# Algorithmic Hallucinations of Near-Surface Winds: Statistical Downscaling with Generative Adversarial Networks to Convection-Permitting Scales

by

Nicolaas J. Annau
BSc, University of Victoria, 2018

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the School of Earth and Ocean Sciences

University of Victoria



We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

Algorithmic Hallucinations of Near-Surface Winds:
Statistical Downscaling with Generative Adversarial Networks to
Convection-Permitting Scales

by

Nicolaas J. Annau
BSc, University of Victoria, 2018

Supervisory Committee

---

Dr. Alex J. Cannon, Co-Supervisor
(School of Earth and Ocean Sciences, University of Victoria)

---

Dr. Adam H. Monahan, Co-Supervisor
(School of Earth and Ocean Sciences, University of Victoria)

---

Dr. Stan E. Dosso, Departmental Member
(School of Earth and Ocean Sciences, University of Victoria)

## ABSTRACT

Providing timely and accurate small-scale information about weather and climate is challenging – especially for variables strongly controlled by processes that are unresolved by low-resolution (LR) models. Motivated by this challenge, this thesis employed emerging artificial intelligence methods from the fields of computer vision and deep learning for the statistical downscaling of surface wind variables to convection-permitting scales. Specifically, Generative Adversarial Networks (GANs) were conditioned on LR inputs to deterministically generate possible arrangements of high-resolution (HR) surface wind patterns using super-resolution (SR). SR models were trained over several subregions in the contiguous United States and southern Canada, allowing for a systematic analysis of the methods. In addition to matching the statistical properties of the target dataset, GANs generate fields with impressive realism and computational efficiency, making them attractive for operational statistical downscaling. However, objectively assessing the realism of the SR models required a careful selection of evaluation metrics. Using power spectra successfully revealed how altered GAN configurations changed spatial structures in the generated fields, where biases in variability originate, and the role of including additional LR covariates in the SR methods. Inspired by recent work in the computer vision field, a novel methodology that separates spatial frequencies in HR fields was used in an attempt to optimize the SR GANs further. This method, called frequency separation, ultimately deteriorated the realism of the generated HR fields. However, frequency separation revealed how spatial structures are influenced by the metrics used to optimize the SR methods. A covariate-sensitivity analysis was also performed, illustrating how physical relationships between the large and small scales are mirrored in the GANs. Furthermore, a methodology was used to evaluate serial dependence in the SR methods. Even without explicitly including temporal information while training, serial dependence in the fine-scale structures of the generated fields was present. Although this thesis did not use fully-stochastic SR models (as has been done in existing work), the results provide valuable insights into variability, biases, covariates, and the optimization problem of SR.

# Table of Contents

# List of Tables

# List of Figures

# ABBREVIATIONS

# ACKNOWLEDGEMENTS

Thank you to:

**My mom, dad, and sister,** for modelling tremendous work ethic, and showing me a prodigious amount of support;

**My partner, Kirsten,** for her continual love and encouragement that motivates me more than she knows;

**My friend, Levy,** for his friendship, support and perspectives, and all those times we went stargazing instead of studying;

**My late grandparents,** who enabled my education;

**My supervisors, Dr. Alex Cannon and Dr. Adam Monahan,** whose incredible knowledge, guidance, and patience taught me the traits of good scientists;

**My dog, Comet,** whose friendship provided respite from pandemic-induced lonesome days – despite his lack of some logical faculties;

**Environment and Climate Change Canada,** for funding my research; and

**The Faculty of Graduate Studies at the University of Victoria,** for awarding me with a scholarship.

*Limits of survival are set by climate, those long drifts of change which a generation may fail to notice. And it is the extremes of climate which set the pattern. Lonely, finite humans may observe climatic provinces, fluctuations of annual weather and occasionally may observe such things as "This is a colder year than I've ever known." Such things are sensible. But humans are seldom alerted to the shifting average through a great span of years. And it is precisely in this alerting that humans learn how to survive on any planet.*

*They must learn climate.*

Frank Herbert, Children of Dune

# DEDICATION

To my parents, who dedicate their lives so that my sister and I can see success on the paths we choose. And, to all the communities enduring the harshness of climate now, and in the years to come.

# Chapter 1

# Introduction

Weather and climate consist of highly variable components that directly influence the habitability of communities. Consequently, providing accurate information about weather and climate saves lives, protects ecosystems, and helps practitioners and decision-makers plan for the future. However, accurately representing all aspects of these dynamic and chaotic components in computational models is a challenging task with imperfect solutions. As a result, relevant climate and weather information is not always available to practitioners. Addressing this problem is the core motivation for this thesis.

Atmospheric patterns exist at spatial scales ranging from micrometres to thousands of kilometres, and temporal scales from microseconds to decades. Accurately representing this range of patterns in climate models forms a massive challenge for computational simulations, especially since patterns at these scales are not necessarily independent of each other. In other words, large-scale circulations can both influence and be influenced by, small-scale processes (Thompson, 1957). Nevertheless, since variability is more pronounced in large-scale atmospheric patterns (e.g. continental variations in weather are more pronounced than inter-city variations), simulating large-scale processes is a necessary starting point.

Earth System Models (ESMs), for example, integrate a host of complex physical Earth System interactions to represent large-scale processes on the order of 100 km. ESMs are invaluable tools for understanding past, and future climate conditions and scenarios (Stocker, 2014). However, variables in ESMs that are strongly influenced by surface heterogeneities (e.g., wind and precipitation), contain significant biases since dominant small-scale processes are unresolved (Whiteman, 2000; Kharin et al., 2007; Stephens et al., 2010; Sillmann et al., 2013; Song et al., 2020).

Explicitly simulating small-scale processes is particularly difficult due to computational constraints, so methods for dealing with this difficulty form the research area of *downscaling*. Downscaling aims to develop models with accurate small-scale information by exploiting the relationships between large scales and small scales and is separated into two approaches: *dynamical* downscaling and *statistical* downscaling.

Traditionally, dynamical downscaling broadly includes models that operate by numerically integrating partial differential equations that describe physical processes provided initial and boundary conditions (Giorgi and Mearns, 1991). These conditions are usually provided by low-resolution (LR) climate models, like ESMs. In order to get smaller-scale information, Regional Climate Models (RCMs), for instance, dynamically downscale ESMs to a finer horizontal resolution on the order of 10 - 100 km (Plummer et al., 2006). RCMs help reduce biases in ESMs due to higher-resolution representations of physical processes and complex topography (Frei et al., 2003; Ban et al., 2015; Torma et al., 2015; Schlager et al., 2019). Nevertheless, biases persist in RCMs because they do not resolve mesoscale and smaller processes, such as mesoscale convective systems (MCSs), and because complex topography is still not adequately represented on their 10 - 100 km horizontal scales (Déqué et al., 2007; Prein et al., 2015; Liu et al., 2017). For the purposes of the present work, ESMs and RCMs are both considered LR models.

On the other hand, statistical downscaling seeks to exploit empirical links between large-scale and small-scale processes using statistical models (Maraun and Widmann, 2018). Meteorological stations, for instance, offer point-wise observations that are not subject to misrepresentations of simulated small-scale processes in models. They can also be interpolated onto gridded datasets using statistical downscaling to produce higher-resolution datasets. However, stations often have incomplete historical records and highly heterogeneous spatial distributions that pose a significant challenge for statistical models. Addressing these incompleteness issues, *reanalysis* data combines past weather forecasts, remote sensing, and station observations through data assimilation to interpolate onto a gridded dataset (Dee et al., 2011; Hersbach et al., 2020). Reanalyses products fill in gaps in station observations and provide a complete global record of past weather and climate. Like RCMs, these products often have a grid-spacing on the order of 10 - 100 km and can be used for statistical or dynamical downscaling, and for bias correction of RCMs and ESMs (Liu et al., 2017; Turco et al., 2017). However, like stations, reanalyses alone cannot represent future conditions.

Since localized assessments of climate conditions are required for adaptation and impact research, additional methods that can effectively model small-scale processes are thus required.

High-resolution (HR) models, such as convection-permitting models (CPMs), dynamically downscale LR models while using them as initial and boundary conditions. At $\leq 4$ km horizontal grid-spacing, CPMs offer improved representations of small-scale variability over ESMs, RCMs, and reanalyses, owing to the explicit resolution of convection and better representations of orography (Kopparla et al., 2013; Prein et al., 2016; Innocenti et al., 2019). While CPMs offer a complete spatiotemporal record of climate variables, due to their computational cost, domains are currently limited to regional to continental scales, and simulations cover relatively short periods. Furthermore, this computational cost does not allow for multiple realizations in large ensemble experiments, which are highly desirable for climate impact studies.

Localized assessments of climate conditions require fast and reliable statistical downscaling methods that can properly represent small-scale processes (Wilby and Wigley, 1997; Cannon, 2008; Sobie and Murdock, 2017; Li et al., 2018). As such, statistical downscaling methods remain an important area of study for researchers, especially as the changing climate threatens infrastructure, societies, entire cultures, and the health of ecosystems (Seneviratne et al., 2012; Watt-Cloutier, 2016; Bush and Lemmen, 2019). The reliability of statistical downscaling, however, depends on the variable and method employed (Maraun and Widmann, 2018).

Machine learning (ML) and deep learning (DL) are branches of artificial intelligence (AI) concerned with getting computers to "learn" how to perform certain tasks. Corresponding models used in DL, such as artificial neural networks (loosely modelled after how biological brains process information), can be used as highly non-linear and flexible statistical models (Gardner and Dorling, 1998). DL models can also be designed to be optimized to perform certain tasks. For example, Convolutional Neural Networks (CNNs) are a class of DL models that are optimized for images and spatial fields (Krizhevsky et al., 2012). CNNs, owing to their capabilities with spatial fields, show promise for use as a downscaling method for several climate variables.

One such promising method, called *single-image super-resolution* (SISR or simply SR), aims to develop models that produce plausible HR details from LR inputs, and using CNNs, SR has achieved significant improvements (Dong et al., 2014; Ledig et al., 2017; Zhang et al., 2018; Zhu et al., 2020). This present work aims to evaluate SR models from computer vision for the multivariate statistical downscaling of surface

wind patterns from a reanalysis product to CPM scales.

Downscaling wind field patterns are chosen for this study because of their improved representations in CPMs compared to LR models. They are also multivariate and highly variable in nature – posing a challenge for statistical downscaling methods so that SR method capabilities and limitations can be examined. As such, the SR models considered in this work can be useful for a variety of practical applications, such as fire weather danger and wildfire behaviour, pollutant and pollen dispersal, infrastructure design, and wind-turbine siting – situations where dynamically downscaled CPMs may be unavailable.

## 1.1 The Challenge of Super-Resolution

From the computer vision field, SR is often concerned with enhancing the resolution of images in a deterministic manner. This determinism implies that, given unique arrangements of LR patterns, there are unique arrangements of HR grid cells or pixels that are produced by the method – i.e. there is exactly one output image for one input image. Similarly, statistical downscaling is often performed deterministically. However, since both SR and statistical downscaling are concerned with enhancing spatial resolution, they are both inherently ill-posed in the presence of a scale gap; the set of possible arrangements of HR patterns associated with fixed arrangements of LR patterns is infinite.

For HR images, this indeterminacy of fine-scale features can be considered an undesirable limitation of SR because only one true reference image exists. However, for CPMs (i.e. the "true" reference in the present work) driven with unique LR models as boundary conditions, non-unique fine-scale features are produced if provided different initial conditions. Since initial conditions are not perfectly known, CPMs produce fine-scale features that *could* exist in the real atmosphere. The reader will note that "true" is put in quotations to highlight that CPMs do not represent the real atmosphere exactly, rather, they show possible approximate representations. It follows that generating fine-scale features (e.g. weather in climate fields) through statistical downscaling that *could* exist in climate fields is a requisite outcome of SR, and can therefore be considered a desirable feature of the method.

In the computer vision field, generating these fine-scale features using SR has been likened to *hallucinating* details that could exist in the data used to train the models, but do not necessarily exactly match on a pixel-to-pixel basis (Zhang et al.,

2020). Although deterministic SR does not sample from the full set of possible HR arrangements associated with given LR fields, it does generate a *possible* individual realization from this set of arrangements. This work designs deterministic SR methods that, when conditioned on unique LR patterns, generate unique arrangements of HR patterns that could exist – even if not explicitly found in the climate simulation.

These *possible* arrangements of HR patterns motivates the careful consideration of error metrics to assess the downscaled fields. Grid-point-based (or pixel-wise) metrics, such as mean squared error (MSE) or mean absolute error (MAE) between SR generated and "true" HR field pixels are frequently exclusively used to optimize and evaluate SR methods in both climate and computer vision contexts. However, when using these metrics to compare fine-scale features in CPMs, enforcing strict adherence to pixel-wise errors penalizes physically realizable – but non-congruent (i.e. mismatched) – high-frequency patterns. This problem is similar to limitations of grid-point based error measures for precipitation fields, known as the *double penalty problem* (Rossa et al.; Michaelides, 2008; Harris et al., 2022). This present work uses novel ML methods from the computer vision field that evaluate both distributional distances and grid-point-based metrics to help mitigate the double penalty problem.

Another challenge of SR and statistical downscaling lies in the relationship between the LR and HR patterns. Implementations of SR in previous computer vision studies frequently coarsen HR images by a factor of four in each spatial dimension to produce the LR inputs for training, i.e. an HR image with dimensions 128px $\times$ 128px would be coarsened to 32px $\times$ 32px. This coarsening approach guarantees bias-free pairs of LR and HR fields for training. Idealized coarsening is generally not a realistic feature of climate datasets since LR and HR patterns may be produced separately in different models – even if they are synchronous in time. When the scale separation is large, the variability of the small-scale features embedded within the large-scale features is also large, leading to a larger range of different kinds of HR fields that LR patterns can produce. This present work involves a scale factor of eight, resulting in the LR scales having weaker control on the HR scales (e.g. the LR fields have only $\frac{1}{8\times 8} \approx 1.56\%$ of the number of grid-points in the HR fields), and also uses LR and HR fields that come from different climate models so that the different scales are not *necessarily* bias-free. In other words, the spatial representation of features in the HR and LR fields may not match exactly.

## 1.2 Contributions

This thesis aims to evaluate the suitability of DL-based SR methods for statistical downscaling in detail, and considers the following topics:

1. Multivariate SR DL models that generate HR horizontal wind field components are designed and trained.

2. The double penalty issue is addressed in this work by using a distribution-based metric (i.e., the Wasserstein distance) in combination with grid-point-based metrics to optimize SR methods. This allows more freedom to generate non-congruent small-scale details.

3. Objectively evaluating the fidelity of the SR models is challenging and requires careful selection of metrics. Spatial frequency scales are utilized to validate the accuracy of generated variability across scales and are found to be sensitive to spatial properties of the generated HR fields.

4. Three climatologically distinct regions of identical size are isolated for SR applications. An additional continental-scale region is considered. By changing regions, the performance of the SR methods in different climatologies (and different HR scale variability) is determined.

5. Sensitivity of results to LR covariates for climate-related SR has not previously been rigorously evaluated. This work proposes a methodology for evaluating the sensitivity of SR wind fields to the inclusion of covariates, thus evaluating the importance of these LR input fields.

6. In this work, SR models are designed and trained that are not explicitly stochastic, i.e. the DL models do not produce a distribution of HR fields when conditioned on unique LR fields. Instead, this work considers SR models that produce unique realizations of HR fields given unique LR conditioning fields. However, the DL models presented here place features where they could be, rather than where they are in the "true" fields. Subsequently, this work identifies sources of biases in this generated variability. As such, it is an important step towards fully stochastic generative DL models.

## 1.3 Terminology

Before proceeding, it is important to acknowledge that computer vision literature generally has its own nomenclature that differs from the climate sciences – even for identical or very closely related concepts. In the present work, what the computer vision field refers to as *images* will be referred to as *fields* to be more consistent with climate science terminology. Downscaling (*upsampling* in the computer vision field) refers to the enhancing of spatial resolution, and upscaling (*downsampling* in the computer vision field) refers to the coarsening of spatial resolution. Colour channels are used to describe the non-spatial axis of the input/output images in the computer vision literature. In this context, channels can be thought of as conditioning climate field variables, or *covariates*. *Pixels* in computer vision naturally refers to *grid-boxes* or *grid-points* in climate models. Further, the terms *fine-scale*, *small-scale*, and *high-frequency* will be used interchangeably in the context of spatial scales. Similarly, *large-scale*, and *low-frequency* will be used interchangeably.

# Chapter 2

# Background

## 2.1 Convolutional Neural Networks

CNNs are a class of highly flexible ML models optimized to represent locally connected structures in data such as those found in images or climate model fields (Krizhevsky et al., 2012). They consist of numerous "trainable" parameters (i.e., weights and biases) to be optimized that are organized as *filters* (or kernels). Their purpose is to activate in the presence of certain structures (e.g., edges or shapes) in the input data by performing hierarchical convolutional operations that communicate this activated information through the network (Karpathy, 2022). Importantly, these filters are not defined before training. Rather, they are "learned" by the CNN during training to detect features that are relevant to the data.

Depending on the design of the architecture, these filters define layers in the CNN, which usually consist of a convolutional layer followed by an activation or transfer function that collectively form *convolutional blocks*. Activation functions are typically a simple function, such as *tanh*, *sigmoid*, or *Rectified Linear Unit (ReLU)*, that transform the output of convolutional layers and are essential for introducing non-linearity into neural networks to help better represent complex patterns (Murphy, 2022, 424-425). The output of a convolutional block is a filtered and activated representation of the input data, with a unique representation for each unique filter. The filtered representations are themselves filtered in subsequent layers in the CNN, i.e., outputs from one layer are used as inputs in the next. In practice, CNNs typically have many layers and hence are a form of DL. All of the weights and biases in the CNN are tuned to minimize a loss function using some form of gradient descent, where gradi-

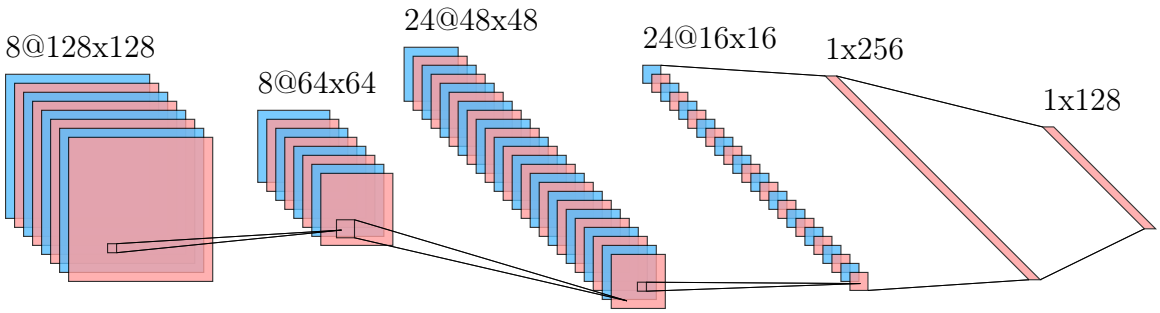


Figure 2.1: An example of a CNN processing input data (with arbitrary shapes and dimensions) for illustrative purposes. Each set of squares represents a convolutional block or layer in the CNN. Connecting lines are meant to demonstrate that previous layers inform the dimensions of subsequent layers as the convolutional blocks transform the shape of the data. The final layers are vectors that are rich in information extracted from the input data. This figure was produced using a tool developed by LeNail (2019).

ents with respect to these weights and biases are calculated through a process called *backpropagation* (Rumelhart et al., 1986).

CNN architectures themselves are very flexible, and can both downsample and/or upsample data to different representations. A general example of the former is shown conceptually in Figure 2.1 (reading left to right), where input data starts at a high-dimensional representation and is downsampled through the network to a lower-dimensional representation. The latter configuration can be thought of as the reverse version of Figure 2.1 (reading right to left) and can be adapted to generate HR fields from LR inputs in SR tasks. Combinations of these architectures can form the backbone of generative DL models (Radford et al., 2015; Pu et al., 2016). CNNs are used extensively in this work; specific CNN architectures and how they are trained are discussed in more detail in Section 3.

## 2.2 Generative Adversarial Networks

Substantial improvements have been made in the SR field by using CNNs (Dong et al., 2015), and by adopting conditional generative adversarial networks (GANs) (Goodfellow et al., 2014; Mirza and Osindero, 2014) for SR tasks (Ledig et al., 2017; Zhu et al., 2020). GANs are an ML architecture that consists of duelling functions (often CNNs) trained simultaneously with opposing objectives. These objectives shape two networks that communicate with each other, namely, the *Generator*, which aims to

generate realistic information, and the *Discriminator*, which aims to judge or critique this generated information and provide feedback to the Generator. The main idea behind GANs comes from the use of the Discriminator to provide an adaptive loss term – the *adversarial loss* – that increases in capacity for evaluating detail as training proceeds. As a result, the Generator receives increasingly relevant feedback as it generates increasingly realistic information. A more detailed discussion of GANs, and how they are adopted for SR problems is included in Section 3.1.

## 2.3 Relevancy of Super-Resolution in the Climate Sciences

Generating SR fields with realistic textures has been a core challenge in the SR field (Ledig et al., 2017) that remains relatively unaddressed for climate fields. In statistical downscaling, there are existing techniques that generate conditional small-scale variability, such as stochastic weather generators (SWGs). SWGs are concerned with generating random sequences of atmospheric variables that are loyal to the observed patterns and distributions of these variables. There exists a vast statistical downscaling literature that includes SWG methods varying from Markov chain models (Gabriel and Neumann, 1962; Richardson and Wright, 1984) to models for estimating conditional occurrence and intensity of precipitation from parametric distributions (Katz, 1977; Chandler and Wheater, 2002; San-Martín et al., 2017). Multi-site weather generators aim to generate spatially-correlated random weather sequences; however, incorporating local variability is a complicated task for multiple sites (and variables) due to the high-dimensional nature of the problem (Chandler, 2005; Cannon, 2008; Jeong et al., 2012; Maraun and Widmann, 2018). Furthermore, the HR information has a high degree of spatial inhomogeneity which makes specifying the distributional structures of the small scales challenging for spatially homogenous models that do not account for spatial non-stationarity (Wilks and Wilby, 1999; Ailliot et al., 2015; Maraun and Widmann, 2018). SR methods – specifically those emerging from ML computer vision fields like GANs using CNNs – are sufficiently flexible for use in a variety of different applications not limited to computer vision. Their ability to represent complex spatial patterns is an attractive feature when modelling high dimensional inter-site dependencies, like those encountered when using SWGs for spatially-correlated weather sequences. Importantly, unlike many SWG approaches,

SR methods do not make assumptions about the underlying distributional nature of the atmospheric variables considered. As such, SR methods can be viewed as non-parametric SWGs.

## 2.4 Existing Work

SR methods have been applied to climate fields with success. For example, GANs have been employed for super resolving wind (Singh et al., 2019; Stengel et al., 2020), and solar irradiance fields (Stengel et al., 2020). Sha et al. (2020) downscales temperature using CNNs over western continental United States, and Wang et al. (2021) uses a deep residual network for downscaling daily precipitation at a variety of resolution gaps for two climatologically unique regions in the continental United States. Additionally, Kumar et al. (2021) uses Super-Resolution Convolutional Neural Network (SRCNN) to downscale rainfall data for regional climate forecasting. However, each of these above studies uses idealized, coarsened versions of the HR data to deterministically downscale the LR fields.

GANs adopted for SR in the climate sciences have also successfully produced ensembles of super-resolved fields using coarse covariates as well as noise to stochastically generate HR fields. This approach has shown promise for precipitation in Leinonen et al. (2020), however, again with idealized LR/HR pairs. Recently, Harris et al. (2022) demonstrated that GANs are an attractive stochastic downscaling method for non-idealized pairs, as is a novel Variational Auto Encoder GAN (VAE-GAN) method. A similar approach to Harris et al. (2022) was taken by Price and Rasp (2022) for stochastic SR with precipitation fields. The present work considers non-idealized LR and HR pairs using advancements in GAN architectures for deterministic SR, although stochastic methods will be revisited for discussion in Section 5.

## 2.5 Super-Resolution Considerations

As discussed earlier, one important consideration when using SR methods is the scale gap between the LR covariates and the HR fields, i.e how many pixels in the HR fields are contained within an LR pixel. Typical applications of SR in the computer vision field achieve a $4\times$ scaling (Dong et al., 2015; Ledig et al., 2017; Wang et al., 2018; Cheng et al., 2020). For the climate sciences, the scale gaps (or scale factors) that

exist between the available LR and HR datasets often exceed 4× (Stengel et al., 2020; Leinonen et al., 2020; Harris et al., 2022). For technical reasons depending on the CNN architectures used, these large-scale factors pose an additional challenge to the GAN networks because more parameters need to be fitted to overcome the scale gap and generate HR images. A further discussion on the networks used in this present work is included in Section 3.5.

Another consideration is how the LR and HR pairs are determined in the first place. Many of the examples used to train GANs for SR in both the computer vision and climate fields come from an artificial coarsening of the HR fields to produce the LR conditioning fields (Dong et al., 2015; Ledig et al., 2017; Wang et al., 2018; Singh et al., 2019; Sha et al., 2020; Cheng et al., 2020; Stengel et al., 2020; Wang et al., 2021; Kumar et al., 2021). This approach guarantees synchronous LR and HR pairs that have idealized consistencies between the scales with no differences at large scales. In practical downscaling applications, the LR fields may not be coarsened versions of the HR fields, so there will be some differences at large scales. Perfect agreement in these coarse scales is unrealistic in this sense. In this study, the Weather Research and Forecasting (WRF) model over the High-Resolution Contiguous United States (HRCONUS) domain (Liu et al., 2017) supplies the HR fields, with synchronous European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-Interim) (Dee et al., 2011) providing the LR fields. A further discussion of the climate models used can be found in Section 3.3.

