Check for
updates

# Forecasting upper atmospheric scalars advection using deep learning: an $O_3$ experiment

**Luiz Angelo Steffenel[1]** (ORCID) **· Vagner Anabor[2] · Damaris Kirsch Pinheiro[2] ·
Lissette Guzman[2] · Gabriela Dornelles Bittencourt[2,3] · Hassan Bencherif[3]**

## Abstract

Weather forecast based on extrapolation methods is gathering a lot of attention due to the advance of artificial intelligence. Recent works on deep neural networks (CNN, RNN, LSTM, etc.) are enabling the development of spatiotemporal prediction models based on the analysis of historical time-series, images, and satellite data. In this paper, we focus on the use of deep learning for the forecast of stratospheric Ozone ($O_3$), especially in the cases of exchanges between the polar vortex and mid-latitudes known as Ozone Secondary Events (OSE). Secondary effects of the Antarctic Ozone Hole are regularly observed above populated zones on South America, south of Africa, and New Zealand, resulting in abrupt reductions in the total ozone column of more than 10% and a consequent increase in UV radiation in densely populated areas. We study different OSE events from the literature, comparing real data with predictions from our model. We obtained interesting results and insights that may lead to accurate and fast prediction models to forecast stratospheric Ozone and the occurrence of OSE.

✉ Luiz Angelo Steffenel
angelo.steffenel@univ-reims.fr

Vagner Anabor
vanabor@ufsm.br

Damaris Kirsch Pinheiro
damaris@ufsm.br

Hassan Bencherif
hassan.bencherif@univ-reunion.fr

[1]  LICIIS Laboratory - LRC CEA DIGIT, Université de Reims Champagne Ardenne, Reims, France

[2]  Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

[3]  LACy Laboratory, Université de la Réunion, Saint Denis, La Réunion, France

# 1 Introduction

Weather forecast has become an essential asset in our society, from the simple temperature and rain trends seen every day on newspaper and TV to domain-specific forecasts such as weather guidance for airports, wind and solar energy generation, emergency alerts for rainfall, flash floods, severe weather, hurricanes, etc. The existing methods for weather forecast can roughly be categorized into two classes (Sun et al. 2014), (i) methods based on numerical weather prediction (NWP) models, and (ii) empirical methods based on extrapolations from storms growing rates, velocities from radar echoes, infrared satellite images, etc.

Forecasts based on NWP use complex numerical integration schemes to solve the governing equations of physical phenomena in the atmosphere. These equations demand initial and boundary conditions coming from large data volumes from ground and satellite observations, and must to be processed in large computational systems in order to solve and create the weather forecast numerical models outputs in time to prevent natural disasters. Nowadays, sophisticate and computationally intensive data assimilation schemes are employed in models as NCEP-CFSv2 (Kistler et al. 2001) and ERA5 (Malardel et al. 2016).

The second approach, based on extrapolation methods, is gathering a lot of attention in recent years. Due to the development of artificial intelligence, especially deep learning, forecasts based on the analysis of historical time-series, images, and satellite data are achieving stunning results.

Two main sub-domains of meteorology ad atmosphere sciences are leading this development. The first one is related to the prediction of extreme weather conditions such as hurricanes (Prabhat et al. 2015; Moradi Kordmahalleh et al. 2016; Racah et al. 2017; Pradhan et al. 2018). The second sub-domain is that of precipitation nowcasting, whose resolution and time accuracy needs are much higher than other traditional forecasting tasks like weekly average temperature prediction. These domains rely on computer vision techniques, which have proven useful for making accurate extrapolation of radar and satellite maps (Sakaino 2013; Shi et al. 2015; Agrawal et al. 2019).

While weather forecasts usually focus on natural events as flash floods, rain, wind gust, and snowstorms, all embedded in the troposphere, the upper atmosphere also demands a special and expensive modeling treatment since this region may play an important role all over weather and climate circulations (Scaife et al. 2012). The stratosphere is a barotropic portion of the Atmosphere and its scalar fields advection (heat, gases, and particulate materials) happens constricted between in isentropic layers, in an almost horizontal displacement, a simpler environment if compared with the troposphere. As the main stratospheric information comes from remotely sensed data from satellites, its modeling may benefit from deep learning extrapolation methods, allowing, for example, to estimate the wind field using the cloudiness displacement and time-difference techniques, i.e., the atmospheric scalar field itself used to estimate the winds that promote its transport.

Among these high atmosphere interest topics, our interest is directed towards the forecast of the Ozone ($O_3$) layer. Ozone is the most important constituent of stratospheric gas traces. Due to its ability to absorb ultraviolet radiation (UV) (Salby 1996; Dobson 1968), $O_3$ is also the most important component in the stratosphere from the point of view of skin protection against harmful UVB solar radiation. Many sources contribute to the monitoring of stratospheric $O_3$. While ground-level equipment can provide information about the Total Ozone Column ($TCO_3$) every couple of minutes and ozone sounding balloons can provide ozone profiles for different heights, most of the data used to estimate the $O3$ global

coverage is originated from satellite sources, which update only once or twice a day. While some NWP models are used to forecast $O_3$ concentration (e.g., NASA's OzoneWatch and ESA's TEMIS websites), no extrapolation methods seem to have been used until now.

In this paper, we aim at the observation and forecast of Ozone total column and notably "Ozone Secondary Events" (OSE), i.e., episodes of exchange between the polar vortex and the mid-latitudes and the tropics (Bencherif et al. 2007; Marchand et al. 2005) in which the polar vortex is deformed and ozone-poor polar air masses move towards mid-latitudes. We aim to provide accurate predictions two to four days before the arrival of OSE above populated areas. This paper largely expands the preliminary results we presented in a poster at EGU General Assembly 2019 (Steffenel et al. 2019). Besides the proposal of a forecasting framework based on deep learning, this work contribution includes an extensive analysis of the results in both domain-specific parameters (interest of the results from the meteorological point of view) and deep learning metrics (image similarity, for example). To our knowledge, this is one of the first attempts to forecast the circulation of the Ozone layer using deep learning. In addition, this study aims to present deep learning as a viable, fast, and computationally inexpensive technique for forecasting, as the only computing-intensive phase is the training phase.

The remainder of this paper is structured as follows: Sect. 2 introduces the Ozone Secondary Event problem and reviews existent works that use Deep Learning techniques to perform forecasting. Section 3 describes our forecasting framework, including data preparation and training steps. Section 4 illustrates the forecasting framework through a detailed analysis of different OSE recorded in the last 10 years. In this section, we perform both domain-specific (meteorology) and domain-agnostic analysis of the results. Section 5 addresses directions to improve the framework in future works. Finally, Sect.6 presents our conclusions.

## 2 Problem description and methodology

### 2.1 Forecasting ozone secondary events

Since its discovery, the "Antarctic ozone hole" has attracted the interest of the scientific community. An ozone hole area is defined as a region with values below 220 DU (Hofmann et al. 1997). The concentration of ozone in a particular region of the Earth is mainly determined by the meridional transport of this element in the stratosphere (Gettelman et al. 2011). The explanation for the higher concentration of ozone found in polar rather than equatorial regions (where there is greater production) is precisely a special type of poleward transport known as the Brewer-Dobson circulation, in which air masses are transported quasi-horizontally from the stratospheric tropical reservoir to polar regions (Brewer 1949; Dobson 1968; Bencherif et al. 2007).

Given the dynamics of the atmosphere, the ozone hole is accompanied by episodes of exchange between the polar vortex and the mid-latitudes and the tropics (Bencherif et al. 2007; Marchand et al. 2005). During these events, the polar vortex is deformed and ozone-poor polar air masses move towards mid-latitudes. We call these episodes of isentropic exchanges "Secondary Effects of the Antarctic Ozone Hole" or "Ozone Secondary Effects" (OSE, for short). This temporary drop in ozone content first was observed by Kirchhoff et al. (1996) above the south of Brazil. Such episodes may last several days and reach the
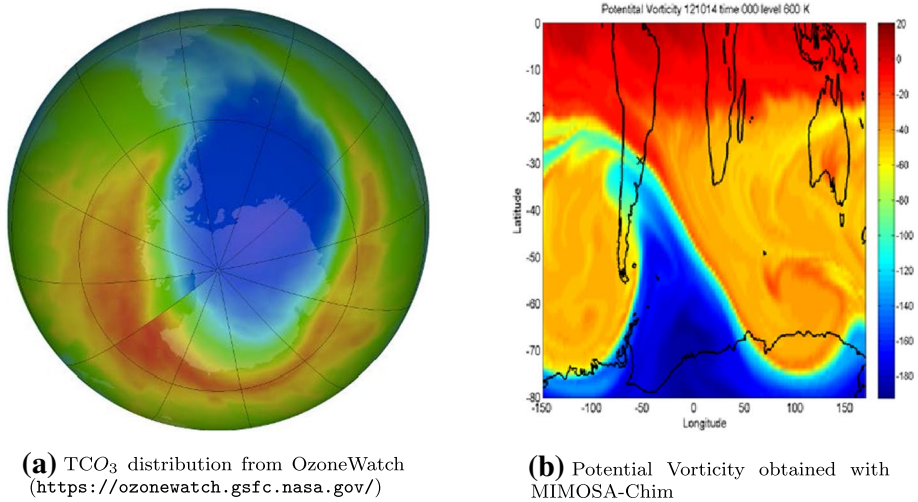
(a) TCO$_3$ distribution from OzoneWatch (https://ozonewatch.gsfc.nasa.gov/)

(b) Potential Vorticity obtained with MIMOSA-Chim

**Fig. 1** OSE event as **a** observed from space on 12 October 2014 and **b** corresponding PV simulated by the MIMOSA-Chim model (Hauchecorne et al. 2002)

tropics, causing significant ozone decreases over these areas and potentially increasing the UV radiation levels at the surface (Bencherif et al. 2007; Casiccia et al. 2008).

