



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE  
CORSO DI LAUREA IN INGEGNERIA INFORMATICA

# **Valutazione del foundation model Aurora per il supporto alle attività operative di ARPA FVG**

**Relatore**

Prof. Varagnolo Damiano

**Correlatore**

Dott. Montanari Francesco

**Laureando**

Stefanuto Giovanni

ANNO ACCADEMICO 2024-2025

Data di laurea 23/09/2025



# Abstract

La tesi documenta un tirocinio svolto presso il Centro Regionale di Modellistica Ambientale di ARPA FVG, incentrato sull'analisi e valutazione del foundation model Aurora (Microsoft) per applicazioni di previsione atmosferica, della qualità dell'aria e delle onde oceaniche. Le attività hanno incluso lo studio della documentazione tecnica, test su dataset selezionati e confronto tra infrastrutture locali e cloud. Il lavoro ha permesso di acquisire competenze su HPC, workflow automation, versionamento software e strumenti AI in Python, offrendo spunti per una valutazione strategica dell'uso di modelli fondazionali in ambito ambientale.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Contesto Operativo: la Previsione Meteorologica in Arpa FVG . . . . .	1
1.1.1	Il Modello WRF: Strumento di Previsione Deterministica . . . . .	2
1.2	La Rivoluzione dei Foundation Models in Ambito Scientifico . . . . .	4
1.3	Stato dell'Arte: il Modello Aurora . . . . .	5
1.3.1	Aurora: un Foundation Model per l'Atmosfera . . . . .	5
1.3.2	Dati e Fasi di Addestramento . . . . .	6
1.3.3	Prestazioni e Benchmark Pubblicati . . . . .	7
<b>2</b>	<b>Scopo della Tesi</b>	<b>9</b>
2.1	Obiettivi e Domande di Ricerca . . . . .	9
2.2	Struttura dell'Elaborato . . . . .	10
<b>3</b>	<b>Metodologia e Ambiente Sperimentale</b>	<b>11</b>
3.1	Infrastrutture a Confronto: Google Colab vs. HPC Locale . . . . .	11
3.1.1	Test Preliminari in Ambiente Cloud: Google Colab . . . . .	11
3.1.2	L'Infrastruttura Principale: il Cluster HPC "Fenice" . . . . .	12
3.2	Pipeline di Elaborazione Adottata: Adattare i Dati WRF per Aurora . . . . .	13
3.3	Metodologia di Valutazione Sequenziale . . . . .	16
3.3.1	La Tecnica del "Rollout" per Previsioni a Lungo Termine . . . . .	16
3.3.2	Criteri di Analisi dei Risultati . . . . .	16
<b>4</b>	<b>Risultati</b>	<b>17</b>
4.1	Benchmark di Prestazione e Limiti Infrastrutturali . . . . .	17
4.1.1	Confronto delle Prestazioni di Inferenza: CPU vs. GPU . . . . .	17
4.1.2	Limiti di Memoria e Validazione dell'Infrastruttura Locale . . . . .	18
4.2	Dalla Teoria alla Pratica: Generazione del Primo Output . . . . .	19
4.3	Il Fallimento del Rollout e l'Ipotesi del Dataset Mismatch . . . . .	21

<b>5</b>	<b>Conclusioni</b>	<b>23</b>
5.1	Sintesi del Lavoro Svolto . . . . .	23
5.2	Risultati Chiave e Risposta alle Domande di Ricerca . . . . .	23
5.3	Discussione: la Lezione Appresa . . . . .	24
5.4	Proposte per Sviluppi Futuri . . . . .	25
5.5	Riflessioni Finali e Competenze Acquisite . . . . .	26
	<b>Bibliografia</b>	<b>27</b>

# 1. Introduzione

## 1.1 Contesto Operativo: la Previsione Meteorologica in Arpa FVG

La previsione meteorologica è un'attività cruciale per la gestione del territorio ed il supporto alle operazioni di Protezione Civile. In Friuli Venezia Giulia questo compito è affidato all'*Agenzia Regionale per la Protezione dell'Ambiente (Arpa FVG)*, con il supporto del suo *Centro Regionale di Modellistica Ambientale (CRMA)*, dove si è svolto il tirocinio oggetto di questa tesi.

Per comprendere a fondo l'ambito di applicazione di questa valutazione, è fondamentale delineare il contesto in cui opera il CRMA.

L'Arpa FVG è l'ente pubblico responsabile del monitoraggio e della protezione ambientale sul territorio regionale. Essa si inserisce nel *Sistema Nazionale per la Protezione dell'Ambiente (SNPA)*, una rete di agenzie coordinate dall'*Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA)*. La struttura organizzativa dell'Agenzia, come delineato nel suo regolamento, si articola in una Direzione Generale da cui dipendono una Direzione Amministrativa e una Direzione Tecnico-Scientifica. A quest'ultima afferiscono diverse Strutture Organizzative Complesse (**S.O.C.**) che ne costituiscono il braccio operativo. [1] Tra queste figurano:

- la S.O.C. "Pressioni sull'ambiente", che si occupa del controllo delle pressioni di origine antropica;
- la S.O.C. "Laboratorio", dedicata alle analisi chimiche e biologiche;
- la S.O.C. "Stato dell'Ambiente", che ha il compito di monitorare e valutare lo stato delle matrici ambientali.

Le attività dell'Agenzia sono concettualmente guidate dal modello **DPSIR** (Determinanti, Pressioni, Stato, Impatti, Risposte), un approccio strutturato per analizzare le problematiche ambientali. [2] Questo schema parte dalle cause di fondo (Determinanti), analizza le attività umane che esercitano un'azione diretta sull'ambiente (Pressioni), ne misura gli effetti sullo stato delle

risorse naturali (Stato), valuta le alterazioni significative (Impatti) e infine definisce le azioni correttive (Risposte). All'interno di questo schema il monitoraggio e la previsione meteorologica rappresentano strumenti fondamentali. Tali funzioni sono affidate specificamente alla S.O.C. “*Osservatorio Meteorologico Regionale*” (**OSMER**), che si occupa dell'osservazione e della previsione dei fenomeni meteorologici, provvedendo alla diffusione delle informazioni e dei servizi derivati. All'interno di OSMER opera il CRMA, il cui compito principale è fornire supporto tecnico agli enti locali e sviluppare, nonché gestire, la modellistica ambientale e atmosferica a servizio delle attività previsionali. L'operato di OSMER è cruciale non solo per la conoscenza dello stato dell'ambiente, ma anche come supporto diretto alle attività della Protezione Civile, rappresentando un pilastro per la gestione delle emergenze e la salvaguardia del territorio. [3]

Proprio per adempiere a questo mandato, il CRMA si avvale di una complessa suite di strumenti di simulazione: l'attività previsionale attuale si basa prevalentemente su complessi modelli numerici deterministici, come il **WRF** (*Weather Research and Forecasting*), che verrà analizzato nel paragrafo successivo.

### 1.1.1 Il Modello WRF: Strumento di Previsione Deterministica

Il principale strumento modellistico su cui si basa l'attività previsionale del CRMA è il **Weather Research and Forecasting (WRF) Model**. [4] Sviluppato e supportato dal *National Center for Atmospheric Research* (NCAR) degli Stati Uniti, il WRF è un modello atmosferico comunitario, progettato sia per la ricerca scientifica che per la previsione meteorologica numerica (NWP, *Numerical Weather Prediction*) [4]

Il WRF è un modello numerico deterministico. Ciò significa che, a partire da un insieme di condizioni iniziali e condizioni al contorno, la sua evoluzione futura è univocamente determinata.[5] Esso risolve numericamente un sistema di equazioni differenziali alle derivate parziali che descrivono i processi fisici e dinamici dell'atmosfera, come fluidodinamica, termodinamica, i processi radiativi e la microfisica delle nubi. Essendo un modello ad area limitata (*limited-area model*), per operare necessita di dati provenienti da modelli a copertura globale (GCM, *Global Circulation Model*), che forniscono le cosiddette “condizioni al contorno” per l'intera durata della simulazione.[4]

L'implementazione operativa in Arpa FVG sfrutta proprio questa caratteristica per ottenere previsioni ad alta risoluzione. La procedura prevede l'utilizzo di tre domini di calcolo annidati (*nested domains*): il primo, più esterno, copre una vasta area geografica con una risoluzione spaziale più grossolana; i successivi si focalizzano su aree progressivamente più piccole con una risoluzione maggiore. Questo approccio multilivello permette di raggiungere un dettaglio orizzontale di 2 km x 2 km sul dominio più interno, che copre la regione chiamata “*AlpeAdria*”.



Le condizioni al contorno per il dominio più esterno sono fornite dal modello globale **GFS** del *National Oceanic and Atmospheric Administration* (NOAA). Il modello viene eseguito una volta al giorno sull'infrastruttura di calcolo ad alte prestazioni (HPC) dell'Agenzia per produrre una previsione a cinque giorni (120 ore) ad ogni run. Come la maggior parte dei modelli di questa categoria, il codice sorgente è scritto principalmente in linguaggio Fortran, scelto per garantire elevate prestazioni computazionali [4].

È importante sottolineare che il WRF, sebbene centrale, è solo uno degli strumenti nella suite modellistica del CRMA. A seconda delle necessità applicative, vengono impiegati modelli specifici per la qualità dell'aria (es. FARM, SPRAY), per la dispersione di inquinanti in emergenza (es. DELFI) o per la simulazione delle condizioni marine (es. SHYFEM).