This work aims to demonstrate the feasibility of GANs for downscaling when challenged by these considerations; that is, large-scale factors and non-idealized LR/HR field pairs. In doing so, it also aims to demonstrate how GAN SR methods have properties resembling those of statistical downscaling and SWG approaches. Inspired by novel work in the computer vision field, it also attempts to improve the performance of GANs for non-idealized pairs by changing the roles of the terms in the objective functions introduced in Section 3.4.1.

# Chapter 3

# Data and Methodology

## 3.1 Utilizing the Wasserstein Distance

The early GAN formulation from Goodfellow et al. (2014) contains duelling objectives in a min-max game between a Generator and Discriminator CNN. In the min-max game, the purpose of the Generator, denoted $G$, is to generate realistic samples that fool the Discriminator, denoted $D$. The Discriminator's purpose is then to distinguish between the "true" and generated samples. Mirza and Osindero (2014) reframed the GAN so that the Generator is supplied conditional information to exert control on generated samples. This conditional approach was adopted by Ledig et al. (2017) in the seminal super-resolution GAN (SRGAN), where LR samples condition the Generator to produce HR samples. If $\mathbf{y}$ is a sample drawn from the "true" distribution, $\mathbf{y} \sim \mathbb{P}_r$, and $G(\mathbf{x})$ is a sample drawn from the generated distribution, $G(\mathbf{x}) \sim \mathbb{P}_g$, then $G$'s purpose is to generate samples conditional on samples from the LR distribution, $\mathbf{x} \sim \mathbb{P}_c$, that are indistinguishable from $\mathbf{y}$. The optimal GAN in this formulation properly estimates the Jensen-Shannon (JS) divergence between $\mathbb{P}_r$ and $\mathbb{P}_g$ through the Discriminator and minimizes it through the Generator.

The success of GANs for SR can be attributed to minimizing a distributional distance between "true" and generated HR samples; rather than, for example, grid-point metrics like the MAE or MSE. However, in practice, the convergence of the min-max game of the early GANs is not guaranteed. Early GANs are notoriously difficult to train and suffer from issues such as *vanishing-gradients*, and *mode-collapse* (Salimans et al., 2016; Arjovsky et al., 2017). The former occurs when the Discriminator vastly outperforms the Generator and does not provide relevant feedback, and the latter

occurs when the Generator exploits "blind spots" in the Discriminator resulting in a significant lack of diversity in the generated fields (i.e. lack of variability in the generated distributions).

Arjovsky et al. (2017), emphasizing how the choice of distance metric can improve GAN behaviour, changed the objective of the GAN to estimate and minimize the Wasserstein distance between $\mathbb{P}_r$ and $\mathbb{P}_g$. The resulting GAN is known as WGAN and is structurally similar to early GANs in that optimal WGANs estimate a distributional distance with a "criticizing" network and minimize this distance with the Generator. Using the Wasserstein distance has the main advantage of more stable convergence without suffering from vanishing gradients resulting from the theoretical benefits of the metric. This improvement in stability can be attributed to the altered role of the Discriminator, which, instead of distinguishing between $\mathbf{y}$ and $G(\mathbf{x})$ and estimating the JS divergence, is optimized to estimate the Wasserstein distance. This change of role of the Discriminator is reflected in a new name for the network in WGAN contexts, namely, the *Critic*, denoted $C$.

For a more illustrative interpretation, if the distributions $\mathbb{P}_r$ and $\mathbb{P}_g$ are analogous to mounds of gravel with the same total mass, the Wasserstein distance is the minimum "effort" required to move gravel within the generated mound to match the target one. The Wasserstein distance can be thought of as measuring the most efficient way of rearranging gravel to have it appear identical to a target distribution. For this reason, the Wasserstein distance is sometimes referred to as the earth mover's distance, or optimal transport, for discrete probability distributions.

One important advantage of using the Wasserstein distance that contributes to its stability in WGANs is that it remains continuous when there is no overlap between $\mathbb{P}_g$ and $\mathbb{P}_r$. This fact was illustrated in Arjovsky et al. (2017) for non-overlapping distributions on low-dimensional manifolds, where the JS divergence shows discontinuities in such situations while the Wasserstein distance does not. Due to these various advantages, within the climate science community, the Wasserstein distance has also been proposed as a means of comparing climate models (Ghil, 2016; Vissio et al., 2020).

If $\mathbb{P}_r$ and $\mathbb{P}_g$ are the "true" and generated distributions respectively, the Wasserstein distance can be written as:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(\mathbf{y}, G(\mathbf{x})) \sim \gamma} \left[ ||\mathbf{y} - G(\mathbf{x})|| \right] \tag{3.1}$$

where $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ is the set of all joint distributions, $\gamma(\mathbf{y}, G(\mathbf{x}))$ with marginals $\mathbb{P}_r$ and $\mathbb{P}_g$. In practice, calculating the infimum in Equation 3.1 is intractable to calculate the entire set of possible joint distributions, but can be reformulated using the Kantorovich-Rubinstein duality (Kantorovich, 2006; Villani, 2009; Arjovsky et al., 2017; Gulrajani et al., 2017):

$$W(\mathbb{P}_r, \mathbb{P}_g) = \sup_{||f||_L \leq 1} \mathbb{E}_{\mathbf{y} \sim \mathbb{P}_r}[f(\mathbf{y})] - \mathbb{E}_{G(\mathbf{x}) \sim \mathbb{P}_g}[f(G(\mathbf{x}))] \tag{3.2}$$

where sup is the supremum of all 1-Lipschitz functions. For $f$ to be 1-Lipschitz, it must satisfy:

$$|f(x_1) - f(x_2)| \leq |x_1 - x_2|, \quad \forall x_1, x_2 \in \mathbb{R} \tag{3.3}$$

In other words, 1-Lipschitz constrains the gradients of $f$ to be no more than 1 everywhere.

The Kantorovich-Rubinstein duality in Equation 3.2 is a theorem that enables easier estimations of the Wasserstein distance by changing it from a minimization over all joint probabilities, to a maximization over 1-Lipschitz functions (Villani, 2009; Arjovsky et al., 2017; Gulrajani et al., 2017). In practice, $f$ is the Critic, and is optimized to find 1-Lipschitz functions that estimate the Wasserstein distance. More intuitively, $f$ enhances small differences between samples from the "true" and generated distributions, and the Lipschitz constraint prevents $f$ from enhancing differences that are too small. As the loss function decreases in the training, the Wasserstein distance gets smaller and the Generator model's output grows closer to the "true" data distribution.

Enforcing 1-Lipshcitz is a key requirement of using the Kantorovich-Rubinstein duality, which Arjovsky et al. (2017) satisfies through gradient clipping. The use of gradient clipping to satisfy 1-Lipschcitz was superseded by WGAN with *gradient penalty* (WGAN-GP) (Gulrajani et al., 2017), which uses a softer and more practical approach to constrain gradients of the Critic. WGAN-GP is the form of GAN used in this work; specific loss functions and further details about the gradient penalty are discussed in the next section

## 3.2 Adapting WGAN-GP for Super-Resolution

Providing conditional information translates an unconditional GAN into a supervised DL task that can take advantage of the Critic. This frames the GAN as an advanced regression model that can be deterministic *or* stochastic depending on the stochasticity of the conditioning information. As discussed earlier, the present work considers a deterministic approach, but the extension to stochastic approaches will be considered in Section 5. Stochastic approaches to climate field SR have already been considered in Leinonen et al. (2020); Price and Rasp (2022); Harris et al. (2022).

The Critic and Generator each have an objective (or loss) function that, when minimized, optimizes their respective weights. For the Critic, the objective function has a Wasserstein distance component and a gradient penalty term that discourages violations of the 1-Lipschitz requirement:

$$\mathcal{L}_C = \underbrace{\mathbb{E}_{G(\mathbf{x})\sim\mathbb{P}_g}[C(G(\mathbf{x}))] - \mathbb{E}_{\mathbf{y}\sim\mathbb{P}_r}[C(\mathbf{y})]}_{\text{Wasserstein component}} + \underbrace{\lambda\,\mathbb{E}_{\hat{\mathbf{x}}\sim\mathbb{P}_{\hat{\mathbf{x}}}}[(||\nabla_{\hat{\mathbf{x}}}C(\hat{\mathbf{x}})||_2 - 1)^2)]}_{\text{Gradient Penalty}} \tag{3.4}$$

where $\lambda$ is a tuneable hyperparameter, and $\hat{\mathbf{x}}$ is a randomly synthesized field interpolated on a straight line between "true" and generated samples as

$$\hat{\mathbf{x}} = \epsilon\mathbf{y} + (1-\epsilon)G(\mathbf{x}) \tag{3.5}$$

where $\epsilon$ is a random number drawn from uniform noise, $\epsilon \sim U[0,1]$, and $G(\mathbf{x})$ is the generated field with LR covariates synchronous with $\mathbf{y}$. A further discussion on the gradient penalty term is in included in Appendix A.1. Importantly, Equation 3.4 optimizes the Critic to learn 1-Lipschitz functions that estimate the Wasserstein distance.

For the Generator, the objective function is as follows:

$$\mathcal{L}_G = \underbrace{-\,\mathbb{E}_{G(\mathbf{x})\sim\mathbb{P}_g}[C(G(\mathbf{x}))]}_{\text{Adversarial component/loss}} + \underbrace{\alpha\,\mathbb{E}_{\mathbf{y}\sim\mathbb{P}_r}\,l_c(\mathbf{y}, G(\mathbf{x}))}_{\text{Content loss}} \tag{3.6}$$

where $\alpha$ is a hyperparameter that weights the relative importance of the content loss, $l_c$, and the adversarial loss. The content loss is meant to guide the generated fields towards $\mathbf{y}$ and is necessary for stability while training the SR models. As will be discussed in Section 3.4.1, in climate applications the content loss should ensure

consistency of only the large-scale features of the HR fields with the conditioning LR fields, since the HR fields will necessarily have grid-point differences at small scales. This present work uses the mean absolute error (MAE) for the content loss as follows:

$$l_c(\mathbf{y}, G(\mathbf{x})) = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} |\mathbf{y}_{i,j} - G(\mathbf{x})_{i,j}| \tag{3.7}$$

where $W$ and $H$ are the width and height of the field in pixels.

In contrast to the supremum in Equation 3.2, minimizing Equation 3.4 also estimates the Wasserstein distance by simply switching the signs of the $C(\mathbf{y})$ and $C(G(\mathbf{x}))$ terms. Furthermore, in order to provide the best estimate of the Wasserstein distance (or adversarial component of the Generator's loss), $C$ in Equation 3.4 should be optimal. Gulrajani et al. (2017) therefore, updates the weights and parameters of $C$ a number of times for every update of the Generator's weights and parameters to ensure that $C$ reasonably estimates the Wasserstein distance. The number of updates is an additional tuneable hyperparameter.

For the Generator's loss, the generated part of the Wasserstein component in Equation 3.4 has an opposing role in Equation 3.6, where it forms the adversarial loss that $G$ aims to minimize. Since Equation 3.4 is minimized, larger values of $C$ correspond to more "realistic" fields because of the negative sign on $C(\mathbf{y})$. In order to guide the Generator to produce more realistic fields, the sign of $C(G(\mathbf{x}))$ is changed in the Generator's objective, and the second term of Equation 3.4 is dropped as its role is redundant and adds unnecessary computational cost.

Importantly, the adversarial component of Equation 3.6 does not compare grid points between "true" and generated fields but is directly related to the distributional distance between collections (i.e. batches) of fields measured by $C$. Because $\mathcal{L}_G$ contains both a distributional and grid-point-based component, even if $G(\mathbf{x})$ is produced without stochastic conditioning information, the adversarial component in Equation 3.6 is responsible for generating similar variability to $\mathbf{y}$, putting fine-scale features where they *could* be, but not necessarily where they are in the "true" HR fields.

## 3.3 Datasets

With the methodology introduced in the preceding sections, deterministic WGAN-GP SR models are developed that generate near-surface wind component fields simulated by the Weather Research and Forecasting (WRF) model over subregions in

the High-Resolution Contiguous United States (HRCONUS) domain (Rasmussen and Liu, 2017; Liu et al., 2017). In these simulations, WRF HRCONUS is configured as a CPM (4 km grid spacing) that is driven using data from six-hourly ERA-Interim (80 km grid spacing) (Dee et al., 2011) as its boundary conditions and initial conditions. Using GANs for SR, this work aims to generate HR fields – conditioned on ERA-Interim input (with the exception of one of these covariates, as discussed in Section 3.3.2) – that are consistent with WRF HRCONUS.

Through its boundary forcing, and spectral nudging at large scales above the boundary layer, WRF HRCONUS is synchronous with ERA-Interim during the historical period October 2000 to September 2013 (Rasmussen and Liu, 2017). This synchronization creates reasonable agreement at large scales between ERA-Interim and WRF HRCONUS. However, due to upscale energy transfers, they may not match exactly since smaller scales can evolve freely. Therefore, WRF HRCONUS and ERA-Interim are well positioned for evaluating the ability of GANs to generate fine-scale details present in WRF HRCONUS through SR.

Using these models also means that LR inputs are not generated by simply applying a coarsening function to the HR fields; thus they represent a non-idealized LR/HR pair. This non-idealized pairing places more responsibility on the GAN to correctly produce details that are consistent with the CPM, thereby testing the extent to which the Critic captures the characteristics of these features and enables the Generator to produce them. Data preparation will be discussed in Section 3.3.2.

### 3.3.1 Training and Test Sets

From WRF HRCONUS, six-hourly 10 m zonal and meridional wind fields, denoted by $u10$ and $v10$ respectively, are selected as the target fields for the GANs to generate. Large-scale variables from ERA-Interim are used as conditioning fields (or covariates). It is noted that WRF HRCONUS is archived at an hourly time step, however, in order for synchroneity with ERA-Interim, six-hourly is used here. There are a total of 18991 available fields in the 2000 - 2013 historical simulation period. Of these fields, 80% are used for training (15704 fields), and 20% (3287 fields) are used for testing and evaluation.

Although wind fields are spatially highly variable, sampling time steps randomly and uniformly could result in dependence between test and training sets due to serial autocorrelation. For example, a wind field in the training set may have very similar

features to an adjacent wind field six hours later that is randomly assigned as a test field. In order to account for this serial dependence and long-term non-stationarity in the wind field patterns, three separate years are used (2000, 2006, and 2010) for testing.

### 3.3.2 Data Preparation

WRF HRCONUS outputs are not on a regular lat/lon grid by default. So, to begin, WRF HRCONUS is re-gridded to a regular grid through nearest neighbour interpolation. While the native WRF HRCONUS grid spacing is ∼4 km, WRF HRCONUS is re-gridded to ∼10 km, resulting in a scale factor of eight with respect to ERA-Interim's 80 km grid spacing. Nearest neighbour interpolation is intentionally used to limit unintended smoothing by other methods, such as bilinear interpolation. Once prepared, the LR and HR field shapes are $16 \times 16$ grid cells and $128 \times 128$ grid cells respectively.

A total of seven covariate fields from reanalysis products were identified for the SR models to generate simultaneous HR $u10$ and $v10$ fields. These are provided to the modified Generator as extra colour channels (see Section 3.5 for a further discussion of the CNNs used). From ERA-Interim, the covariates include coarse $u10$, $v10$, topography, land-sea fraction, surface pressure, and surface roughness length.

Due to the unavailability of convective available potential energy (CAPE) in ERA-Interim at six-hourly time steps, CAPE from ERA5 is used instead (Hersbach et al., 2020) and is interpolated from the native 30 km grid spacing to 80 km. Using ERA5 in this manner increases the potential for mismatches between both ERA-Interim and WRF HRCONUS. However, ERA5 and ERA-Interim represent the same historical atmospheric conditions and so it is assumed that this mismatch is small with both WRF and ERA5 as well as ERA-Interim and ERA5. The motivation for including each additional covariate is the following:

- CAPE is selected for its influence on wind conditions in convective systems, like MCSs;
- topography is a coarse digital elevation map, selected for its role in influencing wind speed and direction;
- land-sea-fraction indicates the ocean-to-land fraction of a coarse grid and influences wind patterns around coastlines;

- surface roughness length provides valuable information about the vertical wind profiles and the physical features at the surface; and
- surface pressure is selected for its role as a pressure gradient force in the surface wind momentum budget.

From the WRF HRCONUS domain, four subregions are identified, each with different climatological conditions. Figure 3.1 shows a representative instance of the wind speed field for the WRF HRCONUS domain, as well as the subregions considered, including:

- the Western region, which covers southern British Columbia, Washington State, and Oregon, is characterized by complex topography that includes mountainous terrain and complex shorelines;
- the Central region, which covers North and South Dakota, as well as Minnesota and northern Iowa, southern Manitoba and southwestern Ontario (not to be confused with the geographical southeastern area of Ontario locally referred to as: "Southwestern Ontario"), has a continental climate with large lakes and relatively frequent fine-scale convective features;
- the Southeast region, which includes Florida, Cuba, and adjacent waters, is subject to tropical cyclones and frequent fine-scale convective features; and
- the Continental region, which includes the entire Central region, southern prairies of Canada, and the northern section of Mexico with patterns influenced by complex topography, shorelines, inland waterways, and fine-scale convective features. The Continental region is evaluated separately from the other regions to preserve consistency.

Each "true" HR and LR variable is standardized to have a mean of zero and a standard deviation of one. The land-sea-fraction covariate is an exception since it is already bounded. The mean of a given variable is calculated over all pixels as:

$$\mu_\theta = \frac{1}{W}\frac{1}{H}\frac{1}{T}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{T}\theta_{i,j,k} \tag{3.8}$$

where $\theta$ represents a variable in $\mathbf{y}$ or $\mathbf{x}$, $T$ is the number of fields or time-steps, $H$ is the height of the fields, and $W$ is the width of the fields in the dataset. The standard deviation can be determined similarly, as

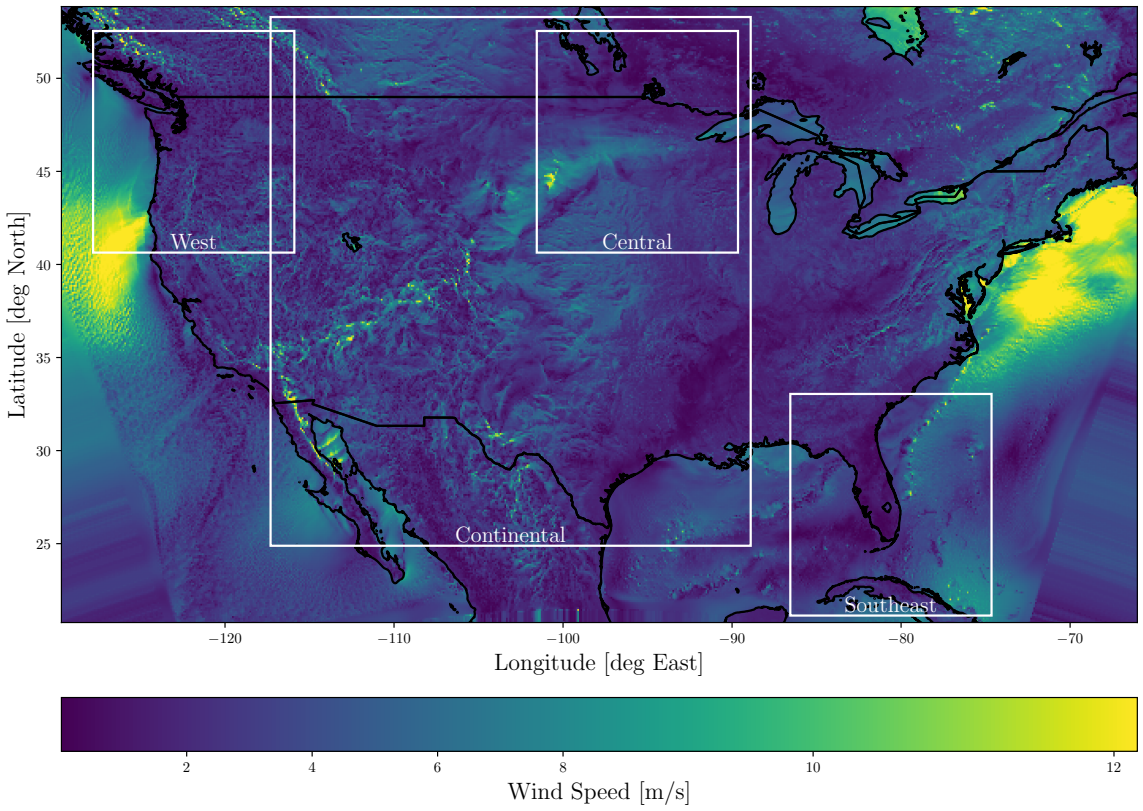


Figure 3.1: Sample wind speed map of WRF HRCONUS for September 28, 2002, at 12:00:00 UTC. Regions selected for this work are in white boxes with their respective names. The southwest and southeast corners of this figure show edge effects caused by the nearest-neighbour interpolation to a regular lat/lon grid.

$$\sigma_\theta = \sqrt{\frac{\frac{1}{W}\frac{1}{H}\frac{1}{T}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{T}(\theta_{i,j,k} - \mu_\theta)^2}{WHT}} \tag{3.9}$$

The standardized set of fields, $\theta'$, can be computed as

$$\theta'_{i,j,k} = \frac{\theta_{i,j,k} - \mu_\theta}{\sigma_\theta} \tag{3.10}$$

Since the models are trained on the standardized variables, the generated HR pixels in, $G(\mathbf{x})$ are de-standardized as:

$$\theta_{i,j,k} = \theta'_{i,j,k}\sigma_\theta + \mu_\theta \tag{3.11}$$

using the respective $\sigma$ and $\mu$ from the "true" HR fields. For notational simplicity, this work will not explicitly distinguish between standardized or de-standardized variables. However, the reader should assume that all variables used while training are standardized, and all results are de-standardized as in Equation 3.11 before they are reported so that they are in the original units of m/s.

## 3.4 Configuring and Training the Super-Resolution Models

### 3.4.1 Frequency Separation

Section 3.2 summarized how the SRGAN framework can be adapted for use with WGAN-GP using Equations 3.4 and 3.6. An important consideration of state-of-the-art SR methods (such as SRGAN) is that they have limited generalization capabilities in practical situations when LR fields suffer from, for instance, aliasing effects, sensor noise, or compression artifacts. This limited capability stems from idealized pairings between the LR and HR fields used for training, as discussed in Section 2, and manifests in the super-resolved HR fields as high-frequency artifacts and distortions (Shocher et al., 2018).

While climate models may not suffer from identical challenges to those that impact LR images in computer vision, LR and HR fields from climate models can contain differences at large scales with one another that impact the generalization capabilities of the SR methods. Fritsche et al. (2019) recognized how non-idealized LR and HR

pairs of images can negatively impact SR methods, and proposed a modification to the objective function that delegated spatial frequencies of the HR fields into a high and low-frequency pair to which the adversarial loss and content loss were applied respectively.

The main idea of *frequency separation* (FS) is to simplify the task of the Critic by recognizing that low-frequency information in the fields is more appropriately evaluated using the content loss (i.e. grid-point error metrics). By focusing only on the high frequencies of the image, the Critic network is no longer tasked with learning potential mismatches in the large scales between the LR and HR fields. Instead, this task is delegated (in the low frequencies) to the content loss. This approach led to perceptually improved results in computer vision with images and is adopted into this present work to examine its influence on wind fields.

More formally, if $\mathscr{L}$ represents a low pass filter, the low-frequency and high-frequency fields are determined with the following:

$$\begin{aligned} \mathbf{y}_h &= \mathbf{y} - \mathscr{L}(\mathbf{y}) \\ G(\mathbf{x})_h &= G(\mathbf{x}) - \mathscr{L}(G(\mathbf{x})) \end{aligned} \tag{3.12}$$

where $l$ and $h$ subscripts indicate low and high-frequency counterparts after applying FS to the HR field.

For this work, $\mathscr{L}$ is set as a spatial averaging kernel. Square matrices are centred on a pixel in the image, and the spatial average of all overlapping pixels is computed. The central pixel value is then replaced with this average. This step pools surrounding pixels and smooths details in the image. More specifically, $\mathscr{L}$ is a 2D averaging kernel applied to the input with stride one (meaning it is applied to each pixel successively). Alternatively, the smoothing of the 2D averaging kernel can be viewed as a convolution operation using a square filter, with each weight set as $1/N$.

The total number of elements in the square averaging kernel with shape $N \times N$ influences the extent of smoothing. As a result, $N$ should be considered a tuneable hyperparameter in the GAN configuration, with values considered for this present work summarized in Table 3.1. Kernels with smaller values of $N$ allow higher frequency information through the filter, and vice versa. The effect of $N$ on image smoothing is shown in Figure 3.2.

For fields with FS (designated with $\mathscr{L}_N$), a reflection pad corresponding to the

half-width of the averaging kernel is applied to preserve the same size between the low-frequency and original fields. Reflection padding extends the boundaries of the image by mirroring pixels close to the edge of the boundary. It offers a more continuous boundary for smoothing than, for example, zero padding.

