

# *IMILAST: a community effort to intercompare extratropical cyclone detection and tracking algorithms*

Article

Published Version

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Coccozza, A., Dacre, H., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L. R., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L. and Wernli, H. (2013) IMILAST: a community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, 94 (4). pp. 529-547. ISSN 0003-0007 doi: <https://doi.org/10.1175/BAMS-D-11-00154.1> Available at <https://centaur.reading.ac.uk/32735/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://dx.doi.org/10.1175/BAMS-D-11-00154.1>

To link to this article DOI: <http://dx.doi.org/10.1175/BAMS-D-11-00154.1>

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# IMILAST

## A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms

BY URS NEU, MIRSEID G. AKPEROV, NINA BELLENBAUM, RASMUS BENESTAD, RICHARD BLENDER, RODRIGO CABALLERO, ANGELA COCOZZA, HELEN F. DACRE, YANG FENG, KLAUS FRAEDRICH, JENS GRIEGER, SERGEY GULEV, JOHN HANLEY, TIM HEWSON, MASARU INATSU, KEVIN KEAY, SARAH F. KEW, INA KINDEM, GREGOR C. LECKEBUSCH, MARGARIDA L. R. LIBERATO, PIERO LIONELLO, IGOR I. MOKHOV, JOAQUIM G. PINTO, CHRISTOPH C. RAIBLE, MARCO REALE, IRINA RUDEVA, MAREIKE SCHUSTER, IAN SIMMONDS, MARK SINCLAIR, MICHAEL SPRENGER, NATALIA D. TILININA, ISABEL F. TRIGO, SVEN ULBRICH, UWE ULBRICH, XIAOLAN L. WANG, AND HEINI WERNLI

An intercomparison experiment involving 15 commonly used detection and tracking algorithms for extratropical cyclones reveals those cyclone characteristics that are robust between different schemes and those that differ markedly.

Extratropical cyclones are fundamental meteorological features and play a key role in a broad range of weather phenomena. They are a central component maintaining the global atmospheric energy, moisture, and momentum budgets. They are on the one hand responsible for an important part of our water supply, and on the other are intimately linked with many natural hazards affecting the middle and high latitudes (wind damage, precipitation-related flooding, storm surges, and marine storminess). Thus, it is important to provide for society an accurate diagnosis of cyclone activity, which includes a baseline climatology of extratropical storms (e.g., Hoskins and Hodges 2002) and also estimates of likely future changes therein. While future changes in some cyclone characteristics such as the total number of cyclones might be small, major signals may still be expected in specific characteristics such as regional storm frequency, intensity, and location (e.g., Leckebusch et al. 2006; Wang et al. 2006; Löptien et al. 2008; Bengtsson et al. 2009; Pinto et al. 2009; Raible et al. 2010; Schneidereit et al. 2010).

Identifying and tracking extratropical cyclones might seem, superficially, to be a straightforward activity, but in reality it is very challenging. In this regard it is useful to compare the situation with tropical cyclones, which possess characteristics that make them relatively easy to identify and track: they occur rarely (making misassociation unlikely), are generally symmetric and slow moving, and have a relatively unambiguous structure. Extratropical cyclones are in a sense the “opposite”: they are much more common, can range greatly in shape and structure (are often asymmetric), differ rather more in size (with diameters ranging from about 100 to well over 1,000 km), and have translational velocities that can vary greatly. Identifying the same physical feature at different times (i.e., tracking) is also complicated by the fact that a single cyclone will sometimes split into separate features, and sometimes two will merge into one. Furthermore, extratropical cyclones occur in very diverse synoptic situations, with some being confined to lower-tropospheric levels and others extending through great depth. This great complexity

reveals why there is no single commonly agreed upon scientific definition of what an extratropical cyclone is, and also why there exists a range of ideas and concepts regarding how to identify and track them.

It could be argued that an in-depth manual reanalysis of cyclone trajectories based on weather maps reconstructed using all available data (e.g., Hewson et al. 2000) would provide the best tracks. However, given the lack of data in some regions and the complexity of cyclone development, such activities inevitably involve some subjective choices being made by the analyst. So there is no accepted single “truth” regarding specific cyclone tracks. Moreover, while careful manual tracking might nonetheless be considered optimal, for quantifying the behavior of all cyclones over many decades it is clearly not feasible, and the application of automated detection and tracking methods to reanalysis data—the thrust of this paper—is indispensable.

Although automated schemes are objective and reproducible, they are based on different understandings of what best characterizes a cyclone. Application of different algorithms provides results that are remarkably similar in some aspects but may be very different in others. Thus, depending on what one is looking for, the selection of a particular method can significantly affect one’s conclusions,

for example, regarding trends in cyclone intensity. Raible et al. (2008), for example, demonstrated that three different algorithms applied to the same input data showed similar interannual variability but considerable differences in total cyclone numbers. While this comparison showed similar trend patterns in the Atlantic, trends found for the Pacific even differed in sign. Indeed, method-associated uncertainties in some cases are quite large, such that equivalent scientific studies may find contradictory climate change signals even when using identical input data (Trigo 2006; Ulbrich et al. 2009). Therefore, it is crucial to know those aspects for which the results are robust with regard to the method used, and those aspects for which there will be large method-related uncertainties. The project Intercomparison of Mid Latitude Storm Diagnostics (IMILAST) is the first comprehensive assessment focusing on this method-related uncertainty.

**STRATEGY AND METHODS.** Over the last two decades, many numerical identification and tracking algorithms have been developed (Murray and Simmonds 1991; Hodges 1995; Serreze 1995; Blender et al. 1997; Sinclair 1997; Simmonds et al. 1999; Lionello et al. 2002; Benestad and Chen 2006; Trigo 2006; Wernli and Schwierz 2006; Akperov

**AFFILIATIONS:** NEU—ProClim, Swiss Academy of Sciences, Bern, Switzerland; AKPEROV AND MOKHOV—A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, Russia; BELLENBAUM, PINTO, AND ULBRICH—Institute of Geophysics and Meteorology, University of Cologne, Cologne, Germany; BENESTAD—Norwegian Meteorological Institute, Oslo, Norway; BLENDER AND FRAEDRICH—Institute of Meteorology, University of Hamburg, Hamburg, Germany; CABALLERO AND HANLEY—Department of Meteorology, Stockholm University, Stockholm, Sweden; COCOZZA AND REALE—Di.S.Te.B.A., University of Salento, Lecce, Italy; DACRE—Department of Meteorology, University of Reading, Reading, United Kingdom; FENG AND WANG—Climate Research Division, Environment Canada, Toronto, Ontario, Canada; GRIEGER, SCHUSTER, AND ULBRICH—Institute of Meteorology, Freie Universität Berlin, Berlin, Germany; GULEV AND TILININA—P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, and Moscow State University, Moscow, Russia; HEWSON—European Centre for Medium-Range Weather Forecasts, Reading, and Met Office, Exeter, United Kingdom; INATSU—Graduate School of Science, Hokkaido University, Sapporo, Japan; KEAY—School of Earth Sciences, University of Melbourne, and Bureau of Meteorology, Melbourne, Victoria, Australia; KEW—Royal Netherlands Meteorological Institute, De Bilt, Netherlands; KINDEM—Bjerknes Centre for Climate Research, Bergen, Norway; LECKEBUSCH—School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom; LIBERATO—Instituto

Dom Luiz, University of Lisbon, Lisbon, and School of Sciences and Technology, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal; LIONELLO—Di.S.Te.B.A., University of Salento, and CMCC, EuroMediterranea Center on Climate Change, Lecce, Italy; RAIBLE—Climate and Environmental Physics, and Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland; RUDEVA—P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia, and School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia; SIMMONDS—School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia; SINCLAIR—Embry-Riddle Aeronautical University, Prescott, Arizona; SPRENGER AND WERNLI—Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland; TRIGO—Instituto Dom Luiz, University of Lisbon, and Institute of Meteorology I. P., Lisbon, Portugal

**CORRESPONDING AUTHOR:** Urs Neu, ProClim, Swiss Academy of Sciences, Schwarztorstrasse 9 CH-3007 Bern, Switzerland  
E-mail: urs.neu@scnat.ch

*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-11-00154.1

Supplements A and B to this article are available online (10.1175/BAMS-D-11-00154.2)

In final form 31 August 2012  
©2013 American Meteorological Society

et al. 2007; Rudeva and Gulev 2007; Inatsu 2009; Kew et al. 2010; Hewson and Tittley 2010; Hanley and Caballero 2012). According to different perceptions of what a cyclone is, tracking may be performed utilizing a number of atmospheric variables (Hoskins and Hodges 2002). One of the most widely discussed algorithmic differences relates to the choice of mean sea level pressure (MSLP) or lower-tropospheric vorticity as a basic identification/tracking metric (e.g., Sinclair 1994; Hodges et al. 2003; Rudeva and Gulev 2007; Ulbrich et al. 2009). These options reflect the different characteristics that one might focus on when examining cyclones: while vorticity is more focused on the wind field and contains more information on the high-frequency synoptic scale, central pressure is linked to the mass field and represents the low-frequency scale better (Hodges et al. 2003). This can lead, for example, to different estimated positions of the cyclone center, since in a westerly airflow the vorticity-based center can sometimes be located a few hundred kilometers equatorward of the related pressure minimum (Sinclair 1994). On other occasions different features are identified, because mobile vorticity centers are not necessarily associated with a pressure minimum. There are many other metrics for assessing cyclone activity, as discussed, for example, by Raible et al. (2008) and Ulbrich et al. (2009).

To quantify the impact on extratropical storm analysis of using different methods, an intercomparison experiment was initiated. In the first activity, on which the results presented in this paper are based, all participating groups computed cyclone tracks (for definitions, see sidebar) for the same period using the same input—the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) dataset (Dee et al. 2011).<sup>1</sup> Space–time resolution of the input data may have a significant impact on cyclone statistics (Blender and Schubert 2000; Pinto et al. 2005; Jung et al. 2006), and high resolution is essential to help capture the full life cycles of cyclones and to ensure that even small cyclonic windstorms can be identified (Pinto et al. 2005; Hewson and Tittley 2010). For this first intercomparison activity, for reasons of availability, we used 1.5° spatial resolution and 6-hourly temporal resolution (except for one method

that uses 12-hourly data) for a 20-yr period from 1 January 1989 to 31 March 2009.

Besides being related to method differences, track uncertainty can also arise from reanalysis inadequacies. However, this aspect has been discussed elsewhere in the literature (e.g., Wang et al. 2006; Benestad and Chen 2006; Raible et al. 2008; Hodges et al. 2011) and is not analyzed here.

