Predicting origins of coherent air mass trajectories using a neural network—the case of dry intrusions

Vered Silverman | Stav Nahum | Shira Raveh-Rubin

Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel

Correspondence
Vered Silverman, Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel.
Email: vered.silverman@weizmann.ac.il

Funding information
Israel Science Foundation, Grant/Award Number: 1347/18

Abstract
Mid-latitude cyclones are complex weather systems that are tightly related to surface weather impacts. Coherent air streams are known to be associated with such systems, in particular dry intrusions (DIs) in which dry air masses descend slantwise from the vicinity of the tropopause equatorward towards the surface. Often, DIs are associated with severe surface winds, heavy precipitation and frontogenesis. Currently, DIs can only be identified in hindsight by costly Lagrangian calculations using high resolution wind field data. Here, we use a novel method aiming to simplify the detection procedure of DI origins to allow their future identification in climate datasets, previously inaccessible for such diagnostic studies. A novel adaptation of a segmentation-oriented neural network model is hereby presented as a successful tool to identify DI origins based solely on three ERA-Interim reanalysis geopotential height fields, representing the state of the atmosphere. The model prediction skill is tested by calculating both the grid-point and DI object based Matthews correlation coefficient. We find the model highly skilful in both reconstructing accurately the climatological distribution and predicting the vast majority of the individual DI origin objects. The skill decreases for relatively small objects and for objects occurring at locations where such cases are relatively less frequent. This indicates that geopotential height variability is related to the dynamic mechanisms involved in DI initiations. The results serve as a proof of concept for predicting DIs and other coherent air mass trajectories even when high resolution wind field data are not available, such as for model output for future climate projections.

KEYWORDS
air mass, air streams, deep learning, dry intrusions, machine learning, neural networks, trajectories, weather prediction

Vered Silverman and Stav Nahum are equally contributing authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Meteorological Applications published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
1 | INTRODUCTION

The weather in the mid-latitudes is dominated by the passage of extra-tropical cyclonic systems (Bjerknes, 1922; Chang et al., 2002). The flow induced by these low-pressure systems is usually characterized by three main different air streams: the warm and cold conveyor belts, and dry air intrusions (DIs) (Carlson, 1980; Browning and Roberts, 1994). The surface impact associated with these flows is often found to be associated with extreme weather events, primarily heavy precipitation, floods and strong wind gusts, causing social and economic damage (Pfahl and Wernli, 2012; Pfahl, 2014; Ludwig et al., 2015; Raveh-Rubin and Wernli, 2016; Dowdy and Catto, 2017). Therefore, a main goal is to improve our understanding of the physical mechanisms responsible for the development of these air streams, which is key for understanding their predictability and potentially improving their forecast. Furthermore, such knowledge is important to interpret future climate projections and their influence on the magnitude and frequency of high impact weather.

To study the air streams associated with extra-tropical weather systems, it is useful to adopt a Lagrangian perspective, allowing the direct study of the dynamic and thermodynamic mechanisms governing the air mass flow evolution naturally. This Lagrangian approach has been extensively used for the study of warm conveyor belts (e.g., Wernli, 1997; Madonna et al., 2014; Pfahl et al., 2014; Binder et al., 2016) as well as of DIs (Raveh-Rubin and Wernli, 2016; Raveh-Rubin, 2017). For this, forward Lagrangian trajectories of air masses are calculated globally. The trajectories of interest are then selected based on specific criteria, often represented by the duration and vertical or horizontal extent from their source to destination. As a result, it is possible to know the timing and location at which the studied air streams will initiate only after they have reached their destination. Moreover, the Lagrangian method requires a large dataset of the wind field with high spatial (horizontal and vertical) and temporal resolution to track the air masses accurately, a condition that is often not met for the openly available global model output of future projections. It is also noteworthy that even when data are available at sufficient resolution, a systematic computation of air mass trajectories involves costly and lengthy computations. In the present study we test the ability of deep neural networks to overcome these limitations, focusing on trajectories of DIs.

DIs are dry air streams descending slantwise from the upper mid-latitude troposphere toward the surface at lower latitudes. The coherent descent of these air parcels makes it possible for them to keep their dynamic and thermodynamic characteristics along their path. As a result, these air streams can introduce strong anomalies in humidity and temperature when reaching to lower altitudes and latitudes (Raveh-Rubin, 2017; Ilotoviz et al., 2021). Indeed, the DI low level outflow is often associated with strong fronts and high impact weather events (Browning and Reynolds, 1994; Raveh-Rubin and Wernli, 2016; Catto and Raveh-Rubin, 2019; Raveh-Rubin and Catto, 2019)—hence the importance of understanding their predictability. The triggering mechanism which initiates the descent of DIs remains an open question. While the DI near surface outflows are clearly tied to steering by the lower tropospheric cyclonic and anticyclonic pressure anomalies typically positioned behind cold fronts, their origin in the upper troposphere is not yet clear. Presumably, DIs originate under specific dynamic environments in the upper troposphere, such that deciphering the environment potentially provides a path to understanding the predictability of the surface weather linked to DIs. In addition to the obvious application for improving weather forecasts, understanding the large-scale environment related to DIs will make it possible to assess changes in their occurrence on climate time scales for different climate projection scenarios. Therefore, we focus on studying DI origins (DIOs), represented as the density of the upper level portion of DI trajectories, gridded on horizontal spatial maps. By training a deep neural network, the DIOs are used by the model to identify the dynamic environment in the upper troposphere under which they occur.

Neural networks come in many different shapes and sizes. As the machine learning field advances owing to readily available big datasets, new forms of networks are being developed and many fields are adopting and adjusting those implementations for their own innovative uses. A specific type of neural network used for machine vision purposes are convolutional neural networks (CNNs). As indicated by their name, CNNs use convolutional operations on an input image to find a compact representation of the features which were found to be important during the training process to produce the desired prediction (e.g., Lawrence et al., 1997). CNNs are already used for different purposes in atmospheric science, such as satellite imagery for agriculture purposes (Li et al., 2020; Segal-Rozenhaimer et al., 2020), cloud classification (Rasp et al., 2020), estimating weather forecast uncertainty (Scher and Messori, 2018) and even weather forecast (Weyn et al., 2019). While most of these applications use CNNs to identify visible objects which can be objectively identified to the common observer, in essence they can be used to try and relate any input to a desired output, and not necessarily images. In this context, Larraondo et al. (2019) used geopotential heights as
input to CNNs, successfully reproducing fully parameterized precipitation as output.

