Supplementary:
Efficient Localized Adaptation of Neural Weather Forecasting: A Case Study in the MENA Region

A More details on GLoRA

GLoRA overall formulates in Eq.2

$$\mathcal{G} = (\mathcal{W}_0 + \mathcal{W}_0 U + V)x + X\mathcal{W}_0 + Yb_0 + Z + b_0 \tag{2}$$

The tensors denoted as U, V, X, Y, and Z serve as trainable support structures for downstream tasks. Notably, during the entire fine-tuning process, the W_0 and b_0 remain fixed. Among these tensors, U is responsible for scaling the weights, while V plays a pivotal role in both scaling the input and shifting the weights. X functions as a layer-wise prompt, akin to the prompt tuning paradigm. Y and Z are used in scaling and shifting the bias, respectively, contributing to the overall functionality of the GLoRA model. For more technical details, we refer the reader to [17].

B Dataset

This dataset, available on a global 0.25° × 0.25° latitude-longitude grid, spans around 40 years, with hourly measurements encompassing 37 altitude levels and the Earth's surface. Comprising 721 × 1440 grid points, the dataset presents altitude levels in terms of pressure levels. ERA5 represents the fifth generation of ECMWF reanalysis, and offers improved spatial and temporal resolution. Reanalysis employs data assimilation techniques to combine model forecasts with observations, yielding a consistent and complete dataset that spans multiple decades. The incorporation of ensemble-based uncertainty estimates enhances the utility of ERA5 for climate-related applications. Furthermore, pre-calculated monthly-mean averages facilitate analysis and interpretation. This comprehensive dataset highlights the critical role of reanalysis in advancing weather modeling and climate analysis.

C Implementation Details

In our experiments, we operate with two levels of resolutions 5.625° (32 × 64 grid points) and 1.40625° (128 × 256 grid points). The dataset is partitioned into training data spanning from 1979 to 2015, validation data for the year 2016, and test data covering 2017 and 2018. All the default variables have been utilized as inputs in our experiments. For a 5.625° resolution, a patch size of 2, and for 1.40625° resolution, a patch size of 4 is used. The learning rate adopted for our experiments is $1.0\text{x}10^{-5}$. The embedding dimension is set to 1024. The training process is conducted using four V100 GPUs, leveraging fp16 floating point precision. We present the results for a wide range of atmospheric variables, including *geopotential_500, temperature_850, 2m_temperature, 10m_u_component_of_wind, 10m_v_component_of_wind, relative_humidity_850, and specific_humidity_850*. We present results using the anomaly correlation coefficient (ACC) and root mean square error (RMSE). For ACC, higher values indicate better performance, while for RMSE, lower values are preferable. These metrics together provide a comprehensive assessment of model performance. The regional setting is the MENA region which is a subset of global grid points.

With the utilization of a comprehensive set of input variables derived from atmospheric and surface data, we focus on forecasting future weather conditions given the current atmospheric state. Specifically with a total of 48 input features, majorly includes geopotential, temperature, wind components, relative humidity, and specific humidity at various pressure levels are crucial for weather prediction tasks due to their direct influence on atmospheric dynamics. The forecasting tasks involve predicting seven target variables: geopotential at 500hPa, 2-meter temperature, and eastward & northward components of the 10m wind, and temperature, relative humidity, & specific humidity at 850hPa, Lead times ranging from 12 hours to 72 hours are considered, encompassing multiple range forecasting scenarios. For training the deep learning models, as described in ClimaX, we utilize a latitude-weighted mean squared error (MSE) loss function and implement early stopping based on validation loss to prevent overfitting. For more details, we refer readers to [7].

