# 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° × 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 hydrometeors 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 NCAR/UCAR RDA (https://rda.ucar.edu/datasets/ds507.5/,

179    last access: Dec 28, 2019). We regrid the original Stage IV precipitation from polar stereographic

180    grids to the 4 km satellite Merged IR grids by using the "neareststod" method from the ESMF

181    'NCL' module (https://www.ncl.ucar.edu/Applications/ESMF.shtml). The "neareststod" method

182    maps each destination point to the closest source point.

183    *2.1.4 ERA5 melting level dataset*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

200  interpolated to regular latitude/longitude grids with a resolution of $0.25° \times 0.25°$. We calculate

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

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

203  to the 4-km satellite Merged IR grids by using the ESMF "nereststod" method.

204      We summarize the basic information of the four types of source datasets in Table A1. And,

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

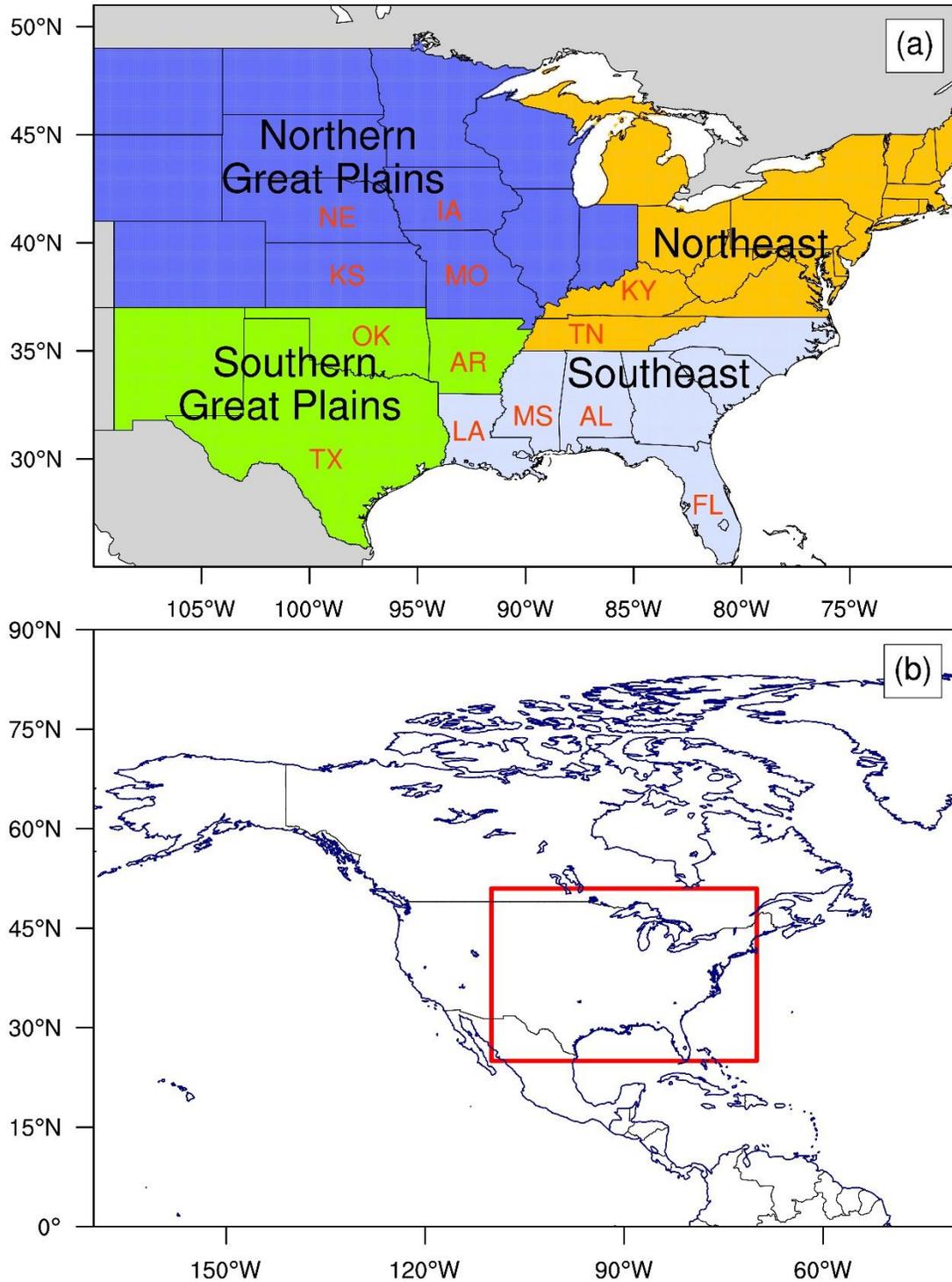
206  (Figure 1), which covers the US east of the Rocky Mountains and excludes the western US. The

