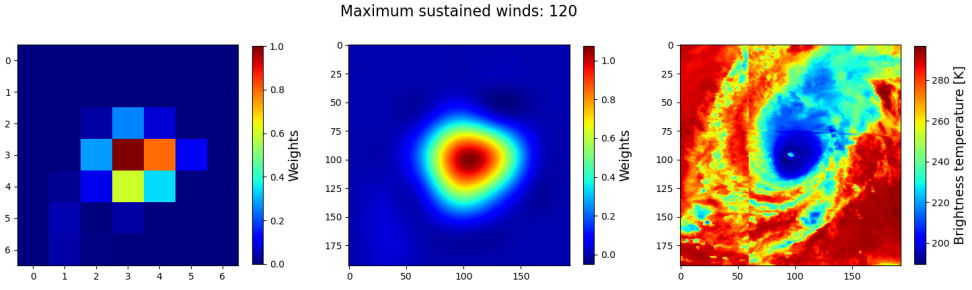


Figure 4.14: Example of heatmap highlighting a circular shape.

5 | Conclusions

In this thesis, we propose new approaches for alleviating deep imbalanced dataset.

With the use of transfer learning, it was possible to train and validate several networks representing the SoTA of image processing. Being pre-trained on ImageNet, deep CNNs were already optimised for feature extraction and only the rebuilt top fully connected layers were re-trained. ResNet, DenseNet, EfficientNet-B0 and EfficientNet-B2 were compared, and EfficientNet-B0 was found to perform better than the others in terms of validation loss.

A Gini-inspired coefficient was introduced, measuring the distance of the actual distribution of data samples from the uniform distribution. In particular, the closer the coefficient is to zero, the more similar the actual distribution is to the uniform one. Vice versa, the closer to one the coefficient, the more imbalanced the distribution. The coefficient proposed was particularly useful when comparing the performance of the CNN on different datasets: if the Gini-inspired coefficient remained the same, one can be certain that the network was trained on datasets with same distribution (and thus the same level of data imbalance).

Two innovations were introduced for TCIE regression with deep learning. First, random erasing, which was thoroughly tested in both its vertical and horizontal and vertical configuration. It was shown to be an effective technique, especially for the least represented and most intense classes, where an increased amount of augmentation is necessary to have an effect. Second, Label Distribution Smoothing and Focal-R loss function. Thanks to this approach it is possible to transfer knowledge from over-represented targets to adjacent ones. In order to exploit this information, Focal-R loss function was adapted for the specific task, making the network give higher weights to the least represented classes, and thus enhancing the learning from their targets.

Finally, heatmaps inspired by Grad-CAM were visualised and interpreted, to understand what the CNN looks for in images of TCs. Preliminary observations allowed to classify heatmaps into three identified classes. In the first class, the network seems to consider peripheral regions, which prompted suspicions that the chosen crop is cutting off part of

the image which could contain relevant information. In the second class, ring-shaped features are clearly distinguishable in the heatmap, highlighting a circular structure around the TC's eye. In the third class, the focus of the network is limited to a very small region characterised by a strong localised gradient of weights and ignoring the vast majority of the region; in some cases, this region corresponds exactly with the TC's eye.

This thesis could be the starting point for a number of future studies. First, the dataset used in this thesis is proposed for further studies concentrating on the issue of TCIE at a global scale; it is the largest available in literature related to ML applications for TCIE.

Second, in light of the fact that several heatmaps highlight as relevant the borders of pictures and considering that the MAE in training and validation loss is decreasing with increasing crop dimensions, further research should be done considering broader scales.

Third, because of time constraints, only few values of the Gini coefficient were considered here. However, when considering random erasing (single or double) and rotation, it was noticed that in generalisation with decreasing G, there was a general decrease of MAE. Further experiments should consider decreasing G further, identifying whether a minimum MAE could be achieved, in order to quantify the capability of this new technique to mimic new images and thus prevent overfitting.

Fourth, the promising results achieved with the introduction of LDS and Focal-R suggest that moving towards a pure regressive network may be the correct intuition to overcome deep imbalance issues in TCIE. Future research may discover whether different frequency-weighted loss functions can achieve better performance. Also, further discoveries may be found dealing with Focal-R hyper-parameter optimization. Once LDS have proven to be effective for the purpose, further developments may consider the implementation of Feature Distribution Smoothing (FDS), which demonstrated to be very effective in Yang et al. (2021) paper, and its coupling with LDS.

Further experiments must be undertaken to confirm the outcomes of this study. Specifically, K-fold cross validation will be performed to better interpret the differences in performances obtained by the various models tested. K-fold cross-validation is a common technique used in ML, performed as follows:

1. The dataset is divided into K equal parts;
2. The model is trained on K-1 folds and tested on the remaining fold;
3. By repeating this procedure K times, each fold is used exactly once as test set;
4. The results of the K experiments are averaged.

By following this procedure, the reliability of the results obtained is thus better assessed, and the validity of the findings better quantified.

Finally, further studies should focus on heatmaps. In this study we focused only on heatmaps belonging to well-predicted images. A better understanding of the relationship between heatmaps and prediction accuracy could help us to better interpret underlying mechanism followed by the CNN to estimate intensity.