Secondary effects of the Antarctic Ozone Hole in mid-latitude zones are regularly observed above populated zones in South America, south of Africa, and New Zealand. Indeed, ground observations reported the occurrence of dozens of OSEs in the last decades above the south of Brazil (Bittencourt et al. 2019), resulting in temporary reductions in the total ozone column of more than 10% over densely populated areas. According to the latest World Meteorological Organization (WMO) reports (Research and Project 2014, 2018), there is a growth trend between the 1980s and 1990s, stabilizing at high rates since the year 2000 despite indications of declining trends in Antarctic ozone in recent years (Solomon et al. 2016).

An example of OSE the event from October 10–14, 2012 presented below in Fig. 1a, which shows the displacement of a polar air mass (in blue) and its extension towards the south of Brazil – around 30°S, driving a 13.7% reduction in the total ozone column (Vaz Peres et al. 2017). Figure 1b, which shows the potential vorticity (PV) map during this event, also illustrates the movement of a polar air mass (in blue) and its extension towards mid-latitudes above South America.

It is also important to note that most works on OSE limit to the study of past events (Canziani et al. 2002; Bencherif et al. 2011; Peres 2013; Bittencourt et al. 2019), whose characterization depends on several indirect factors besides the $O_3$ Total Column. For example, Potential Vorticity (PV) has been used to identify air masses originated in the pole, confirmed by a retroactive trajectory analysis using the HYSPLIT model (Stein et al. 2016). Forecasting OSE was never the main issue up to now.

Predicting sudden reductions in the Ozone coverage like OSE is important as the additional UV radiation may trigger several problems at the public health level. Indeed, reductions of up to 1% in total ozone content in southern Brazil cause an average 1.2% increase in surface ultraviolet radiation (Guarnieri et al. 2004) and, according to UNEP (United
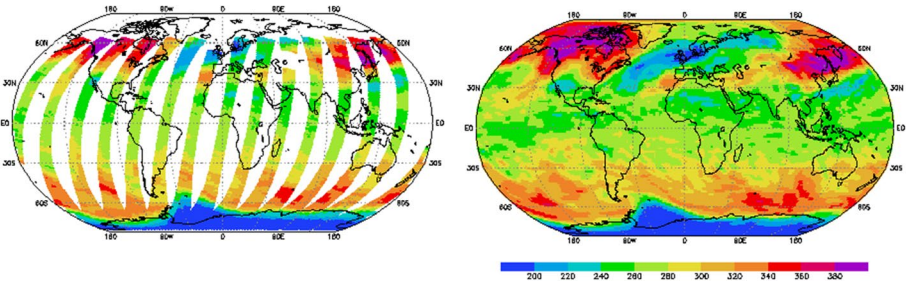
**Fig. 2** A collection of 24 h of GOME data on 30 Nov 1999, and the correspondent assimilated ozone field at 12 GMT [13]

Nations Environment Program[1]), the reduction of 10% of the stratospheric Ozone would cause additional 300,000 cases of carcinoma (malignant skin tumors) and 4500 cases of melanoma (skin cancer) each year, worldwide. Even if OSE hardly last more than 5–7 days, the increase of UV radiation is often in the order of 10% and may happen overnight, surprising individuals that are more exposed or require additional protection. OSE also may have undesirable effects on the flora and fauna, with notable risks for agriculture (Krupa and Jäger 1996) and biodiversity with, for example, the decline in amphibian species due to genetic malformations caused by increased UV radiation levels (Londero et al. 2019).

Unlike other regions of Brazil, the weather conditions in southern Brazil are strongly influenced by transient meteorological systems (Reboita et al. 2010). Examples of such systems are cold and hot fronts, which carry strong westerly winds at high tropospheric levels. Moreover, the upper troposphere–lower stratosphere (UT–LS) region in southern Brazil seems to be the home of many dynamical processes, such as stratosphere-troposphere exchanges and isentropic transport between the tropical stratosphere reservoir, polar vortex, and middle latitude. Indeed, understanding the patterns of the UT–LS is important in understanding transport and exchange processes and the links with tropospheric meteorology (Ohring et al. 2010).

## 2.2 Stratospheric $O_3$ forecast with numerical models

Traditional models for $O_3$ or other atmospheric constituents are often based on the combination of satellite/ground observations and numerical weather prediction (NWP) models. Indeed, the former is used to set demand initial and boundary conditions, while the latter relies on equations aiming to represent different physical phenomena. This integration (also known as *assimilation*) involves complex schemes and parameter tuning. Parameters are chosen to best reflect a physical (and chemical) understanding of the characteristics of the studied phenomenon, often requiring large computational systems to solve the forecast models (Godin-Beekmann 2010).

Nowadays, sophisticated data assimilation schemes for weather forecasts are used in models such as NCEP-CFSv2 (Kistler et al. 2001) and ERA5 (Malardel et al. 2016). In the case of $O_3$, assimilation is also a way to provide a "complete" view of the globe despite the less frequent coverage from the satellites, as illustrated in Fig. 2.

---

[1] https://www.unep.org/.

Eskes et al. (2002) present one of the firsts results from ECMWF traditional NWP data assimilation system, using a three-dimensional ozone advection and near real-time $O_3$ satellite observations as input. They showed that it was possible to forecast $O_3$ concentrations for 6 days in advance in the extratropics and just for 2 days in the tropical region, opening the possibility to the use of NWP to forecast the South Pole Ozone hole dynamics in a 4–5 days time range.

The tropospheric chemistry model in the integrated forecast system of the European Center for Medium-Range Weather Forecasts (ECMWF) includes a stratospheric chemistry module (Huijnen et al. 2016) to improve stratospheric composition compared with the old chemistry module. The stratospheric $O_3$ partial columns (10–100 hPa) show biases smaller than ±20DU when compared to the Aura MSL observations (Errera et al. 2019) and keeps the performance event the Antarctic hole season. However, Davis et al. (2017) evaluates water vapor and $O_3$ data coming from different reanalysis models (ERA-40 and ERA-Interim from Europe, JRA-25 and JRA-55 from Japan, CFSR, MERRA and MERRA-2 from USA), concluding that the handling of ozone varies substantially among reanalyses. While the comparison is not simple, they were able to reproduce the Ozone Total Column (TCO) ~10DU (3%) relatively to the observations. In the high stratosphere, the bias is ±20% of the $O_3$ total column, while in the upper troposphere and lower stratosphere biases increase to ±50% of the $O_3$ total column. Davis et al. (2017) also observes that the use of reanalysis ozone for Antarctic ozone hole studies is problematic, producing reasonable maps when satellite observations are available, and highly biased maps when observations are unavailable.

Despite the high potential biases for $O_3$ estimation, several web sites propose Ozone ($O_3$) global coverage information, including NASA OzoneWatch[2] and the Tropospheric Emission Monitoring Internet Service (TEMIS) from ESA[3]. OzoneWatch relies on the MERRA-2 assimilation model (Gelaro et al. 2017) and the GEOS model, but it does not provide $O_3$ forecasts on the website. TEMIS relies on the TM3-DAM model (Eskes et al. 2002; Elbern et al. 2015), and its website also proposes an 8-days forecast generated from the assimilation model.

Assimilation is therfore a computing intensive activity, requiring not only the integration of satellite (and ground/sounding balloons) data but also the execution of numerical models for both physical (transport) and chemical parameters. While precise details about the operation (execution time, number and type of nodes) of TEMIS or OzoneWatch are not available, the assimilation workflow described in Eskes et al. (2002) let us infer an important computational demand:

> Every day two forecast runs are performed. Directly after completion of the 10-day ECMWF forecast (started at 12:00 UTC) the meteorological fields are extracted from the archive. The wind fields are converted into mass fluxes in a preprocessing step, and the data is sent to KNMI. Upon arrival an analysis and forecast run is started at KNMI, based on the latest near-real time GOME ozone data. Twelve hours later a new forecast run is performed, based on the same meteorological fields, but with an additional 12h of GOME measurements. (Eskes et al. 2002)

Besides global models, regional models are also an alternative for numerical models forecast. WRF (Skamarock et al. 2019) is a well-known simulation tool, which includes a

---

[2] https://ozonewatch.gsfc.nasa.gov/.

[3] http://temis.nl/protocols/o3hole/.

chemical model plugin, WRF-Chem. WRF-Chem can be used to represent $O_3$ transport (Thomas et al. 2019, 40), but is essentially a tropospheric model that relies on fixed values derived from climatology standards to represent the $O_3$ concentration in higher altitudes.

## 2.3  Stratospheric $O_3$ forecast with deep learning

Contrarily to traditional assimilation models, Artificial Intelligence models do not rely on numerical models but try to extract patterns from raw input data, using sophisticate statistical methods. This approach also has the advantage of being computing-intensive only during the "train" phase, where the patterns are extracted. Once consolidated, AI models are often fast to deploy, producing forecasts in a small amount of time, even with reduced computing resources. The recent breakthroughs in hardware and programming methods for artificial intelligence encouraged the development of machine learning strategies to complement (if not to concurrence) existing numerical models.

As stated by Shi et al. (2015), recent advances in deep learning such as recurrent neural network (RNN) and long short-term memory (LSTM) provide useful tools to address this problem. LSTM has been widely used to predict wind regimes (zhi Wang et al. 2017), pollution concentration (Zhang et al. 2020), weather conditions (Miao et al. 2020) or flooding risks (Ding et al. 2020) on specific zones. Recently, Mbatha and Bencherif (2020) developed a hybrid data-driven forecasting model, based on LSTM applied to the total column of ozone time-series recorded at Buenos-Aires, Argentina from 1966 to 2017. However, pure LSTM is not enough to deal with multidimensional data like the spatial distribution of air masses.