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

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

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

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

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

212
213 **Figure 1.** (a) Data product domain and region definitions. Blue shading denotes the Northern
214 Great Plains (NGP), green-yellow shading denotes the Southern Great Plains (SGP), light steel
215 blue shading denotes the Southeast (SE), and orange shading denotes the Northeast (NE). The
216 locations of some US states within each region are also labeled. TX is for Texas, OK for
217 Oklahoma, KS for Kansas, NE for Nebraska, IA for Iowa, MO for Missouri, AR for Arkansas,

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

220  2.2 Algorithm description

221  *2.2.1 SL3D algorithm*

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

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

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

225  following the criteria listed in Table A2. We run the SL3D algorithm for 2004 – 2017 by using

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

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

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

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

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

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

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

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

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

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

236  definitions).

237    *2.2.2 MCS/IDC identification and tracking*

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

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

240    IDC events simultaneously.

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

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

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

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

245    241 K. Cloud systems that do not contain a cold cloud core but with $T_b < 241$ K are also labeled

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

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

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

249    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

250    available) (Feng et al., 2019). We use a $5 \times 5$ pixel moving window to smooth $Z_H$. Figure 2b

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

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

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

254    50%. "Linked" means the CCSs are considered to be from the same cloud systems. FLEXTRKR

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

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

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

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

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

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

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

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

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

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

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

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

267   between consecutive hours (Figures A1 and A2). A track associated with the largest CCS is

268   defined as the main track (Figure A3), and smaller tracks from merges/splits are regarded as

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

270   a "merge"/"split" track associated with an MCS/IDC event must have a CCS-based lifetime of

271   no more than 5 hours. Otherwise, we treat it as an independent track.

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

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

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

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

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

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

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

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

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

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

282    1 hour during the lifecycle of the track when PF and CCF are present (PF and CCF major axis

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

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

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

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

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

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

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

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

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

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

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

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

295    04T03:00:00Z. Figure 2f displays the spatial coverages of MCS/IDC tracks at that time at pixel

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

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

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

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

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

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

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

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

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

305    are also archived in the data product.