Bibliography

R. B. Alley, T. Berntsen, N. L. Bindoff, Z. Chen, A. Chidthaisong, P. Friedlingstein, J. M. Gregory, G. C. Hegerl, M. Heimann, B. Hewitson, B. J. Hoskins, F. Joos, J. Jouzel, V. Kattsov, U. Lohmann, M. Manning, T. Matsuno, M. Molina, N. Nicholls, J. Overpeck, D. Qin, G. Raga, V. Ramaswamy, J. Ren, M. Rusticucci, S. Solomon, and R. Somerville. Summary for policymakers, 2007.

R. L. Bankert and P. M. Tag. An automated method to estimate tropical cyclone intensity using ssm/i imagery, 2002.

M. A. Bender, T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. M. Held. Modeled impact of anthropogenic warming on the frequency of intense atlantic hurricanes. *Science*, 327:454–458, 2010. ISSN 00368075. doi: 10.1126/science.1180568.

L. Bengtsson, K. I. Hodges, M. Esch, N. Keenlyside, L. Kornblueh, J. J. Luo, and T. Yamagata. How may tropical cyclones change in a warmer climate? *Tellus, Series A: Dynamic Meteorology and Oceanography*, 59 A:539–561, 2007. ISSN 02806495. doi: 10.1111/j.1600-0870.2007.00251.x.

M. Bister and K. A. Emanuel. Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, 240:233–240, 1998.

M. Bister and K. A. Emanuel. Low frequency variability of tropical cyclone potential intensity 2. climatology for 1982-1995. *Journal of Geophysical Research: Atmospheres*, 107, 2002.

C. M. Brandon, J. D. Woodruff, D. P. Lane, and J. P. Donnelly. Tropical cyclone wind speed constraints from resultant storm surge deposition: A 2500 year reconstruction of hurricane activity from st. marks, fl. *Geochemistry, Geophysics, Geosystems*, 14: 2993–3008, 2013. ISSN 15252027. doi: 10.1002/ggge.20217.

M. Buda, A. Maki, and M. A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. 10 2017. doi: 10.1016/j.neunet.2018.

07.011. URL http://arxiv.org/abs/1710.05381http://dx.doi.org/10.1016/j.neunet.2018.07.011.

J. Callaghan and S. B. Power. Variability and decline in the number of severe tropical cyclones making land-fall over eastern australia since the late nineteenth century. *Climate Dynamics*, 37:647–662, 2011. ISSN 09307575. doi: 10.1007/s00382-010-0883-2.

S. J. Camargo, A. H. Sobel, A. G. Barnston, and P. J. Klotzbach. The influence of natural climate variability on tropical cyclones, and seasonal forecasts of tropical cyclone activity. *Global Perspectives On Tropical Cyclones: From Science To Mitigation*, pages 325–360, 2010. doi: 10.1142/9789814293488_0011.

J. C. Chan and M. Xu. Inter-annual and inter-decadal variations of landfalling tropical cyclones in east asia. part i: time series analysis johnny. *International Journal of Climatology*, 2029:1285–1293, 2008. doi: 10.1002/joc1782.

F. Chauvin, J.-F. Royer, and M. Deque. Response of hurricane-type vortices to global warming as simulated by arpege-climat at high resolution. *Climate Dynamics*, 27:377–399, 2006.

B. Chen, B. F. Chen, and H. T. Lin. Rotation-blended cnns on a new open dataset for tropical cyclone image-to-intensity regression. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 90–99, 2018. doi: 10.1145/3219819.3219926.

R. Chen, W. Zhang, and X. Wang. Machine learning in tropical cyclone forecast modeling: A review, 7 2020. ISSN 20734433.

S. S. Chen and W. M. Frank. A numerical study of the genesis of extratropical convective mesovortices. part i: Evolution and dynamics. *Journal of atmospheric sciences*, 50: 2401–2426, 1992.

J. H. Christensen, K. K. Kanikicharla, E. Aldrian, S. I. An, I. F. A. Cavalcanti, M. de Castro, W. Dong, P. Goswami, A. Hall, J. K. Kanyanga, A. Kitoh, J. Kossin, N. C. Lau, J. Renwick, D. B. Stephenson, S. P. Xie, T. Zhou, L. Abraham, T. Ambrizzi, B. Anderson, O. Arakawa, R. Arritt, M. Baldwin, M. Barlow, D. Barriopedro, M. Biasutti, S. Biner, D. Bromwich, J. Brown, W. Cai, L. V. Carvalho, P. Chang, X. Chen, J. Choi, O. B. Christensen, C. Deser, K. Emanuel, H. Endo, D. B. Enfield, A. Evan, A. Giannini, N. Gillett, A. Hariharasubramanian, P. Huang, J. Jones, A. Karumuri, J. Katzfey, E. Kjellström, J. Knight, T. Knutson, A. Kulkarni, K. R. Kundeti, W. K. Lau, G. Lenderink, C. Lennard, L. yung Ruby Leung, R. Lin, T. Losada, N. C. Mackellar, V. Magaña, G. Marshall, L. Mearns, G. Meehl, C. Menéndez, H. Murakami, M. J.