While full descriptions of the methods involved in this intercomparison are presented in earlier publications, Table 1 provides a brief outline and supplement A (“General description of the different methods participating in the intercomparison”; available online at <http://dx.doi.org/10.1175/BAMS-D-11-00154.2>) a more extensive overview of method characteristics. Methods differ in a number of aspects, including, for example, variables used for cyclone identification

## DEFINITIONS RELATED TO CYCLONES AND TRACKING

**Cyclone:** There is no accepted universal definition of what a cyclone is or where its exact position is. In this study, “cyclone” refers to a point (the cyclone “center”) identified on the Earth’s surface at a certain time through different approaches, often by searching for a minimum of MSLP or a maximum of lower-tropospheric cyclonic vorticity.

**Track:** A cyclone track consists of a series of cyclones identified in sequential time steps at adjacent locations, which are deemed to represent the same physical feature in reality.

**Number of cyclones:** Count of cyclone tracks over a certain region (globe, hemisphere, etc.).

**Cyclone center counts:** The numbers of cyclone centers identified at each time step, summed over all time steps (in this study, only including cyclones that make up tracks with lifetime  $\geq 24$  h).

**Cyclone center density:** Percentage of cyclone occurrence per time step and per unit area of (1000 km<sup>2</sup>). For example, if at a grid location the cyclone center density is 10%, then in an area of 1000 km  $\times$  1000 km in 100 time steps we find 10 cyclones. A value  $>100\%$  means there is more than one cyclone per time step in that area on average (synonymous with cyclone frequency).

**Track density:** Number of tracks passing a grid cell (with repeated entries of the same track being counted as one).

<sup>1</sup> The key criteria for selecting this input dataset were that it be based on a state-of-the-art model (gridded reanalysis dataset) with a four-dimensional variational data assimilation (4DVAR) scheme, be easy to access, and have a high spatial resolution. Recent intercomparisons (Allen et al. 2010; Hodges et al. 2011) have demonstrated that cyclone characteristics in ERA-Interim are quite comparable with those revealed by, for example, Modern Era Retrospective–Analysis (MERRA) and National Centers for Environmental Prediction–Climate Forecast System Reanalysis (NCEP–CFSR).

(MSLP, vorticity, etc.), the cyclone identification procedures themselves, elimination criteria (thresholds) to filter out weak or “artificial” low pressure systems (e.g., by requiring a minimum travel distance), and algorithms to combine the cyclone centers into a track. In addition to there being disparities between the algorithms due to different concepts of a cyclone, differences in methods also relate to some degree to the different kinds of phenomena the original authors wanted to study. For an intercomparison of different algorithms, it would be desirable to distinguish between these two origins of differences and compare only algorithms that address precisely the same problem. However, this distinction is not clear cut, because algorithm settings (e.g., a threshold for minimum pressure gradient) could have been introduced to match a corresponding cyclone definition (i.e., if the definition contains a gradient threshold) or just to eliminate shallow heat lows. Thus, in our intercomparison we included methods designed to search for extratropical cyclones in general, but excluded algorithms that were clearly designed for very specific types (e.g., looking for extreme cyclones only or for polar lows) to prevent a possible exaggeration of method differences. A further area where one could

argue for standardization is data preprocessing, as performed in some methods, since such steps can lead, for example, to smoothing, which influences the results. However, preprocessing is performed with different purposes in mind, as, for example, addressing inhomogeneous grid field size when using latitude–longitude grids, and is therefore difficult to standardize. Furthermore, preprocessing is only one of many choices and is an integral part of some algorithms. Note finally that some methods originally focused on very different regions. One could argue that this might be a criterion for noninclusion, but given that the fundamental physics of extratropical cyclones is similar everywhere we did not exclude on this basis.

All the schemes participating in this intercomparison are in common use. In that sense our comparison is of most value when they are used in their “standard form” (excepting some adaptations made to accommodate the input data resolution). Therefore, we did not apply any far-reaching standardization. The slight downside is that this approach may increase the difficulty of understanding the reasons behind discrepancies in cyclone climatologies. Moreover, it is difficult to quantify the full range of impacts

**TABLE 1. Different methods and some key characteristics: “variable used” (MSLP: mean sea level pressure; VORT: vorticity or Laplacian of MSLP; VORT Z850: vorticity at 850 hPa as computed by ERA-Interim; Z850: geopotential height at 850 hPa; grad.: gradient of MSLP; min: minimum), and “terrain filtering” (>1000 m; all cyclones positioned over terrain higher than 1,000 m MSL are eliminated).**

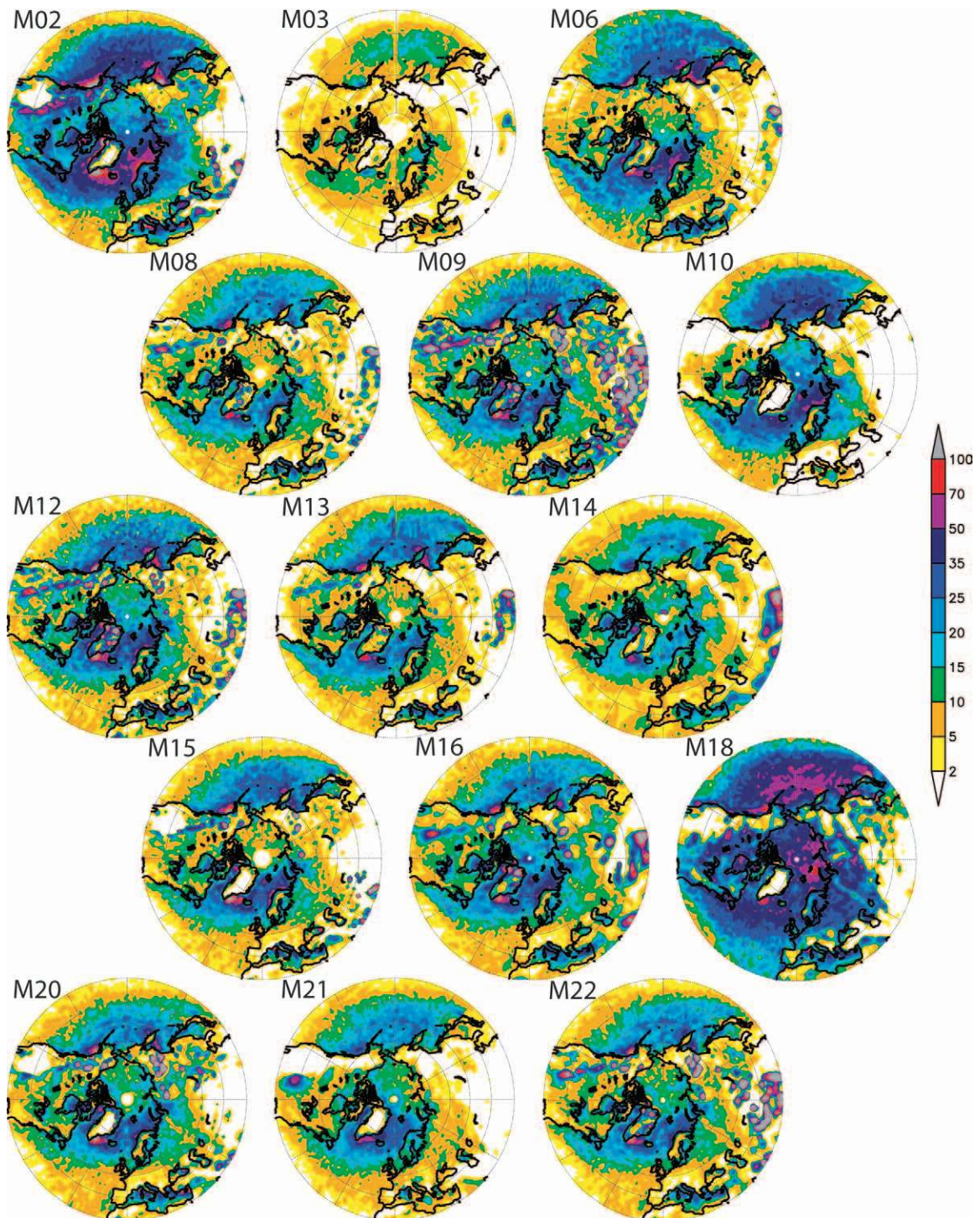
Code*	Main references for method description	Variable used	Terrain filtering
M02	Murray and Simmonds (1991), Pinto et al. (2005)	MSLP (min), VORT	>1500 m
M03	Benestad and Chen (2006)	MSLP (min, grad.)	none
M06	Hewson et al. (1997), Hewson and Titley (2010)	MSLP (min), VORT, wind, fronts	Terrain-following
M08	Trigo (2006)	MSLP (min, grad.)	none
M09	Serreze (1995), Wang et al. (2006)	MSLP (min, grad.), VORT	none
M10	Murray and Simmonds (1991), Simmonds et al. (2008)	MSLP (min), VORT	>1000 m
M12	Zolina and Gulev (2002), Rudeva and Gulev (2007)	MSLP (min)	none
M13	Hanley and Caballero (2012)	MSLP (min)	>1500 m
M14	Kew et al. (2010)	Z850 (min, contour)	none
M15	Blender et al. (1997), Raible et al. (2008)	MSLP (min)	>1000 m
M16	Lionello et al. (2002)	MSLP (min)	none
M18	Sinclair (1994, 1997)	Z850 VORT	>1000 m
M20	Wernli and Schwerz (2006)	MSLP (min)	>1500 m
M21	Inatsu (2009)	Z850 VORT	none
M22	Bardin and Polonsky (2005), Akperov et al. (2007)	MSLP (min, contour)	none

\*Code numbers were assigned at the beginning of the project to research groups interested in participation. A few groups have not (yet) contributed a dataset, and others have since ceased activities in cyclone tracking. Furthermore, some groups use almost identical algorithms, in which case duplicates have been removed. The original code assignment was retained to guarantee compatibility of publications for the whole project duration. Therefore, code numbers are not continuous.

from method differences and search for reasons at the same time. For the latter, sensitivity studies that, for example, change parameter settings in a specific method and compare the corresponding results

would be needed. Such sensitivity studies are foreseen as a next step in the IMILAST project.