It should also be noted that by separating these frequencies, the Wasserstein distance is now computed on a different probability distribution that contains the high frequencies only. These modified samples can be noted as $\mathbf{y}_h \sim \mathbb{P}_{r,h}$ for high frequency "true" fields, and $G(\mathbf{x})_h \sim \mathbb{P}_{g,h}$ for high frequency generated fields. In the objective function of the Critic, this becomes

$$\mathcal{L}_C = \mathbb{E}_{G(\mathbf{x})_h \sim \mathbb{P}_{g,h}}[C(G(\mathbf{x})_h)] - \mathbb{E}_{\mathbf{y}_h \sim \mathbb{P}_{r,h}}[C(\mathbf{y}_h)] + \lambda \mathbb{E}_{\hat{\mathbf{x}}_h \sim \mathbb{P}_{\hat{\mathbf{x}}_h}}[(||\nabla_{\hat{\mathbf{x}}_h} C(\hat{\mathbf{x}}_h)||_2 - 1)^2] \quad (3.13)$$

with $\hat{\mathbf{x}}_h$ representing random samples from the high frequency fields only, i.e. $\hat{\mathbf{x}}_h = \epsilon \mathbf{y}_h + (1-\epsilon) G(\mathbf{x})_h$. Similarly, for the Generator,

$$\mathcal{L}_G = \underbrace{-\mathbb{E}_{G(\mathbf{x})_h \sim \mathbb{P}_{g,h}}[C(G(\mathbf{x})_h)]}_{\text{Adversarial component}} + \underbrace{\alpha\, \mathbb{E}_{\mathbf{y} \sim \mathbb{P}_r}\, l_c(\mathscr{L}_N(\mathbf{y}), \mathscr{L}_N(G(\mathbf{x})))}_{\text{Content loss}} \quad (3.14)$$

### 3.4.2 Training the Super-Resolution Models

The models are trained using a single NVIDIA GTX 1060 GPU with 6 GB of VRAM and with training hyperparameters summarized in Table 3.1. Many of the hyperparameters here are taken directly from existing work, except for $\alpha$ which is discussed in more detail later in this section.

GPUs help optimize DL tasks, and their utilization allows the potential of DL methods to be realized for fast climate impact assessments. The Adam optimizer, a form of stochastic gradient descent (SGD) (Kingma and Ba, 2017), is used to train the models. As discussed in Section 3.2, the Critic in WGAN-GP requires a number of updates for each update of the Generator. Following Gulrajani et al. (2017), this present work updates the Critic *five* times for each update to the Generator. Each GAN takes approximately 48 hours to complete 1000 passes (i.e., epochs) through the entire training set.

As discussed earlier, $\alpha$ weighs the importance of the content loss relative to the adversarial component. While the magnitude of $\mathcal{L}_G$ is unimportant for its optimization, the relative magnitudes of the adversarial component and the content loss are

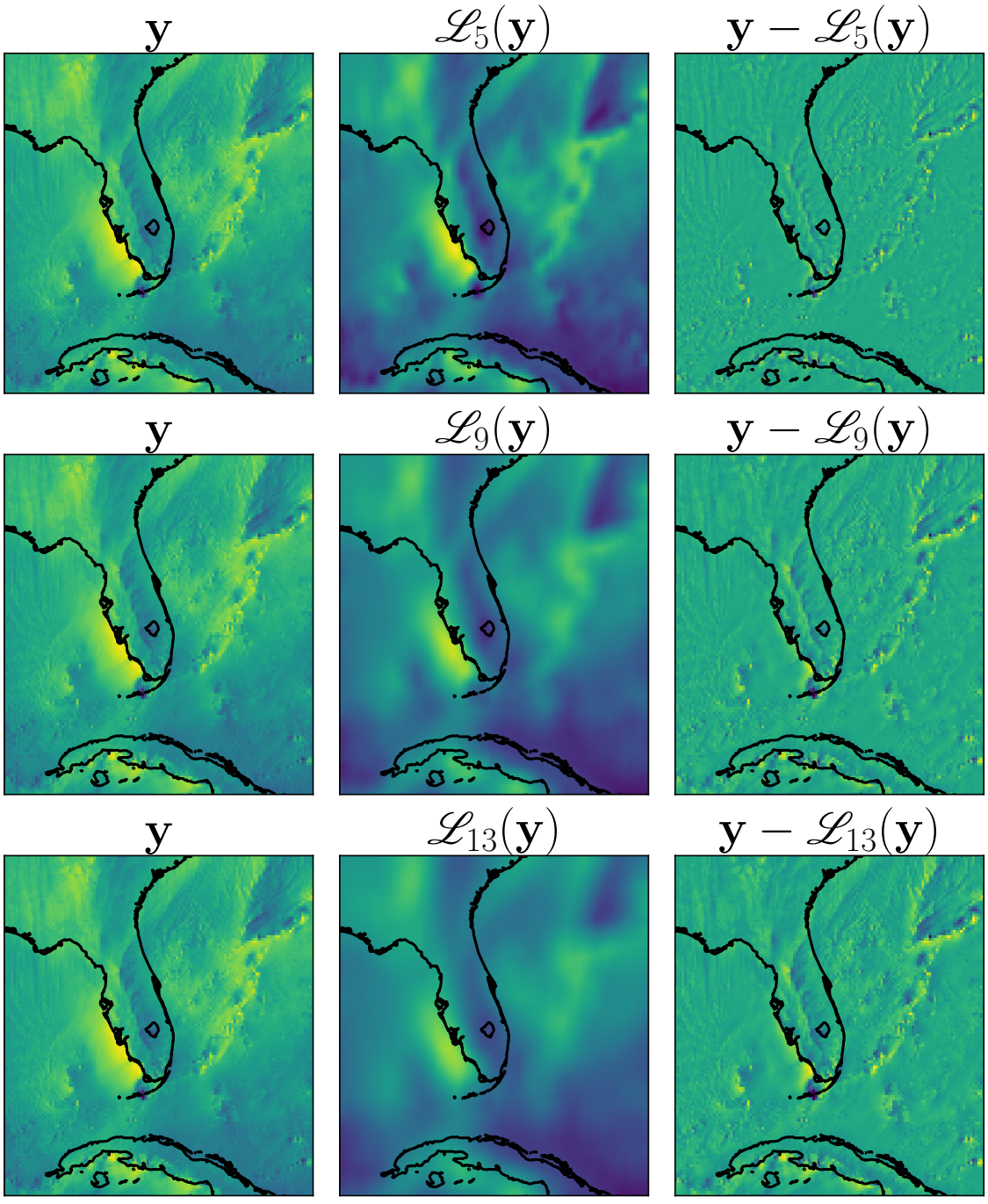


Figure 3.2: Various examples of frequency separation applied to a $u10$ field in the Southeast region. Each row has a different configuration of the filter designated with $\mathscr{L}_N$. The left panel is the unaltered input, the center panel has an average 2D pool filter applied to the input, and the right panel shows the high-frequency field determined by the difference between the left and center panels.

| | Training Hyperparameters | | | | | Adam Optimizer | | | Frequency Separation Avg2DPool | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\lambda$ | $\alpha$ | Epochs | Batch Size | Critic iterations | Learning rate | $\beta_1$ | $\beta_2$ | Filter size ($N$) | Stride | Reflection Padding |
| FS GANs | 10.0 | 500 | 1000 | 64 | 5 | $1 \times 10^{-5}$ | 0.9 | 0.99 | 5, 9, 13 | 1 | 2, 4, 6 |
| NFS GANs | 10.0 | 500 | 1000 | 64 | 5 | $1 \times 10^{-5}$ | 0.9 | 0.99 | **NA** | **NA** | **NA** |
| Pure CNNs | 10.0 | 1 | 1000 | 64 | 5 | $1 \times 10^{-5}$ | 0.9 | 0.99 | **NA** | **NA** | **NA** |

Table 3.1: Summary of hyperparameters.

important. The value chosen for this work is selected such that the content loss and raw output from the Critic are roughly the same magnitudes. Although no optimal values of $\alpha$ are searched for in this present work, the value used here – summarized in Table 3.1 – results in stable training and good performance. Future work will require further investigation of these hyperparameters and this topic will be revisited in Section 5.

For each of the regions, three different values of $N$ are used for $\mathscr{L}_N$ to test various amounts of FS smoothing. These GANs will be collectively referred to as FS GANs and are optimized using Equations 3.13 and 3.14. Additionally, a no-frequency separation GAN (NFS GAN, optimized using Equations 3.4 and 3.6) and pure CNN are also trained for each region. The pure CNN is configured identically to the NFS GAN, however, it is optimized using only the content loss from Equation 3.6 (i.e. the adversarial loss is excluded from the objective function). The pure CNN model is meant to test the role of the adversarial loss on the generated wind fields, and it can be viewed as an extreme case of FS whereby *all* frequencies are delegated to the content loss, and no frequencies are delegated to the adversarial loss. The reader should note that the content loss in Equations 3.6 and 3.7 is applied to all pixels in the HR fields – covering all frequencies – and is the form frequently used (as either the MAE or MSE) in SR in both the computer vision and climate science fields (Ledig et al., 2017; Sha et al., 2020; Stengel et al., 2020; Harris et al., 2022).

## 3.5 Critic and Generator Networks

The Critic is adopted from the SRGAN Discriminator of Ledig et al. (2017), but with some modifications. The batch normalization layers were omitted entirely, as their removal is required for the gradient-penalty term to operate on individual fields rather than batches (Gulrajani et al., 2017). For this work, removing batch normalization is empirically found to stabilize training. Another modification is made to the final linear layer of the SRGAN Discriminator/Critic architecture. Due to GPU

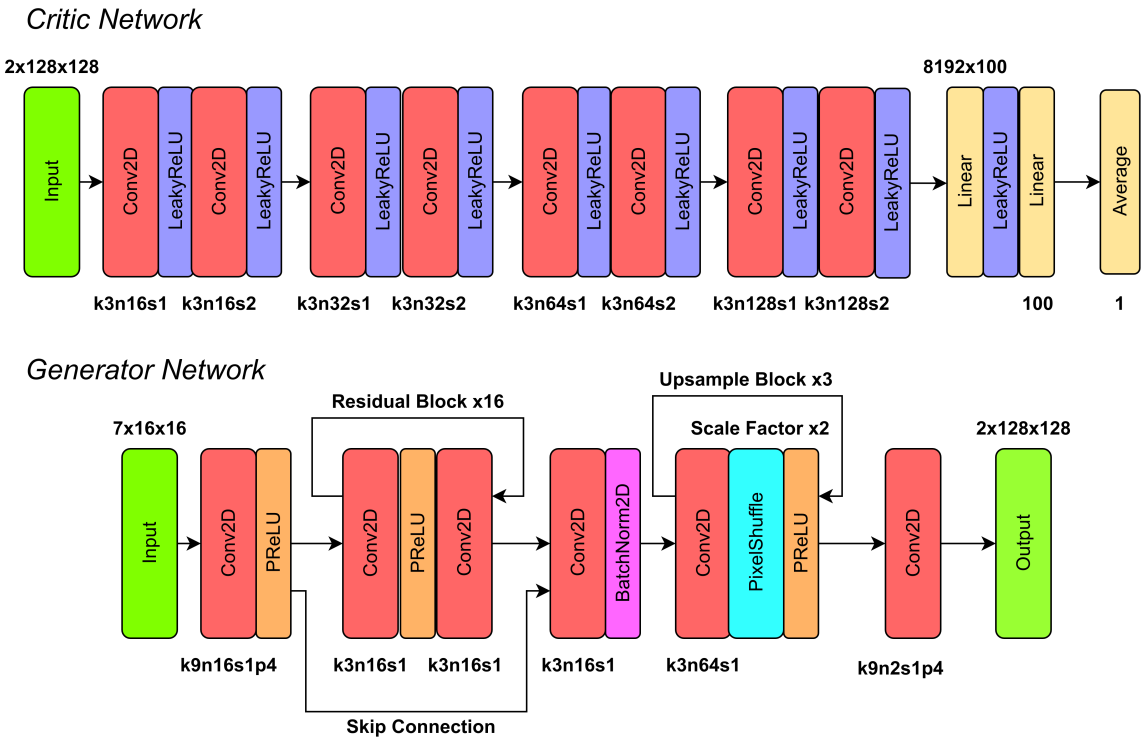


Figure 3.3: Outline of the networks used with WRF HRCONUS (Critic) and ERA (Generator) with WGAN-GP. The inputs to each network are two-dimensional fields. The label for each Conv2D layer describes its function. $k$ indicates the size of the square kernel/filter used, e.g. a kernel size of 3 has 9 trainable parameters (plus one if a bias parameter is fit for that particular filter). $n$ represents the total number of filters in the layer, and $s$ is the stride at which it is applied to the input. Where relevant, the shape of matrices/vectors is included on top of the input and linear layers to indicate the dimensions of the data representations at these stages.

memory constraints, in order to maintain a batch size of 64, the number of linear layer weights is reduced from 1024 in the original SRGAN implementation to 100 in this implementation.

The Generator is very similar to the SRGAN Generator but is adjusted to accept additional covariates and an additional upsampling block that downscales the fields by a scale factor of eight. A diagram of each network is shown in Figure 3.3. The various software packages that implement the CNNs used for SR with WGAN-GP are summarized in Appendix A.4, along with a link to the code repository developed to produce the results in this thesis.

# Chapter 4

# Results

## 4.1 Visual Quality of Generated Fields

A representative set of the $u10$ and $v10$ wind field maps produced using the SR models on October 5th, 2000 at 12:00 UTC are shown in Figure 4.1 and 4.2 respectively. For this particular time-step, Tropical Storm Leslie (a subtropical cyclone forming over eastern Florida) is prominent in the Southeast region; inland features (such as lakes) are found to influence the wind patterns in the Central region; and strong land-sea contrast is shown in the West region with clear evidence of the influence of topography on the flow.

While there are broad consistencies at large scales between WRF HRCONUS and ERA-Interim, differences in the locations of certain structures are present in the fields. For example, the negative $u10$ wind feature in the northeast part of the Southeast region is oriented slightly differently in WRF HRCONUS and ERA-Interim.

The CNN does not represent the fine-scale variability seen in the WRF HRCONUS field for the Southeast and Central region. This fact is most obvious in the Southeast region where fine-scale convective features are not produced by the model and the results appear too smooth. As a pixel-based metric, the MAE alone does not allow the CNN to "hallucinate" fine-scale features. This effect can also be observed in the Central region, although to a lesser extent. The West region shows generally good agreement between the CNN and WRF HRCONUS, in particular with the inland topographical features. However, the CNN for the West region is lacking some of the sharp and well-defined topographical details found in WRF HRCONUS. An interpretation of these CNN results will be revisited in detail in Section 5.

GANs with and without FS show little perceptual difference for the West and Central regions; both show an improvement in fine spatial structure over the pure CNN. In the GAN with $\mathscr{L}_5$ FS for the Southeast region, a noticeable reduction in power in the medium and low spatial frequencies surrounding organized spatial structures in the northeast can be seen (e.g., as illustrated for $u10$ in Figure 4.1 and for $v10$ in Figure 4.2). This effect can be seen as isolated fine-scale features in the $\mathscr{L}_5$ FS GANs when compared to WRF HRCONUS.

GANs with $\mathscr{L}_9$, $\mathscr{L}_{13}$, and no FS (NFS) show little perceptual difference in quality in the realizations. Importantly, these realizations demonstrate that despite each GAN being conditioned on the same coarse ERA-Interim fields, they produce fine-scale structures at different locations than the WRF HRCONUS, and are placing fine-scale weather where it could be, but not necessarily exactly where it is in the WRF HRCONUS fields.

This differing placement of fine-scale structures in the GANs highlights a desirable consequence of using the adversarial loss in $\mathcal{L}_G$, which does not penalize grid-point errors. Furthermore, the increase in the perceptual quality of the GAN-generated fields can be attributed to the Critic's ability to capture the realism of the HR fields. A more extensive set of example patterns can be found in Figures B.2 to B.7.

## 4.2 Evolution of Performance Metrics While Training

Several metrics were monitored during the training of the network configurations used here. Among them are the MAE, MSE, Multi-scale Structural Similarity Index (MS-SSIM), and the Wasserstein distance as approximated by the Critic. MS-SSIM is a metric taken from the computer vision field to objectively measure the quality of images across multiple scales (Wang et al., 2003). A more detailed discussion of MS-SSIM is included in Appendix A.2.

Figure 4.3 shows the evolution of the MAE and the Wasserstein distance on the test sets during the training process while Figure B.1 shows the training evolution on both the test and train sets for MAE, MSE, MS-SSIM, and the Wasserstein distance. $v10$ and $u10$ fields are combined for each of these pixel-wise metrics, so only a single value is calculated for both wind components. One important note about the metrics used is that the MAE, MSE, and MS-SSIM are calculated over all frequencies between

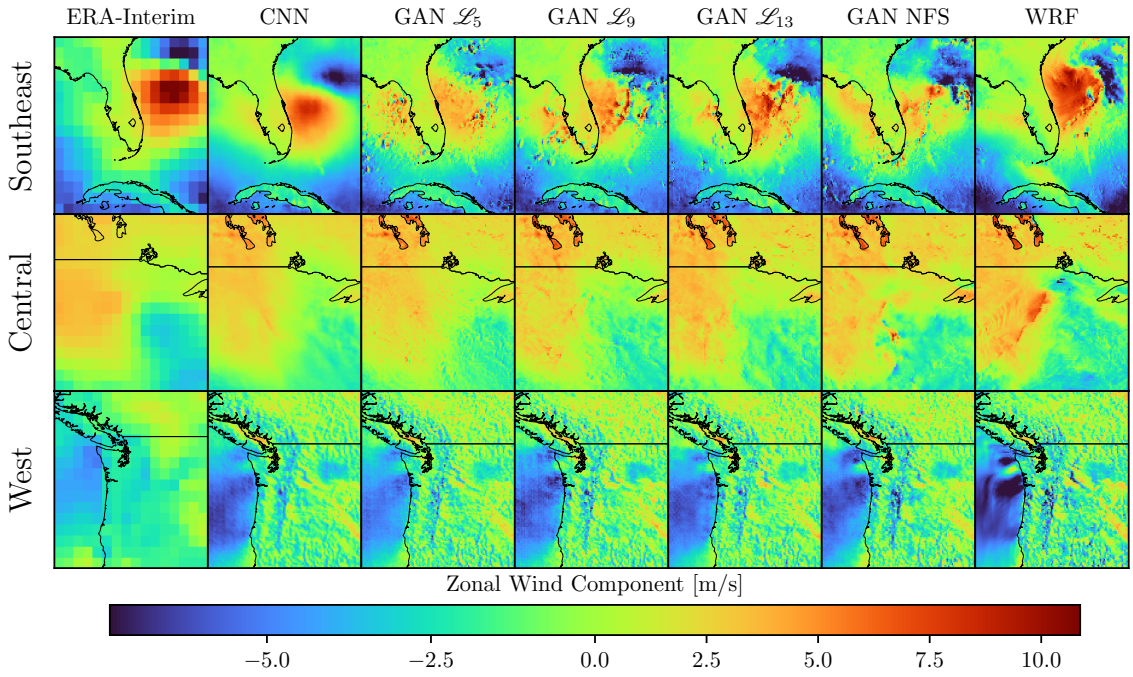


Figure 4.1: Example fields of $u10$ only. Each field is the same time step, on October 5th, 2000 at 12:00:00 UTC. The top row shows the Southeast region, the middle row is the Central region, and the bottom row is the West region. From left to right: the corresponding $u10$ field from LR ERA-Interim, the pure CNN, FS GAN with a $\mathscr{L}_N$ kernel, NFS GAN, followed by WRF HRCONUS. As can be seen in these panels, fine-scale features are not exactly co-located between the GANs as well as the WRF HRCONUS, demonstrating how the GANs generate features that could occur at the fine scales.

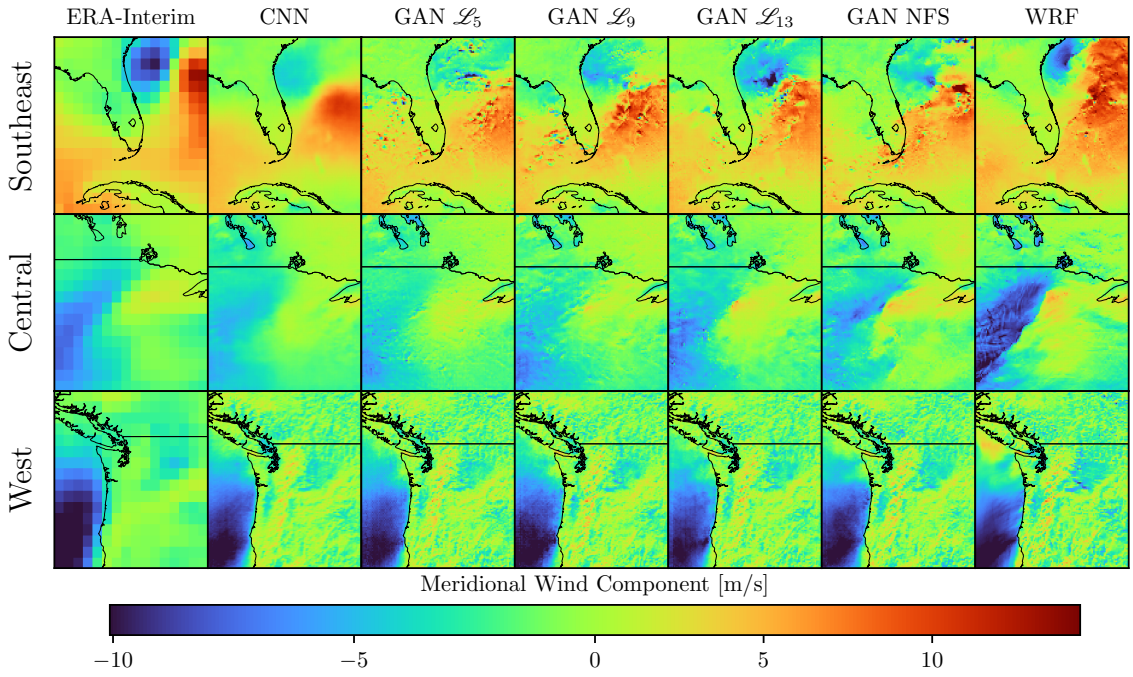


Figure 4.2: Example fields of $v10$ only. Each field is the same time step, on October 5th, 2000 at 12:00:00 UTC. The top row shows the Southeast region, the middle row is the Central region, and the bottom row is the West region. From left to right: the corresponding $v10$ field from LR ERA-Interim, the pure CNN, FS GAN with a $\mathscr{L}_N$ kernel, NFS GAN, followed by WRF HRCONUS. As can be seen in these panels, fine-scale features are not exactly co-located between the GANs as well as the WRF HRCONUS, demonstrating how the GANs generate features that could occur at the fine scales.

pairs of generated and "true" fields, while the Wasserstein distance is approximated between the generated batch and the "true" batch using the Critic. Furthermore, the estimate of the Wasserstein distance is not the same as the value of $\mathscr{L}_C$ in Equation 3.4 or 3.13, rather, it is estimated using the form of Equation 3.2 without the gradient penalty as

$$\mathbb{E}_{\mathbf{y}\sim\mathbb{P}_r}[C(\mathbf{y})] - \mathbb{E}_{G(\mathbf{x})\sim\mathbb{P}_g}[C(G(\mathbf{x}))] \tag{4.1}$$

Individual curves in the Wasserstein distance evolution shown in Figure 4.3 cannot be compared, since applying FS changes the distributions to which the Wasserstein distance estimation is applied. Furthermore, the Wasserstein distance here is an approximation, and so its shape is what deserves attention, rather than the actual values. As such, it has been re-scaled as a ratio of the initial distance encountered for the particular training configuration. Importantly, the Wasserstein distance evolution shows that there is no overfitting of this metric on the test data for any region or GAN, which demonstrates the stability of the WGAN-GP framework even after many epochs.

Conversely, while the MAE reaches a minimum value in the test set after ∼200 epochs for the Southeast and Central regions, the West region does not. This is indicative of overfitting of the Generator in these two regions since no minimum is found in the evolution of MAE on the training set (Figure B.1). At late epochs, the test set MAE does not grow substantially, so overfitting of the MAE is minimal. Interestingly, this result is only present in pixel-wise metrics, not the Wasserstein distance. The slightly larger MAE at later epochs for the Southeast and Central regions may be indicative of the Generator learning to correct large-scale differences or biases specific to the training set only, while not generalizing to the test set. The topic of large-scale differences and biases in variability will be discussed in more detail in Section 4.5.

For the evolution of the MAE, (also in the evolution of MSE, and MS-SSIM to some extent as well, e.g., Figure B.1), there is a clear ordering of performance of the different SR models that is robust across the regions. As shown in Figure 4.3, the best performing model in the MAE sense is the pure CNN, followed by the $\mathscr{L}_5$, $\mathscr{L}_9$, and $\mathscr{L}_{13}$ FS GANs, and finally the NFS GAN shows the largest MAE. This ordering can be explained by examining the role of the components of the Generator's objective function in Equation 3.14.

For example, in pure CNNs, there is no adversarial loss, and the network's weights and parameters are optimized solely to minimize the MAE. For the FS GANs, the amount of smoothing dictates the range of frequencies that the content and adversarial loss will receive and subsequently dictates how the networks are optimized. When there is less smoothing, the optimization problem tends to be more similar to the pure CNNs since a large range of frequencies are delegated to the content loss. This explains why the $\mathscr{L}_5$ kernel has the lowest MAE, and why including any FS leads to a lower MAE than the NFS GAN.

For the NFS GAN, the adversarial loss component of Equation 3.6 encounters the full range of frequency scales of the fields; it is not limited to reproducing pixel-wise variability of the WRF HRCONUS exactly. In a sense, NFS GAN has more freedom to conditionally generate variability across scales. For conditional image generation – where the "hallucination" of fine-scale weather should be encouraged – this increases the MAE because of the double penalty problem.

Additionally, when there is more smoothing in the FS GANs, a larger range of frequencies are delegated to the adversarial loss, which allows for the generation of fine-scale weather across a larger range of frequencies. This explains why the $\mathscr{L}_{13}$ kernel has the highest MAE of the FS GANs for each region. More fundamentally, the ranking shown in Figure 4.3 demonstrates how the double penalty problem is linked to the generated fine-scale features that are encouraged by the adversarial loss. A more detailed discussion on FS is included in Section 5.

