Geosci. Model Dev. Discuss., 8, 7727–7765, 2015
www.geosci-model-dev-discuss.net/8/7727/2015/
doi:10.5194/gmdd-8-7727-2015

Geoscientific Model Development Discussions
Open Access

Stride Search: a general algorithm for storm detection in high resolution climate data

P. A. Bosler[1], E. L. Roesler[2], M. A. Taylor[1], and M. Mundt[1]

[1]Sandia National Laboratories, Center for Computing Research, P.O. Box 5800, Albuquerque NM, 87185-1321, USA
[2]Sandia National Laboratories, Geophysics and Atmospheric Sciences, P.O. Box 5800, Albuquerque NM, 87185-0750, USA

Received: 26 June 2015 – Accepted: 12 August 2015 – Published: 8 September 2015

Correspondence to: P. A. Bosler (pabosle@sandia.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This article discusses the problem of identifying extreme climate events such as intense storms within large climate data sets. The basic storm detection algorithm is reviewed, which splits the problem into two parts: a spatial search followed by a temporal correlation problem. Two specific implementations of the spatial search algorithm are compared. The commonly used grid point search algorithm is reviewed, and a new algorithm called Stride Search is introduced. Stride Search is designed to work at all latitudes, while grid point searches may fail in polar regions. Results from the two algorithms are compared for the application of tropical cyclone detection, and shown to produce similar results for the same set of storm identification criteria. The time required for both algorithms to search the same data set is compared. Stride Search's ability to search extreme latitudes is demonstrated for the case of polar low detection.

1 Introduction

A problem inherent to modern climate science is data analysis. Current global climate models produce dozens of gigabytes of data per simulated month, and as models and computing architectures improve to support higher temporal and spatial resolutions, the amount of data output will increase proportionally. With each individual model, each member of a forecast ensemble, and each model in an intercomparison project producing such a large amount of data, the size and amount of data have become obstacles to scientific analysis. Efficient algorithms and software are necessary to analyze these data.

A basic problem in climate data analysis is the identification of extreme climate events within a data set. For a particular type of event, described quantitatively by a set of *identification criteria*, the data analysis problem is to locate points in space and time within a data set that match or exceed those criteria. With both the size of climate

GMDD
8, 7727–7765, 2015

data and its number of users increasing, search algorithms capable of locating a wide variety of extreme events are increasingly important.

In this paper we address storm detection algorithms, which present solutions to the extreme event identification problem. For simplicity we refer to any extreme climate event in this work as a storm, whether or not "storm" is the most appropriate meteorological term. The paper is organized in the following manner. The remainder of this introduction reviews the fundamental storm detection algorithm, which is based on the typical layout of climate data files, and discusses common strategies for specifying storm identification criteria for a variety of applications. Section 2 reviews the grid point search algorithm, the most common strategy for a spatial search (Vitart et al., 1997; Blender et al., 1997; Wernli and Schwierz, 2006; Kleppek et al., 2008; Raible et al., 2008; Prabhat et al., 2012). Although straightforward to implement, the grid point search algorithm fails at high latitudes because it relies on geometric assumptions that are only valid away from the poles. In Sect. 3 we introduce a new algorithm, called Stride Search, that is capable of detecting storms at any latitude for many applications. To validate the new approach, Sect. 5 compares Stride Search results for the application of tropical cyclone detection to a well known implementation of the grid point search algorithm, the Geophysical Fluid Dynamics Laboratory's (GFDL) TSTORMS (1997) software. Section 6 introduces Stride Search's new polar search capability using the problem of polar low identification. A brief concluding discussion follows.

1.1 Storm detection algorithms

The basic storm detection algorithm is guided by the typical layout of a climate data set. Climate data are commonly represented on uniform latitude-longitude grids with resolution $\Delta\lambda$, with grid points (λ_j, θ_i) located at

$$\lambda_j = j\Delta\lambda, \quad j = 0, \ldots, n_{\text{lon}} - 1, \tag{1a}$$

$$\theta_i = -90 + i\Delta\lambda, \quad i = 0, \ldots, n_{\text{lat}} - 1, \tag{1b}$$

GMDD
8, 7727–7765, 2015

where λ is longitude, θ is latitude, $n_{\text{lon}} = 360/\Delta\lambda$ is the number of longitudinal grid points, and $n_{\text{lat}} = n_{\text{lon}}/2 + 1$ is the number of latitude grid points. Spatial data are separated by time step and grouped consecutively so that each data file represents a discrete time interval. In most cases these data are stored in NetCDF files (Rew et al., 2014).

The layout of the data suggests that the extreme event detection problem may be separated into two steps: a spatial search problem, and a temporal correlation problem (TSTORMS, 1997; Prabhat et al., 2012). The spatial search problem is to find, at each time step, regions of a spatial domain that contain an extreme event, as defined by a set of identification criteria. The temporal correlation problem is to correlate the output of the spatial search across time steps to construct storm tracks.

The basic spatial search is presented as Algorithm 1. Its input is a search domain and a set of per-time step storm identification criteria, as well as the data. A key step in the algorithm is the division of the search domain into a set of search sectors (line 4); we will discuss this process in more detail in the following sections. For each file and each time step, the algorithm compares the data within each sector to the storm identification criteria. If the criteria are met, a storm is recorded to the list L at the current time step.

Identification criteria, particularly those used with a spatial search, are highly application dependent. Cyclones are the most commonly sought after features in climate data. Review articles by Raible et al. (2008) and Neu et al. (2013) discuss extratropical cyclone identification criteria in detail. Detection thresholds may be defined in terms of minimum sea level pressure, maximum cyclonic vorticity, minima in geopotential height surfaces (König et al., 1993; Sinclair, 1994; Blender et al., 1997), or combinations of several independent criteria. Tropical cyclone identification criteria often also include a temperature criterion or a wind shear criterion to filter out baroclinic storms (Vitart et al., 1997; Nguyen and Walsh, 2001; Kleppek et al., 2008; Zarzycki and Jablonowski, 2014). Other studies define criteria to identify extreme weather events using more specialized variables. Examples include using column-integrated water content to identify

GMDD
8, 7727–7765, 2015

Algorithm 1 Spatial search algorithm

Input: search domain R, storm identification criteria (spatial), data
Output: List of potential storms, L, organized by increasing time step

```
1:  set L = empty list
2:  for all files in data set do
3:     n_T = number of time steps per file
4:     Divide search domain into search sectors
5:     for i = 1 to n_T do
6:        for all search sectors at time step i do
7:           if sector meets or exceeds identification criteria then
8:              set l_i = storm data at time step i
9:              add l_i to L
10:          end if
11:       end for
12:    end for
13: end for
```

atmospheric rivers (Ralph et al., 2004) and diagnosing developing polar lows using potential vorticity anomalies (Montgomery and Farrell, 1992; Wu et al., 2011; Føre et al., 2012). An objective definition of a drought is more elusive, as it involves both atmospheric and hydrologic processes (Oladipo, 1985). Keyantash and Dracup (2002) discuss several criteria used to calculate drought indices from forecast data that may be applicable to climate data sets as well. Ideally storm detection software should be flexible enough to allow users to easily define identification criteria relevant to their area of study and should not be limited to any specific geographic region. Users should be able to easily modify the implementation of Algorithm 1, line 7, in code.

The output of the spatial search (Algorithm 1) is a list of candidate storms. This list may contain false positives due to noisy data, topographic effects, or ambiguity within

GMDD
8, 7727–7765, 2015

the identification criteria. The second step of the storm detection, the temporal correlation problem, handles these issues. The temporal correlation algorithm's task is to identify the same storm across adjacent time steps. It does so by building storm tracks and is outlined by Algorithm 2. Users define a maximum travel speed U_{max} appropriate to the type of storm under investigation. The algorithm uses that speed to define $D_{max} = U_{max} \cdot \Delta t$, the maximum possible distance a storm may travel per time step. Beginning with a storm entry in the spatial search output list L at time step i, the algorithm searches all storms detected at time step $i+1$. Any storms at time step $i+1$ separated by a distance less than D_{max} from the storm at time step i are marked as candidate successors. If zero candidates are found, the track is ended at time step i. If one candidate is found, that candidate is linked to its predecessor at time step i and the algorithm continues to build the track by looking for candidates at time step $i+2$. If several candidates are found at time step $i+1$, the algorithm chooses the closest candidate to the entry at time step i and disregards the others. Track building proceeds until either zero candidates are found at the next time step or the data are exhausted.

Storm tracks provide a natural mechanism to count storms and to dismiss false positives. Tracks that consist of only one point, indicating a storm whose duration was only 1 time step may be dismissed as noise. Tracks that persist for many time steps but do not move may possibly be regarded as topographic effects, particularly if the identification criteria use data that are sensitive to topography, such as geopotential height surfaces.

Storm tracks also provide a straightforward method of applying additional identification criteria. A study concerned with identifying regions of cyclogenesis may reject any storms that do not intensify along their track. Temporal criteria may also be used to perform more detailed classification of candidate storms. For example, a vorticity criterion or a vertical wind speed criterion may detect strong convection due to thunderstorms. To make a distinction between typical summertime afternoon thunderstorms and more persistent mesoscale convective complexes, a temporal criterion may be used to neglect storms that do not persist for longer than 12 h.

