A high-resolution unified observational data product of mesoscale convective systems and isolated deep convection in the United States for 2004 – 2017

Jianfeng Li[1*], Zhe Feng[1], Yun Qian[1*], L. Ruby Leung[1]

[1] Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington

Correspondence to Jianfeng Li (jianfeng.li@pnnl.gov) and Yun Qian (yun.qian@pnnl.gov)

Abstract

Deep convection possesses markedly distinct properties at different spatiotemporal scales. We present an original high-resolution (4 km, hourly) unified data product of mesoscale convective systems (MCSs) and isolated deep convection (IDC) in the United States east of the Rocky Mountains and examine their climatological characteristics from 2004 to 2017. The data product is produced by applying an updated Flexible Object Tracker algorithm to hourly satellite brightness temperature, radar reflectivity, and precipitation datasets. Analysis of the data product shows that MCSs are much larger and longer-lasting than IDC, but IDC occurs about 100 times more frequently than MCSs, with a mean convective intensity comparable to that of MCSs. Hence both MCS and IDC are essential contributors to precipitation east of the Rocky Mountains, although their precipitation shows significantly different spatiotemporal characteristics. IDC precipitation concentrates in summer in the Southeast with a peak in the late afternoon, while MCS precipitation is significant in all seasons, especially for spring and summer in the Great Plains. The spatial distribution of MCS precipitation amounts varies by seasons, while diurnally, MCS precipitation generally peaks during nighttime except in the Southeast. Potential uncertainties and limitations of the data product are also discussed. The data product is useful for investigating the atmospheric environments and physical processes associated with different types of convective systems, quantifying the impacts of convection on hydrology, atmospheric chemistry, and severe weather events, and evaluating and improving the representation of convective processes in weather and climate models. The data product is available at http://dx.doi.org/10.25584/1632005 (Li et al., 2020).

1 Introduction

In the atmosphere, deep convection refers to thermally driven turbulent mixing that displaces air parcels from the lower atmosphere to the troposphere above 500 hPa (Davison, 1999), leading to the development of convective storms. The heavy rain-rates associated with deep convection can significantly affect the water cycle (Hu et al., 2020) and other aspects such as soil erosion (Nearing et al., 2004), surface water quality (Carpenter et al., 2018; Motew et al., 2018), and managed and unmanaged ecosystems (Angel et al., 2005; Derbile and Kasei, 2012; Rosenzweig et al., 2002) that are essential elements of the biogeochemical cycle. By redistributing heat, mass, and momentum within the atmosphere, deep convection also has important effects on atmospheric chemistry (Anderson et al., 2017; Andreae et al., 2001; Choi et al., 2014; Grewe, 2007; Thompson et al., 1997; Twohy et al., 2002), large-scale environments (Houze Jr, 2004; Piani et al., 2000; Stensrud, 1996, 2013; Wang, 2003), and radiation balance (Feng et al., 2011; Zhang et al., 2017).

Besides its effects on the energy, water, and biogeochemical cycles, deep convection also has more direct societal impacts. As a significant source of natural hazards such as tornadoes, hail, wind gusts, lightning, and flash flooding, deep convection poses critical threats to human life and property (Brooks et al., 2003; Doswell III et al., 1996; Koehler, 2020; Taszarek et al., 2020). During 1950 – 1994, deep convection thunderstorms produced 47% of annual rainfall and up to 72% of summer rainfall on average east of the Rocky Mountains (Changnon, 2001b). During the same period, both the number of severe thunderstorms and deep convection precipitation has increased in most regions of the contiguous United States (CONUS) (Changnon, 2001a, b; Groisman et al., 2004). Folger and Reed (2013) found that hazards

56 associated with thunderstorms accounted for 57% of annual insured catastrophe losses since

57 1953. Since the 1980s, the inflation-adjusted economic losses due to convective storms increased

58 from about $5 billion to about $20 billion in the recent decade (https://www.iii.org/fact-

59 statistic/facts-statistics-tornadoes-and-thunderstorms). With warmer temperatures, the

60 environments of hazardous convective weather are projected to become more frequent in the

61 future (Diffenbaugh et al., 2013; Seeley and Romps, 2015), although few robust trends have

62 emerged in the recent decades (Houze Jr et al., 2019; Tippett et al., 2015).

63   The crucial roles of deep convection motivate the need for more accurate and

64 comprehensive datasets to improve understanding and modeling of this process and its impacts.

65 To this end, datasets with information on the location and time of occurrence, intensity, and other

66 properties of deep convection are necessary to understand and quantify its impacts on the

67 hydrologic cycle, severe weather hazards, large-scale circulations, etc. While field campaign data

68 can provide detailed information on deep convection properties, they are limited in space-time

69 coverage for statistical analysis. A corresponding reliable long-term dataset is undoubtedly

70 useful for model evaluation and development (Prein et al., 2017; Yang et al., 2017).

71   Deep convection can exist as isolated convective storms or organized storms with

72 mesoscale structures. A mesoscale convective system (MCS) is an aggregate of convective

73 storms organized into a larger and longer-lived system, which is the largest type of deep

74 convection. Due to their much longer duration and broader spatial coverage, MCSs generally

75 have stronger and longer-lasting influences on large-scale circulations than isolated deep

76 convection (IDC) events (Bigelbach et al., 2014; Stensrud, 1996, 2013). MCSs may also produce

77 higher rain rates, larger echo top heights, and greater water and ice masses than IDC (Rowe et

78   al., 2011, 2012). The enhanced rain rates in MCSs might be caused by larger amounts of ice

79   falling out and melting, higher amounts of liquid water below the melting level, and higher

80   concentrations of smaller drops (Rowe et al., 2011, 2012). Rowe et al. (2012) also suggested that

81   the enhanced rainfall from MCSs might be associated with more favorable environmental

82   conditions, such as higher convective available potential energy (CAPE) and wind shear. CAPE

83   and wind shear can impose different impacts on the initiation and evolution of IDC and MCSs

84   (French and Parker, 2008).

85   Considering the significant differences between IDC and MCS events, a reliable long-term

86   dataset not only describing the characteristics of deep convection but also separating IDC events

87   from MCSs is useful. With the deployment of operational remote sensing platforms such as

88   geostationary satellites and ground-based radar network several decades ago, scientists have

89   developed numerical algorithms to automatically detect deep convective systems and track their

90   evolutions over large areas and for long durations on the basis of continuous measurements from

91   remote sensors (Cintineo et al., 2013; Feng et al., 2011; Feng et al., 2012; Futyan and Del Genio,

92   2007; Geerts, 1998; Hodges and Thorncroft, 1997; Liu et al., 2007; Machado et al., 1998).

93   Objective tracking of deep convection has been applied to geostationary satellite data (Cintineo

94   et al., 2013; Sieglaff et al., 2013; Walker et al., 2012) and Next Generation Weather Radar

95   (NEXRAD) data (Haberlie and Ashley, 2019; Pinto et al., 2015) in the United States (US) over

96   different periods. However, a long-term climatological data product of MCS and IDC events

97   over the CONUS has heretofore not been developed.

98   Here, building on the work by Feng et al. (2019), which developed an algorithm for MCS

99   tracking and a dataset for MCSs for eastern CONUS, we produce a unified high-resolution data

100   product of both MCS and IDC events and analyze their characteristics east of the Rocky

101   Mountains for 2004 – 2017. The data product is developed by applying an updated Flexible

102   Object Tracker (FLEXTRKR) algorithm (Feng et al., 2018; Feng et al., 2019) and the Storm

103   Labeling in Three Dimensions (SL3D) algorithm (Starzec et al., 2017) to the NCEP (National

104   Centers for Environmental Prediction) / CPP (the Climate Prediction Center) L3 4 km Global

105   Merged IR V1 brightness temperature (T_b) dataset (Janowiak et al., 2017), the 3-D Gridded

106   NEXRAD Radar (Gridrad) dataset (Homeyer and Bowman, 2017), the NCEP Stage IV

107   precipitation dataset (Lin and Mitchell, 2005), and melting level heights from ERA5 (ECMWF,

108   2018). Section 2 describes the updated FLEXTRKR and SL3D algorithms in detail, as well as

109   the source datasets used by the algorithms. In Section 3, we first compare the climatological

110   characteristics between MCS and IDC events based on the MCS/IDC data product. Then, as an

111   application of the data product, we examine the spatiotemporal precipitation characteristics of

112   MCS and IDC events. In Section 4, we discuss the uncertainties and limitations of the data

113   product. Section 5 provides the availability information of the data product. Finally, we

114   summarize the study in Section 6.

115   **2 Source datasets and algorithms**

116   2.1 Source datasets

117   *2.1.1 Merged 4-km Infrared brightness temperature dataset*

118   In this study, we identify cold clouds associated with MCSs and IDC by using the NOAA

119   NCEP/CPP L3 half-hourly 4 km Global Merged IR V1 infrared T_b data for 2004 – 2017

120   (Janowiak et al., 2017). The dataset is a combination of various geostationary IR satellites with

121   parallax correction and viewing angle correction, therefore, providing continuous coverage

122   globally from 60°S – 60°N with a horizontal resolution of about 4 km and a temporal resolution

123   of 0.5 hours (Janowiak et al., 2001). We only use the hourly T_b data in the FLEXTRKR

124   algorithm discussed below, as all other datasets are only available at an hourly interval.

125   *2.1.2 Three-dimensional Gridded NEXRAD Radar (Gridrad) dataset*

126   Gridrad is an hourly 3-D radar reflectivity (Z_H) mosaic combining individual NEXRAD

127   radar observations to a Cartesian gridded dataset, with a horizontal resolution of $0.02° \times 0.02°$

128   and a vertical resolution of 1 km. The dataset covers 115° W to 69° W in longitude, 25° N to 49°

129   N in latitude, and 1 to 24 km in altitude above sea level (ASL). Homeyer and Bowman (2017)

130   produced the dataset by applying a four-dimensional binning procedure to merge level-2 Z_H data

131   from 125 National Weather Service (NWS) NEXRAD weather radars to Gridrad grid boxes at

132   analysis times. Only the level-2 observations within 300 km of each radar and 3.8 minutes of the

133   analysis time were used in the binning procedure. The Gridrad Z_H was the weighted average of

134   the level-2 observations within the Gridrad grid boxes to reduce the potential loss of information.

135   The weight calculation of each level-2 observation followed a Gaussian scheme in both space

136   and time. Observation weight was negatively correlated with the distance of the observation from

137   the source radar and the time difference between the observation and analysis time. The Gridrad

138   dataset provides the total weight of the level-2 observations within each Gridrad grid box, which

139   is useful for quality control. In addition, the number of level-2 radar observations (N_{obs}) and the

140   number of level-2 radar observations with echoes (N_{echo}) within each Gridrad grid box around

141   analysis times (± 3.8 min) are also available in the Gridrad dataset.

142   We obtain the Gridrad datasets between 2004 and 2017 from NCAR/UCAR Research Data

143   Archive (RDA) (https://rda.ucar.edu/datasets/ds841.0/, last access: Jan 2, 2020). Following the

144   quality control criteria of Homeyer and Bowman (2017) (http://gridrad.org/software.html, last

145   access: Jan 22, 2020), we remove potential low-quality observations, scanning artifacts, and non-

146   meteorological echoes from biological scatters and artifacts. Then we regrid Gridrad Z_H onto the

147   4 km satellite Merged IR grids by using the "bilinear" method from the Earth System Modeling

148   Framework (ESMF) Python module (https://www.earthsystemcog.org/projects/esmpy/) as

149   follows.

150   First, we convert the Gridrad logarithmic reflectivity Z_H to linear reflectivity (Z': mm^6 m^{-3}).

151   We then set Z' in grid boxes with radar observations but no echoes ($N_{obs} > 0$, but $Z_H = NAN$;

152   NAN, Not-A-Number) to 0 ($Z' = 0$). Here the physical interpretation is that NEXRAD scans

153   those grid boxes, but no detectable hydrometers return any echo. The primary motivation of this

154   procedure is to avoid the reduction of the number of valid reflectivity values after re-gridding, as

155   the ESMF bilinear method treats destination point as NAN as long as there is one NAN value in

156   the source points. A common scenario is at the edge between hydrometeor echoes and clear air.

157   Setting Z' of those grid boxes having radar observations but no echoes to NAN would cause all

158   surrounding destination points to become NAN even though all other source points have valid Z'

159   values, which would reduce the number of re-gridded valid Z_H ($Z_H \neq NAN$) by about 20% for

160   2004 – 2017. After the "bilinear" re-gridding of Z', we convert the linear reflectivity Z' back to

161   the logarithmic reflectivity Z_H. And we set Z_H equal to NAN for those grid boxes with Z' equal

162   to 0. Now the NAN values are acceptable and won't affect the SL3D algorithm and FLEXTRKR

163   algorithm discussed below.

164 *2.1.3 NCEP Stage IV precipitation dataset*

165   The NCEP Stage IV precipitation dataset provides hourly rain accumulations over polar

166 stereographic grids across the CONUS with a resolution of 4.76 km at 60°N since 2002. The

167 dataset is a mosaic of precipitation estimates from 12 River Forecast Centers (RFCs) over the

168 CONUS (Stage IV data in Alaska and Puerto Rico are archived separately) (Lin and Mitchell,

169 2005; Nelson et al., 2016). Each RFC produces its precipitation estimates through a combination

170 of radar and rain gauge data based on the multisensory precipitation estimator (MPE) algorithm

171 (for most RFCs), P3 algorithm (for Arkansas-Red basin RFC), or Mountain Mapper algorithm

172 (for California-Nevada, Northwest, and Colorado-basin RFCs with missing radar-derived

173 estimates) (Nelson et al., 2016). Some manual quality control steps are conducted to remove bad

174 radar and gauge data before radar-gauge merging (Lin and Mitchell, 2005; Nelson et al., 2016).

175 The Stage IV dataset has been widely used as a basis to evaluate model simulations, satellite

176 precipitation estimates, and radar precipitation estimates (Davis et al., 2006; Gourley et al., 2011;

177 Kalinga and Gan, 2010; Lopez, 2011; Yuan et al., 2008). Here, we obtain the hourly Stage IV

178 precipitation for 2004 –- 2017 from the NCAR/UCAR RDA

179 (https://rda.ucar.edu/datasets/ds507.5/, last access: Dec 28, 2019). We regrid the original Stage

180 IV precipitation from polar stereographic grids to the 4 km satellite Merged IR grids by using the

181 "neareststod" method from the ESMF 'NCL' module

182 (https://www.ncl.ucar.edu/Applications/ESMF.shtml). The "neareststod" method maps each

183 destination point to the closest source point.

184   *2.1.4 ERA5 melting level dataset*

185   Melting hydrometeors produce intense radar echoes in a horizontal layer about 0.5 km thick

186   located just below the 0°C level (melting level), which is known as "bright band" (Giangrande et

187   al., 2008; Steiner et al., 1995). The bright-band signatures are often pronounced for stratiform

188   precipitation, while convective precipitation produces well-defined vertical cores of maximum

189   reflectivity, diluting bright-band signals (Giangrande et al., 2008; Steiner et al., 1995).