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

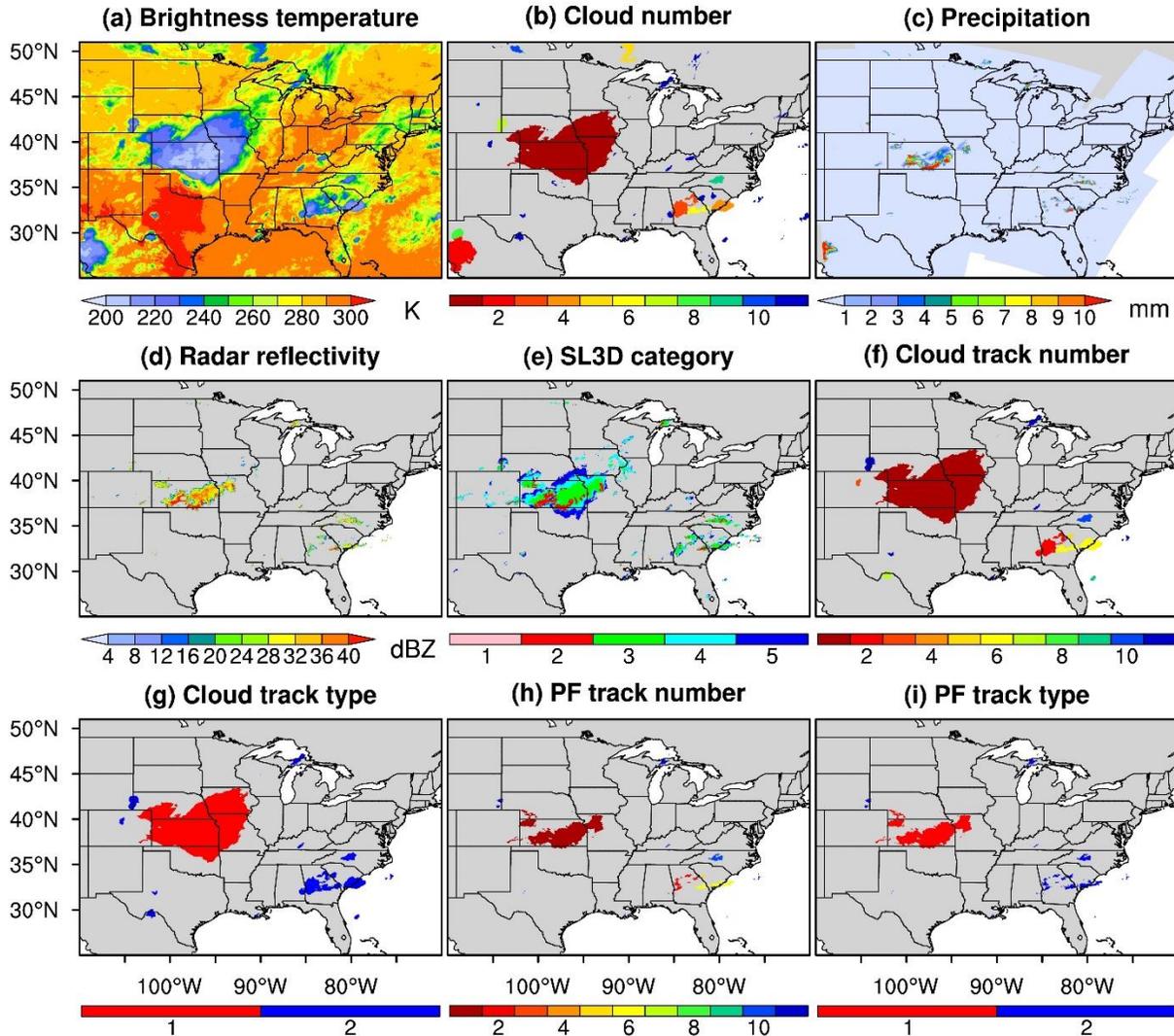
307    time of each continuous tracking is 00Z on 1 January, and the ending time is 23Z on 31