GMDD
8, 7727–7765, 2015

2 Grid point search

Most storm detection studies of climate data use software that implements Algorithm 1 as a grid point search. The Geophysical Fluid Dynamics Laboratory's (GFDL) TSTORMS software, for example, has become the main tool for tropical cyclone detection in high resolution climate data (Vitart et al., 1997; Vitart and Stockdale, 2001; Knutson et al., 2008; Zhao et al., 2009; Prabhat et al., 2012; Zarzycki and Jablonowski, 2014). Grid point searches have also been employed by other studies of mid-latitude extratropical cyclones (Blender et al., 1997; Geng and Sugi, 2001; Wernli and Schwierz, 2006; Kleppek et al., 2008; Raible et al., 2008).

In a grid point search algorithm, each grid point (λ_j, θ_i) in the search domain is a search sector center. A search sector K_{ij} centered at grid point (λ_j, θ_i) is defined as

$$K_{ij} = \{(\lambda_{j\pm k}, \theta_{i\pm l})\} \qquad k, l = 0, \ldots, n-1, \tag{2}$$

where n is a user-specified parameter that corresponds to the scale of a storm in latitude-longitude space. Thus, each sector is a $(2n+1) \times (2n+1)$ square in grid point space. To set up a grid point search, users define the search domain by defining a minimum and maximum latitude, $\theta_{\min}$ and $\theta_{\max}$. They must also select a value for n that relates the spatial scale of the storms they wish to detect to the resolution of the data. Figure 1a shows grid point search sectors along the equator with $n = 2$ for data with resolution $\Delta\lambda = 10^{\circ}$. The sectors are 5×5 boxes in grid point space and span approximately 5000 km × 5000 km on the Earth.

For each sector, the software collects data from the $(2n+1) \times (2n+1)$ points centered at (λ_j, θ_i). The collected data are compared against the storm identification criteria. If the criteria are met or exceeded in the sector, the algorithm checks to see if (λ_j, θ_i) is the location of the storm within that sector. If so, the algorithm records the storm to its output list. If not, the algorithm cycles to the next grid point, say $(\lambda_{j+1}, \theta_i)$, and begins again. In Fig. 1a the blue, horizontally striped sector is centered at $(\lambda_j, \theta_i) = (150^{\circ}\,\mathrm{E},$

GMDD
8, 7727–7765, 2015

Algorithm 2 Temporal correlation algorithm

Input: List L of potential storms from spatial search output, max storm travel speed U_{max}, storm identification criteria (temporal)
Output: List T of storm tracks

set T = empty list
N_T = total number of time steps in data set
for $i = 1$ **to** N_T **do**
  for all elements $I_i \in L$ at time step i **do**
    start new track t at I_i
    continue = `True`
    $j = i$
    while continue **do**
      examine all $I_{j+1} \in L$ at time step $j + 1$ for possible successors to storm I_j
      if successor found **then**
        add I_{j+1} to track t
        $j = j + 1$
      else
        continue = `False`
      end if
    end while
    if track t meets or exceeds identification criteria **then**
      add t to T
    end if
  end for
end for

GMDD
8, 7727–7765, 2015

0° N) and the next two consecutive sectors are shown by the red, vertically striped sector centered at $(\lambda_{j+1}, \theta_i) = (160^\circ\,\mathrm{E}, 0^\circ\,\mathrm{N})$ and black, diagonally striped sector whose center is $(\lambda_{j+2}, \theta_i) = (170^\circ\,\mathrm{E}, 0^\circ\,\mathrm{N})$.

Centering a sector at each grid point in the search domain yields a robust algorithm. It ensures that the entire search domain will be covered and that the same storm will not be recorded twice. Even though a single storm may trigger the identification criteria in several sectors, only the sector whose center corresponds to the storm location will be recorded to output. While the robustness of the grid point search algorithm is an advantage, it comes at the cost of redundant work. The same data points are accessed multiple times because the algorithm only advances one grid point at a time and the overlap of adjacent sectors is considerable.

The data access required by a grid point search and the overlap of adjacent sectors is also illustrated by Fig. 1a. All three sectors read the data from grid points in the region $\{(\lambda, \theta) : 150^\circ\,\mathrm{E} \le \lambda \le 170^\circ\,\mathrm{E}, 20^\circ\,\mathrm{S} \le \theta \le 20^\circ\,\mathrm{N}\}$. For visual clarity we have not plotted the sectors at $(\lambda_{j-1}, \theta_i)$ or $(\lambda_j, \theta_{i\pm1})$, which would also overlap a majority of the same grid points.

We also note that grid point searches perform all computations in the longitude-latitude plane. To compute the distance between two locations (λ_1, θ_1) and (λ_2, θ_2), the grid point algorithm uses the angular distance function

$$\mathrm{dist}_A\big((\lambda_1, \theta_1), (\lambda_2, \theta_2)\big) = \sqrt{(\lambda_2 - \lambda_1)^2 + (\theta_2 - \theta_1)^2}. \quad (3)$$

For GFDL's TSTORMS, whose intended application is in tropical regions, this is a simple and effective strategy because angular distance is a reasonable proxy for geodesic distance near the equator. However, away from the equator this strategy can cause problems. We discuss these issues in Sect. 3.

GMDD
8, 7727–7765, 2015

3 Stride Search

We present Stride Search with the aim of providing an implementation of Algorithm 1 suitable for storm detection applications anywhere on the globe. Instead of Eq. (3), Stride Search uses the geodesic distance function

$$\mathrm{dist}_G\left((\lambda_1,\theta_1),(\lambda_2,\theta_2)\right) = a\arccos\left(\cos\theta_1\cos\theta_2\cos(\lambda_2-\lambda_1)+\sin\theta_1\sin\theta_2\right), \quad (4)$$

where a is the radius of the Earth, to facilitate search applications at any latitude. This allows users to select a spatial scale s in units of distance independently from the resolution of the data. To define s, users select a spatial scale relevant to their application such that a maximum of one storm can be located within any spherical circle of radius s. This scale defines the search sectors.

We divide the search domain into a collection of circles on the sphere, each with the same geodesic radius s. Stride Search sectors, illustrated in Fig. 1b for $s = 2500$ km, are defined by the following procedure.

The number

$$S_{\mathrm{lat}} = \left\lfloor \frac{s\cdot n_{\mathrm{lon}}}{2\pi a} \right\rfloor \quad (5)$$

is the integer number of grid points required to approximately span geodesic distance s in the latitudinal (meridional) direction. We refer to S_{lat} as the *latitude stride* and use it to define lines of constant latitude θ_l, where $l = i_{\min} + (i-1)S_{\mathrm{lat}}$ for $i = 1,\ldots,N_\theta$. The set θ_l divides the search domain into strips using $N_\theta = (i_{\max} - i_{\min})/S_{\mathrm{lat}} + 1$ latitude lines, where the indices $i_{\min}$ and $i_{\max}$ are defined such that $\theta_{i_{\min}} = \theta_{\min}$ and $\theta_{i_{\max}} = \theta_{\max}$.

We also define N_θ *longitude strides*,

$$S_{\mathrm{lon}}^{(l)} = \min\left(\left\lfloor \frac{s\cdot n_{\mathrm{lon}}}{2\pi a\cos\theta_l} \right\rfloor, n_{\mathrm{lon}}\right), \quad (6)$$

GMDD
8, 7727–7765, 2015

so that $S_{\text{lon}}^{(l)}$ are the approximate number of grid points required to span geodesic distance s in the longitudinal (zonal) direction along latitude circle θ_l. The minimum function in Eq. (6) accounts for the case where θ_l is either pole. The longitude stride defines points λ_J along each latitude line θ_l, where $J = (j-1)S_{\text{lon}}^{(l)}$ for $j = 1, \ldots, N_\lambda^{(l)}$. This process creates $N_\lambda^{(l)} = n_{\text{lon}}/S_{\text{lon}}^{(l)}$ longitude points along θ_l.

The points (λ_J, θ_l) define the centers of search sectors K_{lJ}, where K_{lJ} is the set of all points on the sphere lying within a distance s of (λ_J, θ_l),

$$K_{lJ} = \left\{(\lambda,\theta) : \text{dist}_G\left((\lambda,\theta),(\lambda_J,\theta_l)\right) \le s\right\}. \tag{7}$$

Adjacent sectors are separated in grid point space by increments of the latitude stride and the longitude stride to reduce overlap. Hence, by construction, sector K_{lJ} overlaps its neighbors $K_{l\pm1,J}$ and $K_{l,J\pm1}$ by approximately one radius s in physical space. This is a sufficient condition to ensure that the entire search domain will be covered by the circular search sectors.

Stride Search setup is completed by defining the sectors K_{lJ} in terms of the data locations (λ_j, θ_i). Computationally, this involves creating a data structure or method that identifies and links each sector to the data points enclosed by its geographic boundary. For uniform longitude-latitude grids, Eq. (1) applies and the process is straightforward. For unstructured grids either the grid's connectivity information or a *kd*-tree algorithm may be used. Sectors used with high resolution data sets link to more grid points than the same-sized sectors used with lower resolution data. Thus, the stride search parameter s is decoupled from the data resolution and depends only on the application.