190   Therefore, the SL3D algorithm that is described below examines Z_H above the melting level to

191   avoid the false identification of stratiform rain as convective (Starzec et al., 2017). In this study,

192   we use the hourly melting level heights from the ERA5 reanalysis dataset.

193   ERA5, as the successor to ERA-Interim, contains many modeling improvements and more

194   observations based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting

195   System (IFS) at the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5

196   provides hourly estimates of atmospheric variables at a horizontal resolution of 31 km and 137

197   vertical levels from the surface to 0.01 hPa from 1979 to the present (Hersbach et al., 2019). We

198   obtain ERA5 "Zero degree level" (melting level heights above ground) for 2004 – 2017 and

199   "Orography" (geopotential at the ground surface) from the Climate Data Store (CDS) disks

200   (ECMWF, 2018) (last access: Jan 24, 2020). The CDS archived ERA5 variables have been

201   interpolated to regular latitude/longitude grids with a resolution of 0.25° × 0.25°. We calculate

202   melting level heights ASL from "Zero degree level" and "Orography" (divided by 9.80665 m s^{-2}

203   to obtain ground surface height). Finally, we regrid the hourly 0.25° melting level heights ASL

204   to the 4-km satellite Merged IR grids by using the ESMF "neareststod" method.

205   We summarize the basic information of the four types of source datasets in Table S1. And,

206   we define our data product domain as 110°W – 70°W in longitude and 25°N – 51°N in latitude

207   (Figure 1), which covers the US east of the Rocky Mountains and excludes the western US. The

208   domain coverage takes into consideration the availability of the GridRad radar dataset, the

209   relatively scarce radar coverage over the Rocky Mountains, and associated uncertainties in radar-

210   based Stage IV precipitation estimates in complex terrains (Nelson et al., 2016). As shown in

211   Figure 1a, we further define four regions in the domain following Feng et al. (2019): Northern

212   Great Plains (NGP), Southern Great Plains (SGP), Southeast (SE), and Northeast (NE).

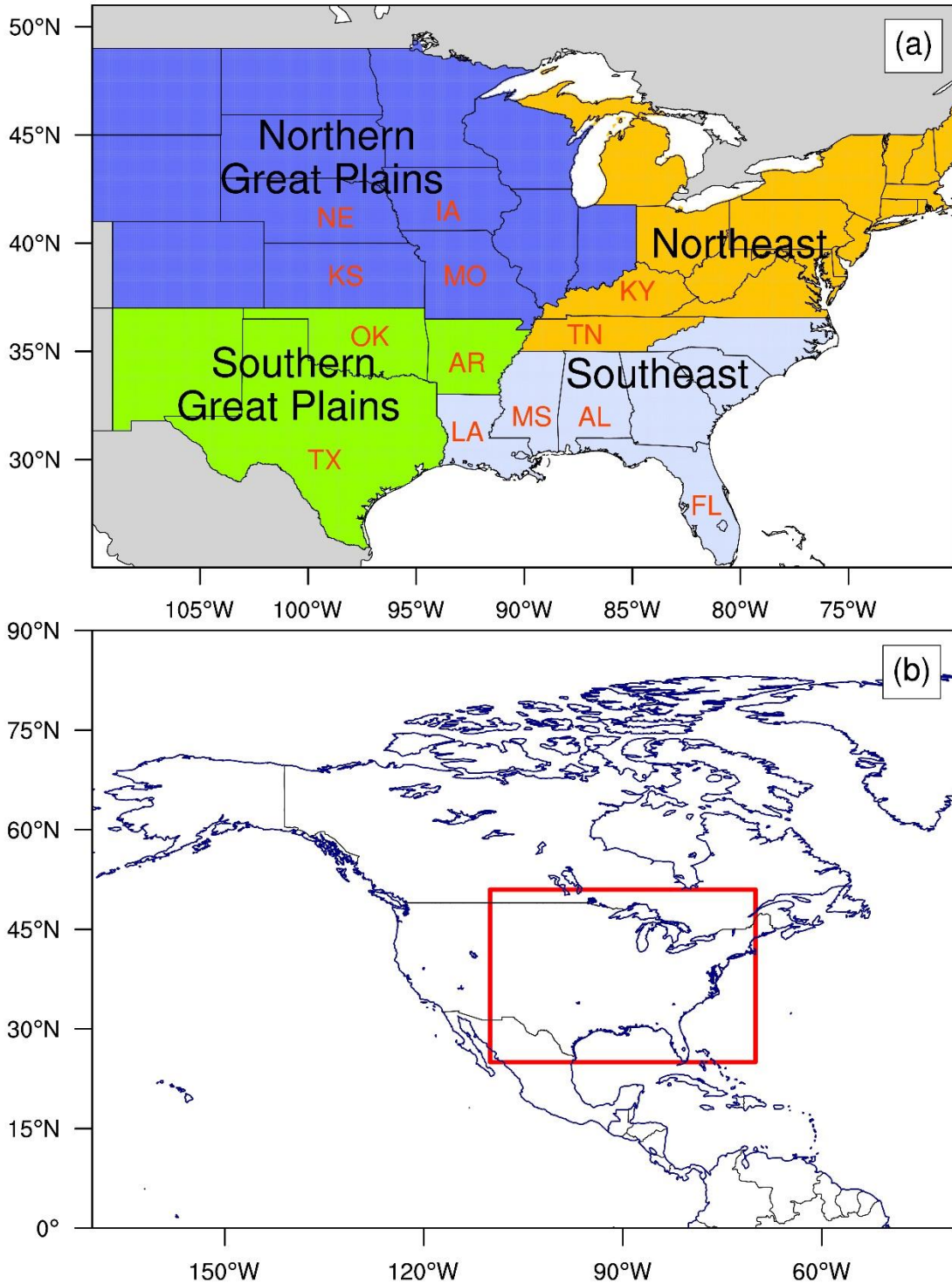



Figure 1. (a) Data product domain and region definitions. Blue shading denotes the Northern
Great Plains (NGP), green-yellow shading denotes the Southern Great Plains (SGP), light steel
blue shading denotes the Southeast (SE), and orange shading denotes the Northeast (NE). The
locations of some US states within each region are also labeled. TX is for Texas, OK for
Oklahoma, KS for Kansas, NE for Nebraska, IA for Iowa, MO for Missouri, AR for Arkansas,

219 LA for Louisiana, MS for Mississippi, AL for Alabama, TN for Tennessee, KY for Kentucky,
220 and FL for Florida. (b) The location of the data product domain (red box) in North America.

221 2.2 Algorithm description

222 *2.2.1 SL3D algorithm*

223   The SL3D algorithm exploits Gridrad Z_H to classify each grid column with radar echoes

224 into five categories: convective, precipitating stratiform, non-precipitating stratiform, anvil, and

225 convective updraft (Starzec et al., 2017). SL3D identifies these five categories successively

226 following the criteria listed in Table S2. We run the SL3D algorithm for 2004 – 2017 by using

227 the re-gridded ERA5 melting level heights and Gridrad Z_H dataset described in Section 2.1.

228 Figure 2e shows an example of the SL3D classification results based on Gridrad Z_H (Figure 2d)

229 at 2005-07-04T03:00:00Z. A sizeable convective system with intense radar echoes and

230 precipitation is observed in Kansas, and many isolated convection events are also observed in the

231 Southeast. The SL3D classification results will be used in the following FLEXTRKR algorithm

232 to identify convective core features (CCFs, continuous updraft/convective areas with

233 precipitation > 0 mm h^{-1}, which are used to indicate the existence of convective activity in the

234 IDC definition; red regions in Figure 3) and precipitation features (PFs, continuous

235 updraft/convective/precipitating-stratiform areas with precipitation > 1 mm h^{-1}; green areas in

236 Figure 3, which are used to denote the sizes of convective systems in the MCS and IDC

237 definitions).

238   *2.2.2 MCS/IDC identification and tracking*

239   The FLEXTRKR algorithm was first developed and used by Feng et al. (2019) to track

240   MCSs. In this study, we further update the algorithm so that it can identify and track MCS and

241   IDC events simultaneously.

242   Figure 3 displays the schematic of FLEXTRKR (Feng et al., 2019). The first step is to

243   identify cold cloud systems (CCSs; continuous areas with $T_b < 241$ K) at each hour by applying a

244   multiple T_b threshold "detect and spread" approach (Futyan and Del Genio, 2007). We search for

245   cold cloud cores with $T_b < 225$ K and spread the cold cloud cores to contiguous areas with $T_b <$

246   241 K. Cloud systems that do not contain a cold cloud core but with $T_b < 241$ K are also labeled

247   as long as they can form continuous areas with at least 64 km^2 (4 pixels). In addition, as

248   described in Feng et al. (2019), CCSs that share the same coherent precipitation feature are

249   combined as a single CCS. A coherent precipitation feature is defined as continuous areas with

250   smoothed Z_H at 2 km > 28 dBZ (if Z_H is not available at 2 km, use Z_H at 3 km instead if it is

251   available) (Feng et al., 2019). We use a 5×5 pixel moving window to smooth Z_H. Figure 2b

252   shows an example of the CCSs identified in the first step based on T_b at 2005-07-04T03:00:00Z.

253   "Cloud 1" in Figure 2b corresponds to a large area of low T_b in the central US (Figure 2a).

254   In step 2, CCSs between two consecutive hours are linked if their spatial overlaps are >

255   50%. "Linked" means the CCSs are considered to be from the same cloud systems. FLEXTRKR

256   produces tracks by extending the link between two consecutive time steps to the entire tracking

257   period, as shown in Figure 3. Each track represents the lifecycle of a cloud system. We calculate

258   a series of CCS summary statistics associated with each track, such as CCS-based lifetime of the

259   track (the duration of the track when CCSs are present), CCS area, CCS major axis length, CCS

260   propagation speed, etc. Besides, SL3D classification (Figure 2e) and Stage IV precipitation

261   (Figures 2c) within the tracked CCS are associated with the tracks and their merges and splits

262   (described below). Then, we can obtain CCF and PF statistics of each track, such as convective

263   and stratiform area, precipitation intensity and coverage, radar-derived echo-top heights, PF

264   major axis length, CCF major axis length, intense convective cells (convective cells with column

265   maximum reflectivity $\geq$ 45 dBZ and precipitation > 1 mm h^{-1}; pink areas in Figure 3, which are

266   used to indicate intense convective activity in the following MCS definition), etc.

267   Merging and splitting refer to situations when two or more CCSs are linked to one CCS

268   between consecutive hours (Figures S1 and S2). A track associated with the largest CCS is

269   defined as the main track (Figure S3), and smaller tracks from merges/splits are regarded as parts

270   of the main track when calculating PF and CCF statistics. In the algorithm, we require that a

271   "merge"/"split" track associated with an MCS/IDC event must have a CCS-based lifetime of no

272   more than 5 hours. Otherwise, we treat it as an independent track.

273   The identification of MCS and IDC is based on the CCS, PF, and CCF statistics of the

274   tracks. Following the definition of MCSs by Feng et al. (2019) (Figure 4), we define a track as an

275   MCS if it satisfies the following criteria: 1) there is at least one pixel of cold cloud core during

276   the whole lifecycle of the track; 2) CCS areas associated with the track surpass 60,000 km^2 for

277   more than six continuous hours; 3) PF major axis length exceeding 100 km and intense

278   convective cell areas of at least 16 km^2 exist for more than five consecutive hours. Considering

279   the lack of a strict and universal MCS definition (Geerts et al., 2017; Haberlie and Ashley, 2019;

280   Pinto et al., 2015; Prein et al., 2017), we evaluate the impact of different MCS definition criteria

281   on the data product in Section 4.4. For the non-MCS tracks, we further identify IDC with the

282   following two criteria (Figure 4): 1) a CCS with at least 64 km^2 (4 pixels) is detected; 2) at least

283   1 hour during the lifecycle of the track when PF and CCF are present (PF and CCF major axis

284   lengths $\geq$ 4 km). In addition, for each IDC event, the CCS-based lifetime of associated merge

285   and split tracks cannot surpass the lifetime of the IDC event. Here, the IDC criteria denote a low

286   limit in convective signals that we can identify by using the FLEXTRKR algorithm and given

287   source datasets. Potential uncertainties associated with the limit are discussed in Section 4.3.

288   Note that while we designate the term IDC to differentiate smaller convective storms from

289   MCSs, there are sub-categories of deep convection within IDC. For example, multicellular

290   convection systems that do not grow large enough or last long enough to meet our MCS

291   definition are defined as IDC in our study, even though they are not necessarily "isolated." Users

292   of the data product can further separate sub-categories within IDC using the derived CCF

293   statistics information to address specific science questions or research objectives.

294   Finally, the FLEXTRKR algorithm maps MCS/IDC track information back to the domain

295   pixels. Figures 2f – 2i give an example of the pixel-level MCS/IDC information at 2005-07-

296   04T03:00:00Z. Figure 2f displays the spatial coverages of MCS/IDC tracks at that time at pixel

297   scale and the corresponding unique numbers of these tracks. From Figure 2f, we know whether a

298   pixel belongs to an MCS/IDC track and the number of the track if the pixel belongs to a track.

299   We can further determine whether the track is an MCS or IDC event from Figure 2g, which

300   shows the types (MCS or IDC) of the tracks in Figure 2f at the pixel scale. Figures 2h and 2i are

301   similar to Figures 2f and 2g, respectively. The difference is that Figures 2h and 2i only show

302   pixels with precipitation > 1 mm h^{-1} in that hour. Together, the track-based CCS, PF, and CCF

303   statistics of MCS and IDC events and the pixel-level dataset constitute the unified high-

304   resolution MCS/IDC data product we develop in this study. Original T_b (Figure 2a), Stage IV

305   precipitation (Figure 2c), Gridrad Z_H at 2 km (Figure 2d), and Gridrad derived echo-top heights

306   are also archived in the data product.

