

# Satellite Observations Guided Diffusion Model for Accurate Meteorological States at Arbitrary Resolution

## Supplementary Material

### 6. Limitations and Future Work.

The training data in SGD consists of ERA5 and GridSat. However, SGD serves as a versatile framework that could incorporate more modalities such as other reanalysis data, observations from polar-orbiting satellites, sounding and radar data. Once these systematic data are all integrated into SGD, more accurate weather conditions near the surface can be achieved.

### 7. Preliminary

Unconditional diffusion model, proposed by [18], is a powerful generative model composed of a forward process and a reverse process. The former aims to gradually introduce random Gaussian noise into the original images over  $T$  diffusion steps, ultimately resulting in pure Gaussian noise  $x_T \sim \mathcal{N}(0, I)$ . The latter, being the reverse of the forward process, intends to denoise and sample the generated images from the pure Gaussian noise through a pre-trained noise estimation network.

The forward process is a Markov chain without learnable parameters. The denoising method for each step is defined by the following equation, where  $\beta_t$  refers to the variance of the forward process, which is experimentally set as a hyperparameter solely dependent on the diffusion steps  $t$ .

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I). \quad (7)$$

For each steps in the reverse process  $p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta I)$ , the mean of the distribution is hard to compute directly as the forward process. Consequently, we necessitate the utilization of a neural network with parameter  $\theta$  to estimate the noise inherent within the image  $x_t$ . By employing Bayes theorem, we can express the mean and variance of the reverse process as follows:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon_\theta(x_t, t)) \quad (8)$$

$$\Sigma_\theta(x_t) = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}\beta_t, \quad (9)$$

Among them,  $\epsilon_\theta(x_t, t)$  represents the noise estimation function, which is pre-trained by utilizing the low-resolution ERA5 maps. It performs real-time estimation and simulation of the noise contained within the maps, thereby enabling denoising to sample  $x_{t-1}$ . The unconditional diffusion model is trained utilizing maximum likelihood estimation, with the objective for each training iteration defined as follows:

tion defined as follows:

$$E_{\epsilon \sim \mathcal{N}(0, I), t \sim [0, T]}[\|\epsilon - \epsilon_\theta(x_t, t)\|^2]. \quad (10)$$

### 8. Patch-based Methods

The scale of the ERA5 maps used as input for SGD reaches  $25km \times 25km$ . To address the downscaling task at this resolution, we employed a patch-based method during the sampling process of the conditional DDPM. The detailed introduction of this method is shown in Algorithm 2, the patch-based method partitions the low-resolution ERA5 maps into several sub-regions based on a fixed stride and size. For each individual sub-region, the gradient term of the distance loss between the ERA5 maps obtained from convolution kernels and the corresponding low-resolution sub-region map is computed separately. Subsequently, the mean of the Gaussian distribution and the parameters of the convolution kernels in each sub-region are updated based on these gradient terms. The overall map is then updated by averaging the updated values across all sub-regions, each weighted by a binary patch mask that quantifies the regional scope, thereby refining the overall mean and convolution kernel parameters of the sampled high-resolution ERA5 map, resulting in smoother generated maps. By leveraging this strategy, SGD is capable of downscaling ERA5 maps to any desired resolution, further enhancing the practicality of the model.

### 9. Pre-trained Encoder

Before utilizing cross attention for feature fusion, SGD necessitates the extraction of features from GridSat maps by an encoder. The pre-trained encoder aims to enhance the feature extraction capabilities of SGD and its downscaling performance. The pre-trained module comprises two components: the encoder and a decoder of symmetric structure. The former is utilized to extract features from GridSat maps into latent space, while the latter aims to reconstruct the encoder's outputs. The encoder module consists of several convolutional layers, employing  $3 \times 3$  convolutional kernels with a padding of 1, elevating the GridSat maps' channel count to 64. Similarly, the decoder also encompasses convolutional layers, responsible for the reconstruction of the extracted features, the detailed structure is shown in Fig. 6. The training objective is to minimize the MSE loss between the input GridSat maps and the output maps post-decoder, with the total training epochs approximating 100.

**Algorithm 2** Patch-based Methods of SGD: Guided diffusion model with the guidance of low-resolution ERA5 map  $z$ . Given a conditional diffusion model pre-trained on ERA5 and GridSat maps  $\epsilon_\theta(x_t, y, t)$ .