Figure 1b illustrates three consecutive sectors beginning with the blue, horizontally striped circle centered at $(\lambda_J, \theta_l) = (150^\circ\,\text{E}, 0^\circ\,\text{N})$. The red, vertically striped sector centered at $(\lambda_{J+1}, \theta_l) = (170^\circ\,\text{E}, 0^\circ\,\text{N})$ and the black, diagonally striped sector centered at $(\lambda_{J+2}, \theta_l) = (170^\circ\,\text{W}, 0^\circ\,\text{N})$ are the next two consecutive sectors. Along the equator, the latitude stride S_{lat} and longitude stride $S_{\text{lon}}^{(l)}$ are both equal to 2 for $s = 2500$ km and $\Delta\lambda = 10^\circ$, hence each of the three sectors' centers are separated by 20°. Since each

GMDD
8, 7727–7765, 2015

Stride Search sector is separated from its immediate neighbors by either a latitude stride or a longitude stride, the overlap between adjacent sectors is much reduced compared to a grid point search. The Stride Search algorithm therefore covers a much larger geographic area with the same number of search sectors.

Reduced overlap leads to improved performance by decreasing the number of redundant data accesses and by reducing the total number of sectors. However, it also creates a new issue. Any sector whose data meet or exceed the storm identification criteria record a storm to a linked-list as an intermediate step prior to final output. Several sectors may detect and record the same storm to the linked-list, as illustrated by Fig. 2. Before the storms detected by Stride Search can be saved to output, the list must be examined to remove duplicate entries.

Duplicates are removed by again referencing the user-specified scale s. Each pair of entries in the linked-list are compared; if a pair are separated by a distance less than s, they are considered duplicates and the less intense entry is deleted. In Fig. 2, this is demonstrated with pressure data. Each of the three search sectors have exceeded the storm identification criteria and independently locate their minimum pressure. The blue (left) sector finds a minimum of 988 hPa, the red (middle) finds 982 hPa, and the black (right) finds 984 hPa. These three entries are clearly separated by a distance less than s, as they are all contained within the red (middle) circle. The duplicate removal procedure will delete the blue (left) and black (right) entries because compared to the red (middle) entry they have higher pressures and are less intense. Only the red 982 hPa entry will be saved to output.

4 Data description

We use data produced by the the spectral element dynamical core of the Community Atmosphere Model, CAM-SE (Neale et al., 2012; Dennis et al., 2012). The model uses a cubed sphere grid and high resolution experiments set 240 elements per face of the cube for a total of 3 110 402 horizontal grid points. This results in a horizontal resolu-

GMDD
8, 7727–7765, 2015

tion of $\Delta\lambda \approx 0.125^{\circ}$ (Worley et al., 2011; Dennis et al., 2012). Due to well-known issues regarding the tuning of physical parameterizations within climate models, this high resolution simulation may not produce a climate that is a good match for the Earth. At such high resolutions, high intensity storms occur with unrealistically high frequencies (Reed and Jablonowski, 2011; Dennis et al., 2012).

The original goals of the high resolution experiments were to demonstrate the parallel scaling of CAM-SE, to document its required run time and related statistics in various high performance computing environments, and to demonstrate the model's capability to produce well-resolved features like tropical cyclones that cannot be represented well in low resolution experiments. Here, we choose this data because its high resolution ensures that small-scale storms will exist within the data in order to provide a good testing environment for storm detection algorithms. The fact that there may be an unrealistically high number of storms in the data is a benefit in this case.

The data set contains 5 years of simulated data that used CAM5 physics and preindustrial (year 1850) initial conditions. Instead of additional model components, the CAM-SE atmospheric dynamical core is coupled to a set of land, ocean, and sea ice data that also correspond to the year 1850 to provide its boundary conditions (Dennis et al., 2012). The land, ocean, and sea ice boundary conditions are periodic, with period 1 year, and simply repeat throughout the 5 year atmospheric simulation.

The model's native cubed sphere data was interpolated to uniform latitude-longitude grids using the regridding software provided by the Earth System Modeling Framework (ESMF) (Balaji et al., 2014). For the tropical cyclone test in Sect. 5, three months of the native data were interpolated to a $\Delta\lambda = 0.25^{\circ}$ mesh with $n_{\text{lon}} = 1440$. This ensures that any poorly resolved "$2\Delta x$" features within the simulation data will be smoothed, and provides the expected data layout of a climate data set on a uniform grid. For the polar search application in Sect. 6 the entire 5 years of data were similarly interpolated to a uniform mesh with $n_{\text{lon}} = 1024$ for a resolution of $\Delta\lambda = 0.35^{\circ}$. To facilitate a timing experiment (presented in the next section), we also interpolate 3 months data to coarser resolutions with $\Delta\lambda = 2, 1^{\circ}$, and 0.5°.

GMDD
8, 7727–7765, 2015

5 Tropical cyclones

In this section we apply both Stride Search and TSTORMS to the problem of tropical cyclone detection. Our aim is to validate the Stride Search by comparison with TSTORMS, which has a proven record. We also discuss the subtle differences between the two algorithms' results that may be of importance to climate researchers.

A tropical cyclone is identified within search sector K_{ij} if the following four criteria are met (Vitart et al., 1997):

1. There is a cyclonic vorticity maximum greater than a threshold value, τ_ζ:

$$\max_{i,j\in K_{ij}}\left[\operatorname{sgn}(\theta_i)\cdot\zeta_{850}(\lambda_j,\theta_i)\right] > \tau_\zeta. \tag{8}$$

2. The distance between the cyclonic vorticity maximum and the sector's sea level pressure minimum is less than a threshold value τ_{D_1}:

$$\operatorname{dist}\big((\lambda_\zeta,\theta_\zeta),(\lambda_\mathrm{P},\theta_\mathrm{P})\big) < \tau_{D_1}, \tag{9}$$

where $(\lambda_\zeta,\theta_\zeta)$ and $(\lambda_\mathrm{P},\theta_\mathrm{P})$ are the locations of sector K_{ij}'s vorticity maximum and sea level pressure minimum, respectively.

3. The difference between the vertically averaged temperature's maximum value and its sector average exceeds a threshold τ_T:

$$\max_{i,j\in K_{ij}}\overline{T}(\lambda_j,\theta_i) - \operatorname{avg}_{K_{ij}}(\overline{T}) > \tau_T, \tag{10}$$

where $\overline{T}$ is defined as

$$\overline{T}(\lambda,\theta) = \frac{1}{2}\big(T_{500}(\lambda,\theta) + T_{200}(\lambda,\theta)\big), \tag{11}$$

GMDD
8, 7727–7765, 2015

and T_{500} and T_{200} are the temperatures the 500 and 200 hPa pressure levels, respectively. The sector average $\mathrm{avg}_{K_{ij}}(\overline{T})$ is computed as a simple arithmetic average,

$$\mathrm{avg}_{K_{ij}}(\overline{T}) = \frac{1}{N_{K_{ij}}} \sum \{\overline{T}(\lambda_j, \theta_i) : (\lambda_j, \theta_i) \in K_{ij}\}, \tag{12}$$

where $N_{K_{ij}}$ is the number of data points in sector K_{ij}.

4. The distance between the maximum vertically averaged temperature and the sea level pressure minimum is less than a threshold value τ_{D_2}:

$$\mathrm{dist}\big((\lambda_T, \theta_T), (\lambda_\mathrm{P}, \theta_\mathrm{P})\big) < \tau_{D_2}, \tag{13}$$

where (λ_T, θ_T) is the location of the sector maximum of $\overline{T}$.

Differences between the two algorithms' detections arise due to the differences in the algorithms themselves. For computing the collocation criteria, Eqs. (9) and (13), TSTORMS uses the angular distance function (3) while Stride Search uses the geodesic distance function (4). Users of TSTORMS must specify τ_{D_1} and τ_{D_2} in angular units, while users of Stride Search must use units of length. The arithmetic averages of the vertically averaged temperature (Eq. 12) will be different from one algorithm to the other, because their sectors will contain a different number of data points. As a result criteria 2, 3, and 4 may behave differently for each algorithm.

5.1 Spatial search results

We apply both algorithms to the same data, three months (18 July – 18 October) from year 4 of the high-resolution simulation described in Sect. 4. We set TSTORMS $n = 12$, Stride Search $s = 450$ km, and use the threshold values shown in Table 1 for Eqs. (8)–(10), and (13).

GMDD
8, 7727–7765, 2015

Results from both spatial search algorithms are plotted in Fig. 3. Each dot represents a storm detected at one time step. All 6 hourly time steps over the entire three months are shown, colored by windspeed-dependent hurricane categories defined by the Saffir-Simpson intensity scale.

Both algorithms produce qualitatively similar results. Visually they appear to agree nearly perfectly on the identifiable storm tracks and intensities. They both have false positives over land and in the Southern Ocean. Stride Search produces more false positives and detects more low-intensity storms than TSTORMS. This is due to the fact that Stride Search sectors typically contain more data points than TSTORMS sectors. The larger number of points per sector reduces the sector average of the vertically averaged temperature (Eq. 12) compared to a TSTORMS sector at the same location. Thus, the warm core temperature excess criterion (Eq. 10) is more easily achieved using Stride Search than TSTORMS for the same value of τ_T, and causes more false positives in the spatial search. These will be eliminated by the temporal correlation algorithm.