307   We run the FLEXTRKR algorithm separately for each year from 2004 to 2017. The starting

308   time of each continuous tracking is 00Z on 1 January, and the ending time is 23Z on 31

309   December. Because winter has the fewest deep convection events, very few MCS/IDC events

310   extend between two different years based on our investigation. Also, the lifetimes of MCS/IDC

311   events are much shorter compared to our tracking period. Therefore, running FLEXTRKR

312   separately for each year rather than continuously for the whole period has little impact on the

313   MCS/IDC statistics.

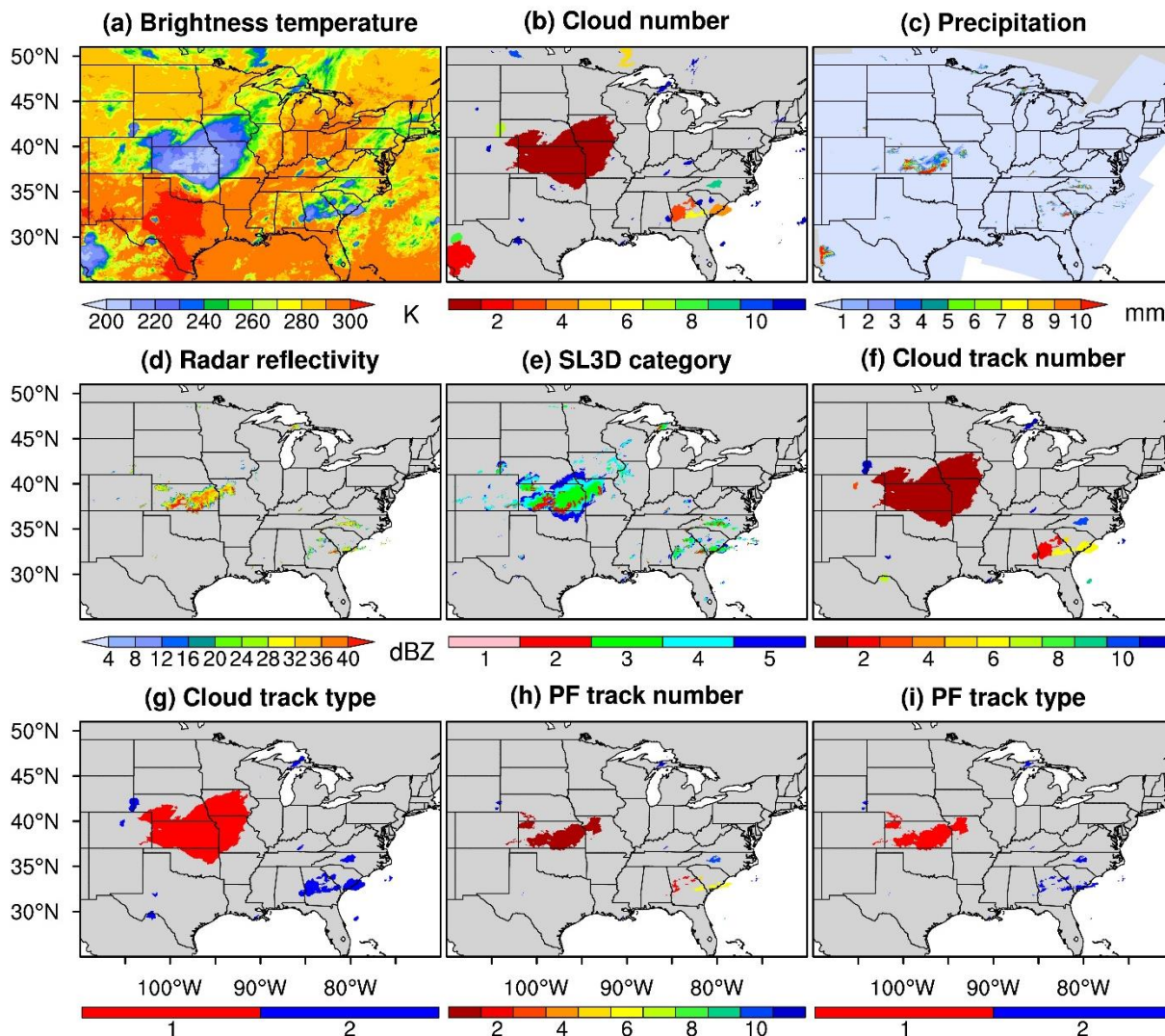
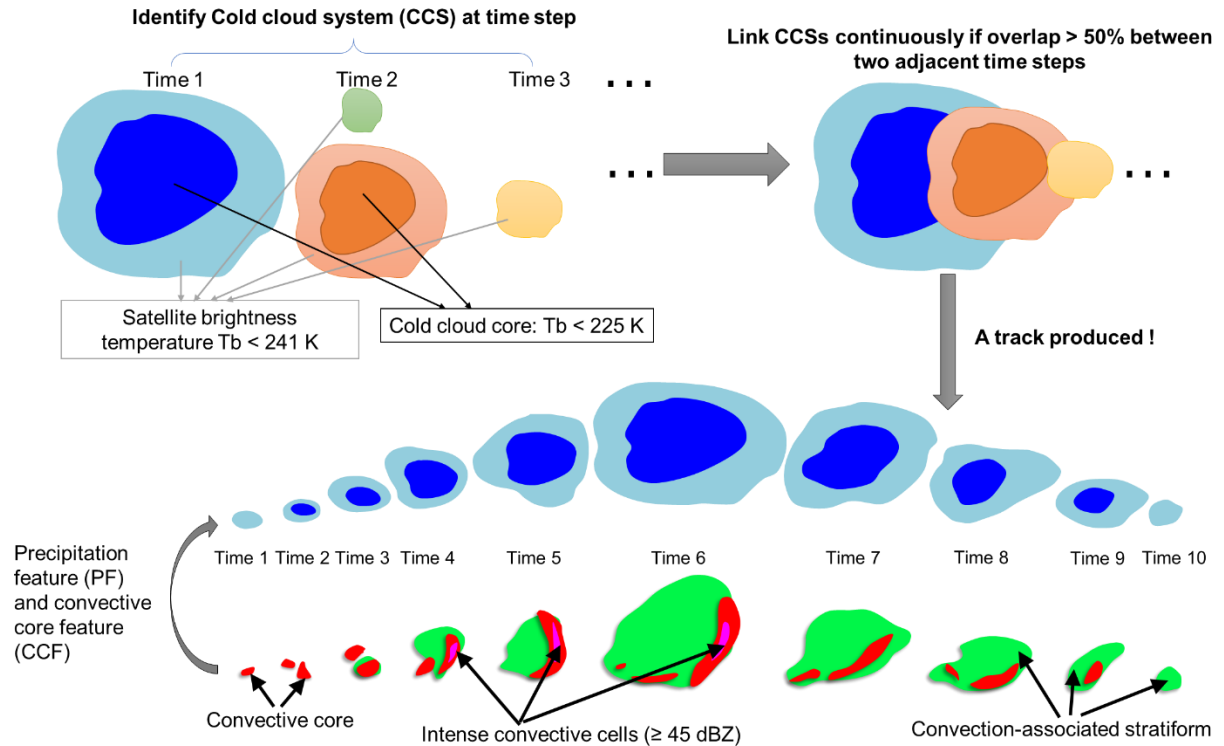
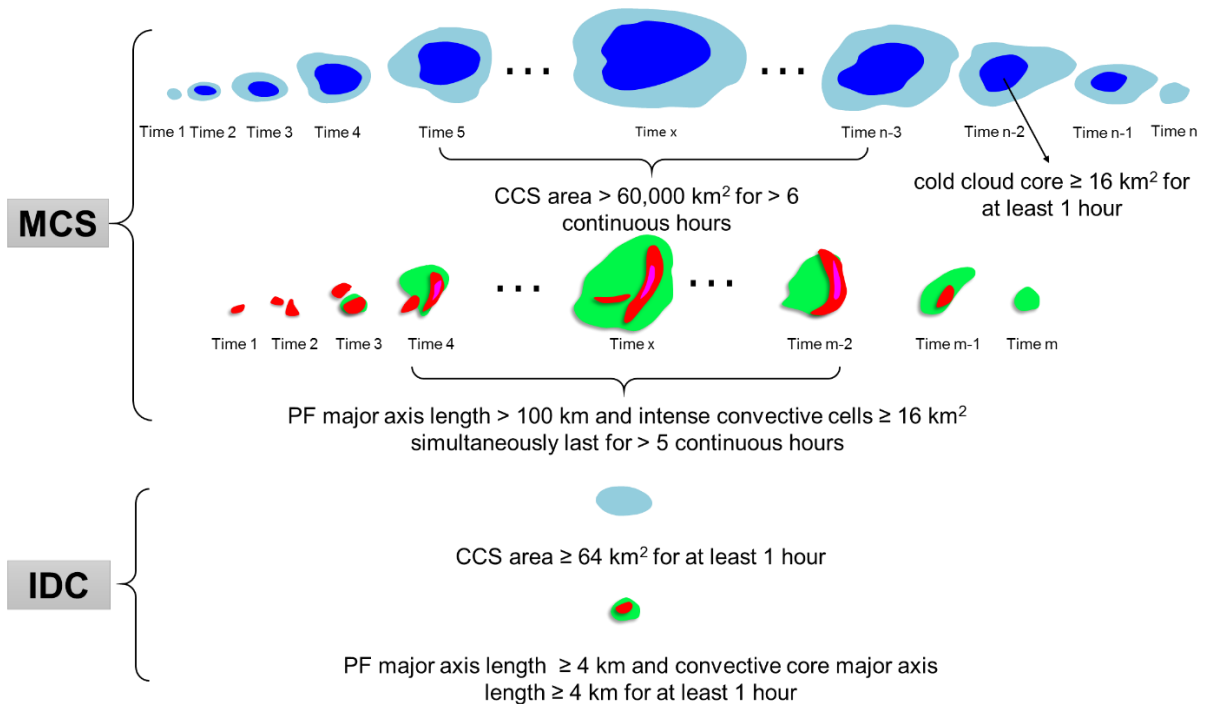



Figure 2. FLEXTRKR pixel-level outputs at 03:00:00Z on July 4, 2005. (a) is satellite T_b. (b) shows identified CCS labels. CCS labels are unique at each hour. (c) is Stage IV hourly accumulated precipitation. (d) is Gridrad Z_H at 2 km (if it is not available, Z_H at 3 km is provided if it is available). (e) is the SL3D classification results: 1, convective updraft; 2, convective; 3, precipitating stratiform; 4, non-precipitating stratiform; 5, anvil. (f) displays the track numbers to which pixels belong. Here, the track numbers are not the real values in the MCS/IDC data product. The track numbers should be unique throughout the whole running period. We adjust the track numbers here to make the figure clear. Similar to "PF track number." (g) gives information on whether the pixels belong to MCS (marked as 1) or IDC (marked as 2) tracks, which correspond to the tracks shown in (f). (h) also displays the track numbers to which the pixels belong, but only for pixels with precipitation > 1 mm h^{-1}. (i) is like (g) but corresponds to (h). All these variables are stored in the FLEXTRKR hourly pixel-level output files.

327
328   Figure 3. Schematic of the FLEXTRKR algorithm.

329
330   Figure 4. Definition of MCSs and IDC.

3 Results and discussions

3.1 Climatological characteristics of MCS and IDC events

According to the MCS/IDC data product, we identify 45,346 IDC and 454 MCS events each year on average between 2004 and 2017 in our data product domain. Summer (June – August) has the most IDC and MCS events with average numbers of 25,073 and 212, while winter has the least with average quantities of 2,545 and 37. During spring and autumn, there are 8,543 and 9,185 IDC events and 122 and 83 MCSs, respectively. The seasonal feature with the most occurrences of MCSs in winter and the least in summer is consistent with the results of Geerts (1998) in the Southeast US and Haberlie and Ashley (2019) over portions of the CONUS east of the Continental Divide (ECONUS).

We compare the climatological characteristics of MCS and IDC events in Table 1. MCSs have much longer lifetimes than IDC, averaging 21.1 hours (CCS-based) and 18.9 hours (PF-based), compared to 2.1 hours (CCS-based) and 1.7 hours (PF-based) for IDC. Here, PF-based lifetime refers to the lifetime determined by the MCS/IDC PFs. Only those hours with a significant PF present (PF major axis length > 20 km for MCSs; $\geq$ 4 km for IDC) are counted during the lifecycle of an MCS/IDC event, which represent the active convective period of a storm. We find that MCSs have the longest PF lifetime in winter (21.3 hours) and the shortest in summer (17.9 hours). In comparison, IDC has the longest PF lifetime in winter (1.9 hours), but the summer lifetime (1.7 hours) is comparable to spring and autumn. We examine the seasonal cumulative distribution functions (CDFs) of PF lifetimes for MCS and IDC events for 2004 – 2017 in Figure S4. Results show winter has the largest fraction of MCS/IDC events with longer lifetimes than other seasons.

353   As expected, MCSs are much larger than IDC events in spatial coverage and precipitation

354   area, as shown in Table 1 by the comparisons of CCS area, PF major axis length, PF

355   convective/stratiform area, CCF area, and CCF major axis length. Generally, on average, winter

356   MCS/IDC events are the largest in overall spatial coverage (both CCS and PF areas), while

357   summer has the smallest. The larger and longer-lived MCSs in winter than in summer were also

358   observed in the Southeast US in 1994 – 1995 by Geerts (1998). The remarkable seasonal

359   difference in MCS/IDC overall spatial coverage is mainly due to stratiform areas. Convective

360   areas are much smaller than stratiform areas. The PF stratiform area of MCSs in winter is 90,513

361   km^2, 2.4 times larger than the area of 26,599 km^2 in summer, but the PF convective area of

362   MCSs in winter is 7,293 km^2, 14% smaller than 8,465 km^2 in summer. Similarly, the IDC PF

363   stratiform area in winter is 3,182 km^2, 2.8 times larger than 828 km^2 in summer, while the IDC

364   PF convective area in winter is 528 km^2, slightly larger (9%) than 483 km^2 in summer. Unlike

365   stratiform areas with the largest value in winter, convective activity is the most intense in

366   summer as indicated by PF mean convective 20-dBZ echo-top height in Table 1. The most

367   intense convective activity reflects the strongest atmospheric thermal instability due to the

368   strongest solar radiation in summer. We further confirm this point by investigating the MCS/IDC

369   initiation time. As shown in Figure S5, most MCS and IDC events initiate in the afternoon of

370   summer when atmospheric instability is the strongest, consistent with Geerts (1998), who found

371   warm-season MCSs generally initiated at 12:00 – 14:00 Local Time in the Southeast US.

372   Although MCSs are much larger than IDC events in spatial coverage, their mean convective

373   20-dBZ echo-top heights, which can be used to represent their mean convective intensities, are

374   similar in Table 1. And their PF mean convective and stratiform rain rates are also comparable.

375   PF mean convective and stratiform rain rates show significant seasonal variations for both MCS

376   and IDC events. Summer MCS and IDC events have the largest rain rates, followed by autumn.