Predicting the shape and movement of air masses (and other meteorological phenomena) is a real challenge for predictive learning. Contrarily to time-series forecasting or object trajectory tracking problems, the speed of air masses are not regular, their trajectories are not always periodical, and their shapes may accumulate, dissipate or change rapidly due to the complex atmospheric environment. Hence, modeling spatial deformation is significant for the prediction of this data.

These technical issues may be addressed by viewing the problem from the machine learning perspective. In essence, the meteorological forecast is a spatiotemporal sequence problem where a set of past maps (radar, satellite) are used as input, and a sequence of maps is produced as output. However, due to the complex structure of atmospheric maps, capturing the spatiotemporal structure of the data would be a hard task with traditional machine learning techniques, especially those based on supervised learning.

For this reason, Shi et al. (2015) formulated a spatiotemporal sequence forecasting problem and proposed the Convolutional Long Short-Term Memory (ConvLSTM) model, which extends the LSTM (Hochreiter and Schmidhuber 1997) to tackle problems such as the precipitation nowcast problem by using radar echo sequences for model training. The ConvLSTM model was further developed in Shi et al. (2017), where GRU (Gated Recurrent Units) are used instead of LSTMs.

In Shi et al. (2015), the radar echo maps are first transformed to grayscale images before being fed to the prediction algorithm. Thus, precipitation forecast can be considered as a type of video prediction problem with a fixed "camera", which is the weather radar. For this reason, methods proposed for predicting future frames in videos are also applicable to weather forecast. Ranzato et al. (2014) proposed the first RNN based model for video prediction, which uses a convolutional RNN to encode the observed frames. Srivastava et al. (2015) proposed the use of a LSTM encoder-decoder network to predict multiple

frames ahead; this model was generalized in Shi et al. (2015) by replacing the fully connected LSTM with ConvLSTM to better represent spatiotemporal correlations. Also Finn et al. (2016) and Jia et al. (2016) extended the ConvLSTM model by making the network predict the input frame instead of the raw pixels. More recent works try to increase the recurrence depth of the networks, all while improving spatial correlations and short-term dynamics. Among such works, we can cite the VPN—Video Pixel Network (Kalchbrenner et al. 2017), the PredRNN, PredRNN++, and Eidetic 3D LSTM models from Wang et al. (2017), Wang et al. (20185), Wang et al. (2019), as well as the Cubic LSTM model (Fan et al. 2019).

Some works try to break free from LSTM, and we can cite (Agrawal et al. 2019), which relies on the U-Net CNN model and (Villegas et al. 2017), which proposed to use both an RNN that captures the motion and a CNN that captures the content to generate the prediction. Along with RNN based models, 2D and 3D CNN based models are also present in the literature (Mathieu et al. 2016; Vondrick et al. 2016). A recent example is the work from Zheng et al. (2020), who create a model based on cascading CNN layers to forecast sea surface temperatures based on satellite data.

Please note that while precipitation nowcast is one of the major target subjects from the literature, other applications may also profit from the same neural networks, as the prediction of air pollution dissemination (Fan et al. 2017), the transportation of volcanic plumes (du Preez et al. 2020) or the estimation of sea surface temperature (Jonnakuti et al. 2020). These models can also complement existing NWP models (Wiegerinck et al. 2019) or, in our case, be applied to forecast the stratospheric ozone circulation.

## 3 Implementing an OSE forecast model

### 3.1 Model description and configuration

For this work, we chose to rely on the PredRNN++ model (Wang et al. 20185). PredRNN++ is a recurrent network model for video predictive learning. Its predecessor, PredRNN, was already used in a precipitation nowcast scenario (Wang et al. 2017), so we adopted PredRNN++ as its performance and stability fits our needs for meteorological forecasting. Also, the source code for the model is freely available at Github (https://github.com/Yunbo426/predrnn-pp), which allows us to concentrate on data preparation and parameter tuning.

PredRNN++ is structured as a network of CausalLSTM and Gradient Highway Unit (GHU) modules. The first ones work as a cascaded mechanism, where the spatial memory is a function of the temporal memory structures, while the GHU organizes the LSTM architecture, preventing the gradients of the objective function from vanishing during back-propagation. Details of the internal organization of PredRNN++ can be seen in Fig. 3.

The PredRNN++ model we used consists of two layers with 128 hidden states each. The convolution filters are set to $1 \times 1$, in order to avoid losing information. For training, the dataset was sliced in consecutive images with a 9-days sliding window. Hence, each sequence consists of 36 frames, 20 for the input (5 days), and 16 for forecasting (4 days).

The choice of the hyperparameters (number of layers and hidden states, filter grids) was deduced empirically from several parameter combinations during a prototyping phase (for example, we considered 64–64, 64–64–64, 128–128, 128–128–128, and 128–64 layers, as well as 1, 2, 4, 16, 64 filters). For instance, adding more layers tended to smooth the
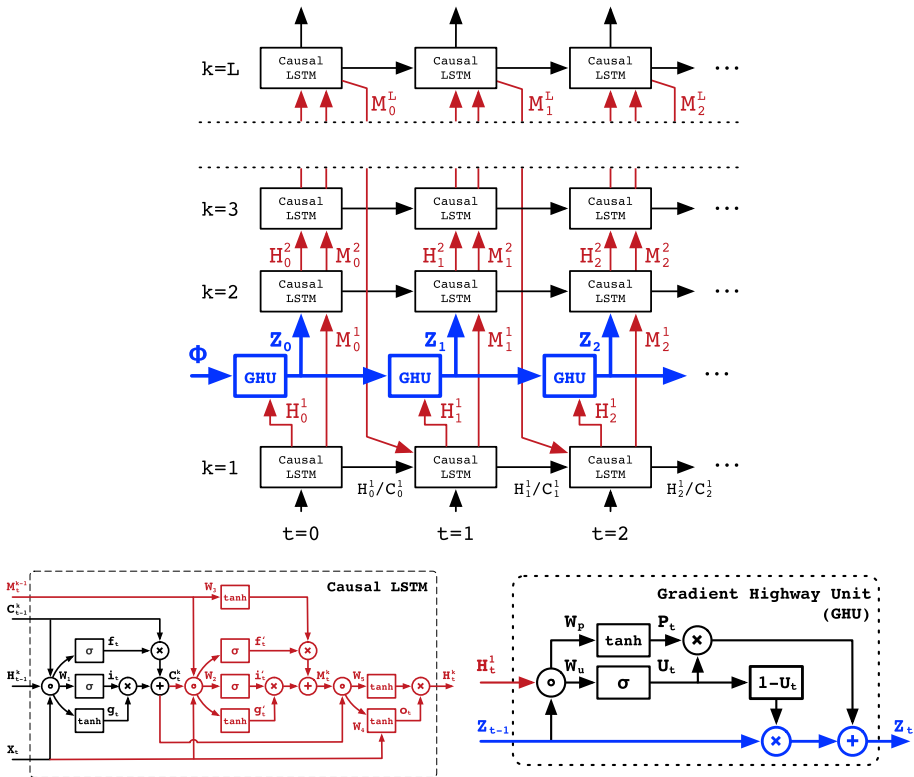
**Fig. 3** PredRNN++ model scheme, associating CausalLSTM and gradient highway units (GHU) (Wang et al. 2018[5])

resulting images, losing resolution on long-term forecasts. Similarly, using 64 hidden states accelerates the training phase but result in less sharp structures, and more than $1 \times 1$ filter grids result in "pixelated" images that don't fit our requirements.
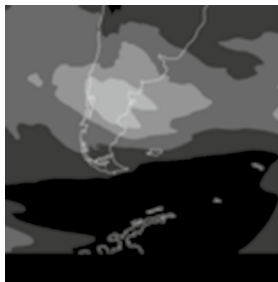
## 3.2 Data preparation and training

We first collect the Total Column Ozone ($TCO_3$) dataset from ERA5 reanalysis[4], which covers the Earth on a 30 km grid. This parameter is the total amount of ozone in a column of air extending from the surface of the Earth to the top of the atmosphere. The ERA5 units for total ozone are kilograms per square meter ($kg/m^2$). Another common unit for Ozone total column is the Dobson Unit (DU), where $1\,DU = 2.1415 \times 10^{-5} kg/m^2$.
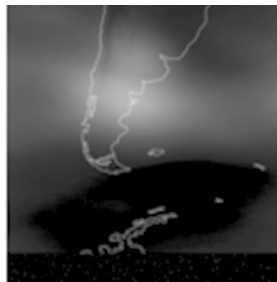
For this work, we delimited the study on the area between $70°S, 100°W$ and $20°S, 30°W$, which includes to the southernmost part of South America and parts of Antarctica. This area allows a good coverage for the air masses coming from the polar vortex, who eventually reach the Southern Spatial Observatory (SSO), located at $29.443752°$ S, $53.823084°$ W (Fig. 4).
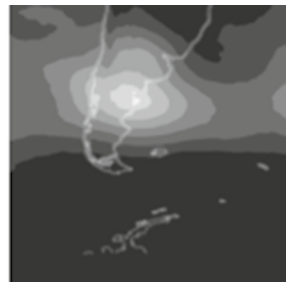
---

[4] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels.

**Fig. 4** Delimited area for $O_3$ forecasting





**(a)** Ground truth        **(b)** prediction output        **(c)** layered output

**Fig. 5** Example of input (**a**) and output (**b**, **c**) images

The dataset used in this experiment consists of about 58500 observations for each coordinate, covering the period from 1980 to 2019 and recorded every 6 h. It was split into two sets: the train dataset, which corresponds to the period between 1980 and 2009, and a test dataset for the period 2010–2019.

As the PredRNN++ model was conceived to accept images as input, we decided to automatize the generation of TCO$_3$ maps using the GrADS application. GrADS is used to generate maps of the $O_3$ concentration from ERA5 NetCDF files, converting them to pixel values which are stored as $128 \times 128$ gray-scale images 5a. As GrADS maps are presented as layered zones (i.e., discretized data), we defined 8 levels covering the usual $O_3$ concentration range in the atmosphere ($5 \times 10^{-3}$ to $8 \times 10^{-3}$ kg/m$^2$).