5.2 Temporal correlation results

In this section we apply the temporal correlation algorithm, Algorithm 2, to the spatial search results. We use a minimum duration criterion, t_{min}, to remove any storms that do not span a length of time of at least t_{min}. Since tropical cyclones are inherently maritime events (Cotton and Anthes, 1989), at this stage we also apply a land mask to remove any tracks whose origins are not over water.

In Fig. 4 we show the storm tracks that correspond to the spatial search output of Fig. 3 with parameters $U_{\text{max}} = 15\,\text{m}\,\text{s}^{-1}$ and $t_{\text{min}} = 2\,\text{days}$. These results show that the temporal correlation algorithm succeeds in eliminating false positives and gives a better representation of the storms within the data set than the raw output from the spatial search algorithm. Table 2 presents the final storm counts for each algorithm for the 3 month data set, separated by hurricane category. Again, both algorithms produce nearly identical results which validates the present work.

GMDD
8, 7727–7765, 2015

The tracking algorithm (Algorithm 2) is sensitive to its two input parameters t_{min} and U_{max}. Choosing a value of t_{min} too low will not eliminate enough false positives, but choosing a value of t_{min} too high will eliminate weak cyclones that did exist but did not intensify enough to last long. The effects of choosing a too high or too low value for U_{max} are more subtle.

A value of U_{max} that is too low will cause storm tracks to fragment into several disjoint pieces. For example, there is a category 5 storm (red) that originates in the central Atlantic and eventually makes landfall in the southeastern US. Setting $U_{max} = 10.5\,m\,s^{-1}$ causes this track to split into two separate tracks – one a category 1 storm that ends over Cuba, and a second track that originates north of Cuba and develops into a category 5 storm. Setting $U_{max} = 11.5\,m\,s^{-1}$ keeps the track as one. This indicates that this particular storm increased in speed from ≈ 10 to $\approx 11\,m\,s^{-1}$ as it crossed Cuba. The lower value of U_{max} would result in this storm being erroneously counted twice. By contrast, choosing an unrealistically high value for U_{max} could cause the tracking algorithm to merge two storms that are in reality separate entities.

5.3 Performance and timing

As discussed previously, climate data sets are already large and are expected to increase in size as climate models run at high resolutions with $\Delta\lambda < 0.5^{\circ}$. Storm detection algorithms must be both accurate and efficient. In this section we document the dependence of each search algorithm's run time on the resolution of its input data set.

Differences in the structure of the two algorithms result in notable differences in the number of search sectors and the number of times each grid point is read from a file per time step. In the top row of Table 3 we present the total number of search sectors required by each algorithm to search the tropical domain, $\theta_{min} = 40^{\circ}$ S, $\theta_{max} = 40^{\circ}$ N for each data resolution. The number of TSTORMS's search sectors increases by a factor of 4 as the data resolution is halved, whereas the number of Stride Search sectors remains approximately constant across all data resolutions. The different shapes of each algorithm's sectors also cause differences in the number of data points associated

GMDD
8, 7727–7765, 2015

with each sector, and these numbers are presented in the second row of the table. The circular Stride Search sectors are linked to more data points than the corresponding TSTORMS sectors because the number of points in a Stride Search sector grows as a function of latitude.

In the last row of the table the total number of per-time step data accesses required by each algorithm are given; these numbers are the product of the number of sectors and the number of points per sector. The reduced overlap between sectors in the Stride Search algorithm result in many fewer data accesses (by orders of magnitude) than TSTORMS.

Our implementation of Stride Search in code favors simplicity over performance. Both TSTORMS and the current implementation of Stride Search run in serial on a single thread. Figure 5 shows the wall clock time required by each algorithm to search the 3 month data set for tropical cyclones on a standard desktop workstation using GNU's Fortran compiler. The same tests were run on one node of Sandia's Red Sky High Performance Computing cluster using the Intel Fortran compiler and produced similar results. Despite the differences presented in Table 3 between the two algorithms, we do not see a large advantage of one over the other. Profiling each code reveals that both algorithms are limited by their output subroutines, writing the list L from Algorithm 1 to disk.

TSTORMS performs slightly faster than Stride Search for low resolution data with $\Delta\lambda > 0.5^{\circ}$. Since TSTORMS performs all of its computations in data space it has very little computational overhead upon startup, while regardless of the data resolution Stride Search sectors have to be initialized prior to reading the data. This overhead causes Stride Search to perform slightly slower than TSTORMS for coarse data sets. At high resolutions, however, Stride Search performs faster than TSTORMS. Stride Search timing scales at the expected rate of $O(\Delta\lambda^{-2})$, while TSTORMS appears to scale at a higher rate.

GMDD
8, 7727–7765, 2015

6 Polar search

A key motivation for developing Stride Search was to provide a detection algorithm capable of searching all latitudes, including polar regions. The Arctic and Antarctic climates are becoming increasingly frequent subjects of study due to recent significant changes in these environments (Stocker et al., 2013), and a detection algorithm capable of searching data near the poles is necessary. Although grid point searches have been used at midlatitudes up to $\approx 60^\circ$ N and 60° S (König et al., 1993; Raible and Blender, 2004), their results at higher latitudes are questionable. Applying the grid point algorithm to polar search applications may miss storms because the zonal width of a search sector approaches zero at high latitudes while the spatial scale of a storm, of course, does not.

Figure 6a shows three consecutive grid point search sectors along $\theta = 60^\circ$ N. As in Fig. 1, the grid of data points has resolution $\Delta\lambda = 10^\circ$ and we have used the same $n = 2$ to set up grid point search sectors. The blue (horizontal stripes) sector is centered at $(\lambda_j, \theta_i) = (150^\circ \text{ E}, 60^\circ \text{ N})$ and the red (vertical stripes) and black (diagonal stripes) sectors are at $(\lambda_{j+1}, \theta_i) = (160^\circ \text{ E}, 60^\circ \text{ N})$ and $(\lambda_{j+2}, \theta_i) = (170^\circ \text{ E}, 60^\circ \text{ N})$, respectively. Each grid point search sector spans a distance of 5 grid points in latitude, or approximately 5000 km south to north.

The square grid point search sectors may appear correct in the left plot of Fig. 6a, a Mercator projection, but the problem with them is clear in the polar stereographic projection to the right. Far from the equator, angular distance is no longer approximately proportional to geodesic distance. The southern boundary of each sector lies along the 40° N latitude circle, where 5 grid points in longitude span 4000 km east to west. But the northern boundary of each sector, along 80° N, spans only about 1000 km east to west.

The constant geodesic radius of Stride Search sectors solves this problem. Figure 6b shows three Stride Search sectors along $\theta = 60^\circ$ N, with $s = 2500$ km. The blue (horizontally striped) sector is centered at $(\lambda_J, \theta_I) = (150^\circ \text{ E}, 60^\circ \text{ N})$. The red (vertically

GMDD
8, 7727–7765, 2015

striped) and black (diagonally striped) sectors are at $(\lambda_{j+1}, \theta_i) = (170^\circ\,\mathrm{W},\ 60^\circ\,\mathrm{N})$ and $(\lambda_{j+2}, \theta_i) = (130^\circ\,\mathrm{W},\ 60^\circ\,\mathrm{N})$, respectively. The longitude stride along $\theta = 60^\circ\,\mathrm{N}$ is twice as large as the longitude stride along the equator, so the three consecutive sectors in Fig. 6b cover twice as many longitude lines than the three sectors in Fig. 1b. The shapes of each sector in the left plot are due to the effects of the Mercator projection. All sectors are still circles on the sphere, as shown in the polar stereographic projection (right). Stride Search sectors – even one centered at the pole – have the same geographic size regardless of latitude, and are therefore capable of searching polar regions as effectively as midlatitude and tropical regions.

As an example application, we consider polar lows. A polar low is defined as a "small, but fairly intense maritime cyclone that forms poleward of the main baroclinic zone (the polar front or other main baroclinic zone). The horizontal scale of the polar low is approximately between 200 and 1000 km and surface winds near or above gale force," (Rasmussen and Turner, 2003). Polar lows may be quite destructive to coastal towns and to offshore drilling platforms, but typically decay rapidly over land or sea ice. They also contribute to the break-up of sea ice which has significant implications for the polar climate. Polar lows are distinct from synoptic low pressure systems due to their different developmental forcing and a typical lack of associated fronts (Montgomery and Farrell, 1992; Rasmussen and Turner, 2003).

Individual polar lows may develop a barotropic structure more similar to hurricanes than to baroclinic midlatitude storms (Nordeng and Rasmussen, 1992; Føre et al., 2012). These particularly interesting polar lows are typically smaller in diameter and duration than tropical cyclones, and therefore require a high resolution model such as the one described in Sect. 4 to resolve. A companion study will investigate the Arctic climatology of variable resolution climate models and their ability to simulate "hurricane-like" polar lows (Roesler et al., 2015).

A set of objective identification criteria for polar lows are given by Bracegirdle and Gray (2008), which we adapt to the Stride Search algorithm. Storm intensity is measured with vorticity and pressure and, as was the case with the tropical cyclone identifi-

GMDD
8, 7727–7765, 2015

cation criteria, a collocation requirement is applied. New to this application is a criterion that identifies regions of increased instability associated with cold air outbreaks over relatively warm ocean water.