Nath, J. D. Neelin, G. J. van Oldenborgh, M. Olesen, J. Polcher, Y. Qian, S. Ray, K. D. Reich, B. R. de Fonseca, P. Ruti, J. Screen, J. Sedláček, S. Solman, M. Stendel, S. Stevenson, I. Takayabu, J. Turner, C. Ummenhofer, K. Walsh, B. Wang, C. Wang, I. Watterson, M. Widlansky, A. Wittenberg, T. Woollings, S. W. Yeh, C. Zhang, L. Zhang, X. Zheng, and L. Zou. Climate phenomena and their relevance for future regional climate change, 1 2013. Publisher Copyright: Intergovernmental Panel on Climate Change 2014.

J. S. Combinido, J. R. Mendoza, and J. Aborot. *A Convolutional Neural Network Approach for Estimating Tropical Cyclone Intensity Using Satellite-based Infrared Images; A Convolutional Neural Network Approach for Estimating Tropical Cyclone Intensity Using Satellite-based Infrared Images*. 2018. ISBN 9781538637883. doi: 10.0/Linux-x86_64.

M. DeMaria, J. A. Knaff, and B. H. Connell. A tropical cyclone genesis parameter for the tropical atlantic. *Weather and forecasting*, 16:219–233, 2001.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. pages 248–255. IEEE, 6 2009. ISBN 978-1-4244-3992-8. doi: 10.1109/CVPR.2009.5206848.

J. F. Dutton, C. J. Poulsen, and J. L. Evans. The effect of global climate change on the regions of tropical convection in csm1. *Geophysical Research Letters*, 27:3049–3052, 2000.

V. F. Dvorak. Tropical cyclone intensity analysis and forecasting from satellite imagery. *Monthly Weather Review*, 103:420–430, 1975. ISSN 0027-0644. doi: 10.1175/1520-0493(1975)103<0420:tciaaf>2.0.co;2.

V. F. Dvorak. *Tropical cyclone intensity analysis using satellite data*, volume 1. 1984. ISBN 3663537137.

V. F. Dvorak. Tropical clouds and cloud systems observed in satellite imagery: Tropical cyclones. *Workbook vol. 2*, 1995.

J. B. Elsner, J. P. Kossin, and T. H. Jagger. The increasing intensity of the strongest tropical cyclones. *Nature*, 455:92–95, 2008. ISSN 14764687. doi: 10.1038/nature07234.

K. Emanuel. A statistical analysis of tropical cyclone intensity. *Monthly Weather Review*, 128:1139–1152, 2000.

K. Emanuel. Tropical cyclones. *Annual Review of Earth and Planetary Sciences*, 31: 75–104, 2003. ISSN 00846597. doi: 10.1146/annurev.earth.31.100901.141259.

K. Emanuel. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436:686–688, 2005. ISSN 00280836. doi: 10.1038/nature03906.

K. Emanuel. Environmental factors affecting tropical cyclone power dissipation. *Journal of Climate*, 20:5497–5509, 2007. ISSN 08948755. doi: 10.1175/2007JCLI1571.1.

K. Emanuel, R. Sundararajan, and J. Williams. Hurricanes and global warming: Results from downscaling ipcc ar4 simulations. *Bulletin of the American Meteorological Society*, 89:347–368, 2008.

K. A. Emanuel. An air-sea interaction theory for tropical cyclones. part i: Steady-state maintenance. *Journal of atmospheric sciences1*, 43:585–605, 1986. doi: https://doi.org/10.1175/1520-0469(1986)043<0585:AASITF>2.0.CO;2.

K. A. Emanuel. The dependence of hurricane intensity on climate. *Nature*, 326:483–485, 1987. doi: https://doi.org/10.1038/326483a0.

K. A. Emanuel. The maximum intensity of hurricanes. *Journal of atmospheric sciences*, 45:1143–1155, 1988.

K. A. Emanuel. Tropical cyclone activity downscaled from noaa-cires reanalysis, 1908-1958. *Journal of Advances in Modeling Earth Systems*, 2, 2010.

F. A. Farris. The gini index and measures of inequality. *American Mathematical Monthly*, 117:851–864, 12 2010. ISSN 00029890. doi: 10.4169/000298910X523344.

R. W. Fett. Some characteristics of the formative stage of typhoon development: a satellite study. page 10. U.S. Weather Bureau, 1964.

N. Gillett, P. Stott, and B. Santer. Attribution of cyclogenesis region sea surface temperature change to anthropogenic influence. *Geophysical Research Letters*, 35, 2008.

S. B. Goldenberg and L. J. Shapiro. Physical mechanisms for the association of el ni$\tilde{\{n\}}$o and west african rainfall with atlantic major hurricane activity. *Journal of Climate*, 9: 1169–1187, 1996.

S. B. Goldenberg, C. W. Landsea, A. Mestas-Nunez, and W. M. Gray. The recent increase in atlantic hurricane activity: Causes and implications. *Science*, 293:474–479, 2001.