During the training phase, the mini-batch size is set to 16, and the training process is stopped after 300,000 iterations. The training phase required about 22 h on a node from

the ROMEO Computing Center (each node has 2× Intel® Xeon™ Gold "Skylake" 6132 - 2 × 14 cores 2.60 GHz, 96GB DDR4 RAM, and 4x NVidia Tesla P100/16GB SXM2).

While the training phase is computing intensive even in a high-end machine such as the ROMEO supercomputer, the trained model can be easily deployed and executed, even in machines without GPUs. For instance, making a forecast takes about 1 min in a CPU-only machine, with most of the time dedicated to the load of the model.

The resulting forecast is presented as a set of images with gradient values (Fig. 5b). To improve the readability and the analysis of the structure of the air masses, we also provide the output images as contour layered maps (using Matplotlib), as presented in Fig. 5c.

## 4 Results

At first, we notice how complex is the atmospheric motion: contrarily to most examples in the literature [moving MNIST Srivastava et al. 2015, KTH action Schuldt et al. 2004], atmospheric motion is variating due to interactions in the atmosphere. As a result, forecasting more than a few days is still very difficult. The length of the input dataset is also a problem, as too many frames favor large-scale events (and require a lot of memory), while too few frames don't provide enough information for the learning model. In this work, we decided to provide 20 input frames (equivalent to 5 days of measures) and generate frames for the next 4 days (16 frames).

In this section, we concentrate our analysis on a list of 24 OSE compiled by Bittencourt et al. (2019). These events occur during the Austral Spring and are tightly related to the "opening" of the South Pole Ozone hole and the breaking of the polar vortex. As a consequence, such an event may present unusual atmosphere interactions that make forecasts more complex. Although the model can be applied to any period in the year, we concentrate in these cases as the prediction and quantification of OSE events is the main goal of our project MESO[5]. For this reason, the following examples concern predictions with data up to 3 days before the events. In our understanding, this gives a reasonable margin to raise alerts to the population. Furthermore, many satellite datasets are not provided in real-time, so predictions must also take into account this delay.

In order to evaluate the accuracy of the machine learning model, we present 3 different modes of evaluations that will be developed in the next sections:

1. *Graphical similarity* analysis of graphical metrics between ground truth images and predicted ones;
2. *Frame structures* analysis of the frame structures from the meteorological point of view;
3. *Scalar predictions* capability of the framework to estimate $TCO_3$ values at a given coordinate.

### 4.1 Graphical similarity

Before conducting a domain-wise analysis, i.e., the evaluation of the quality of the forecasts from the meteorological point of view, we present a generic analysis of the results based on image similarity metrics. These metrics are often used to quantify the noise or

---

[5] https://meso.univ-reims.fr.

**Table 1** Average metric values for different OSE forecasts (3 days before). These metrics are averaged over the 16 predicted frames

| Event | MSE | MAE | PSNR | SSIM |
|---|---|---|---|---|
| **12/08/2010** | 9121.559 | 2328.866 | 15.002 | 0.479 |
| **12/09/2010** | 8084.251 | 2012.055 | 15.904 | 0.493 |
| **17/10/2010** | 8993.567 | 2101.920 | 15.152 | 0.486 |
| **26/10/2010** | 4997.680 | 1485.948 | 17.699 | 0.549 |
| **09/09/2011** | 8979.653 | 2021.981 | 15.601 | 0.525 |
| **25/10/2011** | 6479.053 | 1679.782 | 16.990 | 0.509 |
| **18/09/2012** | 7386.710 | 1789.323 | 16.288 | 0.514 |
| *26/09/2012* | **3098.222** | **1216.431** | **19.485** | **0.606** |
| **18/10/2012** | 15166.162 | 2339.554 | 13.939 | 0.483 |
| **27/10/2013** | 6004.239 | 1771.038 | 16.600 | 0.483 |
| **14/08/2014** | 10787.351 | 2310.118 | 14.548 | 0.475 |
| **26/08/2014** | 8854.087 | 2011.818 | 14.944 | 0.457 |
| **16/10/2014** | 5646.543 | 1599.670 | 17.060 | 0.552 |
| *07/11/2014* | 4912.641 | 1547.895 | 17.705 | 0.556 |
| **15/09/2015** | 9144.110 | 2167.729 | 15.242 | 0.465 |
| **26/09/2015** | 3776.599 | 1348.961 | 18.756 | 0.574 |
| **29/08/2016** | 7516.148 | 1960.775 | 15.698 | 0.483 |
| **09/09/2016** | 12524.553 | 2292.583 | 14.022 | 0.484 |
| *16/09/2016* | 13736.578 | 2740.289 | 13.179 | **0.421** |
| **24/10/2016** | 15928.443 | 2653.982 | 14.436 | **0.421** |
| *30/08/2017* | 17962.563 | **2809.668** | **12.918** | 0.478 |
| **22/09/2017** | **18180.069** | 2613.564 | 13.795 | 0.485 |
| **21/09/2018** | 7220.709 | 1838.054 | 16.707 | 0.563 |
| **15/11/2018** | 7559.428 | 1948.871 | 16.135 | 0.494 |

the divergence between images and can be applied to our results in order to verify how close are the predictions concerning validation data.

In Table 1 we list the per-frame mean square error (MSE) and the mean absolute error (MAE), the peak noise-signal ratio (PSNR), and the structural similarity index measure (SSIM) (Wang et al., 2004). In the case of MSE, MAE, low values indicate less difference or "noise" with respect to the reference image. In the case of PSNR and SSIM, higher values are better.

For each metric, Table 1 emphasizes the maximum and minimum values. From this list, we selected some OSE for further analysis (shaded cells in Table 1). The selected events include most of the maximum and minimum values cited above (or are sufficiently close to these extremes), as well as an average case.

To better evaluate these events, Fig. 6 shows the frame-wise evolution of each metric. For instance, the OSE cases from September 2016 and August 2017 shows a rapid degradation. The other examples are much more stable, keeping high structural similarity (SSIM) and low levels of MSE and MAE.

When generalizing the analysis to the whole validation period (2010–2019), we obtain the summary statistics from Table 2. Please note that the case from November 2014 cited above stands very close to the average.
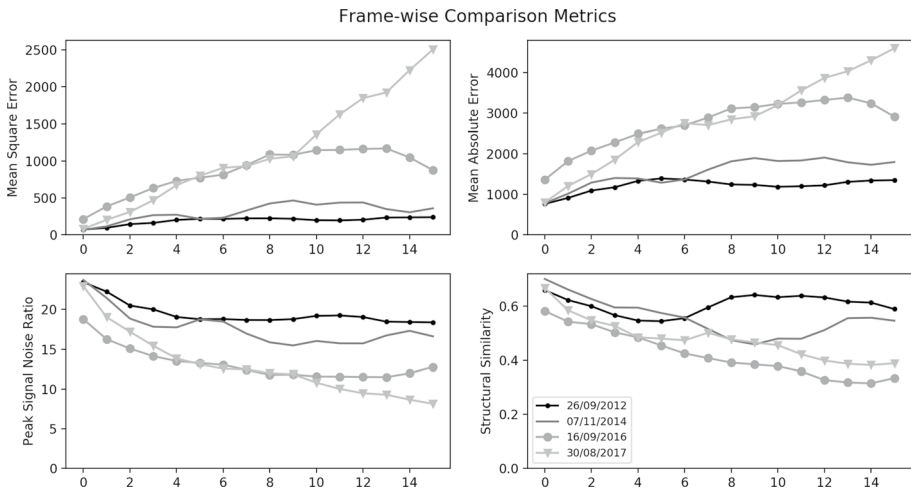
**Fig. 6** Frame-wise comparison of different predictions with metrics MSE and MAE (lower values are better), PSNR and SSIM (higher values are better)

**Table 2** Image metrics distribution from 2010–2019

|  | MSE | MAE | PSNR | SSIM |
|---|---|---|---|---|
| Mean | 6933.573966 | 1789.254265 | 17.101077 | 0.544447 |
| Std | 4111.933260 | 523.887336 | 2.419963 | 0.072979 |
| Min | 895.956512 | 641.903100 | 10.250528 | 0.339540 |
| 25% | 3707.098116 | 1360.617900 | 15.301739 | 0.488440 |
| 50% | 6176.086090 | 1757.214500 | 16.847370 | 0.536342 |
| 75% | 8984.420170 | 2138.030600 | 18.985342 | 0.595424 |
| Max | 32167.737152 | 3831.909200 | 24.816187 | 0.741482 |

In addition, Fig. 7 shows this distribution by year. The boxplots indicate that MSE and MAE have too many outliers, which may compromise the search for reference parameters for an eventual automatic quality evaluator. On the opposite side, PSNR and SSIM are more stable metrics, with more balanced quartiles that may help the setup of an automatic quality assessor.

## 4.2 Frame structural analysis

In the previous section, we compared forecasts using standard image similarity metrics. While such comparison provides useful insights into the raw efficiency of the deep learning algorithms, it doesn't include domain-specific knowledge about the observed events. For this reason, this section provides a "qualitative analysis" based on visual insights from the resulting forecasts.