**Input:** Conditional input GridSat satellite observation map  $y$ , low-resolution ERA5 map  $z$ . Downscaling convolutional kernel  $\mathcal{D}$  with parameter  $\varphi$ . Pre-trained encoder module  $f$  with parameter  $\phi$ . Learning rate  $l$  and guidance scale  $s$ . Distance measure function  $\mathcal{L}$ . Overlapping patch stride  $r$ , overlapping patch size  $v = 720 \times 1440$ . Overlapping patch set  $K$ , each patch commences its traversal from the top-left block of the  $720 \times 1440$  grid on the maps, progressing sequentially with a displacement of stride  $r$ . A binary patch mask set  $\{P^k\}, k \in K$ .

**Output:** Output high resolution ERA5 map  $x_0$ .

- 1: Sample  $x_T$  from  $\mathcal{N}(0, I)$
- 2:  $y' = f_\phi(y)$
- 3: **for all**  $t$  from  $T$  to  $1$  **do**
- 4:  $\tilde{x}_0 = \frac{x_t}{\sqrt{\alpha_t}} - \frac{\sqrt{1-\alpha_t}\epsilon_\theta(x_t, t)}{\sqrt{\alpha_t}}$
- 5: **for all**  $i$  from  $1$  to  $|K|$  **do**
- 6:  $\mathcal{L}_{\varphi^i, \tilde{x}_0} = \mathcal{L}(z \circ P^i, \mathcal{D}^{\varphi^i}(\tilde{x}_0 \circ P^i))$
- 7:  $\varphi^i \leftarrow \varphi^i - l \nabla_{\varphi^i} \mathcal{L}_{\varphi^i, \tilde{x}_0}$
- 8:  $\tilde{x}_0^i \leftarrow \tilde{x}_0^i - \frac{s(1-\alpha_t)}{\sqrt{\alpha_{t-1}\beta_t}} \nabla_{\tilde{x}_0} \mathcal{L}_{\varphi^i, \tilde{x}_0}$
- 9:  $\tilde{\mu}_t^i = \frac{\sqrt{\alpha_{t-1}\beta_t}}{1-\alpha_t} \tilde{x}_0^i + \frac{\sqrt{\alpha_t}(1-\alpha_{t-1})}{1-\alpha_t} x_t$
- 10: **end for**
- 11:  $\varphi = \frac{1}{|K|} \sum_{j=1}^{|K|} \varphi^j \circ P^j$
- 12:  $\tilde{\mu}_t = \frac{1}{|K|} \sum_{j=1}^{|K|} \tilde{\mu}_t^j \circ P^j$
- 13:  $\tilde{\beta}_t = \frac{1-\alpha_{t-1}}{1-\alpha_t} \beta_t$
- 14: **end for**
- 15: Sample  $x_{t-1}$  from  $\mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$
- return**  $x_0$

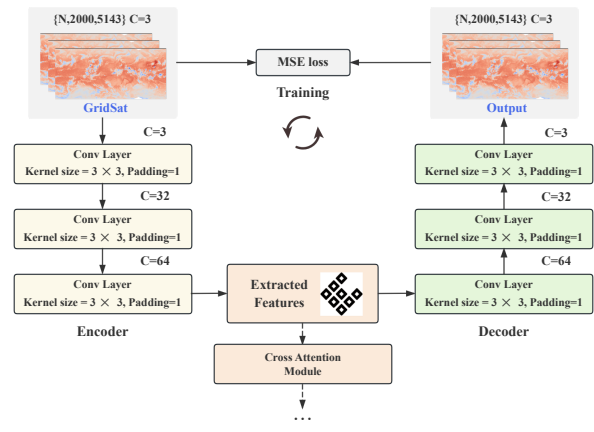


Figure 6. Overall architecture of the encoder and decoder modules used in SGD.

## 10. Additional Visualization Results

In this section, we present the downscaling results of SGD for the variables  $V_{10}$  and  $MSL$ . Fig. 7 shows that SGD exhibits more faithful details in the maps as compared to interpolation-based and diffusion-based methods. Furthermore, SGD exhibits no discernible disparity in overall intensity when compared to ERA5 at a scale of  $25km \times 25km$ . Combining the results of the other two variables in the main text, it is validated that SGD is capable of producing highly satisfactory downscaling results across all four variables.