**L'Importanza del Pre e Post-Processing** L'esecuzione di un modello numerico come il WRF è solo una parte di un flusso di lavoro più complesso, che include fasi cruciali di pre-processing e post-processing.

Il **pre-processing** è gestito da un insieme di utility noto come *WRF Preprocessing System* (WPS) [4]. Questa fase è dedicata alla preparazione degli input per il modello: vengono definite le aree geografiche dei domini di calcolo, si acquisiscono i dati statici come la topografia e l'uso del suolo, e si interpolano i dati meteorologici del modello globale (le condizioni iniziali e al contorno) sulla griglia di calcolo del WRF.

Ancor più rilevante per le finalità operative dell'Agenzia è l'attività di **post-processing**. L'output grezzo del modello è costituito da matrici numeriche multi-dimensionali, di difficile interpretazione per un utente non esperto. Questa fase è quindi fondamentale per trasformare i dati in informazione utile e comunicabile, sia per i bollettini destinati al pubblico e alla Protezione Civile, sia per ulteriori analisi. In Arpa FVG può succedere che i dati prodotti dal modello vengono spesso elaborati e integrati con le misure provenienti dalla rete di stazioni al suolo. Attraverso tecniche avanzate, come il *Kriging* [6], è possibile combinare l'informazione misurata (puntuale e accurata) con quella modellistica (distribuita spazialmente ma soggetta a incertezza), ottenendo una stima più robusta e rappresentativa dello stato dell'atmosfera sul territorio.

**Limiti Operativi e Costi Computazionali** Pur garantendo un'elevata affidabilità, i modelli numerici deterministici come il WRF presentano sfide operative significative, in particolare per quanto riguarda i costi computazionali. L'esecuzione di una singola previsione a cinque giorni richiede un notevole dispendio di risorse: l'infrastruttura di ARPA FVG impegna quotidianamente 160 core di calcolo per circa 2.5 ore. Questo si traduce in un costo operativo considerevole che, se esternalizzato, ammonterebbe a circa 54.000 euro all'anno, giustificando quindi gli investimenti in cluster di calcolo proprietari, come ad esempio il **cluster Fenice** (oggetto di discussione nei prossimi paragrafi).

Tali costi e i lunghi tempi di esecuzione impongono una rigida pianificazione operativa: un programma specifico per pipeline automatizzate (**ecFlow**) che scarica durante la notte i dati delle condizioni al contorno dal modello globale GFS e lancia l'esecuzione del WRF nelle prime ore del mattino.[7] Questo flusso di lavoro è necessario per garantire che le previsioni aggiornate siano disponibili per l'analisi e la sintesi da parte del previsore all'inizio della giornata lavorativa. L'introduzione di modelli basati sull'intelligenza artificiale potrebbe potenzialmente abbattere questo limite temporale, un aspetto che verrà approfondito nei capitoli successivi di questa tesi.

A queste sfide si aggiungono limiti intrinseci legati alla natura di questi modelli. La risoluzione spaziale, per quanto elevata, impedisce al modello di "vedere" i fenomeni che si sviluppano a scale inferiori, come le dinamiche iper-locali all'interno di strette vallate alpine, rendendo l'output grezzo insufficiente per previsioni di dettaglio. Inoltre, sebbene questi modelli possano identificare condizioni generali di instabilità, spesso faticano a determinare con precisione il punto e l'istante esatto di innesco di eventi estremi e localizzati, come un temporale. [5]

L'emergere di nuovi approcci basati sull'intelligenza artificiale, e in particolare dei *foundation model*, offre una prospettiva promettente per superare tali limiti, affiancando o potenziando gli strumenti tradizionali. Questo lavoro di tesi si inserisce esattamente in tale contesto, con l'obiettivo di **valutare il potenziale del foundation model Aurora per il supporto alle attività operative di Arpa FVG**.

## 1.2 La Rivoluzione dei Foundation Models in Ambito Scientifico

Parallelamente all'evoluzione dei metodi numerici tradizionali, negli ultimi anni è emerso un nuovo paradigma nel campo dell'intelligenza artificiale, che promette di superare alcune delle sfide storiche della modellistica complessa: quello dei **foundation model**, o modelli fondazionali. [8]

Un foundation model è una rete neurale su larga scala la cui peculiarità risiede in un processo di addestramento articolato in due fasi distinte: *pre-training* e *fine-tuning*. Durante il pre-training, il modello viene esposto a una vastissima e diversificata quantità di dati senza essere ottimizzato per un compito specifico. L'obiettivo di questa fase, che è la più onerosa dal punto di vista computazionale, è quello di apprendere rappresentazioni generali e di estrarre informazioni fondamentali e riutilizzabili dal dominio di interesse. Successivamente, nella fase di *fine-tuning*, queste rappresentazioni generali vengono sfruttate per adattare, o meglio specializzare, il modello a compiti specifici con uno sforzo computazionale notevolmente inferiore e senza modifiche strutturali. Questa filosofia, basata sul concetto di *transfer learning*, segna un

cambiamento radicale rispetto all'approccio classico, che prevedeva lo sviluppo di un modello ad-hoc per ogni singolo problema. [9]

Questo approccio ha ottenuto un successo straordinario nel campo del trattamento del linguaggio naturale (NLP), dove modelli come GPT-3 di OpenAI hanno ridefinito lo stato dell'arte.[10] Tuttavia, la vera rivoluzione è stata l'applicazione di questo paradigma a domini scientifici complessi. Esempi significativi includono **AlphaFold** di DeepMind, che ha rivoluzionato la predizione della struttura delle proteine. [11] Un altro esempio importante è **Stable Diffusion**, che ha segnato un punto di svolta nella generazione di immagini ad alta risoluzione partendo da descrizioni testuali. [12]

Il potenziale di questa rivoluzione tecnologica per la previsione ambientale è immenso. La crescente disponibilità di dati, infatti, introduce sfide critiche: la loro vastità ne rende difficile un'assimilazione concreta e la gestione dell'incertezza statistica diventa un ostacolo che i modelli deterministici faticano a superare, portando a errori significativi. In questo contesto, la capacità dei modelli di intelligenza artificiale di apprendere da dati eterogenei, di generalizzare a nuovi compiti con un addestramento mirato e di migliorare l'accuratezza nella previsione, rappresenta una delle motivazioni principali per la valutazione proposta in questa tesi.

## 1.3 Stato dell'Arte: il Modello Aurora

Il modello al centro di questa tesi, **Aurora**, rappresenta lo stato dell'arte nell'applicazione dei *foundation model* alla previsione del sistema Terra.

Sviluppato da un team di ricerca di *Microsoft Research AI4Science* e presentato in un articolo pubblicato online il 21 maggio 2025, Aurora si propone come un unico modello generalista, capace di affrontare un'ampia gamma di compiti previsionali, dalla meteorologia alla chimica atmosferica, superando (nei casi documentati) le prestazioni dei sistemi operativi specializzati, ma con un costo computazionale drasticamente inferiore. [9]

### 1.3.1 Aurora: un Foundation Model per l'Atmosfera

Aurora è un *foundation model* su larga scala progettato specificamente per la modellazione atmosferica. La sua architettura è quella di un *Vision Transformer* (ViT) con una struttura di tipo *Encoder-Decoder*. Questo paradigma architetturale è stato scelto per la sua flessibilità nel processare dati eterogenei.[13]

- L'**Encoder** ha il compito di trasformare input di diversa natura (variabili, risoluzioni, fonti) in una rappresentazione latente standardizzata, ovvero una griglia tridimensionale unificata.

- Il **Backbone**, un trasformatore basato sull'architettura *3D Swin Transformer*, agisce come un integratore temporale, facendo evolvere lo stato latente nel tempo.[14]
- Il **Decoder**, infine, riconverte la rappresentazione latente evoluta nelle variabili fisiche di interesse, proiettandole nuovamente sulla griglia di output desiderata.

Questa architettura permette al modello di generare previsioni in modo autoregressivo: l'output a un determinato istante temporale viene riutilizzato come input per predire lo stato successivo, estendendo l'orizzonte previsionale fino a 10 giorni e oltre. [9]

### 1.3.2 Dati e Fasi di Addestramento

Seguendo la filosofia dei *foundation model*, il processo di addestramento di Aurora si articola in due fasi distinte e sequenziali, progettate per massimizzare la generalizzazione e l'efficienza.

**Pre-training** La prima fase, il *pre-training*, è la più intensiva dal punto di vista computazionale e mira a far apprendere al modello una comprensione generale delle dinamiche che governano il sistema Terra. Per raggiungere questo obiettivo Aurora è stato esposto a un corpus di dati immenso ed eterogeneo, quantificato in oltre un milione di ore di dati geofisici, per un volume totale di circa 1,2 petabyte. Questo dataset include una combinazione di:

- **Dati di rianalisi**, come ERA5, che fondono osservazioni reali e modelli fisici per creare una ricostruzione storicamente coerente del clima. [15]
- **Previsioni operative** da modelli numerici allo stato dell'arte, come IFS (ECMWF) e GFS (NOAA). [16],[17]
- **Simulazioni climatiche** da progetti come CMIP6. [18]

L'obiettivo di questa fase non è ottimizzare il modello per un compito specifico, ma esporlo a un'ampia variabilità di condizioni iniziali, bias modellistici e dinamiche caotiche, affinché impari a mediare questi effetti e a estrarre rappresentazioni fisicamente significative e generalizzabili. [9]

**Fine-tuning** Una volta completato il pre-training, il modello generalista viene specializzato per compiti specifici attraverso una seconda fase, molto più rapida ed efficiente, chiamata *fine-tuning*. In questa fase, il modello viene adattato a un dataset specifico (ad esempio, previsioni di inquinanti o meteorologia ad alta risoluzione) aggiornando solo un sottoinsieme dei suoi parametri. Per rendere questo processo ancora più efficiente, Aurora sfrutta la tecnica della **Low-Rank Adaptation (LoRA)**, che consiste nel "congelare" i pesi pre-addestrati e nell'addestrare solo

delle piccole matrici di correzione a basso rango.[19] Questo approccio riduce drasticamente il numero di parametri da aggiornare, permettendo di specializzare un modello da oltre un miliardo di parametri con risorse computazionali relativamente contenute, stimate in circa 5 giorni su 8 GPU. [9]

### 1.3.3 Prestazioni e Benchmark Pubblicati

Le capacità di Aurora sono state validate attraverso un confronto rigoroso con i principali sistemi operativi e modelli di intelligenza artificiale esistenti, dimostrando prestazioni allo stato dell'arte in diversi domini critici.