Metric	Lead time	geop@500	2m_temp	r_hum@850	s_hum@850	temp@850	10m_u_wind	10m_v_wind
ACC	**12**	0.985	0.917	0.870	0.899	0.954	0.950	0.952
	24	0.951	0.914	0.797	0.846	0.927	0.907	0.903
	36	0.876	0.880	0.705	0.782	0.868	0.829	0.815
	48	0.779	0.868	0.631	0.727	0.792	0.742	0.712
	60	0.674	0.814	0.554	0.668	0.696	0.650	0.602
	72	0.585	0.804	0.502	0.623	0.620	0.570	0.517
RMSE	**12**	87.689	0.961	10.721	0.001	0.877	1.077	1.125
	24	152.173	0.973	13.151	0.001	1.090	1.465	1.579
	36	237.892	1.223	15.450	0.001	1.412	1.957	2.138
	48	311.773	1.258	16.939	0.002	1.800	2.357	2.604
	60	370.279	1.489	18.206	0.002	2.142	2.692	2.982
	72	411.125	1.518	18.945	0.002	2.366	2.931	3.219

Table 3: ***Regional models over different prediction ranges to show trends regarding lead time with respective models.***: Forecasting on MENA region at resolution 1.40652°. Results reported in anomaly correlation coefficient (ACC) and root mean square error (RMSE).

Metric	Rank (r)	geop@500	2m_temp	temp@850
ACC (↑)	2	0.575	0.824	0.612
	4	0.576	0.824	0.612
	8	0.575	0.824	0.612
	16	0.582	0.823	0.614
	32	0.579	0.823	0.612
RMSE (↓)	2	439.717	1.400	2.482
	4	439.969	1.401	2.483
	8	440.129	1.395	2.479
	16	438.213	1.414	2.476
	32	439.885	1.417	2.482

Table 4: ***Impact of rank (r) in LoRA module***: Forecasting on MENA region with the regional model for 72 hrs prediction range. Resolution is 5.625°. This table shows the trends that how rank value impacts the overall performance. We report three atmospheric variables here.

D More Results

Predictions over ranges: For more accurate predictions, it was observed that a specialized model with a specific lead time stands out. With this approach, we train multiple models for specific lead time ranges from 12 hours to 3 days. In Table 3, we report results on different ranges of predictions that show the behavior of atmospheric variables over time. We can observe that as the time varies from short to medium, it becomes very crucial to predict the accurate forecast.

D.1 Ablations

In this section, we delve into two different mechanisms. Firstly, we explore the influence of the rank number on model behavior to understand how it changes with varying rank numbers. Secondly, we investigate the integration of LoRA with components other than the attention module, aiming to assess the impact of such integration on model performance.

Impact of rank r: In investigating the impact of rank r in the Lora module, we have observed that within the range of Lora ranks, optimal results are achievable by integrating rank 16. Specifically, in the context of low-resolution settings, our examination of rank r behavior for regional forecasting over a 72-hour prediction range is shown in Table 4.

Lora's significance on attention and fc layers: We also seek to investigate the effectiveness of the LoRA module in conjunction with the feed-forward network, rather than the attention module only. Surprisingly, we observe a degradation in performance with the integration of LoRA into a model of this size. Specifically, we aim to incorporate LoRA with the feed-forward network ($Lf1$ & $Lf12$) and the results of these experiments are detailed in Table 5.

Metric	**ACC** (↑)			**RMSE** (↓)		
Model	geop@500	2m_temp	temp@850	geop@500	2m_temp	temp@850
Regional$_{Lora}$	0.582	0.823	0.614	438.213	1.414	2.476
Regional$_{Lora+Lf1}$	0.540	0.808	0.578	453.996	1.459	2.576
Regional$_{Lora+Lf12}$	0.540	0.808	0.578	454.228	1.458	2.574

Table 5: ***LoRA module in addition to attention module***: Forecasting on MENA region with the regional model for 72 hrs prediction range. Resolution is $5.625°$. It is observed that in addition to the attention module when Lora is extended to the feed-forward network, performance degrades.

Model	Params (M)	Convergence (Hrs)	GPU Memory (GB)
Regional$_{fft}$	108.0	$\sim$ 8.6	15.2
Regional$_{Lora}$	16.2	$\sim$ 4.5	9.3

Table 6: ***Parameters in Lora and full fine tuning***: Resolution is $5.652°$. LoRA requires fewer parameters for training, uses less GPU memory for computations, and converges faster (in terms of training time) compared to the full fine-tuning (fft) approach.

Memory and time comparison: Compared to full fine-tuning, LoRA in a parameter-efficient fine-tuning paradigm, significantly reduces the number of trainable parameters and accelerates convergence. Furthermore, during training, the memory consumption on the GPU is noticeably lower in LoRA, demonstrating 38.82% memory reduction and 85.0% parameter reduction relative to the full fine-tuning model. For more details see Table 6.