377   Winter has the lowest rain rates compared to other seasons.

378   The high-resolution nature of the MCS/IDC data product enables a detailed examination of

379   the 3-D evolutions of MCS/IDC events to investigate the relationships between atmospheric

380   environments and MCS/IDC characteristics and to examine the impacts of MCSs and IDC on

381   hydrology, atmospheric chemistry, and severe weather hazards. The data product can also be

382   used to evaluate and improve the representation of MCS/IDC processes in weather and climate

383   models. As an example of the application of the MCS/IDC data product, in Section 3.2, we

384   investigate the contributions of MCS and IDC events to precipitation east of the Rocky

385   Mountains for 2004 – 2017.

386 **Table 1.** Annual and seasonal mean characteristics of MCS and IDC events in the data product domain for 2004 – 2017

	MCS					IDC				
	Annual	spring	Summer	autumn	winter	annual	spring	summer	autumn	winter
CCS-based lifetime / hour	21.1	21.5	19.9	22.1	24.3	2.1	2.1	2.0	2.0	2.7
CCS area[1] / km²	185,436	223,230	130,769	185,246	373,220	6,775	9,400	4,542	6,515	20,902
CCS major axis length / km	693	774	568	726	1,067	99	117	86	100	169
PF-based lifetime[2] / hour	18.9	19.3	17.9	19.7	21.3	1.7	1.7	1.7	1.7	1.9
Major axis length of the largest PF[3] / km	397	426	325	436	620	63	69	56	69	93
PF convective area[4] / km²	8,273	8,589	8,465	7,752	7,293	494	509	483	502	528
PF stratiform area / km²	41,336	47,241	26,559	48,376	90,513	1,261	1,610	828	1,583	3,182
PF mean convective rain rate / mm h⁻¹	4.4	3.9	4.7	4.5	3.8	4.2	3.4	4.5	4.3	3.0
PF mean stratiform rain rate /mm h⁻¹	2.6	2.4	2.8	2.6	2.2	2.8	2.5	3.0	2.9	2.3
PF mean convective 20-dBZ echo-top height / km	6.5	6.2	7.2	6.0	4.9	6.6	6.1	7.0	6.2	5.0
Area of the largest CCF / km²	2,578	2,515	2,983	2,068	1,606	343	359	339	340	349
Major axis length of the largest CCF / km	109	109	117	100	92	29	30	29	29	31

387
388
389
390
391
392
393
394
395
396
397
398

[1] In this table, for hourly characteristics (all variables except for CCS-based lifetime and PF-based lifetime), we generally first calculate the average values of the characteristics during the duration of each MCS/IDC event except for the max 30/40-dBZ echo-top heights, which are the maximum values of the attributes within the period. Then we calculate the mean values of the characteristics of all MCS/IDC events. For example, an MCS has a CCS-based lifetime of 10 hours. During its duration, it has a CCS at each hour. We calculate the average CCS area during the 10 hours, which is the average CCS area of the MCS. Then, we average all MCSs identified during a period to derive the values shown in this row.

[2] Lifetimes of MCS/IDC events determined by PFs. Only count those hours of an MCS/IDC event with a significant PF present (PF major axis length > 20 km for MCSs; ≥ 4 km for IDC).

[3] There can be multiple PFs and CCFs at a given time for an MCS/IDC event. "Largest" means only the largest PF or CCF is used in the calculation.

[4] There can be multiple PFs and CCFs at a given time for an MCS/IDC event. If not specified, all PFs/CCFs are considered. For example, convective areas of all PFs at a given time are summed to represent the PF convective area of an MCS/IDC event at that time. Similarly, the convective rain rates of all PFs at the given time are averaged to represent the PF mean convective rain rate of the MCS/IDC at that time.

399   3.2 Precipitation characteristics from different sources

400   Here we only consider hourly data with precipitation > 1 mm h^{-1} (Feng et al., 2019). At 4

401   km resolution, precipitation less than 1 mm h^{-1} accounts for less than 19% of the total

402   precipitation, and the uncertainty of radar-derived precipitation at such low rainfall intensity is

403   typically large. Including hourly data with precipitation ≤ 1 mm h^{-1} in the calculation will change

404   the values shown in this study but will neither affect the comparison among MCS, IDC, and non-

405   convective (NC) precipitation nor their spatial distribution patterns. Here, NC precipitation refers

406   to precipitation not associated with any MCS or IDC events and is mainly from stratiform rain.

407   Total precipitation is the sum of MCS, IDC, and NC precipitation. It is noteworthy that NC

408   precipitation may contain some convection-associated rain due to the limitation of the source

409   datasets and the algorithms used in this study. More relevant details are discussed in Section

410   3.2.3 and Section 4.

411   *3.2.1 Annual spatial distributions of different types of precipitation*

412   According to the MCS/IDC data product, the annual average total precipitation east of the

413   Rocky Mountains in the US (US grid cells in Figure 1) is 691 mm between 2004 and 2017 with a

414   mean precipitation intensity of 3.6 mm h^{-1}. MCSs contribute the most to the total precipitation

415   with a fraction of 45%, followed by NC (30%) and IDC (25%). And the mean precipitation

416   intensities of MCSs (4.4 mm h^{-1}) and IDC (3.8 mm h^{-1}) are much larger than NC (2.7 mm h^{-1}).

417   Our MCS precipitation fraction (45%) is higher than that (~30%) from Haberlie and Ashley

418   (2019) over the ECONUS due to their different algorithms and stricter criteria to track and define

419   MCSs.

420   Figure 5 displays the spatial distributions of annual mean precipitation amounts and

421   intensities for different precipitation types for 2004 – 2017. We also calculate the distributions of

422   the fractions of different types of precipitation in Figure 6. MCS precipitation strongly affects the

423   whole eastern US (105°W – 70°W, MCS precipitation fractions: 46% ± 12%), especially in the

424   South Central US (MCS precipitation fractions: ~60%). The spatial distribution patterns of MCS

425   annual precipitation amounts and fractions in Figure 5 are similar to those from Haberlie and

426   Ashley (2019), although their MCS precipitation fractions are generally lower than our results.

427   IDC precipitation is concentrated in the SE and NE coastal areas, with peak values in Florida.

428   NC precipitation is substantial in the eastern and southern regions with ample moisture supply

429   and contributes over 35% to the total precipitation across most of the NE region. The coastal area

430   near Louisiana, which is significantly affected by all three types of precipitation, has the most

431   total precipitation with annual amounts of over 1,350 mm. The annual total precipitation

432   amounts in most regions of SE also exceed 1,050 mm due to MCS contributions. While the total

433   precipitation amounts in most regions of Florida are also over 1,050 mm, they are mainly

434   attributed to IDC.

435   The spatial patterns of precipitation intensities are somewhat different from those of

436   precipitation amounts (Figure 5). Generally, the southern regions, especially in the coastal areas,

437   have larger precipitation intensities than the northern areas. The MCS precipitation intensities are

438   the largest in Texas, Louisiana, Oklahoma, and Kansas, significantly shifting west compared to

439   MCS precipitation amounts. Unlike IDC precipitation amounts concentrating in the SE and NE

440   coastal areas, IDC precipitation intensities are the largest over the SGP and SE. IDC precipitation

441   intensities over the NE are much smaller compared to the SGP and SE, similar to NC

442   precipitation intensities. We summarize the annual mean precipitation amounts and intensities of

443   different types of precipitation in the NGP, SGP, SE, and NE in Table S3.

444   The distributions of MCS/IDC precipitation amounts are mainly determined by the

445   distributions of MCS/IDC hours (Figures 5 and 7). Here, the MCS/IDC hour of a grid cell during

446   a period is the number of hours when any MCS/IDC events produce > 1 mm hourly accumulated

447   rainfall in the grid cell. The distributions of MCS/IDC precipitation intensities, although not the

448   main factor, can also affect the distributions of MCS/IDC precipitation amounts. For example,

449   the maximum MCS hours are located around Missouri (Figures 7a), but the maximum MCS

450   precipitation amount is in the coastal area of Louisiana (Figure 5c). The larger MCS precipitation

451   intensities in the southern regions contribute more to the MCS precipitation amount in the

452   southern US. In addition, a large number of IDC events (IDC hours > 60 h yr^{-1}) occur in the NE

453   region along the Appalachian Mountains (Figure 7b), but IDC in that region only contributes to

454   20% – 30% of the total precipitation amount (Figure 6b) due to the low precipitation intensities

455   (Figure 5f).

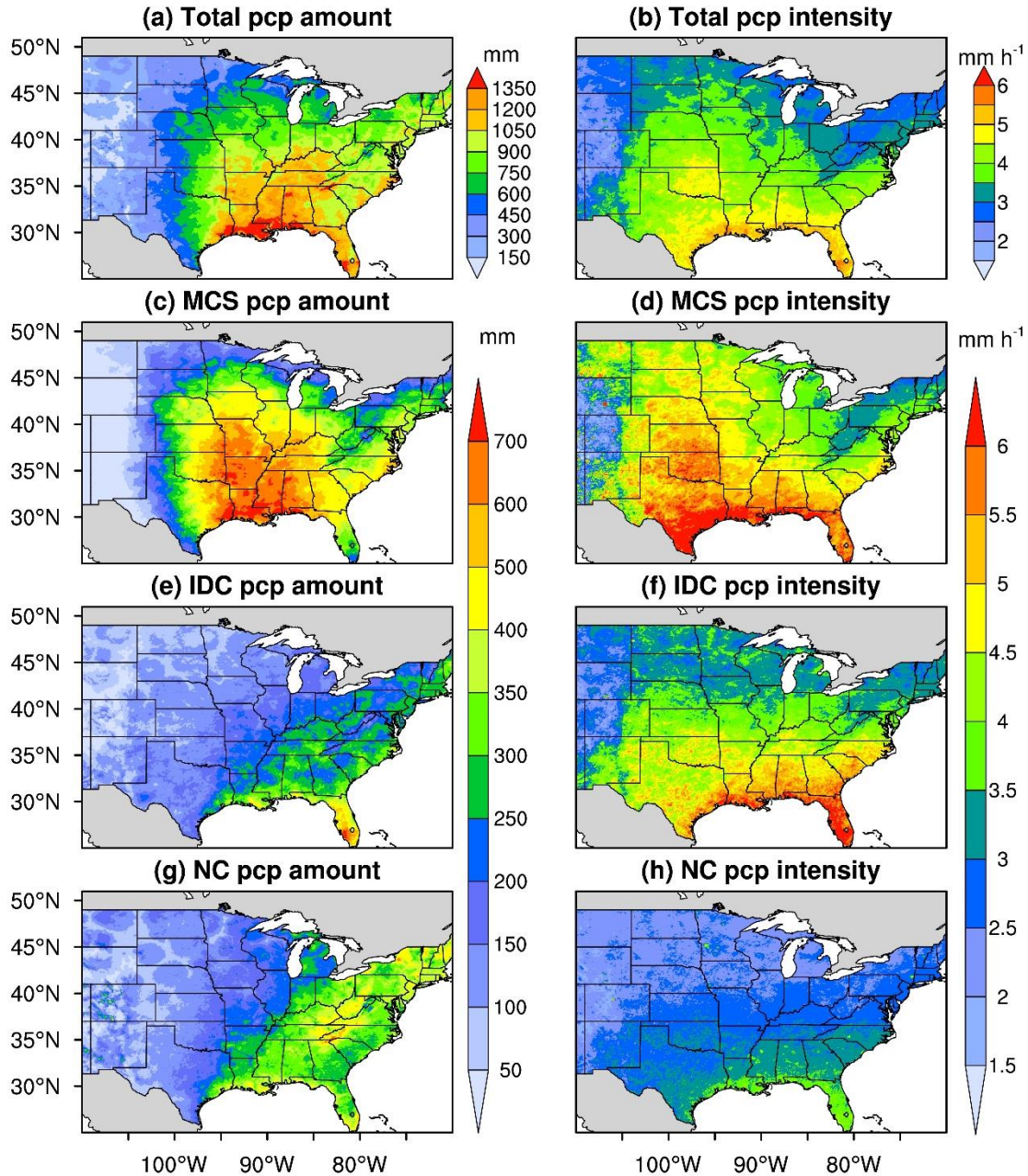



Figure 5. Distributions of annual mean precipitation amounts (a, c, e, g) and intensities (b, d, f, h) for different types of precipitation for 2004 – 2017. (a) and (b) are for total precipitation, (c) and (d) are for MCS precipitation, (e) and (f) are for IDC precipitation, and (g) and (h) are for NC precipitation. We only include hourly data with precipitation > 1 mm h^{-1} in the calculation.