- **Chimica atmosferica e qualità dell'aria:** nel compito di prevedere la composizione atmosferica globale a 5 giorni, Aurora si è dimostrato competitivo con il sistema di riferimento CAMS (*Copernicus Atmosphere Monitoring Service*), eguagliandolo o superandolo nel 74% delle variabili target analizzate (inclusendo diversi inquinanti a varie quote e orizzonti temporali). [9]
- **Previsione meteorologica ad alta risoluzione:** Aurora è stato il primo modello AI a operare con successo a una risoluzione di  $0,1^\circ$  (circa 9 km), superando le prestazioni del modello numerico di riferimento, l'IFS-HRES dell'ECMWF, in oltre il 92% dei target a orizzonti temporali superiori alle 12 ore. [9]
- **Confronto con altri modelli AI:** in un confronto diretto a parità di risoluzione ( $0,25^\circ$ ) con GraphCast di Google DeepMind, considerato uno dei modelli AI più avanzati, Aurora ha mostrato una superiorità sistematica, superandolo su circa il 94% dei target valutati. [9]
- **Previsione di eventi estremi:** la robustezza di un modello si misura anche sulla sua capacità di prevedere gli eventi più rari e intensi. In un caso di studio sulla **Tempesta Ciarán**, che ha colpito l'Europa nel 2023, Aurora è stato l'unico modello di intelligenza artificiale, tra quelli analizzati in uno studio pubblicato su *Nature*, in grado di prevedere correttamente il picco di intensità massima del vento, avvicinandosi molto alla "verità" simulata dal modello IFS, a differenza di altri modelli che tendevano a sottostimare sistematicamente l'evento. [20] [9]

Questi risultati, ottenuti con tempi di inferenza che sono di alcuni ordini di grandezza inferiori a quelli dei modelli numerici tradizionali (minuti su una singola GPU vs ore su supercomputer), evidenziano il potenziale trasformativo di questo approccio per il futuro della previsione ambientale. [9]



## 2. Scopo della Tesi

Il capitolo precedente ha delineato il contesto operativo di ARPA FVG, evidenziando la centralità del modello deterministico WRF e le sfide computazionali ad esso associate. Successivamente, è stata presentata l'architettura del *foundation model* Aurora, illustrandone il potenziale innovativo basato sull'apprendimento da vasti ed eterogenei dataset geofisici. Questo elaborato si colloca all'intersezione di queste due realtà, investigando la possibilità di integrare una tecnologia di frontiera basata su Intelligenza Artificiale all'interno di un flusso di lavoro meteorologico consolidato e operativo.

### 2.1 Obiettivi e Domande di Ricerca

L'obiettivo primario di questa tesi è **valutare la fattibilità tecnica e l'utilità operativa del modello Aurora per supportare le attività di previsione di ARPA FVG**. L'analisi non si limita a un semplice confronto di performance, ma si propone di esplorare in profondità le sfide pratiche e le implicazioni metodologiche che emergono dall'accoppiamento di due sistemi concettualmente molto diversi: un modello numerico basato sulla fisica (WRF) e un modello *data-driven* (Aurora).[21]

Per guidare questa indagine, sono state formulate le seguenti domande di ricerca chiave:

- **Integrazione e Compatibilità:** è tecnicamente possibile integrare Aurora nel flusso di lavoro previsionale di ARPA FVG, utilizzando l'output del modello WRF come condizione iniziale? Quali passaggi di pre-processing sono necessari per rendere compatibili i dati?
- **Sfide Tecniche e Infrastrutturali:** quali sono le principali sfide legate all'adattamento dei dati, come la gestione di diverse proiezioni cartografiche, risoluzioni spaziali, valori mancanti e NaN? L'infrastruttura di calcolo locale (cluster HPC Fenice) è adeguata per l'inferenza del modello, e quali sono i suoi limiti rispetto ad ambienti cloud dotati di GPU?
- **Limiti e Opportunità Operative:** quali sono i limiti intrinseci di un modello pre-addestrato su dati globali quando applicato a un dominio locale ad alta risoluzione? Il

fenomeno del "*dataset mismatch*", ovvero l'incompatibilità statistica tra i dati di WRF e il corpus di addestramento di Aurora, rappresenta un ostacolo insormontabile?[22] Al di là di un utilizzo diretto, quali sono le reali opportunità offerte da Aurora, ad esempio in termini di efficienza computazionale per le previsioni di tipo *ensemble*?

La risposta a queste domande permetterà non solo di valutare Aurora in sé, ma anche di trarre conclusioni più ampie sul ruolo che i *foundation models* possono aspirare a ricoprire nei centri meteorologici regionali.

## 2.2 Struttura dell'Elaborato

Per rispondere in modo sistematico agli obiettivi prefissati, il presente elaborato è strutturato come segue. Il **Capitolo 3**, "Metodologia e Ambiente Sperimentale", descriverà in dettaglio l'approccio tecnico adottato. Verranno presentate le infrastrutture di calcolo utilizzate, dal cloud di Google Colab al cluster HPC "Fenice", e sarà illustrata la pipeline di elaborazione dati sviluppata per adattare gli output di WRF, con un focus sugli strumenti NetCDF e CDO. Infine, verrà esposta la metodologia di valutazione sequenziale basata sulla tecnica del *rollout* autoregressivo. Il **Capitolo 4**, "Risultati", presenterà e analizzerà criticamente gli esiti della sperimentazione. Saranno discussi i benchmark computazionali, il superamento del problema dei valori mancanti che ha portato alla generazione del primo output plausibile ed il successivo fallimento della tecnica del *rollout*, interpretato alla luce dell'ipotesi del *dataset mismatch* e documentato da artefatti visivi. Infine, il **Capitolo 5**, "Conclusioni", sintetizzerà il lavoro svolto, fornendo una risposta puntuale alle domande di ricerca. Verrà discussa la lezione chiave appresa — l'importanza della coerenza statistica dei dati — e saranno esplorate le opportunità strategiche che un modello come Aurora può comunque offrire ad ARPA FVG, concludendo con proposte per sviluppi futuri e una riflessione sulle competenze acquisite.



## 3. Metodologia e Ambiente Sperimentale

Questo capitolo descrive l’approccio metodologico e l’ambiente sperimentale utilizzati per valutare il modello Aurora. Verranno dapprima presentate le infrastrutture di calcolo, dal cloud di Google Colab al cluster HPC locale. Successivamente, sarà illustrata in dettaglio la pipeline di elaborazione dati sviluppata per adattare gli output del modello WRF ai requisiti di Aurora. Infine, verrà esposta la metodologia di valutazione sequenziale adottata per testare le capacità previsionali del modello.

### 3.1 Infrastrutture a Confronto: Google Colab vs. HPC Locale

Per la valutazione del modello Aurora sono stati utilizzati due ambienti di calcolo distinti e complementari. La fase preliminare di studio e benchmark è stata condotta in un ambiente cloud flessibile, per poi spostare le elaborazioni più onerose sull’infrastruttura di calcolo ad alte prestazioni (HPC) di ARPA FVG.

#### 3.1.1 Test Preliminari in Ambiente Cloud: Google Colab

I primi test pratici sono stati svolti utilizzando la piattaforma Google Colab, scelta per la sua immediata disponibilità e per l’accesso a risorse hardware dedicate all’intelligenza artificiale. [23] L’obiettivo di questa fase era verificare la fattibilità tecnica dell’inferenza, familiarizzare con la libreria software di Aurora e ottenere benchmark prestazionali preliminari.

L’ambiente utilizzato era una sessione Colab Pro in prova gratuita, che forniva accesso a una GPU NVIDIA T4. I test hanno confermato che l’ambiente è idoneo per eseguire inferenze con le versioni più piccole del modello e su griglie a bassa risoluzione, evidenziando un notevole vantaggio prestazionale rispetto alla CPU. Tuttavia, sono emersi anche limiti infrastrutturali critici: durante i tentativi di eseguire il modello pre-addestrato sulla griglia a risoluzione nativa di 0.25 circa (dimensione di  $721 \times 1440$  punti) con dati di input generati casualmente, la memoria della GPU (14.74 GiB) si è rivelata insufficiente, generando un errore di tipo `OutOfMemoryError`.

Questa limitazione ha reso necessario il passaggio a un'infrastruttura più potente per le analisi su domini e risoluzioni significative. Per una trattazione dettagliata dei risultati si rimanda al Capitolo Risultati.