D.2 Qualitative

Extreme Weather Event: In Fig. 3, it is shown that our Lora-tuned model performs better and gives minimum error in predictions with respect to ground truth. This also signifies the stability of the model for extreme weather conditions that happened in June 2017 in Kuwait [23].

Non-MENA region: To evaluate the generalizability of the regional model, we assessed its performance on the Non-MENA region (primarily China and its surroundings), where it had not encountered data over the historical range of years. Our regional model demonstrated superior performance as shown in Fig. 4. This improved generalizability may be due to the model's focus on a specific region, allowing it to learn more consistent and relevant features and develop a coherent understanding of local data attributes. Consequently, the regional model can perform better on unseen data from other regions with similar patterns. In contrast, the global model might face challenges due to the diverse and complex data it encounters from multiple regions.

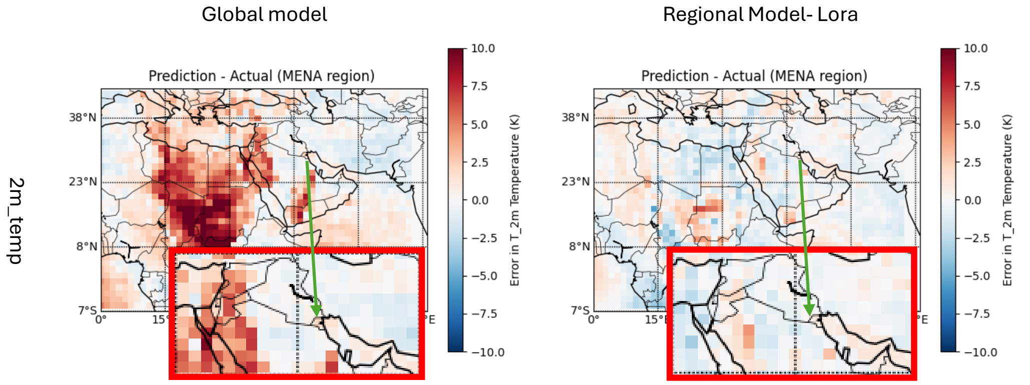
Global model
Regional Model- Lora
2m_temp
Prediction - Actual (MENA region)
Prediction - Actual (MENA region)
38°N
23°N
8°N
7°S
0°
15°
Error in T_2m Temperature (K)
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0

Figure 3: Qualitative: Error/Bias in Predictions and Actual measurements for temperature_2m (K), with reference to [23]. Dated, 22^{nd} June 2017.

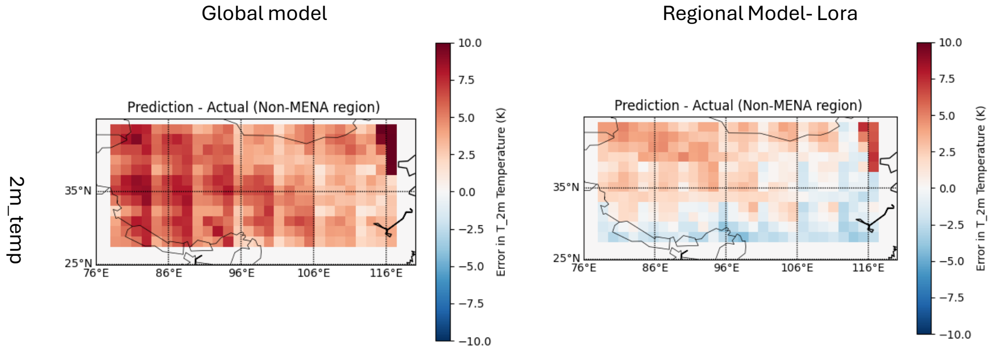
Global model
Regional Model- Lora
2m_temp
Prediction - Actual (Non-MENA region)
Prediction - Actual (Non-MENA region)
35°N
25°N
76°E
86°E
96°E
106°E
116°E
Error in T_2m Temperature (K)
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0

Figure 4: Qualitative: Error/Bias in Predictions and Actual measurements for temperature_2m (K), on Non-MENA region. Dated, 20^{th} May 2017.