However, we did standardize one aspect throughout—namely, a minimum lifetime. This was



**FIG. 1.** Total cyclone center density in the NH for cyclones lasting 24 h or more (percentage of cyclone occurrence per time step and area of 1000 km<sup>2</sup>; see sidebar) for all detection and tracking methods in DJF. The results of methods M03, M09, and M14 (which extrapolate the input data to a higher resolution for their calculation) are interpolated to the 1.5° × 1.5° grid.

fixed to 24 h for all methods (i.e., to be retained a cyclone has to exist for a minimum of five 6-hourly time steps, or three 12-hourly time steps for method M06). This standardization was seen as permissible and desirable because in most methods the lifetime threshold is a parameter of arbitrary choice with a rather straightforward impact: a shorter lifetime

threshold would increase the number of cyclones considerably and a longer one would decrease it. The inclusion or exclusion of cyclones over elevated topography has not been standardized a priori because this is an integral part of some methods and not necessarily problem oriented. However, for comparisons of hemispheric statistics (see sections below)

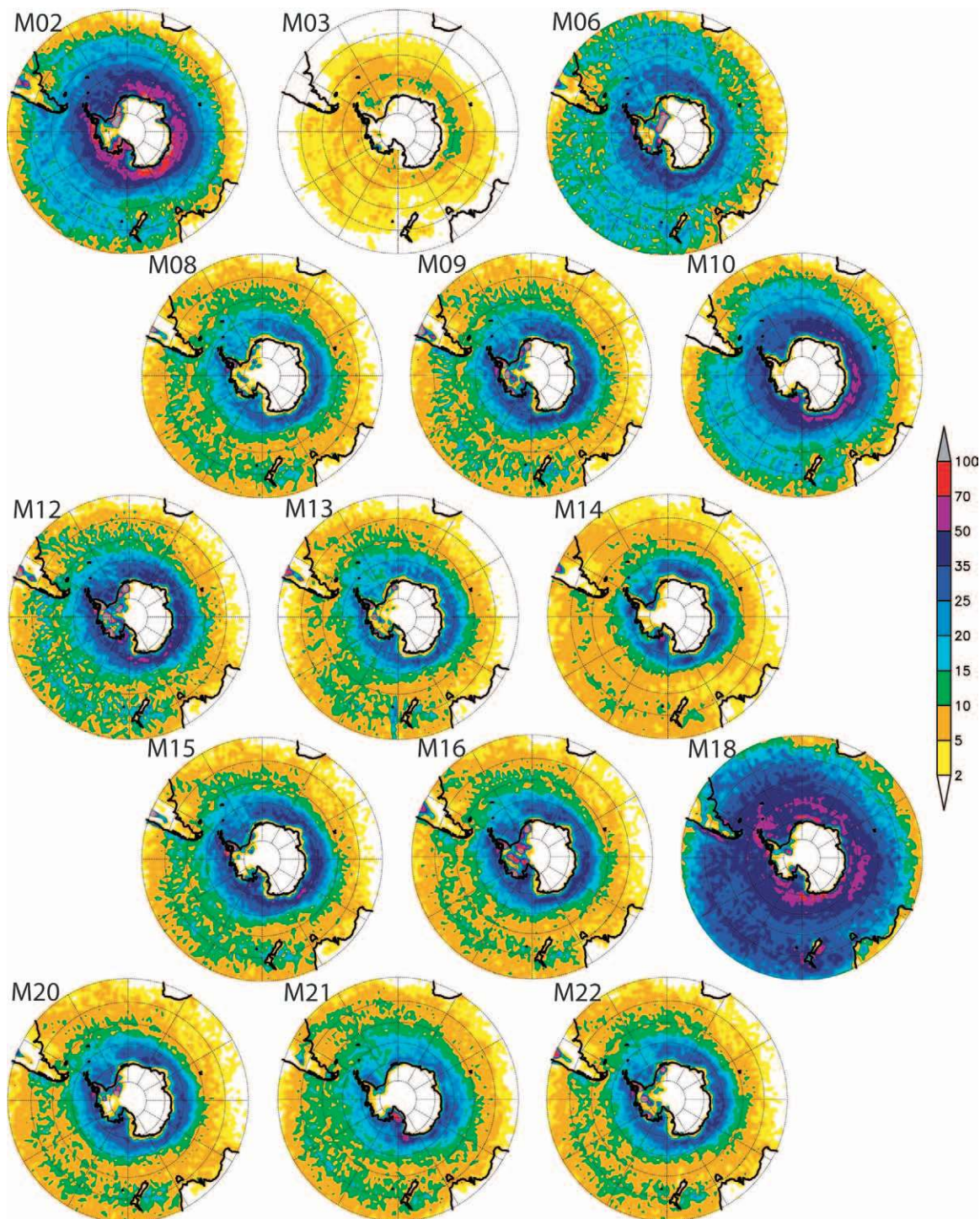


FIG. 2. As in Fig. 1, but for the SH in JJA.



a standardized set of track data where cyclones over mountainous terrain were eliminated a posteriori has been used. We did not introduce any other standardization, as this would not have been straightforward and indeed was seen as rather arbitrary.

In the following we compare and analyze the track datasets derived by the 15 different methods, focusing on a number of important cyclone characteristics: climatological frequency (section “Climatology of midlatitude cyclone characteristics”), life cycle aspects (section “Cyclone life cycle characteristics”), case studies (section “Case studies”), and interannual variability and trends (section “Interannual variability and trends”). A short discussion concludes this paper. As stated above we have no best-track datasets to define truth in this study. Therefore, the study cannot assess any kind of “quality” of individual methods, so the reader should view results accordingly and not be prejudiced for or against method(s) exhibiting outlier behavior. The purpose of this experiment is to assess the range of variability for different cyclone characteristics and to highlight the robustness of different characteristics with regard to method differences.

### **CLIMATOLOGY OF MIDLATITUDE CYCLONE CHARACTERISTICS.**

In this section we examine some aspects of cyclone climatologies obtained from the different detection and tracking methods. Figures 1 and 2 present the spatial patterns of cyclone center density (or cyclone frequency) for each method for the winter season of each hemisphere. Overall, a qualitative agreement in the spatial structures is found, as all methods identify the major oceanic cyclone activity areas east of Greenland and along the Scandinavian coast line, the two centers in the North Pacific in the Northern Hemisphere (NH), and regional maxima in the Indian Ocean sector, the Amundsen Sea, and the Drake Passage in the Southern Hemisphere (SH). In contrast, there are noteworthy discrepancies throughout the Mediterranean, which is of particular societal relevance given the high population density here. Some specific differences between individual methods can be explained—for example, a somewhat smoother pattern in M21 due to the nature of the preprocessing prior to the cyclone identification and tracking. Any kind of smoothing of input data has a similar effect as using lower-resolution data. Different studies have shown that using lower-resolution data in general decreases the number of detected cyclones (e.g., Blender and Schubert 2000; Pinto et al. 2005). The generally low numbers detected by method M03 relate mainly to one aspect of its unique approach: the application of a cutoff to retain only the 25 cyclones

with lowest central pressure in each hemisphere and at each time step (if the number initially detected exceeds 25). Test calculations for a period of 100 days with different settings subsequently showed that a higher cutoff value of 50 instead of 25 cyclones approximately doubled the number of cyclones detected, which shows that the cutoff threshold of 25 indeed considerably reduces the number of detected cyclones.

Although methods M02 and M10 are both based on the algorithm of Murray and Simmonds (1991), they include different updates and parameter settings (Pinto et al. 2005; Simmonds et al. 2008) and do not show a strong level of agreement that we might expect. How far the differences in parameter settings are related to the fact that the two algorithms were developed for different hemispheres is not clear (since cyclones are the same physical phenomenon in both hemispheres) and has to be evaluated. Large deviations are apparent between the various methods over mountain areas. This is mainly because there are different strategies for dealing with mountains; in some methods, for example, such regions are excluded a priori.

Despite qualitatively consistent spatial patterns across different methods, quantitative differences in the total numbers of extratropical cyclones are relatively large in both hemispheres. For the NH, total numbers range from about 6,000 (M03) to 21,000 (M18) during winter [December–February (DJF); first row in Table 2]. In summer [June–August (JJA); first column in Table 2] the range is between about 5,000 (M03) and 28,000 (M09). Interestingly, the seasonal (winter to summer) changes vary quite strongly between methods. Some methods increase the number of cyclones from winter to summer by more than 50%, whereas others decrease it by a few percent. Both discrepancies are related to differences in algorithms that are not easy to disentangle. One factor is the extent to which shallow cyclones (of which some can be attributed to summertime heat lows) are excluded. This influences both the total number of cyclones found and the seasonal changes. Other factors are, for example, the inclusion or exclusion of so-called open systems (without closed pressure contours), which influences primarily the total number, or the choice of a minimum distance between two cyclone centers, which influences the total number but might also influence the difference between summer and winter because cyclones tend to be larger in winter.

In the SH, somewhat smaller ranges and deviations of total numbers are found (Table 3, first row and column for summer and winter, respectively). More cyclone tracks are on average detected in austral

**TABLE 2. Number of cyclones in the NH (30°–90°N) for summer (JJA, first column) and winter (DJF, first row) detected by each method, and track agreement between methods for summer (lower-left triangular matrix) and winter (top-right triangular matrix). Values denote a nominal percentage agreement (relative to the lower number of tracks produced by the two methods) when both methods detect a track at a similar place and time (see supplement B “Method of track-to-track comparison” for more details). Values ≥50% are shaded (blue for winter, red for summer) with dark shading for ≥70%. In deriving this table, mountain areas (>1,500 m MSL) have been excluded.**

Method	JJA	M02	M03	M06	M08	M09	M10	M12	M13	M14	M15	M16	M18	M20	M21	M22
DJF	×100	147	57	205	95	168	111	124	72	70	120	111	214	158	105	102
M02	123	100	68	53	65	52	60	53	67	66	61	57	50	45	39	59
M03	51	52	100	72	68	74	67	66	54	50	69	65	68	67	41	63
M06	207	51	63	100	68	49	65	59	71	66	60	60	44	45	61	61
M08	125	40	61	56	100	80	63	67	67	64	70	69	62	70	35	65
M09	285	55	73	48	77	100	66	66	75	74	71	77	45	60	38	80
M10	99	38	52	62	57	71	100	55	64	60	58	55	63	55	34	55
M12	282	51	62	44	65	50	60	100	71	65	56	64	50	58	31	62
M13	82	46	46	61	59	74	47	68	100	53	68	68	65	67	39	69
M14	82	48	43	60	59	76	45	71	45	100	66	70	65	68	39	65
M15	132	47	62	50	53	69	50	59	55	55	100	57	55	57	36	61
M16	155	44	60	49	61	74	56	66	61	66	51	100	56	69	33	69
M18	183	39	54	42	43	40	57	41	50	52	40	40	100	42	48	57
M20	236	50	65	42	67	60	62	53	66	71	58	67	38	100	35	72
M21	87	42	52	61	57	68	58	58	44	44	56	55	55	59	100	32
M22	147	39	44	48	37	52	33	47	42	47	39	38	32	46	35	100

**TABLE 3. As in Table 2, but for the SH (30°–90°S).**

Method	JJA	M02	M03	M06	M08	M09	M10	M12	M13	M14	M15	M16	M18	M20	M21	M22
DJF	×100	89	43	177	91	129	104	154	60	58	88	88	222	113	77	72
M02	133	100	60	56	51	57	55	54	55	57	56	48	59	49	37	56
M03	38	54	100	70	63	71	65	57	44	47	69	55	72	60	39	54
M06	204	55	77	100	62	54	63	41	63	66	65	58	50	51	68	66
M08	91	44	70	70	100	75	61	65	64	68	65	60	60	62	37	71
M09	120	39	77	62	79	100	61	57	70	74	76	73	51	64	40	81
M10	120	37	72	65	66	58	100	53	56	62	59	55	66	51	41	63
M12	108	41	61	59	63	63	55	100	68	73	62	71	42	60	40	71
M13	59	59	42	72	54	57	56	57	100	53	61	62	61	62	35	61
M14	61	54	55	74	74	79	70	68	46	100	63	65	67	69	38	65
M15	111	42	79	65	74	72	73	59	57	74	100	56	64	60	39	66
M16	86	46	61	68	67	77	63	66	52	71	69	100	56	64	34	69
M18	244	45	80	51	67	61	68	58	64	73	64	65	100	50	55	63
M20	101	40	66	61	65	70	56	56	53	72	63	66	59	100	36	71
M21	103	41	73	66	62	62	60	49	52	73	61	60	67	53	100	34
M22	78	47	63	71	73	82	69	69	52	71	75	73	69	71	66	100

winter than in austral summer (by 9 out of the 15 schemes), probably because heat lows play a more minor role in the SH due to reduced landmass.