Indeed, we try here to evaluate the ability of the model to provide convincing (and precise) forecasts based on a set of historical spatiotemporal data. Therefore, Figs. 8 to 11 present a visual summary from the predictions for the OSE studied in the previous section.
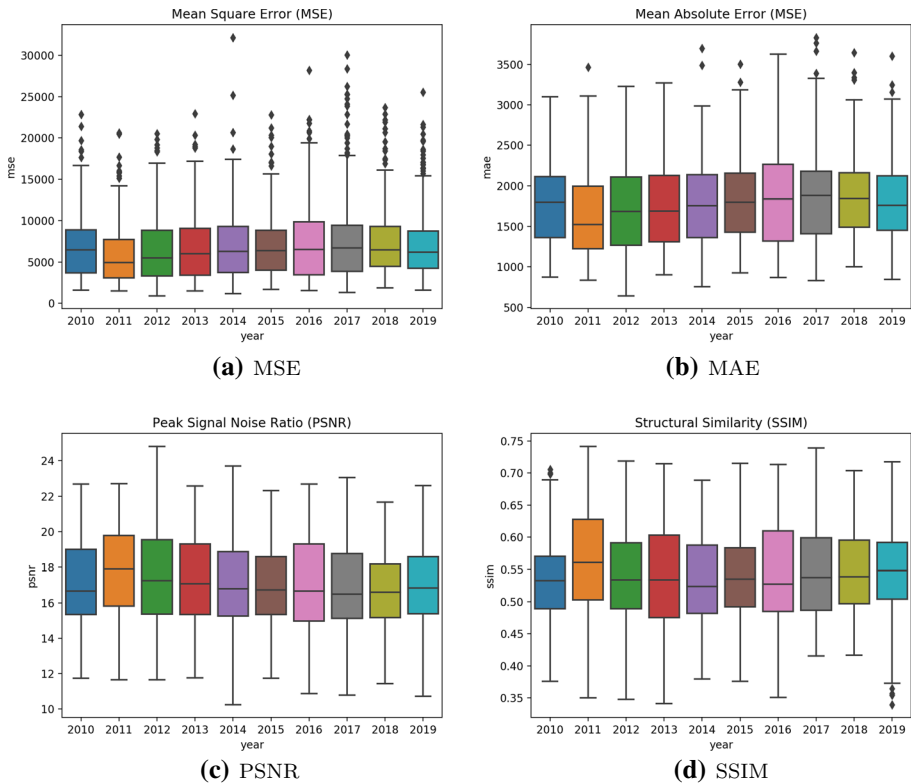
**(a)** MSE

**(b)** MAE

**(c)** PSNR

**(d)** SSIM

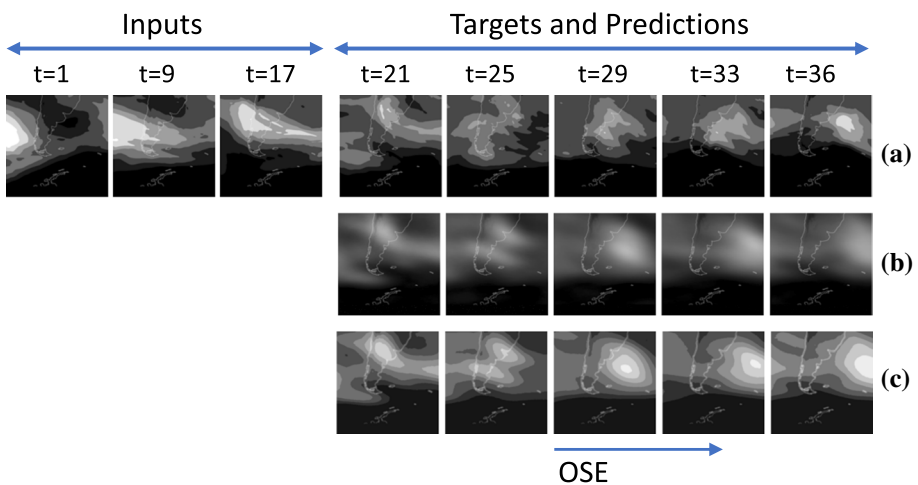**Fig. 7** Example of input (**a**) and output (**b**, **c**) images



**Fig. 8** $TCO_3$ prediction 3 days prior the September 26th, 2012 OSE. **a** Ground truth; **b** prediction output; **c** layered output
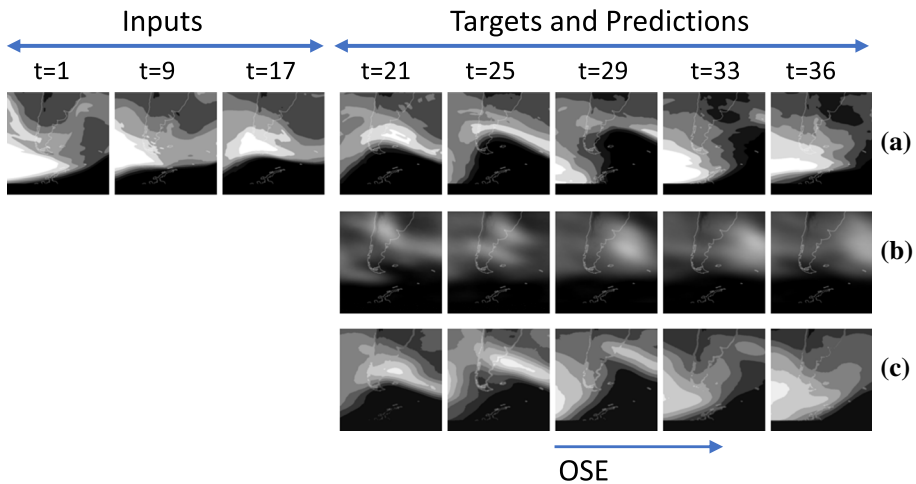
**Fig. 9** $TCO_3$ prediction 3 days prior the November 7th, 2014 OSE. **a** Ground truth; **b** prediction output; **c** layered output

The average atmospheric flow on the Southern Hemisphere mid-latitudes goes from West to East, or also called westerly winds. In Fig. 8, $t = 1$, it is possible to observe that an area of high $TCO_3$ values is present at windward of the Andes is embedded downstream of a trough that will soon cross the mountain chain. This high $TCO_3$ concentration air mass is related to the convergence produced by the trough and do not cross at once, but fractionally due to the resistance offered by the mountain range. In the following frames, it is possible to observe that this mass crosses towards the leeward side of the mountain and presents itself as a high concentration $TCO_3$, probably due to convergence produced by a cyclonic vortex circulation at the trough axis over Argentina and South Brazil.

As in the training phase, 20 frames from the past are used to predict 16 frames in the future. In the case from Fig. 8, which represents the OSE from September 26th, 2012, the forecast renders the air masses with enough similarities in both shape and intensity to the real data (ground truth), even when considering several frames in the future. Please remember that the deep learning model only has access to the "Inputs" part of the ground truth frames, the "Targets and Predictions" frames on the right are presented only for comparison purposes.

A similar result is presented in Fig. 9. Not only the model is able to render the $TCO_3$ concentration and the trough pattern visible at $t = 25$ and $t = 29$ but is also predicts a new ozone-rich mass arriving at $t = 36$ from the west.

Although the prediction represents valuable information to short term weather analysis, not all long-term forecasts perform that well. In some cases, the model degrades too fast or diverges after a few frames, which is mostly due to the non-linear nature of the atmospheric physical phenomena. In other cases, complex atmospheric structures such as ridges and troughs, wave trains, and vortexes are clearly present, tending to be smoothed, losing sharpness, although still presenting valuable information for weather forecasters.

We can observe this in Fig. 10, which represents $TCO_3$ predictions for November 2014. While the model provides frames with similar forms up to $t = 25$ and $t = 29$, the remainder frames keep a close $TCO_3$ concentration shape but present a faster displacement eastward relative to the ground truth, as a new front with reduced Ozone concentration comes from
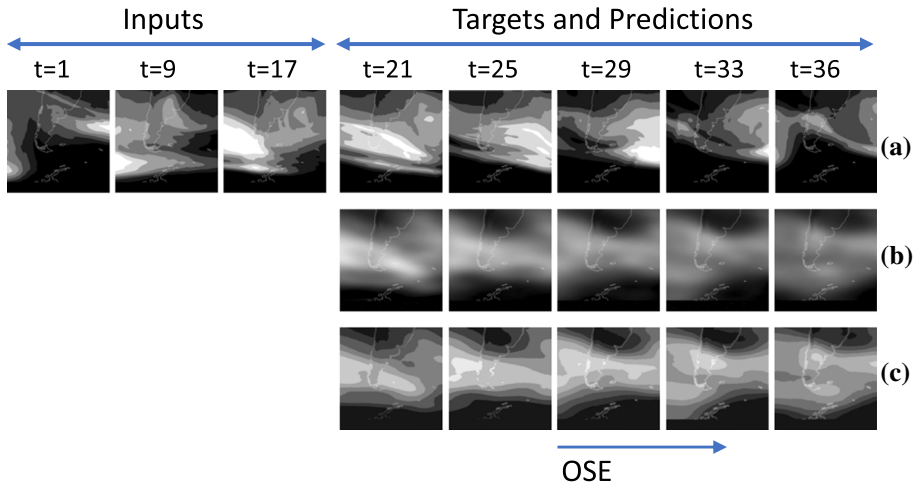
**Fig. 10** $TCO_3$ prediction 3 days prior the September 16th, 2016 OSE. **a** Ground truth; **b** prediction output; **c** layered output
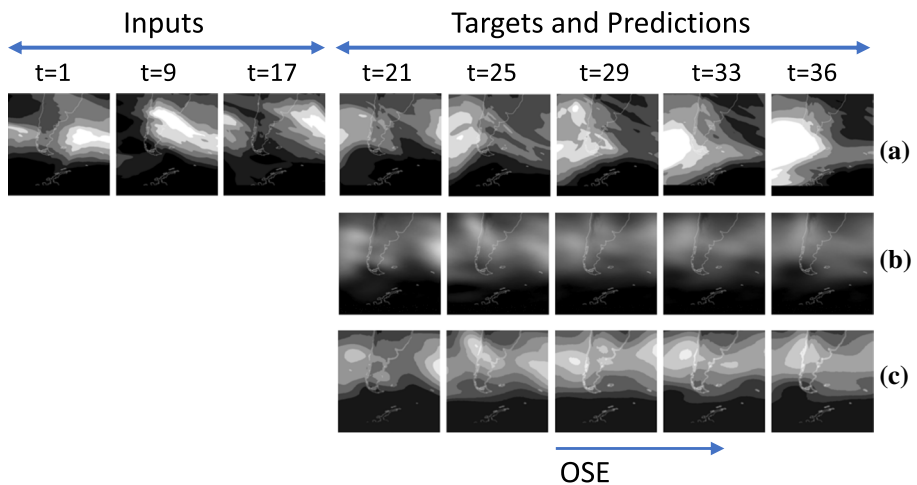


**Fig. 11** $TCO_3$ prediction 3 days prior the August 30st, 2017. **a** Ground truth; **b** prediction output; **c** layered output

the west. Also, the input frames present filaments and complex structures that are too specific to be followed by the model, at least in the current level of training.

