Making Numerical Weather Predictions Portable

Compression of Weather Data for Use in Radar Propagation Modeling

Vincent van Leijen, Marijn Boone
Defence Materiel Organisation (DMO)
JIVC/KIXS
Den Helder, The Netherlands
AV.v.Leijen@mindef.nl, T.Boone@mindef.nl

Katherine Horgan
Naval Surface Warfare Center
Dahlgren Division
Dahlgren, VA, USA
Katherine.Horgan@navy.mil

Abstract— **Would it be possible to reduce large numerical weather predictions to the size of an email attachment? This question was given a positive answer in a recent internship after which a prototype for (de-)compression was tested in two field experiments. Large volumes of weather were shown to be reduced to less than a percent of the original datasets but this came with the price of information loss. In this study a method for lossy compression of weather data is described by representing the data as a sparse and adaptive subset. Adaptive sampling is then solving an optimization problem for the minimal loss of information for applications in radar propagation modeling.**

Keywords—compression; optimization; numerical weather prediction; radar propagation

I. INTRODUCTION

Numerical Weather Predictions (NWP) are a scientific product of ever increasing accuracy. The current state of the art is that 4 model runs become available each day in 3km grids covering more than 60 model levels in 1 hour forecast steps. Already the supply of NWP has grown to large data volumes and increases in resolutions are to be expected [1].

Most if not all operational users of NWP do not demand the full set of available data. The obvious control of data volume is to obtain relevant parameters only, for a limited set of forecast hours and within some geographical boundaries. But even then there could be good reasons for users to reduce the volume of weather data. NWP is a time-critical product so it makes sense to use compression to save space and transmission times. Data formats such as NetCDF and Grib2 support lossless compression while maintaining quick access to the data [2].

This paper aims to go beyond these measures in support of those users that require substantially smaller datasets in exchange for some loss of information.

II. LOSSY COMPRESSION BY ADAPTIVE SAMPLING

A. Intended application: use of NWP at sea

A typical case of time-critical use of NWP that would benefit from the large data reduction that lossy compression might offer is planning of activities by ships at sea. The relative slow, limited and at times interrupted availability of satellite communication makes that ships in general do not have access to NWP. The most demanding use of NWP is probably in applications for radar propagation modeling as these use data from all the model layers – so this will be the use case studied here. Lossy compression can be formulated as an optimization problem: compress a large NWP to a specified and much smaller data volume (based on the available bandwidth), where the loss of information has the lowest impact on radar propagation modeling. In other words: given a reduction factor, suppress the data degradation as much as possible.

B. General approach of compression and decompression.

The dynamic character of local weather conditions can have a strong impact on radar propagation. It has been shown that linear down-sampling (or decimation) from data sets with high range resolution can introduce various artifacts [3]. A better approach for compression would therefore be to replace a large grid of data by a subset of well-chosen samples. Decompression is then a matter of restoring the data grid by interpolation from the sample points. Compression with less information loss can be achieved by adaptive sampling, a procedure that aims to select those samples that will enable the best approximation of the original grid [4].

C. Measuring information loss

To measure the information loss of compression one has to compare the original data Y with the decompressed data $\hat{Y}$. Two common measures are the mean squared error (MSE) and the peak-signal to noise ratio (PSNR).

$$MSE = 1/n \sum_{i=1..n} (Y_i - \hat{Y}_i)^2 \qquad (1)$$

$$PSNR = 10log_{10} ((Y_{max}-Y_{min})^2 / MSE) \qquad (2)$$

In (1) Y and $\hat{Y}$ count n data points and in (2) Y_{max} and Y_{min} are the largest and smallest values in Y. The (root) MSE is an intuitive measure: the unit is the same as for Y and $\hat{Y}$, and the MSE drops to zero when there is no information loss. The PSNR is preferred when losses are to be compared over different data sets.
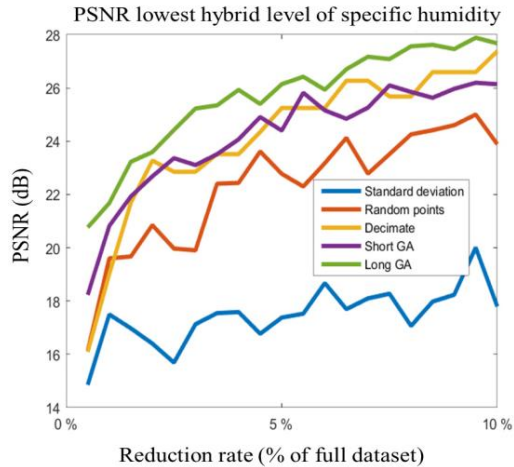
Fig. 1. Comparison of adaptive sampling methods for varying reduction rates. A long running Genetic Algorithm (GA) gives the highest PSNR and thus the least loss of information.
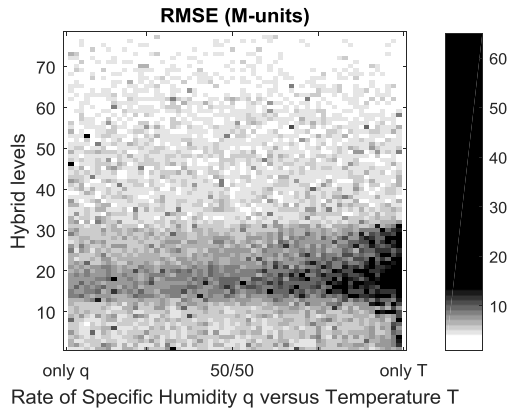