This study aims for a novel application of CNNs, originally designed for image segmentation purposes, relating specific atmospheric fields to specific properties of air mass trajectories. A CNN is trained to predict where and when DIs will be initiated, given a relatively simple and commonly used set of atmospheric parameters: horizontal 2D maps of geopotential height. A successful prediction of the model provides multiple gains. Contrary to the limitations of the Lagrangian method which is currently used to identify DIOs, the trained model does not require high resolution data and costly computations. In addition, we gain knowledge about the most relevant parameters needed to make a good prediction, also indicating the dynamic mechanisms involved in the process. Lastly, making a forecast based on instantaneous atmospheric maps, for the first time, makes it possible to predict DIOs multiple days prior to their actual descent.

The paper is constructed as follows. A detailed description of the method is presented in Section 2. Technical details about the predictors and DI database used to train the model are provided in Section 2.1. The model architecture, training procedure and performance metrics are presented in Sections 2.2-2.4. The resulting model performance is analysed in detail in Section 3, based on both specific case studies and a climatological evaluation. Finally, suggested applications and outlook are discussed in Section 4.

## 2 | METHODS

The spatio-temporal variability of DIOs indicates certain regions and seasons of increased frequency (see Raveh-Rubin, 2017, for further spatial analysis). Each DI trajectory (Figure 1a) was traced back to its origin in the upper troposphere, thus creating a subset composed solely of Eulerian objects indicating the DIO areas (Figure 1a, black contours) for each time step (see Section 2.1, item 2, for more data processing details). Climatologically, DIOs peak during Northern Hemisphere wintertime (December, January and February) (Figure 1b) with distinct preferred locations across the hemisphere. Yet, it is unknown what the coherent and recognizable dynamic environment conducive to the development of DIO events is. Choosing and designing a neural network model that can identify meaningful details in a given
dataset has no previous preset biases and is capable of flexible learning in a modular layer-like way (Section 2.2), thus being an excellent engine for this task.

Considering that geopotential height surfaces provide a first approximation for the upper tropospheric circulation, and are commonly available outputs among various numerical weather prediction and general circulation models, they were selected as the dynamic fields for which we aim to reveal whether their variability is possibly linked to DIOs. We aim to gain useful insights on whether these specific dynamic fields hold any of the necessary information that can contribute to our understanding of DIO events.

2.1 Atmospheric data

The data used for training the model are the European Centre for Medium-Range Weather Forecasts ERA-Interim reanalysis (Dee et al., 2011), for the Northern Hemisphere winter (December, January and February) during 1979–2018. The 3D wind field and geopotential height variables are available in a 6 hr temporal resolution, interpolated to 1° × 1° horizontal resolution on 60 hybrid vertical model levels. To avoid boundary-related computational artefacts due to the cyclic nature of the data in the longitude direction, the maps are extended on both ends (west/east) by 76° on each side. Latitudinally, the domain stretches between 20° N and 84° N, covering the latitudinal band in which DI trajectories initiate (Figure 1b). The choice of the domain size obeys the requirement of being a power of 2 in each dimension, resulting in maps of size 64 × 512.

1. **Model input:** The data used as model input are 2D maps of geopotential height anomalies of three geopotential height surfaces; 300, 500, 700 hPa. The 6 hr geopotential height anomalies at each grid-point were obtained by removing the instantaneous Northern Hemisphere mean and dividing by its standard deviation.

2. **Model target** (DIOs): To calculate DI trajectories, air parcels were tracked globally with the Lagrangian analysis tool LAGRANTO v2.0 (Sprenger and Wernli, 2015), using the three wind components (u, v, w). The DI trajectories were selected according to specific constraints: air parcels which descended more than 400 hPa in 48 hr (Raveh-Rubin, 2017). An illustrative example of the DI trajectories is given in Figure 1a, showing DI trajectories during 48 hr, with colours indicating the pressure levels along their southeastward descent. To study the initiating mechanism of DIs, we define DIOs as the upper tropospheric part of the trajectories. For this, the portion of the trajectories located at pressures lower than 600 hPa are gridded into 2D Eulerian DIO objects, represented by the horizontal trajectory density, for each time step. This method ensures that the DIs are captured in the right geographical location during their initiation and not during the DI later stages to the southeast.

To avoid potential noise in the learning process, small isolated DIO objects were removed. For this, the size of all DIO objects in the dataset was calculated, and the lower 15th percentile patches (less than 12 grid-points) were removed. In addition, neighbouring DIO grid-points were grouped together if separated by one grid-point. The resulting winter frequency of DIO objects is shown in Figure 1b. Peak DIO frequencies exceed 20% in East Asia at 40° N, and exceed 12% in NE and SW North America, while DIOs also occur ~5–10% of the time over large parts of the North Atlantic and Pacific Oceans, Europe, the Mediterranean and parts of Asia. The model (detailed in Section 2.2) is therefore trained to differentiate between the two discrete classes, namely DIO and non-DIO, at each grid-point. The DIO labels were given to grid-points at which the DI trajectory density was non-zero (Figure 1a, enclosed within the black contours), while all other grid-points were labelled as non-DIO. The mean fraction of the entire domain that DIOs occupy in a single time step is approximately 4%, with a standard deviation of 1.5%.

2.2 Model architecture

To predict the location and probability of a DIO we used a neural network model architecture, commonly used for semantic image segmentation. The U-Net (Ronneberger et al., 2015) is a CNN first introduced as a way to increase the efficiency and accuracy of other conventional CNNs used for segmentation tasks. Our model architecture is inspired by the original U-Net, modified to be applied on atmospheric data rather than RGB images (Figure 2), in order to classify each pixel in the input domain with a specific label. The original U-Net gets an image as an input, and is trained to classify each pixel with a unique label (“dog” or “background”) as the model output. In this work, the methodology is adapted in a non-trivial manner as illustrated in Figure 2 (bottom panel). To predict DIOs, we use 2D geopotential height surface maps as the input (Figure 2, bottom left) and DIOs as the pixel labels (Figure 2, bottom right).