Stride Search records a polar low in sector K_{ij} if the following criteria are met:

1. A sea level pressure minimum of sufficient intensity exists,

$$\min_{i,j\in K_{ij}} P_{sl}(\lambda_j, \theta_i) < \tau_P. \tag{14}$$

2. The storm originates over the ocean (not land or ice),

$$\min_{i,j\in K_{ij}} F(\lambda_j, \theta_i) \leq \tau_{\text{ice}}, \tag{15}$$

where F is the ice fraction and $\tau_{\text{ice}} \in [0, 1)$.

3. A cold air outbreak exists,

$$\min_{i,j\in K_{ij}} \left[\theta_{700}(\lambda_j, \theta_i) - \text{SST}(\lambda_j, \theta_i)\right] \leq \tau_T. \tag{16}$$

4. A cyclonic vorticity maximum of sufficient strength exists,

$$\max_{i,j\in K_{ij}} \left[\text{sgn}(\theta_i)\zeta(\lambda_j, \theta_i)\right] > \tau_\zeta. \tag{17}$$

5. The vorticity maximum must be collocated with the pressure minimum,

$$\text{dist}\big((\lambda_{\text{P}}, \theta_{\text{P}}), (\lambda_\zeta, \theta_\zeta)\big) < \tau_D. \tag{18}$$

Stride Search set up uses a sector radius $s = 450\,\text{km}$, and search region boundaries $\theta_{\min} = 50^\circ\,\text{N}$, and $\theta_{\max} = 90^\circ\,\text{N}$. Threshold values are set at $\tau_{\text{P}} = 980\,\text{hPa}$, $\tau_{\text{ice}} = 0$, $\tau_T = 7\,\text{K}$, $\tau_\zeta = 2.0 \times 10^{-4}\,\text{s}^{-1}$ and $\tau_D = 450\,\text{km}$. To the temporal correlation algorithm we add a minimum duration $t_{\min} = 12\,\text{h}$ and set $U_{\max} = 20\,\text{m s}^{-1}$.

GMDD
8, 7727–7765, 2015

The data include only temperature (not potential temperature) and do not include the 700 hPa pressure level. The required θ_{700} data are approximated as

$$\theta_{700} = \frac{1}{2}(\theta_{850} + \theta_{500}), \tag{19}$$

where $\theta_{850} = T_{850}\left(\frac{1000}{850}\right)^{0.286}$ and $\theta_{500} = T_{500}\left(\frac{1000}{500}\right)^{0.286}$.

Results from the entire 5 year data set are presented in Fig. 7, separated by season. Storm tracks are colored by their maximum strength on the U.S. National Weather Service's maritime warning scale (Bowditch, 2002); gale force storms (black) have maximum wind speeds $17.5 \le u_{\max} < 24.5\,\mathrm{m\,s^{-1}}$, storm force (blue) has $24.5 \le u_{\max} < 33\,\mathrm{m\,s^{-1}}$, and hurricane force storms (red) have maximum wind speeds greater than $33\,\mathrm{m\,s^{-1}}$. The results show the expected seasonal variation of storm frequencies, with the maximum number of storms and the maximum intensity of storms occurring in the winter (DJF) months. Spring (MAM) months show more activity over the pole than the fall (SON) months, and there are few storms in the summer (JJA).

Once storm tracks are built users may investigate individual storms more easily. In Fig. 8, for example, the 850 hPa relative vorticity of a "hurricane-like" polar low is shown. At the plotted time step, 25 December, year 3, 12:00 UTC, the storm is located at 86° N, 017° E. The perimeter of the plot is the 80° N latitude circle; hence, the plot's radius is approximately 1000 km, and the storm's diameter is approximately 500 km. This storm was chosen because of its symmetric structure, small spatial scale, and because it nearly traveled directly over the pole. We therefore see that Stride Search is capable of finding the event we sought, even at such extreme latitudes.

7 Conclusions

We have introduced the Stride Search algorithm for detection of extreme climate events within large climate data sets. The algorithm assumes data are given on uniform

GMDD
8, 7727–7765, 2015

latitude-longitude grids separated by time step, as is typical for standard NetCDF climate data files. Stride Search setup is based on search sectors that are circles of constant geodesic radius on the surface of the sphere, regardless of latitude. By contrast, a grid point search algorithm uses sectors that are squares in the latitude-longitude plane and cannot be used in polar regions.

Stride Search was designed to be flexible software, capable of searching data sets for a variety of extreme events. Such events must be described by a set of quantifiable identification criteria, and as examples we show results for detections of tropical cyclones and polar lows within a high-resolution data set. The capability to search polar latitudes is a new feature introduced by Stride Search, and was the primary motivation for its development.

To validate Stride Search we compare its output to the output of GFDL's TSTORMS software, the current standard tool for tropical cyclone detection. TSTORMS performs faster at coarse resolutions while Stride Search performs faster for high resolutions. As climate data sets move toward higher resolutions with $\Delta\lambda \ll 1^{\circ}$, the speed advantage of Stride Search over a grid point search will become more noticeable. Due to the differences in each algorithm, the specific threshold values for a particular application's identification criteria may vary between a grid point search and Stride Search. For example, the differently-shaped search sectors cause the tropical cyclone warm core criterion to behave slightly differently for TSTORMS than for Stride Search. However, even with these differences, our results show that both algorithms produce similar results and can be relied upon to produce similar tropical cyclone climatology statistics for the same data sets.

Both Stride Search and TSTORMS software may be improved via parallelization. It is already common to take advantage of temporal parallelism by applying the spatial search algorithm to multiple time steps and multiple files concurrently using several compute nodes. This may be implemented with customized run scripts or dedicated software such as NASA's Portable Distributed Script (PoDS) software (Kouatchou and Oloso, 2014), GNU Parallel (Tange, 2011), and the Toolkit for Extreme Climate Analysis

GMDD
8, 7727–7765, 2015

(Prabhat et al., 2012). However, there also remains a significant amount of unexploited parallelism in the climate search problem, as individual search sectors at the same time step may be distributed across intranode threads. We mark the parallel development of the Stride Search software as an item for future work.

Code availability

Our implementation of Stride Search is written in Fortran for data on uniform latitude-longitude grids. The software and its documentation are undergoing copyright review and will be released on GitHub (www.github.com) once this process is complete. Preliminary versions may be requested individually from the corresponding author.

Acknowledgements. This work was supported by Sandia National Laboratories' John von Neumann Postdoctoral Fellowship and by the Water Cycle and Climate Extremes Modeling project which is supported by the Office of Biological and Environmental Research in the DOE Office of Science. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

The authors thank GFDL and the creators of the TSTORMS software for providing and maintaining this publicly available software.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2015-4839J.

References

Balaji, V., Boville, B., Cheung, S., Clune, T., Collins, N., Craig, T., Cruz, C., da Silva, A., DeLuca, C., de Fainchtein, R., Eaton, B., Hallberg, B., Henderson, T., Hill, C., Iredell, M., Jacob, R., Jones, P., Kluzek, E., Kauffman, B., Larson, J., Li, P., Liu, F., Michalakes, J., Murphy, S., Neckels, D., O'Kuinghttons, R., Oehmke, B., Panaccione, C., Rosinski, J., Sawyer, W., Schwab, E., Smithline, S., Spector, W., Stark, D., Suarez, M., Swift, S., Theurich, G.,

GMDD
8, 7727–7765, 2015

Trayanov, A., Vasquez, S., Wolfe, J., Yang, W., Young, M., and Zaslavsky, L.: Earth System Modeling Framework (ESMF) version 6.3, available at: www.earthsystemmodeling.org (last access: January 2015), 2014. 7739

Blender, R., Fraedrich, K., and Lunkeit, F.: Identification of cyclone-track regimes in the North Atlantic, Q. J. Roy. Meteor. Soc., 123, 727–741, 1997. 7729, 7730, 7733

Bowditch, N.: The American Practical Navigator, Pub. No. 9, National Imagery and Mapping Agency, bicentennial edn., available at: http://www.nws.noaa.gov/om/marine/home.htm (last access: June 2015), 2002. 7748

Bracegirdle, T. J. and Gray, S. L.: An objective climatology of the dynamical forcing of polar lows in the Nordic seas, Int. J. Climatol., 28, 1903–1919, 2008. 7746

Cotton, W. R. and Anthes, R. A.: Storm and Cloud Dynamics, Academic Press, San Diego CA, USA, 1989. 7742

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., St-Cyr, A., Taylor, M. A., and Worley, P. H.: CAM-SE: a scalable spectral element dynamical core for the Community Atmosphere Model, Int. J. High Perform. C., 26, 74–89, 2012. 7738, 7739

Føre, I., Kristjánsson, J. E., Kolstad, E. W., Bracegirdle, T. J., Saetra, Ø., and Røsting, B.: A "hurricane-like" polar low fuelled by sensible heat flux: high-resolution numerical simulations, Q. J. Roy. Meteor. Soc., 138, 1308–1324, 2012. 7731, 7746