The case of Fig. 11 is even more evident, with the model being unable to represent the evolution of the air masses. The succession of a low $TCO_3$ air mass in the south followed by a high $TCO_3$ from the west confused the model, which finally tends towards "average" values that minimize the overall distance.

From this analysis, we can affirm that the framework provides accurate frames for many OSE, but there are still some situations that require a model improvement. Nevertheless, the qualitative information provided may be a reliable resource to weather
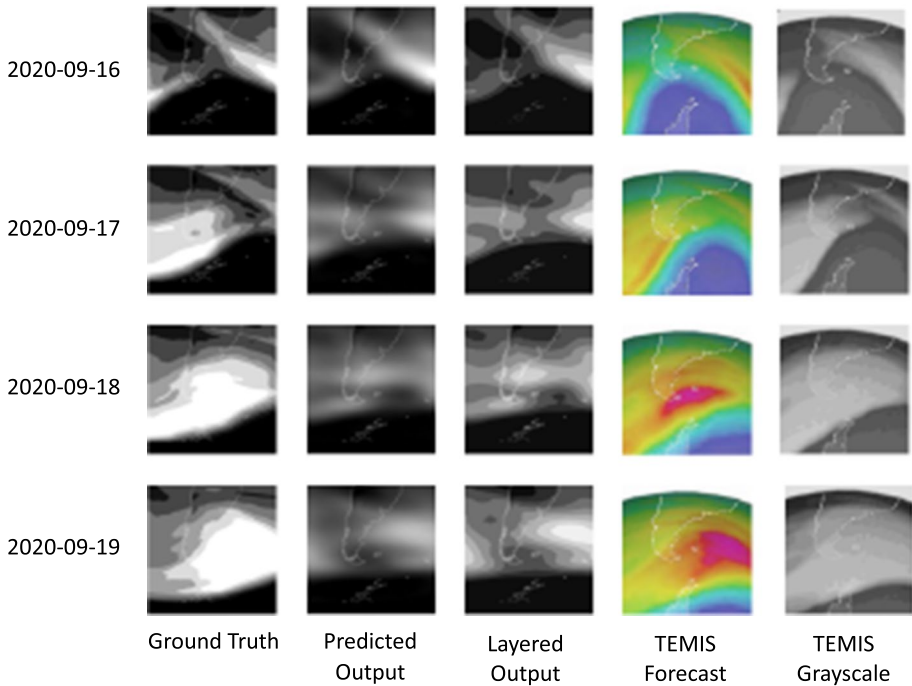
**Fig. 12** TCO$_3$ predictions from September 16,2020. Ground truth, our forecasts and TEMIS forecasts (in color and grayscale)

analysis operational services. Among the strategies to improve the model (also discussed later, in the Sect. 5), we can cite additional training rounds, larger coverage zones, data enrichment with other parameters besides TCO$_3$ and perhaps the use of GANs (Generative adversarial networks) (Goodfellow et al. 2014).

Besides the comparison with ground truth (i.e., the observed satellite images), we also tried to compare our predictions against forecasts from a numerical model in order to establish a comparison baseline with existing forecast methods.

This task proved to be hard for several reasons. First, only TEMIS publishes forecasts on its web site, but these forecasts are not archived for further analysis. Hence, we could not access archived data for the same events presented above. Nonetheless, we started a collecting campaign, storing daily forecast since the beginning of September 2020.

We selected a scenario observed from September 16th, 2020 and, using PredRNN++, we produced forecasts for the same period. As a result, Fig. 12 presents the resulting forecasts side by side, together with the satellite ground truth. The main elements from the satellite images are present in all forecasts. First, the OSE ribbon is seen crossing Argentina on September 16, dissipating the next day (this is better seen on TEMIS in grayscale). The arrival of an $O_3$ air mass is better predicted by the global model (TEMIS) than by our model based only on regional images, but the subsequent concentration of $O_3$ at the Atlantic coast on September 19th is found in both TEMIS and PredRNN++ outputs.

While this single example does not constitute a comparison baseline, it emphasizes the potential of our approach. In truth, OSE forecasting (and stratospheric $O_3$,
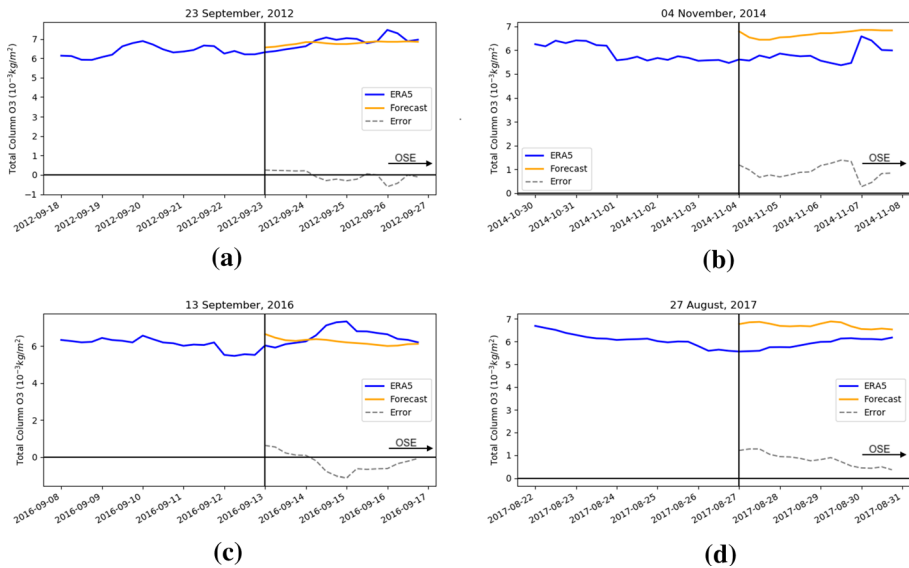
**Fig. 13** Comparison between predicted TCO$_3$ values and raw data from ERA5 for the location of the Southern Space Observatory—SSO (29.443752°$S$,53.823084°$W$)

in general) is an evolving target as both $O_3$ production, destruction, and transport are influenced by external factors from both human and natural sources. Historical averages can't show the trends of $O_3$ over the years, neither explains the increasing number of OSE in recent years. However, identifying recurrent atmospheric transport patterns, something that Artificial Intelligence models excel at, may help to predict future events.

### 4.3 Scalar prediction analysis

This last part is dedicated to the evaluation of the scalar precision of the forecast model. While the previous sections demonstrate that a good compromise can be achieved when regarding the structural similarity between the forecast frames and the observed data, it is also important to evaluate the accuracy of the predicted $0_3$ column values.

To perform this evaluation, we compare the TCO$_3$ raw values from ERA5 over one interest location—the Southern Space Observatory in Brazil—and compared with data extracted from the forecast frames. This last step is achieved by reversing the input process, which translates the input matrix from ERA5. For recall, TCO$_3$ data is normalized and transformed into layered grayscale images by the GrADS software, then feed to the deep learning algorithm. The output is presented as gradient values, that once denormalized give values in similar ranges from ERA5.

Taking again the same four OSE examples from the previous sections, we obtain the results presented in Fig. 13. Here, the blue line represents the daily measures from ERA5, the vertical line indicates the beginning of the forecast and the orange line represents the predicted values. As before, we start predicting 3 days before the registered OSE (the small arrow in the lower right). In the case of the September 20th, 2012 event (Fig. 13a), predictions fit really well the observed data. The remainder cases (Fig. 13b–d) show a relative

error in the $TCO_3$ estimation but in most cases (except Fig. 13b), the forecasts tend towards the same values levels as the observed data.

We believe that the differences are mostly caused by the use of layered images from GrADS. Indeed, values are discretized in a limited number of layers, which introduce bias from the real values and fool the prediction algorithm. Using plain normalized data from ERA5 may help improve the precision of the forecasts.

## 5 Discussion and future works

The analyses from the previous section allow us to identify several strategies to improve the quality of the forecasts. Indeed, predictions are not only expected to keep a high similarity level of the structures (air masses) over time but also to allow the estimation of TCO $_3$ over a locality with a reduced error.

Finding a good combination of input data and hyperparameters is the first step to do. The experiments presented here consider 20 frames (5 days) input dataset, which is a minimum threshold to capture the atmospheric circulation patterns. Adding extra input frames can help the model to understand additional interactions, at the expense of more memory. Also, as discussed earlier, this may reinforce the weight of large scale events, at the expense of rare or episodic patterns. Extra training epochs may help improve this resolution, but we must be extremely attentive to overfitting during the training phase.

Since we perform predictions on independent geographical tiles (i.e., a regional coverage), border effects are also a problem. When air masses originate outside the boundaries of a tile, the model can only guess their arrival based on recurrent historical patterns. Figure 11 shows an instance of this where the model is unable to predict a mass rich in Ozone coming from the west. By enlarging the tile towards the west would include more information about air masses coming to a given interest point. Another possible approach is to provide images containing a South Pole stereo-polar view, where an circular continuous flow can be observed, giving therefore a synoptic observational field to the PredRNN++ training system.