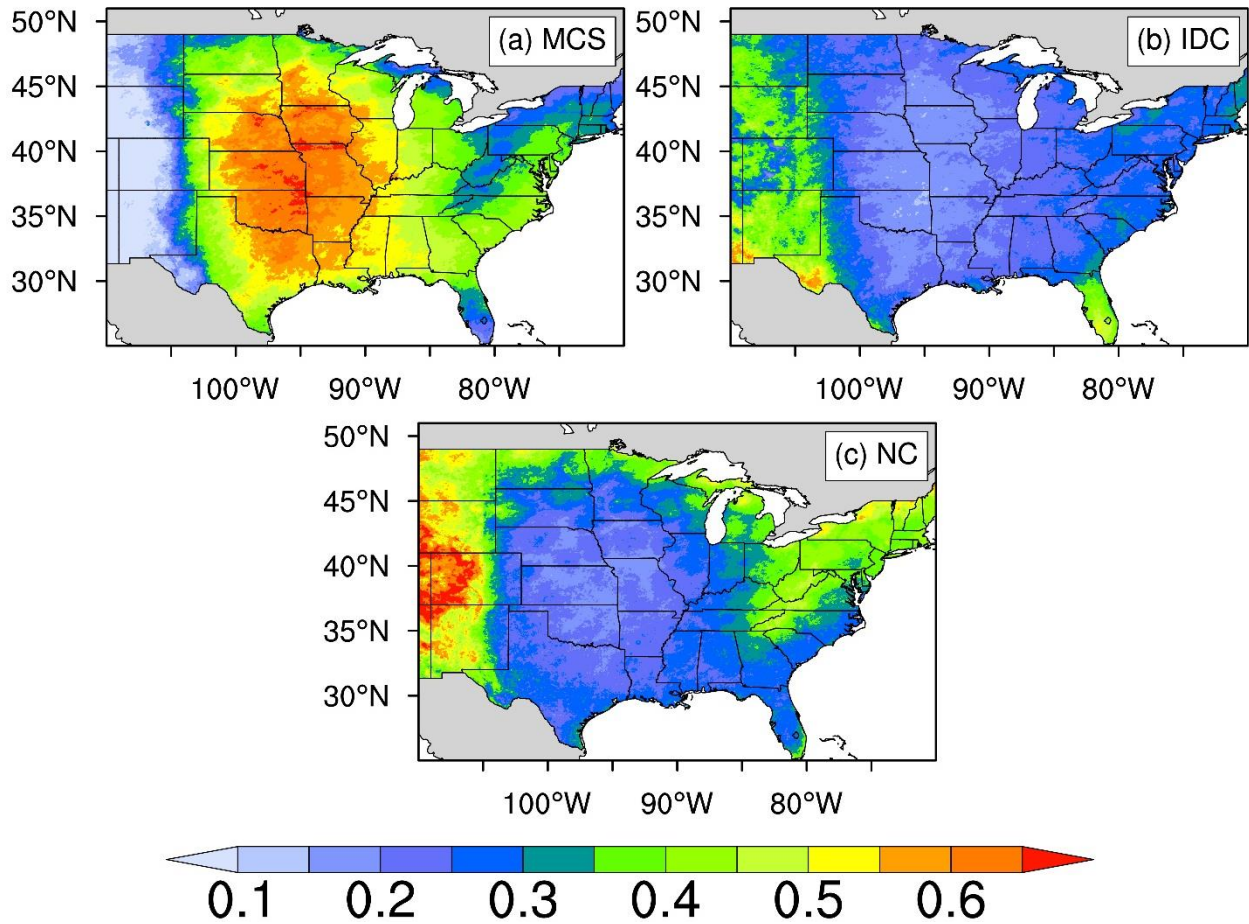
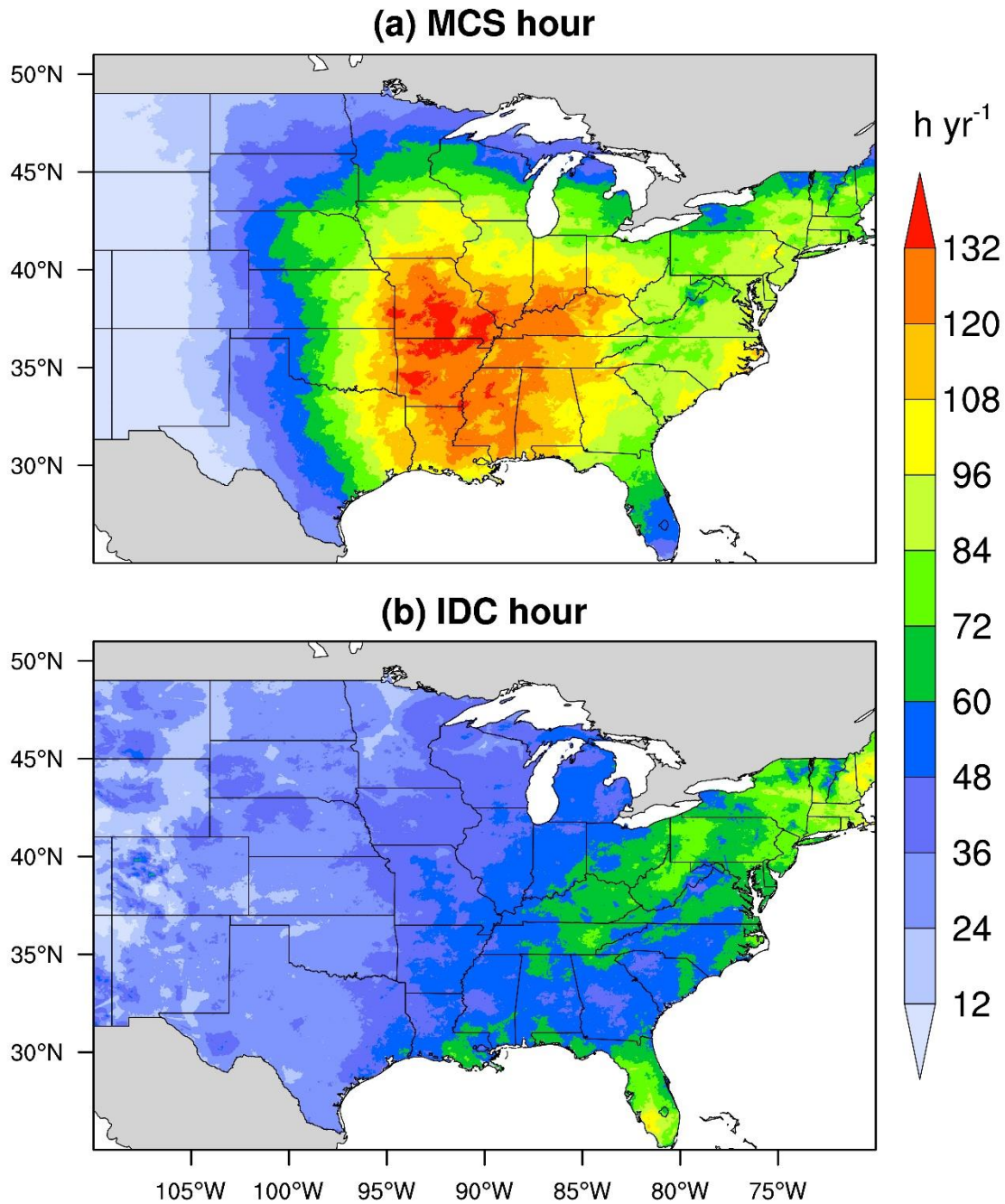



Figure 6. Distributions of the fractions of different types of precipitation (MCS, IDC, NC). Here, precipitation refers to annual mean values for 2004 – 2017. We exclude hourly data with precipitation ≤ 1 mm h^{-1} in the calculation.

465
Figure 7. Spatial distributions of annual mean MCS/IDC hours for 2004 – 2017. (a) is for MCS,
467 and (b) is for IDC. The annual mean MCS/IDC hour of a grid cell is the number of hours per
468 year when any MCS/IDC events produce > 1 mm hourly accumulated rainfall in the grid cell.

469   *3.2.2 Seasonal spatial distributions of different types of precipitation*

470   Figures 8, S6, and S7 display the mean seasonal distributions of precipitation amounts,

471   precipitation fractions, and precipitation intensities for different types of precipitation in 2004 –

472   2017. The MCS precipitation center migrates northwards from Arkansas in spring to northern

473   Missouri and Iowa in summer, followed by a southward migration to Louisiana in autumn, and

474   finally to Mississippi and Alabama in the Southeast (Figures 8e – 8h) in winter. The seasonal

475   shift of the MCS precipitation center agrees with the study of Haberlie and Ashley (2019),

476   showing different MCS precipitation distributions between warm and cold seasons over the

477   ECONUS. Spring and summer have much larger MCS precipitation amounts (~100 mm) than

478   autumn (~62 mm) and winter (~50 mm). The mean MCS precipitation amount in spring is close

479   to that in summer. However, the total number of identified MCSs in summer (212) is much

480   higher than that in spring (122), as discussed in Section 3.1; and the mean MCS precipitation

481   intensity in summer (5.2 mm h^{-1}) is also larger than that in spring (4.1 mm h^{-1}) (Figure S7). The

482   inconsistency is because MCSs in spring occur in more favorable large-scale environments with

483   strong baroclinic forcing and low-level moisture convergence (Feng et al., 2019; Song et al.,

484   2019). As a result, spring MCSs are larger and longer-lasting, and they produce more rainfall per

485   MCS event compared to those in summer (Table 1), compensating for the fewer number of MCS

486   events and lower precipitation intensities in spring. The fractions of MCS precipitation amounts

487   are generally > 35% over the Northern and Southern Great Plains in spring and summer and can

488   reach up to over 70% within the MCS precipitation center (Figures S6a – S6b). The results are

489   roughly consistent with Fritsch et al. (1986), which showed that MCSs accounted for about 30%

490   – 70% of the warm-season (April-September) precipitation over much of the region between the

491   Rocky Mountains and the Mississippi River. The results are also consistent with Haberlie and

492 Ashley (2019) showing MCS precipitation fractions generally > 30% with a peak > 60% over the

493 Great Plains between May and August. Due to the low precipitation amounts of IDC and NC, the

494 fractions of MCS precipitation amounts in autumn and winter are also large, showing over 50%

495 within the MCS precipitation center (Figures S6c – S6d).

496   The IDC precipitation amounts reach a maximum in summer, centered in the coastal areas

497 of the SE, where IDC precipitation contributes to more than 40% of the total precipitation

498 amounts (Figures 8i – 8l and S6e – S6h). Winter has the least IDC precipitation. Areas of high

499 IDC precipitation do not show much seasonal variability, suggesting that IDC is constrained by

500 local conditions such as moisture availability, local solar radiation, and land-atmosphere

501 interactions. The NC precipitation amount also peaks in summer, followed by autumn,

502 particularly in the NE (Figures 8m – 8p). However, because both MCS and IDC precipitation

503 amounts are very high in summer, the fraction of the NC precipitation amount in summer (28%)

504 is smaller than that of winter (32%) (Figures S6i – S6l). Winter NC precipitation center occurs in

505 the SE coastal areas (Figure 8p).

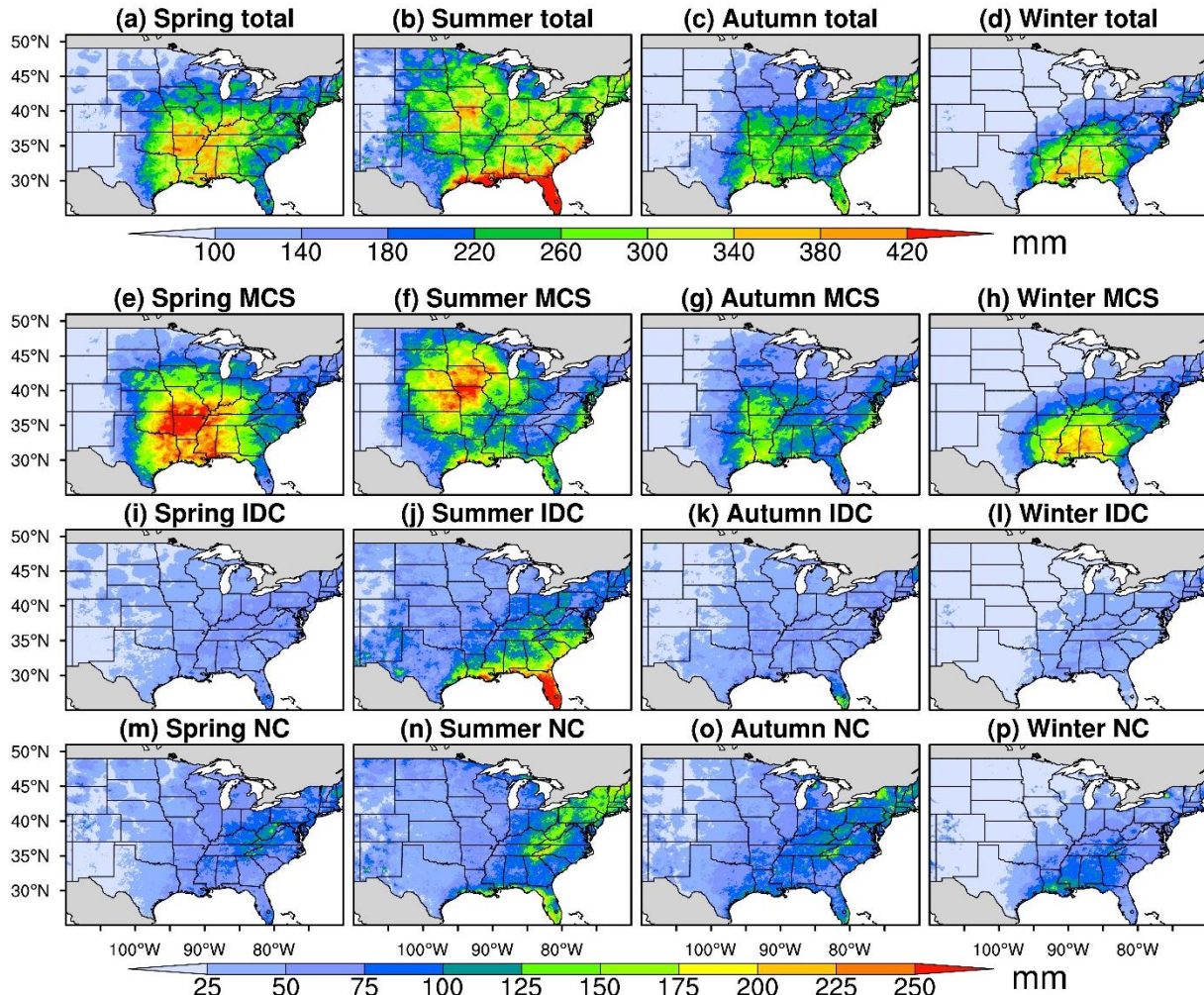



Figure 8. Distributions of annual mean seasonal precipitation amounts for different types of precipitation for 2004 – 2017. The first row is for total precipitation, the second for MCS precipitation, the third row for IDC precipitation, and the fourth row for NC precipitation. The first column shows spring precipitation, the second column for summer, the third column for autumn, and the fourth column for winter. MCS, IDC, and NC precipitation share the same label bar. We exclude hourly data with precipitation ≤ 1 mm h^{-1} in the calculation.

The precipitation intensities of all three types peak in summer and reach minimums in winter (Figure S7). In each season, precipitation intensities in the south are larger than those in the north except for MCS precipitation intensities in summer, which maximize in Oklahoma. We summarize the mean seasonal precipitation amounts and intensities of different types of precipitation over the 4 climate regions of Figure 1 in Table S4.

518   *3.2.3 Diurnal cycles of different types of precipitation*

519   Figure 9 shows the monthly mean diurnal cycles of precipitation amounts from MCSs, IDC,

520   and NC in the NGP, SGP, SE, and NE, respectively. Generally, MCS precipitation peaks during

521   nighttime in the NGP, SGP, and NE. The seasonal shift of the peaks from spring in the SGP to

522   summer in the NGP reflects the northward migration of the MCS precipitation center in the

523   Great Plains (Figures 8e and 8f).

524   The SE has significantly different diurnal cycles of MCS precipitation from other regions.

525   In spring, SE MCS precipitation is mainly located in the western areas (Figure 8e), showing

526   similar diurnal characteristics as the SGP MCS precipitation but with peaks in the early morning

527   and late afternoon (Figures 9d and 9g). Besides, the SGP MCS precipitation peaks in May

528   (Figure 9d), while SE peaks in April (Figure 9g), suggesting that the MCS precipitation center

529   first appears in the western SE regions (Alabama, Mississippi, and Louisiana) in April, and then

530   moves northwards to Arkansas in May. In summer, the SE MCS precipitation diurnal cycles are

531   more like those of IDC (Figures 9g and 9h), peaking in the late afternoon and much different

532   from those in the Great Plains. The significantly different precipitation diurnal variations

533   between the Great Plains and SE were also identified by Haberlie and Ashley (2019). We find

534   that most summer MCS precipitation over the SE occurs near the coastal areas (Figure 8f), far

535   from the MCS precipitation center in northern Missouri and Iowa, suggesting either a different

536   MCS genesis mechanism in the SE from those in the SGP and NGP (Feng et al., 2019) or long-

537   duration deep convective systems showing MCS characteristics (Geerts, 1998). In autumn, the

538   SE MCS precipitation peaks in the morning (Figure 9g). The diurnal cycle of MCS precipitation

539   in September shows mixed features of summer and autumn with peaks both in the morning and

540   the afternoon. In winter months, the diurnal cycle of the SE MCS precipitation shifts from the

541   autumn feature to the spring feature, with peaks shifting from the morning to the afternoon. The

542   distinct diurnal cycles of SE MCS precipitation in different seasons in Figure 9g are roughly

543   consistent with the corresponding seasonal diurnal variations of MCS occurrence frequencies

544   from Geerts (1998), where the occurrence time of an MCS was defined as the central time

545   between the initiation and decay of the MCS.