## 4.3 Climatological Statistics

### 4.3.1 Marginal Distributions

The marginal distributions determined over all test set time and space points of the $u10$ and $v10$ components, as well as wind speed, are shown in Figure 4.4 for each training configuration. For both wind components, the overall shape, extremes (1st and 99th percentiles), 1st and 3rd quartiles, as well as median show reasonable agreement between the "true" and the trained models, especially between the 1st quartile, median, and 3rd quartile. These distributions reveal that all the SR models have small biases in the variance of the wind components that lead to a more pronounced negative bias in the quartiles and 99th percentile when the components are transformed non-linearly to the wind speed. The CNN produces $u10$ and $v10$ fields that

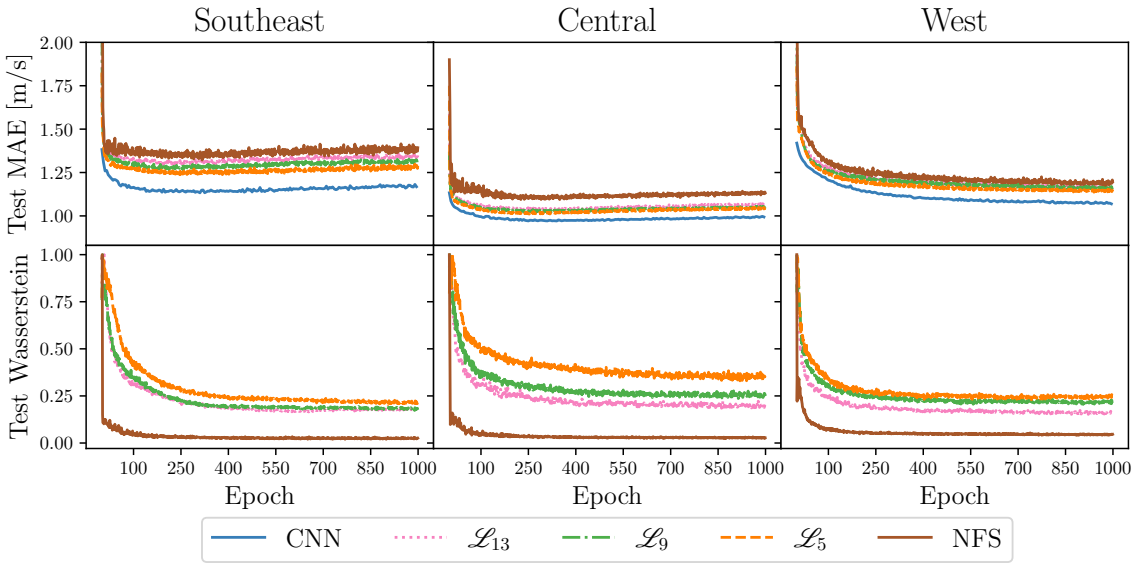


Figure 4.3: Evolution of the MAE (top row) and the Wasserstein distance (bottom row) for 1000 epochs on the test data. Each SR model is indicated in the legend. The MAE is calculated pixel-wise for all pixels combined from the $u10$ and $v10$ HR fields. Note that no Wasserstein distance is calculated for the CNN baseline since no Critic network is trained to calculate it. The Wasserstein distance estimate is scaled by the initial distance.

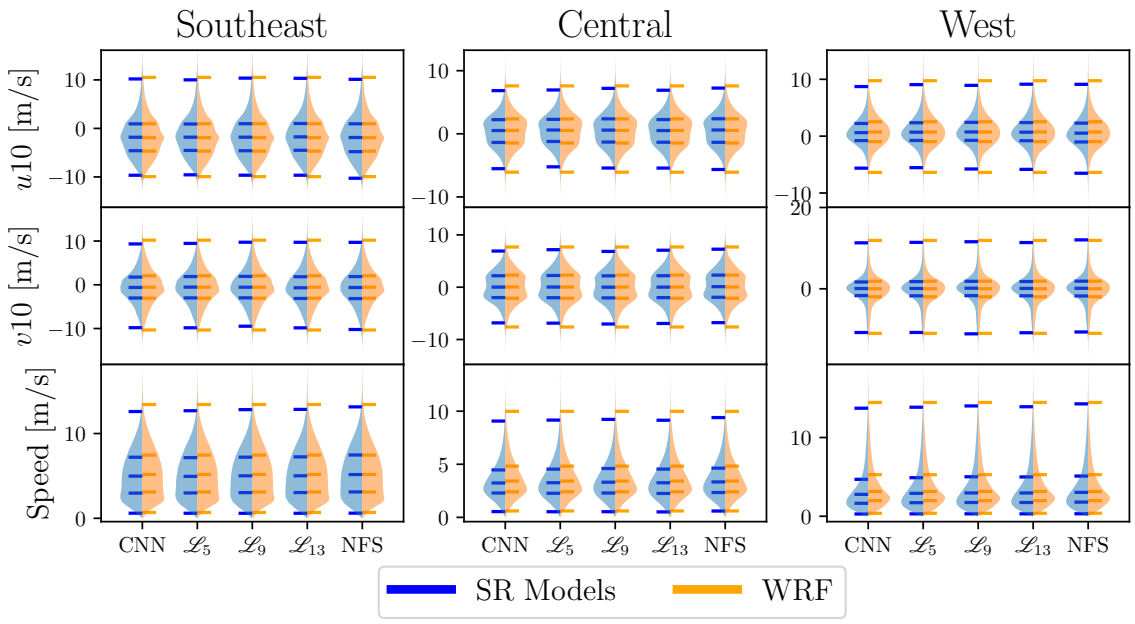


Figure 4.4: Violin plots of each pixel in the test set for each region and model. Each distribution (or "violin") is split vertically with the SR model on the left, and WRF HRCONUS on the right. From lowest value to largest value, horizontal ticks indicate the 1st, 25th, 50th, 75th, and 99th percentile values from the test sets of their respective distributions.

have a low-variance bias due to the lack of high spatial frequency information in the fields. Conversely, the GAN fields generate more fine-scale variability, so the low-variance bias is smaller. The quality of the spatial structure of the generated fields, however, is not shown in these marginal distributions.

### 4.3.2 Maps of Biases

The panels of Figure 4.5 show maps of biases between the WRF HRCONUS and generated fields for the temporal mean, standard deviation, and the 90th percentile of the wind speed for the Central region test set. Wind speed offers a more stringent test of the biases since it is a non-linear function of the wind components.

Good agreement is found between the statistics of the generated fields from the SR models and WRF HRCONUS, with a spatial mean bias $< 1$ m/s for each panel, and each statistic. Figure 4.5 shows that the spatial mean bias is consistently smallest in the NFS GAN. However, more importantly, the panels reveal systematic patterns of biases in the statistics of each grid.

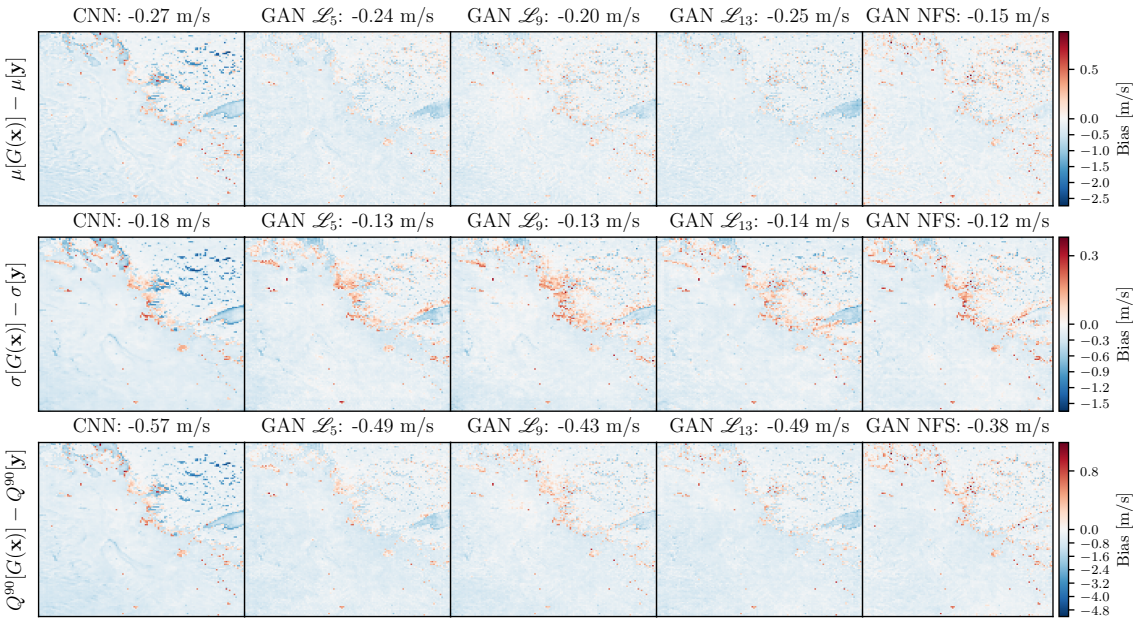


Figure 4.5: Map of the Central region wind speed bias in the mean (top row), standard deviation (middle row), and 90th percentile (bottom row) for each SR model. Statistics are calculated along the temporal dimension of the test set. The spatial mean of the bias is reported in the title of each panel.

Land features and coastlines can be clearly identified in the maps, indicating a bias associated with these features in the field. These identifiable land features, such as the western tip of Lake Superior (and smaller lakes to the north) are more pronounced in the CNN model (the left-most column of Figure 4.5) and tend to be less prominent in the NFS GAN, with each FS GAN showing improvement over the CNN. This result can be attributed to the role of the adversarial loss in reducing the low-variance bias of the generated fields, especially for these fixed features.

Just as in Figure 4.4, a negative bias in the spatial mean can be seen in all the wind speed panels, and can be explained by a low-variance bias of the generated wind components (rather than biases at particular locations) under the transformation from components to speed. Similar results were found for the Southeast and West region (Figures B.8 and B.9).

## 4.4 Power Spectra

As noted previously, quantifying the perceptual quality of the generated fields is challenging, given that the fine-scale features of the SR fields are not necessarily captured by the MAE and MSE, and those that perform well on these metrics show larger biases in the fields and lower qualitative fidelity. The HR spatial patterns reflect spatial correlation structures that can instead be naturally measured using the power spectra of the HR fields.

For wind fields, measures of spatial correlation through power spectra can be used to evaluate the trained models as done previously for wind field SR (Singh et al., 2019; Stengel et al., 2020; Kashinath et al., 2021). In the wind components, large-scale (low spatial frequency) processes have large power at small wavenumbers, while small-scale (high spatial frequency) processes have low power at large wavenumbers. Each wind component for each time step has its own power spectrum which is then averaged to produce a final mean spectrum shown in Figure 4.6. A more detailed discussion of the power spectrum methodology is provided in Appendix A.3.

Figure 4.6 shows the power spectra of each SR model, including WRF HRCONUS as a reference field. At large scales, each model generally agrees with the WRF HRCONUS spectrum; however, the CNN is clearly lacking power at the small scales, which is consistent with the qualitative assessment of the field realizations in Section 4.1. The reduction of small-scale power is owed to the missing adversarial loss in the pure CNN's objective function since the adversarial loss is responsible for generating fine-scale weather.

The power reduction in the pure CNN at small scales is more pronounced in the Central and Southeast regions where fine-scale convection is more common. The NFS GAN consistently provides the closest match to the WRF HRCONUS spectrum for each region and wind component, with the FS GAN's spectra falling between those of the CNN and NFS GAN. Here, the consequences of changing the optimization problem with FS become more apparent. The $\mathscr{L}_5$ GAN delegates a larger portion of the field's frequencies to the MAE, and so it tends to match the CNN spectrum at large to medium scales. However, it "breaks away" from the CNN curve at the frequency cutoff point (dictated by the size of the FS kernel) where it becomes more consistent with the small-scale structures found in the NFS GAN and WRF HRCONUS.

The high-spatial frequency details are delegated to the adversarial loss for the NFS GANs, which leads to this variability being more accurately produced in these

models. The "breaking away" from the WRF HRCONUS spectra can be clearly seen in Figure B.10 by taking the ratio relative to WRF HRCONUS spectra.

This behaviour is robust across each region for each wind component. As the cutoff frequency changes as the FS kernel size increases, the range over which the respective NFS GAN more closely matches the CNN shifts to larger scales. This shifting to match larger scales is a robust result shown for both the $\mathscr{L}_9$ and $\mathscr{L}_{13}$ kernels for each region and wind component.

Furthermore, the power spectra of each SR model show a reduction in the energy across all scales – particularly visible in the Central region. This scale-wide reduction in energy is another representation of the observed biases of the variance discussed earlier. As will be discussed in the next section, as well as Section 5, the cause of this bias across scales can be attributed to differences in the large scales between ERA-Interim and WRF HRCONUS. Most importantly, Figure 4.6 reveals that the generated fields more accurately match WRF HRCONUS when the adversarial loss is provided with the full range of frequencies, despite the increase in pixel-wise metric errors when doing so. The GAN SR models, therefore, produce the right amount of high-frequency variability but do not necessarily place the corresponding features in the exact location as in WRF HRCONUS (i.e., it puts them where they could occur).

## 4.5 Idealized Covariates

The power spectra are a valuable tool for evaluating the relative performance of the generated fields. They provide scale-dependent information about variability biases between the WRF HRCONUS and DL models, as shown in the FS experiments and pure CNNs. For example, large-scale differences are present between ERA-Interim and WRF HRCONUS due to the CPM's freedom to produce fine-scale features. In other words, the upscale energy transfer from the fine scales results in a mismatch at coarse scales between ERA-Interim and WRF HRONCUS. This mismatch is a clear demonstration that large-scale processes are not fully independent of small-scale processes.

To test the extent to which differences between ERA-Interim and WRF HRCONUS may impact the performance of GANs, two additional NFS GANs were trained using just the $u10$ and $v10$ fields as coarse covariates. One of these GANs is conditioned with ERA-Interim, just as before but without the additional covariates, and the other uses artificially coarsened (by a scale factor of eight) WRF HRCONUS $u10$ and $v10$

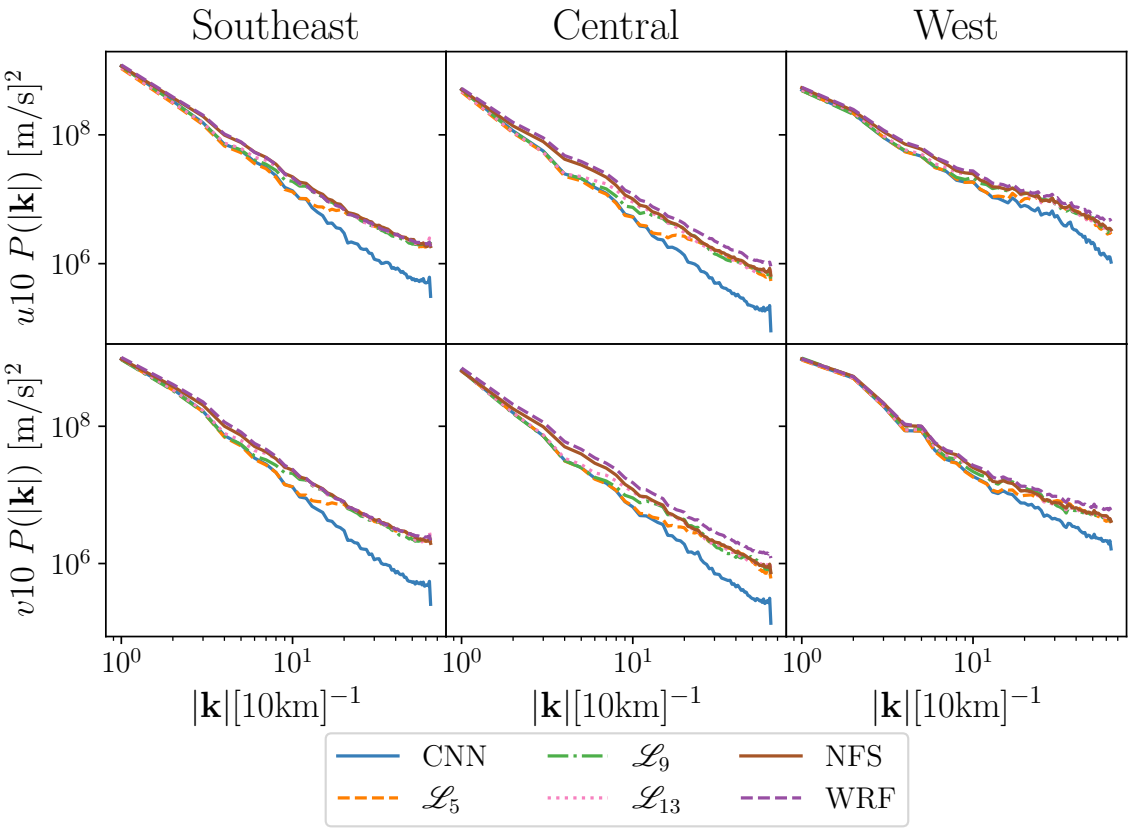


Figure 4.6: Power spectra of the SR models and WRF HRCONUS represented as lines. The top row shows the $u10$ component, and the bottom row shows the $v10$ component. Each column is a separate region.

HR fields. The resulting "coarsened WRF" LR fields are unbiased at large scales with the HR WRF HRCONUS fields. The latter configuration simulates approaches common to both the computer vision and climate literature (Ledig et al., 2017; Singh et al., 2019; Sha et al., 2020; Leinonen et al., 2020; Stengel et al., 2020; Kumar et al., 2021; Wang et al., 2021).

Although not explicitly shown in this work, idealizing the coarse covariates resulted in a training evolution of the MAE and MSE without signs of overfitting. The MAE and MSE simply plateau at late epochs. This result is in contrast to the overfitting seen in Figure 4.3 and Figure B.1 and supports the hypothesis that the SR models are overfitting large-scale differences between ERA-Interim and WRF HRCONUS in the training set only.

Figure 4.7 shows the power spectra ratio (relative to WRF HRCONUS) of $u10$ and $v10$ wind fields with the idealized GAN (GAN with Coarsened WRF), the GAN with just $u10$ and $v10$ ERA-Interim covariates (GAN with ERA), and the NFS GAN from earlier with all seven covariate fields (GAN with ERA +).

For GANs that use the ERA-Interim covariates, at small wavenumbers $|k|$, there is a difference in the power compared to WRF HRCONUS. Namely, there is less power (i.e. a "low-power bias") at these scales when compared to the coarsened WRF covariate GAN.

This result is due to large-scale differences between WRF HRCONUS and ERA-Interim and is shown clearly in Figure 4.7 where the low-power bias at small wavenumbers almost entirely vanishes in the coarsened WRF GAN. At high frequencies, there is less of a difference seen between the coarsened WRF GAN and ERA GAN with just $u10$ and $v10$.

Interestingly, by including all seven covariates (GAN with ERA + in Figure 4.7), there is a minor improvement in this bias at large scales, but more importantly, there is a significant improvement in the small scales. This demonstrates that the additional covariates are improving the Generator's ability to reconstruct realistic high spatial frequency information. This is a robust result seen in both the $u10$ and $v10$ fields for each region.

## 4.6 Covariate Sensitivity of GAN Spectra

As shown in Figure 4.7, covariates play an important role in the generation of realistic fine-scale features of the generated fields. To further explore this result, an experi-

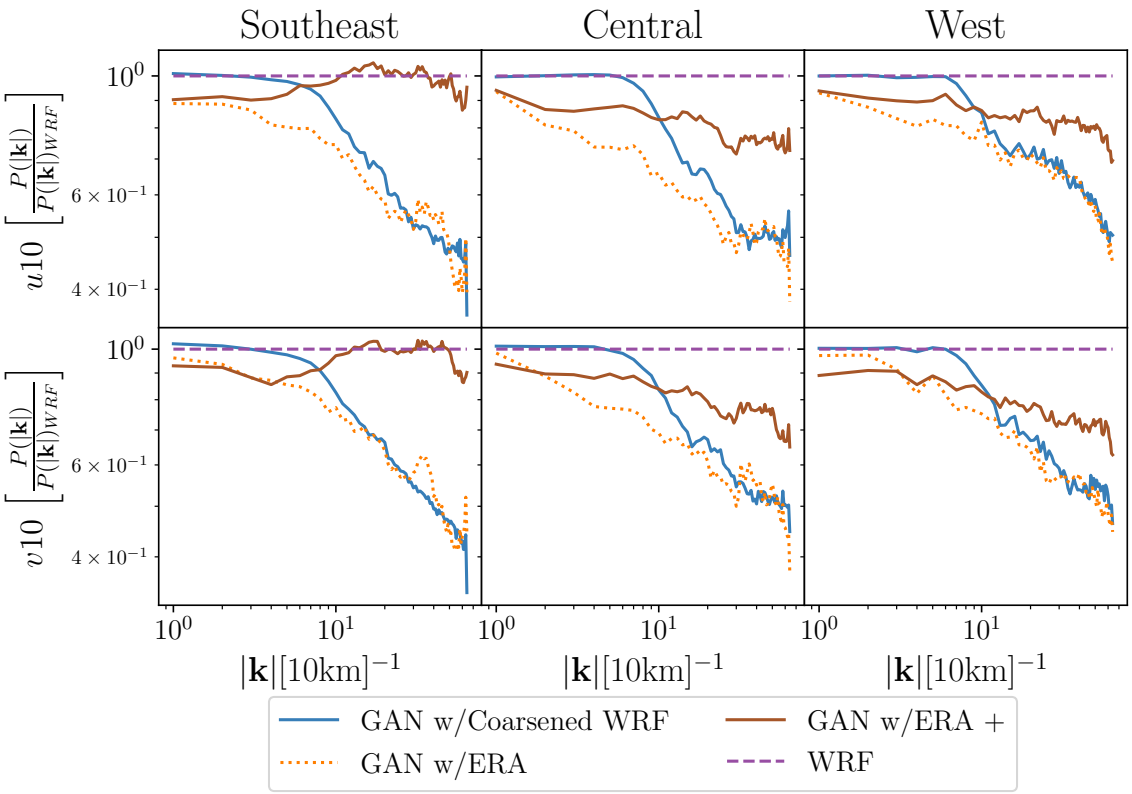


Figure 4.7: Ratio of power spectra of the SR models with idealized covariates, as well as the NFS GAN with all covariates (GAN with ERA +) represented as different lines. The top row shows the $u10$ component, and the bottom row shows the $v10$ component. Each column is a separate region.

ment is devised to perturb four covariate fields of the already-trained NFS GAN, and measure across wavenumbers the resulting changes in the spectra. Since perturbing time-invariant fields results in no changes, only $u10$, $v10$, surface pressure, and CAPE are isolated in this way. The experiment proceeds:

1. From the test set, isolate a covariate to modify and replace each time step with a randomly sampled one from somewhere else in the test set. This modifies the selected covariate to have six correct covariate fields, and one randomly sampled one from another time-step in the set.

2. Generate HR fields with the modified covariate input and compute the power spectrum for each realization in the test set. Simultaneously compute an unmodified power spectrum of the field as a baseline.

3. Repeat Steps 1 and 2 25 times to generate many HR fields with randomly sampled covariates.

4. For each wind component, as well as each generated HR field with randomly sampled covariates, calculate the relative differences, $RD$, between the power spectra of the modified, $P_s(\mathbf{k})$, and unmodified baseline, $P_b(\mathbf{k})$ as

$$RD(P_s, P_b) = \frac{\log P_s(\mathbf{k}) - \log P_b(\mathbf{k})}{\log P_b(\mathbf{k})} \tag{4.2}$$

5. Compute the variance in $RD$ as $\widehat{\mathrm{Var}}\left[RD(P_s, P_b)\right]$ at each wavenumber as a measure of model sensitivity.

6. Repeat Steps 1 - 5 for each isolated coarse covariate field.

The resulting variance at each wavenumber for each perturbed covariate is shown in Figure 4.8.

The methodology reveals the role that the covariates play for each region, and how a region's climatology dictates the sensitivity of the covariates. For example, $u10$ and $v10$ wind components are most sensitive to changes to the respective coarse $u10$ and $v10$ covariates, especially at large scales. This is not surprising given that the large scales in WRF HRCONUS are synchronous with ERA-Interim, and so the networks are making direct use of this large-scale information. This same result can be seen for each region and each wind component.

Interestingly, the degree to which each wind component is sensitive to the other (i.e. the sensitivity of HR $v10$ fields to LR $u10$ fields, and vice versa) is less than that seen for some of the other covariates, such as CAPE in the Southeast region. This particular result is expected since CAPE is highly correlated with convective processes that dominate the high spatial frequencies of the generated wind components (Houze Jr, 2004) in the Southeast region where fine-scale convective features are common. Surface pressure also plays a moderate role in the Southeast region, possibly correlating with weather systems that are accompanied by small-scale variability caused by squall lines or multi-cell storms.

For the West region, the generated fields are most sensitive to the $u10$ and $v10$ coarse covariates but are not sensitive to CAPE and surface pressure. This result can also be understood in the context of the West region's climatology, where convective storms are rare. Instead, strong influences of land-sea boundaries and complex topography are better predicted by the coarse $u10$ and $v10$ fields themselves.

The Central region shows sensitivity to $u10$, $v10$, and CAPE. Like the Southeast region, convective features are also common in the Central region, which explains the observed sensitivity to CAPE. The frequency of convective systems in WRF HRCONUS in the Central region may not be quite as high as in the Southeast region, which could explain why there is a slightly higher relative sensitivity in the Southeast region.

## 4.7 Continental Scale Super-Resolution

The size of the regions selected for this work is limited by practical GPU memory constraints. To test the scalability of the SR methods used in this work, a state-of-the-art GPU (NVIDIA A100) with 40 GB VRAM is utilized to develop a GAN that can generate HR fields at the continental scale.

