⊂⊃ ECMWF Feature article

EARTH SYSTEM SCIENCE

··

Red sky at night...
producing weather forecasts
directly from observations

··

Red sky at night... producing weather forecasts directly from observations

Tony McNally, Christian Lessig, Peter Lean, Matthew Chantry, Mihai Alexe, Simon Lang

ECMWF is embarking on a radical and ambitious project to investigate if weather forecasts can be made directly from meteorological observations, harnessing the power of machine learning (ML).

Highly skilful ML weather forecasts have challenged our approach to numerical weather prediction (NWP), prompting the development at ECMWF of our own ML forecasting system called AIFS (see the article on the AIFS in this Newsletter). Its performance is already highly competitive with that of established systems like Pangu-Weather and GraphCast. ML prediction systems (including the AIFS) have been trained to forecast future weather by learning from long historical records of past weather, typically provided by ECMWF reanalyses, such as ERA5. These datasets are well suited to training ML forecasts: they are highly accurate descriptions of the atmosphere and they provide conveniently gridded values of all required parameters, available at all locations and at all times for very long historical periods. In addition, our reanalysis datasets have been freely available to the wider research community, including the commercial sector. This has been a major factor in the rapid rise of ML forecast systems and the impressive levels of accuracy they have already achieved. However, while reanalyses, and initial conditions generated through data assimilation, are currently still crucial for ML forecasts, it is unclear if this will remain so in the future. Fundamentally, atmospheric analyses are just a fusion of an existing short-range forecast with the available meteorological observations, and an obvious question is if ML forecast systems could be trained and initialised directly from these observations. It is an intriguing science question, but also one that could potentially have huge implications for how weather forecasting is done in the future.
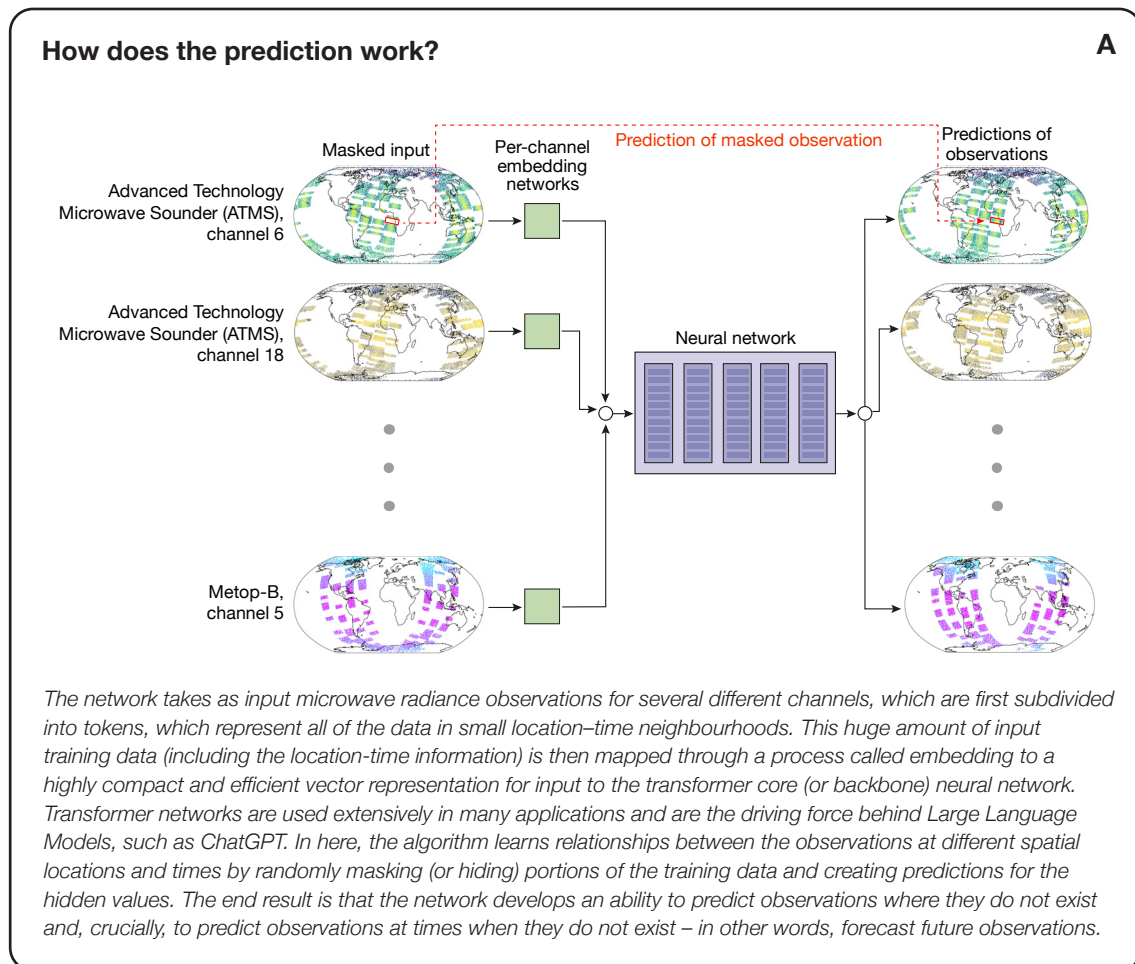
The use of observations in conventional forecast models