### 3.1.2 L'Infrastruttura Principale: il Cluster HPC "Fenice"

L'infrastruttura principale per la sperimentazione è stata il cluster di calcolo ad alte prestazioni (HPC) "Fenice" della Regione Friuli Venezia Giulia e dell'ARPA FVG.[24] L'accesso avviene tramite il client MobaXterm, che stabilisce una connessione SSH a uno dei nodi di accesso ridondanti. Questi nodi, dotati di circa 20 processori, fungono da punto di ingresso per l'autenticazione basata su chiavi pubblica/privata, la gestione dei file e la sottomissione dei job di calcolo al sistema a code.

Le componenti chiave del cluster utilizzate in questo lavoro sono:

- **Nodi di Calcolo e Code:** il calcolo vero e proprio viene eseguito su nodi specializzati, interconnessi da una rete ad alta velocità InfiniBand e organizzati in diverse code. Per le attività di sviluppo e test interattivi è stata utilizzata principalmente la coda arpa, che dispone di 2 server da 96 processori ciascuno. Altre code rilevanti includono la coda hp (4 server da 32 processori) per job occasionali e la coda julia (8 server da 40 processori) per calcoli automatici a lunga esecuzione. L'infrastruttura di calcolo è cresciuta nel tempo; nel 2027 è previsto il prossimo intervento di sostituzione dei server diventati obsoleti.[24]
- **Sistema di Gestione dei Job:** la gestione delle risorse e la schedulazione dei job sono affidate al gestore Portable Batch System (PBS) Professional. Gli script di calcolo vengono sottomessi tramite il comando qsub e includono direttive #PBS per specificare le risorse richieste (coda, numero di processori, memoria, tempo massimo di esecuzione o *walltime*). Lo stato dei job (es. R per in esecuzione, Q per in attesa) viene monitorato con qstat ed è possibile rimuovere un job con qdel. Per le catene operative complesse, come le previsioni meteorologiche, PBS è orchestrato da ecFlow, un gestore di workflow più avanzato di un semplice crontab.[25][7]
- **File System Dedicati:** l'infrastruttura è dotata di file system distribuiti con scopi specifici.
  - **/u:** ospita le *home directory* degli utenti (es. /u/arpa/stefanotog) con una quota di circa 10 GB. Viene utilizzata per file personali, script e software leggero ed è soggetta a backup regolare.[24]
  - **/lustre:** questo spazio è un file system parallelo ad alte prestazioni (es. /lustre/arpa/stefanotog) con una quota di circa 2 TB. Funge da spazio di lavoro.

ro temporaneo (*scratch*) per i dati pesanti, i modelli e i risultati dei calcoli. Questo spazio non è soggetto ad alcun backup.[24]

- **/atlas e /archive:** per l’archiviazione a lungo termine sono disponibili il file system rapido /atlas (240 TB) per dati importanti e accessibili, e /archive per dati storici a cui si accede raramente.[24]
- **Ambiente Software Modulare:** Per evitare conflitti tra le dipendenze software dei vari utenti, il cluster utilizza un sistema di moduli. L’ambiente di lavoro viene configurato dinamicamente caricando le versioni software necessarie (es. `cdo`, `ncview`) tramite il comando `module load <nome_modulo>`.[24]

## 3.2 Pipeline di Elaborazione Adottata: Adattare i Dati WRF per Aurora

Per poter funzionare, Aurora necessita innanzitutto di una griglia di partenza, ossia una struttura regolare con rigidi vincoli sui dati di ingresso. In azienda la soluzione più immediata è stata quella di utilizzare la griglia di output del modello WRF, perché rappresentava la scelta più vicina a ciò che Aurora richiedeva in termini di dati iniziali.[9][4] Questa decisione ha tuttavia evidenziato una delle principali sfide tecniche del progetto: i due modelli operano su griglie con caratteristiche profondamente diverse. WRF utilizza una proiezione conica conforme di Lambert, mentre Aurora richiede una griglia geografica regolare basata su coordinate di latitudine e longitudine.[26] È stato quindi necessario sviluppare una pipeline di pre-processing capace di adattare i dati prodotti da WRF ai requisiti specifici di Aurora.

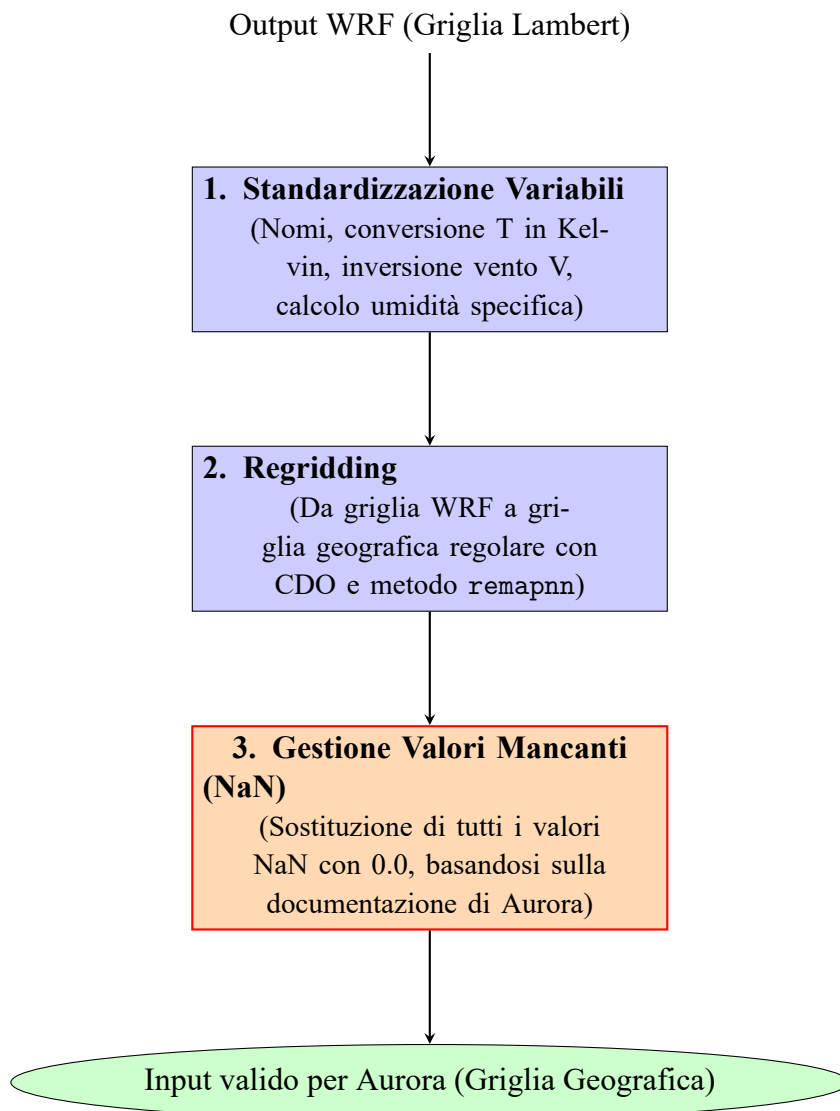
Gli strumenti software principali utilizzati per la gestione e la manipolazione dei dati sono stati:

- **NetCDF:** tutti i dati di input e output sono stati gestiti nel formato NetCDF (`.nc`), uno standard per dati scientifici auto-descrittivi che incapsula in un unico file sia i dati grezzi che i metadati (dimensioni, variabili, unità di misura).[27]
- **CDO (Climate Data Operators):** per le operazioni di manipolazione delle griglie (re-gridding) è stata utilizzata la suite di comandi da terminale CDO, uno strumento standard e altamente performante nel settore della modellistica climatica e meteorologica.[28]

Il flusso di pre-processing è stato articolato nelle seguenti fasi sequenziali:

1. **Standardizzazione delle variabili:** la prima fase è consistita nell'uniformare i dati grezzi di WRF secondo le convenzioni richieste da Aurora.[29] Questo ha incluso la ridenominazione delle variabili, la conversione della temperatura da Celsius a Kelvin, l'inversione di segno della componente meridionale del vento (V) e il calcolo di variabili derivate come l'umidità specifica.
2. **Regridding:** l'operazione successiva è stata la mappatura dei dati dalla griglia WRF a una griglia geografica regolare compatibile con Aurora. Un primo tentativo con interpolazione bilineare (`remapbil`) si è rivelato problematico, in quanto introduceva un numero eccessivo di valori mancanti (sostituiti da NaN) ai bordi del dominio. Si è quindi optato per un'interpolazione basata sul vicino più prossimo (*Nearest Neighbor*, `remapnn`). Sebbene questo metodo sia statisticamente più semplice, è stato scelto perché evita la generazione di artefatti da interpolazione, garantendo che ogni punto della nuova griglia ricevesse un valore fisicamente presente nei dati originali.[30]
3. **Gestione dei valori mancanti:** il problema più critico riscontrato è stata la gestione dei valori mancanti, una sfida comune e ben documentata nell'analisi di dati climatici incompleti [31]. I dati WRF contengono *missing value* fisicamente corretti, ad esempio a quote elevate dove i livelli di pressione standard si trovano al di sotto dell'orografia del terreno. È noto che la presenza di valori mancanti può degradare significativamente le prestazioni dei modelli di machine learning [32], producendo NaN che, se non gestiti, corrompono l'intero processo di inferenza di Aurora, producendo output inutilizzabili. La svolta è avvenuta scoprendo, tramite l'analisi della documentazione scientifica di Aurora, che il modello pre-addestrato è in grado di gestire intrinsecamente i dati mancanti a patto che questi vengano sostituiti con il valore zero. L'implementazione di questa semplice tecnica di imputazione, ovvero la conversione di tutti i NaN in 0.0, si è rivelata decisiva e ha permesso di ottenere il primo output fisicamente plausibile.