## 11. Station observation-guided downscaling bias with stations in Weather2K

In this section, we endeavor to integrate the MSE loss from ERA5 LR maps and MAE loss from the observation stations in Weather5K within the distance function utilized in the sampling process. Subsequently, we evaluated the high-resolution ERA5 maps derived from SGD with this setting across all stations within the Weather2K dataset, thereby further assessing the efficacy of the guided sampling and the accuracy of the downscaling results.

Weather2K dataset [46] is a benchmark dataset that aims to address the shortcomings of existing weather forecasting datasets in terms of real-time relevance, reliability, and diversity, as well as the critical impediment posed by data quality. The data is available from January 2017 to August 2021. It encompasses the meteorological data from 2130 ground weather stations across 40896 time steps, with each data incorporates 3 position variables and 20 meteorological variables.

Specifically, we incorporate the MAE loss between the generated HR ERA5 maps and station observations from the Weather5K dataset [15] with equal weights into our distance function to measure bias. Subsequently, we calculate the biases between the downscaling results obtained under this setting with the meteorological data at the stations from the Weather2K dataset. The evaluation metrics we employed are the MSE and MAE loss of the variable  $T_{2M}$ .

We compared our results with those of interpolation-based and diffusion-based methods using the same metrics. As shown in Tab. 5, the discrepancy between ERA5+station guided SGD and Weather2K stations is smaller, indicating that using ERA5 and Weather5K with equal weights as the distance function yields more ideal downscaling results for stations beyond Weather5K.

Fig. 8 illustrates the differences in downscaling results among various methods at parts of the stations within Weather2K, with darker colors indicating smaller discrepancies at the stations. In terms of the bias between the downscaled results at the station locations in the image and the actual observations, SGD with mixed guidance down-

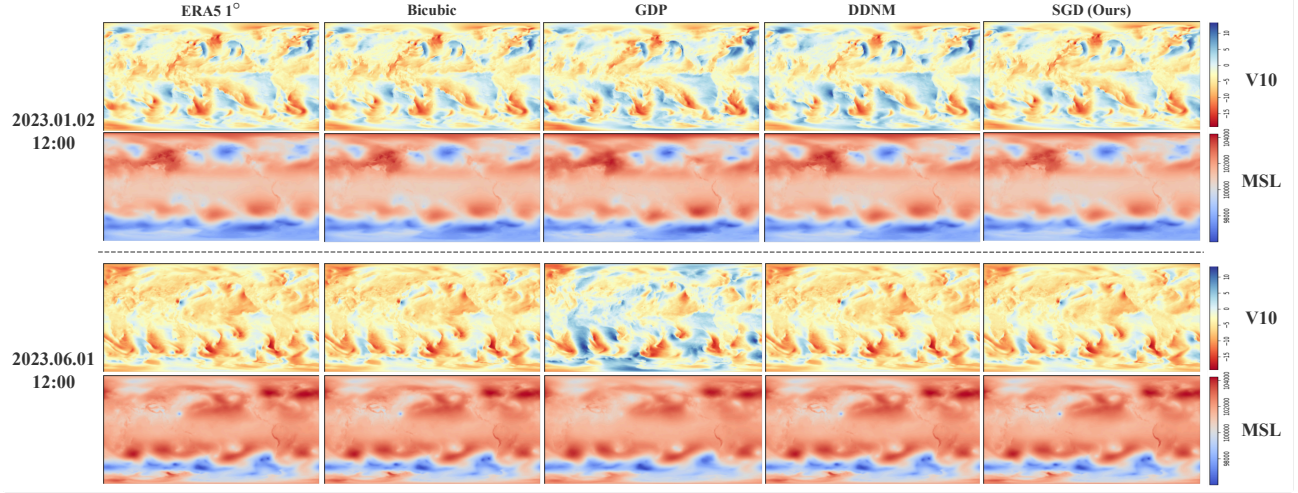


Figure 7. Visualization comparison of different interpolation-based and diffusion-based downscaling results in various time stamps. We use different colors to distinguish  $V_{10}$  and  $MSL$ .

scaling results has less extreme bias stations, which is symbolized as yellow-labeled stations. Moreover, the overall station coloration appears deeper. This suggests that utilizing weather5k as guidance can enhance the model’s performance in downscaling at the local scale, aligning more closely with the real conditions.