The main advantage of the U-Net architecture is its distinct structure which is composed of two main branches and an additional unique connecting component.
First, the encoder branch (Figure 3, descending branch) applies a set of convolution operations followed by nonlinear activation functions (Figure 3, blue arrows), whilst also reducing dimensionality (maximum pooling operations; Figure 3, pink arrows); together these operations detect the meaningful features but unfortunately lose some information about their location. Second, the decoder branch (Figure 3, ascending branch) combines Up-Conv2d operations (Figure 3, green arrows) to enhance the resolution lost, alongside concatenation of data from the descending branch (Figure 3, yellow blocks) to restore the information about the feature’s location. At the end of the decoder branch, for each input, the model produces a 2D probability map in which the value at each grid-point represents the probability of it being associated with a specific label, in our case a DIO (e.g., Figure 4, shading). The resulting model outputs of 2D probability maps for the predicted DIOs have the same resolution as the model input, unlike the original U-Net which outputs a lower resolution prediction.

The model components were set with the following configurations: the Conv + ReLU operation was set as a 3 × 3 convolution matrix with a stride of 1 and a rectified linear unit (ReLU) activation function. For the MaxPool operation, we used a 2 × 2 matrix with a stride of 2, and for the Up + Conv operation we used a 3 × 3 matrix. A detailed description of the model’s configuration is summarized in Table 1.

### 2.3 Training

In essence, most neural network models operate in a similar way: without any prior knowledge, they seek to minimize a predefined loss function in an iterative learning process by constantly optimizing a set of weights (coefficients) that are nonlinearly bounded to a given set of inputs in order to produce a prediction that is as close as possible to the matching target. The lower the loss function value is, the closer the prediction is to the target. By repeating this process on many different samples (pairs of input and a matching target), if the learning process is successful, the model will converge, meaning that it will minimize the loss function in an optimal way, resulting in successfully extracting the meaningful features hidden in the input and in formulating a multi-dimensional
nonlinear function which associates the input (meteorological data) to the target (Eulerian objects; DIOs).

The 6 hr data record (December, January, February, 1979–2018) was divided into four different sets for use in the training and testing procedures, as presented in Table 2. The training set, as implicated, is the sub-dataset on which the model will be trained. During training, the model constantly evaluates the loss values, aiming to minimize them. For the training stage, 4,000 time steps (28% of the data) were randomly selected, while avoiding consecutive time steps. For validation, 1,000 random time steps (7% of the data) were used as a side test during the training. The validation set is not used to train/adjust the model’s weights but is only being used to evaluate the loss function during learning. As the model is trained to classify between DIO and non-DIO grid-points, we use binary cross-entropy as a loss function.

A successful training stage is considered when the model is able to produce a prediction that is remarkably similar to the target. During training, the loss function is constantly evaluated on both the training and validation sets. An indication for reaching the optimal number of
TABLE 1  An overview of the configuration used in the model for each component and of the resulting data in each step

| Level | Components | Kernel size, stride | Lon × Lat × channel |
|-------|------------|---------------------|---------------------|
| Input |            |                     | 64 × 512 × 3        |
| DOWN1 | Conv2D + ReLU | 3 × 3, stride 1 | 64 × 512 × 64       |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 64 × 512 × 64       |
|        | Maxpool     | 2 × 2, stride 2   | 32 × 256 × 64       |
| DOWN2 | Conv2D + ReLU | 3 × 3, stride 1 | 32 × 256 × 128      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 32 × 256 × 128      |
|        | Maxpool     | 2 × 2, stride 2   | 16 × 128 × 128      |
| DOWN3 | Conv2D + ReLU | 3 × 3, stride 1 | 16 × 128 × 256      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 16 × 128 × 256      |
|        | Maxpool     | 2 × 2, stride 2   | 8 × 64 × 256        |
| DOWN4 | Conv2D + ReLU | 3 × 3, stride 1 | 8 × 64 × 512        |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 8 × 64 × 512        |
|        | Maxpool     | 2 × 2, stride 2   | 4 × 32 × 512        |
| BASE0 | Conv2D + ReLU | 3 × 3, stride 1 | 4 × 32 × 1,024      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 4 × 32 × 1,024      |
|        | Up + Conv2D | 3 × 3, stride 1   | 8 × 64 × 512        |
| UP1   | Concatenate |                     | 8 × 64 × 1,024      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 8 × 64 × 512        |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 8 × 64 × 512        |
|        | Up + Conv2D | 3 × 3, stride 1   | 16 × 128 × 256      |
| UP2   | Concatenate |                     | 16 × 128 × 512      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 16 × 128 × 256      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 16 × 128 × 256      |
|        | Up + Conv2D | 3 × 3, stride 1   | 32 × 256 × 128      |
| UP3   | Concatenate |                     | 32 × 256 × 256      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 32 × 256 × 128      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 32 × 256 × 128      |
|        | Up + Conv2D | 3 × 3, stride 1   | 64 × 512 × 64       |
| UP4   | Concatenate |                     | 64 × 512 × 128      |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 64 × 512 × 64       |
|        | Conv2D + ReLU | 3 × 3, stride 1 | 64 × 512 × 64       |
|        | Conv2D     | 3 × 3, stride 1   | 64 × 512 × 1 (output) |

Note: The level column corresponds to the level numbering shown in Figure 3 (in pink), the components column describes the operation executed in each step. The right column details the resulting dimension of the data in each step.