Figura 3.1: Diagramma di flusso della pipeline di pre-processing per adattare i dati WRF ad Aurora.



### 3.3 Metodologia di Valutazione Sequenziale

Una volta ottenuti input validi, si è proceduto a testare le performance previsionali del modello attraverso una metodologia incrementale.

#### 3.3.1 La Tecnica del "Rollout" per Previsioni a Lungo Termine

Per generare previsioni su un orizzonte temporale esteso (es. 5 giorni), è stata implementata la tecnica del "rollout" o previsione autoregressiva. Aurora è addestrato a prendere in input due stati atmosferici a 6 ore di distanza ( $t_0$  e  $t_{6h}$ ) per predire lo stato successivo ( $t_{12h}$ ). La tecnica del rollout consiste nell'utilizzare l'output generato dal modello come input per il passo successivo. Ad esempio, la previsione a  $t_{12h}$  viene combinata con lo stato a  $t_{6h}$  per generare la previsione a  $t_{18h}$ , e così via, estendendo iterativamente l'orizzonte previsionale.

#### 3.3.2 Criteri di Analisi dei Risultati

Data la natura esplorativa della sperimentazione, l'analisi dei risultati si è basata su i seguenti criteri qualitativi:

- **Valutazione qualitativa visuale:** il primo e più immediato criterio di valutazione è stato l'esame visivo degli output tramite l'interfaccia grafica di `ncview`. [33] Questo ha permesso di confrontare la coerenza spaziale e fisica dei campi previsti da Aurora con quelli di riferimento del modello WRF, verificando la plausibilità delle strutture meteorologiche generate.
- **Analisi dei gradienti per l'identificazione di artefatti:** per investigare la presenza di artefatti spaziali, è stata condotta un'analisi dei gradienti orizzontali dei campi previsti. Calcolando la somma dei gradienti lungo le righe e le colonne della griglia, è stato possibile identificare la presenza di discontinuità regolari, o "patch effect", un noto artefatto dei modelli basati su architettura Transformer che può indicare problemi nell'elaborazione spaziale dei dati. [34]

## 4. Risultati

Questo capitolo presenta i risultati ottenuti durante la fase sperimentale del tirocinio. L'obiettivo primario era valutare la fattibilità di implementare il modello Aurora sull'infrastruttura HPC aziendale, dotata esclusivamente di CPU. Vengono descritti sia i successi che i fallimenti incontrati, interpretandoli alla luce degli obiettivi prefissati. L'analisi parte dai benchmark prestazionali che definiscono i limiti infrastrutturali, per poi descrivere il superamento delle difficoltà tecniche che ha portato alla generazione del primo output fisicamente plausibile. Infine, viene analizzato il fallimento della tecnica di previsione a lungo termine (*rollout*), formulando un'ipotesi basata sul concetto di *dataset mismatch*. [22]

### 4.1 Benchmark di Prestazione e Limiti Infrastrutturali

La valutazione delle performance computazionali è stata una fase preliminare fondamentale per determinare la fattibilità del progetto. L'obiettivo era comprendere i requisiti hardware del modello Aurora per poi valutare la sua applicabilità sull'infrastruttura HPC aziendale, dotata unicamente di CPU. A tale scopo, i primi test sono stati condotti in un ambiente cloud (Google Colab) per stabilire un benchmark di riferimento, confrontando le prestazioni di CPU e GPU.

#### 4.1.1 Confronto delle Prestazioni di Inferenza: CPU vs. GPU

Per quantificare il vantaggio computazionale offerto dall'hardware specializzato, sono stati eseguiti dei benchmark comparativi in cloud tra CPU e GPU (NVIDIA T4) utilizzando il modello AuroraPretrained. I test, condotti su griglie di input di dimensioni crescenti, hanno permesso di definire un riferimento prestazionale. I risultati, riassunti nella Tabella 4.1, evidenziano un'accelerazione notevole grazie all'uso della GPU. Questa differenza prestazionale è dovuta alla loro architettura: le CPU sono ottimizzate per eseguire compiti in modo sequenziale, mentre le GPU sono progettate specificamente per il calcolo parallelo. Grazie a migliaia di core che operano simultaneamente, le GPU possono processare in parallelo l'enorme quantità di calcoli matriciali richiesti dai modelli di deep learning, accelerando drasticamente le operazioni. [35]

<b>Risoluzione</b>	<b>Tempo CPU (s)</b>	<b>Tempo GPU (s)</b>	<b>Speed up GPU</b>
$17 \times 32$	0.1732	0.0325	$\sim 5.3\times$
$128 \times 256$	6.5700	0.1438	$\sim 45.7\times$
$256 \times 512$	27.1666	0.5109	$\sim 53.2\times$
$512 \times 1024$	110.9088	1.9776	$\sim 56.1\times$

Tabella 4.1: Tempi medi di inferenza per una singola previsione a 6 ore del modello Aurora, confrontando l'esecuzione su CPU e GPU a diverse risoluzioni di input. Lo speed up evidenzia il guadagno prestazionale ottenuto con la GPU.

Come si evince dalla tabella, il vantaggio della GPU cresce in modo significativo con l'aumentare della risoluzione. Questi risultati confermano che, sebbene l'inferenza su CPU sia tecnicamente fattibile, l'uso di una GPU è determinante per lavorare a risoluzioni elevate in tempi operativamente accettabili.[35]

#### 4.1.2 Limiti di Memoria e Validazione dell'Infrastruttura Locale

Oltre al confronto sulla velocità di esecuzione, i test preliminari in ambiente cloud sono stati utili per definire i requisiti di memoria del modello in condizioni estreme e, di conseguenza, per contestualizzare le capacità dell'infrastruttura aziendale. È stato condotto uno stress test spingendo deliberatamente le dimensioni della griglia di input fino al limite massimo sostenibile dall'hardware disponibile.

Durante il tentativo di eseguire l'inferenza del modello AuroraPretrained sulla sua griglia nativa di  $0,25^\circ$  ( $721 \times 1440$  punti), la GPU NVIDIA T4 di Google Colab, con una capacità di 14.74 GiB, si è rivelata insufficiente, generando un errore di tipo `OutOfMemoryError`. Questo esperimento ha permesso di stabilire un limite tecnico preciso: per operare alla massima risoluzione per cui Aurora è stato progettato, è necessaria un'infrastruttura dotata di acceleratori con memoria video superiore ai 16 GB.

Tale scoperta, tuttavia, non ha rappresentato un ostacolo per il progetto, il cui obiettivo primario era proprio la valutazione della fattibilità sull'infrastruttura HPC aziendale, basata esclusivamente su CPU. Le analisi richieste per il tirocinio, infatti, necessitavano di griglie di dimensioni notevolmente inferiori rispetto a quella nativa del modello. In questo contesto operativo, il cluster HPC locale si è dimostrato non solo adeguato, ma pienamente performante. Ad esempio, l'inferenza su una griglia di  $256 \times 256$  punti è stata completata in circa 9 secondi su una singola CPU, un tempo considerato del tutto accettabile per le finalità del progetto.

In conclusione, i test hanno fornito un doppio risultato: da un lato, hanno chiarito che l'uso di soluzioni cloud a basso costo è precluso per analisi a piena risoluzione; dall'altro, hanno validato l'infrastruttura HPC aziendale come una soluzione efficace e affidabile per gli scopi prefissati,



dimostrando che le prestazioni della CPU erano ampiamente sufficienti per le esigenze operative del tirocinio.

## 4.2 Dalla Teoria alla Pratica: Generazione del Primo Output

Superata la fase di benchmark, la sfida principale è diventata quella di fornire al modello dati di input validi, partendo dagli output del modello WRF. Questo processo si è scontrato con un ostacolo critico: la gestione dei valori mancanti (*Not a Number*, NaN).

La svolta, definita negli appunti di lavoro come "la vittoria dell'11 Agosto", è arrivata dopo un'attenta analisi del problema. I dati WRF, soprattutto dopo il processo di regridding, contenevano due tipologie di valori NaN:

- **NaN da interpolazione**, artefatti introdotti dal processo di mappatura dalla griglia nativa di WRF a quella geografica, localizzati principalmente ai bordi del dominio.
- **NaN fisicamente corretti**, ovvero valori assenti perché privi di significato fisico. Un esempio significativo riguarda i livelli di pressione inferiori (es. 1000 hPa o 925 hPa) in corrispondenza di aree montuose, dove il suolo si trova a una quota superiore a quella del livello di pressione stesso.

La presenza massiccia di questi valori (fino al 76% dei dati a 1000 hPa) corrompeva l'intero processo di inferenza, poiché qualsiasi operazione matematica con un NaN produce un altro NaN, portando a un output completamente inutilizzabile.

La soluzione è stata individuata analizzando la documentazione scientifica di Aurora [9]. Sebbene il paper descriva una tecnica esplicita per la gestione dei dati mancanti tramite "density layers" (utilizzata per le previsioni marine), si è scoperto empiricamente che il modello pre-addestrato possiede una capacità intrinseca di gestire i valori mancanti a patto che questi vengano sostituiti con il valore zero. L'ipotesi è che il modello, essendo stato addestrato su un dataset globale, abbia imparato a riconoscere e ignorare valori nulli in contesti fisicamente incoerenti (es. zero umidità a bassa quota), trattandoli di fatto come dati assenti.

