

# Storm tracking and eScience

BY LIZZIE FROUDE\*

*Environmental Systems Science Centre (ESSC), University of Reading,  
Reading RG6 6AL, UK*

A new ‘storm-tracking approach’ to analysing the prediction of storms by different forecast systems has recently been developed. This paper provides a brief illustration of the type of results/information that can be obtained using the approach. It also describes in detail how eScience methodologies have been used to help apply the storm-tracking approach to very large datasets.

**Keywords:** storm; tracking; prediction; forecast; OPENDAP; CONDOR

## 1. Introduction and background

Extratropical cyclones (often referred to simply as storms) are low-pressure weather systems that are fundamental to the weather in the mid-latitudes. In the presence of these cyclones, weather conditions are generally unsettled, wet and windy; and in their absence, the weather is more settled and dry. Extratropical cyclones can be beneficial, in that they provide a majority of the precipitation received in the mid-latitudes and are therefore important for human activities such as agriculture. They can also be very damaging, since under certain conditions they can intensify more than usual, bringing very heavy rainfall and extremely strong winds. It is therefore important that these cyclones are predicted as accurately and as far in advance as possible by numerical weather prediction (NWP).

The prediction of storms by NWP has been explored using the storm identification and tracking software TRACK (Hodges 1995, 1999, see <http://www.nerc-essc.ac.uk/~kih/TRACK/Track.html>). The software works by identifying cyclones as extrema in a time series of data and then links these points together to form trajectories of the storm tracks. TRACK has been used to develop a new ‘storm-tracking approach’ to analyse the prediction of storms by different forecast systems (Bengtsson *et al.* 2005; Froude *et al.* 2007*a,b*; Froude *in press*). The approach involves in the use of TRACK for the identification and tracking of extratropical cyclones along forecast trajectories. Statistics can then be generated to determine the rates at which the position, intensity and other properties of the forecast cyclones diverge from those of analysed cyclones with increasing forecast time. The methodology has revealed some interesting

\*lsrf@mail.nerc-essc.ac.uk

One contribution of 24 to a Discussion Meeting Issue ‘The environmental eScience revolution’.

scientific results, showing, for example, that the intensity and propagation speed of storms are more difficult to predict accurately than is the direction the storm takes and that in general forecast storms propagate too slowly.

Meteorological datasets, such as those obtained from NWP and analysed using the storm-tracking method, are continually getting larger. This is a result of increasing computer power allowing models to be run at higher resolutions and for longer time periods. The increasing size of these datasets has resulted in more distributed archiving, and it is consequently becoming more difficult to analyse these datasets at a single location. Storing the required data locally may not be possible because of the enormous amounts of disk space required, and transferring the data from its remote source to a local resource can be considerably time consuming. The vast amount of CPU required to process and analyse such large amounts of data presents another difficulty and the use of one computer for such a task may be completely infeasible.

Such difficulties arose in the storm-tracking studies of Bengtsson *et al.* (2005), Froude *et al.* (2007*a,b*) and Froude (in press) mentioned above, which analysed some very large datasets. This was particularly the case in the studies of Froude *et al.* (2007*a*) and Froude (in press), which analysed the prediction of storms by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) ensemble prediction systems (EPS; Toth & Kalnay 1993, 1997; Buizza & Palmer 1995; Molteni *et al.* 1996). In the past, NWP involved the integration of a single model from a single initial state. More recently, ensemble prediction techniques have been introduced, in which multiple integrations of a model (or multiple models) are performed. It is clear that the output from EPS will constitute dramatically larger datasets than those obtained from the older deterministic approach. The difficulties associated with analysing these large datasets motivated the development of a TRACK Web application (Froude 2008), which allows the TRACK program to be executed from a Web browser, with remotely stored datasets, using distributed computing.

This paper has two main objectives. Firstly, it aims to illustrate the type of results and information that can be obtained from the storm-tracking methodology. The results for this are taken from Froude (in press). Secondly, this paper aims to describe the TRACK Web application of Froude (2008) developed and used to help with the computation involved in the storm-tracking analysis of large datasets. The paper continues with a discussion of some of the science results of Froude (in press) in §2, a description of the TRACK Web application is given in §3 and the paper finishes with a summary and discussion of future work in §4.