The analysis of graphical similarity metrics also shows that Structural Similarity (SSIM) gives much more valuable information about the quality of the forecasts. Adapting SSIM as a loss function on PredRNN++ may improve the accuracy of the model, especially for long-term forecasts.

Improvements may also be made concerning the input data quality. As explained before, the current framework uses data transformed by the GrADS application, which creates layered maps with flatten values categories. While this seems to accelerate the learning process, it induces a bias on the output values for a given location.

Data quality can also be enriched through the addition of other parameters of interest. While the current experiment uses only $TCO_3$, at least two other atmospheric components are known for their correlation with the Ozone circulation: the temperature and the potential vorticity (PV). Instead of a simple channel in a grayscale image, these other values could be added as additional image channels, in a similar way as RGB images can be handled.

Another direction that may be explored to refine the topological structure and hyperparameters of the neural network is the use of Generative Adversarial Networks (GANs) (Goodfellow et al. 2014), which have shown excellent results in problems in which the output have quality requirements to match.

Despite such improvements, we strongly believe that AI-based methods are not intended to replace existing models but to complement them. Forecasts generated by deep-learning methods may represent an asset for traditional assimilation methods, highlighting patterns that are hard to identify and model otherwise. Besides, deep-learning models offer a fast and non-expensive way to forecast scalar advection, providing first-hand elements for decision making while numerical-based models forecasts are not yet available.

Our team also plans to expand the usage of PredRNN++ and future algorithms towards other usage cases. Indeed, scalar advection can be applied to other subjects, if properly trained. Wildfire smoke and volcanic plums forecasting are among the subjects currently studied by our laboratories, and we believe that the algorithms used in this paper may be easily transposed to solve these problems.

## 6 Conclusion

In this paper, we explore the usage of deep learning to handle stratospheric Ozone ($O_3$) spatiotemporal forecast. The motivation is that the Ozone layer lies in a zone in the middle atmosphere (stratosphere) with peculiar characteristics that are prone to deep learning extrapolation methods, instead of traditional numerical simulation methods. Indeed, the stratosphere is a barotropic portion of the atmosphere whose scalar fields advection (heat, gases, and particulate materials) are constricted between in isentropic layers, in an almost horizontal displacement. Data comes essentially satellite observations, which can easily be handled as matrix or images.

We focused our study on the forecast of Ozone Secondary Events (OSE), episodes of exchange between the polar vortex and the mid-latitudes and the tropics in which the polar vortex is deformed and ozone-poor polar air masses move towards mid-latitudes, resulting in drastic increases in the UV radiation. We aim to provide prediction tools to early identify such events.

In this paper, we leveraged a spatiotemporal algorithm initially developed for video frame predictions, adapting it to accept satellite data. We evaluate the performance of the algorithm output under both deep learning metrics (image similarity, for example) and domain-specific requirements (interest of the results from the meteorological point of view), showing interesting results and directions for improvement. To our knowledge, this is one of the first attempts to forecast the circulation of the Ozone layer using deep learning.

Another interest of deep learning-based forecasts is the amount of computing resources they require. If the training phase of a model may be long and computing-intensive, performing a prediction is very fast and requires almost no computing resources. By comparison, forecasts based in traditional numerical models often require dedicated infrastructures and several computing hours to process the data assimilation and the forecast models. Therefore, deep learning models allow almost immediate updates of the forecast with good short-term prediction quality, which is especially interesting for new satellite sources such as GOES-16 that produce hourly updates. If the total transition from traditional forecast methods in favor of deep learning remains an open question, the combined usage of both methods in an ensemble forecast approach is a promising choice to consider.

# References

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., & Hickey, J. (2019). Machine learning for precipitation nowcasting from radar images. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019).

Bencherif, H., Amraoui, L. E., Semane, N., Massart, S., Charyulu, D. V., Hauchecorne, A., et al. (2007). Examination of the 2002 major warming in the southern hemisphere using ground-based and ODIN/SMR assimilated data: stratospheric ozone distributions and tropic/mid-latitude exchange. *Canadian Journal of Physics*, *85*(11), 1287–1300. https://doi.org/10.1139/p07-143.

Bencherif, H., El Amraoui, L., Kirgis, G., Leclair De Bellevue, J., Hauchecorne, A., Mzé, N., Portafaix, T., Pazmino, A., & Goutail, F. (2011). Analysis of a rapid increase of stratospheric ozone during late austral summer 2008 over kerguelen (49.4° s, 70.3° e). *Atmospheric Chemistry and Physics, 11*(1), 363–373. https://doi.org/10.5194/acp-11-363-2011.

Bittencourt, G. D., Pinheiro, D. K., Bageston, J. V., Bencherif, H., Steffenel, L. A., & Vaz Peres, L. (2019). Investigation of the behavior of the atmospheric dynamics during occurrences of the ozone hole's secondary effect in southern brazil. *Annales Geophysicae*, *37*(6), 1049–1061. https://doi.org/10.5194/angeo-37-1049-2019.

Brewer, A. (1949). Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, *75*(326), 351–363.

Canziani, P. O., Compagnucci, R. H., Bischoff, S. A., & Legnani, W. E. (2002). A study of impacts of tropospheric synoptic processes on the genesis and evolution of extreme total ozone anomalies over southern south america. *Journal of Geophysical Research: Atmospheres*, *107*(D24), ACL 2–1–ACL 2–25. https://doi.org/10.1029/2001JD000965.

Casiccia, C., Zamorano, F., & Hernandez, A. (2008). Erythemal irradiance at the magellan's region and antarctic ozone hole 1999–2005. *Atmosfera*, *21*(1), 2–12.

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., et al. (2017). Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmospheric Chemistry and Physics*, *17*(20), 12743–12778. https://doi.org/10.5194/acp-17-12743-2017.

du Preez, D. J., Bencherif, H., Bègue, N., Clarisse, L., Hoffman, R., & Wright, C. (2020). Investigating the large-scale transport of a volcanic plume and the impact on a secondary site. *Atmosphere*, *11*, https://doi.org/10.3390/atmos11050548.

Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*, *403*, 348–359. https://doi.org/10.1016/j.neucom.2020.04.110.

Dobson, G. M. B. (1968). Forty years' research on atmospheric ozone at oxford: A history. *Applied Optics*, *7*(3), 387–405. https://doi.org/10.1364/AO.7.000387.

Elbern, H., Agusti-Panareda, A., Benedetti, A., et al. (2015). Assimilation of satellite data for atmospheric composition. In *Seminar on use of satellite observations in numerical weather prediction, 8–12 September 2014. Shinfield Park, Reading*. https://www.ecmwf.int/node/9274.

Errera, Q., Chabrillat, S., Christophe, Y., Debosscher, J., Hubert, D., Lahoz, W., et al. (2019). Technical note: Reanalysis of aura MLS chemical observations. *Atmospheric Chemistry and Physics*, *19*(21), 13647–13679. https://doi.org/10.5194/acp-19-13647-2019.

Eskes, H. (2009). TM3-DAM: Assimilated ozone fields based on GOME data. http://www.temis.nl/gofap/tm3doc/tm3dam.html.

Eskes, H. J., van Velthoven, P. F. J., & Kelder, H. M. (2002). Global ozone forecasting based on ERS-2 gome observations. *Atmospheric Chemistry and Physics*, *2*(4), 271–278. https://doi.org/10.5194/acp-2-271-2002.

Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., & Lin, S. (2017). A spatiotemporal prediction framework for air pollution based on deep RNN. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 15.

Fan, H., Zhu, L., & Yang, Y. (2019). Cubic LSTMS for video prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 8263–8270.

Finn, C., Goodfellow, I., & Levine, S. (2016). Unsupervised learning for physical interaction through video prediction. In *Advances in neural information processing systems* (pp. 64–72).

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., & Birner, T. (2011). The extratropical upper troposphere and lower stratosphere. *Reviews of Geophysics,*. https://doi.org/10.1029/2011R G000355.

Godin-Beekmann, S. (2010). Spatial observation of the ozone layer. *Comptes Rendus Geoscience*, *342*(4), 339–348. https://doi.org/10.1016/j.crte.2009.10.012.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Proceedings of the 27th international conference on neural information processing systems (NIPS'14)* (Vol. 2, pp. 2672–2680). Cambridge, MA, USA: MIT Press.

Guarnieri, R., Padilha, L., Guarnieri, F., Echer, E., Makita, K., Pinheiro, D., et al. (2004). A study of the anticorrelations between ozone and UV-B radiation using linear and exponential fits in southern brazil. *Advances in Space Research*, *34*(4), 764–76.

Hauchecorne, A., Godin, S., Marchand, M., Heese, B., & Souprayen, C. (2002). Quantification of the transport of chemical constituents from the polar vortex to midlatitudes in the lower stratosphere using the high-resolution advection model mimosa and effective diffusivity. *Journal of Geophysical Research: Atmospheres*, *107*(D20), SOL 32–1–SOL 32–13. https://doi.org/10.1029/2001JD000491.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hofmann, D., Oltmans, S., Harris, J., Johnson, B., & Lathrop, J. (1997). Ten years of ozonesonde measurements at the south pole: Implications for recovery of springtime antarctic ozone. *Journal of Geophysical Research: Atmospheres*, *102*(D7), 8931–8943.

Huijnen, V., Flemming, J., Chabrillat, S., Errera, Q., Christophe, Y., Blechschmidt, A. M., et al. (2016). C-ifs-cb05-bascoe: Stratospheric chemistry in the integrated forecasting system of ECMWF. *Geoscientific Model Development*, *9*(9), 3071–3091. https://doi.org/10.5194/gmd-9-3071-2016.

Jia, X., De Brabandere, B., Tuytelaars, T., & Gool, L. V. (2016). Dynamic filter networks. In *Advances in neural information processing systems* (pp. 667–675).