Geng, Q. and Sugi, M.: Variability of the North Atlantic cyclone activity in winter analyzed from NCEP-NCAR reanalysis data, J. Climate, 14, 3863–3873, 2001. 7733

Keyantash, J. and Dracup, J. A.: The quantification of drought: an evaluation of drought indices, B. Am. Meteorol. Soc., 83, 1167–1180, 2002. 7731

Kleppek, S., Muccione, V., Raible, C. C., Bresch, D. N., Koellner-Heck, P., and Stocker, T. F.: Tropical cyclones in ERA-40: a detection and tracking method, Geophys. Res. Lett., 35, 5, doi:10.1029/2008GL033880, 2008. 7729, 7730, 7733

Knutson, T. R., Sirutis, J. J., Garner, S. T., Vecchi, G. A., and Held, I. M.: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions, Nat. Geosci., 1, 359–364, 2008. 7733

König, W., Sausen, R., and Sielmann, F.: Objective identification of cyclones in GCM simulations, J. Climate, 6, 2217–2231, 1993. 7730, 7745

Kouatchou, J. and Oloso, A.: Portable Distributed Script (PoDS), Tech. Rep. GSC-16531-1, NASA, Goddard Space Flight Center, Greenbelt MD, USA, 2014. 7749

GMDD
8, 7727–7765, 2015

Montgomery, M. T. and Farrell, B. F.: Polar low dynamics, J. Atmos. Sci., 49, 2484–2505, 1992. 7731, 7746

Neale, R., Chen, C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., Lamarque, J., Marsh, D., Mills, M., Smith, A. K., Tilmes, S., Vitt, F., Morrison, H., Cameron-Smith, P., Collins, W. D., Iacono, M. J., Easter, R. C., Ghan, S. J., Liu, X., Rasch, P. J., and Taylor, M. A.: Description of the NCAR Community Atmosphere Model (CAM 5.0), Tech. Rep. NCAR/TN-486+STR, NCAR, Boulder CO, USA, 2012. 7738

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., 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.: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms, B. Am. Meteorol. Soc., 94, 529–547, 2013. 7730

Nguyen, K. C. and Walsh, K. J. E.: Interannual, decadal, and transient greenhouse simulation of tropical cyclone-like vortices in a regional climate model of the South Pacific, J. Climate, 14, 3043–3054, 2001. 7730

Nordeng, T. E. and Rasmussen, E. A.: A most beautiful polar low. A case study of a polar low development in the Bear Island region, Tellus A, 44, 81–99, 1992. 7746

Oladipo, E. O.: A comparitive performance analysis of three meteorological drought indices, J. Climatol., 5, 655–664, 1985. 7731

Prabhat, Ubel, O., Byna, S., Wu, K., Li, F., Wehner, M., and Bethel, E. W.: TECA: A parallel toolkit for extreme climate analysis, in: Third Workshop on Data Mining in Earth System Science (DMESS 2012), LBNL-5352E, International Conference on Computational Science (ICCS 2012), available at: http://vis.lbl.gov/Vignettes/2012-climatePatterns-vignette/ (last access: January 2015), Omaha NE, 2012. 7729, 7730, 7733, 7750

Raible, C. C. and Blender, R.: Northern hemisphere midlatitude cyclone variability in GCM simulations with different ocean representations, Clim. Dynam., 22, 239–248, 2004. 7745

Raible, C. C., Della-Marta, P. M., Schwierz, C., Wernli, H., and Blender, R.: Northern hemisphere extratropical cyclones: a comparison of detection and tracking methods and different reanalyses, Mon. Weather Rev., 136, 880–897, 2008. 7729, 7730, 7733

GMDD
8, 7727–7765, 2015

Ralph, F. M., Nieman, P. J., and Wick, G. A.: Satellite and aircraft observations of atmospheric rivers over the eastern North Pacific ocean during the winter of 1997/98, Mon. Weather Rev., 132, 1721–1745, 2004. 7731

Rasmussen, E. A. and Turner, J. (Eds.): Polar Lows: Mesoscale Weather Systems in the Polar Regions, Cambridge University Press, Cambridge, UK, 2003. 7746

Reed, K. A. and Jablonowski, C.: Assessing the uncertainty in tropical cyclone simulations in NCAR's community atmosphere model, J. Adv. Mod. Earth Sys., 3, 16, doi:10.1029/2011MS000076, 2011. 7739

Rew, R., Davis, G., Emmerson, S., Davies, H., Hartnett, E., Heimberger, D., and Fisher, W.: NetCDF 4.3.2, available at: http://www.unidata.ucar.edu/software/netcdf/ (last access: January 2015), 2014. 7730

Roesler, E. L., Bosler, P., and Taylor, M. A.: Arctic storms in a regionally refined atmospheric general circulation model, Geosci. Model Devel., in preparation, 2015. 7746

Sinclair, M. R.: An objective cyclone climatology for the Southern Hemisphere, Mon. Weather Rev., 122, 2239–2256, 1994. 7730

Stocker, T., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., Bréon, F.-M., Church, J., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J., Hartmann, D., Jansen, E., Kirtman, B., Knutti, R., Krishna Kumar, K., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G., Mokhov, I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L., Vaughan, D., and Xie, S.-P.: Technical summary, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., IPCC, Cambridge, UK, 2013. 7745

Tange, O.: GNU parallel – the command-line power tool, The USENIX Magazine, 36, 42–47, 2011. 7749

TSTORMS: Diagnosis and Detection of Tropical Storms in High-Resolution Atmospheric Models, available at: http://www.gfdl.noaa.gov/tstorms, (last access: 27 January, 2015) 1997. 7729, 7730

Vitart, F. and Stockdale, T. N.: Seasonal forecasting of tropical storms using coupled GCM integrations, Mon. Weather Rev., 129, 2521–2537, 2001. 7733

Vitart, F., Anderson, J. L., and Stern, W. F.: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations, J. Climate, 10, 745–760, 1997. 7729, 7730, 7733, 7740

GMDD
8, 7727–7765, 2015

Wernli, H. and Schwierz, C.: Surface cyclones in the ERA-40 dataset (1958-2001). Part I: Novel identification method and global climatology, J. Atmos. Sci., 63, 2486–2507, 2006. 7729, 7733

Worley, P. H., Craig, A. P., Dennis, J. M., Mirin, A. A., Taylor, M. A., and Vertenstein, M.: Performance of the community Earth system model, in: International Conference for High Performance Computing, Networking, Storage and Analysis (SC11), November 2011, Seattle WA, USA. doi:10.1145/2063384.2063457, 2011. 7739

Wu, L., Martin, J. E., and Petty, G. W.: Piecewise potential vorticity diagnosis of the development of a polar low over the Sea of Japan, Tellus A, 63, 198–211, 2011. 7731

Zarzycki, C. M. and Jablonowski, C. J.: A multidecadal simulation of Atlantic tropical cyclones using a variable resolution global atmosphere general circulation model, J. Adv. Mod. Earth Sys., 6, 805–828, 2014. 7730, 7733

Zhao, M., Held, I. M., Lin, S. J., and Vecchi, G. A.: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM, J. Climate, 22, 6653–6678, 2009. 7733

GMDD
8, 7727–7765, 2015

Table 1. Threshold values used for tropical cyclone detection.

	τ_ζ	τ_{D_1}	τ_T	τ_{D_2}
TSTORMS	$8.5 \times 10^{-4}\ s^{-1}$	4°	2 K	2°
Stride Search	$8.5 \times 10^{-4}\ s^{-1}$	450 km	2 K	225 km

GMDD
8, 7727–7765, 2015

Table 2. Storm counts by hurricane category for each data set; this table corresponds to the storm tracks shown in Fig. 4.

max. wind ($m\,s^{-1}$)	cat 0 $\lvert u\rvert < 33$	cat 1 $33 \le \lvert u\rvert < 43$	cat 2 $43 \le \lvert u\rvert < 50$	cat 3 $50 \le \lvert u\rvert < 58$	cat 4 $58 \le \lvert u\rvert < 70$	cat 5 $\lvert u\rvert \ge 70$	total
TSTORMS	1	4	4	5	8	5	27
Stride Search	1	6	4	6	8	5	30

GMDD
8, 7727–7765, 2015

Table 3. Numbers of sectors, maximum data points per sector, and total number of data accesses per time step to search tropical domain $\theta \in [40^\circ\text{S}, 40^\circ\text{N}]$ vs. data resolution. TSTORMS used $n = \{2, 3, 6, 12\}$ for the corresponding $\Delta\lambda$, and Stride Search used $s = 450$ km.