The many millions of meteorological observations made each day require highly sophisticated data assimilation (DA) systems to transform the raw measurements at irregular times and locations into data on a regular grid and into the variables required to initialise forecast models. This is an extremely challenging task because traditional forecast models demand initialisation on a fine spatial grid over the entire globe, even where there may be no observations, and with meteorological variables that may not be directly measured. For example, the overwhelming majority of observations come from weather satellites which measure thermal radiation emanating from broad vertical layers of the atmosphere. They do not directly measure temperature or humidity, and the extraction of this information requires a detailed understanding of complex radiative transfer processes in the atmosphere. Downward looking satellites cannot provide information on fine vertical scales, either. To meet the demands of forecast model initialisation, the measurements must be carefully blended with gridded fine-scale information from a previous short-range forecast background. ECMWF has been extremely successful in the development of its data assimilation system and has ambitious plans to extend its capability in a number of exciting directions. However, exploiting observations in this way remains a highly demanding scientific and technical activity.

Is there an alternative?

Much of the complexity described above stems from the fact that conventional forecast models demand initial conditions for all meteorological variables on a regular spatial grid. The strategy employed by conventional models to produce accurate weather forecasts is to represent the real atmosphere as comprehensively as possible, explicitly describing a myriad of fine-scale physical processes and interactions between variables at every grid point over the entire globe, from the surface up to the mesosphere. However, it has been demonstrated recently that ML systems operating with far fewer variables than conventional models and on significantly coarser spatial grids are capable of producing highly skilful medium-range forecasts of important weather parameters. This begs the intriguing question if ML forecast models could learn and be initialised directly from observations, obviating the need to convert observations to a fine grid of unmeasured physical variables dictated by the NWP model. If training of a ML model from observations was successful, large fractions of the conventional data assimilation process could be circumvented, potentially allowing forecasts to be produced more quickly as soon as observations are available. Direct Observation Prediction (DOP) with ML models may also be able to exploit additional observations that are currently not used by conventional data assimilation

systems. For example, satellite radiance measurements in the visible part of the spectrum have such complex radiative transfer that they are not yet assimilated in global NWP systems. Yet in any animation of visible imagery one can clearly see the movement of weather patterns around the globe, and it seems entirely plausible that an ML forecast system could readily exploit this information to predict how weather patterns will evolve in the future. Another aspect is that an observation-based forecasting system does not require the approximative modelling of unresolved physical processes. These 'parametrizations' are a source of considerable uncertainties in conventional models. Instead, the ML model would rely on the feedback from small-scale processes that is captured in the observations to implicitly resolve processes that, for example, take place at small scales.

How does the prediction work?   **A**

The network takes as input microwave radiance observations for several different channels, which are first subdivided into tokens, which represent all of the data in small location–time neighbourhoods. This huge amount of input training data (including the location-time information) is then mapped through a process called embedding to a highly compact and efficient vector representation for input to the transformer core (or backbone) neural network. Transformer networks are used extensively in many applications and are the driving force behind Large Language Models, such as ChatGPT. In here, the algorithm learns relationships between the observations at different spatial locations and times by randomly masking (or hiding) portions of the training data and creating predictions for the hidden values. The end result is that the network develops an ability to predict observations where they do not exist and, crucially, to predict observations at times when they do not exist – in other words, forecast future observations.