TABLE 2  Distribution of the 6 hr ERA-Interim dataset for the training procedure

| Number of steps | Fraction from total (%) | Time span                                      |
|-----------------|-------------------------|-----------------------------------------------|
| All time steps  | 14,304                  | 100                                          |
| Training        | 4,000                   | 28                                           |
| Validation      | 1,000                   | 7                                            |
| Test            | 6,977                   | 48                                           |
| Test—isolated   | 1,864                   | 13                                           |

December, January, February, 1979–2018
Random
Remaining
December, January, February, 1984/1985, 2004/2005, 2015/2018
Training epochs followed a common rule of thumb: when the loss function goes down on the training set but begins to rise on the validation set. In that manner we ensure that the model, in its current configuration, reached its optimal learning/prediction capabilities. Similarly to other function fitting tasks, any excess learning process can result in over-fitting of the model too closely to the training set at the expense of allowing it certain “flexibility” that will enable it to produce accurate predictions on datasets that it has not yet encountered.

The test set is a separate dataset which is not used at all during the learning procedure. It is used to evaluate the final prediction abilities of the model on data that it was not trained or validated on. Since synoptic patterns, geopotential height and DIOs in our case, of consecutive time steps often resemble each other, and since the selections of training, validation and test were randomized, presumably similar consecutive time steps ended up in different sets, adding bias to the learning procedure. For this reason, three complete winter sets were isolated from the training set and added to the test set (see Table 2).

### 2.4 Skill metrics

The trained neural network produces 2D spatial maps where each value indicates the probability of a specific grid-point being classified as “DIO,” ranging from 0 to 1, where 1 is a positive match (illustrated for a single time step in Figure 4, shading). The skill of the neural network method to predict DIOs, based on 2D instantaneous geopotential maps, is evaluated against the trajectory-detected DIOs (which require knowledge of the flow field 48 hr ahead). Namely, the model prediction is compared with the reference model target, that is, binary segmentation (Figure 4, black contours; see Section 2.1, item 2). As the model predicts probabilities, it is necessary to choose a certain threshold for transforming to a binary classification. For this purpose, a contour algorithm was used to set a threshold, marking larger/smaller values as “DIO”/“No-DIO” (Figure 4, green contours). The 0.2 contour level was finally selected for the binary classification for the entire dataset based on sensitivity tests to find the value for which most objects are correctly identified.
based on the methods discussed below (see comment on the sensitivity to this choice in Section 3.3).

To evaluate the performance of the model, each of the grid-points is classified as true/false positive (TP/FP) or negative (TN/FN) classification (Figure 5a,b). The spatial distribution of the model target (ground truth) and the model output for a single time step are illustrated in Figure 4. The resulting occurrences are representative in the sense that the “DIO” class events (black/green contours) take a significantly smaller fraction than those in the “No-DIO” class (background). In such cases, measuring the model skill using a standard correlation approach can be misleading if these differences in the overall identified class distributions are not well accounted for. Therefore, to measure the model skill in predicting the relatively sparse DIO objects, we use the Matthews correlation coefficient (MCC) (Baldi et al., 2000; Perruchet and Peerenman, 2004; Powers, 2011). The MCC is a single score metric which can be reliably interpreted as a success or failure of the classifier model, regardless of differences in class representations. The MCC is calculated as follows:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (1)$$

showing the correlation between the target and model prediction, the MCC ranges from −1 to 1, while a perfect match (FP = FN = 0) gives the maximum correlation 1. The score is calculated in a grid-point manner by summing the true and false grid-point identifications at each time step.

The common grid-point based metrics used for computer vision purposes are not necessarily the best choice for meteorological phenomena (Ebert-Uphoff and Hilburn, 2020). As indicated in Figure 4, the DIO grid-points are grouped in specific areas which can be enclosed by a bounding contour and can be referred to as a “DIO object”. Therefore, it is worth evaluating the model performance by finding true/false classifications based on definitions of identified DIO objects rather than considering each grid-point separately. To identify the DIO objects, a clustering algorithm is applied on the binary masks of both target and model output, grouping nearby DIOs and marking them with a unique label. The DIO objects predicted by the model are matched with the target, where a TP event (object based TP) is determined if an intersection between the model and target objects of at least 30% exists, accounting for either the model or the target (Figure 5c). Otherwise, the predicted DIO object is considered a FP event (object based FP, Figure 5d). Similarly, DIO objects in the target map which have no corresponding model prediction are marked as FN events (object based FN). The MCC is then calculated by considering the object-based grid-point labelling to produce the object-based MCC.

In case multiple DIO target objects are associated with a single model output object, the target objects are grouped to a single DIO event and the overlap fraction is calculated accordingly. This is demonstrated in the DIO objects located at 60° W, 100° E, as well as for the elongated DIO object extending from Greenland to the Mediterranean (Figure 4). Such situations can occur due to the choice of threshold, where its contour surrounds several probability maxima (as in the DIO objects located at 100° E and Greenland/Mediterranean). Nonetheless, the object grouping in such cases actually reflects a true relationship in reality, which is often found to be grouped into a single object in the target when inspecting previous time steps, as is the situation in both of the aforementioned DIO objects (not shown).

The MCC will be calculated in both grid-point and object-based manners. To illustrate the differences between the two methods, an exemplary case is shown (January 13, 1979, 1800 UTC, Figure 4), yielding a grid-point based MCC of 0.47 while the object based MCC scored a higher 0.81. The differences emerge from the grid-points associated with object-based TPs in the model output and are considered a correct identification, while in the grid-point based metric they are marked as FP. Also, some grid-points are marked FN instead of TP if the objects overlap less than 30% of the target area. However, this does not have a large impact as will be shown in Section 3.

In summary, the prediction skill of the model is evaluated using the MCC calculated for both grid-point and object-based methods in several manners. First, the MCC is calculated per time step, considering the entire map domain, indicating the model skill for an instantaneous input. Next, the spatial variability of the model performance is tested by calculating the MCC per grid-point, considering all time steps, for finding the target and model intersections. The precision of the true DIO object identifications is evaluated by calculating the MCC for individual TP DIO object events, considering only the grid-points of the matched DIO event. To evaluate the skill of the model to identify the majority of significant DIO events, two additional complementary analyses are provided as a decomposition into the TP/FP/FN of (a) the climatological frequency of the model in both grid-point based and object-based methods and (b) the size distribution of the DIO objects using the object-based method (Section 3.2.2).