546   The diurnal cycles of IDC precipitation are consistent in all regions (Figures 9b, 9e, 9h, and

547   9k), peaking in the late afternoon in summer (Tian et al., 2005), again reflecting the impact of

548   local instability driven by the solar forcing on IDC development. NC precipitation (Figures 9c,

549   9f, 9i, and 9l) shows some diurnal cycle characteristics similar to IDC precipitation. It may be

550   caused by the limitation of the temporal resolution of the datasets used in the FLEXTRKR

551   algorithm. Weak IDC events that are shorter than 1 hour could be missed by Gridrad in

552   identifying CCFs, as Gridrad Z_H only considers reflectivities within $\pm$ 3.8 minutes of the analysis

553   time. These weak IDC could be aliased to NC precipitation, therefore showing some similar

554   diurnal cycles as IDC. Another possible reason is that the FLEXTRKR algorithm may miss some

555   parts of IDC clouds with $T_b \geq 241$ K, which are then classified as NC, so the NC precipitation

556   exhibits some IDC characteristics.

557   The monthly diurnal cycles of precipitation intensities for MCSs, IDC, and NC are

558   generally similar among all regions, peaking in the late afternoon and early morning in the warm

559   season (Figure S8).

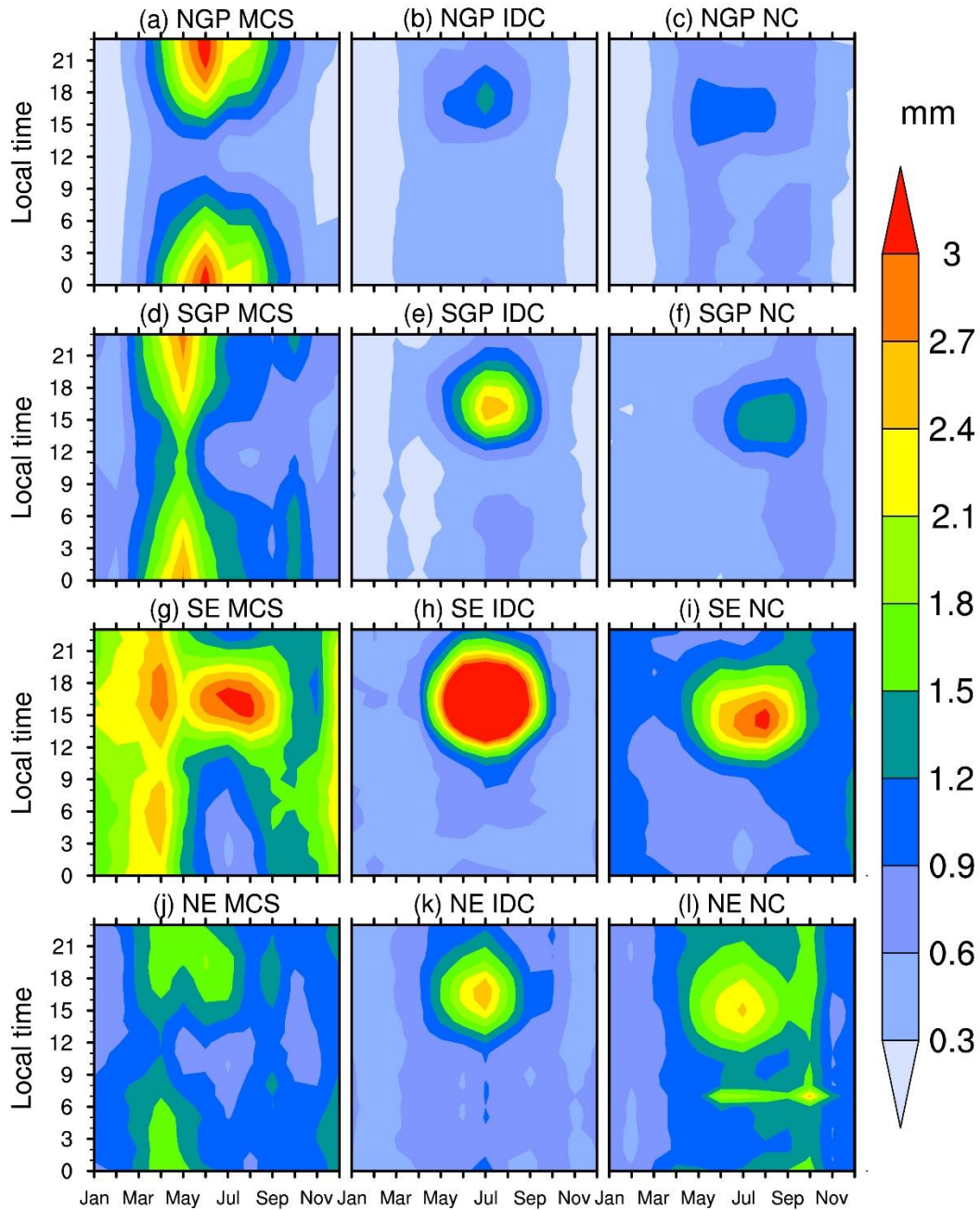



Figure 9. Monthly mean diurnal cycles of precipitation amounts from MCSs (a, d, g, j), IDC (b, e, h, k), and NC (c, f, i, l) in the NGP (a, b, c), SGP (d, e, f), SE (g, h, i), and NE (j, k, l) during 2004 – 2017.

4 Uncertainties of the data product

4.1 Uncertainties from source datasets

The NCEP/CPP L3 4 km Global Merged IR V1 T_b dataset has been view-angle corrected and re-navigated for parallax (Janowiak et al., 2001) to reduce errors. However, the US continent is covered by two series of geostationary IR satellites (GOES-W and GEOS-E). During the production of the T_b dataset, the value with the smaller zenith angle is adopted when duplicate data are available in a grid pixel. Measurements from different satellites may be inconsistent. Janowiak et al. (2001) suggest this type of inconsistency to be considered minor.

For the Gridrad radar dataset, some bad volumes have been removed during the production of Gridrad Z_H. We further filter out potential low-quality observations, scanning artifacts, and non-meteorological echoes from biological scatters and artifacts following the approaches of Homeyer and Bowman (2017). However, there is another source of error from anomalous propagation caused by non-standard refractions of radar signals in the lower atmosphere, which cannot be mitigated during the filtering procedure. Non-standard refractions can result in underestimation or overestimation of the true radar beam altitude, thus affecting the location of radar reflectivity for binning. Estimating the corresponding uncertainties is out of the scope of this study. However, anomalous propagation is typically limited to radar beams traveling long distances in the boundary layer (Homeyer and Bowman, 2017).

Stage IV precipitation is a mosaic of precipitation estimates based on a combination of NEXRAD and gauge data from 12 RFCs. Therefore, the errors of Stage IV are from several sources, such as inherent NEXRAD biases, radar quantitative precipitation estimate (QPE)

585   algorithm biases, bad gauge data removal inconsistency among different RFCs, multisensory

586   processing algorithm inconsistency among different RFCs, and mosaicking border

587   discontinuities (Nelson et al., 2016). The most severe errors occur in the western US, where

588   NEXRAD data are limited, and a gauge-only rainfall estimation algorithm is used (Nelson et al.,

589   2016; Smalley et al., 2014). Hence our data product has a geographical focus east of the Rocky

590   Mountains, with the best NEXRAD coverage in the US. After regridding the Stage IV

591   precipitation into our 4-km domain, we further manually filter out certain "erroneous

592   precipitation" hours and set all precipitation in those hours to missing values. "Erroneous

593   precipitation" is defined as sudden appearance and disappearance of a large contiguous area (>

594   4,800 km^2) with intense precipitation (> 40 mm h^{-1}) (Figure S9), which is physically not

595   possible. There are 40 hours in total in the period 2004 – 2017 containing such "erroneous

596   precipitation."

597   As the FLEXTRKR algorithm is applied to a combination of three independent types of

598   remote sensing datasets, we identify the most robust MCS/IDC events satisfying all the criteria

599   based on the three datasets. It reduces the potential false classification of tracks as MCSs or IDC

600   based on any single dataset. And to consider the potential error of ERA5 melting level heights,

601   we require $Z_H \geq 45$ dBZ above ($Z_{melt} + 1$) km for convective classification in the SL3D algorithm

602   (Table S2).

603   4.2 The impact of missing data

604   In the CCS identification step of the FLEXTRKR algorithm, we require the fraction of

605   missing satellite T_b in the domain at each hour to be less than 20%. Otherwise, the hour is

606   excluded from our data product. During 2004 – 2017, we excluded 716 hours with missing

607   satellite T_b data, accounting for less than 0.6% of the total period. The year with the most

608   missing satellite data is 2008, with 206 missing hours (2.3%), followed by 2004 with 154 hours

609   (1.8%). All other years have no more than 57 missing hours. During the link procedure of the

610   FLEXTRKR algorithm, we search the next hour if a missing hour is encountered, as long as the

611   time gap between the two "linked" hours is less than 4 hours. Otherwise, we start new tracks

612   from the next available hour. This method aims to reduce the impact of the missing hours.

613   Considering the high completeness of the satellite T_b data in 2004 – 2017, we conclude that the

614   missing satellite data have little effect on the data product.

615   We show the distribution of the fractions of valid Stage IV precipitation data in 2004 – 2017

616   in Figure S10. The fractions are over 97% for all grid cells of the US in the domain. Most grid

617   cells in the US have less than 2% missing hours, which should have a negligible impact on the

618   data product.

619   Figure S11 shows the fractions of available Gridrad reflectivity data from 2004 to 2017

620   between 1 km and 12 km ASL. The fractions are relatively high over the majority of the

621   troposphere except for 1 km ASL. Based on the criteria of the SL3D algorithm, Z_H at 1 km is

622   rarely used and can be easily substituted by Z_H at 2 km. Generally, Gridrad has good spatial

623   coverage during the period with most grid cells east of the Rocky Mountains having fractions >

624   90% between 2 and 9 km and 80% between 10 and 12 km. The completeness of the Gridrad

625   dataset is relatively lower compared to the satellite T_b and Stage IV precipitation datasets, and

626   Gridrad Z_H is a crucial variable in the SL3D classification and MCS/IDC identification.

627   Therefore, the missing data of Gridrad Z_H should have some impacts on our data product.

628   However, as an advanced long-term high-resolution 3-D radar reflectivity dataset, Gridrad is

629   valuable for constructing a climatological MCS/IDC data product.

630   4.3 Temporal resolution limitation of the source datasets

631   As we discussed in Section 3.2.3, the diurnal cycles of NC precipitation show some possible

632   aliasing from IDC precipitation. Some weak IDC events are so short that the hourly data cannot

633   properly capture their occurrence, especially for Gridrad Z_H, which only includes reflectivities

634   within ± 3.8 minutes of each hour. We calculate the cumulative distribution functions of PF-

635   based lifetimes for MCS and IDC events and their associated precipitation in the data product for

636   2004 – 2017, as shown in Figure 10. About 75% of IDC events have a PF-based lifetime of 1

637   hour. Therefore, it is almost certain that we miss some IDC events shorter than 1 hour in the data

638   product. Here we give an estimate of the probability p that a given IDC event with a convective

639   signal duration of x minutes is detected by radar, as expressed below:

640    $$p = \frac{2 \times 3.8}{60 - x}$$ (1)

641   where the numerator is the time window of Gridrad observation in each hour, and x is the

642   duration of the IDC event. The detection probability is only about 25% when $x = 30$ minutes. To

643   obtain a detection probability of 50%, we require $x \geq 45$ minutes. Hence, we cannot assess the

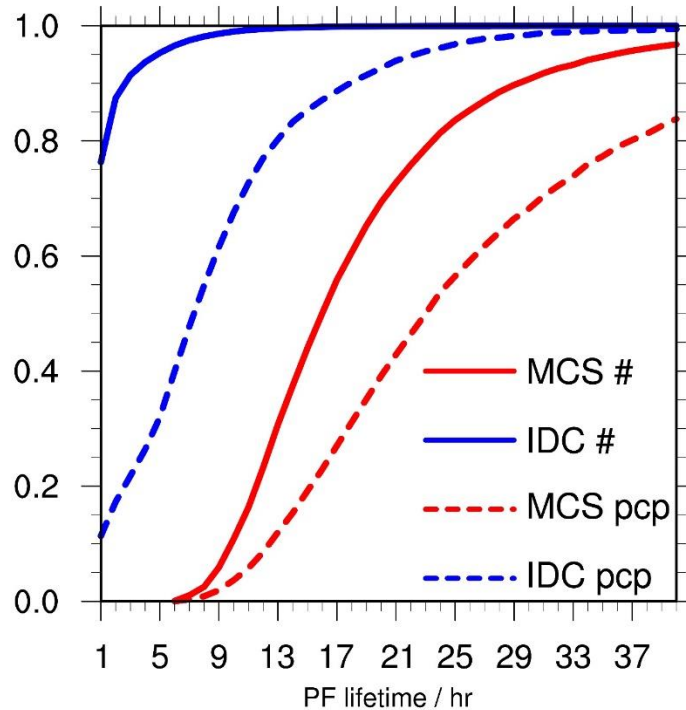
644   distribution of IDC convective signals with durations less than 1 hour using the currently

645   available datasets. Higher-resolution datasets, such as individual NEXRAD radar data, which

646   typically has an update cycle of 4-5 min, are necessary to derive the information. However, as

647   shown in Figure 10, we find that precipitation from IDC events with a 1-hour PF lifetime only

648 accounts for about 10% of the total IDC precipitation. Therefore, IDC events with PF lifetimes

649 less than 1 hour should have a relatively small impact on precipitation.

650
651 Figure 10. Cumulative distribution functions of PF-based lifetimes for MCS and IDC events and
652 their associated precipitation in the data product domain for 2004 – 2017. The red solid line is for
653 the number of MCSs, the red dash line for MCS associated precipitation, the blue solid line for
654 the number of IDC events, and the blue dash line for IDC associated precipitation.

655 4.4 The impact of MCS and IDC definition criteria

656   The separation between MCSs and long-lasting IDC events is somewhat fuzzy (Feng et al.,

657 2019; Geerts et al., 2017; Haberlie and Ashley, 2019; Pinto et al., 2015; Prein et al., 2017). Here,

658 we briefly examine the impact of different MCS/IDC definition criteria on the data product. We

659 change the definition of MCSs to relax the CCS and PF size and duration thresholds.