	TSTORMS				StrideSearch			
$\Delta\lambda$	2°	1°	0.5°	0.25°	2°	1°	0.5°	0.25°
number of sectors	7.38E3	2.92E4	1.16E5	4.62E5	1890	1818	1778	1736
max points per sector	25	49	169	625	17	63	251	1017
max data accesses	1.84E5	1.43E6	1.96E7	2.89E8	3.21E4	1.14E5	4.46E5	1.77E6

GMDD
8, 7727–7765, 2015

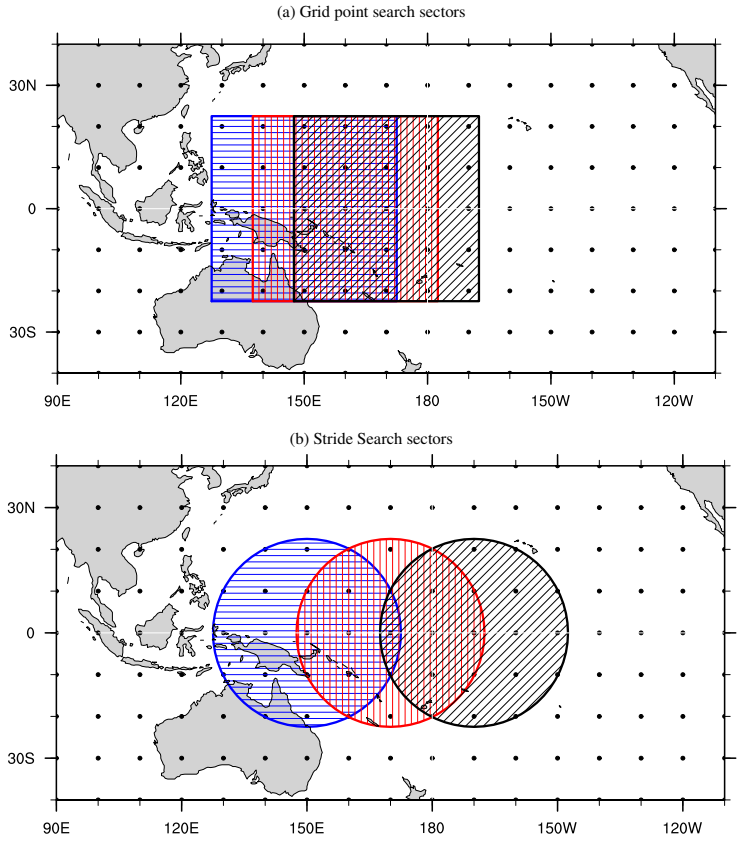

Figure 1. Search sectors along the equator. Black dots represent data points with resolution $\Delta\lambda = 10^{\circ}$. Blue sector (horizontal striping) center is $\lambda = 150^{\circ}$ E, $\theta = 0^{\circ}$ N. Red (vertical striping) and black (diagonal striping) sectors are the next two consecutive searches; **(a)** grid point search, $n = 2$; **(b)** Stride Search, $s = 2500$ km.

GMDD
8, 7727–7765, 2015

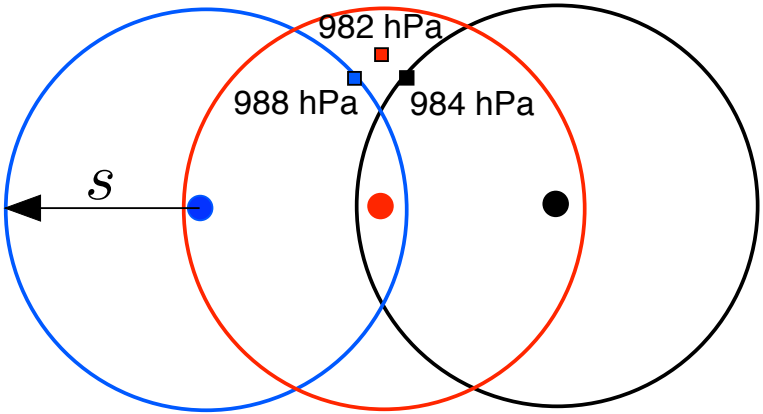

Figure 2. Duplicate detections of the same storm using the Stride Search algorithm. Each sector (blue/left, red/middle, and black/right) locates its minimum sea level pressure (corresponding squares). Duplicate entries are removed prior to the output stage; only the red 982 hPa entry will remain.

GMDD
8, 7727–7765, 2015

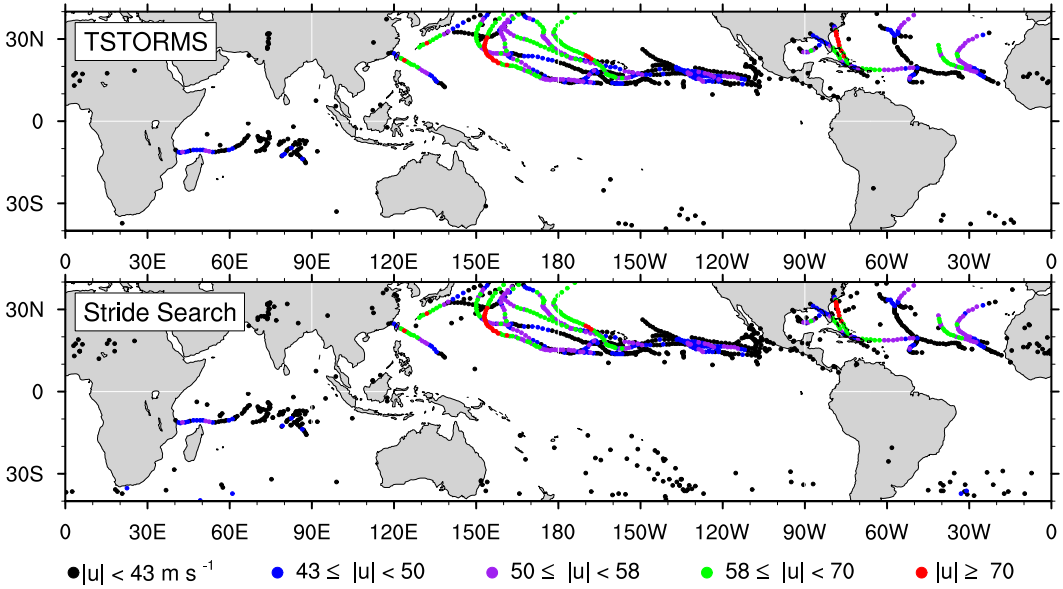

Figure 3. Spatial search results. Each dot is a storm detected at one time step; colors correspond to the categories of the Saffir-Simpson hurricane scale.

GMDD
8, 7727–7765, 2015

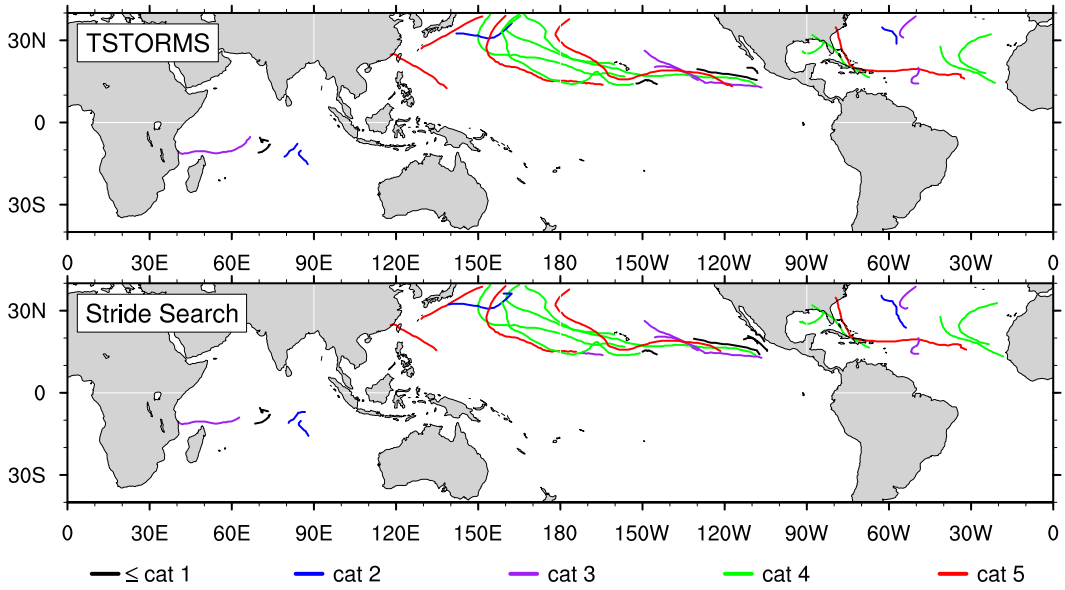

Figure 4. Storm tracks. The ultimate output of each algorithm for the 3 month data set. Each track is colored by the hurricane category corresponding to the maximum wind achieved over its lifetime.