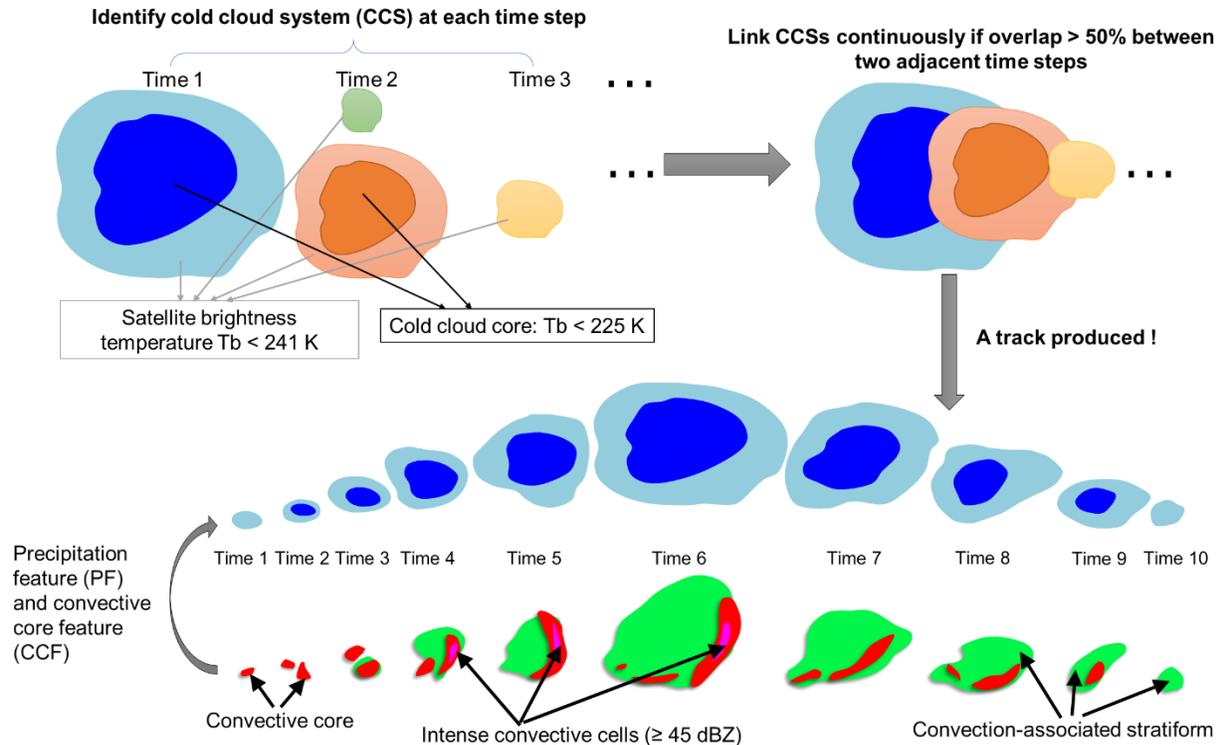
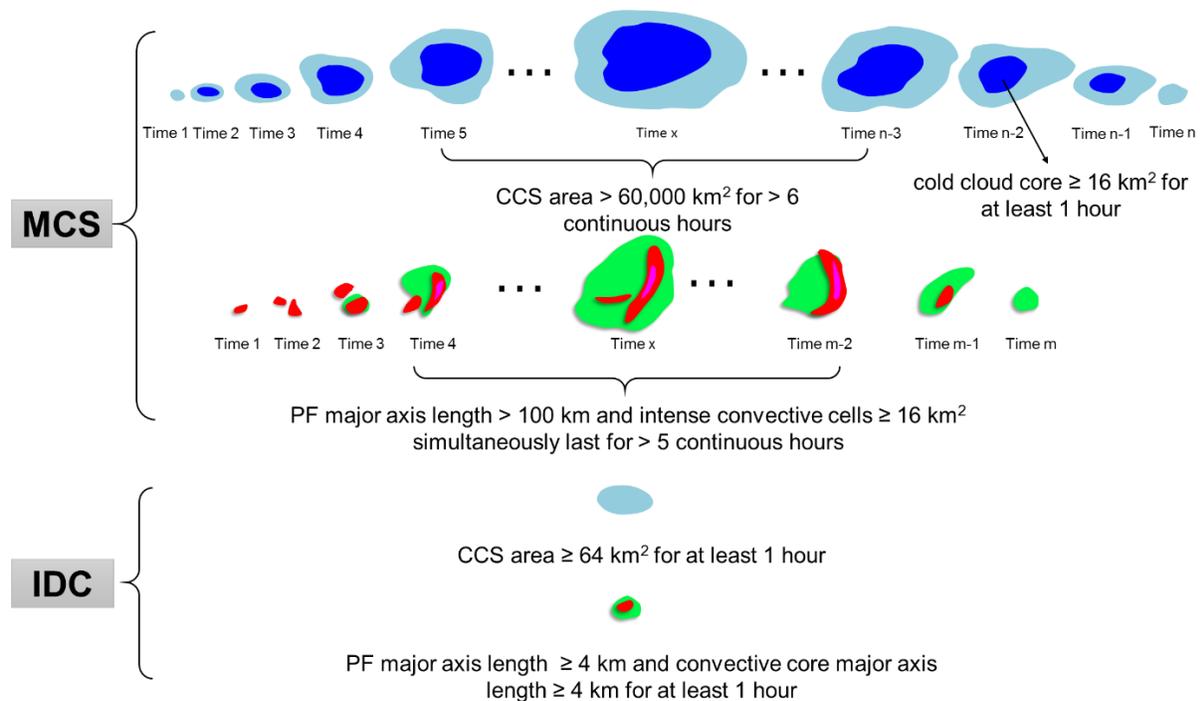
308    December. Because winter has the fewest deep convection events, very few MCS/IDC events