Here, the configuration of the hyperparameters is the same for the smaller sub-regions, with the exception of batch size which is reduced to 32 to reduce memory requirements. The region chosen is 3040 km $\times$ 3040 km at the same resolution gap as earlier (i.e. 80 km LR grids, 10 km HR grids). Although training requires a larger-memory GPU, the generation of these fields can be done on smaller-memory GPUs in seconds. An example realization is shown in Figure 4.9. A similar analysis is repeated for this continental scale GAN, with biases shown in Figure B.11 and the power spectra shown in Figure B.12. These results are similar to those for the

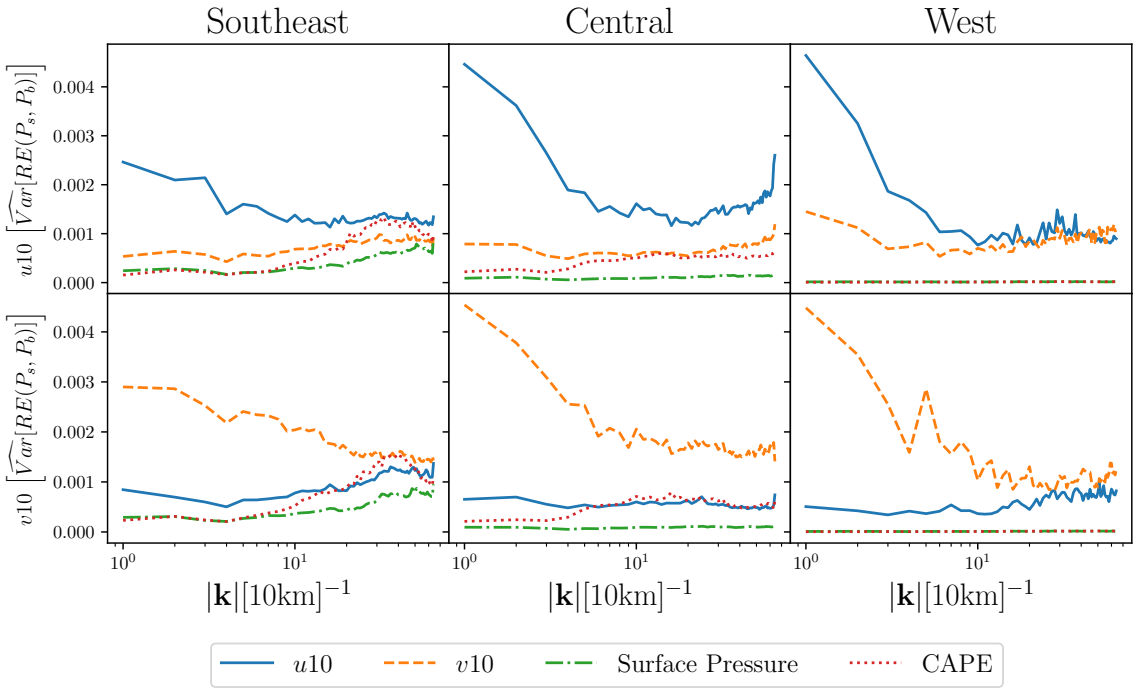


Figure 4.8: Sensitivity of the HR field power spectrum in response to modified LR conditioning fields. Each line is a different covariate, with the variance reported at different wavenumbers. The top row is for $u10$ wind, and the bottom is for the $v10$ wind. Each column is a different region.

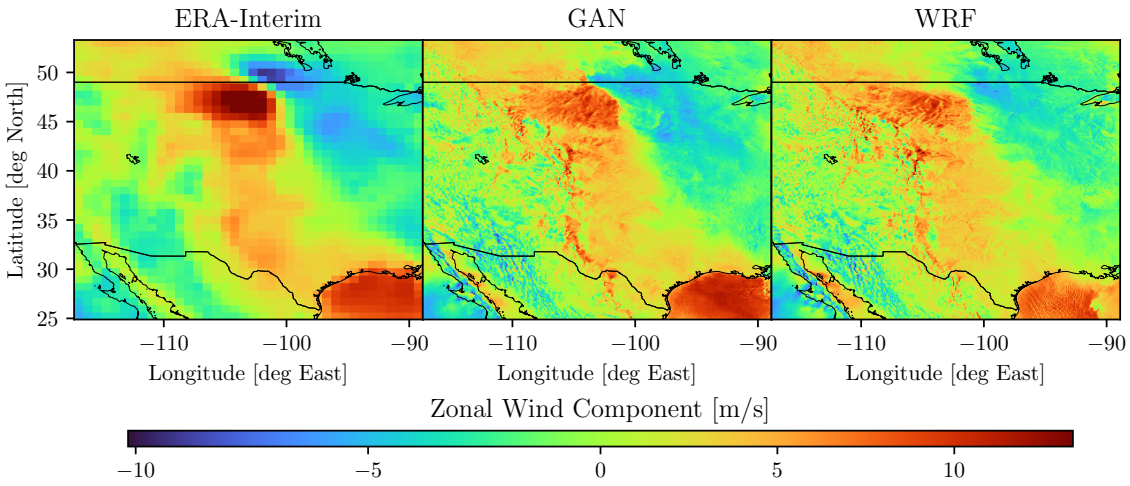


Figure 4.9: Random example realization of continental scale SR for the $u10$ wind component. The left panel shows ERA-Interim, the center panel shows the results of the NFS GAN, and the right panel shows WRF HRCONUS.

smaller regions, and importantly, show that the WGAN-GP framework can be scaled to continental-scale regions.

## 4.8 Super-Resolving a Serially Dependent Covariate Space

Results from the GAN SR models presented thus far have demonstrated that the characteristics of generated fine-scale features are quite similar to those simulated by WRF HRCONUS. However, actual weather evolves continuously through time at a range of scales, and so far, this characteristic has not been demonstrated by the deterministic SR models trained in this work. Inspired by seminal methods from Radford et al. (2016) using Deep Convolutional GANs (DCGANs), the following section explores interpolated GAN inputs and asks the question: is the Generator locally continuous around neighbouring LR fields? Using the NFS GAN, generated fields can be manipulated by interpolating temporally adjacent covariate fields to simulate the evolution of actual weather. In doing so, the continuity (or discontinuity) of the small-scale features produced by the GANs can be assessed.

By exploiting the temporal persistence in the large-scale features, linear interpolation between adjacent six-hourly LR covariates is performed as

$$X_{\tilde{t}} = (1-\epsilon)X_t + \epsilon X_{t+1}; \quad 0 \leq \epsilon < 1 \tag{4.3}$$

where $X_{\tilde{t}}$ denotes the interpolated coarse set of covariates, $X_t$ represents covariates at time $t$, and $X_{t+1}$ represents the covariates six hours later. By conditioning the GAN on $X_{\tilde{t}}$, a temporal interpolation can be performed in the HR fields via $G(X_{\tilde{t}})$. Before proceeding, it should be clear that no claims are made about the physical plausibility of generated temporal persistence, and doing so would require further investigation beyond the present scope of this thesis.

An animation of HR wind speed is made by interpolating the six hourly covariates to five minutes using Equation 4.3 for the continental scale GAN. To view this video, see this link, or go to `https://youtu.be/2-InALQLb-A` (for best results, please set the video playback to a high-resolution). The interpolated fields are compared to the native hourly WRF HRCONUS fields and reconstruct realistic fine-scale features that continuously evolve through time – much like the actual evolution of weather. The continuity of the fine-scale features is particularly interesting since no information about serial dependence is used during training. In other words, serial dependence in the HR fields is inherited only from the simulated serial dependence of the conditioning LR fields.

Another demonstration of this serial dependence is shown in Figure 4.10 where the NFS GAN is used with the LR fields interpolated to hourly for direct comparison with hourly WRF HRCONUS. Figure 4.10 shows how the temporal evolution evolves continuously in the NFS GAN in a similar fashion to the hourly WRF HRCONUS. The time series is broadly consistent with the six-hourly ERA-Interim time series; however, both the NFS GAN and WRF HRCONUS series demonstrate their own variability in time. This variability demonstrates temporal evolution that *could* exist, but doesn't necessarily occur exactly in WRF HRCONUS.

Deterministic GANs demonstrate that there is continuity and serial dependence of the fine-scale features. Stochastic GANs, that produce fine-scale features stemming from random draws, may not show this property unless temporal dependence is somehow incorporated into the methodology. For example, Leinonen et al. (2020), uses ConvGRU layers which are a special convolutional layer that incorporates temporal dependence of the fields during optimization (Ballas et al., 2015). As discussed earlier, (and in contrast to the aforementioned work that utilizes ConvGRU layers), no explicit serially dependent information is included while training the GAN from

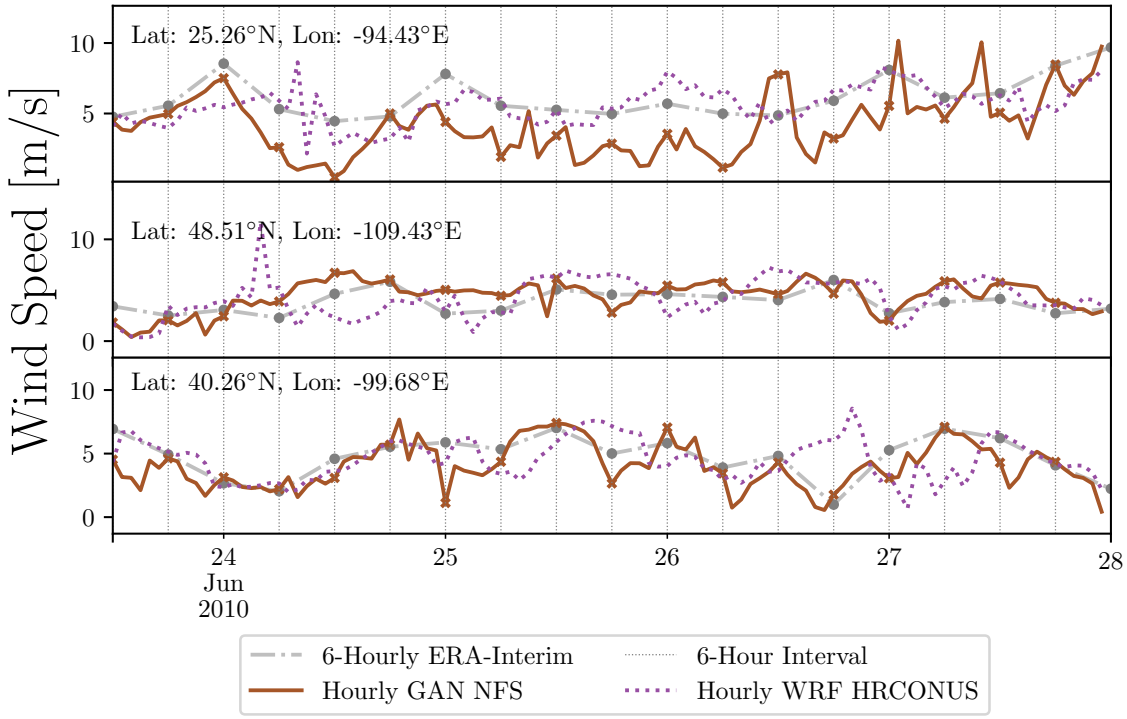


Figure 4.10: Time series of wind speed for three random grid cells (rows) from the Continental region with their location reported in each panel. The HR grid cell is located near the center of a randomly selected LR grid cell so that the LR and HR grid cells are co-located. The lines represent the different models, with vertical markers and points on the lines representing the six-hourly intervals in the training set.

this present work, yet serial dependence is observed.

# Chapter 5

# Discussion

Adopting GANs for SR shows impressive capabilities in downscaling multivariate wind fields to convection-permitting scales. Importantly, the GANs generate fine-scale variability that is statistically very similar to the "true" variability simulated by the CPM. Extensive dynamical downscaling by CPMs is operationally infeasible due to computational costs, which makes statistical downscaling using GANs a very attractive and practical alternative.

Once the GANs are trained to mimic the CPM, multivariable HR fields can be produced very quickly (on the order of seconds). As such, the SR models for wind fields considered in this work can be useful for a variety of practical applications, such as: estimating fire weather danger and wildfire behaviour, modelling pollutant and pollen dispersal, infrastructure design, and wind-turbine siting.

While the SR methods can be configured with various variables as covariates and predictands, multivariable wind fields offer a challenge due to the high degree of spatial inhomogeneity in the fields. The success of the GANs with multivariate wind fields (Singh et al., 2019; Stengel et al., 2020), as well as other spatially inhomogeneous fields such as precipitation (Leinonen et al., 2020; Harris et al., 2022; Price and Rasp, 2022), bodes well for their operational feasibility in statistically downscaling other variables as well.

Important improvements to GAN architectures that are often overlooked or taken for granted in existing work are those benefits provided by WGAN-GP over the classical min-max game from early GANs. As shown in this work, minimizing the Wasserstein distance between generated and "true" samples leads to impressive training stability. This stability is persistent despite altering loss functions, regions, and conditioning fields. WGAN-GP therefore serves as a solid framework for continuing work

in building stochastic implementations of SR, like in Leinonen et al. (2020); Harris et al. (2022); Price and Rasp (2022).

## 5.1 Frequency Separation

In addition to adopting WGAN-GP for SR applications of wind fields, a novel application of spatial FS is introduced that acknowledges the strengths and limitations of terms in the loss functions used in SR tasks. FS recognizes that generated large-scale features should match "true" large-scale features. The MAE applied to low-frequencies is therefore appropriate, with a caveat about coarse scale differences between the LR and "true" HR fields to be discussed later in this section. Conversely, evaluating the generated fine-scale features leads to the double penalty problem using grid-point-based metrics.

The adversarial loss is not a grid-point-based metric. In FS, it evaluates the Wasserstein distance between the generated and "true" high-frequency fine-scale features only – it is, therefore, responsible for generating high-frequency variability that could have existed in the "true" HR fields. It is shown, however, that FS alters the optimization problem of SR – and not necessarily for the better.

As discussed in Section 4.2, the CNN consistently performs best in terms of MAE, whereas the performance of the FS GANs scale with kernel size, with NFS GAN showing the largest MAE. The ranking can be explained by considering how FS changes the optimization problem to "converge" to the CNN depending on the range of frequencies that are delegated to the content loss. The results of FS can therefore be deceiving if one fails to consider how variability is generated across spatial scales for example as shown in Figures 4.6 and B.10 where the NFS GAN (with largest MAE) best matches the variability of WRF HRCONUS across all spatial scales and regions.

This altering of the optimization problem discussed above can be compared to more classical statistical downscaling approaches. Pure deterministic CNN SR models can be viewed as a classical regression problem that provides a single predicted value from the conditional distribution. If optimized to find the lowest MAE, this predicted value can be viewed as the conditional median of the predictive distribution (with the lowest MSE, corresponding to the conditional mean) (Hsieh, 2009, 127-132). This leads to a more spatially smooth arrangement of predictions – as can be clearly seen in the pure CNN approach – since only the median of the conditional distribution is taken

as the predicted value. Much like SWG regression models, where random variability is introduced by sampling from a distribution defined by the residuals about the best-fit predicted value, deterministic SR GANs are analogous to generating a single predicted value (provided by the content loss) with a residual component that superimposes an instance of random variability sampled from the conditional distribution (provided by the adversarial loss) onto this predicted value. One important difference between many SWG regression approaches and SR GANs is that the adversarial component in the latter can model the entire distribution, without assuming a parametric form of the conditional distribution from which to sample the random variability.

The goal of the content loss in the Generator's objective function is to get *single realizations* to look like the conditional median, which drives the output to appear more like the pure CNN by suppressing variability in the single realizations. Simultaneously, using the Critic, the goal of the adversarial loss is to ensure that single realizations are drawn from the entire distribution of possible draws. Crucially, for the deterministic GANs, only a single realization is provided (i.e. one sample from the conditional distribution), which provides a poor estimate of the conditional median, and explains why the MAE is larger than the pure CNN cases.

## 5.2 Stochastic Super-Resolution

Stochastic SR can be achieved by introducing noise into the conditioning information. One approach is to simply add a noise-only LR field (with the same shape as the LR covariates) to the set of fixed LR covariates. Providing multiple realizations of noise fields (with fixed LR covariates) would then ideally allow for the sampling of multiple representations of fine-scale features from the conditional distribution.

Harris et al. (2022) using a stochastic approach for precipitation fields, calculates the content loss (in this case, using the MSE) on the ensemble mean of generated fields sampled from the conditional distribution. This sampling of multiple realizations provides a better estimate of the conditional mean; small-scale features are averaged out so that the ensemble mean is not over-penalized. Therefore, extending this work stochastic SR should include calculating the content loss between the "true" HR fields and the ensemble mean of realizations, rather than individual SR-generated HR fields.

Furthermore, the lack of freedom in deterministic SR to generate more variability imposed by the content loss demonstrates a key limitation in weather and climate and explains why biases may persist in the power spectra of deterministically generated

fields.

Indeed, applying the content loss to the ensemble average in stochastic SR is similar to smoothing a single realization using a low-frequency filter, since high-frequency differences between realizations in a stochastic approach would smooth out given multiple realizations. However, the smoothing effect of the content loss is different from frequency separation, since in Harris et al. (2022) all frequencies are included in the adversarial component, whereas FS in this present work supplies only the high frequencies.

One possible way to "mimic" the effect of ensemble averaging in deterministic approaches would be to apply FS and delegate the low frequencies to the content loss while allowing all frequencies to be passed to the adversarial loss, resulting in a "partial" FS GAN (the term "partial" implying that only part of the Generator's objective function uses FS). A "partial" FS approach is expected to result in power spectra similar to the NFS GANs, however, with a reduced bias in the power spectra since the content loss is not suppressing valuable variability. At the same time, values of the MAE (calculated over all frequencies) may show a similar ordering as the FS GANs reflecting how the size of the low-frequency filter changes the optimization. In a sense, "partial" FS may also help identify the limits of the GANs' ability to correct large-scale differences between the LR and HR training fields. Training and evaluating such "partial" FS GANs is left for future work.

## 5.3 Low-power Biases in the Spectra

As can be seen in Figure B.10, a more prominent low-power bias exists between WRF HRCONUS and the FS GANs than WRF HRCONUS and the NFS GANs in each region. This occurs for the same reason that pure CNNs exhibit low-power bias. The absence of the adversarial loss at larger scales means that variability is not sampled from the conditional distribution. In plainer terms, the adversarial loss more successfully produces variability that it "sees" during training than the content loss. FS, in a sense, blinds the adversarial component of the model to this variability, which is undesirable.

There is also a low-power bias between WRF HRCONUS and the NFS GAN (GAN with ERA +) shown in Figure 4.7. This bias is separate from the low-power bias at low frequencies caused by predicting the conditional median (seen at low wavenumbers in the FS GANs and the pure CNN) since the adversarial loss *does* allow for sampling

of variability at the large scales. To an extent, this sampling "corrects" the low-power bias of the generated fields to match WRF HRCONUS more closely. But still, why does a reduced bias persist at these scales? These biases likely result from differences in the large scales between the WRF HRCONUS and ERA-Interim wind components.

To test this idea, additional GANs are trained with only $u10$ and $v10$ LR fields from coarsened WRF HRCONUS (i.e. the "idealized" version) and are compared to GANs trained with only $u10$ and $v10$ LR fields from ERA-Interim. Idealized coarsening is common to SR applications in climate sciences and computer vision, and Figure 4.7 demonstrates how coarsening changes the power spectra biases at low frequencies. Namely, idealized coarsening dramatically reduces the low-power bias at large scales (low wavenumbers).

## 5.4 The Role of Additional Covariates

To the best of the author's knowledge, this is the first application of SR to climate fields that demonstrates the value of including additional covariates. Namely, a robust reduction in low-power bias is seen across all frequencies as a result of including additional covariates – especially at high frequencies. This result is seen clearly in Figure 4.7. Interestingly, low-power biases at high frequencies are quite similar between the idealized and non-idealized GANs using just $u10$ and $v10$ as the LR covariate fields suggesting that additional covariates play an important role in generating high-frequency variability.

The observed low-power bias for each SR model is consistent with biases in the spatial patterns of the means, standard deviations, and 90th percentiles of wind speed, as well as the marginal distributions of $u10$ and $v10$. However, these low-power biases are shown more clearly in the power spectra.

Additionally, a methodology is proposed to determine the relative importance of some of the coarse covariates. This sensitivity of the generated fields on these coarse variables appears consistent with the climatological characteristics of each region. For example, in the West region, surface pressure and CAPE appear to play a minimal role in the generated spectra, so including them may not be necessary for this region. Meanwhile, the Southeast region where fine-scale convective features are common appears to be very sensitive to CAPE. It is noteworthy, however, that the method developed to analyze the covariate sensitivity does not translate to invariant covariates, such as the surface roughness length, topography, or land-sea-mask, which

may play an important role in properly representing high-frequency variability. Future work that analyzes the power spectra of new GANs by eliminating those covariates may reveal their importance. While novel CNN architectures in GANs that make use of HR invariant covariates may significantly improve generated fields (as in Harris et al. (2022) for topography), a comparison with and without these additions has not yet been demonstrated to the best of the author's knowledge.

Importantly, small-scale bias is reduced by providing relevant covariates to the Generator. This improvement is an important finding for designing SR models for climate and weather fields; the physical relationships between covariates and the predictands are mirrored in the GANs.

Over the Continental region in the video provided in Section 4.8, longer-living organized convective systems in WRF HRCONUS do not appear to be represented as well in the evolving generated fields. Vertical wind shear can play an important role in organized and long-living convection, severe gusts, and tornadoes (Pucik et al., 2021). Including the appropriate vertical wind shear information (in the form of LR covariates) could improve the SR methods' ability to generate realistic convective structures. Further, since high-frequency details are especially sensitive to well-chosen covariates, adding wind shear may improve the low-power biases by representing convective systems associated with strong vertical wind shears.

In existing studies, stochastic methods have demonstrated under-dispersion in the rank statistics of the generated fields for both idealized (Leinonen et al., 2020), and non-idealized (Price and Rasp, 2022; Harris et al., 2022) LR and HR pairs. Harris et al. (2022) performed an idealized experiment using covariates derived from the HR fields (similar to this work), which revealed that idealized covariates improve the calibration and continuous ranked probability score (CRPS). Moreover, Harris et al. (2022) attributed large-scale differences between the LR and HR fields as being the main limiting factor in the GAN performance. Such an interpretation is consistent with the results found here after examining the power spectra of several SR models. The differences between LR and HR that lead to low-power biases suggest that correction to the LR fields prior to SR is a reasonable approach to help alleviate this issue, and indeed Price and Rasp (2022) includes an additional *corrector* step to correct the variability of the LR patterns resulting better performance compared to the "pure" SR GAN implementation. The results from these stochastic SR methods indicate that reducing the low-frequency bias can improve the range of generated variability in the stochastic HR fields. This present work complements these results

from stochastic methods by using deterministic GANs to carefully examine how the low-frequency bias is represented in the spectra of the generated fields, and, importantly, how the content loss can influence this bias. Additionally, since appropriately chosen covariates help with high-frequency biases, choosing the proper covariates may play an important role in the rank statistics of stochastic methods and may help limit under-dispersion or poor CRPS scores.

Besides bias-correcting, the LR fields to reduce low-power bias at large scales (Price and Rasp, 2022), and including appropriate covariates to reduce high-frequency bias, one additional way to help overcome the biases in the spectra may be to include an additional loss term that directly evaluates spectra of batches of generated and "true" fields. Such an approach was introduced in Kashinath et al. (2021) by replacing the adversarial loss entirely with this "spectral loss". This approach performs well, but its implementation in Kashinath et al. (2021) comes with caveats – i.e. it was tested with idealized covariates and a scale gap of $4\times$. Moreover, the success of this approach may also be strongly dictated by the regional climatology of the fields considered, e.g. regions with plentiful unresolved fine-scale convective features in the LR fields demand more generated variability from the SR models. For WGANs, leaving out the adversarial loss means not evaluating the Wasserstein distance, which is a critical component for converging to the entire target distribution and for generating fine-scale features in the presence of a large-scale gap. In future work, rather than replace components of the Generator's loss function entirely, this "spectral loss" could serve an auxiliary role that supports the generation of variability across spatial scales in addition to the content and adversarial components. If successful, this could lead to better CRPS and rank statistics of the generated ensembles in stochastic approaches.

## 5.5 Tuning the GANs

One caveat of this work is that hyperparameter tuning is not used to find the optimal SR model. While this should not affect the main insights provided here, the SR models should not be considered "optimal". For instance, reducing the content loss weighting in Equation 3.6 or 3.14 may result in different power spectra than those presented here – specifically, in the reduction of the low-power bias present across the regions. In other words, weighting the adversarial loss in the Generator's loss function more heavily may promote the generation of variability across scales and reduce some of the biases found in the spectra. Removing the content loss entirely

from Equation 3.6 (i.e. relying solely on the adversarial loss), although not reported here, leads to stability issues while training and results in generated fields with poor quality. Clearly, a trade-off between GAN stability and increased variability must be met. Future work can balance the stability of the SR method with the weighting of the content loss, potentially improving the biases that the adversarial loss helps alleviate.

It should be worth noting that Enhanced SRGAN (ESRGAN) from Wang et al. (2018) is a state-of-the-art improvement to SRGAN; however, it introduces additional complexity to the Generator network that increases GPU hardware requirements and pushes the GPU used in this work beyond its practical capabilities. The SRGAN Generator is used in the present work purely for practical reasons. However, as demonstrated in previous work (e.g. Wang et al. (2018); Fritsche et al. (2019)), improvements to the quality of reconstructed fields may be found by using ESRGAN instead.

## 5.6 Extrapolation and Projected Periods

Despite the successes of SR methods for climate fields thus far, it remains unknown how SR models can extrapolate to projected periods (given anthropogenic forcing) when trained using model outputs from a historical period (Chantry et al., 2021). An important consideration for developing SR models that generalize to different climate states is the careful choice of covariates. If the LR covariates have a strong physical relationship with the predictand, and this relationship does not change in the future, then the SR models have a better chance of generalizing to "new" climate conditions. Further, there is still a deficiency in the limited availability of synchronous (but non-idealized) pairs of LR and HR climate models due to the availability of HR models, such as CPMs. In this work, covariates from ERA-Interim which supplied boundary conditions to WRF HRCONUS were selected to maximize synchroneity and reduce differences. However, as demonstrated, ERA-Interim and WRF HRCONUS still show some differences that are represented in the generated fields. Even with these differences, however, the GANs perform very well in generating realistic HR fields.