Jonnakuti, P. K., & Tata Venkata Sai, U. B. (2020). A hybrid cnn-lstm based model for the prediction of sea surface temperature using time-series satellite data. In *EGU general assembly 2020*. https://doi.org/10.5194/egusphere-egu2020-817.

Kalchbrenner, N., van den Oord, A., Simonyan, K., Danihelka, I., Vinyals, O., Graves, A., & Kavukcuoglu, K. (2017). Video pixel networks. In D. Precup, & Y. W. Teh (Eds.,) textitProceedings of the 34th international conference on machine learning Proceedings of machine learning research (Vol. 70, pp. 1771–1779). Sydney, Australia: PMLR, International Convention Centre.

Kirchhoff, V. W. J. H., Schuch, N., Pinheiro, D., & Harris, J. (1996). Evidence for an ozone hole perturbation at 30° south. *Athmospheric Environment*, *33*(9), 1481–1488.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., et al. (2001). The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, *82*(2), 247–268. .

Krupa, S. V., & Jäger, H. J. (1996). Adverse effects of elevated levels of ultraviolet (UV)-B radiation and ozone (O3) on crop growth and productivity. In W. S. F. Bazzaz (Ed.), *Global Climate Change and Agricultural Production* (pp. 141–169). Chichester: Wiley.

Londero, J.E.L., dos Santos, M.B., & Schuch, A.P. (2019). Impact of solar uv radiation on amphibians: Focus on genotoxic stress. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* **842**, 14 – 21. https://doi.org/10.1016/j.mrgentox.2019.03.003. Detection of Genotoxins in Aquatic and Terrestric Ecosystems.

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kuehnlein, C., Mozdzynski, G., Hamrud, M., & Smolarkiewicz, P. (2016). A new grid for the ifs. *ECMWF newsletter* (pp. 23–28). https://doi.org/10.21957/zwdu9u5i.

Marchand, M., Bekki, S., Pazmino, A., Lefèvre, F., Godin-Beekmann, S., & Hauchecorne, A. (2005). Model simulations of the impact of the 2002 antarctic ozone hole on the midlatitudes. *Jounal of the Atmospheric Sciences*, *62*, 871–884.

Mathieu, M., Couprie, C., & LeCun, Y. (2016). Deep multi-scale video prediction beyond mean square error. In *4th international conference on learning representations (ICLR 2016)*.

Mbatha, N., & Bencherif, H. (2020). Time series analysis and forecasting using a novel hybrid LSTM data-driven model based on empirical wavelet transform applied to total column of ozone at Buenos Aires, Argentina (1966–2017). *Atmosphere*, *11*(5), 457.

Miao, K., Han, T., Yao, Y. Q., Lu, H., Chen, P., Wang, B., et al. (2020). Application of LSTM for short term fog forecasting based on meteorological elements. *Neurocomputing,*. https://doi.org/10.1016/j.neuco m.2019.12.129.

Moradi Kordmahalleh, M., Gorji Sefidmazgi, M., & Homaifar, A. (2016). A sparse recurrent neural network for trajectory prediction of atlantic hurricanes. In *Proceedings of the genetic and evolutionary computation conference 2016 (GECCO '16)* (pp. 957–964). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2908812.2908834.

NOAA: Rap chem model fields. https://rapidrefresh.noaa.gov/RAPchem/.

Ohring, G., Bojkov, R., Bolle, H. J., Hudson, R., & Volkert, H. (2010). Radiation and ozone: Catalysts for advancing international atmospheric science programmes for over half a century. *Space Research Today*, *177*, 16–31. https://doi.org/10.1016/j.srt.2010.03.004.

Peres, L. V. (2013). Efeito Secundário do Buraco de Ozônio Antártico Sobre o Sul do Brasil. Universidade Federal de Santa Maria, Brazil.

Prabhat, B., Vishwanath, V., Dart, E., Wehner, M., & Collins, W. D. (2015). TECA: Petascale pattern recognition for climate science. In G. Azzopardi & N. Petkov (Eds.), *Computer analysis of images and patterns* (pp. 426–436). Cham: Springer.

Pradhan, R., Aygun, R. S., Maskey, M., Ramachandran, R., & Cecil, D. J. (2018). Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Transactions on Image Processing*, *27*(2), 692–702. https://doi.org/10.1109/TIP.2017.2766358.

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., & Prabhat, P. C. (2017). Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Proceedings of the 31st international conference on neural information processing systems (NIPS'17)* (pp. 3405–3416). Red Hook, NY, USA: Curran Associates Inc.

Ranzato, M., Szlam, A., Bruna, J., Mathieu, M., Collobert, R., & Chopra, S. (2014). Video (language) modeling: A baseline for generative models of natural videos. arXiv:1412.6604.

Reboita, M., Gan, M., Da Rocha, R., Ambrizzi, T., et al. (2010). Precipitation regimes in south america: A bibliography review. *Revista Brasileira de Meteorologia*, *25*(2), 185–204.

Research, G. O., & Project, M. (2014). Scientific assessment of ozone depletion: 2018. *Tech. Rep. Report No. 55*. Geneva, Switzerland: WMO (World Meteorological Organization).

Research, G.O., & Project, M. (2018). Scientific assessment of ozone depletion: 2018. *Tech. Rep. Report No. 58*. Geneva, Switzerland: WMO (World Meteorological Organization).

Sakaino, H. (2013). Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(5), 3023–3036. https://doi.org/10.1109/TGRS.2012.2212201.

Salby, M. L. (1996). Fundamentals of atmospheric physics. In *International geophysics* (Vol. 61). Academic Press. https://doi.org/10.1016/S0074-6142(96)80037-4.

Scaife, A. A., Spangehl, T., Fereday, D. R., Cubasch, U., Langematz, U., Akiyoshi, H., et al. (2012). Climate change projections and stratosphere-troposphere interaction. *Climate Dynamics*, *38*, 2089–2097. https://doi.org/10.1007/s00382-011-1080-7.

Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: A local SVM approach. In *Proceedings 17th international conference on pattern recognition (ICPR'04)* (Vol. 3, pp. 32–36). USA: IEEE Computer Society.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Proceedings of the 28th international conference on neural information processing systems (NIPS'15)* (pp. 802–810). Cambridge, MA, USA: MIT Press.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D., Wong, W., et al. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30: Annual conference on neural information processing systems 2017, 4–9 December 2017* (pp. 5617–5627). USA: Long Beach, CA.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). A description of the advanced research wrf version 4. *Tech. Rep. NCAR/TN-556+STR, NCAR*. https://doi.org/10.5065/1dfh-6p97.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., & Schmidt, A. (2016). Emergence of healing in the antarctic ozone layer. *Science*, *353*(6296), 269–274. https://doi.org/10.1126/science.aae0061.

Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using LSTMs. In *International conference on machine learning* (pp. 843–852).

Steffenel, L. A., Rasera, G., Begue, N., Damaris Kirsch-Pinheiro, D., & Bencherif, H. (2019). Spatio-temporal LSTM forecasting of ozone secondary events. In: E. G. Assembly (ed.) *Geophysical research abstracts* (Vol. 21, pp. EGU2019–8075). Vienna, Austria.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2016). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., et al. (2014). Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bulletin of the American Meteorological Society*, *95*(3), 409–426. https://doi.org/10.1175/BAMS-D-11-00263.1.

Thomas, A., Huff, A. K., Hu, X. M., & Zhang, F. (2019). Quantifying uncertainties of ground-level ozone within WRF-CHEM simulations in the mid-atlantic region of the united states as a response to variability. *Journal of Advances in Modeling Earth Systems*, *11*(4), 1100–1116. https://doi.org/10.1029/2018MS001457.

Vaz Peres, L., Bencherif, H., Mbatha, N., Passaglia Schuch, A., Toihir, A. M., Bègue, N., et al. (2017). Measurements of the total ozone column using a brewer spectrophotometer and toms and omi satellite instruments over the southern space observatory in brazil. *Annales Geophysicae*, *35*(1), 25–37. https://doi.org/10.5194/angeo-35-25-2017.

Villegas, R., Yang, J., Hong, S., Lin, X., & Lee, H. (2017). Decomposing motion and content for natural video sequence prediction. In *ICLR*.

Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Generating videos with scene dynamics. In *Advances in neural information processing systems* (pp. 613–621).

Wang, Y., Gao, Z., Long, M., Wang, J., & Yu, P. S. (2018). Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *35th international conference on machine learning*.

Wang, Y., Jiang, L., Yang, M. H., Li, L. J., Long, M., & Fei-Fei, L. (2019). Eidetic 3d LSTM: A model for video prediction and beyond. In *7th international conference on learning representations (ICLR 2019)*.

Wang, Y., Long, M., Wang, J., Gao, Z., & Yu, P. S. (2017). Predrnn: Recurrent neural networks for predictive learning using spatiotemporal LSTMS. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.,) *Advances in neural information processing systems* (Vol. 30, pp. 879–888). Curran Associates, Inc.

Wiegerinck, W., Thalmeier, D., & Selten, F. (2019). Deep learning for weather forecasting? a proof of principle. In: E.G. Assembly (Ed.), *Geophysical research abstracts* (Vol. 21, pp. EGU2019–11443). Vienna, Austria.

Zhang, B., Zhang, H., Zhao, G., & Lian, J. (2020). Constructing a pm2.5 concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environmental Modelling & Software,* **124**, 104600. https://doi.org/10.1016/j.envsoft.2019.104600.

Zheng, G., Li, X., Zhang, R. H., & Liu, B. (2020). Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Science Advances,*. https://doi.org/10.1126/sciadv.aba1482.

Zhi Wang, H., Qiang Li, G., Bin Wang, G., Chun Peng, J., Jiang, H., & Tao Liu, Y. (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, *188*, 56–70. https://doi.org/10.1016/j.apenergy.2016.11.111.