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

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

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

312    MCS/IDC statistics.

313
314 **Figure 2.** FLEXTRKR pixel-level outputs at 03:00:00Z on July 4, 2005. (a) is satellite $T_b$. (b)
315 shows identified CCS labels. CCS labels are unique at each hour. (c) is Stage IV hourly
316 accumulated precipitation. (d) is Gridrad $Z_H$ at 2 km (if it is not available, $Z_H$ at 3 km is provided
317 if it is available). (e) is the SL3D classification results: 1, convective updraft; 2, convective; 3,
318 precipitating stratiform; 4, non-precipitating stratiform; 5, anvil. (f) displays the track numbers to
319 which pixels belong. Here, the track numbers are not the real values in the MCS/IDC data
320 product. The track numbers should be unique throughout the whole running period. We adjust
321 the track numbers here to make the figure clear. Similar to "PF track number." (g) gives
322 information on whether the pixels belong to MCS (marked as 1) or IDC (marked as 2) tracks,
323 which correspond to the tracks shown in (f). (h) also displays the track numbers to which the
324 pixels belong, but only for pixels with precipitation > 1 mm h$^{-1}$. (i) is like (g) but corresponds to
325 (h). All these variables are stored in the FLEXTRKR hourly pixel-level output files.

326

327
**Figure 3.** Schematic of the FLEXTRKR algorithm highlighting three key steps in the algorithm:
(1) the identification of CCS (upper left), (2) linking of overlapping CCSs (upper right), and (3)
the tracking of both PF and CCF (bottom).

331
**Figure 4.** Definition of MCS and IDC based on the FLEXTRKR algorithm shown in Figure 3
and the specific threshold values used in the algorithm.

## 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 A4. Results show winter has the largest fraction of MCS/IDC events with longer lifetimes than other seasons.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

372    initiation time. As shown in Figure A5, most MCS and IDC events initiate in the afternoon of

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

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

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

376    20-dBZ echo-top heights, which can be used to represent their mean convective intensities, are

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

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

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

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

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

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

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

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

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

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

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

388    Mountains for 2004 – 2017.

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$^1$ / km$^2$ | 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$^4$ / km$^2$ | 8,273 | 8,589 | 8,465 | 7,752 | 7,293 | 494 | 509 | 483 | 502 | 528 |
| PF stratiform area / km$^2$ | 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 / km$^2$ | 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 |

[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.

402    3.2 Precipitation characteristics from different sources

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

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

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

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

407    change the values shown in this study but will neither affect the comparison among MCS, IDC,

408    and non-convective (NC) precipitation nor their spatial distribution patterns. Here, NC

409    precipitation refers to precipitation not associated with any MCS or IDC events and is mainly

410    from stratiform rain. Total precipitation is the sum of MCS, IDC, and NC precipitation. It is

411    noteworthy that NC precipitation may contain some convection-associated rain due to the

412    limitation of the source datasets and the algorithms used in this study. More relevant details are

413    discussed in Section 3.2.3 and Section 4.

414    *3.2.1 Annual spatial distributions of different types of precipitation*

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

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

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

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

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

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

421    (2019) over the ECONUS due to their different algorithms and stricter criteria to track and

422    define MCSs.

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

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

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

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

427   especially in the South Central US (MCS precipitation fractions: ~60%). The spatial

428   distribution patterns of MCS annual precipitation amounts and fractions in Figure 5 are similar

429   to those from Haberlie and Ashley (2019), although their MCS precipitation fractions are

430   generally lower than our results. IDC precipitation is concentrated in the SE and NE coastal

431   areas, with peak values in Florida. NC precipitation is substantial in the eastern and southern

432   regions with ample moisture supply and contributes over 35% to the total precipitation across

433   most of the NE region. The coastal area near Louisiana, which is significantly affected by all

434   three types of precipitation, has the most total precipitation with annual amounts of over 1,350

435   mm. The annual total precipitation amounts in most regions of SE also exceed 1,050 mm due to

436   MCS contributions. While the total precipitation amounts in most regions of Florida are also

437   over 1,050 mm, they are mainly attributed to IDC.