660 Specifically, the second and third criteria listed in Section 2.2.2 are modified as follows: 2) CCS

661 areas associated with the track surpass 40,000 km^2 for more than 4 continuous hours; 3) PF

662 major axis length exceeding 80 km and intense convective cell areas $\geq$ 16 km^2 exist for more

663   than 3 consecutive hours. And we also require that each merge/split-track associated with

664   MCS/IDC events must have a CCS-based lifetime of no more than 3 hours. We keep the

665   definition of IDC the same as described in Section 2.2.2, which is a limit for IDC that we can

666   identify based on the source datasets.

667   By using the new definition, as expected, the lifetimes and spatial coverages of MCSs are

668   reduced, and those of IDC change little because most IDC events cannot satisfy the new MCS

669   criteria (Tables 1 and S5). The annual number of MCSs identified in 2004 − 2017 increases from

670   454 to 857. The number increases from 122 to 207 in spring, 212 to 434 in summer, 83 to 151 in

671   autumn, and 37 to 62 in winter. As PF-based lifetimes of MCS/IDC events in summer are the

672   shortest (Table 1), the new definition has the most significant impact in summer. The annual

673   number of IDC decreases from 45,346 to 45,225. Reducing the merge/split lifetime limit retains

674   more independent IDC events, which is the reason why the decrease in the number of IDC events

675   is smaller than the increase in the number of MCSs. Annual mean MCS precipitation east of the

676   Rocky Mountains increases from 313 mm to 353 mm, while IDC precipitation decreases from

677   170 mm to 130 mm. The fraction of MCS precipitation only increases by 6% (from 45% to

678   51%), compared to the almost doubling of MCS number (from 454 to 857), suggesting the MCS

679   definition in the original data product is capable of capturing most of the important MCSs with

680   heavy precipitation. Similar to MCS numbers, summer has the most increase in MCS

681   precipitation amount, from 100 mm to 119 mm. And annual mean MCS and IDC precipitation

682   intensities decrease slightly as MCS precipitation intensities are somewhat larger than IDC in

683   most regions (Tables S3, S4, S6, and S7). We summarize the regional precipitation statistics of

684   the NGP, SGP, SE, and NE based on the new definition in Tables S6 and S7.

685   Although the new definition changes the absolute values of MCS/IDC characteristics, the

686   contrast between MCS and IDC events is still present. The new definition has small impacts on

687   the spatial distribution patterns of MCS/IDC precipitation. And NC precipitation characteristics

688   are almost the same as before. Therefore, our original definition captures the essential

689   characteristics of MCS and IDC events. In addition, the original data product is complete and

690   flexible. We store all criteria variables of MCS/IDC events in the data product. Users can easily

691   change the definition of MCSs and switch between tracks that are attributed to MCS and IDC

692   without re-running the FLEXTRKR algorithm. There is no need to change the "track" and

693   "merge" lifetime criterion as we do above because they have little impact on the climatological

694   characteristics of MCS and IDC events.

695   4.5 Recommendations for the usage of the MCS/IDC data product

696   Considering the limitations and uncertainties mentioned above, we generally recommend

697   using the data product for observational analyses and model evaluations of convection statistics

698   and characteristics over relatively long periods such as a month, a season, or longer to fully take

699   advantage of the long term dataset, although analysis of individual weather events is also

700   possible as supported by the hourly temporal resolution of the data product. In addition, since the

701   completeness and quality of the source radar dataset degrade dramatically beyond the US border

702   and over the Rocky Mountains (Figure S11), we recommend the usage of the data product within

703   the CONUS east of the Rocky Mountains to alleviate the impact of the termination of MCS/IDC

704   tracks due to poor radar coverage and missing radar data beyond their maximum scan range.

705   Detailed investigation of a short period or a specific MCS/IDC event is acceptable, but

706   cautions should be taken when encountering missing data around the track during the period.

707   Due to the complexity of the algorithms used to develop the data product, it is difficult to

708   quantify the impact of missing data on the MCS/IDC track. Therefore, we do not recommend

709   examining a specific MCS/IDC track if there are too many missing data (precipitation, T_b, or Z_H)

710   along the track. Users planning to apply the data product for a specific case study should

711   examine the availability of the source data first, which are also stored in the data product except

712   for 3-D Z_H due to the large data volume. Users can access the original 3-D Z_H at

713   https://rda.ucar.edu/datasets/ds841.0/ (Table S1).

714   Lastly, although our sensitivity test in Section 4.4 shows that precipitation characteristics

715   are similar between two different sets of MCS/IDC definition criteria, we still recommend users

716   conduct further sensitivity tests and examine the impact of different definition criteria on the

717   results if the data product is applied to other studies, such as the effects of MCS and IDC events

718   on atmospheric circulation, environmental conditions associated with the initiation and evolution

719   of MCS and IDC events, and MCS/IDC associated weather hazards.

720 5 Data availability

721   The high-resolution (4 km hourly) MCS/IDC data product and the corresponding user guide

722   document are available at http://dx.doi.org/10.25584/1632005 (Li et al., 2020). The original

723   format of the data files is NetCDF-4, and we archive them as compressed files for each year so

724   that the data product is easily accessible. The user guide contains a brief explanation about the

725   approach to develop the data product and a detailed description of the data file content to help

726   users understand the data product.

6 Conclusions

Here we present a unified high-resolution (4 km, hourly) data product that describes the spatiotemporal characteristics of MCS and IDC events from 2004 to 2017 east of the Rocky Mountains over the CONUS. We produce the data product by applying an updated FLEXTRKR algorithm to the NCEP/CPP L3 4 km Global Merged IR V1 T_b dataset, ERA5 melting level heights, the 3-D Gridrad radar reflectivity dataset, and the Stage IV precipitation dataset. Climatological features of the MCS and IDC events from the data product are compared, with a focus on their precipitation characteristics. Consistent with our definitions of MCSs and IDC in the FLEXTRKR algorithm, we find that MCSs have much broader spatial coverage and longer duration than IDC events. While there are many more frequent IDC occurrences than MCSs, the mean convective intensities of IDC events are comparable to those of MCSs. MCS and IDC events both contribute significantly to precipitation east of the Rocky Mountains but with distinct spatiotemporal variabilities. MCS precipitation affects most regions of the eastern US in all seasons, especially in spring and summer. The MCS precipitation center migrates northwards from Arkansas in spring to northern Missouri and Iowa in summer, followed by a southward migration to Louisiana in autumn, and finally to Mississippi and Alabama in the Southeast in winter. IDC precipitation mostly concentrates in the Southeast in summer. IDC precipitation shows a significant diurnal cycle in summer months with a peak around 16:00 – 17:00 Local Time over all regions east of the Rocky Mountains. In contrast, MCS precipitation peaks during nighttime in spring and summer for most regions except for the Southeast, where MCS precipitation peaks in the late afternoon in summer, similar to IDC precipitation. Lastly, we analyze the potential uncertainties of the data product and the sensitivity of the dataset to MCS definitions and give our recommendations for the usage of the data product. The data product

750   will be useful for investigating the atmospheric environments and physical processes associated

751   with convective systems, quantifying the impacts of convection on hydrology, atmospheric

752   chemistry, severe weather hazards, and other aspects of the energy, water, and biogeochemical

753   cycles, and improving the representation of convective processes in weather and climate models.

754   **Author contributions**

755   JL and ZF updated the FLEXTRKR algorithm and prepared the source datasets. JL ran the SL3D

756   and updated FLEXTRKR algorithms for 2004 – 2017. JL collected and archived the MCS/IDC

757   data product and did the analyses. JL led the writing of the manuscript with input from ZF, YQ,

758   and LRL. YQ and LRL guided the development of the data product. JL, ZF, YQ, and LRL

759   reviewed the manuscript.

760   **Competing interests**

761   The authors declare that they have no conflict of interest.

762   **Acknowledgments**

763   This research was supported by the US Department of Energy Office of Science Biological and

764   Environmental Research as part of the Regional and Global Modeling and Analysis (RGMA)

765   program area through the Integrated Coastal Modeling (ICoM) project. LRL and ZF were also

766   partly supported by the Water Cycle and Climate Extremes Modeling (WACCEM) Scientific

767   Focus Area funded by RGMA. The research used computational resources from the National

768   Energy Research Scientific Computing Center (NERSC), a DOE User Facility supported by the

769   Office of Science under Contract DE-AC02-05CH11231. PNNL is operated for DOE by Battelle

770 Memorial Institute under Contract DE-AC05-76RL01830. We thank Dr. Cameron R. Homeyer

771 from the University of Oklahoma for helping us understand the Gridrad dataset and Dr. Jingyu

772 Wang from PNNL for identifying the existence of erroneous Stage IV precipitation. We obtain

773 the NCEP/CPP L3 half-hourly 4 km Global Merged IR V1 brightness temperature dataset from

774 https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_1/summary (last access: Dec 28, 2019). The

775 3D Gridrad dataset is from https://rda.ucar.edu/datasets/ds841.0/ (last access: Jan 2, 2020). We

776 download hourly Stage IV precipitation data from https://rda.ucar.edu/datasets/ds507.5/ (last

777 access: Dec 28, 2019), and the ERA5 melting level height data was downloaded from

778 https://doi.org/10.24381/cds.adbb2d47 (last access: Jan 24, 2020).

779

References

Anderson, J. G., Weisenstein, D. K., Bowman, K. P., Homeyer, C. R., Smith, J. B., Wilmouth, D. M., Sayres, D. S., Klobas, J. E., Leroy, S. S., and Dykema, J. A.: Stratospheric ozone over the United States in summer linked to observations of convection and temperature via chlorine and bromine catalysis, Proc. Natl. Acad. Sci. U.S.A., 114, E4905-E4913, https://doi.org/10.1073/pnas.1619318114, 2017.

Andreae, M. O., Artaxo, P., Fischer, H., Freitas, S., Grégoire, J. M., Hansel, A., Hoor, P., Kormann, R., Krejci, R., and Lange, L.: Transport of biomass burning smoke to the upper troposphere by deep convection in the equatorial region, Geophys. Res. Lett., 28, 951-954, https://doi.org/10.1029/2000GL012391, 2001.

Angel, J. R., Palecki, M. A., and Hollinger, S. E.: Storm precipitation in the United States. Part II: Soil erosion characteristics, Journal of Applied Meteorology, 44, 947-959, https://doi.org/10.1175/JAM2242.1, 2005.

Bigelbach, B., Mullendore, G., and Starzec, M.: Differences in deep convective transport characteristics between quasi‑isolated strong convection and mesoscale convective systems using seasonal WRF simulations, J. Geophys. Res.-Atmos., 119, 11,445-411,455, https://doi.org/10.1002/2014JD021875, 2014.

Brooks, H. E., Doswell III, C. A., and Kay, M. P.: Climatological estimates of local daily tornado probability for the United States, Weather and Forecasting, 18, 626-640, https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2, 2003.

Carpenter, S. R., Booth, E. G., and Kucharik, C. J.: Extreme precipitation and phosphorus loads from two agricultural watersheds, Limnol. Oceanogr., 63, 1221-1233, https://doi.org/10.1002/lno.10767, 2018.

Changnon, S. A.: Damaging thunderstorm activity in the United States, Bulletin of the American Meteorological Society, 82, 597-608, https://doi.org/10.1175/1520-0477(2001)082<0597:DTAITU>2.3.CO;2, 2001a.

Changnon, S. A.: Thunderstorm rainfall in the conterminous United States, Bulletin of the American Meteorological Society, 82, 1925-1940, https://doi.org/10.1175/1520-0477(2001)082<1925:TRITCU>2.3.CO;2, 2001b.

Choi, S., Joiner, J., Choi, Y., Duncan, B., Vasilkov, A., Krotkov, N., and Bucsela, E.: First estimates of global free-tropospheric NO_2 abundances derived using a cloud-slicing technique applied to satellite observations from the Aura Ozone Monitoring Instrument (OMI), Atmos. Chem. Phys, 14, 10,565-510,588, https://doi.org/10.5194/acp-14-10565-2014, 2014.

Cintineo, J. L., Pavolonis, M. J., Sieglaff, J. M., and Heidinger, A. K.: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties, Journal of applied meteorology and climatology, 52, 2009-2023, https://doi.org/10.1175/JAMC-D-12-0330.1, 2013.

Davis, C., Brown, B., and Bullock, R.: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas, Monthly Weather Review, 134, 1772-1784, https://doi.org/10.1175/MWR3145.1, 2006.

817   Davison, M.: Shallow/Deep Convection:
818   https://www.wpc.ncep.noaa.gov/international/training/deep/index.htm, access: April 9, 2020, 1999.

819   Derbile, E. K., and Kasei, R. A.: Vulnerability of crop production to heavy precipitation in north-eastern
820   Ghana, International Journal of Climate Change Strategies and Management,
821   https://doi.org/10.1108/17568691211200209, 2012.

822   Diffenbaugh, N. S., Scherer, M., and Trapp, R. J.: Robust increases in severe thunderstorm environments
823   in response to greenhouse forcing, Proc. Natl. Acad. Sci. U.S.A., 110, 16361-16366,
824   https://doi.org/10.1073/pnas.1307758110, 2013.

825   Doswell III, C. A., Brooks, H. E., and Maddox, R. A.: Flash flood forecasting: An ingredients-based
826   methodology, Weather and Forecasting, 11, 560-581, https://doi.org/10.1175/1520-
827   0434(1996)011<0560:FFFAIB>2.0.CO;2, 1996.

828   ECMWF: ERA5 hourly data on single levels from 1979 to present, available at
829   https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview,
830   https://doi.org/10.24381/cds.adbb2d47, 2018 (last access: Jan 24, 2020).

831   Feng, Z., Dong, X., Xi, B., Schumacher, C., Minnis, P., and Khaiyer, M.: Top‑of‑atmosphere radiation
832   budget of convective core/stratiform rain and anvil clouds from deep convective systems, J. Geophys.
833   Res.-Atmos., 116, https://doi.org/10.1029/2011JD016451, 2011.

834   Feng, Z., Dong, X., Xi, B., McFarlane, S. A., Kennedy, A., Lin, B., and Minnis, P.: Life cycle of
835   midlatitude deep convective systems in a Lagrangian framework, J. Geophys. Res.-Atmos., 117,
836   https://doi.org/10.1029/2012JD018362, 2012.

837   Feng, Z., Leung, L. R., Houze Jr, R. A., Hagos, S., Hardin, J., Yang, Q., Han, B., and Fan, J.: Structure
838   and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection‑
839   permitting simulations over the United States, Journal of Advances in Modeling Earth Systems, 10, 1470-
840   1494, https://doi.org/10.1029/2018MS001305, 2018.

841   Feng, Z., Houze Jr, R. A., Leung, L. R., Song, F., Hardin, J. C., Wang, J., Gustafson Jr, W. I., and
842   Homeyer, C. R.: Spatiotemporal characteristics and large-scale environments of mesoscale convective
843   systems east of the Rocky Mountains, J. Clim., 32, 7303-7328, https://doi.org/10.1175/JCLI-D-19-
844   0137.1, 2019.