W. M. Gray. Global view of the origin of tropical disturbances and storms. *Monthly Weather Review*, 96:669–700, 1968.

W. M. Gray. Cumulus convection and larger scale circulations i. broadscale and mesoscale considerations. *Monthly Weather Review*, 101:839–855, 1973.

W. M. Gray. Tropical cyclone genesis, 1975.

W. M. Gray. Tropical cyclone intensity determination through upper-tropospheric aircraft reconnaissance. *Bulletin of the American Meteorological Society*, 60:1069–1074, 1979.

W. M. Gray. Atlantic seasonal hurricane frequency. part i : El ni{̃n}o and 30-mb quasi-bienniel oscillation influences. *Monthly Weather Review*, 112:1649–1668, 1984. doi: https://doi.org/10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2.

W. M. Gray. The formation of tropical cyclones. *Meteorology and Atmospheric Physics*, 67:37–69, 1998. ISSN 01777971. doi: 10.1007/BF01277501.

S. Gualdi, E. Scoccimarro, and A. Navarra. Changes in tropical cyclone activity due to global warming: Results from a high-resolution coupled general circulation model. *Journal of Climate*, 21:5204–5228, 2008.

J. Haig, J. Nott, and G. J. Reichart. Australian tropical cyclone activity lower than at any time over the past 550-1,500 years. *Nature*, 505:667–670, 2014. ISSN 00280836. doi: 10.1038/nature12882. URL `http://dx.doi.org/10.1038/nature12882`.

H. He and E. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21:1263–1284, 9 2009. ISSN 1041-4347. doi: 10.1109/TKDE.2008.239.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. 12 2015.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. pages 770–778. IEEE, 6 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90.

C. G. Hebert, R. A. Weinzapfel, and M. A. Chambers. *Hurricane Severity Index: A new way of estimating a tropical cyclone's destructive potential.* Inc., Houston, TX, 2010.

I. M. Held and B. J. Soden. Robust responses of the hydrological cycle to global warming. *Journal of Climate*, 19:5686–5699, 2006.

A. Henderson-Sellers, H. Zhang, G. Berz, K. Emanuel, W. Gray, C. Landsea, G. Holland, J. Lighthill, S. L. Shieh, P. Webster, and K. McGuffie. Tropical cyclones and global climate change: A post-ipcc assessment. *Bulletin of the American Meteorological Society*, 79:19–38, 1998. ISSN 00030007. doi: 10.1175/1520-0477(1998)079<0019:TCAGCC>2.0.CO;2.

G. Holland and C. L. Bruyère. Recent intense hurricane response to global climate change. *Climate Dynamics*, 42:617–627, 2014. ISSN 09307575. doi: 10.1007/s00382-013-1713-0.

G. J. Holland. The maximum potential intensity of tropical cyclones. *Journal of atmospheric sciences*, 54:2519–2541, 1997.

G. J. Holland and P. J. Webster. Heightened tropical cyclone activity in the north atlantic: Natural variability or climate trend? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365:2695–2716, 2007. ISSN 1364503X. doi: 10.1098/rsta.2007.2083.

G. J. Holland, J. M. Done, R. Douglas, G. R. Saville, and M. Ge. Global tropical cyclone damage potential. *Springer*, pages 23–42, 2019. doi: 10.1007/978-3-030-02402-4_2.

G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. 8 2016.

IPCC. Technical summary, 5 2022.

N. Japkowicz, C. E. Myers, and M. A. Gluck. A novelty detection approach to classification. 1995.

N. C. Johnson and S.-P. Xie. Changes in the sea surface temperature threshold for tropical convection. *Nature Geoscience*, 3:842–845, 2010.

R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie. Cmorph: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, 5:487–503, 6 2004. ISSN 1525-755X. doi: 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

L. Kantha. Time to replace the saffir-simpson hurricane scale? *Eos, Transactions American Geophysical Union*, 87:3–6, 2006.

D. J. Karoly and Q. Wu. Detection of regional surface temperature trends. *Journal of Climate*, 18:4337–4343, 2005.

P. J. Klotzbach. The madden–julian oscillation's impacts on worldwide tropical cyclone activity. *Journal of Climate*, 27:2317–2330, 3 2014. ISSN 0894-8755. doi: 10.1175/JCLI-D-13-00483.1.

P. P. Klotzbach, S. G. Bowen, R. G. R. PielKe, and M. Bell. Continental u.s. hurricane landfall frequency and associated damage: Observations and future risks. *Bulletin of the American Meteorological Society*, 99:1359–1376, 2018. ISSN 00030007. doi: 10.1175/BAMS-D-17-0184.1.

K. R. Knapp. Scientific data stewardship of international satellite cloud climatology

project b1 global geostationary observations. *Journal of Applied Remote Sensing*, 2: 023548, 11 2008. ISSN 1931-3195. doi: 10.1117/1.3043461.

K. R. Knapp and J. P. Kossin. New global tropical cyclone data set from isccp b1 geostationary satellite observations. *Journal of Applied Remote Sensing*, 1:013505, 2007. ISSN 1931-3195. doi: 10.1117/1.2712816.