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

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

440   areas, have larger precipitation intensities than the northern areas. The MCS precipitation

441   intensities are the largest in Texas, Louisiana, Oklahoma, and Kansas, significantly shifting

442   west compared to MCS precipitation amounts. Unlike IDC precipitation amounts concentrating

443   in the SE and NE coastal areas, IDC precipitation intensities are the largest over the SGP and

444   SE. IDC precipitation intensities over the NE are much smaller compared to the SGP and SE,

445   similar to NC precipitation intensities. We summarize the annual mean precipitation amounts

446   and intensities of different types of precipitation in the NGP, SGP, SE, and NE in Table A3.

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

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

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

450    accumulated rainfall in the grid cell. The distributions of MCS/IDC precipitation intensities,

451    although not the main factor, can also affect the distributions of MCS/IDC precipitation

452    amounts. For example, the maximum MCS hours are located around Missouri (Figures 7a), but

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

454    larger MCS precipitation intensities in the southern regions contribute more to the MCS

455    precipitation amount in the southern US. In addition, a large number of IDC events (IDC

456    hours > 60 h $yr^{-1}$) occur in the NE region along the Appalachian Mountains (Figure 7b), but

457    IDC in that region only contributes to 20% – 30% of the total precipitation amount (Figure 6b)

458    due to the low precipitation intensities (Figure 5f).