In Tables 2 and 3 we show the results of a track-by-track comparison following the approach of Blender and Schubert (2000) (for a description see

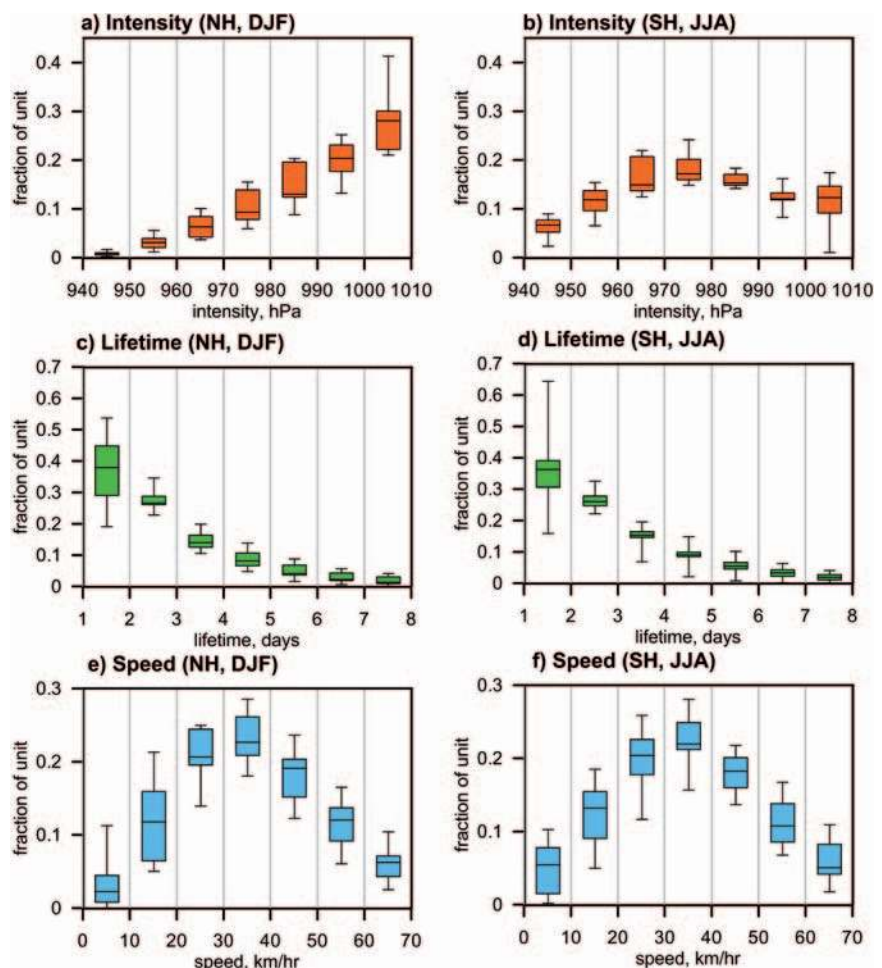
supplement B “Method of track-to-track comparison,” available online at <http://dx.doi.org/10.1175/BAMS-D-11-00154.2>) that matches up individual cyclone trajectories generated by the different methods. Overall, the matching rate ranges from roughly 50% to 70%. This seems reasonable in view of the differences in the methods’ approaches. The lower matching values for M21 might be again attributable to its preprocessing of the input data. It is worth noting that M06, which is the only method using 12-hourly input data, exhibits “average” matching relative to all other methods, suggesting that the larger time resolution does not produce particularly large differences. On the other hand, methods M02 and M10, both based on the same initial algorithm, do not exhibit particularly high matching rates. This shows that differences between methods cannot be immediately attributed to specific method features but are more likely the result of a complex interplay between the different approaches, the different threshold parameters used, and the different thresholds applied with those parameters.

In general, the methods agree better for winter cyclones than for summer cyclones in both hemispheres. Part of the reason for this might be that cyclones in winter are deeper and thus more easily detected.

### CYCLONE LIFE CYCLE CHARACTERISTICS.

In this section, we look at statistics of the life cycles of the detected cyclones. Figure 3 shows the normalized occurrence distributions of cyclone intensity

(minimum central pressure), lifetime, and propagation speed for the respective winter season in both hemispheres. The distribution of intensities (Figs. 3a,b) does not exhibit particularly large variations across methods. The largest variability occurs for the weakest category. The application of a statistical [Kolmogorov–Smirnov (k–s)] test shows that in the NH four schemes (M02, M03, M13, and



**FIG. 3.** Normalized distribution and statistical spread between methods of different cyclone life cycle characteristics, in box-whisker format: (a),(b) cyclone intensity (= minimum central pressure), (c),(d) lifetime, and (e),(f) speed (= mean propagation speed of cyclone center) for the (left) NH (DJF) and (right) SH (JJA) winter season (box-whisker plots). Each bar indicates the fraction of cyclones detected in the parameter range denoted by the x-axis labels on either side. The horizontal line indicates the mode of all methods and the box the standard deviation; the whiskers extend to the maximum and minimum values. For example, in the upper-left panel, the bar between 970 and 980 hPa indicates that on average about 10% (= fraction of 0.10 indicated on the y axis) of the cyclones detected in NH winter have a minimum central pressure between 970 and 980 hPa. The standard deviation of this value between methods is 8%–13%; the values for all methods are between 6% and 15%. The tails of the distributions beyond the given ranges are not shown; they include, for example, extreme values, which are of importance but require special analysis.

M18) have maximum intensity distributions that are significantly different (95% level) from those of the other methods. In the SH only two of them (M02 and M18) have distributions that are significantly different from the others.

The percentage of “deep” cyclones (defined as those with core pressure <960 hPa in the NH and <950 hPa in the SH, with the different thresholds reflecting deeper cyclones seen generally in SH) compared to the total number varies from 2% to 8% (M18 and M22, respectively) in NH winter, and from 4% to 12% (M18 and M13, respectively) in the SH winter. The low percentage of deep cyclones in method M18 can be explained by the highest overall winter number of detected cyclones by this method (see Tables 2 and 3), since methods generally have better skill in identifying deep cyclones compared to shallow ones. Therefore, high numbers in certain methods might plausibly originate from there being higher numbers of moderate and shallow cyclones. However, while this reasoning seems adequate for M18, this result does not apply generally. For example, only a few methods showing the highest fraction of very weak cyclones demonstrate high total numbers at the same time. Thus, there must be other important method-related influences on this distribution. Moreover, the exclusion of cyclones over mountainous terrain does not result in statistically significant changes in the distributions (not shown). One exception to this picture was a slight shift toward deeper cyclones, especially in winter, as we would expect since cyclones over high terrain are shallower in general.

The analysis of the lifetime distribution (Figs. 3c,d) also shows the largest spread for short-living cyclones. Those schemes (M03, M09, and M21) that produce a high percentage of short-living transients (1–2 day) in both NH and SH winter might, for example, be (i) more restrictive in capturing the first and/or the last stages of cyclone lifespan or (ii) tend to repeatedly produce tracks associated with short-living local and often weak depressions. For those methods showing remarkably smaller counts of short-living cyclones and higher fractions of longer-living transients (M02 in NH and SH, M12 and M14 in NH), the respective opposite might apply. However, there is no evidence from our results that either of the two suggested reasons above are a major factor: on the one hand, if the main reason for more short-living cyclones were related to the identification of cyclones at an early stage, one would expect that the schemes using relative vorticity would be more skillful in identifying cyclones at a very early

stage and, thus, demonstrate potentially a somewhat longer lifetime. However, a robust corresponding pattern discriminating methods using vorticity or MSLP could not be detected in the analysis (not shown). On the other hand, if the reason was related to the identification of short-living local depressions, one would expect the three methods producing high fractions of short-living cyclones to be among those with high numbers of shallow cyclones, but this is only the case for one of the three.

Moreover, there is also the possibility that some methods connect features that are actually not the same physical entity, which would artificially skew the lifetime distribution toward longer lifetimes.

The distribution of the mean propagation velocities (Figs. 3e,f) also shows somewhat higher variance in the classes of lower velocities. Some strange behavior is also apparent for very high system velocities (beyond about 80 km h<sup>-1</sup>; not shown). Note that in extreme cases cyclones can move at over 110 km h<sup>-1</sup>. Differences in the distributions of propagation velocities might be partly associated with the tendency of some schemes to terminate the trajectories when rapid cyclone translation occurs, which is a known problem in many algorithms. Different tracking approaches, both those looking for the nearest neighbors in a defined distance as well as those extrapolating cyclone velocities, can miss fast-moving cyclones. In the first case this is due to cyclones moving farther than the predefined distance, and in the second case due to rapid acceleration or deceleration of cyclones. Some fast-moving cyclones, by virtue of being associated with stronger upper-level jets, are more likely, because of energy conversions, to give rise to an extreme windstorm, so the correct handling of system velocities is clearly important. Identification of the specific features of different algorithms responsible for the above is a challenging task and is left for future in-depth analysis.

**CASE STUDIES.** In section “Climatology of midlatitude cyclone characteristics” we showed that track-to-track matching rates for different schemes are in most cases at the 50%–70% level. In this section, two case studies, one from each hemisphere, are analyzed in detail to assess the performance of the different tracking algorithms for two fast-developing, high-impact storms.