K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann. The international best track archive for climate stewardship (ibtracs). *Bulletin of the American Meteorological Society*, 91:363–376, 2010. ISSN 00030007. doi: 10.1175/2009BAMS2755.1.

K. R. Knapp, S. Ansari, C. L. Bain, M. A. Bourassa, M. J. Dickinson, C. Funk, C. N. Helms, C. C. Hennon, C. D. Holmes, G. J. Huffman, J. P. Kossin, H.-T. Lee, A. Loew, and G. Magnusdottir. Globally gridded satellite observations for climate studies. *Bulletin of the American Meteorological Society*, 92:893–907, 7 2011. ISSN 0003-0007. doi: 10.1175/2011BAMS3039.1.

T. R. Knutson and R. E. Tuleya. Increased hurricane intensities with co2-induced warming as simulated using the gfdl hurricane prediction system. *Climate Dynamics*, 15:503–519, 1999.

T. R. Knutson and R. E. Tuleya. Impact of co 2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *Journal of Climate*, 17, 2004.

T. R. Knutson, R. E. Tuleya, W. Shen, and I. Ginis. Impact of co 2-induced warming on hurricane intensities as simulated in a hurricane model with ocean coupling. *Journal of Climate*, 14:2458–2468, 2001.

T. R. Knutson, T. Delworth, K. Dixon, I. Held, J. Lu, V. Ramaswamy, M. Schwarzkopf, G. Stenchikov, and R. Stouffer. Assessment of twentieth-century regional surface temperature trends using the gfdl cm2 coupled models. *Journal of Climate*, 19:1624–1651, 2006.

T. R. Knutson, J. J. Sirutis, S. T. Garner, I. M. Held, and R. E. Tuleya. Simulation of the recent multidecadal increase of atlantic hurricane activity using an 18-km-grid regional model. *Bulletin of the American Meteorological Society*, 88:1549–1565, 2007.

T. R. Knutson, J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held. Simulated reduction in atlantic hurricane frequency under twenty-first-century warming conditions. *Nature Geoscience*, 1:359–364, 2008.

T. R. Knutson, J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. K. Srivastava, and M. Sugi. Tropical cyclones and climate change. *Nature Geoscience*, 3:157–163, 2010. ISSN 17520894. doi: 10.1038/ngeo779.

R. L. Korty, S. J. Camargo, and J. Galewsky. Tropical cyclone genesis factors in simulations of the last glacial maximum. *Journal of Climate*, 25:4348–4365, 2012a. ISSN 08948755. doi: 10.1175/JCLI-D-11-00517.1.

R. L. Korty, S. J. Camargo, and J. Galewsky. Variations in tropical cyclone genesis factors in simulations of the holocene epoch. *Journal of Climate*, 25:8196–8211, 2012b. ISSN 08948755. doi: 10.1175/JCLI-D-12-00033.1.

J. P. Kossin and C. S. Velden. A pronounced bias in tropical cyclone minimum sea level pressure estimation based on the dvorak technique. *Monthly weather review*, 132: 165–173, 2004.

J. P. Kossin and D. J. Vimont. A more general framework for understanding atlantic hurricane variability and trends. *Bulletin of the American Meteorological Society*, 88: 1767–1782, 2007.

J. P. Kossin, K. R. Knapp, D. J. Vimont, R. J. Murnane, and B. A. Harper. A globally consistent reanalysis of hurricane variability and trends. *Geophysical Research Letters*, 34, 2007. ISSN 00948276. doi: 10.1029/2006GL028836.

J. P. Kossin, K. A. Emanuel, and G. A. Vecchi. The poleward migration of the location of tropical cyclone maximum intensity. *Nature*, 509:349–352, 2014. ISSN 14764687. doi: 10.1038/nature13278.

M. E. Kozar and V. Misra. Statistical prediction of integrated kinetic energy in north atlantic tropical cyclones. *Monthly Weather Review*, 142:4646–4657, 2014.

B. Krawczyk. Learning from imbalanced data: open challenges and future directions, 11 2016. ISSN 21926360.

A. Krizhevsky. Learning multiple layers of features from tiny images. 2009.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. 2012.

H. Kubota and J. C. Chan. Interdecadal variability of tropical cyclone landfall in the philippines from 1902 to 2005. *Geophysical Research Letters*, 36:2–5, 2009. ISSN 00948276. doi: 10.1029/2009GL038108.

A. Kulkarni, R. Bankert, and M. Hadjimichael. Tropical cyclone intensity estimation using

neural networks. ASPRS, 4 . URL https://www.researchgate.net/publication/268417973.

K. Kunkel, P. Bromirski, H. Brooks, T. Cavazos, A. Douglas, D. Easterling, K. Emanuel, P. Groisman, G. Holland, T. Knutson, J. Kossin, P. Komar, D. Levinson, and R. Smith. Observed changes in weather and climate extremes., 2008.

C. W. Landsea. Counting atlantic tropical cyclones back to 1900. *Eos*, 88:2005–2008, 2007. ISSN 00963941. doi: 10.1029/2007EO180001.

C. W. Landsea and W. M. Gray. The strong association between western sahelian monsoon rainfall and intense atlantic hurricanes. *Journal of Climate1*, 5:435–453, 1992.