GMDD
8, 7727–7765, 2015

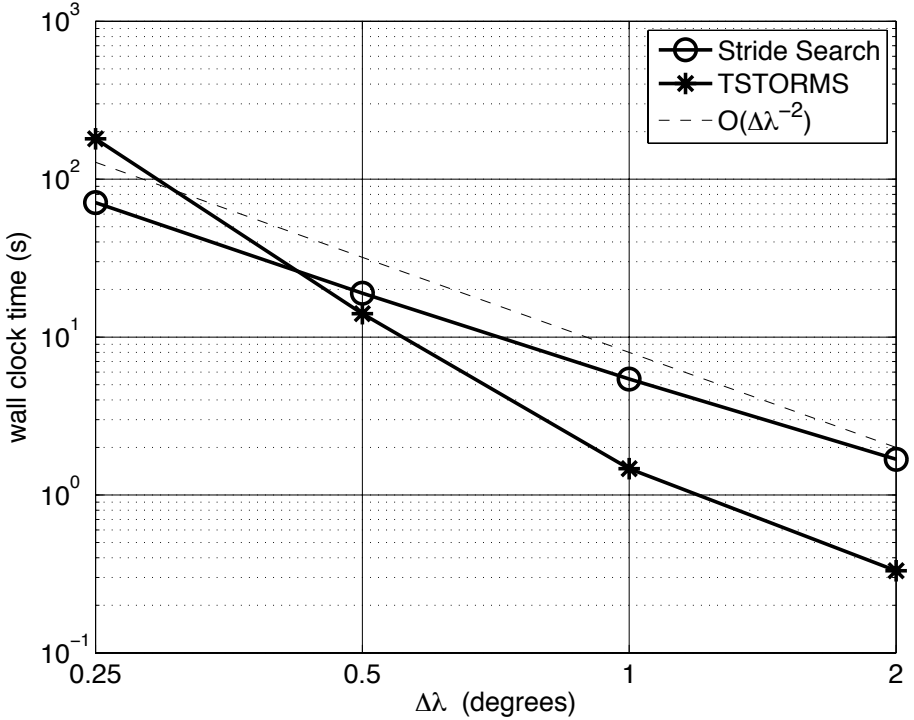

Figure 5. Wall clock time required to search three months of data for tropical cyclones, θ_{min} = 40° S, θ_{max} = 40° N, as a function of data resolution $\Delta\lambda$.

GMDD
8, 7727–7765, 2015

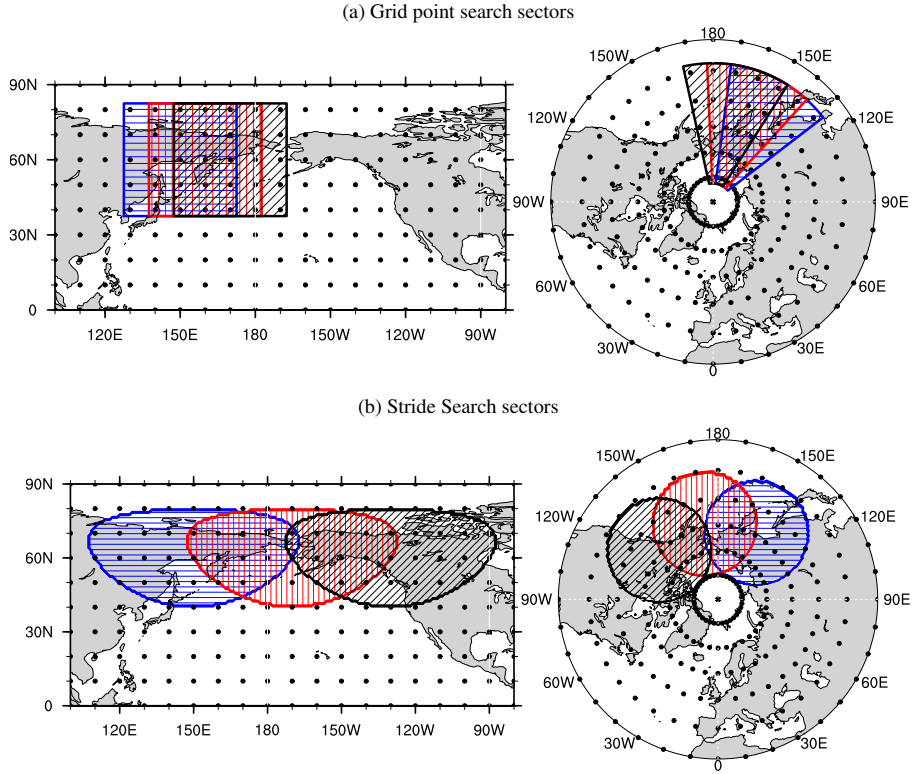

Figure 6. Search sectors along 60° N. Black dots represent data points with resolution $\Delta\lambda = 10°$. Blue sector center is $\lambda = 150°$ E, $\theta = 60°$ N. Red and black sectors are the next two consecutive searches; **(a)** TSTORMS, $n = 2$; **(b)** Stride Search, $s = 2500$ km.

GMDD
8, 7727–7765, 2015

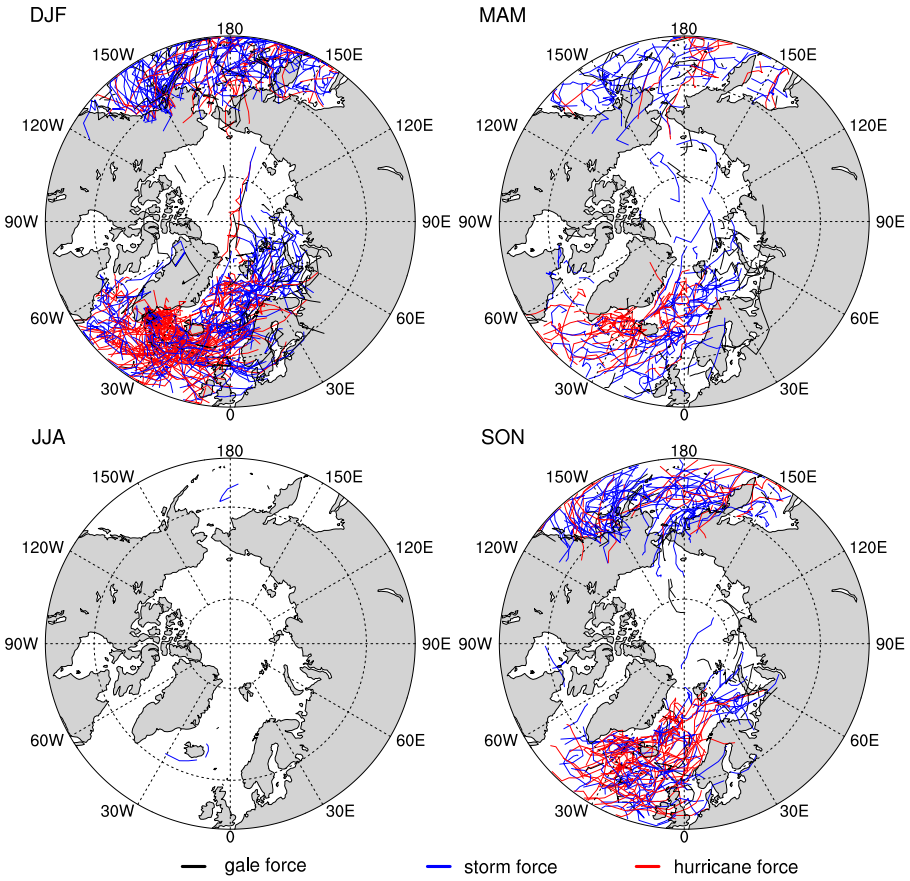

Figure 7. Northern Hemisphere polar low storm tracks. 5 years worth of data, separated by season; winter (DJF), spring (MAM), summer (JJA), fall (SON). Tracks are colored by the NOAA warning category corresponding to their lifetime maximum wind speed.

GMDD
8, 7727–7765, 2015

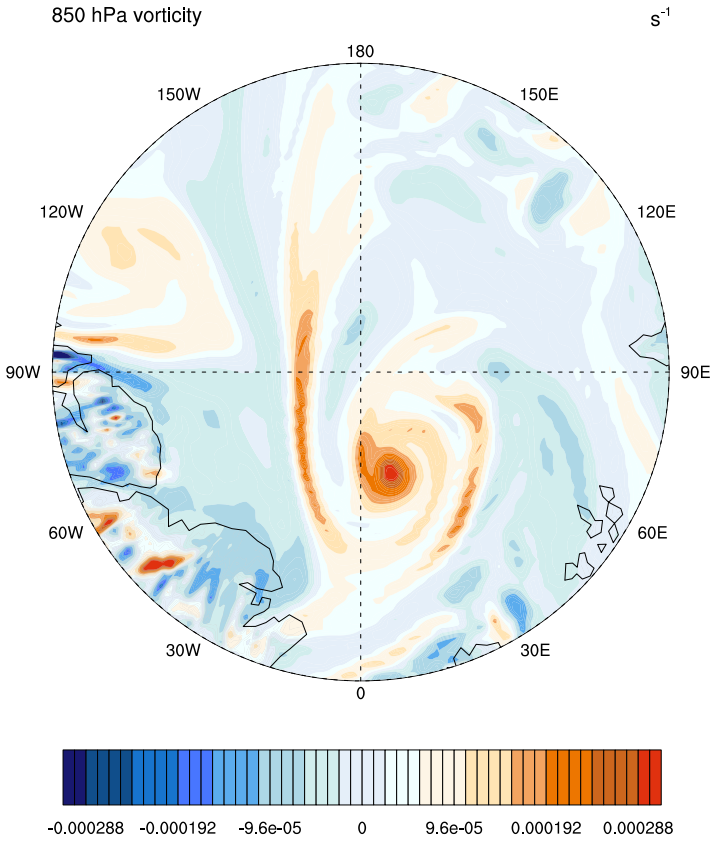

Figure 8. An example polar low located at 86° N 17° E with structural similarities to a tropical cyclone. The perimiter of the plot is the 80° N latitude circle.

GMDD
8, 7727–7765, 2015