The first case is the NH storm “Klaus,” which hit southwestern Europe in late January 2009 (Liberato et al. 2011; Bertotti et al. 2012) with particularly strong impacts over northern Spain and southwestern France. The cyclone developed over the subtropical

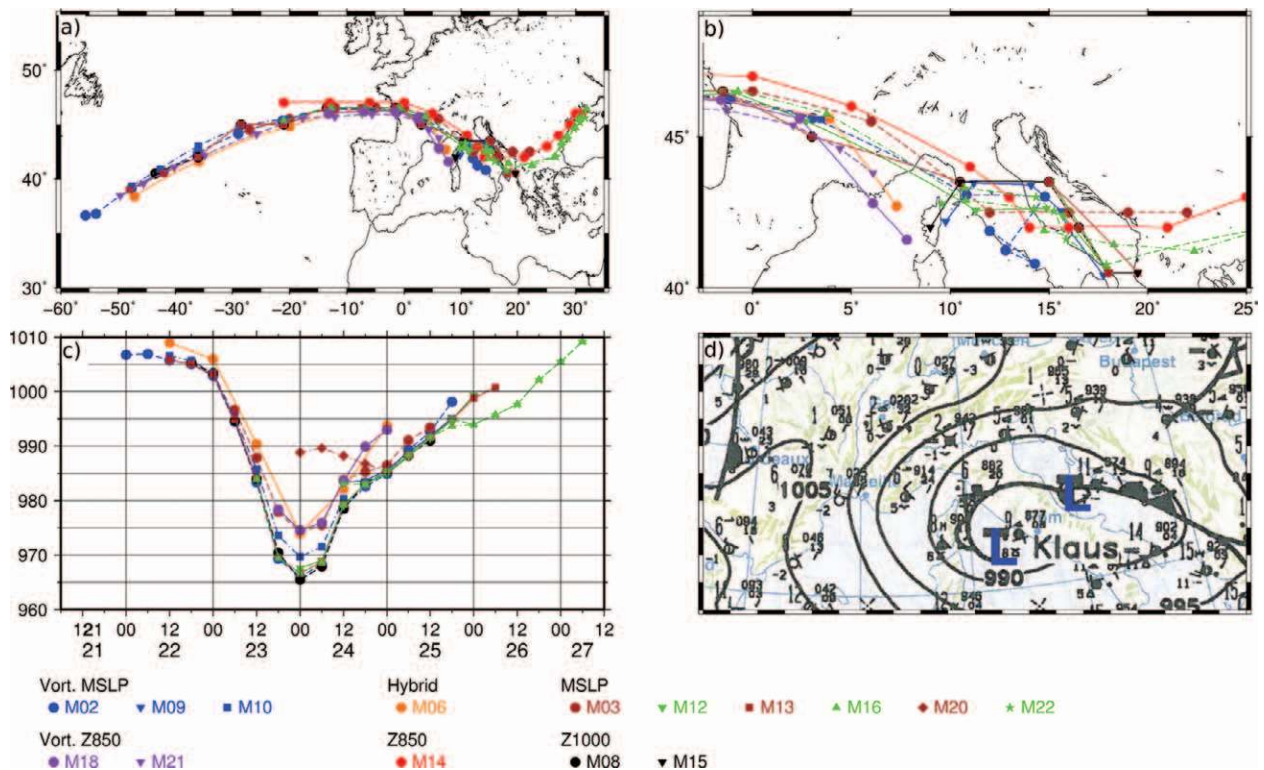
North Atlantic Ocean on 21–22 January 2009, moved eastward embedded in the strong westerly flow, and underwent explosive development during 23 January 2009. The storm then moved rapidly into the Bay of Biscay and then propagated further to the western Mediterranean.

Most methods agree well in the identification of the positions of this storm throughout most of its life cycle (Fig. 4a), particularly during the phase of explosive development and its propagation into the western Mediterranean on 23 and 24 January. However, there are some significant differences in the details of the life cycle, for example, in lifetime: many methods do not identify a cyclone track at the earlier stage of the development (22 January), and there is substantial disagreement in the lysis (dissolution of the cyclone) position. In particular, the earliest identification of the storm is at 0000 UTC 22 January (by method M02), and the latest is on 0600 UTC 27 January (M16). Considerable differences were found in the exact location of the cyclone positions over the western and central Mediterranean (Fig. 4b), resulting in different minimum central pressures (Fig. 4c).

Discrepancies in the cyclone positions appear mainly between 1200 UTC 24 January and 1800 UTC

25 January, where the system was characterized by two pressure minima: one located over southern France and a deeper one over the Gulf of Genoa (Liberato et al. 2011; see Fig. 4d). At this stage, some methods started to follow different minima when building the tracks, resulting in the spread of trajectories and central core pressure as displayed in Figs. 4b and 4c. This is likely due to method differences in the tracking procedure. However, the “choice” of which cyclone to follow in the case of splitting of cyclones or in case of generation of a new cyclone near an existing one made by an automated algorithm is not clear, and differences will always occur.

With most methods, the storm exhibited deepening rates of  $35 \text{ hPa (24 h)}^{-1}$  during its maturing stage. Although all methods except for one agree on the time of the minimum central pressure, the corresponding values vary within a range of 9 hPa (965.5–974.5 hPa). The reasons for this lie partially in the preprocessing of the input data, and also the manner in which a scheme interpolates to the location of the lowest pressure. For example, the preprocessing in M21 that leads to smoothing may be of importance in areas with weak large-scale pressure gradients like the Mediterranean.



**FIG. 4.** Detailed track comparison for NH storm Klaus (22–27 Jan 2009). The tracks of available methods are shown in different colors (see legend). (a) Complete track, (b) detail for the Alpine/north Mediterranean area, and (c) core pressure evolution. (d) The weather chart shows the synoptic situation on 0000 UTC 25 Jan with two cyclones over the Mediterranean.

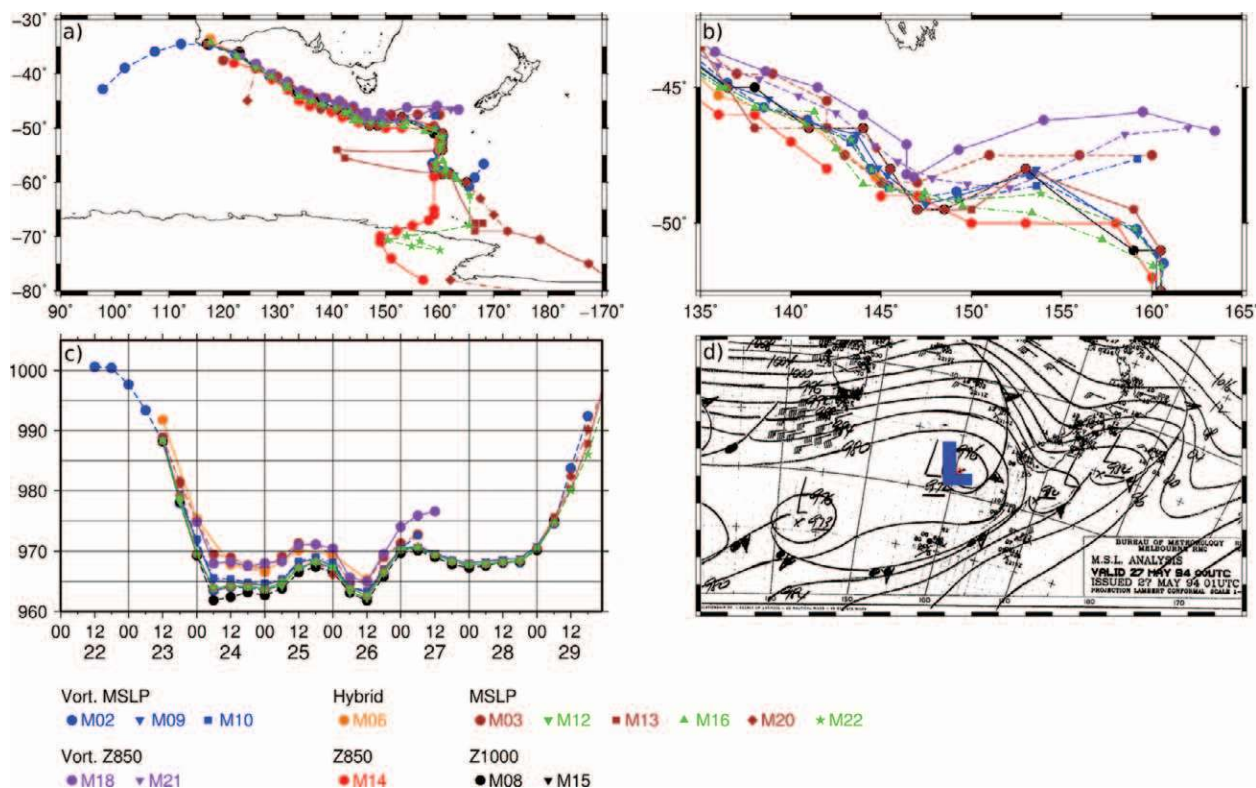
The second selected case study (Fig. 5) involves a deep cyclonic storm that affected southwest Western Australia and southeast Australia in late May 1994. It induced high winds and significant rainfall during 23 and 24 May 1994. Over the next days the cyclone moved southeastward across the Great Australian Bight. On 24 and 25 May a strong cold front associated with the cyclone caused a large dust storm that affected parts of South Australia, New South Wales, and Victoria (McInnes and Hubbert 1996; Trewin 2002).

The cyclone was successfully tracked by all methods (Fig. 5a). A drop in central pressure of about 30 hPa (24 h)<sup>-1</sup> was consistently captured by the different schemes from 0600 UTC 23 May to 0600 UTC 24 May. All algorithms also perform similarly during the intense phase (23–26 May), although differences in positions span about two grid lengths (relative to input data resolution; ~300 km). There are much larger differences in general at the stages of genesis and lysis. For instance, M02 captures a storm earliest, while M20 and M22 follow a track for longest during storm decay. Most methods identify the storm track from 26 to 29 May 1994 after

its reintensification (Fig. 5b). A large disagreement between the methods appears on 26 May, where some tracks start to follow a path toward New Zealand (M03, M10, M18, and M21) while some others turn to the south. As in the first case study, the handling of this splitting (see Fig. 5b) by the methods might be quite sensitive and can be induced by only minor algorithm differences.

Figure 5c displays the central core pressure between 0000 UTC 22 and 0000 UTC 29 May of different methods. Generally, its evolution is similar among the different methods. However, after the period of the most intense development at 0600 UTC 24 May, the minimum central pressure differs by 10 hPa (between 961.9 hPa in M08 and 971.6 hPa in M06; interpolated in time for M06). For the second pressure minimum during reintensification on 1200 UTC 26 May the spread of 3.5 hPa is smaller, ranging between 961.8 hPa (M08) and 965.3 hPa (M06).