J. Lee, J. Im, D.-H. Cha, H. Park, and S. Sim. Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data. *Remote Sensing*, 12:108, 12 2019. ISSN 2072-4292. doi: 10.3390/rs12010108. URL https://www.mdpi.com/2072-4292/12/1/108.

T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft coco: Common objects in context. 5 2014.

T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. 8 2017.

R. E. Lopez. Cumulus convection and larger scale circulations ii. cumulus and mesoscale interactions. *Monthly Weather Review*, 101:856–870, 1973.

M. O. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9:209–219, 1905.

M. E. Mann, K. A. Emanuel, G. J. Holland, and P. J. Webster. Atlantic tropical cyclones revisited. *Eos*, 88:349–350, 2007. ISSN 00963941. doi: 10.1029/2007EO360002.

E. J. L. Margaret A. Jorgensen, David P.; Zipser. Vertical motions in intense hurricanes. *American Meteorological Society*, pages 839–856, 1985. ISSN 17549469.

J. L. McBride. Observational analysis of tropical cyclone formation. part i: Basic description of data sets. *Journal of atmospheric sciences*, 38:1117–1131, 1981.

J. L. McBride and R. Zehr. Observational analysis of tropical cyclone formation. part ii: comparison of non-developing versus developing systems, 1981.

G. Meehl, T. Stocker, W. Collins, P. Friedlingstein, A. Gaye, G. Gregory, A. Kitoh, R. Knutti, J. Murphy, A. Noda, S. Raper, I. Watterson, A. Weaver, and Z. Zhao. Global

climate projections. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 2007.

H. Murakami, B. Wang, and A. Kitoh. Future change of western north pacific typhoons: Projections by a 20-km-mesh global atmospheric model. *Journal of Climate*, 24:1154–1169, 2011.

R. J. Murnane and J. B. Elsner. Maximum wind speeds and us hurricane losses. *Geophysical Research Letters*, 39, 2012.

J. Namias. Tropical cyclones related to the atmosphere's general circulation. *Trans. NY Acad. Sci., Ser. II*, 17:346–349, 1955.

E. Neumayer and F. Barthel. Normalizing economic loss from natural disasters: A global analysis. *Global Environmental Change*, 21:13–24, 2011.

NOAA. https://www.nhc.noaa.gov/aboutsshws.php.

W. D. Nordhaus. The economics of hurricanes and implications of global warming. *Climate Change Economics*, 1:1–20, 2010. ISSN 20100086. doi: 10.1142/S2010007810000054.

T. L. Olander and C. S. Velden. The advanced dvorak technique: Continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery. *Weather and forecasting*, 22:287–298, 2007.

T. L. Olander, C. Velden, and J. Kossin. The advanced objective dvorak technique (aodt)-latest upgrades and future directions. pages 294–295, 2004.

K. Oouchi, J. Yoshimura, H. Yoshimura, R. Mizuta, S. Kusunoki, and A. Noda. Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: Frequency and wind intensity analyses. *Journal of the Meteorological Society of Japan. Ser. II*, 84:259–276, 2006.

R. A. Pielke and C. N. Landsea. La ni$\tilde{\{n\}}$a, el ni$\tilde{\{n\}}$o, and atlantic hurricane damages in the united states. *Bulletin of the American Meteorological Society*, 80:2027–2033, 1999. ISSN 00030007. doi: 10.1175/1520-0477(1999)080<2027:LNAENO>2.0.CO;2.

R. A. Pielke and C. W. Landsea. Normalized hurricane damages in the united states: 1925-95. *Weather and forecasting*, 13:621–631, 1998.

R. A. Pielke, J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin. Normalized hurricane damage in the united states: 1900–2005. *Natural Hazards Review*, 9:29–42, 2008. ISSN 1527-6988. doi: 10.1061/(asce)1527-6988(2008)9:1(29).

M. F. Pineros, E. A. Ritchie, and J. S. Tyo. Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data. *IEEE Transactions on Geoscience and Remote Sensing*, 46:3574–3580, 11 2008. ISSN 0196-2892. doi: 10.1109/TGRS.2008.2000819.

M. F. Pineros, E. A. Ritchie, and J. S. Tyo. Detecting tropical cyclone genesis from remotely sensed infrared image data. *IEEE Geoscience and Remote Sensing Letters*, 7: 826–830, 10 2010. ISSN 1545-598X. doi: 10.1109/LGRS.2010.2048694.

M. F. Pineros, E. A. Ritchie, and J. S. Tyo. Estimating tropical cyclone intensity from infrared image data. *Weather and Forecasting*, 26:690–698, 10 2011. ISSN 0882-8156. doi: 10.1175/WAF-D-10-05062.1.

M. D. Powell and T. A. Reinhold. Tropical cyclone destructive potential by integrated kinetic energy. *Bulletin of the American Meteorological Society*, 88:513–526, 2007.

R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil. Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Transactions on Image Processing*, 27:692–702, 2018. ISSN 10577149. doi: 10.1109/TIP.2017.2766358.

H. A. Ramsay and A. H. Sobel. Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *Journal of Climate*, 24:183–193, 2011.