3 | RESULTS

We start the presentation of the model performance by inspecting single case studies, and then proceed to a
hemisphere-wide evaluation of how well the model did in reconstructing Northern Hemisphere DIO climatology.

3.1 | Case studies

In order to assess the prediction skill of the model, a focus on two selected case studies is brought here. To test the impartial capabilities of the model, the chosen dates were not part of any training process and were taken from the isolated subset (see Table 2). The first case serves as an example for a good model prediction while discussing the different measures used for success evaluation, as well as some physical insights. The second case study was chosen to analyse the lowest scoring time step, revealing the interesting features that demonstrate the complexity in evaluating the success of the model.

3.1.1 | Case 1—Good MCC score

This case was chosen as it highlights a good performance of the model in several geographical areas; the presented time step is February 28, 2018, 0000 UTC. Figure 6 presents a comparison of the DIOs predicted by the model (green contours) and the DIO target (black contours). As a reference, the model input of geopotential height at 500 hPa is shown (red-blue shading).

The model predicted five large DIO objects (green contours), validated as TP events, and a smaller FP DIO object over Italy. All predicted patches are positioned west of a trough, aligned with previous work (Raveh-Rubin, 2017). Moreover, the model was able to make a distinction between troughs associated with DIOs and those which are not (i.e., the trough west of Spain), indicating that it was able to detect meaningful information hidden in the three geopotential height surfaces, assisting it to produce correct predictions. This consistent behaviour of the model suggests possible dynamic features identified by the model serving as a good DIO predictor.

Overall, the model captured the main patterns of DIO events very well—hence, the high MCC score on both object-based (0.87) and grid-point based (0.58) approaches (for a detailed examination of score distributions see Section 3.2.2). Two large patches, over west and east North America (black contours), were accurately predicted by the model (green contours), the western patch prediction having almost no variation from its target while the eastern patch prediction is slightly misaligned. The two large DIO areas over East Asia are also well predicted by the model, despite some deviations in their extent and shape. A fifth large DIO area spreading from east Russia towards Alaska is also identified; however, here the deviation in the areal coverage is more significant. This deviation is representative with respect to the location dependence of the model precision, which is discussed in Section 3.2, and is found in a location where the DIO frequency is relatively low.

The fact that the false prediction of a small DIO event spotted over Europe is also found at a westward flank of a trough invites us to re-examine the air mass trajectories in that region, under the assumption that some physically meaningful DI-related features were identified by the trained model. The Lagrangian analysis was repeated with a relaxed DI criterion for descent of 350 hPa (instead of 400 hPa) in 48 hr, which showed that such trajectories indeed exist in this false predicted DIO region. Also, inspecting the model-predicted DIO probability shows that this is a region with a relatively low value. These findings indicate that FP predictions can also hold a physical meaning and, considering the model-predicted DIO probability, can also be useful despite them being FPs.

Lastly, there is a small DIO area located in southeast Asia (black contours) which was not recognized by the model, and hence defined as FN. However, examining the previous time steps shows that this DIO patch, which
was located slightly to the north, was previously predicted by the model and was previously connected to the larger DIO object located to the north (not shown).

### 3.1.2 Case 2—Lowest MCC score

This case study highlights some insights regarding the interpretation of the model predictions of DIOs.

The lowest scoring time step in the isolated dataset is shown in Figure 7 (middle), indicating the zero score resulting from seven FP DIO events. To understand the source of these false predictions, a Lagrangian analysis was performed to identify trajectories meeting a relaxed DI criterion (descent of 350 hPa in 48 hr), which was actually found in all the FP patches (not shown). In addition, preceding and following time steps are provided as reference (Figure 7, top and bottom), with capital letters indicating their corresponding matching object in the central date (January 7, 2016, 0600 UTC). From this it follows that events B and E are related to a previous DIO event which continued further than detected by the calculated DI threshold. In contrast, events A, C, F and D are early identification of DIO events found later by the calculated DI threshold. For events C and B, there are also small DIO regions which were filtered out during the training process. Nonetheless, the trained model is able to identify the regions where DIOs are expected and predicts them even when the smaller events were filtered out of the training data. In summary, despite the low MCC score, all false model predictions indeed identify meaningful information.

With this notion, it is considered that the performance of the model is better done for several consecutive
time steps as the false identifications can be related to true prolonged events, a feature that was clearly not considered during the training process. A further discussion regarding this is provided in Section 4.

3.2 | Climatology

3.2.1 | Geographical distribution

Producing a DIO climatology from the model’s predictions is crucial for estimating whether the model is able to capture the overall spatio-temporal variability. The Northern Hemisphere winter climatology can be produced by two complementary approaches. The first is a climatology map which is based on the predicted 2D probabilities map of the model (Figure 8). This is done by computing the mean forecast probability of a DIO issued by the CNN at each time step for each grid-point, that is, without setting a threshold that classifies the prediction maps into our two classes (DIO and non-DIO, see Section 2.4). The rationale of composing this kind of grid-point based climatology is its independence from the threshold choice. Thus, choosing to inspect the raw probability maps directly highly simplifies both the model application and the results interpretation stage.

The climatology of DIO frequency is shown in Figure 8 for the model predicted probabilities (shading) and target DIOs (black contours). It is apparent that the shape and orientation of the target climatology is accurately predicted, from the high frequency areas over North America and East Asia and down to smaller scale weaker variations over Europe and towards the east over the Persian Gulf. Furthermore, the predicted climatology slightly overestimates/underestimates the magnitudes of frequencies by about 3–7% in the DIO peaks/off-peaks, respectively.

Next, the feature-based approach is used to construct the climatology by entailing the binary classification, as shown in Figure 9a. Since DIOs are not a grid-point phenomenon but rather have a spatial structure, a binary classification into DIO/non-DIO classes was defined using a threshold for the predicted model probabilities (see Section 2.4). Setting a threshold to the probabilistic predictions of the model allows us to quantitatively evaluate the model performance with regard to DIO events and their characteristic features and further evaluate the spatial accuracy of DIO objects by estimating the extent of the FP/TP regions. The overall predicted climatological pattern (Figure 9a, coloured shading) captures very well the ground-truth target climatology (Figure 9a, black contours) with its detailed spatial structure, in agreement with Figure 8. Here, the predictions are overestimated everywhere except near Greenland (Figure 9b, shading). However, in this case, using the precise definition of the two classes, we can inspect the influencing components to get a more coherent view of the results.