## 2. Example of science results

To provide an illustration of the type of information that can be obtained from the storm-tracking analysis methodology, some results from Froude (in press) will be presented and discussed. The Froude (in press) study investigated the regional differences in the prediction of storms by the ECMWF EPS. The cyclones were identified and tracked along the 6 hourly forecast trajectories of each of the ensemble members of EPS data for the 1 year period of 6 January

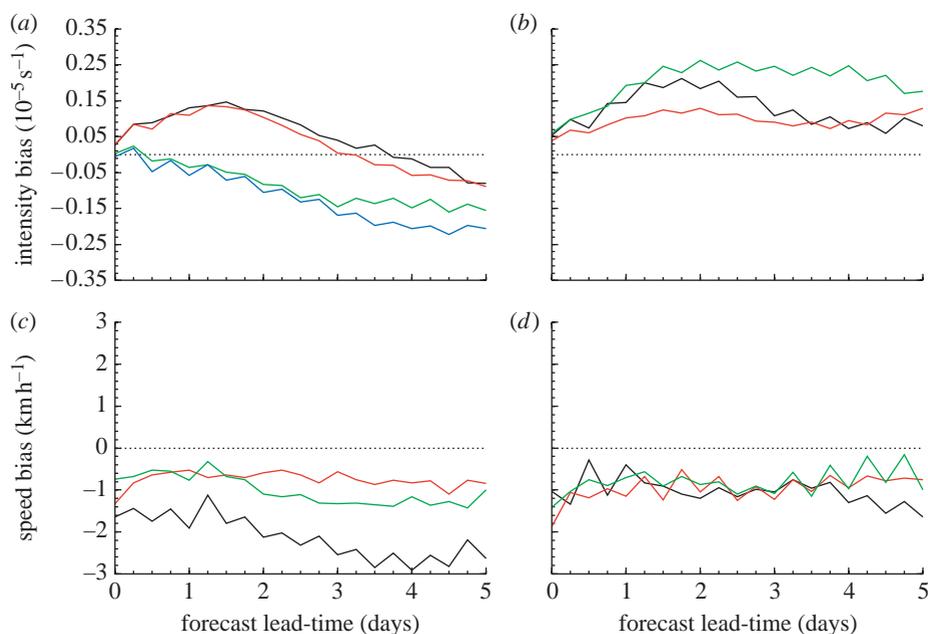


Figure 1. Biases in intensity and propagation speed of storms forecast by the ECMWF EPS from Froude (in press). (a,b) The signed intensity difference between the storms predicted by the ensemble members and the analysed storms in the Northern Hemisphere and Southern Hemisphere, respectively. (c,d) As in (a,b) but the signed difference in propagation speed. (a,c) Black, Atlantic; red, Pacific; green, Eurasia; blue, North America; (b,d) black, Atlantic, red, Pacific, green, Indian.

2005–5 January 2006. The tracking was also performed along the ECMWF operational analysis for the same time period. This resulted in a set of ensemble forecast storm tracks, which could then be validated against the analysis storm tracks. For details of how this verification was performed, please see Froude (in press).

Figure 1a,b shows the signed intensity difference between the ensemble members storm tracks and the analysis storm tracks for different regions in the Northern and Southern Hemispheres (NH, SH), respectively. A positive/negative difference corresponds to an overprediction/underprediction of the storms' intensity. The diagnostics show that the EPS overpredicts cyclone intensity over the ocean regions (Atlantic and Pacific in NH and all regions in SH) and underpredicts the intensity over the land. Figure 1c,d shows the signed difference in propagation speed between the storms predicted by the ensembled members and the analysed storms. It was not possible to show the results for North America as the data sample was insufficient for this particular diagnostic. There is a negative bias, corresponding to the forecast storms propagating too slowly, for all of the regions. The NH Atlantic region stands out, having a bias of twice the magnitude of the other regions. We believe that the biases of figure 1 are due to errors in the vertical structure (specifically the tilt) of the predicted storms. This is discussed in more detail in Froude (in press) and is being investigated as future work. The NH Atlantic region is perhaps subjected to larger biases

because of the dramatic differences between the observing network over North America compared with the Atlantic Ocean. There are large numbers of radiosonde observations over North America, but as the storms move over the ocean the observations become mainly satellite. For further discussion, please see Froude (in press).

### 3. TRACK Web application

The TRACK Web application described in Froude (2008) currently enables users to compute storm tracks from the NCEP reanalysis (Kalnay *et al.* 1996) and NCEP EPS (Toth & Kalnay 1993, 1997) datasets, which are both archived in the USA and made freely available via the Internet. A list of jobs can be constructed and executed across multiple computers to reduce computation time. The progress of each job can be monitored and, once completed, the computed storm tracks can be downloaded and plotted in a Web browser.