## 12. Running Time and Resource Consumption

Tab. 6 shows the running time and resource consumption of SGD during the training and the sampling process. To enhance the inference efficiency, we have also tested our SGD on 50-step DDIM sampling to generate HR ERA5 maps within one minute, making it a feasible approach for practical use.

## 13. Ablation Studies on the Relationship of Variables

SGD utilizes the coupling relationship between data from satellite observation and ERA5 maps as a condition. For details, ERA5, as a reanalysis dataset, is derived from satellite observations and other data. Among them, the brightness temperature from satellite observations provides temperature variations being the primary driver of atmospheric changes. Therefore, the high-quality brightness temperature data from GridSat play a crucial role in the ERA5 reanalysis process. Moreover, atmospheric state variables influence observations through radiative processes, while observational data, in turn, feed back into ERA5 via data assimilation systems.

To reveal the impact of variables in GridSat on the downscaling of ERA5 maps, an ablation study on the variable

relationship was conducted to quantify the degree of influence. When only the brightness temperature variables from GridSat (IrWin\_Cdr or IrWin\_VZA\_Adj) are employed as the condition, a satisfactory performance can be obtained, demonstrating that brightness temperature is an important condition for ERA5 maps downscaling (Tab. 7). All variables from GridSat could guide the SGD to yield higher-quality HR ERA5 maps. When incorporating GridSat as conditions and utilizing only a single LR ERA5 variable as guidance to generate the single HR ERA5 variable, the introduction of GridSat yields the most significant enhancement for the ERA5 temperature variable ( $T_{2m}$ ) (Tab. 8). Considering the correlation between sea-level pressure and brightness temperature, the incorporation of GridSat as conditions for generating the single  $MSL$  variable also contributes to enhancing the  $MSL$  downscaling.

Table 5. Station-level downscaling results for  $T_{2M}$ , which utilize the stations from Weather2k to assess the bias between the downscaling maps and Weather2k station observation values. ERA5 guided and ERA5 + station guided SGD respectively denote the SGD models that employ the MSE loss between the generated maps and ERA5 maps as the sole distance function, and the SGD model that integrates the Weather5k station observations into its distance function.

Variable	Metrics	ERA5 1°	ERA5 0.25°	GDP	ERA5 Guided SGD	ERA5+Station Guided SGD
$T_{2M}$	MSE	17.51	17.80	18.87	18.08	<b>15.61</b>
	MAE	407.81	420.38	466.00	431.33	<b>355.31</b>

Table 6. The running time and resource consumption of SGD.

Mode	SGD Training	SGD Sampling	SGD Sampling with DDIM
Running Time	48h	6min	1min
Resource Consumption	$5 \times 10^4$ MiB	$1.8 \times 10^4$ MiB	$1.6 \times 10^4$ MiB

Table 7. Ablation study employing a single GridSat variable.

Methods	$U_{10}$		$V_{10}$		$T_{2m}$		$MSL$	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Only IrWin.Cdr	55.74	6.07	46.11	5.76	191.42	11.04	412.11	16.70
Only IrWin.VZA.Adj	56.08	5.99	47.53	5.84	194.05	10.87	405.74	16.55
Only IrWVP	60.42	6.74	55.08	6.10	214.07	12.07	424.17	17.52
All Variables in GridSat	<b>51.65</b>	<b>5.84</b>	<b>39.82</b>	<b>5.05</b>	<b>187.69</b>	<b>10.63</b>	<b>374.39</b>	<b>14.49</b>

Table 8. Ablation study employing a single ERA5 variable.