DIOs are synoptic scale phenomena spanning over 1,000 km with variable shapes. Through object-based analysis, entire predicted DIO objects can be readily compared to ground-truth objects. By doing so, a score can be given to each prediction based on accuracy: Did the model predict an event at all? Did it overestimate or underestimate its size? By how much? Figure 9c–h demonstrates the breakdown of those components, object-based analyses (left column) versus the grid-point based analyses (right column); the three rows are TP, FP and FN accordingly. The first row shows the climatology of the TP identifications. Strong similarities between the grid-point based and object based approaches (Figure 9c,d) reveal that indeed the majority of the events are captured by the model (80%/90% for grid-point/object based methods, respectively). The FP maps offer a complementary meaningful view. While the grid-point based FP (Figure 9f) map shows widespread areas with a high frequency of FP identifications, peaking locally off Japan and suggesting a frequency double that observed, the object-based FP map (Figure 9e) shows a significantly lower frequency, mostly 0–3%, and locally up to 6%/7% (off Japan), accounting for

FIGURE 8  Climatology of dry air intrusion origin (DIO) probability as predicted by the trained model (shading) and the frequency of target DIOs (black contours), shown as a percentage of 6 hr time steps for December, January, February, 1979–2018
an over-prediction of 20% in those regions. The lower object-based FP frequency suggests that a majority of the over-predictions indicated by the grid-point based FP are related to DIO objects which were correctly identified; however, the accuracy in terms of their exact shape is not perfect.

Another indication that the model captures most of the calculated DI events is provided by examining the missed DIO events given by FN identifications, being relatively low in both grid-point and object based methods (Figure 9g,h). The object-FN identifications observed near Greenland account for about 40% of the DIO events in that region. Considering the object-FP frequency in that region, it is possible that at least some of the missed events are due to a lower precision of the model in the region. In the DIO hotspot regions, the FN frequency accounts for less than 10% of the calculated DIO events.

To summarize, the model is able to capture most of the calculated DIO events in the DIO hotspot regions, while the accuracy in their spatial structure can be improved. The DIO object size and accuracy are further discussed in the next section.

3.2.2 DIO object characteristics

The model performance is next examined in terms of the DIO object size distributions (Figure 10). By construction,
TP objects can be examined in both the model and target, while FN/FP object sizes can be examined at the target/model only. The model-predicted DIO objects are overall overestimated in size, as indicated by the size distribution of the true identifications in the target/model (Figure 10a, blue/orange). This finding supports the notion that the low DIO frequencies shown by the grid-point based approach (Figure 9f,h) compared to object based (Figure 9e,g) can be explained by the overestimated sizes of the predicted objects.

The model performs better in predicting larger DIO objects, as shown by the size distribution of the FN/FP DIO object model predictions (Figure 10a, green/red), which peak at the smallest size bin and decrease sharply for larger DIO objects, having practically no objects larger than 400 1° pixels. In contrast, the size distribution of successful model predictions (TP) peaks at larger sizes (~100 grid-points) and its tail extends to much larger sizes up to 1,000 grid-points. As discussed above (and in Section 3.2.1), it is evident that the accuracy of the true model prediction is not perfect in terms of areal coverage. Here, we inspect how this accuracy varies depending upon the size of the DIO object by evaluating the MCC score individually for each TP object and not for the entire time step. The dependence of the MCC calculated for TP objects upon the corresponding object size at the target is displayed in the 2D density map (Figure 10b). It is evident that the lower MCC scores dominate the smaller DI objects, while the larger objects are mostly associated with a higher MCC score (Figure 10b). Overall, the model skill is higher for larger DI events, both in being able to identify them and in capturing their spatial structure.

The distribution of the object-based MCC per time step (Figure 10c, blue) indicates that the model scores are high most of the time, while the grid-point based scores are shifted to lower values (orange), in agreement with the previous findings. The average scores for the model performance are summarized in Table 3, for both grid-point and object based methods, as well as for the average score of all the positive identifications. The spatial distribution of the MCC score can reveal how the...
model performance varies geographically (Figure 11, shading). The grid-point based MCC score (Figure 11a) varies from 0.2 to 0.7, indicating a positive but not perfect local correlation between the model prediction skill and the target DIO frequencies, consistent with the previously mentioned results (Figures 9 and 10b). The object-based MCC map (Figure 11b) shows higher scores: 0.5–0.9. Using both grid-point and object based MCC approaches, the spatial distribution is not uniform. Areas of high MCC scores can be found mostly in the regions where the DIs occur more frequently (Figure 11, black contours). The highest DIO frequency is found over East Asia and the Pacific Ocean, co-located with high MCC scores larger than 0.8. The same can be seen for the lower frequency DIO hotspots over Europe and the Middle East. In North America a slight deviation between the peak frequencies and the MCC highest peak is found. The DIO contour, which extends into Greenland and is expressed in a MCC high score zone, indicates that the true DI predictions in that region tend to extend northwards. Given the model is adjusted towards predicting DI events during the training, perhaps it is not surprising that the score is higher in the regions where the events are more frequent.

3.3 Sensitivity tests

The results discussed above are based on the specific contour level which was selected for converting the model probabilities to the binary DIO classification. The choice for the contour level and overlapping threshold to define a TP was selected by inspecting individual steps for different cases. Then, a sensitivity test was conducted to find where the largest differences occur in the MCC and FN/FP rate. Too low (high) thresholds result in a higher FP (FN) identification rate.