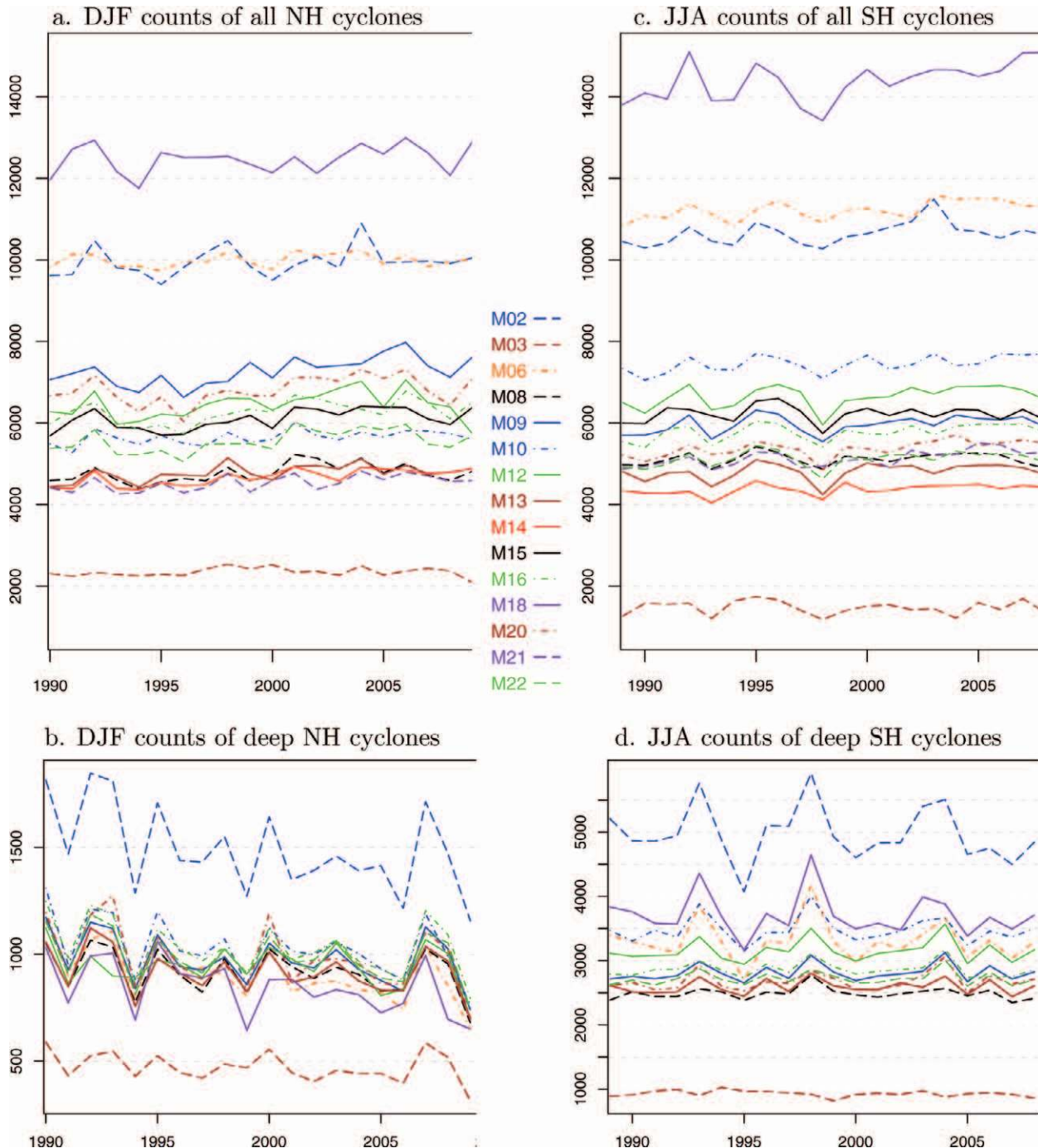
For these two examples all methods agree in the replication of the main segment of the track of a large cyclone, associated with the mature stage including explosive development. However, large scheme-to-scheme differences in both position and central core



**FIG. 5.** Detailed track comparison for an SH storm (unnamed; 22–29 May 1994). The available methods are shown in different colors (see legend). (a) Complete track, (b) detail for the central section of the track, and (c) core pressure evolution. (d) The weather chart shows the synoptic situation at the time after the splitting of the tracks on 0000 UTC 27 May.

pressure exist during the genesis and the dissolution phases, showing that the end-to-end tracking of high-impact weather systems has many nontrivial aspects. Future work will reveal if for other cases the explosive development phase is captured equally well, as this may not always be the case.

**INTERANNUAL VARIABILITY AND TRENDS.** Past and future trends in cyclone characteristics are an important issue in the discussion of climate change impacts. In this context, the knowledge of interannual and decadal variability is indispensable for assessing the importance of observed or



**FIG. 6.** Time series of (a),(b) NH and (c),(d) SH cyclone center counts in NH (DJF) and SH (JJA) winter, respectively. Deep cyclones here are those with core pressure  $\leq 980$  hPa (only for methods that output cyclone core pressure; for “deep cyclones” only time steps where the pressure is  $\leq 980$  hPa are counted). Note that for M06 cyclone counts have been doubled to account for its unique 12-h time step.

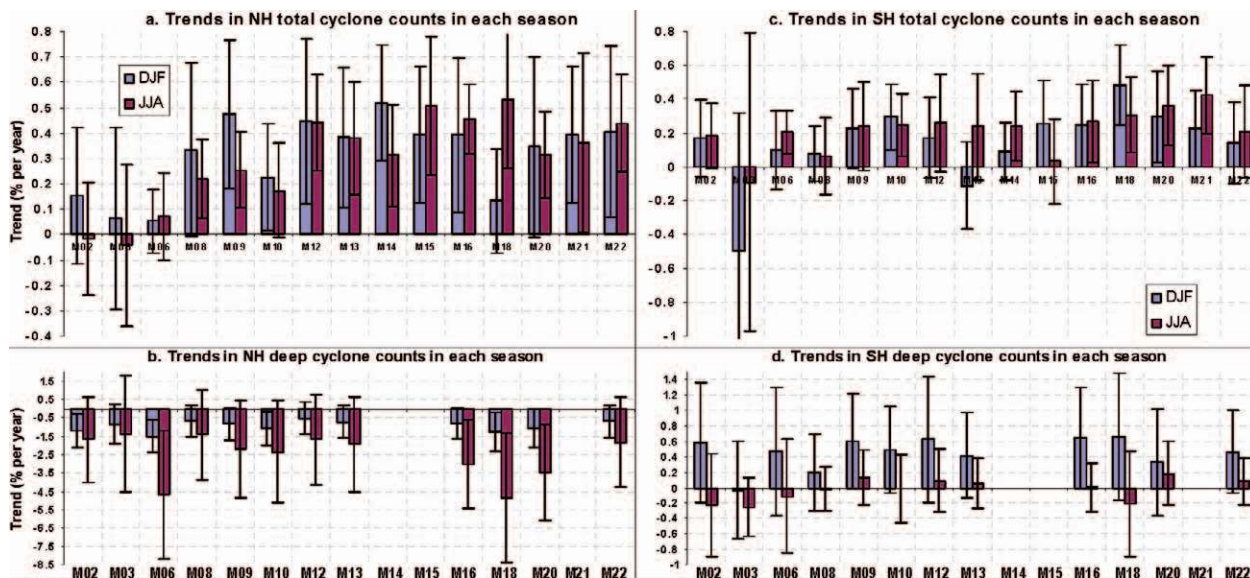
projected trends. Although a comprehensive diagnosis of long-term trends versus natural variability is not possible in our study, given the short 20-yr dataset used, it is important to quantify whether an overall agreement (or disagreement) of the trend sign and magnitude as well as interannual variability over this time interval exists between different tracking schemes. This provides information on the robustness against method uncertainties of long-term trend signals detected in studies using a single method.

Time series of hemispheric seasonal cyclone center counts for the NH and SH winter season are shown in Fig. 6. Note that “count” here means all centers found at any time step (see sidebar). Complementary to the discussion in section “Climatology of midlatitude cyclone characteristics,” Fig. 6 shows that although the scheme-to-scheme spread in total numbers of cyclones is wide, different methods are rather robust in counting deep cyclones throughout the 20-yr period, albeit with two outliers in both hemispheres (M02 and M03, Figs. 6b and 6d). The reason for these outliers at first sight seems to be the high and the low overall numbers. However, as we argued earlier, the correlation between total numbers and the number of deep cyclones is low, because we expect all algorithms to detect deep cyclones more effectively. In case of M03, the restriction of the number of cyclones per analysis (which selects the 25 strongest cyclones per time step) seems to reduce not only the overall number but also the number of

deep cyclones. The source of the particularly high number of deep cyclones in M02 might be associated with the deeper core pressures detected by M02 compared to other methods. The specific reason for that will be further analyzed in future work. Figure 6 demonstrates a striking similarity in the year-to-year variability between methods, especially for deep cyclones. Interannual variations of cyclone counts, in percentage terms, thus seem to depend very little on the method chosen.

The analysis of the 20-yr trends of seasonal hemispheric cyclone center counts shows that in the NH most methods identify a significant increase in the total number of cyclone centers (Fig. 7a) over the 1989–2009 period, but with considerable quantitative differences. The number of deep cyclone centers, however, consistently decreases (although mostly insignificantly) with all methods (Fig. 7b), with a somewhat smaller spread between methods than for total numbers. In the SH, all but one method find positive trends for total cyclone center counts, as well as a weak (mostly statistically insignificant) increase in the number of deep cyclone centers in austral summer (Figs. 7c,d). The spread between trend estimates derived from different methods is generally larger in the SH than in the NH, and SH trends estimated by different methods for the same season could have opposite signs.