## 5.7 Using a Different Reanalysis Model

ERA5 (from which CAPE is taken) is a reanalysis that has superseded ERA-Interim and exists at a higher-spatiotemporal resolution (i.e. hourly and ∼30 km). While ERA5 does not drive WRF HRCONUS, it may serve as a good LR option since it reduces the scale gap, and is archived at the same hourly time-step as the WRF HRCONUS fields. Although ERA5 and ERA-Interim may have slight differences, the benefits of a reduced scale gap and more training data may outweigh these differences when it comes to the quality of the generated fields – especially since the GANs demonstrate an ability to overcome some of the ERA-Interim to WRF HRCONUS differences. Since the differences are most prevalent in the large scales of the generated fields, ERA5's better spatiotemporal resolution may very well actually improve these large-scale low-power biases of the generated fields.

## 5.8 Ethical Artificial Intelligence and the Limitations of Super-Resolution

SR methods are capable of generating results with uncanny realism. Here, the generated surface wind fields are particularly impressive since the SR methods required only minor modifications for compatibility with climate models. As such, the adaptability of the SR methods presents exciting opportunities for use beyond computer vision and climate sciences (e.g., Yamashita and Markov (2020)). However, SR can potentially be misused, and this brief section aims to acknowledge this possibility.

In some existing operational applications of AI, *algorithmic biases* can reflect or reinforce societal biases. For instance, algorithmic biases have been found in facial recognition software (Klare et al., 2012; Buolamwini and Gebru, 2018), tools for criminal justice and law (Angwin et al., 2016; Yeung et al., 2021), and even in healthcare settings (Obermeyer et al., 2019).

As discussed in detail in McGovern et al. (2022), AI methods in the environmental sciences are *not* immune to algorithmic biases that can manifest iMuch like SWG regression models, where random variability is introduced by sampling from a distribution defined by the residuals about the best-fit predicted value, deterministic SR GANs are analogous to generating a single predicted value (provided by the content loss) with a residual component that superimposes an instance of random variability

sampled from the conditional distribution (provided by the adversarial loss) onto this predicted value. One important difference between many SWG regression approaches and SR GANs is that the adversarial component in the latter can model the entire distribution, without assuming a parametric form of the conditional distribution from which to sample the random variability.n various ways and are exacerbated if practitioners misplace confidence in the objectivity of the methods. As AI becomes more advanced and interdisciplinary, researchers should be transparent and clear in communicating the limitations of their methods – especially if said methods could introduce or reinforce societal biases that lead to discriminatory practices (O'Neil, 2017; Benjamin, 2019; Cataleta, 2020).

Perhaps the most obvious limitation of SR in the environmental sciences concerns limitations within the target data. With WGAN-GP and SR, the goal is to minimize the Wasserstein distance between the generated distribution and the target distribution. Therefore, shortcomings in the HR target data itself could lead to shortcomings in the generated HR data, described as *coded bias* (O'Neil, 2017; McGovern et al., 2022). For example, there are deficiencies in WRF HRCONUS like the central United States dry bias in mid-summer rainfall and the under-representation of wintertime heavy-precipitation events in the southeast (Liu et al., 2017). These biases could simply be replicated if using SR for precipitation. Outside of climate and weather, one could also imagine coded biases in environmental science datasets stemming from the non-representative geographies, populations, or observations leading to misleading or poor results. As models and frameworks increase in their capacity to represent target data, they could commensurately embody coded biases. This is a key consideration when using SR, and recognizing shortcomings in the target data could help avoid its misuse.

Furthermore, misunderstandings may arise if complicated AI methods are not interpretable or appear as *black-boxes* to practitioners. Efforts that help reveal how complicated AI models process information (and subsequently help reduce algorithmic biases) form the research area of explainable AI (XAI) (Vedaldi et al., 2019). The sensitivity analysis in this thesis is inspired by XAI techniques and illustrates how relevant physical information can help generate more realistic fields. Although the analysis does not fully explain how the CNNs process information, the results are still useful for interpreting how variability is generated and is therefore a step towards a more interpretable implementation of SR.

Another limitation may occur when humans interpret SR-generated data, espe-

cially because it could appear quite realistic even without the use of physical constraints (e.g., without the right set of covariates). For climate and weather fields, *appearing* realistic is not necessarily the same as *being* physically realistic. In the absence of physical constraints, this distinction should be made clear. In this work, although physical relationships between the LR and HR covariates are mirrored in the GAN by additional covariates, no explicit physical constraints were imposed on the SR models while training. Incorporating explicit physical constraints on 2D fields – especially on small scales – is a challenging task that depends on the variables considered. As a result, end-users should not misplace confidence in the realism of any single SR-generated field. Fully stochastic SR methods that provide uncertainty estimates if properly calibrated, and also impose physical constraints can help increase confidence in the realism of these SR-generated realizations. As such, fully stochastic SR that makes use of physical constraints is an important research area for future methods.

# Chapter 6

# Conclusions

This work has put forward a comprehensive analysis of deterministic SR models for surface wind fields applied to different regions in a convection-permitting model. The GANs used with the WGAN-GP framework show impressive realism and feasibility for practical applications. Despite only training deterministic SR GANs, the findings from this work can be generalized to stochastic methods.

Large scale gaps demanded the generated fields to contain fine-scale features that *could* exist in WRF HRCONUS, rather than necessarily the fine-scale features that did exist in the training data, putting forward challenges in selecting appropriate error metrics. This work demonstrated that the content loss can negatively influence the variability of the generated fields for deterministic applications of SR. Utilizing power spectra is critically important for understanding biases in the variability.

Additionally, this work introduced a novel application of frequency separation to climate fields – based on work in the computer vision field – that delegated low and high spatial frequency information to terms in the Generator's objective function. Frequency separation did not result in a more realistic GAN but did reveal how modifying the Generator's objective function changed the power spectra of the generated fields.

The role of covariates was closely examined in the power spectra of the generated fields. Specifically, this work revealed that sources of biases can come from existing LR/HR differences that become represented in the generated fields. This work also showed that carefully chosen covariates help reduce biases at all spatial scales, but especially in the fine-scale features.

To further investigate the role of covariates, a sensitivity experiment was conducted to demonstrate the value added to spatial structures in the generated fields

by each of the covariates. The importance of the covariates differed between regions, providing an interpretation consistent with their respective climatology.

The GAN SR model was scaled up to an additional continental-size task, demonstrating the stability of WGAN-GP with the SR methods used, and showing impressive fidelity in spatial patterns of their bias of multiple statistics, as well as power spectra. Interestingly, performing a temporal interpolation of the LR fields results in serial dependence of the generated continental scale fields despite the lack of explicit serial dependence information used for training the GAN networks. Such a temporal interpolation is novel to GAN SR for climate sciences,

Finally, future avenues of work have been identified, namely: (1.) stochastic methods, that can estimate the ensemble mean/median, could be used in training to reduce biases associated with the content loss; (2.) for deterministic approaches, a better estimate of the conditional median can be mimicked using low frequencies for the content loss and might improve the biases in the power spectra; (3.) a sensitivity analysis by systematically omitting invariant covariates like the topography, surface roughness length, and land-sea mask in the training of new GANs may help reveal their importance (especially by examining their power spectra), (4.) novel CNN architectures that can make use of HR invariant information might improve GAN performance and generalization capabilities; (5.) an explicit "spectral loss" that measures differences in spectra of batches may help reduce some of the low-power biases found in this work; (6.) using newer GAN networks, like ESRGAN, may show some benefits over the SRGAN networks used; (7.) how well GANs might generalize to future periods remains to be seen, but carefully chosen covariates might help; and (8.) newer datasets, like ERA5, may be well-suited for GAN SR with WRF HRCONUS especially since ERA5 and WRF HRCONUS are both hourly and provide a much larger training and test set.

*

# Appendix A

# Additional Descriptions of Methods

## A.1 Gradient Penalty

The gradient penalty term introduced in Equation 3.4 will be discussed in more detail here. WGAN-GP replaces weight-clipping from Arjovsky et al. (2017) as a way to restrict the gradients of $C$ to be 1-Lipschitz (1-Lipschitz is a requirement of utilizing the Kantorovich-Rubinstein duality). Weight-clipping leads to vanishing or exploding gradients, as well as the Critic learning simple functions (Gulrajani et al., 2017). For WGAN-GP, gradients are calculated on the output of the Critic with respect to the inputs, $\hat{x}$, as $\nabla_{\hat{x}} C(\hat{x})$. The optimal Critic has unit gradient norm everywhere, however in practice, the gradients are computed on straight lines between $\mathbb{P}_r$ and $\mathbb{P}_g$. As discussed in the main text, loss functions, $\mathcal{L}$, are provided batches of data from the training set. For the gradient penalty term, a single randomly sampled field, $\hat{x} = \epsilon x + (1 - \epsilon) G(\mathbf{x})$ with $\epsilon \sim U[0, 1]$, is produced for each batch member. The gradient penalty term is then taken as

$$GP = \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(||\nabla_{\hat{x}} C(\hat{x})||_2 - 1)^2] \tag{A.1}$$

where it can be seen that deviations from unit gradient norm contribute a positive value to the minimization posed in Equation 3.4.

## A.2 Multi-Scale Structural Similarity Index

The Multi-Scale Structural Similarity Index (MS-SSIM) is an extension of the Structural Similarity Index (SSIM), which are metrics developed in the computer vision

field for an objective measurement of the quality of images relative to a reference image (Wang et al., 2003). SSIM uses a single spatial scale (or window size) to quantify three components of image quality: a luminance, contrast, and structure term that are combined. MS-SSIM builds on SSIM by iteratively downsampling and applying low-pass filters to the input images. Downsampling provides multiple scales for which the SSIM can be evaluated and combined to produce the MS-SSIM. MS-SSIM is thought to be an improvement to measurements like peak signal-to-noise ratio (PSNR) and mean squared error (MSE) which don't correlate well with perceived quality (Eskicioglu and Fisher, 1995; Zhou Wang and A. C. Bovik, 2002; Wang et al., 2004). Compared to SSIM, MS-SSIM requires less configuration of parameters which makes it a more generalized approach.

## A.3 Power Spectra

Consider a field $x$ with horizontal (longitudinal) and vertical (latitudinal) axis $\phi$ and $\theta$ respectively such that $x(\phi, \theta)$. The 2D Fourier transform changes the basis to the frequency domain $\boldsymbol{\chi}$ with frequency components $l$ and $k$ with:

$$\boldsymbol{\chi}(l,k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(\phi,\theta)e^{-j2\pi(l\phi+k\theta)}d\phi d\theta \tag{A.2}$$

$$\phi \Leftrightarrow l \tag{A.3}$$

$$\theta \Leftrightarrow k \tag{A.4}$$

$$\boldsymbol{\chi}(l,k) \Leftrightarrow x(\phi,\theta) \tag{A.5}$$

where, $\boldsymbol{\chi}(l,k)$ is complex with frequency amplitudes $|\boldsymbol{\chi}(l,k)|^2$ (Press, 1986). Assuming zero mean amplitude, $|\boldsymbol{\chi}(l,k)|^2$ represents the variance across pixels for a given wavenumber, $|\mathbf{k}| = \langle l,k\rangle$. To compute the power spectral density, $P(\mathbf{k})$, the amplitudes are then weighted by the total pixel surface area (number of pixels per wavenumber) and binned and averaged by wavenumber.

## A.4 Software Packages

Here is a summary of the main software packages used in this thesis from the open-source community. The code-base for this thesis, titled *DOwnscaling WasserteiN*

*Generative Adversarial Network (DoWnGAN)* can be found at this URL: `https://github.com/nannau/DoWnGAN`.

- Python version 3.8.10 (Van Rossum and Drake, 2009)
- PyTorch version 1.12.1 (Paszke et al., 2017)
- MLflow version 1.27.0 (Chen et al., 2020)
- Matplotlib version 3.5.3 (Hunter, 2007)
- SciencePlots version 1.0.9 (Garrett, 2021)
- NumPy version 1.23.1 (Harris et al., 2020)
- SciPy version 1.9.0 (Virtanen et al., 2020)
- Climate Data Operators (CDO) version 2.1.0 (Schulzweida, 2022)

An exhaustive list can be found in the DoWnGAN repository.

# Appendix B

# Additional Figures

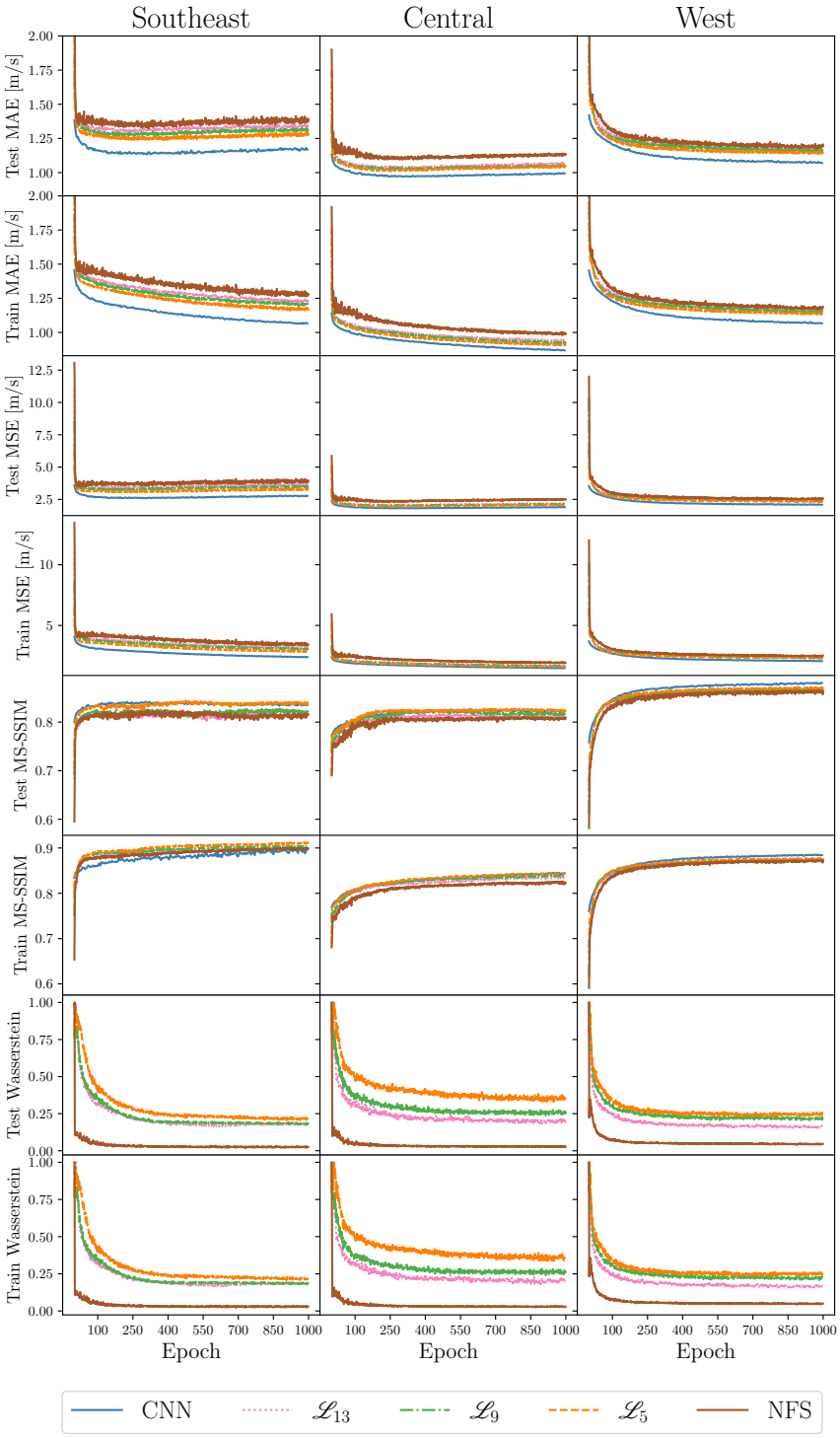


Figure B.1: Training evolution of the MAE, MSE, MS-SSIM, and Wasserstein distance for 1000 epochs on both the test and training data. Each SR model is indicated in the legend. The MAE, MSE, and MS-SSIM are calculated grid-wise with combined $u10$ and $v10$ HR fields. The Wasserstein distance estimate is scaled by the initial distance. Note that no Wasserstein distance is calculated for the pure CNN, since no Critic network was trained. This figure is supplementary to Figure 3.2

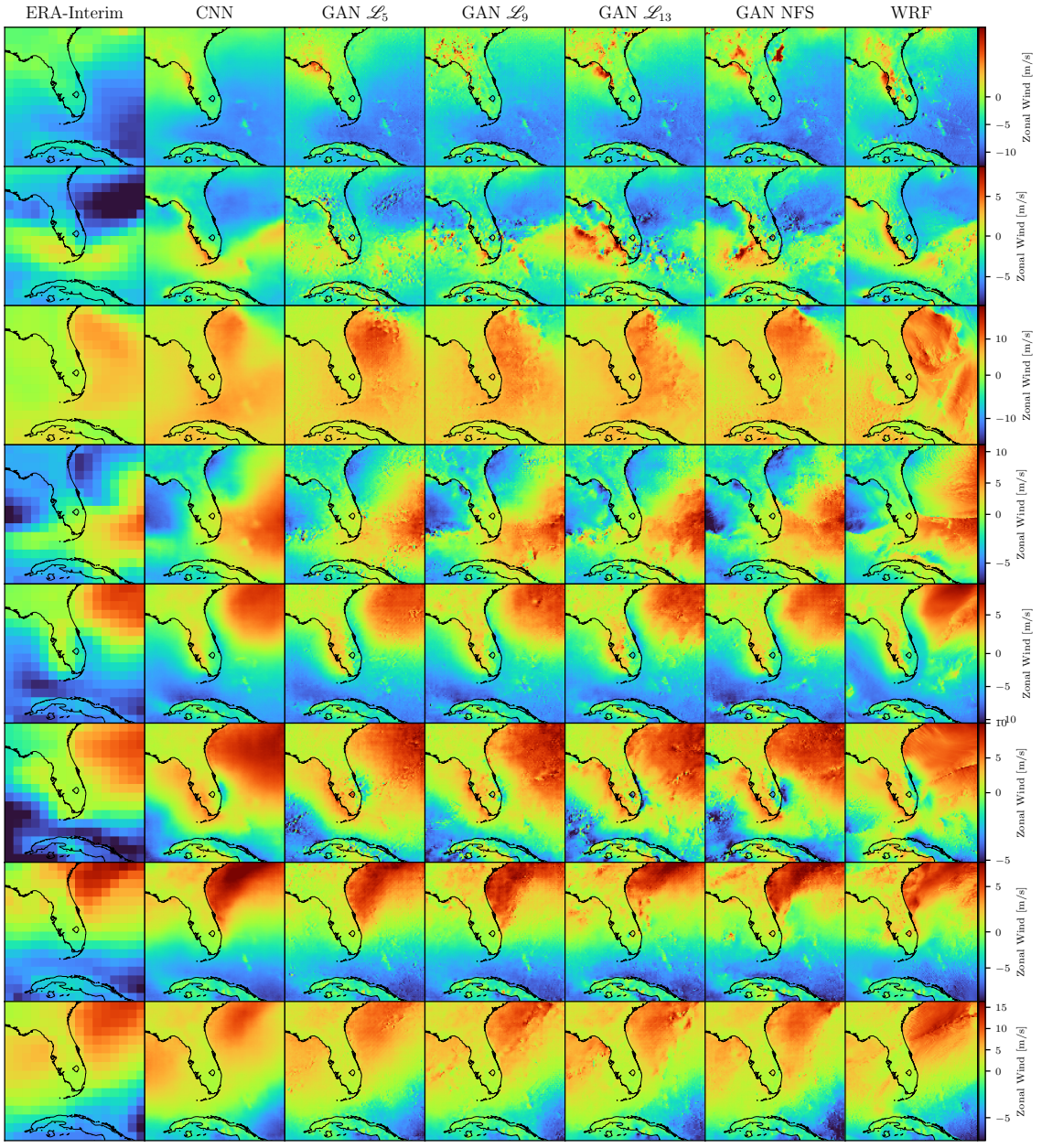


Figure B.2: Example realizations of $u10$ wind for the Southeast region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

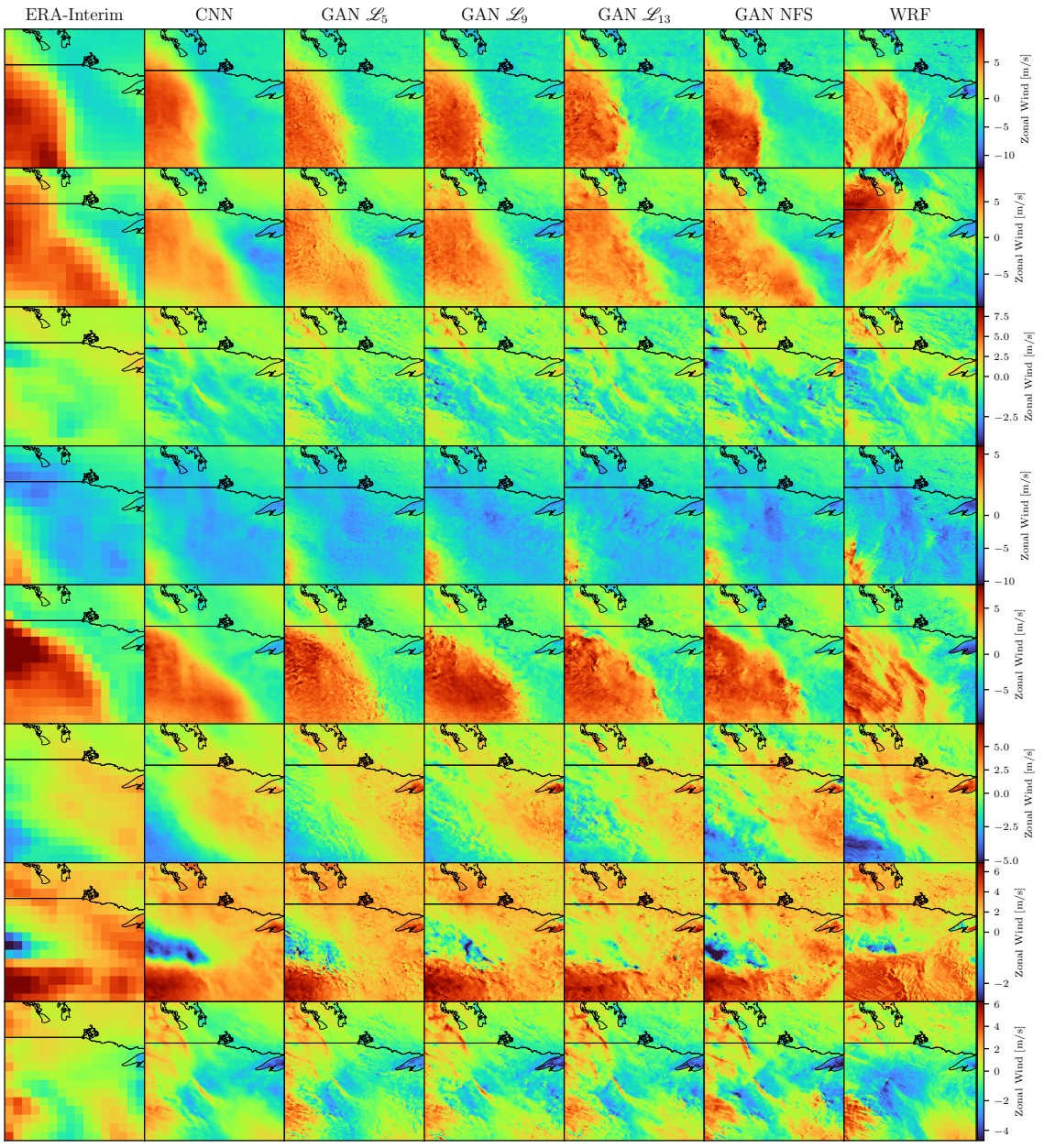


Figure B.3: Example realizations of $u10$ wind for the Central region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