The Web application was written using Java Servlets/Java Server Pages (Hall 1999). It accesses the remote data using the Open-source Project for a Network Data Access Protocol (OPENDAP, <http://www.opendap.org>). OPENDAP allows data to be accessed over the Internet by using a client-server model in which the client requests some data from an OPENDAP server and the server replies by returning the data. Data analysis programs that use data access application programming interfaces (APIs) such as NETCDF can be converted to OPENDAP clients by re-linking them with the OPENDAP versions of the API libraries. Remote data provided via an OPENDAP server can then be accessed by the data analysis program in effectively the same way as locally stored datasets by using a URL instead of a filename. OPENDAP also has a subsampling facility, so that a specific part of the data can be requested by appending information to the end of the URL that references the data. This allows the user to download just the parts of the data they require, rather than downloading the entire data file. The TRACK program uses the NETCDF API and was re-linked with the OPENDAP versions of the libraries. It can now be used to compute storm tracks from remote datasets in the same way as locally stored data, but using URLs instead of filenames. The OPENDAP subsampling facility is used to request specific meteorological fields and time periods requested by the user in their job list. This use of sub-setting dramatically reduces the amount of data that need to be stored locally. For example, the NCEP EPS data files include a large number of meteorological fields at a large number of different pressure levels. For the storm-tracking analysis only mean sea-level pressure or vorticity at the 850 hPa level were required. These fields are selected with the subsampling facility rather than downloading the entire file.

The TRACK Web application allows users to submit a list of jobs to the CONDOR (Thain *et al.* 2005) pool in ESSC. CONDOR is a software system that manages a collection of jobs by making use of the computational power of machines over a network. Users can submit a list of multiple jobs to CONDOR, which chooses where and when to run them. For each job, CONDOR determines if there is a suitable machine available and, if there is, it begins to run the job on that machine. Each job in the user's job list is submitted as a separate job to the CONDOR pool and is run on a different machine. This allows a much faster throughput than using just a single machine.

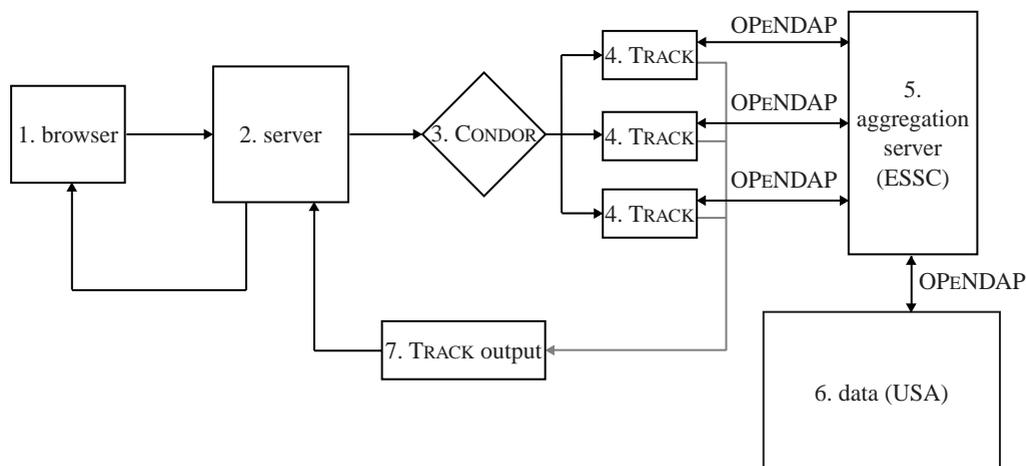


Figure 2. Flow chart to illustrate how TRACK Web application works, adapted from Froude (2008).

The NCEP re-analysis data at the Climate Diagnostic Center (CDC) is stored in yearly files (January–December). The OPENDAP subsampling facility allows users to select a time period within a given year (i.e. the same file). It is not, however, possible to select a period that begins in one year and ends in another (e.g. a December–February season) because the data for such a period are split across two files. To overcome this problem, the OPENDAP aggregation server (<http://www.opendap.org/server/agg-html/agg.html>) was used. This is a piece of software that can be used to create aggregated datasets by effectively merging individual files so they appear as one large file. These individual files do not have to be local files; they can also be remote files that are provided by an OPENDAP server. Once a number of files have been aggregated to appear as one large file, the OPENDAP subsampling facility can be used to access a section of data that overlaps multiple files. The OPENDAP aggregation server has already been installed at ESSC. The NCEP re-analysis dataset was aggregated so that it could be treated as one large 50 year file rather than as 50 smaller 1 year files. The aggregation of the data means that the user is able to run TRACK with NCEP re-analysis data from any time period between 1943 and the present.