$U_{10}$			$V_{10}$			$T_{2m}$			$MSL$		
Methods	MSE	MAE	Methods	MSE	MAE	Methods	MSE	MAE	Methods	MSE	MAE
ERA5 1°	53.18	5.95	Era5 1°	38.51	4.95	ERA5 1°	216.27	11.39	Era5 1°	470.06	15.78
Only $U_{10}$	56.72	6.45	Only $V_{10}$	47.28	5.84	Only $T_{2m}$	194.15	10.71	Only $MSL$	398.05	14.90



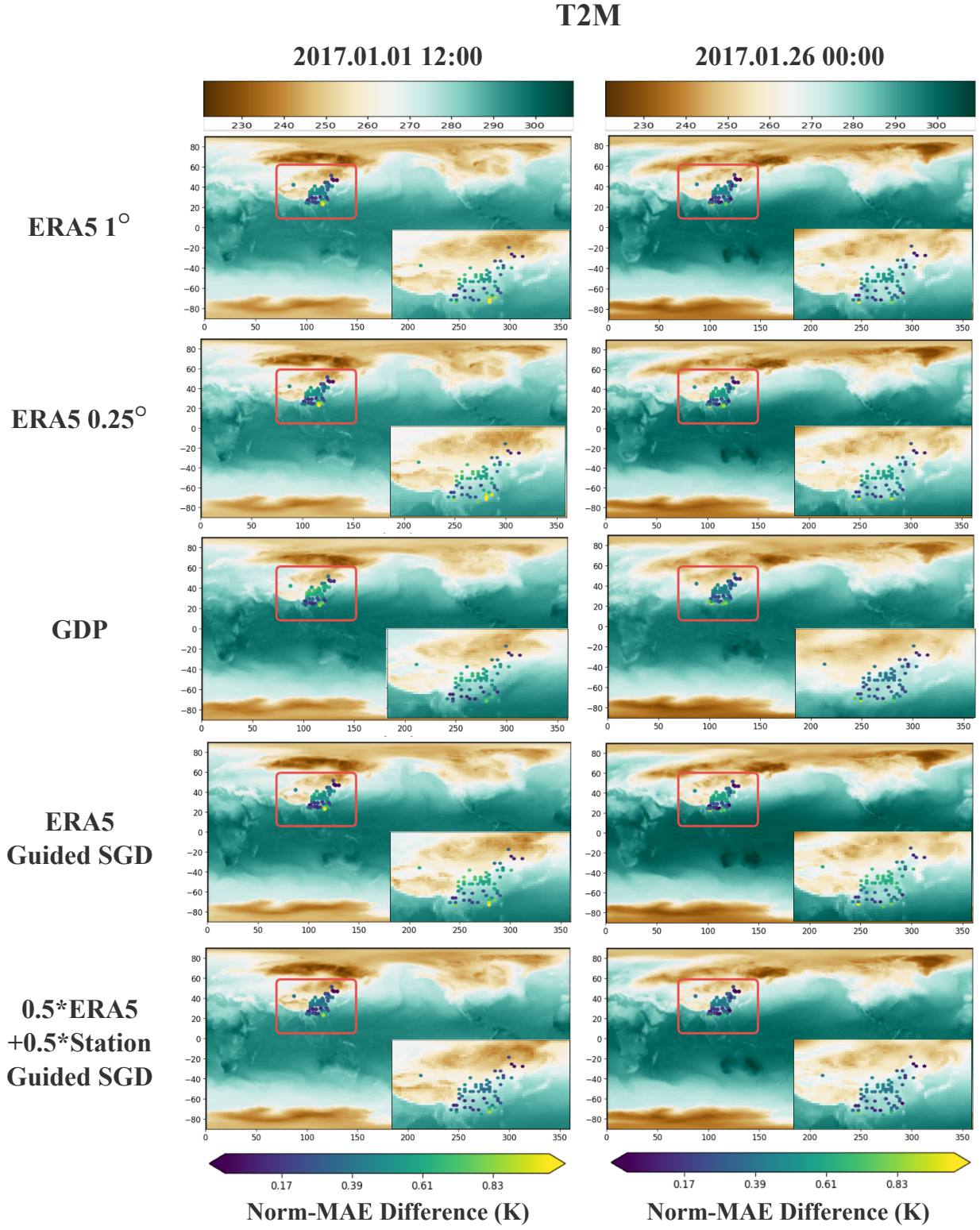


Figure 8. Visualization comparison of SGD downscaling to station-scale employing various distance function, where the coloration of each Weather2k observation station signifies the MAE loss between the downscaled results and their corresponding observed values.