### Preliminary investigations: observations predicting future observations

Research is at a very early stage, and the first question we are attempting to answer is how much predictive skill can be achieved when training on observations alone. When models like AIFS and GraphCast are trained from reanalysis datasets like ERA5, they are learning to predict the weather from past observations, but also from the conventional forecast model (the Integrated Forecasting System, IFS) used in the ERA5 data assimilation process. This conventional model plays an important role in the production of a coherent gridded representation of the atmospheric state. As a preliminary step, we are therefore investigating if an ML algorithm with only observations as input can be trained to predict future observations, completely removing any influence of a conventional forecast model in the training process. Our first experiments use microwave radiance measurements from satellites. These have the advantage that they provide high-quality observations with homogeneous coverage over long historical periods of time. However, they also come with the limitation of rather coarse vertical and horizontal resolution. A prototype Transformer Neural Network (Box a) has been trained with ten years of real satellite observations to make 12-hour predictions, essentially using observations in one 12-hour assimilation time window as input and predicting the observations that would be obtained in the next 12-hour window.

It can be seen in Figure 1 that the 12-hour predicted values are certainly realistic in terms of structure and variability compared to the real observations obtained 12 hours later. Furthermore, one can clearly see the movement of conspicuous meteorological features in the real observations from one 12-hour window to the next, and that the ML prediction correctly captures this movement.
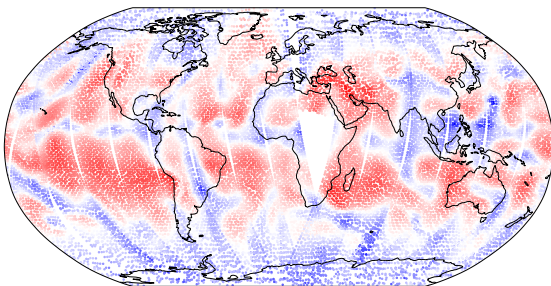
a ATMS radiances in 12-hour window
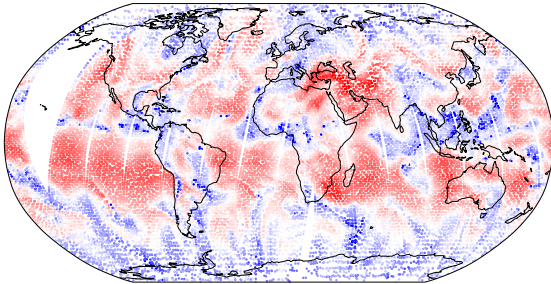
b ATMS radiances in subsequent 12-hour window
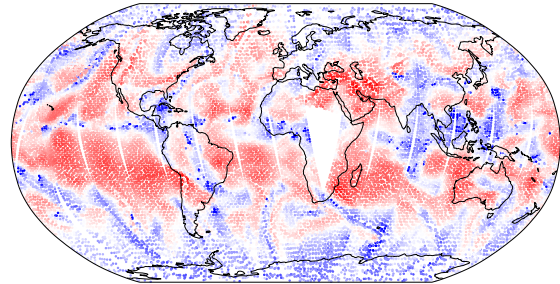
c ML predicted values

Figure 1 Here we show (a) observed Advanced Technology Microwave Sounder (ATMS) channel 18 radiances in an arbitrary 12-hour window provided as input to the ML prediction, (b) ATMS observations obtained in the subsequent 12-hour window, and (c) the ML predicted values.

These results are encouraging and suggest there is indeed predictive skill (at least over 12 hours) that can be learned from the observations alone. A next step is to challenge the algorithm with completely different types of observations. We will use land-surface temperature measurements from SYNOP weather stations, to test if meteorological variation can still be learned from the data in the presence of a strong diurnal cycle. This will also bridge the gap between the radiances used in the current experiments and variables of relevance to users, such as 2 m temperature or wind. Finally, in this preliminary part of the project we will investigate to what extent complementary information from different observing systems enhances the quality of the prediction. An exciting example of this will be combining visible and infrared imagery with SYNOP surface temperature observations. We know that in nature the presence of cloud cover can have a strong impact upon the magnitude of the land surface temperature diurnal cycle. Can the algorithm improve its prediction of surface temperature using cloud information learned from the imagery?

How might we predict weather parameters directly from observations?