Figure 2 shows a flow chart, from Froude (2008), illustrating how the different components of the TRACK Web application fit together. The user constructs a list of TRACK jobs in their Web browser (labelled 1), which is then sent to the server (labelled 2). The server then submits this list of jobs to the CONDOR pool in ESSC (labelled 3). CONDOR puts the jobs into a queue and then sends the jobs to different computers (labelled 4) as and when they become available. The TRACK program accesses the data using the OPENDAP protocol. The NCEP EPS data are accessed directly from the OPENDAP server (labelled 6), whereas the re-analysis data are accessed via the aggregation server at ESSC (labelled 5). Once all the jobs have finished running, the output from TRACK (labelled 7) is put onto the server for the user to download or plot. While a set of jobs are running, the user is able to check the progress of each individual job from their Web browser. For further details of the TRACK Web application, please see Froude (2008).

#### 4. Recent/future work and summary

In recent work, an even larger dataset is being analysed. This dataset is known as the THORPEX interactive grand global ensemble and consists of ensemble forecast data from 10 operational forecast centres around the world. To compare the prediction of storms by the different ensemble forecast systems will require vast amounts of data processing. In order to reduce the computation time of this data processing, the University of Reading Campus Grid is being used. This consists of a CONDOR pool of approximately 150 Linux machines and therefore dramatically speeds up the data processing. The importance of CONDOR to the storm-track analysis work in general cannot be overemphasized. Without it, the data processing would have been extremely difficult with the facilities available. CONDOR was particularly well suited to the processing of the EPS data, since the storm tracking for each ensemble member could be performed on a different computer.

In summary, the storm-tracking methodology (Bengtsson *et al.* 2005; Froude *et al.* 2007*a,b*; Froude *in press*) required very large samples of data. Without the use of eScience methodologies, it would not have been possible to store and analyse such large amounts of data and thereby obtain new and detailed information about the prediction of storms. We are currently working with the oil/gas consultancy Schlumberger (<http://www.slb.com>) to explore how storm prediction information can be incorporated into their information systems. This information is potentially very valuable to the management of operations both on- and offshore. It is anticipated that the use of eScience will help with this task considerably.

The author would like to thank ECMWF and NCEP for providing the data for us to carry out this research. Thanks also go to Prof. Robert Gurney, Dr Kevin Hodges and Prof. Lennart Bengtsson for their helpful comments and advice. The author would also like to acknowledge both NERC and Schlumberger for funding this research.

#### References

- Bengtsson, L., Hodges, K. I. & Froude, L. S. R. 2005 Global observations and forecast skill. *Tellus A* **57**, 515–527. (doi:10.1111/j.1600-0870.2005.00138.x)
- Buizza, R. & Palmer, T. N. 1995 The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**, 1434–1456. (doi:10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2)
- Froude, L. S. R. 2008 Storm tracking with remote data and distributed computing. *Comput. Geosci.* **34**, 1621–1630. (doi:10.1016/j.cageo.2007.11.004)
- Froude, L. S. R. *In press*. Regional differences in the prediction of extratropical cyclones by the ECMWF ensemble prediction system. *Mon. Weather Rev.* (doi:10.1175/2008MWR2610.1)
- Froude, L. S. R., Bengtsson, L. & Hodges, K. I. 2007*a* The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Mon. Weather Rev.* **135**, 315–333. (doi:10.1175/MWR3274.1)
- Froude, L. S. R., Bengtsson, L. & Hodges, K. I. 2007*b* The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Mon. Weather Rev.* **135**, 2545–2567. (doi:10.1175/MWR3422.1)
- Hall, M. 1999 Servlets and Java Server Pages (JSP) 1.0: a tutorial. See <http://www.apl.jhu.edu/~hall/java/Servlet-Tutorial/>.

- Hodges, K. I. 1995 Feature tracking on the unit sphere. *Mon. Weather Rev.* **123**, 3458–3465. (doi:10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2)
- Hodges, K. I. 1999 Adaptive constraints for feature tracking. *Mon. Weather Rev.* **127**, 1362–1373. (doi:10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2)
- Kalnay, E., Kanamitsu, R. & Kistler, R. 1996 The NCEP/NCAR 40-year re-analysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471. (doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Molteni, F., Buizza, R., Palmer, T. N. & Petroliagis, T. 1996 The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119. (doi:10.1002/qj.49712252905)
- Thain, D., Tannenbaum, T. & Livny, M. 2005 Distributed computing in practice: the Condor experience. *Concurr. Comput. Pract. Exp.* **17**, 323–356. (doi:10.1002/cpe.938)
- Toth, Z. & Kalnay, E. 1993 Ensemble forecasting at the NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330. (doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2)
- Toth, Z. & Kalnay, E. 1997 Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **125**, 3297–3319. (doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2)