L'implementazione di questa semplice tecnica di imputazione – la conversione di tutti i valori NaN in 0.0 – si è rivelata decisiva. Eseguendo nuovamente l'inferenza, il modello ha prodotto per la prima volta un output fisicamente plausibile e visivamente coerente, come mostrato in Figura 4.1. L'analisi qualitativa, condotta tramite lo strumento ncview, ha confermato che la distribuzione spaziale delle variabili meteorologiche (es. temperatura, pressione) rispecchiava, a bassa risoluzione, quella del modello WRF di riferimento, segnando il primo successo significativo del progetto.

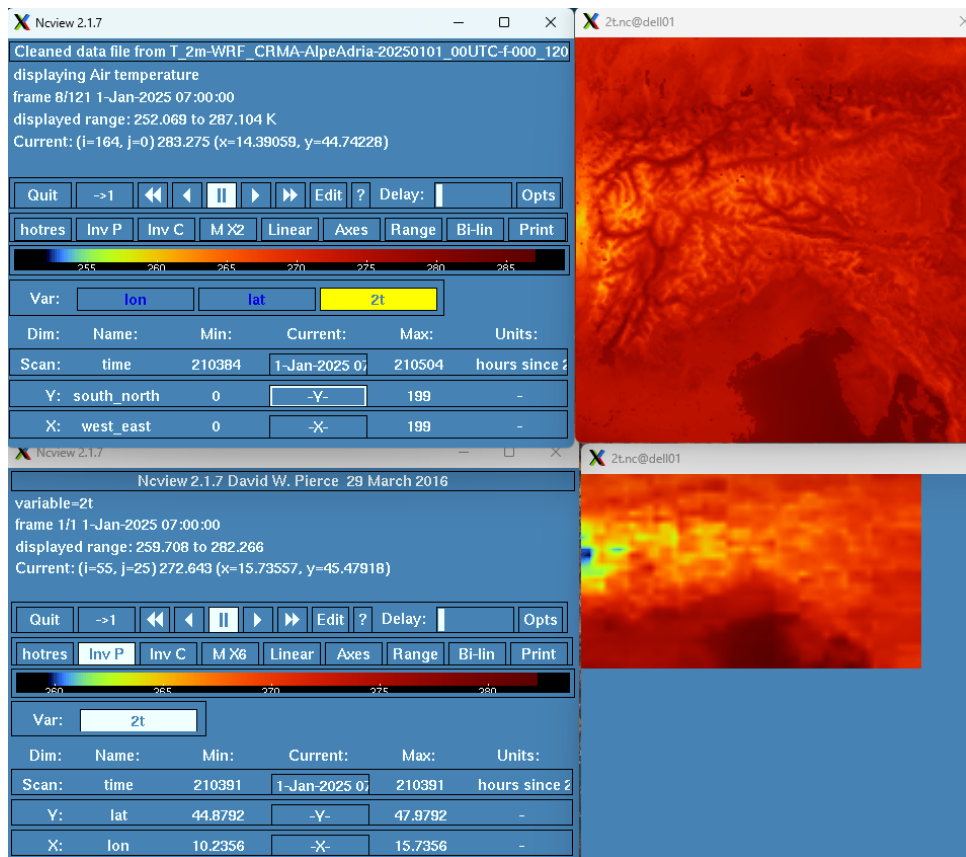


Figura 4.1: Confronto visivo qualitativo tra il campo di temperatura a 2 metri (2t) del modello WRF di riferimento (sopra) e il primo output fisicamente plausibile generato da Aurora (sotto) dopo la corretta gestione dei valori NaN. Sebbene a una risoluzione inferiore e con un bias di temperatura, la struttura geografica e la distribuzione meteorologica generale risultano coerenti.

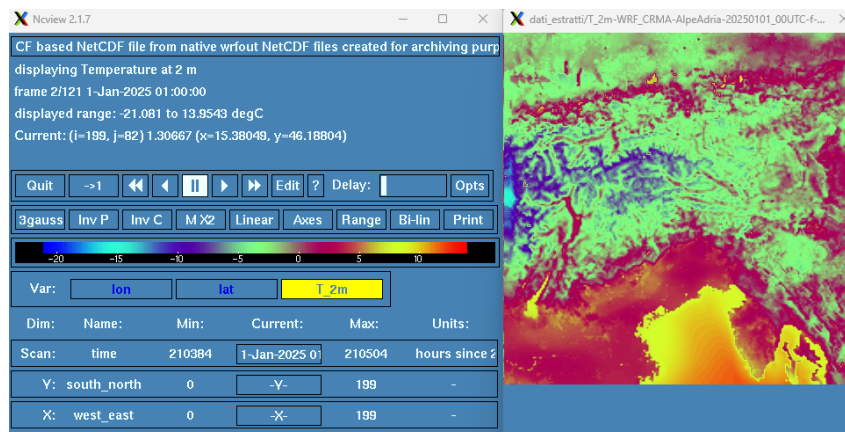


Figura 4.2: Visualizzazione dei dati di input preparati per Aurora, dopo il regridding e la sostituzione dei valori NaN con zero. Questi dati rappresentano lo stato iniziale da cui il modello ha generato la sua prima previsione.

### 4.3 Il Fallimento del Rollout e l’Ipotesi del Dataset Mismatch

Ottenuto un singolo output corretto, il passo successivo è stato implementare la tecnica del *rollout* per generare previsioni a lungo termine. Come descritto nel paper di riferimento [9], Aurora opera in modo autoregressivo: prende in input due stati atmosferici a 6 ore di distanza ( $t_0$  e  $t_{6h}$ ) per predire lo stato successivo ( $t_{12h}$ ). La previsione generata viene poi riutilizzata come input per il passo successivo, estendendo iterativamente l’orizzonte previsionale.

Tuttavia, l’applicazione di questa tecnica ha portato al principale fallimento sperimentale della tesi. Sebbene l’intervallo temporale di 6 ore fosse rispettato, la qualità delle previsioni degradava molto rapidamente: dopo appena 2 o 3 passi iterativi, l’output diventava fisicamente irrealistico e non più utilizzabile.

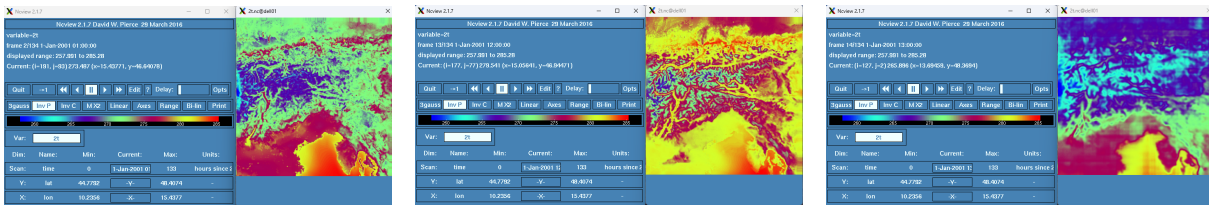
L’analisi critica di questo fallimento ha portato a formulare un’ipotesi basata su un fondamentale **dataset mismatch**. Il modello Aurora è stato pre-addestrato su un corpus di dati immenso che include principalmente dati di rianalisi (ERA5) e previsioni operative da modelli globali come IFS dell’ECMWF [9]. I dati forniti in questo esperimento, invece, provenivano dal modello ad area limitata WRF. Sebbene rappresentino le stesse variabili fisiche, i due dataset possiedono evidentemente una distribuzione statistica e caratteristiche intrinseche differenti, dovute a:

- **Diverse parametrizzazioni fisiche** utilizzate dai modelli.
- **Diverse condizioni al contorno** (GFS per WRF vs. IFS per il training di Aurora).
- **Artefatti introdotti dal regridding** da un dominio ad area limitata.

Il modello Aurora, non avendo mai “visto” dati con le caratteristiche statistiche di WRF durante il suo addestramento, non è stato in grado di generalizzare. L’errore introdotto al primo passo, per quanto piccolo, veniva amplificato a ogni iterazione successiva del rollout, portando a un rapido collasso della previsione.

La prova più evidente di questo fallimento è emersa dall’analisi visiva degli output. Variabili come l’umidità e il vento hanno manifestato la comparsa di vistosi artefatti a blocchi, come illustrato nella sequenza di Figura 4.3. Queste strutture, cromaticamente uniformi e prive di qualsiasi gradiente fisico realistico, sono il sintomo di un’incapacità del modello di interpretare correttamente l’input e di propagarlo nel tempo in modo coerente.

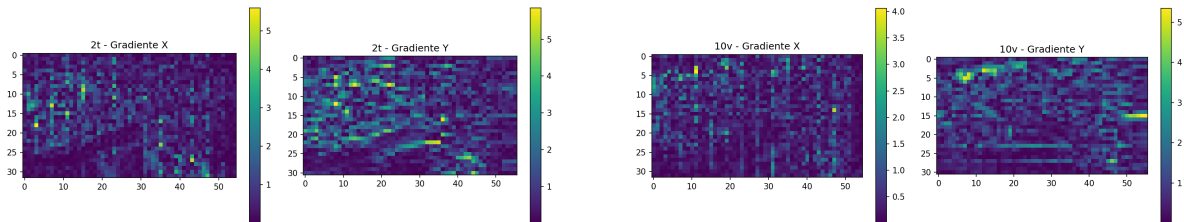
Questo risultato, seppur negativo dal punto di vista operativo, rappresenta una conclusione scientifica fondamentale di questo lavoro di tesi: l’accoppiamento diretto di un *foundation model* pre-addestrato con un modello numerico locale, senza una fase di *fine-tuning* specifica, non è tecnicamente fattibile a causa dell’incompatibilità statistica tra i rispettivi domini di dati. A conferma di un problema strutturale nell’elaborazione spaziale, è stata condotta un’analisi del



(a) Primo passo del rollout ( $t + 6h$ ). (b) Dodicesimo passo del rollout ( $t + 72h$ ). (c) Tredicesimo passo del rollout ( $t + 78h$ ).