In the experiments described above, the ML algorithm is learning to predict future observations at real observation locations and times. However, provided there is sufficient information in the training data and with a suitable network architecture, ML algorithms are capable of learning spatial and temporal relationships between observations. Through this, they can generalise the predictions to locations and times where there are no direct measurements. Thus, if the algorithm can successfully learn to predict the surface temperature patterns from SYNOP stations, it should be possible to predict surface temperature at any location, even locations where there are no SYNOP stations. The same approach would apply for upper air information if the algorithm can learn to successfully predict radiosonde observations. All of this remains speculative at the moment. But with the existing network architecture and training protocol, we have identified a plausible mechanism allowing weather information for any time or location to be generated from a prediction algorithm trained directly on observations. There are, however, a number of different options for DOP that can be explored in the future (see Box b).

Learning from observations: different approaches that could be taken   B

The current generation of analysis-driven ML models takes gridded (re)analysis data as input and is trained to predict the same fields at a future time (as shown under 1). Learning directly from observational data could take several different forms as illustrated in the figure. Firstly, observations could be used as a target 'truth' in the training, while still using gridded analysis data as input (shown under 2). For example, the existing AIFS could be trained to predict SYNOP observations – free from any systematic model biases present in the analysis datasets.

Alternatively, the ML model could be initialised directly from observational data (shown under 3–6). An ML model could be trained to map from input observations to the gridded 4D-Var analysis valid at the same time (shown under 3). By emulating 4D-Var in this way, a gridded analysis could be generated far more quickly than through the current computationally

expensive data assimilation system. However, any model biases present in the existing system would be inherited by the ML model, and such an approach would never be able to surpass the quality of the current 4D-Var analysis.

Panel 4 shows a scenario where the ML model takes observations as input but is trained to predict a gridded state in the future. This has the advantage that gridded forecasts could be generated for all model variables. However, a physical model and data assimilation would still be needed to generate the training dataset.

In the exploratory work discussed in this article, we are using a neural network which takes observations as input and is trained to predict the future state using observations as the ground truth (panel 5). This approach only includes observational data in the training dataset.

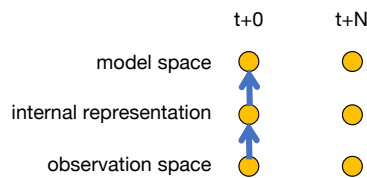**1 Predictions from analysis**
e.g. current AIFS

2 Predict observation targets from analysis input
e.g. fine tune AIFS to predict SYNOPS

3 Learn the analysis
emulate 4D-Var

4 Predictions from observations
make predictions in model space,
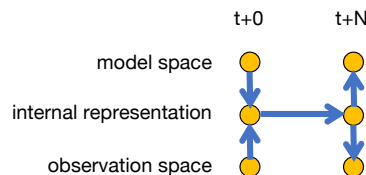use reanalysis as truth

5 Predict future observations from observations
make predictions in observation space,
use observations as truth

6 Other combinations

Moving towards longer-range predictions

The acid test of any observation-based prediction system is if it can be extended to forecast ranges of multiple days or even weeks. Data-driven forecast models (such as the AIFS, GraphCast and Pangu-Weather) trained from reanalysis data have clearly demonstrated this capability, so we have good reason for optimism with DOP. To extend the range of predictions, we will investigate the use of the different types of neural network available (e.g. the Transformer Network described here, but we will also explore the Graph Networks employed by the AIFS). We will also explore different options for training longer-range predictions (auto-regressive vs fixed forecast length) to assess which is most suitable for longer-range direct observation forecasting. Here, our efforts will be accelerated by drawing extensively upon the experience gained in the development of the AIFS. Another interesting area of exploration will be enhancing the information content of the initialisation of the DOP algorithm with windows significantly longer than 12 hours (e.g. several days of observations). The aim will be to better inform on the past and current trajectory of weather systems and to provide an analogue of the background information used in conventional data assimilation. Here we will directly benefit from the highly sophisticated scientific and technical observation handling infrastructure that already exists at ECMWF.

Summary

We believe there is a strong strategic and scientific motivation for exploring this exciting new approach to weather forecasting. Direct observation-driven prediction is a very radical approach and, as such, pursuing this direction comes with no guarantee of success. However, we believe that our observation handling and assimilation experience, combined with our rapid uptake of ML technology, makes ECMWF very well placed to explore this pioneering and potentially paradigm-shifting area of research. This effort will very soon be complemented by a joint project with ECMWF Member States, where ML-assisted data assimilation will be addressed as one of the topics.