The mean MCC score varies by approximately 3% for the object-based score, and less than 2% for the

| TABLE 3 Performance metrics |
|-----------------------------|
| Grid-point based | Object based | True positive (TP) events |
| Mean Matthews correlation coefficient (MCC) | 0.527 | 0.796 | 0.608 |

FIGURE 11 Longitude–latitude distribution of the Matthews correlation coefficient (MCC) score (shading) calculated for the grid-point based (top) and object based (bottom) methods (details provided in Section 2.4). The target dry air intrusion origin (DIO) climatological frequency (%) is shown as black contours.
grid-point-based score, when increasing the contour level by 0.05. The grid-point-based metric is improved by up to 0.25 while the object-based metric decreases. While the pixel-based MCC is increasing due to fewer FP grid-points, the object based MCC is decreased due to the DIO objects being smaller, which dominates over the reduction on FP grid-point count, composed mostly of smaller DIO objects (as in Figure 10a). Increasing the threshold further results in a lower grid-point-based MCC, with more FN DIO objects found, increasing in size. Therefore, the 0.2 level selected was a reasonable threshold which captures the DIO events.

The effect of changing the threshold on the climatological DIO frequency is relatively small; increasing the level to 0.25 reduces DIO frequency by up to 7% in the East Pacific and up to 3% in the West Atlantic. Accordingly, TP is reduced by ~3% in the East Pacific and 1.5% in the West Atlantic. The FN rate is increased by less than 1%.

There is a trade-off between the contour level and overlap thresholds: taking a lower contour level threshold results in larger DIO objects compared to their matching target (Lagrangian calculated) objects. Therefore, a lower overlap fraction is required to ensure matching objects are not dismissed. For this reason, the MCC score calculated per object is provided (Figure 10c) as a metric for the object matching precision, which depends upon the overlap fraction.

In principle, it is possible to train the neural network to produce the binary mask directly and avoid these decisions; however, since the phenomenon is not binary in nature, here we explored the results from inspecting the probability map.

4 | SUMMARY AND CONCLUSIONS

This work examined the use and application of a neural network based model for predicting the occurrence of atmospheric phenomena, related to specific features of their trajectories, based on three instantaneous 2D maps of geopotential height. A U-Net inspired model was successful in extracting the meaningful features from the input data and to resolve the relationships between dry intrusion origins (DIOs) and geopotential height input fields. By formulating a complex (and largely unknown) relationship between the geopotential structure and the probability of a DIO occurrence, the trained model is able to predict individual DIO events based on limited instantaneous atmospheric data, which could otherwise be found only in retrospect after extensive calculations of Lagrangian air mass trajectories.

The Northern Hemisphere winter climatology was accurately reconstructed (Figures 8 and 9b), including the detailed spatial structure and frequency, along with the main DIO hotspots. Using an object based evaluation of the model performance revealed that most of the DIO events are captured, while the precision in the event spatial structure can be improved, especially in regions where the DIO frequency is lower.

The Matthews correlation coefficient (MCC) metric was used to evaluate the model in both grid-point and object-based manners. The object-based approach showed overall higher MCC scores compared to the grid-point based approach, shown here in terms of both the MCC temporal and spatial distributions. Owing to the object-based approach we learned that model predictions are more accurate (higher MCC score) for larger DIO events. Possible explanations can be that (a) the larger/stronger atmospheric signature of larger DIO events make it more straightforward for the model to predict them, (b) the limited temporal duration of the smaller events prevents the model from identifying them, and (c) a fraction of small events were in fact identified by the model but were a mismatch in terms of location (such as in Figure 6) resulting in the prediction being accounted for as both a false positive (FP) and a false negative (FN).

The geopotential height fields were shown to be a successful predictor of DIOs and to hold valuable information on D1 initiation. The evaluation of representative case studies (Section 3.1) indicated that DIOs were positioned by the model on the western flank of a trough, as well as some of the FP objects, agreeing with previous observations (Raveh-Rubin, 2017). Nonetheless, the model was able to make a distinction between troughs associated with DIOs and those that are not. Geopotential height provides the model with the most straightforward representation of upper tropospheric flows. This is an encouraging and applicable result, as geopotential height is a common output for atmospheric models. A future application is to evaluate the DIO phenomena on climate model outputs (e.g., CMIP, Eyring et al., 2016), under different future climate scenarios. Indeed, most of these models do not provide the high spatio-temporal resolution data required for Lagrangian trajectory calculations. By training a model based on geopotential height maps, it will be possible to project future climatic changes in DIs.

It is critical to emphasize that it is not trivial that geopotential height fields alone turned out to be a very good predictor. DIOs have a large spatio-temporal variability which may be driven by several atmospheric processes, potentially varying among regions with diverse topographies. Since different synoptic regimes across the Northern Hemisphere may apply different triggering mechanisms to DIOs, the model could be further developed, to adapt
uniquely to each region during the training process. If this results in higher prediction skill it would reinforce the notion that under different synoptic regimes which vary across regions DIOs do act differently.

In this work DIs were treated in a binary way, without considering their air mass flux or 3D spatial/temporal evolution. Future construction of a more complex model which would account for DIO spatial density and/or their space–time evolution could provide a more accurate detailed prediction of DIs and their impact.

Adapting a neural network model primarily developed for computer vision purposes and treating it as a tool to find the right predictors for an atmospheric phenomenon has proven useful and strongly inspires confidence in reviewing further methods and further implementation that can be borrowed from the realm of computer science. It serves as a valid proof of concept to an innovative way to predict weather-related objects based on limited atmospheric data.

ACKNOWLEDGEMENTS

We are grateful to Jonathan Shlomi and Eilam Gross (Weizmann Institute of Science) for inspiring this work and for the helpful discussions. The ECMWF and Israel Meteorological Service are acknowledged for providing access to ERA Interim. This research was supported by the Israel Science Foundation (grant no. 1347/18). We thank Ambrogio Volonté and an anonymous reviewer whose constructive suggestions helped to improve the clarity of the manuscript.

ORCID

Vered Silverman https://orcid.org/0000-0002-4679-3785
Stav Nahum https://orcid.org/0000-0001-5260-0882
Shira Raveh-Rubin https://orcid.org/0000-0001-6244-8693

REFERENCES

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. and Nielsen, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics, 16, 412–424.

Binder, H., Boettcher, M., Joos, H. and Wernli, H. (2016) The role of warm conveyor belts for the intensification of extratropical cyclones in northern hemisphere winter. Journal of the Atmospheric Sciences, 73, 3997–4020.