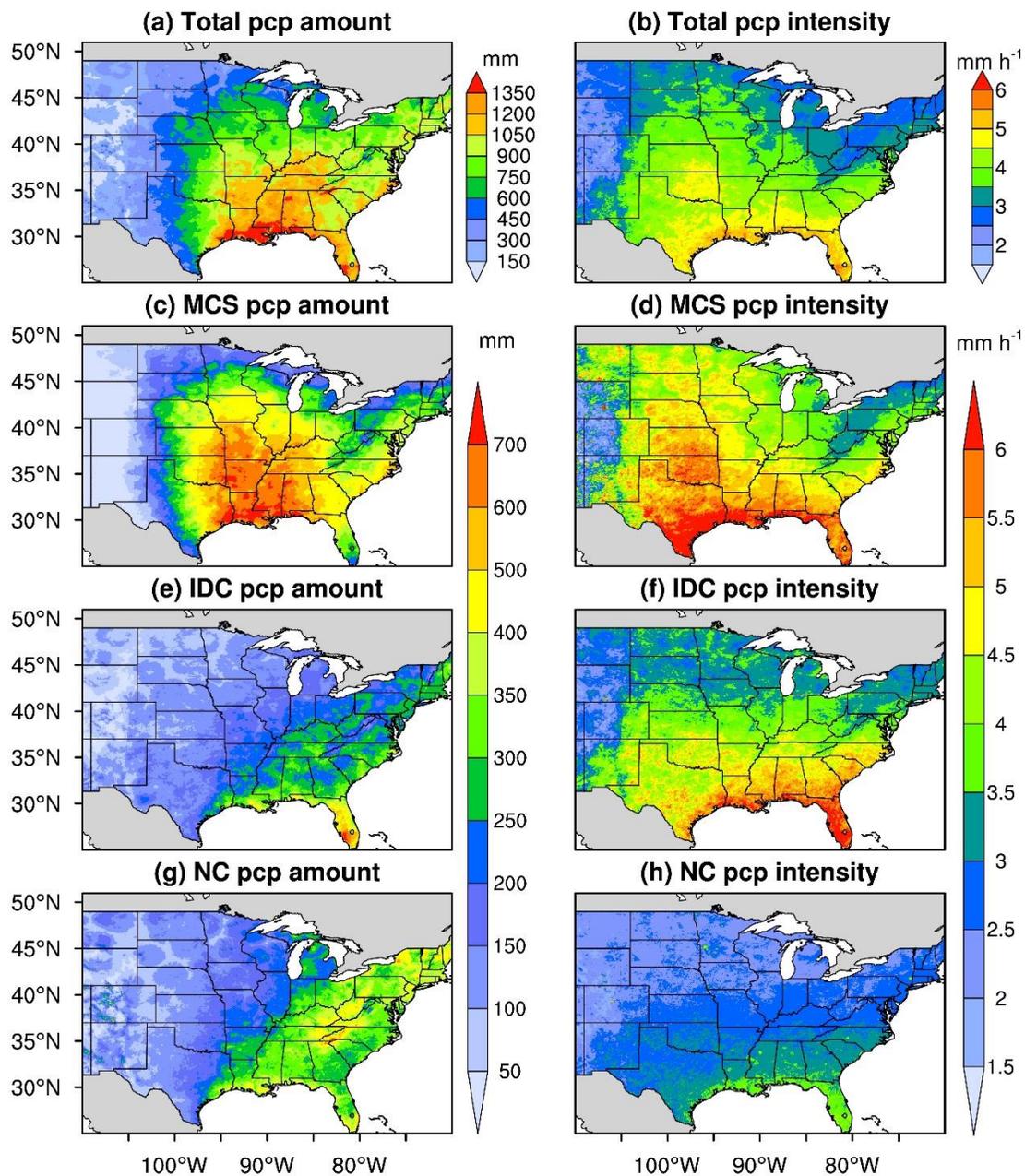
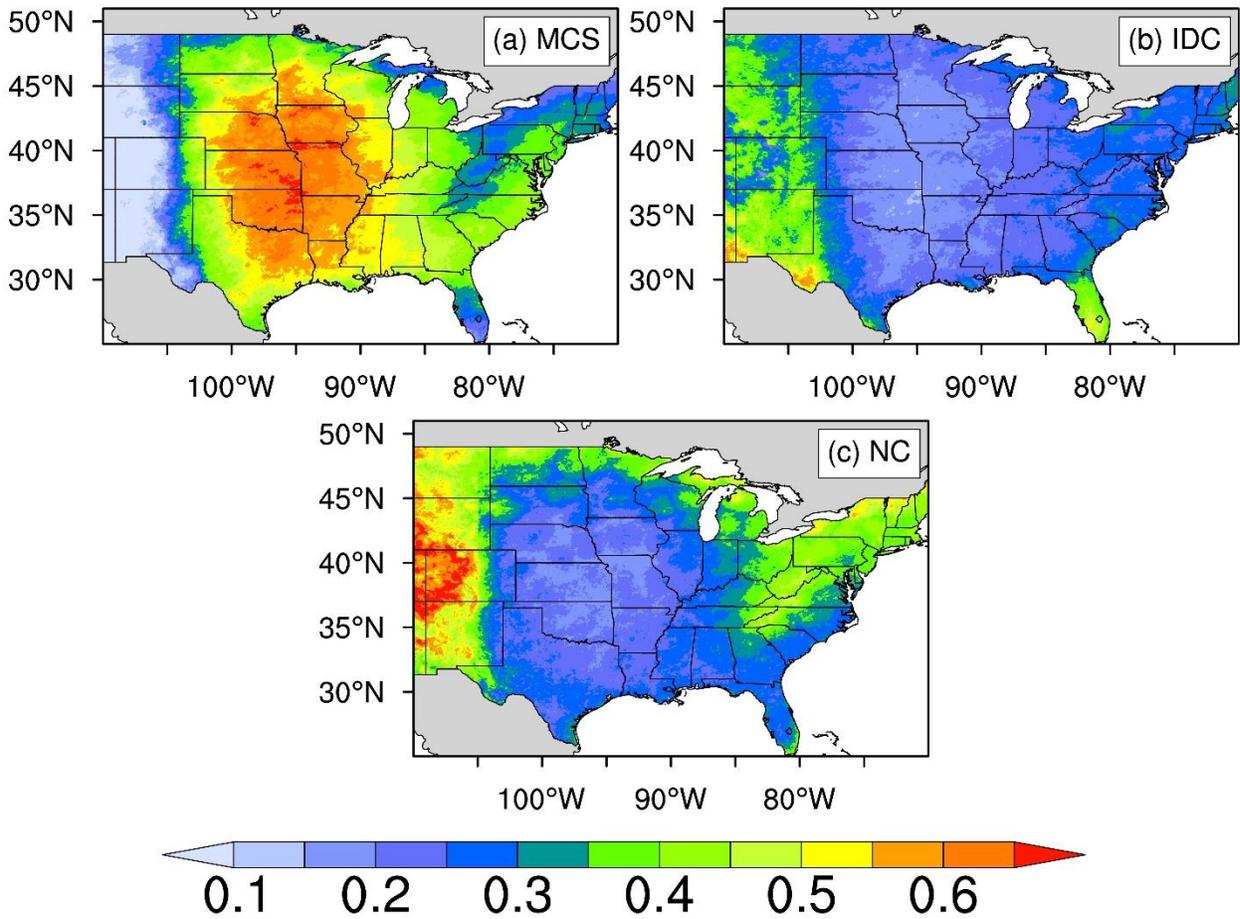