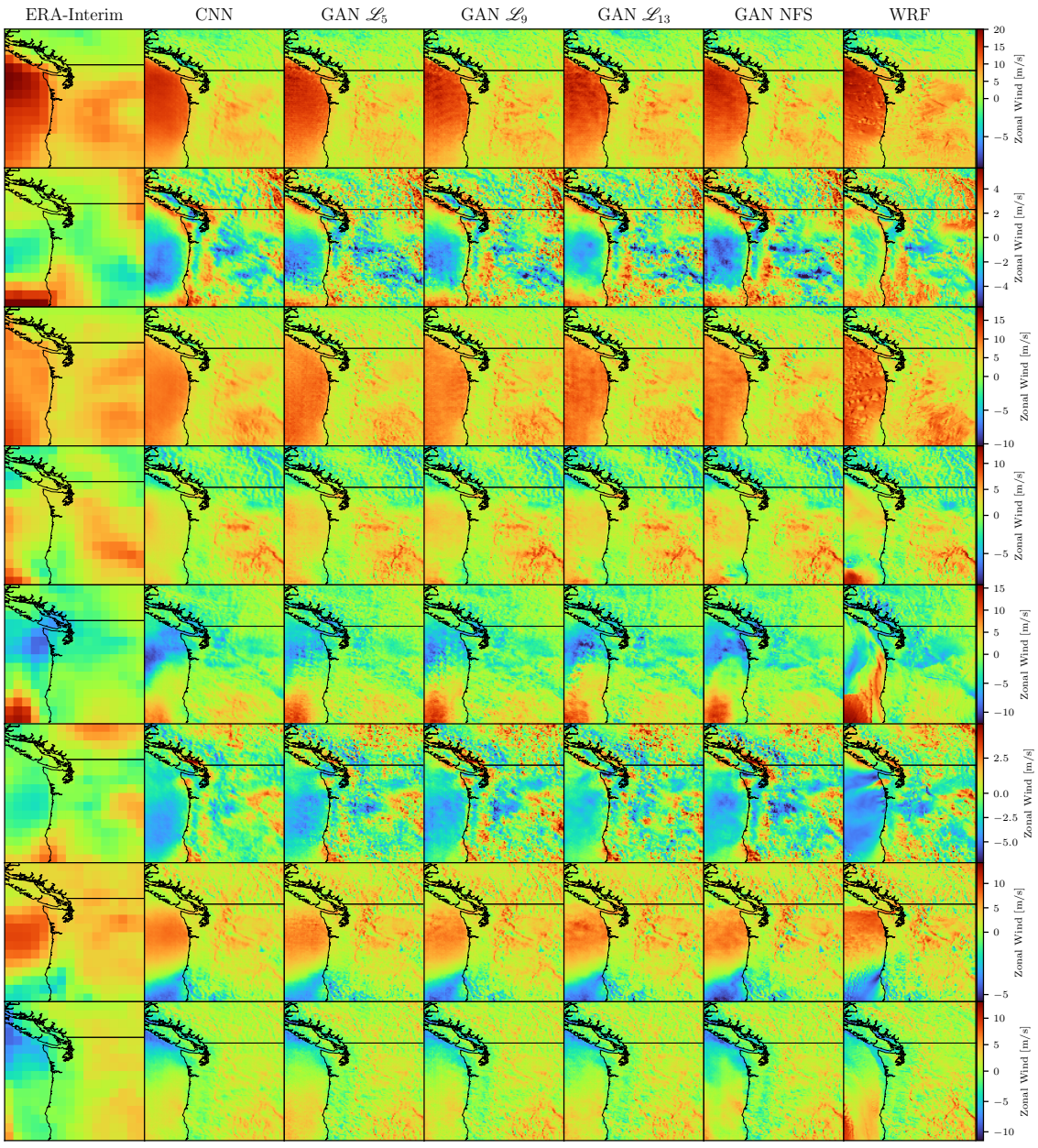


Figure B.4: Example realizations of $u10$ wind for the West region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

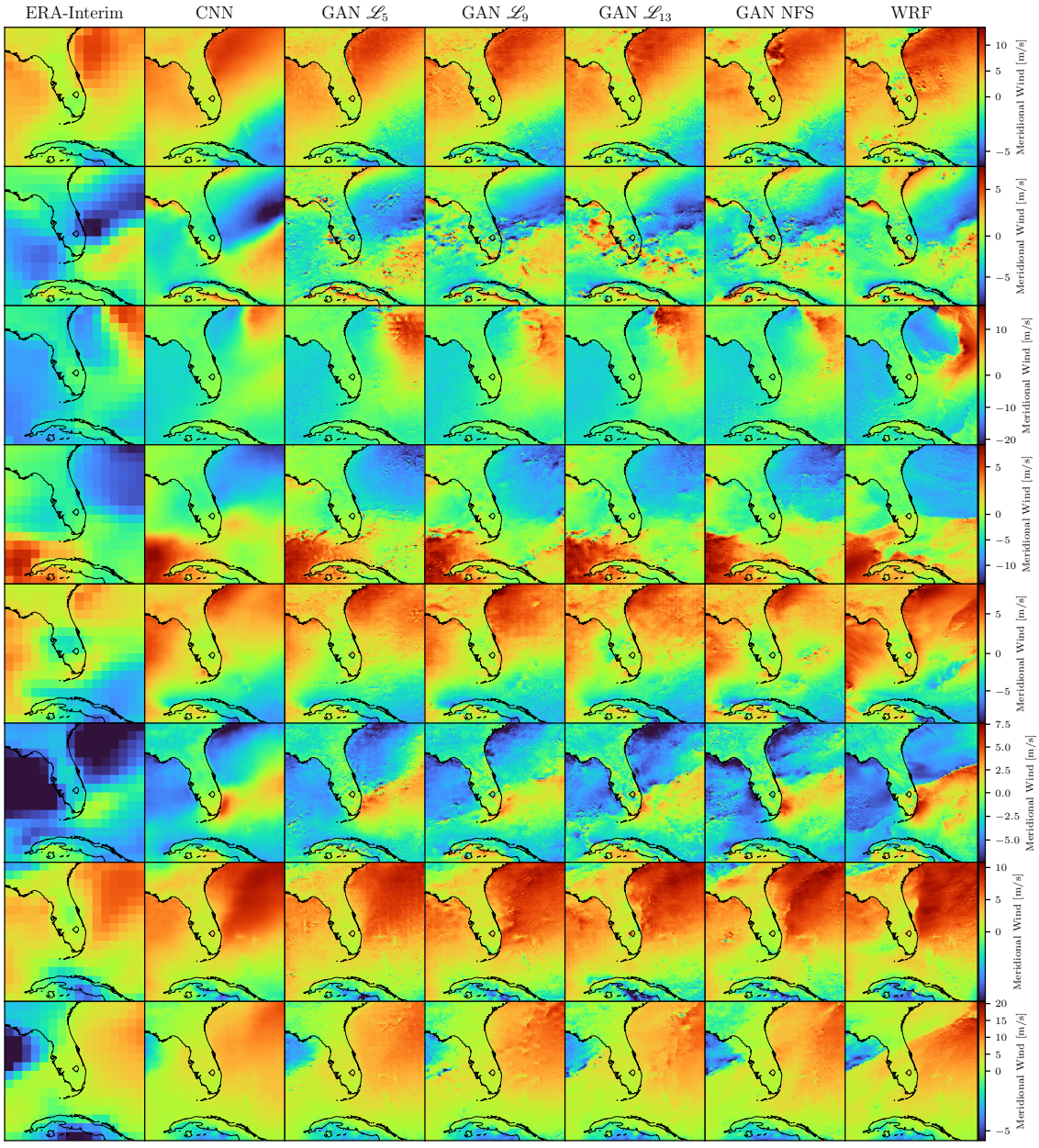


Figure B.5: Example realizations of $v10$ wind for the Southeast region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

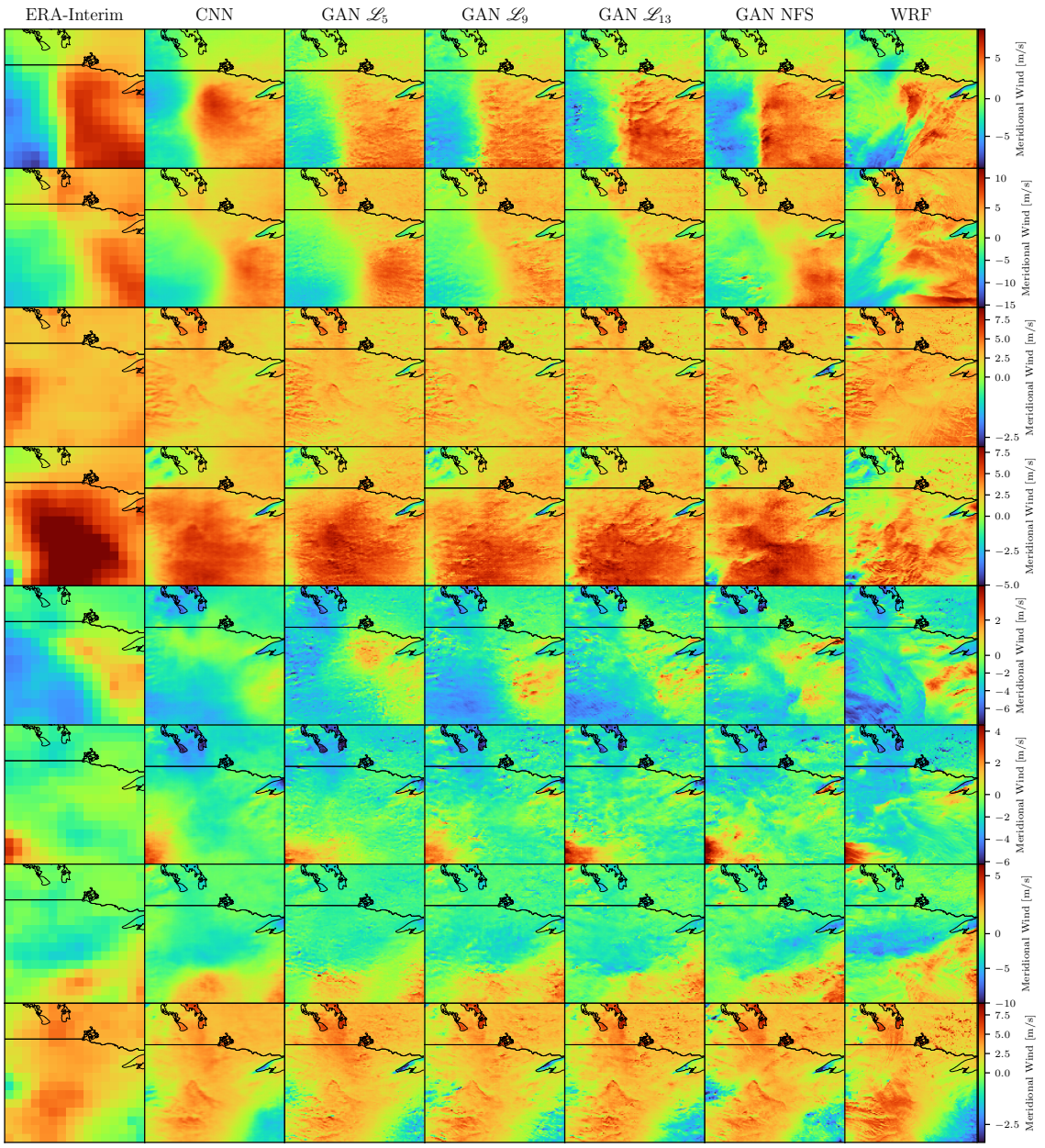


Figure B.6: Example realizations of $v10$ wind for the Central region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

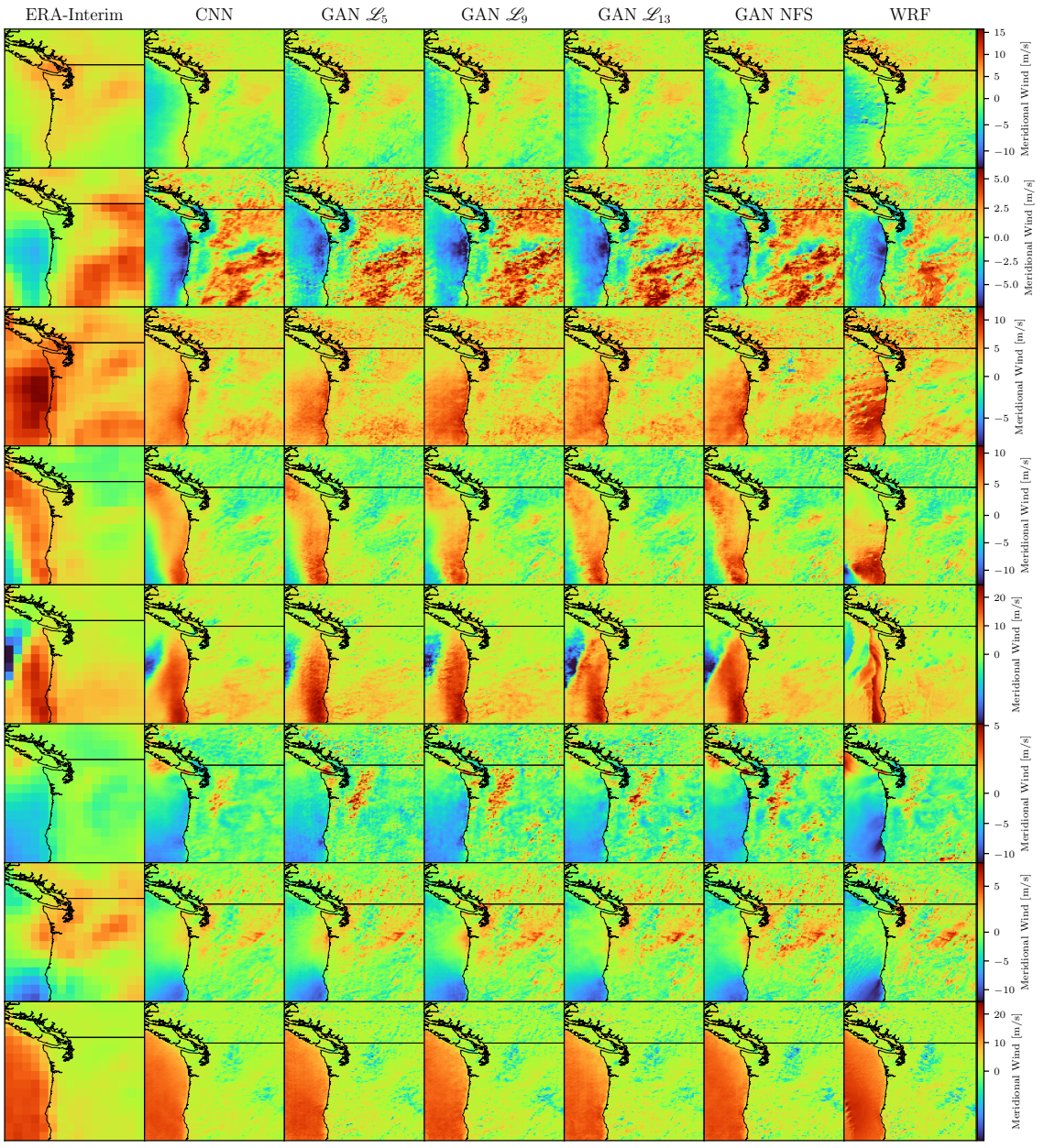


Figure B.7: Example realizations of $v10$ wind for the West region. Each row represents a randomized time step from the test set, while each column is a different model considered in this work.

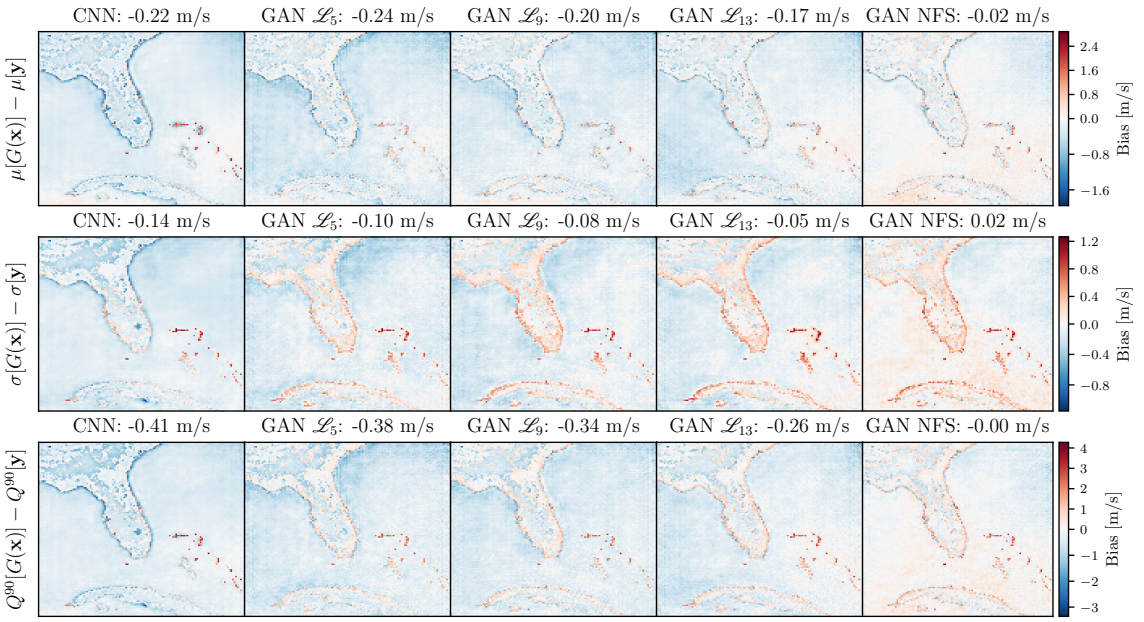


Figure B.8: Map of the Southeast region wind speed bias in the mean (top row), standard deviation (middle row), and 90th percentile (bottom row) for each SR model. Statistics are calculated along the temporal dimension of the test set. The spatial mean of the bias is reported in the title of each panel.

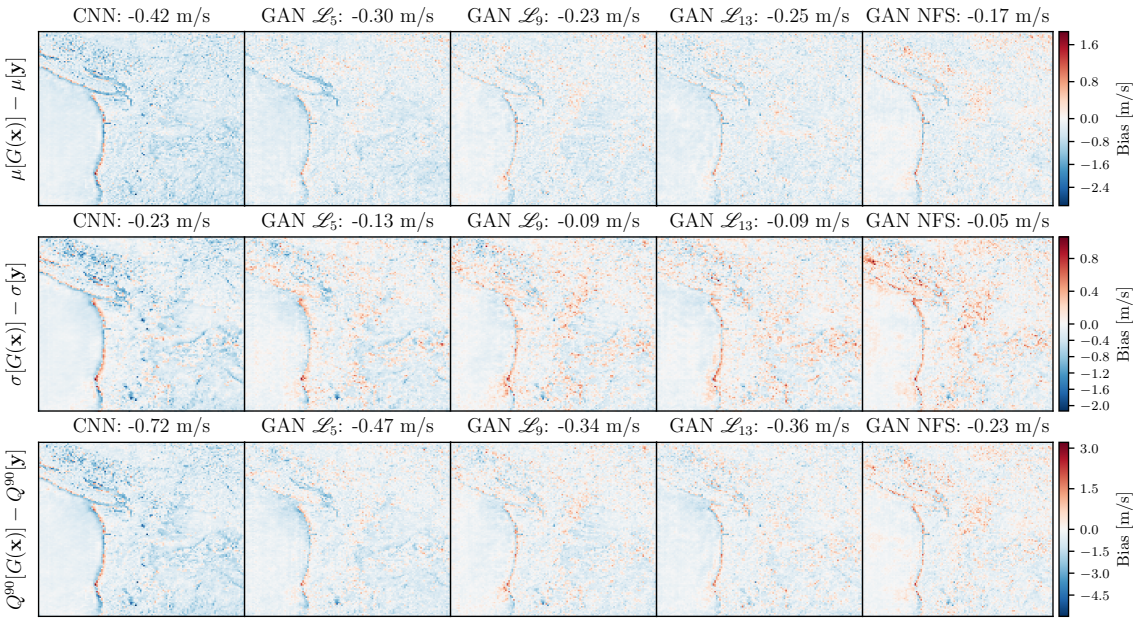


Figure B.9: Map of the West region wind speed bias in the mean (top row), standard deviation (middle row), and 90th percentile (bottom row) for each SR model. Statistics are calculated along the temporal dimension of the test set. The spatial mean of the bias is reported in the title of each panel.

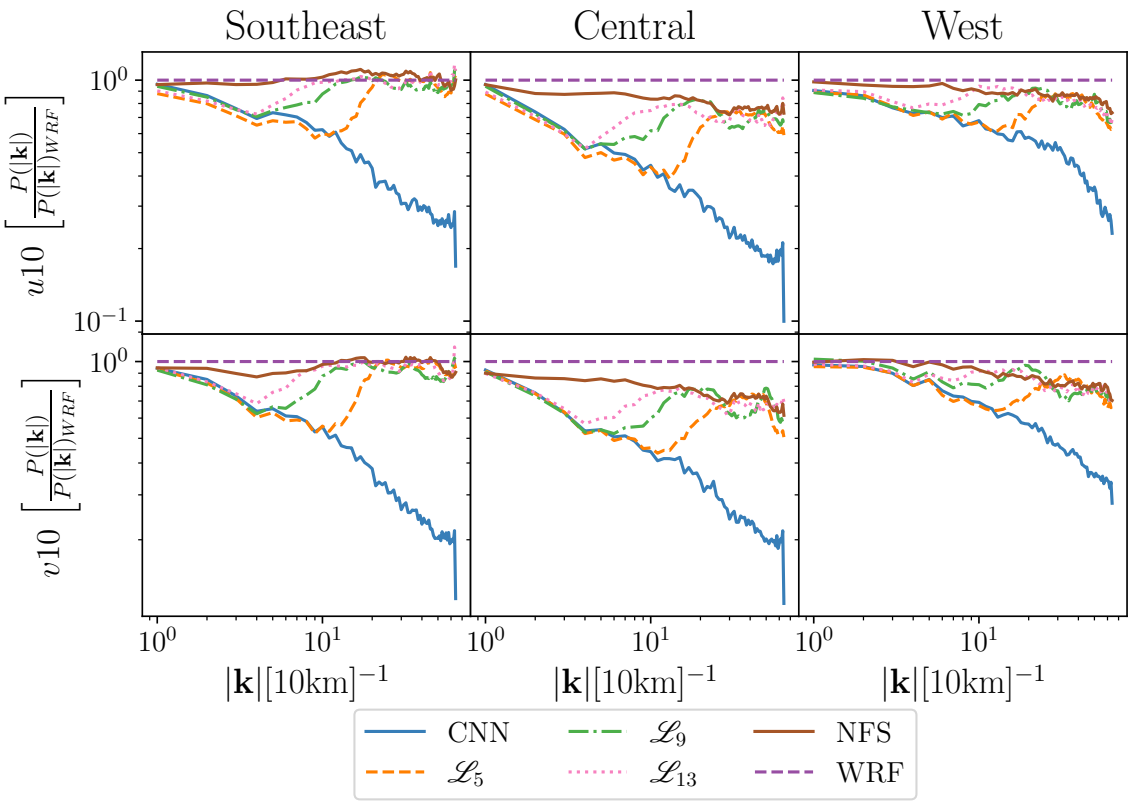


Figure B.10: Ratio power spectra of the SR models with respect to the WRF HRCONUS spectra. The top row shows the $u10$ component, and the bottom row shows the $v10$ component. Each column is a separate region.

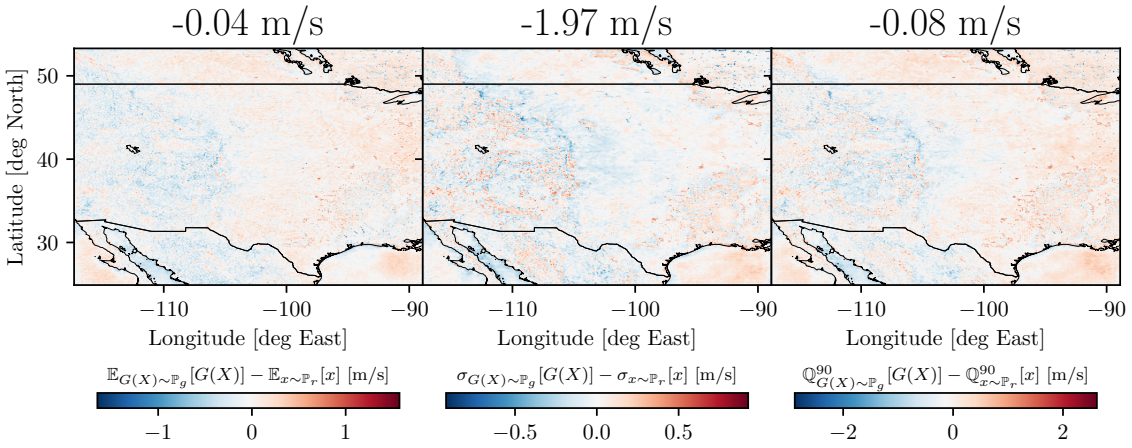


Figure B.11: Map of the Continental region wind speed bias. The left panel shows the bias in the mean, the center panel shows biases in the standard deviation, and the right panel shows biases in the 90th percentile. Only the NFS GAN was trained for the Continental region.

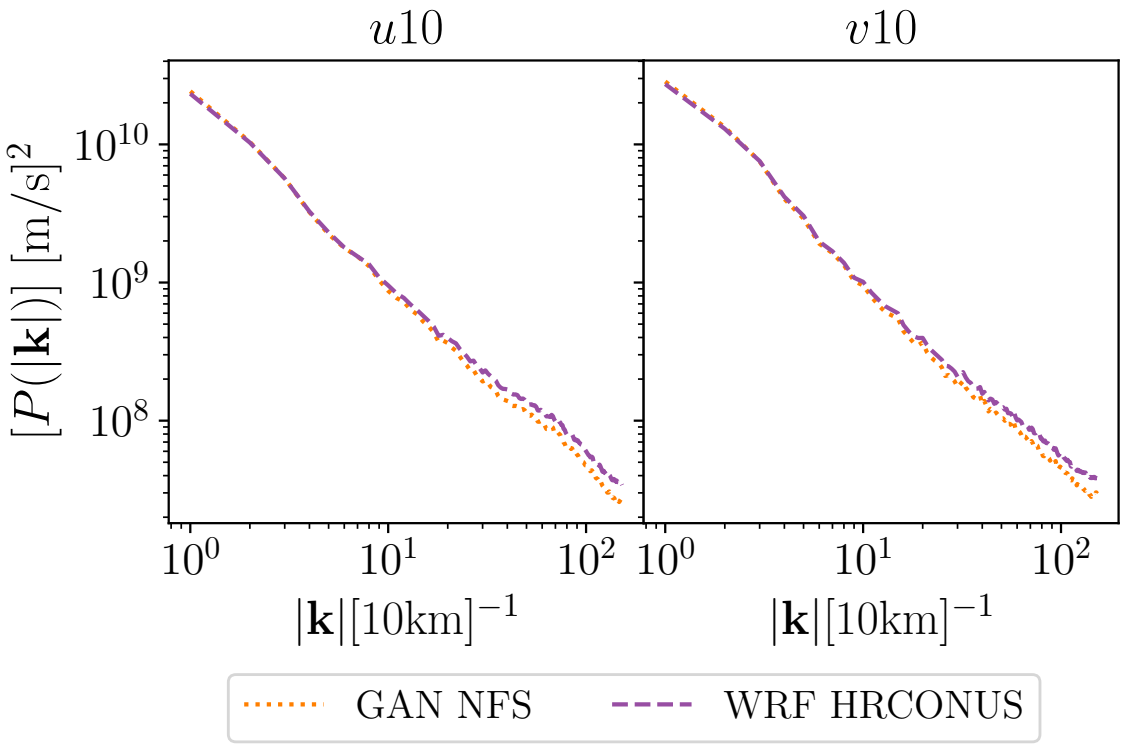


Figure B.12: Power spectra of each wind component for the continental scale SR fields.