Bjerknes, J. (1922) Life cycle of cyclones and the polar front theory of atmospheric circulation. Geophysik Publikation, 3, 1–18.

Browning, K. and Roberts, N. (1994) Structure of a frontal cyclone. Quarterly Journal of the Royal Meteorological Society, 120, 1535–1557.

Browning, K.A. and Reynolds, R. (1994) Diagnostic study of a narrow cold-frontal rainband and severe winds associated with a stratospheric intrusion. Quarterly Journal of the Royal Meteorological Society, 120, 235–257.

Carlson, T.N. (1980) Airflow through midlatitude cyclones and the comma cloud pattern. Monthly Weather Review, 108, 1498–1509.

Catto, J.L. and Raveh-Rubin, S. (2019) Climatology and dynamics of the link between dry intrusions and cold fronts during winter. Part I: Global climatology. Climate Dynamics, 53(3), 1873–1892. https://doi.org/10.1007/s00382-019-04745-w.

Chang, E.K., Lee, S. and Swanson, K.L. (2002) Storm track dynamics. Journal of Climate, 15, 2163–2183.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balssemma, M.A., Balssemma, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragnani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksegen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137, 553–597.

Dowdy, A.I. and Catto, J.L. (2017) Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences. Scientific Reports, 7, 40359.

Ebert-Uphoff, I. and Hilburn, K. (2020) Evaluation, Tuning, and Interpretation of Neural Networks for Working with Images in Meteorological Applications. Bulletin of the American Meteorological Society, 101(12), E2149–E2170. http://dx.doi.org/10.1175/bams-d-20-0097.1.

Eyting, V., Bony, S., Méheurl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9, 1937–1958.

Ilotoviz, E., Ghate, V.P. and Raveh-Rubin, S. (2021) The impact of slant-wise descending dry intrusions on the marine boundary layer and air–sea interface over the ARM eastern North Atlantic site. Journal of Geophysical Research: Atmospheres, 126, e2020JD033879.

Larson, P.R., Renzullo, L.J., Inza, I. and Lozano, J.A. (2019) A data-driven approach to precipitation parameterizations using convolutional encoder–decoder neural networks. arXiv preprint arXiv:1903.10274.

Lawrence, S., Giles, C.L., Tsoi, A.C. and Back, A.D. (1997) Face recognition: a convolutional neural-network approach. IEEE Transactions on Neural Networks, 8, 98–113.

Li, A.S., Chirayath, V., Segal-Rozenhaimer, M., Torres-Perez, J.L., van den Bergh, J. (2020) NASA NeMO-Net’s Convolutional Neural Network: Mapping Marine Habitats with Spectrally Heterogeneous Remote Sensing Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 5115–5133. http://dx.doi.org/10.1109/jstars.2020.3018719.

Ludwig, P., Pinto, J.G., Hoepp, S.A., Fink, A.H. and Gray, S.L. (2015) Secondary cyclogenesis along an occluded front leading to damaging wind gusts: Stormwind Kyrill, January 2007. Monthly Weather Review, 143, 1417–1437.

Madonna, E., Wernli, H., Joos, H. and Martius, O. (2014) Warm conveyor belts for the intensification of extratropical cyclones. Monthly Weather Review, 143, 1537–1557.

Perruchet, P. and Peereeman, R. (2004) The exploitation of distributional information in syllable processing. Journal of Neurolinguistics, 17, 97–119.
Pfahl, S. (2014) Characterising the relationship between weather extremes in Europe and synoptic circulation features. *Natural Hazards and Earth System Sciences*, 14, 1461–1475.

Pfahl, S., Madonna, E., Boettcher, M., Joos, H. and Wernli, H. (2014) Warm conveyor belts in the ERA-interim dataset (1979–2010). Part II: Moisture origin and relevance for precipitation. *Journal of Climate*, 27, 27–40.

Pfahl, S. and Wernli, H. (2012) Quantifying the relevance of cyclones for precipitation extremes. *Journal of Climate*, 25, 6770–6780.

Powers, D. M. (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Rasp, S., Schulz, H., Bony, S. and Stevens, B. (2020) Combining Crowdsourcing and Deep Learning to Explore the Mesoscale Organization of Shallow Convection. *Bulletin of the American Meteorological Society*, 101(11), E1980–E1995. https://doi.org/10.1175/BAMS-D-19-0324.1.

Raveh-Rubin, S. (2017) Dry intrusions: Lagrangian climatology and dynamical impact on the planetary boundary layer. *Journal of Climate*, 30, 6661–6682.

Raveh-Rubin, S. and Catto, J.L. (2019) Climatology and dynamics of the link between dry intrusions and cold fronts during winter, part II: front-centred perspective. *Climate Dynamics*, 53, 1893–1909.

Raveh-Rubin, S. and Wernli, H. (2016) Large-scale wind and precipitation extremes in the Mediterranean: dynamical aspects of five selected cyclone events. *Quarterly Journal of the Royal Meteorological Society*, 142, 3097–3114.

Ronneberger, O., Fischer, P. and Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer.

Scher, S. and Messori, G. (2018) Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, 144, 2830–2841.

Segal-Rozenhaimer, M., Li, A., Das, K. and Chirayath, V. (2020) Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN). *Remote Sensing of Environment*, 237, 111446.

Sprenger, M. and Wernli, H. (2015) The LAGRANTO Lagrangian analysis tool—version 2.0. *Geoscientific Model Development*, 8, 2569–2586.

Wernli, H. (1997) A Lagrangian-based analysis of extratropical cyclones. II: A detailed case-study. *Quarterly Journal of the Royal Meteorological Society*, 123, 1677–1706.

Weyn, J.A., Durran, D.R. and Caruana, R. (2019) Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11, 2680–2693.

**How to cite this article:** Silverman V, Nahum S, Raveh-Rubin S. Predicting origins of coherent air mass trajectories using a neural network—the case of dry intrusions. *Meteorol Appl*. 2021;28:e1986. https://doi.org/10.1002/met.1986