Figura 4.3: Degrado progressivo della previsione della temperatura a 2 metri (2t) durante il rollout. Mentre il primo passo (a) è ancora fisicamente plausibile, dopo 12 passi (b) iniziano a comparire artefatti evidenti. Al tredicesimo passo (c), la previsione è completamente collassata, mostrando i caratteristici "quadrati" e perdendo ogni coerenza fisica.

gradiente su tutto il campo previsto, come descritto nella metodologia. Come visibile in Figura 4.4, la mappa della magnitudo del gradiente non presenta un andamento omogeneo, ma denota una sorta di "reazione" nei cambi di intensità che si manifesta come una griglia di discontinuità sull'intera immagine. Questo "patch effect" è un'ulteriore prova che il modello non riesce a garantire la continuità spaziale durante il processo di inferenza autoregressiva. Ciò accade perché l'architettura di tipo Vision Transformer, propria di Aurora, non elabora l'immagine nella sua interezza, ma la scompone in una griglia di blocchi ("patch") più piccoli; [9] le discontinuità visibili emergono proprio lungo i confini di questi blocchi quando il modello fallisce nel raccordarli in modo fisicamente coerente.



(a) Analisi dei gradienti per la temperatura a 2m (2t). (b) Analisi dei gradienti per il vento meridionale a 10m (10v).

Figura 4.4: Visualizzazione della magnitudo del gradiente calcolata sugli output degradati del rollout. Le nette divisioni a "blocchi" visibili sull'intera mappa (a, b) indicano la presenza di discontinuità spaziali ("patch effect"), confermando un fallimento del modello nell'elaborazione spaziale coerente di dati non familiari.

## 5. Conclusioni

Il presente lavoro di tesi ha esplorato la possibilità di integrare un *foundation model* di frontiera, Aurora, nel flusso operativo di un centro meteorologico regionale, ARPA FVG, che si affida a modelli numerici deterministici consolidati come WRF. Questo capitolo finale intende ripercorrere il percorso di ricerca, sintetizzare i risultati chiave in risposta alle domande iniziali e discutere le lezioni apprese, le opportunità future e le riflessioni personali maturate durante l'esperienza di tirocinio.

### 5.1 Sintesi del Lavoro Svolto

Il percorso di questa tesi si è articolato in diverse fasi sequenziali, partendo da un'analisi del contesto per arrivare a una sperimentazione pratica. Inizialmente, è stato studiato il modello Aurora, approfondendone l'architettura, le fasi di addestramento e le performance documentate in letteratura. Parallelamente, è stato analizzato l'ambiente operativo di ARPA FVG, con un focus sul modello WRF e sull'infrastruttura di calcolo HPC "Fenice".

La fase successiva si è concentrata sullo sviluppo di una pipeline di pre-processing per superare l'incompatibilità tra i due modelli. Questa attività ha richiesto la standardizzazione delle variabili, la gestione di diverse proiezioni cartografiche e, soprattutto, la risoluzione del problema critico dei valori mancanti (NaN). Una volta ottenuto un input valido, si è proceduto alla valutazione delle capacità previsionali di Aurora, implementando la tecnica del *rollout* autoregressivo per generare previsioni a lungo termine. I risultati, sia positivi che negativi, sono stati infine analizzati criticamente per trarre delle conclusioni operative e scientifiche.

### 5.2 Risultati Chiave e Risposta alle Domande di Ricerca

L'obiettivo primario era valutare la fattibilità e l'utilità di Aurora per le attività di ARPA FVG. Le domande di ricerca che hanno guidato l'indagine trovano qui una risposta basata sull'evidenza sperimentale raccolta.

- **Integrazione e Compatibilità:** si è dimostrato che è tecnicamente possibile integrare Aurora nel flusso di lavoro esistente, utilizzando l'output di WRF come input. Tuttavia, questo richiede una complessa pipeline di pre-processing, che include la ridenominazione delle variabili, la conversione delle unità di misura, il calcolo di variabili derivate e, soprattutto, un'attenta gestione delle griglie e dei valori mancanti.
- **Sfide Tecniche e Infrastrutturali:** la sfida tecnica più rilevante è stata la gestione dei valori NaN, superata sostituendoli con il valore zero, una soluzione empirica derivata dalla documentazione di Aurora che si è rivelata efficace per ottenere un primo output plausibile. Per quanto riguarda l'infrastruttura, i benchmark hanno confermato che l'inferenza su CPU è praticabile per le griglie a risoluzione ridotta utilizzate nel progetto, validando l'adeguatezza del cluster HPC "Fenice" per gli scopi del tirocinio. Tuttavia, è emerso chiaramente che per operare a risoluzioni maggiori sono necessarie GPU con una memoria video superiore ai 16 GB, precludendo l'uso di soluzioni cloud a basso costo.
- **Limiti e Opportunità Operative:** il limite operativo più grande è emerso con il fallimento della tecnica del *rollout*. Questo insuccesso ha portato a identificare nel **dataset mismatch** l'ostacolo principale: l'incompatibilità statistica tra i dati del modello locale WRF e il corpus di addestramento globale di Aurora (basato su IFS/ERA5) impedisce una generalizzazione efficace e porta al rapido collasso della previsione. Di conseguenza, l'accoppiamento diretto WRF-Aurora non risulta fattibile senza una specifica fase di fine-tuning. Ciononostante, la straordinaria efficienza computazionale di Aurora apre a un'opportunità strategica di grande valore: il suo impiego per le previsioni di insieme (*ensemble*), che permetterebbe di esplorare l'incertezza previsionale a un costo oggi insostenibile con i modelli tradizionali. Generare decine o centinaia di "universi" previsionali alternativi, perturbando leggermente le condizioni iniziali, diventerebbe fattibile in tempi rapidi e con risorse contenute. Questo consentirebbe di passare da una singola previsione deterministica a una mappa di probabilità, offrendo ai meteorologi uno strumento molto più potente per valutare il rischio di eventi severi e comunicare l'affidabilità della previsione al pubblico e alla Protezione Civile. L'adozione di un approccio ibrido, dove il modello tradizionale fornisce il dettaglio fisico e l'IA esplora l'incertezza, rappresenta quindi la via più promettente per il futuro della meteorologia operativa.

## 5.3 Discussione: la Lezione Appresa

Al di là dei singoli risultati, l'esito di questa sperimentazione suggerisce una lezione chiave di carattere più generale: nell'accoppiare modelli data-driven con simulatori fisici, **la coerenza**

**statistica dei dati può essere più importante dell'architettura del modello stesso.** Il fallimento del rollout, infatti, non è imputabile a un difetto di Aurora, le cui performance sono comprovate, ma alla sua applicazione al di fuori del dominio di dati per cui è stato addestrato.

È tuttavia importante interpretare i risultati di questa tesi con cautela. Il fallimento osservato è probabilmente il prodotto di una concomitanza di fattori. Se il dataset mismatch rappresenta l'ipotesi più solida, non si può escludere che altre variabili abbiano contribuito. La soluzione adottata per la gestione dei valori NaN, pur avendo permesso di superare un blocco operativo, era una soluzione pragmatica basata su un'interpretazione della documentazione; non vi è certezza assoluta che questa tecnica non abbia introdotto un errore sistematico, poi amplificato dal processo autoregressivo.

Infine, un fattore da considerare è l'inevitabile curva di apprendimento personale. In qualità di studente tirocinante, ho affrontato questo progetto con competenze in via di formazione. L'arco temporale limitato del tirocinio non ha permesso di esplorare ogni possibile soluzione con la profondità che un ricercatore esperto avrebbe potuto dedicare. Le scelte metodologiche sono state il frutto di un processo di apprendimento sul campo, un tentativo di trovare soluzioni pratiche a problemi complessi in un tempo definito.

## 5.4 Proposte per Sviluppi Futuri

Le conclusioni di questo lavoro non rappresentano un punto di arrivo, ma una base solida per future investigazioni. Le direzioni più promettenti includono:

- **A breve termine:** implementare una metodologia di validazione statistica per confrontare non solo i valori medi, ma le intere distribuzioni prodotte da Aurora e WRF. Questo permetterebbe di quantificare in modo più rigoroso il "dataset mismatch" e di capire se Aurora è in grado di catturare almeno la variabilità sub-griglia di WRF, come discusso in una delle ipotesi di lavoro con altri operatori dell'Arpa.
- **A lungo termine:** la strada maestra per superare il mismatch è il **fine-tuning** di Aurora su un archivio storico di dati WRF. Sebbene computazionalmente oneroso, questo processo adatterebbe il modello alle specifiche caratteristiche statistiche dell'output di WRF, rendendo l'accoppiamento diretto potenzialmente di successo. Inoltre, si potrebbe esplorare l'applicazione di Aurora ad altri ambiti operativi di ARPA FVG, come la previsione della qualità dell'aria o delle condizioni marine, dove il modello ha già dimostrato capacità notevoli.

## 5.5 Riflessioni Finali e Competenze Acquisite

Al di là dei risultati tecnici, l'esperienza di tirocinio in ARPA FVG è stata per me un'opportunità di crescita di eccezionale valore, sia a livello professionale che personale. Ho avuto il privilegio di essere accolto in un ambiente di lavoro stimolante e sereno, dove ho potuto imparare non solo da un punto di vista tecnico, ma anche umano. La disponibilità dei colleghi e dei supervisori nel guidarmi, insegnandomi un approccio metodico alla ricerca e al problem solving, è stata fondamentale.