# References

P. Ailliot, D. Allard, V. Monbet, and P. Naveau. Stochastic weather generators: an overview of weather type models. *Journal de la société française de statistique*, 156 (1):101–113, 2015. URL http://www.numdam.org/item/JSFS_2015__156_1_101_0/. Publisher: Société française de statistique.

J. Angwin, J. Larson, S. Mattu, and L. Kirchner. Machine Bias. *ProPublica*, May 2016. URL https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv:1701.07875 [cs, stat]*, Dec. 2017. URL http://arxiv.org/abs/1701.07875. arXiv: 1701.07875.

N. Ballas, L. Yao, C. Pal, and A. Courville. Delving deeper into convolutional networks for learning video representations. *arXiv preprint arXiv:1511.06432*, 2015.

N. Ban, J. Schmidli, and C. Schär. Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, 42 (4):1165–1172, 2015. ISBN: 0094-8276 Publisher: Wiley Online Library.

R. Benjamin. *Race after technology : abolitionist tools for the New Jim Code*. Polity Press, Cambridge, UK ;, 2019. ISBN 978-1-5095-2639-0.

J. Buolamwini and T. Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Machine Learning Research 81:1–15, 2018*, 2018.

E. Bush and D. Lemmen. *Canada's changing climate report.* Environment and Climate Change Canada = Environnement et changement climatique Canada, Gatineau, QC, 2019. ISBN 978-0-660-30222-5.

A. J. Cannon. Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network. *Journal of Hydrometeorology*, 9(6):1284–1300, Dec. 2008. ISSN 1525-7541, 1525-755X. doi: 10.1175/2008JHM960.1. URL http://journals.ametsoc.org/view/journals/hydr/9/6/2008jhm960_1.xml. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

M. S. Cataleta. Humane Artificial Intelligence. Technical report, East-West Center, 2020. URL http://www.jstor.org.ezproxy.library.uvic.ca/stable/resrep25514.

R. E. Chandler. On the use of generalized linear models for interpreting climate variability. *Environmetrics*, 16(7):699–715, Nov. 2005. ISSN 1180-4009. doi: 10.1002/env.731. URL https://doi.org/10.1002/env.731. Publisher: John Wiley & Sons, Ltd.

R. E. Chandler and H. S. Wheater. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, 38(10):10–1–10–11, 2002. doi: https://doi.org/10.1029/2001WR000906. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001WR000906. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001WR000906.

M. Chantry, H. Christensen, P. Dueben, and T. Palmer. Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200083, Apr. 2021. doi: 10.1098/rsta.2020.0083. URL https://doi.org/10.1098/rsta.2020.0083. Publisher: Royal Society.

A. Chen, A. Chow, A. Davidson, A. DCunha, A. Ghodsi, S. A. Hong, A. Konwinski, C. Mewald, S. Murching, T. Nykodym, P. Ogilvie, M. Parkhe, A. Singh, F. Xie, M. Zaharia, R. Zang, J. Zheng, and C. Zumar. Developments in MLflow: A System to Accelerate the Machine Learning Lifecycle. In *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning*, DEEM'20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-8023-2. doi: 10.1145/3399579.3399867. URL https://doi-org.ezproxy.library.uvic.ca/10.1145/3399579.3399867. event-place: Portland, OR, USA.

J. Cheng, J. Liu, Z. Xu, C. Shen, and Q. Kuang. Generating High-Resolution Climate Prediction through Generative Adversarial Network. *Procedia Computer Science*, 174:123–127, 2020. ISSN 18770509. doi: 10.1016/j.procs.2020.06.067. URL `https://linkinghub.elsevier.com/retrieve/pii/S1877050920315817`.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011. ISSN 1477-870X. doi: 10.1002/qj.828. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.828`. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.828.

C. Dong, C. C. Loy, K. He, and X. Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, Lecture Notes in Computer Science, pages 184–199, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10593-2. doi: 10.1007/978-3-319-10593-2_13.

C. Dong, C. C. Loy, K. He, and X. Tang. Image Super-Resolution Using Deep Convolutional Networks. *arXiv:1501.00092 [cs]*, July 2015. URL `http://arxiv.org/abs/1501.00092`. arXiv: 1501.00092.

M. Déqué, D. P. Rowell, D. Lüthi, F. Giorgi, J. H. Christensen, B. Rockel, D. Jacob, E. Kjellström, M. de Castro, and B. van den Hurk. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change*, 81(1):53–70, May 2007. ISSN 1573-1480. doi: 10.1007/s10584-006-9228-x. URL `https://doi.org/10.1007/s10584-006-9228-x`.

A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE transactions on communications*, 43(12):2959–2965, 1995. ISSN 0090-6778. doi: 10.1109/26.477498. URL `https://go.exlibris.link/8C2hrTHk`. Publisher: IEEE.

C. Frei, J. H. Christensen, M. Déqué, D. Jacob, R. G. Jones, and P. L. Vidale. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research: Atmospheres*, 108(D3), 2003. ISBN: 0148-0227 Publisher: Wiley Online Library.

M. Fritsche, S. Gu, and R. Timofte. Frequency Separation for Real-World Super-Resolution. *arXiv:1911.07850 [cs, eess]*, Nov. 2019. URL `http://arxiv.org/abs/1911.07850`. arXiv: 1911.07850.

K. R. Gabriel and J. Neumann. A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375):90–95, 1962. ISSN 1477-870X. doi: 10.1002/qj.49708837511. URL `http://onlinelibrary.wiley.com/doi/abs/10.1002/qj.49708837511`. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49708837511.

M. W. Gardner and S. R. Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14):2627–2636, Aug. 1998. ISSN 1352-2310. doi: 10.1016/S1352-2310(97)00447-0. URL `https://www.sciencedirect.com/science/article/pii/S1352231097004470`.

J. D. Garrett. garrettj403/SciencePlots. Sept. 2021. doi: 10.5281/zenodo.4106649. URL `http://doi.org/10.5281/zenodo.4106649`. Publisher: Zenodo Version Number: 1.0.9.

M. Ghil. A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability? In *Climate change: Multidecadal and beyond*, pages 31–51. World Scientific, 2016.

F. Giorgi and L. O. Mearns. Approaches to the simulation of regional climate change: A review. *Reviews of Geophysics*, 29(2): 191–216, 1991. ISSN 1944-9208. doi: 10.1029/90RG02636. URL `http://onlinelibrary.wiley.com/doi/abs/10.1029/90RG02636`. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/90RG02636.

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014. URL `http://arxiv.org/abs/1406.2661`. arXiv: 1406.2661.

I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved Training of Wasserstein GANs. *arXiv:1704.00028 [cs, stat]*, Dec. 2017. URL `http://arxiv.org/abs/1704.00028`. arXiv: 1704.00028.

C. R. Harris, K. J. Millman, S. J. v. d. Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. v. Kerkwijk, M. Brett, A. Haldane, J. F. d. Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`. Publisher: Springer Science and Business Media LLC.

L. Harris, A. T. T. McRae, M. Chantry, P. D. Dueben, and T. N. Palmer. A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts. 2022.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, July 2020. ISSN 0035-9009. doi: 10.1002/qj.3803. URL `https://doi.org/10.1002/qj.3803`. Publisher: John Wiley & Sons, Ltd.

R. A. Houze Jr. Mesoscale convective systems. *Reviews of geophysics (1985)*, 42 (4):RG4003–n/a, 2004. ISSN 8755-1209. doi: 10.1029/2004RG000150. Edition: Houze, R. A.Jr. (2004), Mesoscale convective systems, Rev. Geophys., 42, RG4003, doi:10.1029/2004RG000150. Publisher: American Geophysical Union.

W. W. Hsieh. *Machine learning methods in the environmental sciences : neural networks and kernels*. Cambridge University Press, Cambridge, 2009. ISBN 1-108-45690-1. Publication Title: Machine learning methods in the environmental sciences : neural networks and kernels.

J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55. Publisher: IEEE COMPUTER SOC.

S. Innocenti, A. Mailhot, A. Frigon, A. J. Cannon, and M. Leduc. Observed and Simulated Precipitation over Northeastern North America: How Do Daily and Subdaily Extremes Scale in Space and Time? *Journal of Climate*, 32(24):8563–8582, Dec. 2019. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-19-0021.1. URL https://journals.ametsoc.org/view/journals/clim/32/24/jcli-d-19-0021.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

D. I. Jeong, A. St-Hilaire, T. B. M. J. Ouarda, and P. Gachon. Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator. *Climatic Change*, 114(3):567–591, Oct. 2012. ISSN 1573-1480. doi: 10.1007/s10584-012-0451-3. URL https://doi.org/10.1007/s10584-012-0451-3.

L. V. Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4):1381–1382, 2006. Publisher: Kluwer Academic Publishers-Consultants Bureau.

A. Karpathy. CS231n Convolutional Neural Networks for Visual Recognition, Oct. 2022. URL https://cs231n.github.io/convolutional-networks/#add.

K. Kashinath, M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and Prabhat. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200093, Apr. 2021. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2020.0093. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0093.

R. W. Katz. Precipitation as a Chain-Dependent Process. *Journal of Applied Meteorology and Climatology*, 16(7):671–676, July 1977. ISSN 1520-0450. doi: 10.1175/1520-0450(1977)016<0671:PAACDP>2.0.CO;2. URL http://journals.ametsoc.org/view/journals/apme/16/7/1520-0450_1977_

016_0671_paacdp_2_0_co_2.xml. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

V. V. Kharin, F. W. Zwiers, X. Zhang, and G. C. Hegerl. Changes in Temperature and Precipitation Extremes in the IPCC Ensemble of Global Coupled Model Simulations. *Journal of Climate*, 20(8):1419–1444, Apr. 2007. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI4066.1. URL https://journals.ametsoc.org/view/journals/clim/20/8/jcli4066.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, Jan. 2017. URL http://arxiv.org/abs/1412.6980. arXiv: 1412.6980.

B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge, and A. K. Jain. Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security*, 7(6):1789–1801, 2012. doi: 10.1109/TIFS.2012.2214212.

P. Kopparla, E. M. Fischer, C. Hannay, and R. Knutti. Improved simulation of extreme precipitation in a high-resolution atmosphere model. *Geophysical Research Letters*, 40(21):5803–5808, 2013. ISSN 1944-8007. doi: 10.1002/2013GL057866. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2013GL057866. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL057866.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

B. Kumar, R. Chattopadhyay, M. Singh, N. Chaudhari, K. Kodari, and A. Barve. Deep learning–based downscaling of summer monsoon rainfall data over Indian region. *Theoretical and Applied Climatology*, 143(3):1145–1156, Feb. 2021. ISSN 1434-4483. doi: 10.1007/s00704-020-03489-6. URL https://doi.org/10.1007/s00704-020-03489-6.

C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *arXiv:1609.04802 [cs, stat]*, May 2017. URL http://arxiv.org/abs/1609.04802. arXiv: 1609.04802.

J. Leinonen, D. Nerini, and A. Berne. Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13, 2020. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2020.3032790. URL https://ieeexplore.ieee.org/document/9246532/.

A. LeNail. NN-SVG: Publication-Ready Neural Network Architecture Schematics. *Journal of Open Source Software*, 4(33):747, 2019. doi: 10.21105/joss.00747. URL https://doi.org/10.21105/joss.00747. Publisher: The Open Journal.

G. Li, X. Zhang, A. J. Cannon, T. Murdock, S. Sobie, F. Zwiers, K. Anderson, and B. Qian. Indices of Canada's future climate for general and agricultural adaptation applications. *Climatic Change*, 148(1):249–263, May 2018. ISSN 1573-1480. doi: 10.1007/s10584-018-2199-x. URL https://doi.org/10.1007/s10584-018-2199-x.

C. Liu, K. Ikeda, R. Rasmussen, M. Barlage, A. J. Newman, A. F. Prein, F. Chen, L. Chen, M. Clark, A. Dai, J. Dudhia, T. Eidhammer, D. Gochis, E. Gutmann, S. Kurkute, Y. Li, G. Thompson, and D. Yates. Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, 49(1):71–95, July 2017. ISSN 1432-0894. doi: 10.1007/s00382-016-3327-9. URL https://doi.org/10.1007/s00382-016-3327-9.

D. Maraun and M. Widmann. *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, Cambridge, 2018. ISBN 978-1-107-06605-2. doi: 10.1017/9781107588783. URL https://www.cambridge.org/core/books/statistical-downscaling-and-bias-correction-for-climate-research/4ED479BAA8309C7ECBE6136236E3960F.

A. McGovern, I. Ebert-Uphoff, D. J. Gagne, and A. Bostrom. Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science*, 1:e6, 2022. doi: 10.1017/

eds.2022.5. URL https://doi.org/10.1017/eds.2022.5. Edition: 2022/04/13 Publisher: Cambridge University Press.

S. C. Michaelides. *Precipitation: Advances in Measurement, Estimation and Prediction*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1st ed. 2008. edition, 2008. ISBN 1-281-23156-8. doi: 10.1007/978-3-540-77655-0.

M. Mirza and S. Osindero. Conditional Generative Adversarial Nets. *arXiv:1411.1784 [cs, stat]*, Nov. 2014. URL http://arxiv.org/abs/1411.1784. arXiv: 1411.1784.

K. P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL probml.ai.

Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science (American Association for the Advancement of Science)*, 366(6464):447–453, 2019. ISSN 0036-8075. doi: 10.1126/science.aax2342. Place: United States Publisher: American Association for the Advancement of Science.

C. O'Neil. *Weapons of math destruction : how big data increases inequality and threatens democracy*. Broadway Books, New York, NY, first paperback edition. edition, 2017. ISBN 978-0-553-41883-5.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in PyTorch. 2017.

D. A. Plummer, D. Caya, A. Frigon, H. Côté, M. Giguère, D. Paquin, S. Biner, R. Harvey, and R. De Elia. Climate and climate change over North America as simulated by the Canadian RCM. *Journal of climate*, 19(13):3112–3132, 2006. ISBN: 1520-0442.

A. F. Prein, W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tölle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N. P. M. v. Lipzig, and R. Leung. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2):323–361, 2015. ISSN 1944-9208. doi: https://doi.org/10.1002/2014RG000475. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000475. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014RG000475.

A. F. Prein, A. Gobiet, H. Truhetz, K. Keuler, K. Goergen, C. Teichmann, C. Fox Maule, E. van Meijgaard, M. Déqué, G. Nikulin, R. Vautard, A. Colette, E. Kjellström, and D. Jacob. Precipitation in the EURO-CORDEX 0.11 deg and 0.44 deg simulations: high resolution, high benefits? *Climate Dynamics*, 46(1):383–412, Jan. 2016. ISSN 1432-0894. doi: 10.1007/s00382-015-2589-y. URL https://doi.org/10.1007/s00382-015-2589-y.

W. H. Press. *Numerical recipes : the art of scientific computing*. Cambridge University Press, Cambridge [Cambridgeshire] ;, 1986. ISBN 0-521-30811-9. Publication Title: Numerical recipes : the art of scientific computing.

I. Price and S. Rasp. Increasing the accuracy and resolution of precipitation forecasts using deep generative models. In *International Conference on Artificial Intelligence and Statistics*, pages 10555–10571. PMLR, 2022. ISBN 2640-3498.

Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin. Variational Autoencoder for Deep Learning of Images, Labels and Captions. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/eb86d510361fc23b59f18c1bc9802cc6-Paper.pdf.

T. Pucik, P. Groenemeijer, and I. Tsonevsky. Vertical wind shear and convective storms, Jan. 2021. URL https://www.ecmwf.int/node/19905. Issue: 879 Publication Title: ECMWF Technical Memoranda.

A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2015.

A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434 [cs]*, Jan. 2016. URL http://arxiv.org/abs/1511.06434. arXiv: 1511.06434.

R. Rasmussen and C. Liu. High Resolution WRF Simulations of the Current and Future Climate of North America. Feb. 2017. doi: 10.5065/D6V40SXP. URL https://rda.ucar.edu/datasets/ds612.0/. Publisher: UCAR/NCAR - Research Data Archive Type: dataset.

C. Richardson and D. Wright. WGEN: A model for generating daily weather variables. ARS-8. *US Department of Agriculture, Agricultural Research Service, Washington, DC*, page 83, 1984.

A. Rossa, P. Nurmi, and E. Ebert. *Overview of methods for the verification of quantitative precipitation forecasts*. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 3-540-77654-0. doi: 10.1007/978-3-540-77655-0_16.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct. 1986. ISSN 1476-4687. doi: 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

D. San-Martín, R. Manzanas, S. Brands, S. Herrera, and J. M. Gutiérrez. Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods. *Journal of Climate*, 30(1):203–223, Jan. 2017. ISSN 0894-8755, 1520-0442. doi: 10.1175/JCLI-D-16-0366.1. URL http://journals.ametsoc.org/view/journals/clim/30/1/jcli-d-16-0366.1.xml. Publisher: American Meteorological Society Section: Journal of Climate.

C. Schlager, G. Kirchengast, J. Fuchsberger, A. Kann, and H. Truhetz. A spatial evaluation of high-resolution wind fields from empirical and dynamical modeling in hilly and mountainous terrain. *Geoscientific Model Development*, 12(7):2855–2873, July 2019. ISSN 1991-9603. doi: 10.5194/gmd-12-2855-2019. URL https://gmd.copernicus.org/articles/12/2855/2019/.

U. Schulzweida. CDO User Guide, Oct. 2022. URL https://doi.org/10.5281/zenodo.7112925. Version Number: 2.1.0.

S. I. Seneviratne, N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, X. Zhang, M. Rusticucci, V. Semenov, L. V. Alexander, S. Allen, G. Benito, T. Cavazos, J. Clague, D. Conway, P. M. Della-Marta, M. Gerber, S. Gong, B. N. Goswami, M. Hemer, C. Huggel, B. van den Hurk, V. V. Kharin, A. Kitoh, A. M. K.

Tank, G. Li, S. Mason, W. McGuire, G. J. van Oldenborgh, B. Orlowsky, S. Smith, W. Thiaw, A. Velegrakis, P. Yiou, T. Zhang, T. Zhou, and F. W. Zwiers. Changes in Climate Extremes and their Impacts on the Natural Physical Environment. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, pages 109–230. 2012. ISBN 1-107-60780-9.

Y. Sha, D. J. G. II, G. West, and R. Stull. Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part I: Daily Maximum and Minimum 2-m Temperature. *Journal of Applied Meteorology and Climatology*, 59(12):2057 – 2073, 2020. doi: 10.1175/JAMC-D-20-0057.1. URL `https://journals.ametsoc.org/view/journals/apme/59/12/jamc-d-20-0057.1.xml`. Place: Boston MA, USA Publisher: American Meteorological Society.

A. Shocher, N. Cohen, and M. Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126, 2018.

J. Sillmann, V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, 118(4):1716–1733, 2013. ISSN 2169-8996. doi: 10.1002/jgrd.50203. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50203`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50203.

A. Singh, A. Albert, and B. White. Downscaling Numerical Weather Models with GANs. page 4, École Normale Supérieure, Paris, France, 2019.

S. R. Sobie and T. Q. Murdock. High-Resolution Statistical Downscaling in Southwestern British Columbia. *Journal of Applied Meteorology and Climatology*, 56(6):1625–1641, June 2017. ISSN 1558-8424, 1558-8432. doi: 10.1175/JAMC-D-16-0287.1. URL `https://journals.ametsoc.org/view/journals/apme/56/6/jamc-d-16-0287.1.xml`. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

J.-H. Song, Y. Her, S. Shin, J. Cho, R. Paudel, Y. P. Khare, J. Obeysekera, and C. J. Martinez. Evaluating the performance of climate models in reproducing the hydrological characteristics of rainfall events. *Hydrological sciences journal*, 65(9):

1490–1511, 2020. ISSN 0262-6667. doi: 10.1080/02626667.2020.1750616. Publisher: Taylor & Francis.

K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, July 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1918964117. URL https://www.pnas.org/content/117/29/16805. Publisher: National Academy of Sciences Section: Physical Sciences.

G. L. Stephens, T. L'Ecuyer, R. Forbes, A. Gettelmen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes. Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres*, 115(D24), 2010. ISSN 2156-2202. doi: 10.1029/2010JD014532. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014532. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014532.

T. Stocker. *Climate change 2013 : the physical science basis : Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 2014. ISBN 1-139-90550-3.

P. D. Thompson. Uncertainty of Initial State as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns. *Tellus*, 9(3):275–295, Aug. 1957. ISSN 0040-2826. doi: 10.1111/j.2153-3490.1957.tb01885.x. URL https://doi.org/10.1111/j.2153-3490.1957.tb01885.x. Publisher: John Wiley & Sons, Ltd.

C. Torma, F. Giorgi, and E. Coppola. Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps. *Journal of geophysical research. Atmospheres*, 120(9):3957–3972, 2015. ISSN 2169-897X. doi: 10.1002/2014JD022781. Edition: Torma, Cs., F. Giorgi, and E. Coppola (2015), Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps, J. Geophys. Res. Atmos., 120, 3957-3972. doi: 10.1002/2014JD022781. Publisher: Blackwell Publishing Ltd.

M. Turco, M. C. Llasat, S. Herrera, and J. M. Gutiérrez. Bias correction and downscaling of future RCM precipitation projections using a MOS-Analog technique. *Journal of Geophysical Research: Atmospheres*, 122(5):2631–2648, Mar. 2017. ISSN 2169-897X. doi: 10.1002/2016JD025724. URL https://doi.org/10.1002/2016JD025724. Publisher: John Wiley & Sons, Ltd.

G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1-4414-1269-7.

A. Vedaldi, G. Montavon, L. K. Hansen, W. Samek, and K.-R. Muller. *Explainable AI: interpreting, explaining and visualizing deep learning*, volume 11700 of *LNCS sublibrary. SL 7, Artificial intelligence*. Springer, 2019. ISBN 3-030-28953-2.

C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg, 2009. ISBN 978-3-540-71049-3. doi: 10.1007/978-3-540-71050-9. URL `https://www.springer.com/gp/book/9783540710493`.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

G. Vissio, V. Lembo, V. Lucarini, and M. Ghil. Evaluating the Performance of Climate Models Based on Wasserstein Distance. *Geophysical Research Letters*, 47 (21):e2020GL089385, Nov. 2020. ISSN 0094-8276. doi: 10.1029/2020GL089385. URL `https://doi.org/10.1029/2020GL089385`. Publisher: John Wiley & Sons, Ltd.

F. Wang, D. Tian, L. Lowe, L. Kalin, and J. Lehrter. Deep Learning for Daily Precipitation and Temperature Downscaling. *Water Resources Research*, 57(4): e2020WR029308, Apr. 2021. ISSN 0043-1397. doi: 10.1029/2020WR029308. URL `https://doi.org/10.1029/2020WR029308`. Publisher: John Wiley & Sons, Ltd.

X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *arXiv:1809.00219 [cs]*, Sept. 2018. URL `http://arxiv.org/abs/1809.00219`. arXiv: 1809.00219.

Z. Wang, E. Simoncelli, and A. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, pages 1398–1402, Pacific Grove, CA, USA, 2003. IEEE. ISBN 978-0-7803-8104-9. doi: 10.1109/ACSSC.2003.1292216. URL http://ieeexplore.ieee.org/document/1292216/.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861. URL https://go.exlibris.link/gpQ4fmk4. Place: New York, NY Publisher: IEEE.

S. Watt-Cloutier. *The right to be cold : one woman's story of protecting her culture, the Arctic, and the whole planet*. Penguin, Toronto, Ontario, Canada, 2016. ISBN 978-0-14-318764-6.

C. D. Whiteman. Mountain Climates of North America. In *Mountain Climates of North America*. Oxford University Press, 2000. doi: 10.1093/oso/9780195132717.003.0008. URL https://oxford.universitypressscholarship.com/10.1093/oso/9780195132717.001.0001/isbn-9780195132717-book-part-8.

R. L. Wilby and T. M. L. Wigley. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, 21:530 – 548, 1997.

D. S. Wilks and R. L. Wilby. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography: Earth and Environment*, 23(3): 329–357, Sept. 1999. ISSN 0309-1333. doi: 10.1177/030913339902300302. URL https://doi.org/10.1177/030913339902300302. Publisher: SAGE Publications Ltd.

K. Yamashita and K. Markov. Medical Image Enhancement Using Super Resolution Methods. In V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, editors, *Computational Science – ICCS 2020*, pages 496–508, Cham, 2020. Springer International Publishing. ISBN 978-3-030-50426-7.

D. Yeung, I. Khan, N. Kalra, and O. A. Osoba. Identifying Systemic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Applications. Tech-

nical report, RAND Corporation, 2021. URL http://www.jstor.org.ezproxy.library.uvic.ca/stable/resrep29576.

Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual Dense Network for Image Super-Resolution. *arXiv:1802.08797 [cs]*, Mar. 2018. URL http://arxiv.org/abs/1802.08797. arXiv: 1802.08797.

Y. Zhang, Z. Zhang, S. DiVerdi, Z. Wang, J. Echevarria, and Y. Fu. Texture hallucination for large-factor painting super-resolution. In *European Conference on Computer Vision*, pages 209–225. Springer, 2020.

Zhou Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, Mar. 2002. ISSN 1558-2361. doi: 10.1109/97.995823.

X. Zhu, L. Zhang, L. Zhang, X. Liu, Y. Shen, and S. Zhao. GAN-Based Image Super-Resolution with a Novel Quality Loss. *Mathematical Problems in Engineering*, 2020:e5217429, Feb. 2020. ISSN 1024-123X. doi: 10.1155/2020/5217429. URL https://www.hindawi.com/journals/mpe/2020/5217429/. Publisher: Hindawi.