D. A. Randall, R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, Jayaraman, and others Srinivasan. Climate models and their evaluation. *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)*, pages 589–562, 2007.

M. Re. Natcatservice database. *Munich: Munich Reinsurance Company, Geo Risks Research*, 2018.

E. A. Ritchie, K. M. Wood, O. G. Rodriguez-Herrera, M. F. P. {n}eros, and J. S. Tyo. Satellite-derived tropical cyclone intensity in the north pacific ocean using the deviation-angle variance technique. *Weather and Forecasting*, 29:505–516, 6 2014. ISSN 0882-8156. doi: 10.1175/WAF-D-13-00133.1.

B. F. Ryan, I. G. Watterson, and J. L. Evans. Tropical cyclone frequencies inferred from gray's yearly genesis parameter: Validation of gcm tropical climates. *Geophysical Research Letters*, 19:1831–1834, 1992.

J. C. Sadler. On the origin of tropical vortices. *Working Panel on Trop. Dyn. Meteor. Naval Post Graduate School, Monterey*, pages 39–75, 1967.

J. C. Sadler. A role of the tropical upper tropospheric trough in early season typhoon development. *Monthly Weather Review*, 104:1266–1278, 1976. doi: 10.1175/1520-0493(1976)104<1266:AROTTU>2.0.CO;2.

H. S. Saffir. Hurricane wind and storm surge. *The Military Engineer*, 423:4–5, 1973.

T. Sakuragi, S. Hoshino, and N. Kitabatake. Development and verification of a tropical cyclone intensity estimation method reflecting the variety of trmm/tmi brightness temperature distribution. URL `http://daac.gsfc.nasa.gov`.

B. D. Santer, T. Wigley, P. Gleckler, C. Bonfils, M. Wehner, K. AchutaRao, T. Barnett, J. Boyle, W. Br{ü}ggemann, M. Fiorino, and Others. Forced and unforced ocean temperature changes in atlantic and pacific tropical cyclogenesis regions. *Proceedings of the National Academy of Sciences*, 103:13905–13910, 2006.

M. A. Saunders and A. R. Harris. Statistical evidence links exceptional 1995 atlantic hurricane season to record sea warming. *Geophysical Research Letters*, 24:1255–1258, 1997.

S. Schmidt, C. Kemfert, and P. H{ö}ppe. The impact of socio-economics and climate change on tropical cyclone losses in the usa. *Regional Environmental Change*, 10:13–26, 2010.

C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 40:185–197, 1 2010. ISSN 1083-4427. doi: 10.1109/TSMCA.2009.2029559.

R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. 10 2016. doi: 10.1007/s11263-019-01228-7.

S. Seneviratne, N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang. Changes in climate extremes and their impacts on the natural physical environment: An overview of the ipcc srex report. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*, pages 109–230, 2012.

L. J. Shapiro. Hurrican climatic fluctuation. part i: Patterns and cycles. *Monthly Weather Revie*, 110:1007–1013, 1982a.

L. J. Shapiro. Hurricane climatic fluctuation. part ii: Relation to large-scale circulation. *Monthly Weather Review*, 110:1014–1023, 1982b.

L. J. Shapiro and S. B. Goldenberg. Atlantic sea surface temperatures and tropical cyclone formation. *Journal of Climate*, 11:578–590, 1998.

J. D. Simon Banholzer, Sandra; Kossin. The impact of climate change on natural disasters. *Springer*, Preventing, 2014. doi: 10.1007/978-94-017-8598-3.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. 9 2014.

R. Simpson. The hurricane disaster—potential scale. *Weatherwise*, 27:169–186, 1974. ISSN 0043-1672. doi: 10.1080/00431672.1974.9931702.

R. H. Simpson and H. Riehl. *The hurricane and its impact*. Louisiana State University Press, 1981.

S. Smith. Managing catastrophic risk: beyond cat bonds. weather risk management: a guide for corporations, hedge funds and investors. *Incisive Financial Publishing*, 2010.

M. Sugi, A. Noda, and N. Sato. Influence of the global warming on tropical cyclone climatology: An experiment with the jma global model. *Journal of the Meteorological Society of Japan. Ser. II*, 80:249–272, 2002.

M. Sugi, H. Murakami, and J. Yoshimura. A reduction in global tropical cyclone frequency due to global warming. *Sola*, 5:164–167, 2009.

M. Sugi, K. Yoshida, and H. Murakami. More tropical cyclones in a cooler climate? *Geophysical Research Letters*, 42:6780–6784, 2015. ISSN 19448007. doi: 10.1002/2015GL064929.

K. L. Swanson. Nonlocality of atlantic tropical cyclone intensities. *Geochemistry, Geophysics, Geosystems*, 9, 2008. ISSN 15252027. doi: 10.1029/2007GC001844.

M. Tan and Q. V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. 5 2019.

A. Timchalk, L. F. Hubert, and S. Fritz. *Wind speeds from TIROS pictures of storms in the tropics*. US Department of Commerce, Weather Bureau, 1965.