Hemispherically averaged trends do not provide information about regional shifts of cyclone

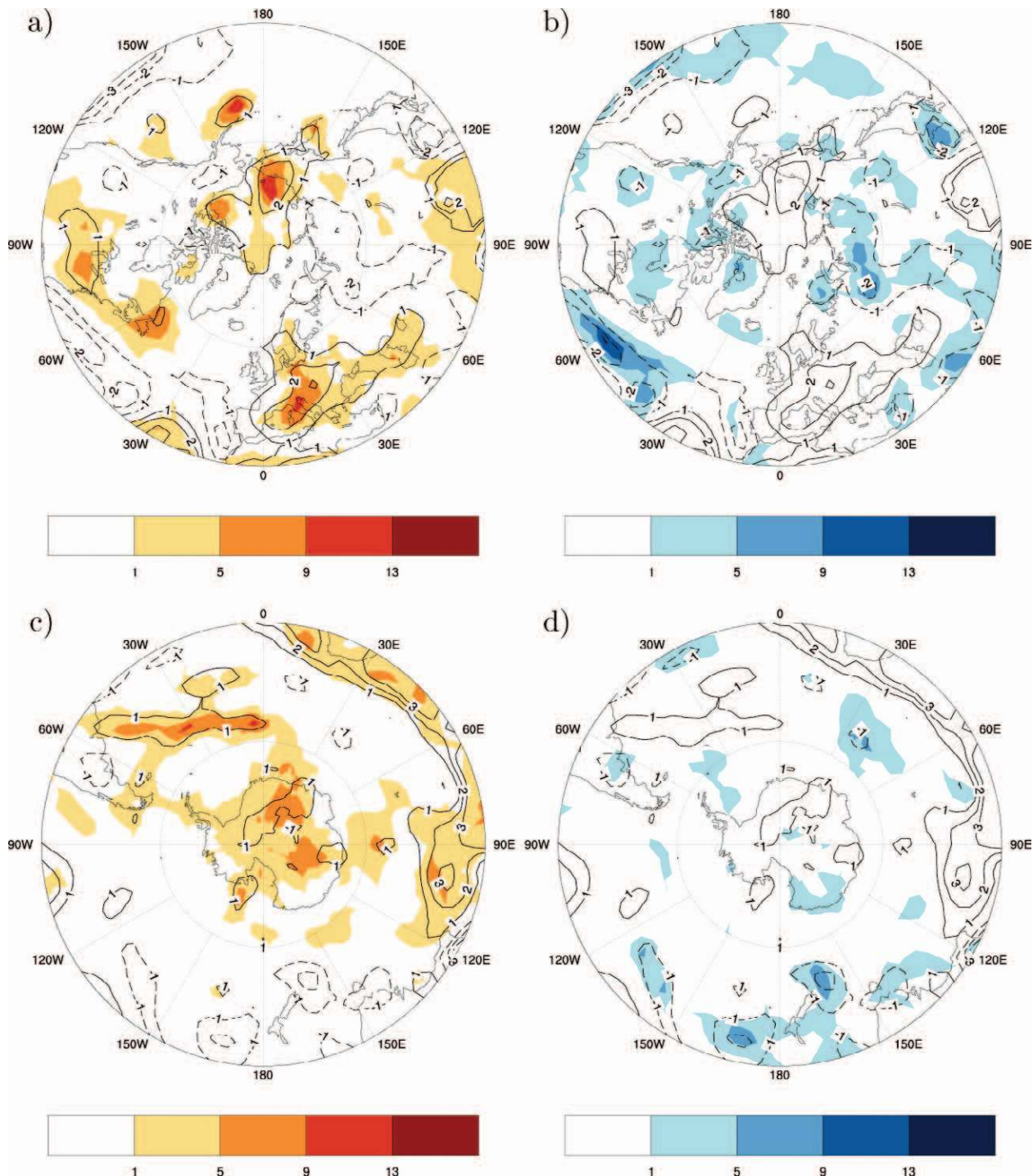


**FIG. 7.** Trends (per year, in % of the mean; least squares estimates) in time series of (a) NH total cyclone center counts, (b) NH deep cyclone center (core pressure  $\leq 980$  hPa) counts, (c) SH total cyclone center counts, and (d) SH deep cyclone center counts in DJF and JJA (series shown in Fig. 6). The error bars represent the 95% confidence range of the trend estimate (a trend is significant at 5% level if the error bar does not include zero).



occurrence. Because of their regional importance, we also investigated spatial patterns of trends in winter track density. Figure 8 shows the winter trend patterns of the multimethod ensemble (average of all

methods) of track density for the NH (Figs. 8a,b) and SH (Figs. 8c,d; contour lines). The agreement between the methods is analyzed by examining the number of methods that exhibit a significant positive (Figs. 8a,c)



**FIG. 8.** Geographical patterns of 20-yr winter trends of cyclone track density. The color scale represents the number of methods with a significant (a) positive and (b) negative relative trend (significance level of  $p > 90\%$ ) in the Northern Hemisphere (DJF); the contour lines represent the magnitude of the method ensemble-mean relative trend (in % change per year, of the average yearly track density). (c),(d) As in (a) and (b), respectively, but for SH (JJA).

and negative (Figs. 8b,d) trend sign, respectively (color scale). In the NH, several regions show a relatively large ensemble trend (contour lines), for example, over the Atlantic (negative; Fig. 8b), central Europe, and the northeast Pacific (positive; Fig. 8a). The closest agreement between methods was found in these areas of most distinct signals where a high number of methods show a significant positive or negative trend, respectively (Figs. 8a,b), and all methods show the same sign of trend (not shown). In the SH, ensemble trends in cyclone track density reveal regions with strong positive signals (Fig. 8c) in the Atlantic sector (60°W–0° at around 40°–50°S), in the Indian Ocean sector (about 90°E, 45°–55°S), and north of the Ross Sea. As in the NH, regions with strong ensemble trends coincide with regions of a close method-to-method agreement with high numbers of methods with a corresponding significant trend (Figs. 8c,d), and all methods exhibiting the same trend sign. In regions where the ensemble trend is weak, trends may often differ in sign across methods, but, reassuringly, there are very few areas where there are methods that exhibit significant trends with opposite sign.

Large areas with significant trends of the same sign across the methods present evidence for the presence of physically meaningful regional trend signals, because (i) there is overwhelming method consistency and (ii) large structures are unlikely to be generated by noise alone, whereas small areas of significant trends could occur by chance. The comparison of the method agreement in trend sign between track density, cyclone genesis, and lysis (not shown) indicates that signals detected for the field of track density are more consistent than for genesis and lysis locations. The location of the beginning or end of a track is thus more sensitive to the cyclone identification method used, as suggested earlier in the paper.

## DISCUSSION AND CONCLUSIONS.

Fifteen cyclone detection and tracking methods are compared using the same input dataset in order to assess their similarities and differences. For cyclone characteristics, for which the results from different methods are robust, estimates from a single method can be taken with a certain confidence. Consistency across the methods is generally higher for deep (or strong) cyclones than for shallow ones. This conclusion seems to hold also for cyclone frequency and life cycle, as well as for characteristics of interannual variability and trends. In the two cyclone case studies consistency across methods was best for the most intense part of the life cycles, rather than the periods of development and lysis. To some extent this is an

expected result, since intense cyclones show distinct values for most variables that different methods might use for identification, and thus they will be captured by most methods. Thus, the identification of intense cyclones and the part of their life cycle with intense development looks to be most robust with respect to choice of the method. However, even for these intense events, there can be significant differences in life cycle characteristics, in particular during genesis and lysis phases of the cyclones and, related to these, lifetime. Furthermore, the robustness of estimates of cyclone propagation speed is a concern, since there is a hint that some schemes might not recognize rapid movement. This will be investigated in future work.

With respect to numbers of cyclones, a qualitative “pattern matching” agreement between methods is obtained in terms of interannual variability and in geographical distribution, although there are some important differences in certain regions, notably the Mediterranean. Differences in absolute total numbers of cyclones are particularly large and imply caution when comparing corresponding results from studies using single but different methods. Analysis of life cycle characteristics in general shows a reasonable agreement. The largest spread in the frequency distributions was found for short-living, shallow, and slowly moving cyclones, whose detection is more sensitive to the choice of scheme. Differences in the distributions are generally larger in the NH than in the SH and are larger over parts of continents (e.g., Europe, North America, the Mediterranean), which are regions of high interest because storm impacts are high there.

An important result specifically relevant to the analysis of climate change impacts is the qualitative consistency shown for geographical linear trend patterns, where regions with strong trends show a good agreement, at least in sign, over most methods. This is an important consideration when trying to disentangle genuine trends in cyclone activity from natural variability and method uncertainty, and accordingly trying to quantify the statistical significance of any long-term trends being highlighted (Hodges 2008; Löptien et al. 2008; Della-Marta and Pinto 2009; Sienz et al. 2010).

Another key result is that it has so far proved difficult to clearly associate differences in the identified cyclone characteristics with features of the different schemes. In a few cases, outlier behavior can be explained by specific features of certain methods, like the preprocessing in method M21 or the restriction of the number of cyclones in method M03. In general, and somewhat surprisingly, we have

found little evidence of clustering of cyclone statistics according to algorithm features (e.g., vorticity schemes vs. MSLP schemes). Also, we see no detectable outlier performance for the sole method using a lower time resolution (M06) and no significantly better agreement in the results of two algorithms that are based on the same original method (M02 and M10). Furthermore, in the two case studies four methods clearly show shallower cyclones than the others, though the algorithms are unrelated. This all indicates that the documented differences result from a complex interplay between different aspects of each method. Threshold settings for identifying centers are believed to be one key aspect, while the algorithms that build tracks, and the pre- or postprocessing of input and output data also are very relevant.

When dealing with the complex and multifaceted character of these synoptic features, this study has shown that an ensemble of cyclone schemes has the potential to extract the key relevant features. Nevertheless, some findings might look somewhat discouraging in view of the importance of extratropical cyclones and their impact, but this has to be expected given the complexity of these features. Since there is no universal agreement upon cyclone definition, we cannot “judge” the algorithms or say that a specific one delivers “incorrect” results. They are all “right” in some sense. The many different approaches each have their own strengths and weaknesses, and each brings valuable perspectives to bear. They are all based on a similar physical understanding of complex processes, but deal with this in different ways. However, this makes it somewhat problematic for users of such results; they may lack guidance when assessing the results from different studies using different schemes, especially if the results contradict each other. In this sense, the findings of this study constitute important information for the interpretation of results of any extratropical cyclone analysis that uses only one identification and tracking algorithm. In particular, our study shows which aspects of cyclone identification and tracking are likely to be independent of the method used (and thus deserve higher confidence) and which aspects should be treated with caution. Thus, if using a single method, one should be aware of the sensitivity of results to the method, in particular with respect to total cyclone counts and the role of weak cyclones in the statistics.

Diagnosing the key reasons for differences in portrayed characteristics will involve more detailed study, especially sensitivity studies for specific parameters with a number of individual methods, and

will be undertaken in the next phase of this ongoing project. For now, IMILAST provides the community with a unique, comprehensive, and updatable database for analyzing the performance of cyclone tracking algorithms. We anticipate that further analysis of regional features, of other cyclone characteristics, and other case studies will involve a still wider community. Not all existing algorithms are included in this paper; for example, the algorithm from Hodges et al. (1995) is missing, but we hope to add more in further work.