Questo progetto, pur non avendo raggiunto l'obiettivo operativo sperato, mi ha permesso di acquisire e consolidare un'ampia gamma di competenze pratiche:

- La gestione di un'infrastruttura di Calcolo ad Alte Prestazioni (HPC), inclusa l'interazione con il sistema a code PBS e la navigazione in file system distribuiti.
- La manipolazione di dati scientifici complessi in formato NetCDF tramite strumenti standard del settore come CDO e librerie Python.
- L'utilizzo avanzato della shell Linux e lo sviluppo di script per l'automazione di pipeline di elaborazione dati.
- L'applicazione pratica di modelli di Intelligenza Artificiale a un problema reale del dominio scientifico.

Ho imparato ad affrontare un progetto di ricerca partendo da un'idea, anche senza avere una chiara visione del risultato finale, e a navigare tra successi e fallimenti come parti integranti del processo di scoperta. Questa tesi non è solo la documentazione di un esperimento, ma il racconto di un percorso di apprendimento che ha arricchito profondamente la mia formazione come futuro ingegnere informatico.



# Bibliografia

- [1] *Regolamento di Organizzazione dell'Agenzia Regionale per la Protezione dell'Ambiente del Friuli Venezia Giulia*, Deliberazione del Direttore Generale N. 66 del 08/06/2015, 2015.
- [2] E. Smeets e R. Weterings, «Environmental indicators: Typology and overview,» European Environment Agency, Copenhagen, rapp. tecn. 25, 1999.
- [3] *Legge Regionale 18 giugno 2007, n. 16. Norme in materia di tutela dall'inquinamento atmosferico e dall'inquinamento acustico*, Bollettino Ufficiale della Regione n. 26 del 27 giugno 2007, Testo vigente. L'istituzione del Centro Regionale di Modellistica Ambientale (CRMA) è disposta all'Art. 5, comma 5, 2007.
- [4] J. G. Powers et al., «The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions,» *Bulletin of the American Meteorological Society*, vol. 98, n. 8, pp. 1717–1737, 2017. DOI: 10.1175/BAMS-D-15-00308.1
- [5] J. Smagorinsky, «Problems and promises of deterministic extended range forecasting,» *Bulletin of the American Meteorological Society*, vol. 50, n. 5, pp. 286–312, 1969. DOI: 10.1175/1520-0477-50.5.286
- [6] J. Haqq-Misra, E. T. Wolf, T. J. Fauchez e R. K. Kopparapu, «Interpolation and synthesis of sparse samples in exoplanet atmospheric modeling,» *arXiv preprint arXiv:2405.14693*, 2024.
- [7] D. Giaiotti, «Workflow manager ecFlow e utente operativo della FENICE,» ARPA FVG - Centro Regionale di Modellistica Ambientale (CRMA), Articolo tecnico-scientifico Tec-Sci 005/2015, set. 2015. indirizzo: <https://www.arpafvg.it/temi/temi/modellistica-ambientale-crma/pubblicazioni/workflow-manager-ecflow-e-utente-operativo-della-fenice/>
- [8] Y. Choi et al., «Defining Foundation Models for Computational Science: A Call for Clarity and Rigor,» *arXiv preprint arXiv:2505.22904*, 2025. indirizzo: <https://arxiv.org/abs/2505.22904v2>

- [9] C. Bodnar et al., «A foundation model for the Earth system,» *Nature*, vol. 641, pp. 1180–1187, 2025. DOI: 10.1038/s41586-025-09005-y
- [10] T. B. Brown et al., «Language Models are Few-Shot Learners,» *arXiv preprint arXiv:2005.14165*, 2020. arXiv: 2005.14165.
- [11] J. Jumper et al., «Highly accurate protein structure prediction with AlphaFold,» *Nature*, vol. 596, pp. 583–589, 2021. DOI: 10.1038/s41586-021-03819-2
- [12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser e B. Ommer, «High-Resolution Image Synthesis with Latent Diffusion Models,» 2022. arXiv: 2112.10752 [cs.CV].
- [13] A. Dosovitskiy et al., «An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,» in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021. indirizzo: <https://arxiv.org/abs/2010.11929>
- [14] Z. Liu et al., «Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,» in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2021, pp. 10 012–10 022. DOI: 10.1109/ICCV48922.2021.00986 indirizzo: <https://arxiv.org/abs/2103.14030>
- [15] H. Hersbach et al., «The ERA5 global reanalysis,» *Quarterly Journal of the Royal Meteorological Society*, vol. 146, n. 730, pp. 1999–2049, 2020. DOI: 10.1002/qj.3803
- [16] European Centre for Medium-Range Weather Forecasts (ECMWF), «IFS Documentation – Cy48r1, Part VIII: Atmospheric Composition,» ECMWF, rapp. tecn., 2023, Operational implementation 27 June 2023.
- [17] Environmental Modeling Center, «The Global Forecast System (GFS) Version 16 (GFSv16) Implementation,» NOAA/NWS/NCEP/EMC, rapp. tecn., 2021, NOAA/NWS/NCEP/EMC GLOBAL MODEL TECHNICAL NOTE. indirizzo: <https://vlab.noaa.gov/web/gfs/documentation>
- [18] V. Eyring et al., «Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization,» *Geoscientific Model Development*, vol. 9, n. 5, pp. 1937–1958, 2016. DOI: 10.5194/gmd-9-1937-2016
- [19] E. Hu et al., «LoRA: Low-Rank Adaptation of Large Language Models,» *arXiv preprint arXiv:2106.09685*, 2021. arXiv: 2106.09685.
- [20] A. J. Charlton-Perez et al., «Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán,» *npj Climate and Atmospheric Science*, 2024. DOI: 10.1038/s41612-024-00638-w indirizzo: <https://doi.org/10.1038/s41612-024-00638-w>

- [21] A. Quarteroni, P. Gervasio e F. Regazzoni, «Combining physics-based and data-driven models: advancing the frontiers of research with Scientific Machine Learning,» *Mathematical Models and Methods in Applied Sciences (M2AS)*, 2025. DOI: 10.1142/S0218202525500125 arXiv: 2501.18708 [math.NA].
- [22] H. Yu, J. Liu, X. Zhang, J. Wu e P. Cui, «A Survey on Evaluation of Out-of-Distribution Generalization,» *arXiv preprint arXiv:2403.01874*, 2024. arXiv: 2403.01874 [cs.LG].
- [23] Google Research, *Google Colaboratory*, [Online], Accessed: YYYY-MM-DD. indirizzo: <https://colab.research.google.com/>
- [24] D. Giaioti, «Documentazione sulla strutturazione filesystem FENICE,» ARPA FVG - Centro Regionale di Modellistica Ambientale (CRMA), rapp. tecn. Art. Tec-Sci No. 003/2015, ago. 2015, Articolo Tecnico-Scientifico.
- [25] Altair Engineering, Inc., *Altair PBS Professional 2022.1 User's Guide*, Aggiornato 19/07/22, Altair Engineering, Inc., Troy, MI, USA, lug 2022. indirizzo: <https://www.altair.com>
- [26] E. Ghaderpour, «Map Projection,» *arXiv preprint arXiv:1412.7690*, 2014. arXiv: 1412.7690 [physics.geo-ph].
- [27] R. Rew, G. Davis, S. Emmerson, H. Davies e E. Hartne, *The NetCDF Users Guide: Data Model, Programming Interfaces, and Format for Self-Describing, Portable Data*, NetCDF Version 4.0.1, mar. 2009.
- [28] S. Al-Dabbagh, «Climate Data Operators CDO,» rapp. tecn., ott. 2018. DOI: 10.13140/RG.2.2.12068.40320
- [29] Microsoft, *Aurora: A Foundation Model for Earth System Forecasting*, Accessed: 2025-09-08, 2025. indirizzo: <https://microsoft.github.io/aurora/intro.html>
- [30] C. R. Rajulapati, S. M. Papalexiou, M. P. Clark e J. W. Pomeroy, «The Perils of Regriding: Examples Using a Global Precipitation Dataset,» *Journal of Applied Meteorology and Climatology*, vol. 60, n. 11, pp. 1561–1573, 2021. DOI: 10.1175/JAMC-D-20-0259.1
- [31] T. Schneider, «Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values,» *Journal of Climate*, vol. 14, n. 5, pp. 853–871, 2001. DOI: 10.1175/1520-0442(2001)014<0853:A0ICDE>2.0.CO;2
- [32] E. Acuna e C. Rodriguez, «The Treatment of Missing Values and its Effect on Classifier Accuracy,» in *Classification, Clustering, and Data Mining Applications*, Springer, 2004, pp. 639–647.

- [33] D. W. Pierce, *Ncview: A NetCDF Visualizer*, 2024. indirizzo: <http://cirrus.ucsd.edu/~pierce/ncview/>
- [34] S. d’Ascoli, H. Touvron, M. L. Leavitt, A. S. Morcos, G. Biroli e L. Sagun, «ConViT: improving vision transformers with soft convolutional inductive biases\*,» *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2022, n. 11, p. 114 005, nov. 2022, ISSN: 1742-5468. DOI: 10 . 1088/1742-5468/ac9830 indirizzo: <http://dx.doi.org/10.1088/1742-5468/ac9830>
- [35] D. Gyawali, *Comparative Analysis of CPU and GPU Profiling for Deep Learning Models*, 2023. arXiv: 2309.02521 [cs.DC]. indirizzo: <https://arxiv.org/abs/2309.02521>