H. Tonkin, G. J. Holland, N. Holbrook, and A. Henderson-Sellers. An evaluation of thermodynamic estimates of climatological maximum potential tropical cyclone intensity. *Monthly Weather Review*, 128:746–762, 2000.

K. E. Trenberth, J. Fasullo, and L. Smith. Trends and variability in column-integrated atmospheric water vapor. *Climate Dynamics*, 24:741–758, 2005. ISSN 09307575. doi: 10.1007/s00382-005-0017-4.

K. University. Gms/goes9/mtsat data archive for research and education., 2015. URL `http://weather.is.kochi-u.ac.jp/archive-e.html`.

G. A. Vecchi and B. J. Soden. Increased tropical atlantic wind shear in model projections of global warming. *Geophysical Research Letters*, 34, 2007a. ISSN 00948276. doi: 10.1029/2006GL028905.

G. A. Vecchi and B. J. Soden. Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, 450:1066–1070, 2007b.

G. A. Vecchi, B. J. Soden, A. T. Wittenberg, I. M. Held, A. Leetmaa, and M. J. Harrison. Weakening of tropical pacific atmospheric circulation due to anthropogenic forcing. *Nature*, 441:73–76, 2006.

G. A. Vecchi, K. L. Swanson, and B. J. Soden. Whither hurricane activity? *Science*, 322: 687–689, 2008.

C. Velden, B. Harper, F. Wells, J. L. Beven, R. Zehr, T. Olander, M. Mayfield, C. Guard, M. Lander, R. Edson, and Others. The dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, 87:1195–1210, 2006.

C. S. Velden, T. L. Olander, and R. M. Zehr. Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery. *Weather and Forecasting*, 13:172–186, 1998.

K. Walsh. Tropical cyclones and climate change: Unresolved issues. *Climate Research*, 27:77–83, 2004. ISSN 0936577X. doi: 10.3354/cr027077.

K. J. Walsh and B. F. Ryan. Tropical cyclone intensity increase near australia as a result of climate change. *Journal of Climate*, 13:3029–3036, 2000.

K. J. Walsh, J. L. Mcbride, P. J. Klotzbach, S. Balachandran, S. J. Camargo, G. Holland, T. R. Knutson, J. P. Kossin, T. cheung Lee, A. Sobel, and M. Sugi. Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 7:65–89, 2016. ISSN 17577799. doi: 10.1002/wcc.371.

P. J. Webster, G. J. Holland, J. A. Curry, and H. R. Chang. Atmospheric science: Changes

in tropical cyclone number, duration, and intensity in a warming environment. *Science*, 309:1844–1846, 2005. ISSN 00368075. doi: 10.1126/science.1116448.

H. Wexler. Radar storm detection. *Transactions, American Geophysical Union*, 28, 1947.

I. WGII. Climate change 2014: Impacts, adaptation, and vulnerability. *Summary for Policymakers*, pages 1–44, 2014.

H. Willoughby, J. Clos, and M. Shoreibah. Concentric eye walls, secondary wind maxima, and the evolution of the hurricane vortex. *Journal of atmospheric sciences*, 39:395–411, 1982.

A. A. Wing, A. H. Sobel, and S. J. Camargo. Relationship between the potential and actual intensities of tropical cyclones on interannual time scales. *Geophysical Research Letters*, 34, 2007.

S.-P. Xie, C. Deser, G. A. Vecchi, J. Ma, H. Teng, and A. T. Wittenberg. Global warming pattern formation: Sea surface temperature and rainfall. *Journal of Climate*, 23:966–986, 2010.

M. Yanai, S. K. Esbensen, and J.-H. Chu. Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *Journal of the Atmospheric Sciences*, 30:611–627, 1973.

Y. Yang, K. Zha, Y.-C. Chen, H. Wang, and D. Katabi. Delving into deep imbalanced regression. 2 2021. URL `http://arxiv.org/abs/2102.09554`.

J. T. D. YDJG Houghton, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. Johnson. Climate change 2001. the scientific basis, 2001. URL `https://www.ipcc.ch/report/ar3/wg1/`.

M. Ye, J. Wu, W. Liu, X. He, and C. Wang. Dependence of tropical cyclone damage on maximum wind speed and socioeconomic factors. *Environmental Research Letters*, 15, 2020. ISSN 17489326. doi: 10.1088/1748-9326/ab9be2.

J. Yoshimura, M. Sugi, and A. Noda. Influence of greenhouse warming on tropical cyclone frequency. *Journal of the Meteorological Society of Japan. Ser. II*, 84:405–428, 2006.

R. Zehr. Improving objective satellite estimates of tropical cyclone intensity. 1989.

M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks, 2014.

M. Zhao, I. M. Held, S.-J. Lin, and G. A. Vecchi. Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution gcm. *Journal of Climate*, 22:6653–6678, 2009.

Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang. Random erasing data augmentation. 8 2017. URL http://arxiv.org/abs/1708.04896.

Z.-H. Zhou and X.-Y. Liu. On multi-class cost-sensitive learning. *Computational Intelligence*, 26:232–257, 7 2010. ISSN 08247935. doi: 10.1111/j.1467-8640.2010.00358.x.

List of Figures

List of Tables