Fig. 2. Information loss depends on the distribution of available samples over q and T. For the lower hybrid levels the dark area to the right indicates that information loss is more sensitive to specific humidity than to temperature.
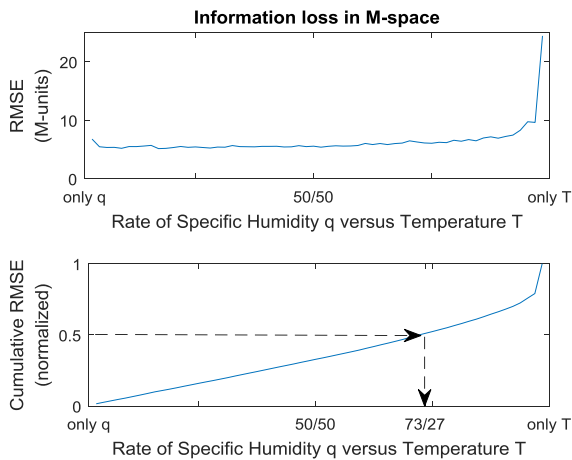

Fig. 3. The cumulative distribution of the total information loss gives Specific Humidity emphasis over Temperature with a factor 73/27≈2.8.

Radar propagation depends on profiles of modified refractivity. These M-profiles can be calculated from specific humidity, temperature, pressure and geometric height. The last two parameters can be recalculated from the lowest layer, and so omitting layers two and higher up is already a good start for lossless compression. Other surface-layer parameters like wind and sea-surface temperature are useful for assessment of the evaporation duct [5], but as these are single-layer parameters most reduction can be expected from the many layers of specific humidity and temperature.

III. RESULTS

Various methods for optimal adaptive sampling have been studied in [4]. From these, Fig.1 shows that a long running genetic algorithm outperforms the other studied methods. This method is a guided search strategy that requires the longest compression time (times are not shown).

Using the inferior but faster random sampling method, Fig.2 depicts how the information loss depends on the balance between samples of specific humidity and temperature. In the studied NWP example, most information loss of the M-profiles occurred in the lower hybrid levels, which hints that more samples should be used for lower levels than for higher levels.

Fig.3 shows the information loss of the M-profiles for the integral atmosphere and in the same dataset. The cumulative distribution of the RMSE points out that the specific humidity should be given more weight than the temperature. In this example the ratio for adaptive distribution of samples between specific humidity and temperature is 2.8.

IV. CONCLUSION

Large datasets (>> 100MB) of numerical weather predictions can be reduced to the size of an email attachment (<< 10MB). The loss of information can be measured and minimized by use of adaptive sampling. It has been shown that specific humidity has more significance than temperature, as do model layers at the lower heights.

REFERENCES

[1] R. E. Marshall, W. D. Thornton, G. Lefurjah, and T. S. Casey, "Modeling and simulation of notional future radar in non/standard propagation environment facilitated by mesoscale numerical weather prediction modeling,," in Naval Engineers Journal, vol. 120, 2008, pp. 55–66.

[2] Sullivan, S. "Comparison of Netcdf4 (HDF5) and Grib2 compression methods for meteorological data", UCAR report, June, 2011.

[3] Derksen, J., Radar Propagation Modelling - determination of the atmospheric horizontal and temporal data resolution requirements for radar coverage assessment: case studies of North Sea environments, M.Sc. thesis, Delft University of Technology, January 2015.

[4] Boone, M., De Koninklijke Marine voor de wind - Het comprimeren van datasets van atmosfeermodellen om operationeel gebruik op zee mogelijk te maken, B.Sc. thesis, Amsterdam University of Applied Sciences, June 2016 (in Dutch).

[5] Babin, S.M. and Dockey, G.D., "LKB-based evaporation duct model comparison with buoy data," Journal of Applied Meteorology, vol. 401, pp. 434-446, April 2002.