There are many future developments that could play a positive role in improving our knowledge of cyclone climatology, for the benefit of society. One would be generating a “best-track dataset” for extreme extratropical cyclones, as already exists for tropical cyclones. This would be a good basis for the next step within IMILAST, which is to extend the comparison of method-related differences for a set of extreme cyclones (i.e., a set of further case studies). Others would be to improve reanalysis datasets to the point where they are better able to represent fine-scale structures of extratropical cyclones, or the extension of cyclone identification schemes to examine vertical structures of cyclones (e.g., Dacre et al. 2012; Kouroutzoglou et al. 2012; Čampa and Wernli 2012). These developments will require concerted international efforts for realization.

**ACKNOWLEDGMENTS.** We thank Swiss Re for sponsoring the project (coordination office and workshops) and ECWMF for providing the input data of ERA Interim. C. C. Raible is supported by NCCR Climate, funded by the Swiss National Science Foundation. M. L. R. Liberato was supported by the project STORMEx (FCOMP-01-0124-FEDER-019524), funded by FCT and cofunded by FEDER. N. Bellenbaum, J. G. Pinto, and S. Ulbrich thank AON Benfield Impact Forecasting for support over the EUWS project. J. Grieger and M. Schuster are supported by the DFG project SACAI (DFG-LE1865/1-3). M. G. Akperov and I. I. Mokhov are supported by the Russian Ministry of Education and Science (11.519.11.5004). We appreciate the lead authorship of C. C. Raible, S. Gulev, J. G. Pinto, G. C. Leckebusch, and X. L. Wang, respectively, for the different analysis sections of this paper.

## REFERENCES

Akperov, M. G., M. Yu, E. Bardin, M. Volodin, G. S. Golitsyn, and I. I. Mokhov, 2007: Probability distributions for cyclones and anticyclones from the NCEP/NCAR reanalysis data and the INM RAS climate model. *Izv., Atmos. Oceanic Phys.*, **43**, 705–712.

- Allen, J. T., A. B. Pezza, and M. T. Black, 2010: Explosive cyclogenesis: A global climatology comparing multiple reanalyses. *J. Climate*, **23**, 6468–6484.
- Bardin, M. Yu., and A. B. Polonsky, 2005: North Atlantic oscillation and synoptic variability in the European-Atlantic region in winter. *Izv., Atmos. Oceanic Phys.*, **41**, 127–136.
- Benestad, R. E., and D. Chen, 2006: The use of a calculus-based cyclone identification method for generating storm statistics. *Tellus*, **58A**, 473–486.
- Bengtsson, L., K. I. Hodges, and N. Keenlyside, 2009: Will extratropical storms intensify in a warmer climate? *J. Climate*, **22**, 2276–2301.
- Bertotti, L., and Coauthors, 2012: Performance of different forecast systems in an exceptional storm in the Western Mediterranean Sea. *Quart. J. Roy. Meteor. Soc.*, **138**, 34–55.
- Blender, R., and M. Schubert, 2000: Cyclone tracking in different spatial and temporal resolutions. *Mon. Wea. Rev.*, **128**, 377–384.
- , K. Fraedrich, and F. Lunkeit, 1997: Identification of cyclone-track regimes in the North Atlantic. *Quart. J. Roy. Meteor. Soc.*, **123**, 727–741.
- Čampa, J., and H. Wernli, 2012: A PV perspective on the vertical structure of mature midlatitude cyclones in the Northern Hemisphere. *J. Atmos. Sci.*, **69**, 725–740.
- Dacre, H. F., M. K. Hawcroft, M. A. Stringer, and K. I. Hodges, 2012: An extratropical cyclone atlas: A tool for illustrating cyclone structure and evolution. *Bull. Amer. Meteor. Soc.*, **93**, 1497–1502.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 1972–1990.
- Della-Marta, P. M., and J. G. Pinto, 2009: Statistical uncertainty of changes in winter storms over the North Atlantic and Europe in an ensemble of transient climate simulations. *Geophys. Res. Lett.*, **36**, L14703, doi:10.1029/2009GL038557.
- Hanley, J., and R. Caballero, 2012: Objective identification and tracking of multicentre cyclones in the ERA-Interim reanalysis data set. *Quart. J. Roy. Meteor. Soc.*, **138**, 612–625.
- Hewson, T. D., 1997: Objective identification of frontal wave cyclones. *Meteor. Appl.*, **4**, 311–315.
- , and H. A. Titley, 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteor. Appl.*, **17**, 355–381.
- , G. C. Craig, and C. Claud, 2000: Evolution and mesoscale structure of a polar low outbreak. *Quart. J. Roy. Meteor. Soc.*, **126**, 1031–1063.
- Hodges, K. I., 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465.
- , 2008: Confidence intervals and significance tests for spherical data derived from feature tracking. *Mon. Wea. Rev.*, **136**, 1758–1777.
- , B. J. Hoskins, J. Boyle, and C. Thorncroft, 2003: A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Mon. Wea. Rev.*, **131**, 2012–2037.
- , R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Climate*, **24**, 4888–4906.
- Hoskins, B. J., and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061.
- Inatsu, M., 2009: The neighbor enclosed area tracking algorithm for extratropical wintertime cyclones. *Atmos. Sci. Lett.*, **10**, 267–272.
- Jung, T., S. K. Gulev, I. Rudeva, and V. Soloviev, 2006: Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **132**, 1839–1858.
- Kew, S. F., M. Sprenger, and H. C. Davies, 2010: Potential vorticity anomalies of the lowermost stratosphere: A 10-yr winter climatology. *Mon. Wea. Rev.*, **138**, 1234–1249.
- Kouroutzoglou, J., H. A. I. Flocas, K. Keay, I. Simmonds, and M. Hatzaki, 2012: On the vertical structure of Mediterranean explosive cyclones. *Theor. Appl. Climatol.*, **110**, 155–176, doi:10.1007/s00704-012-0620-3.
- Leckebusch, G. C., B. Koffi, U. Ulbrich, J. G. Pinto, T. Spanghel, and S. Zacharias, 2006: Analysis of frequency and intensity of winter storm events in Europe on synoptic and regional scales from a multimodel perspective. *Climate Res.*, **31**, 59–74.
- Liberato, M. R. L., J. G. Pinto, I. F. Trigo, and R. M. Trigo, 2011: Klaus—An exceptional winter storm over northern Iberia and southern France. *Weather*, **66**, 330–334.
- Lionello, P., F. Dalan, and E. Elvini, 2002: Cyclones in the Mediterranean region: The present and the doubled CO<sub>2</sub> climate scenarios. *Climate Res.*, **22**, 147–159.
- Löptien, U., O. Zolina, S. K. Gulev, M. Latif, and V. Soloviev, 2008: Cyclone life cycle characteristics over the Northern Hemisphere in coupled GCMs. *Climate Dyn.*, **31**, 507–532.
- McInnes, K. L., and G. D. Hubbert, 1996: Extreme events and the impact of climate change on Victoria's coastline. Environment Protection Authority, State Government of Victoria Publ. 488, 69 pp.
- Murray, R. J., and I. Simmonds, 1991: A numerical scheme for tracking cyclone centers from digital data.

- Part I: Development and operation of the scheme. *Aust. Meteor. Mag.*, **39**, 155–166.
- Pinto, J. G., T. Spanghel, U. Ulbrich, and P. Speth, 2005: Sensitivities of a cyclone detection and tracking algorithm: Individual tracks and climatology. *Meteor. Z.*, **14**, 823–838.
- , S. Zacharias, A. H. Fink, G. C. Leckebusch, and U. Ulbrich, 2009: Factors contributing to the development of extreme North Atlantic cyclones and their relation with the NAO. *Climate Dyn.*, **32**, 711–737.
- Raible, C. C., P. M. Della-Marta, C. Schwierz, H. Wernli, and R. Blender, 2008: Northern Hemisphere extratropical cyclones: A comparison of detection and tracking methods and different reanalyses. *Mon. Wea. Rev.*, **136**, 880–897.
- , B. Ziv, H. Saaroni, and M. Wild, 2010: Winter synoptic-scale variability over the Mediterranean Basin under future climate conditions as simulated by the ECHAM5. *Climate Dyn.*, **35**, 473–488.
- Rudeva, I., and S. K. Gulev, 2007: Climatology of cyclone size characteristics and their changes during the cyclone life cycle. *Mon. Wea. Rev.*, **135**, 2568–2587.
- Schneiderreit, A., R. Blende, and K. Fraedrich, 2010: A radius–depth model for midlatitude cyclones in reanalysis data and simulations. *Quart. J. Roy. Meteor. Soc.*, **136**, 50–60.
- Serreze, M. C., 1995: Climatological aspects of cyclone development and decay in the Arctic. *Atmos.–Ocean*, **33**, 1–23.
- Sienz, F., A. Schneiderreit, R. Blender, K. Fraedrich, and F. Lunkeit, 2010: Extreme value statistics for North Atlantic cyclones. *Tellus*, **62A**, 347–360.
- Simmonds, I., R. J. Murray, and R. M. Leighton, 1999: A refinement of cyclone tracking methods with data from FROST. *Aust. Meteor. Mag.*, **Special Issue**, 35–49.
- , C. Burke, and K. Keay, 2008: Arctic climate change as manifest in cyclone behavior. *J. Climate*, **21**, 5777–5796.
- Sinclair, M. R., 1994: An objective cyclone climatology for the Southern Hemisphere. *Mon. Wea. Rev.*, **122**, 2239–2256.
- , 1997: Objective identification of cyclones and their circulation intensity, and climatology. *Wea. Forecasting*, **12**, 591–608.
- Trewin, B., 2002: Charts from the past—24 May, 1994. *Bull. Aust. Meteor. Ocean. Soc.*, **17**, 74.
- Trigo, I. F., 2006: Climatology and interannual variability of storm-tracks in the Euro-Atlantic sector: A comparison between ERA-40 and NCEP/NCAR reanalyses. *Climate Dyn.*, **26**, 127–143.
- Ulbrich, U., G. C. Leckebusch, and J. G. Pinto, 2009: Extra-tropical cyclones in the present and future climate: A review. *Theor. Appl. Climatol.*, **96**, 117–131.
- Wang, X. L., V. R. Swail, and F. W. Zwiers, 2006: Climatology and changes of extratropical cyclone activity: Comparison of ERA-40 with NCEP/NCAR reanalysis for 1958–2001. *J. Climate*, **19**, 3145–3166.
- Wernli, H., and C. Schwierz, 2006: Surface cyclones in the ERA-40 dataset (1958–2001). Part I: Novel identification method and global climatology. *J. Atmos. Sci.*, **63**, 2486–2507.
- Zolina, O., and S. K. Gulev, 2002: Improving accuracy of mapping cyclone numbers and frequencies. *Mon. Wea. Rev.*, **130**, 748–759.