845   Folger, P., and Reed, A.: Severe thunderstorms and tornadoes in the United States, Congressional
846   Research Service, 2013.

847   French, A. J., and Parker, M. D.: The initiation and evolution of multiple modes of convection within a
848   meso-alpha-scale region, Weather and forecasting, 23, 1221-1252,
849   https://doi.org/10.1175/2008WAF2222136.1, 2008.

850   Fritsch, J. M., Kane, R. J., and Chelius, C. R.: The contribution of mesoscale convective weather systems
851   to the warm-season precipitation in the United States, Journal of Applied Meteorology and Climatology,
852   25, 1333-1345, https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2, 1986.

853   Futyan, J. M., and Del Genio, A. D.: Deep convective system evolution over Africa and the tropical
854   Atlantic, J. Clim., 20, 5041-5060, https://doi.org/10.1175/JCLI4297.1, 2007.

855   Geerts, B.: Mesoscale convective systems in the southeast United States during 1994–95: A survey,
856   Weather and Forecasting, 13, 860-869, https://doi.org/10.1175/1520-
857   0434(1998)013<0860:MCSITS>2.0.CO;2, 1998.

858   Geerts, B., Parsons, D., Ziegler, C. L., Weckwerth, T. M., Biggerstaff, M. I., Clark, R. D., Coniglio, M.
859   C., Demoz, B. B., Ferrare, R. A., and Gallus Jr, W. A.: The 2015 plains elevated convection at night field
860   project, Bulletin of the American Meteorological Society, 98, 767-786, https://doi.org/10.1175/BAMS-D-
861   15-00257.1, 2017.

862   Giangrande, S. E., Krause, J. M., and Ryzhkov, A. V.: Automatic designation of the melting layer with a
863   polarimetric prototype of the WSR-88D radar, Journal of Applied Meteorology and Climatology, 47,
864   1354-1364, https://doi.org/10.1175/2007JAMC1634.1, 2008.

865   Gourley, J. J., Hong, Y., Flamig, Z. L., Wang, J., Vergara, H., and Anagnostou, E. N.: Hydrologic
866   evaluation of rainfall estimates from radar, satellite, gauge, and combinations on Ft. Cobb basin,
867   Oklahoma, Journal of Hydrometeorology, 12, 973-988, https://doi.org/10.1175/2011JHM1287.1, 2011.

868   Grewe, V.: Impact of climate variability on tropospheric ozone, Sci. Total Environ., 374, 167-181,
869   https://doi.org/10.1016/j.scitotenv.2007.01.032, 2007.

870   Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B., and Lawrimore, J. H.:
871   Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from
872   in situ observations, Journal of hydrometeorology, 5, 64-85, https://doi.org/10.1175/1525-
873   7541(2004)005<0064:CCOTHC>2.0.CO;2, 2004.

874   Haberlie, A. M., and Ashley, W. S.: A radar-based climatology of mesoscale convective systems in the
875   United States, J. Clim., 32, 1591-1606, https://doi.org/10.1175/JCLI-D-18-0559.1, 2019.

876   Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J. M., Nicolas, J., Radu, R., Schepers, D.,
877   Simmons, A., and Soci, C.: Global reanalysis: goodbye ERA-Interim, hello ERA5, in: ECMWF Newsl,
878   159, 17-24, https://doi.org/10.21957/vf291hehd7, 2019.

879   Hodges, K. I., and Thorncroft, C.: Distribution and statistics of African mesoscale convective weather
880   systems based on the ISCCP Meteosat imagery, Monthly Weather Review, 125, 2821-2837,
881   https://doi.org/10.1175/1520-0493(1997)125<2821:DASOAM>2.0.CO;2, 1997.

882   Homeyer, C. R., and Bowman, K. P.: Algorithm Description Document for Version 3.1 of the Three-
883   Dimensional Gridded NEXRAD WSR-88D Radar (GridRad) Dataset, available at
884   http://gridrad.org/pdf/GridRad-v3.1-Algorithm-Description.pdf, 23, 2017.

885   Houze Jr, R. A.: Mesoscale convective systems, Rev. Geophys., 42,
886   https://doi.org/10.1029/2004RG000150, 2004.

887   Houze Jr, R. A., Wang, J., Fan, J., Brodzik, S., and Feng, Z.: Extreme convective storms over high‐
888   latitude continental areas where maximum warming is occurring, Geophys. Res. Lett., 46, 4059-4065,
889   https://doi.org/10.1029/2019GL082414, 2019.

Hu, H., Leung, L. R., and Feng, Z.: Observed Warm‑Season Characteristics of MCS and Non‑MCS Rainfall and Their Recent Changes in the Central United States, Geophys. Res. Lett., 47, e2019GL086783, https://doi.org/10.1029/2019GL086783, 2020.

Janowiak, J., Joyce, B., and Xie, P.: NCEP/CPC L3 half hourly 4 km global (60 S–60 N) merged IR V1, available at https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_1/summary, https://doi.org/10.5067/P4HZB9N27EKU, 2017 (last access: Dec 28, 2019).

Janowiak, J. E., Joyce, R. J., and Yarosh, Y.: A real-time global half-hourly pixel-resolution infrared dataset and its applications, Bulletin of the American Meteorological Society, 82, 205-218, https://doi.org/10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2, 2001.

Kalinga, O. A., and Gan, T. Y.: Estimation of rainfall from infrared‑microwave satellite data for basin‑scale hydrologic modelling, Hydrological processes, 24, 2068-2086, https://doi.org/10.1002/hyp.7626, 2010.

Koehler, T. L.: Cloud-to-Ground Lightning Flash Density and Thunderstorm Day Distributions over the Contiguous United States Derived from NLDN Measurements: 1993–2018, Monthly Weather Review, 148, 313-332, https://doi.org/10.1175/MWR-D-19-0211.1, 2020.

Li, J., Feng, Z., Qian, Y., and Leung, L. R.: MCSs and IDC in the US for 2004 – 2017, available at https://data.pnnl.gov/dataset/13218, http://dx.doi.org/10.25584/1632005, 2020 (last access: Jun 18, 2020).

Lin, Y., and Mitchell, K. E.: 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications, 19th Conf. Hydrology, American Meteorological Society, San Diego, CA, USA, 2005,

Liu, C., Zipser, E. J., and Nesbitt, S. W.: Global distribution of tropical deep convection: Different perspectives from TRMM infrared and radar data, J. Clim., 20, 489-503, https://doi.org/10.1175/JCLI4023.1, 2007.

Lopez, P.: Direct 4D-Var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF, Monthly Weather Review, 139, 2098-2116, https://doi.org/10.1175/2010MWR3565.1, 2011.

Machado, L., Rossow, W., Guedes, R., and Walker, A.: Life cycle variations of mesoscale convective systems over the Americas, Monthly Weather Review, 126, 1630-1654, https://doi.org/10.1175/1520-0493(1998)126<1630:LCVOMC>2.0.CO;2, 1998.

Motew, M., Booth, E. G., Carpenter, S. R., Chen, X., and Kucharik, C. J.: The synergistic effect of manure supply and extreme precipitation on surface water quality, Environmental Research Letters, 13, 044016, https://doi.org/10.1088/1748-9326/aaade6, 2018.

Nearing, M., Pruski, F., and O'neal, M.: Expected climate change impacts on soil erosion rates: a review, Journal of soil and water conservation, 59, 43-50, 2004.

Nelson, B. R., Prat, O. P., Seo, D.-J., and Habib, E.: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons, Weather and Forecasting, 31, 371-394, https://doi.org/10.1175/WAF-D-14-00112.1, 2016.

926   Piani, C., Durran, D., Alexander, M., and Holton, J.: A numerical study of three-dimensional gravity
927   waves triggered by deep tropical convection and their role in the dynamics of the QBO, Journal of the
928   atmospheric sciences, 57, 3689-3702, https://doi.org/10.1175/1520-
929   0469(2000)057<3689:ANSOTD>2.0.CO;2, 2000.

930   Pinto, J. O., Grim, J. A., and Steiner, M.: Assessment of the High-Resolution Rapid Refresh model's
931   ability to predict mesoscale convective systems using object-based evaluation, Weather and Forecasting,
932   30, 892-913, https://doi.org/10.1175/WAF-D-14-00118.1, 2015.

933   Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., and Clark, M.: Simulating
934   North American mesoscale convective systems with a convection-permitting climate model, Climate
935   Dynamics, 1-16, https://doi.org/10.1007/s00382-017-3993-2, 2017.

936   Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., and Bloomfield, J.: Increased crop damage in the
937   US from excess precipitation under climate change, Global Environ. Change, 12, 197-202,
938   https://doi.org/10.1016/S0959-3780(02)00008-0, 2002.

939   Rowe, A. K., Rutledge, S. A., and Lang, T. J.: Investigation of microphysical processes occurring in
940   isolated convection during NAME, Monthly weather review, 139, 424-443,
941   https://doi.org/10.1175/2010MWR3494.1, 2011.

942   Rowe, A. K., Rutledge, S. A., and Lang, T. J.: Investigation of microphysical processes occurring in
943   organized convection during NAME, Monthly weather review, 140, 2168-2187,
944   https://doi.org/10.1175/MWR-D-11-00124.1, 2012.

945   Seeley, J. T., and Romps, D. M.: The effect of global warming on severe thunderstorms in the United
946   States, J. Clim., 28, 2443-2458, https://doi.org/10.1175/JCLI-D-14-00382.1, 2015.

947   Sieglaff, J. M., Hartung, D. C., Feltz, W. F., Cronce, L. M., and Lakshmanan, V.: A satellite-based
948   convective cloud object tracking and multipurpose data fusion tool with application to developing
949   convection, Journal of Atmospheric and Oceanic Technology, 30, 510-525,
950   https://doi.org/10.1175/JTECH-D-12-00114.1, 2013.

951   Smalley, M., L'Ecuyer, T., Lebsock, M., and Haynes, J.: A comparison of precipitation occurrence from
952   the NCEP Stage IV QPE product and the CloudSat Cloud Profiling Radar, Journal of hydrometeorology,
953   15, 444-458, https://doi.org/10.1175/JHM-D-13-048.1, 2014.

954   Song, F., Feng, Z., Leung, L. R., Houze Jr, R. A., Wang, J., Hardin, J., and Homeyer, C. R.: Contrasting
955   spring and summer large-scale environments associated with mesoscale convective systems over the US
956   Great Plains, J. Clim., 32, 6749-6767, https://doi.org/10.1175/JCLI-D-18-0839.1, 2019.

957   Starzec, M., Homeyer, C. R., and Mullendore, G. L.: Storm labeling in three dimensions (SL3D): A
958   volumetric radar echo and dual-polarization updraft classification algorithm, Monthly Weather Review,
959   145, 1127-1145, https://doi.org/10.1175/MWR-D-16-0089.1, 2017.

960   Steiner, M., Houze Jr, R. A., and Yuter, S. E.: Climatological characterization of three-dimensional storm
961   structure from operational radar and rain gauge data, Journal of Applied Meteorology, 34, 1978-2007,
962   https://doi.org/10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2, 1995.

963 Stensrud, D. J.: Effects of persistent, midlatitude mesoscale regions of convection on the large-scale
964 environment during the warm season, Journal of the atmospheric sciences, 53, 3503-3527,
965 https://doi.org/10.1175/1520-0469(1996)053<3503:EOPMMR>2.0.CO;2, 1996.

966 Stensrud, D. J.: Upscale effects of deep convection during the North American monsoon, Journal of the
967 atmospheric sciences, 70, 2681-2695, https://doi.org/10.1175/JAS-D-13-063.1, 2013.

968 Taszarek, M., Allen, J. T., Púčik, T., Hoogewind, K. A., and Brooks, H. E.: Severe convective storms
969 across Europe and the United States. Part II: ERA5 environments associated with lightning, large hail,
970 severe wind, and tornadoes, J. Clim., 33, 10263-10286, https://doi.org/10.1175/JCLI-D-20-0346.1, 2020.

971 Thompson, A. M., Tao, W.-K., Pickering, K. E., Scala, J. R., and Simpson, J.: Tropical deep convection
972 and ozone formation, Bulletin of the American Meteorological Society, 78, 1043-1054,
973 https://doi.org/10.1175/1520-0477(1997)078<1043:TDCAOF>2.0.CO;2, 1997.

974 Tian, B., Held, I. M., Lau, N. C., and Soden, B. J.: Diurnal cycle of summertime deep convection over
975 North America: A satellite perspective, J. Geophys. Res.-Atmos., 110,
976 https://doi.org/10.1029/2004JD005275, 2005.

977 Tippett, M. K., Allen, J. T., Gensini, V. A., and Brooks, H. E.: Climate and hazardous convective
978 weather, Current Climate Change Reports, 1, 60-73, https://doi.org/10.1007/s40641-015-0006-6, 2015.

979 Twohy, C. H., Clement, C. F., Gandrud, B. W., Weinheimer, A. J., Campos, T. L., Baumgardner, D.,
980 Brune, W. H., Faloona, I., Sachse, G. W., and Vay, S. A.: Deep convection as a source of new particles in
981 the midlatitude upper troposphere, J. Geophys. Res.-Atmos., 107, AAC 6-1-AAC 6-10,
982 https://doi.org/10.1029/2001JD000323, 2002.

983 Walker, J. R., MacKenzie Jr, W. M., Mecikalski, J. R., and Jewett, C. P.: An enhanced geostationary
984 satellite–based convective initiation algorithm for 0–2-h nowcasting with object tracking, Journal of
985 Applied Meteorology and Climatology, 51, 1931-1949, https://doi.org/10.1175/JAMC-D-11-0246.1,
986 2012.

987 Wang, P. K.: Moisture plumes above thunderstorm anvils and their contributions to cross‐tropopause
988 transport of water vapor in midlatitudes, J. Geophys. Res.-Atmos., 108,
989 https://doi.org/10.1029/2002JD002581, 2003.

990 Yang, Q., Houze Jr, R. A., Leung, L. R., and Feng, Z.: Environments of long‐lived mesoscale
991 convective systems over the central United States in convection permitting climate simulations, J.
992 Geophys. Res.-Atmos., 122, 13,288-213,307, https://doi.org/10.1002/2017JD027033, 2017.

993 Yuan, H., McGinley, J. A., Schultz, P. J., Anderson, C. J., and Lu, C.: Short-range precipitation forecasts
994 from time-lagged multimodel ensembles during the HMT-West-2006 campaign, Journal of
995 Hydrometeorology, 9, 477-491, https://doi.org/10.1175/2007JHM879.1, 2008.

996 Zhang, K., Randel, W. J., and Fu, R.: Relationships between outgoing longwave radiation and diabatic
997 heating in reanalyses, Climate Dynamics, 49, 2911-2929, https://doi.org/10.1007/s00382-016-3501-0,
998 2017.

999