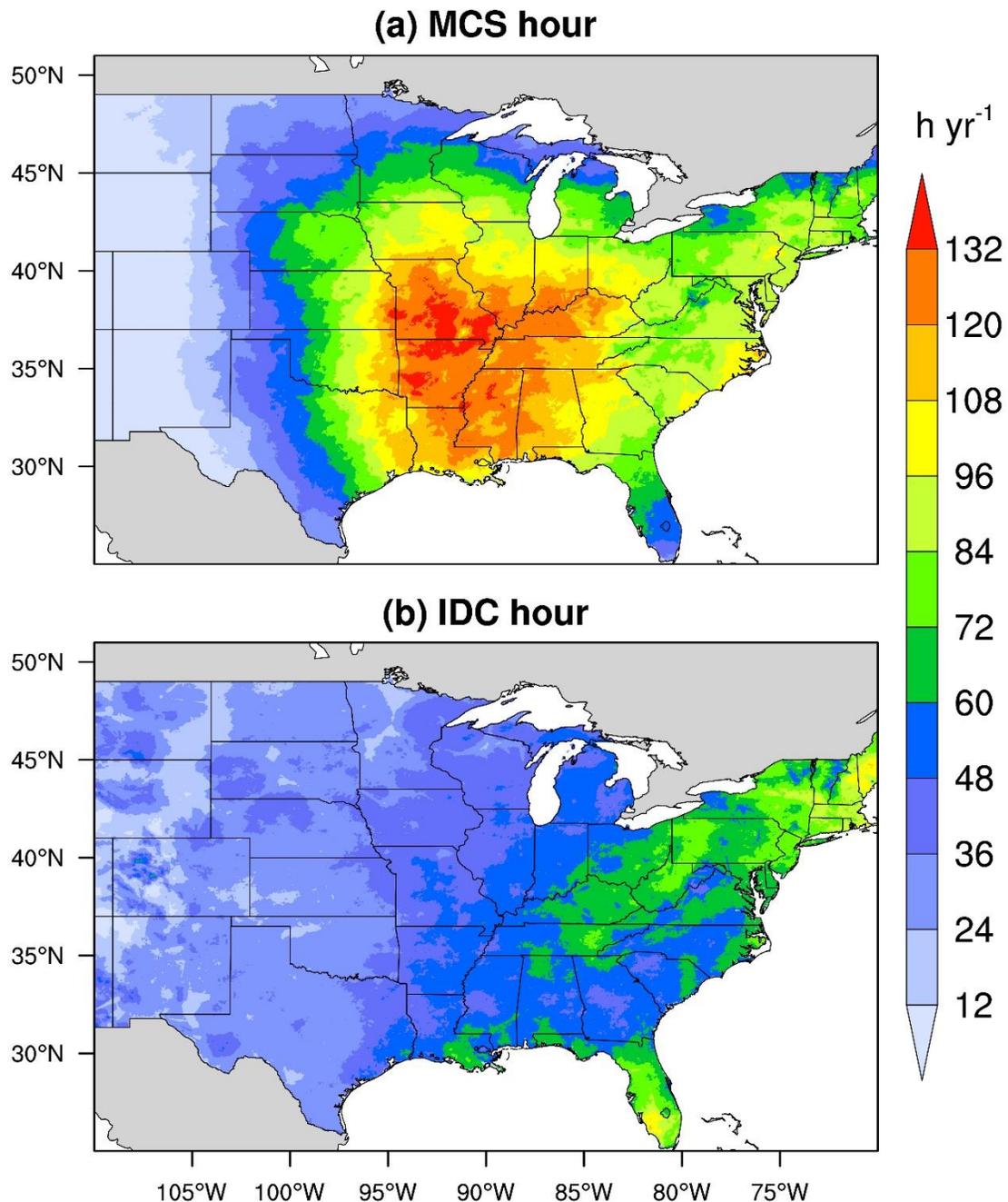



**Figure 5.** Distributions of annual mean precipitation amounts (a, c, e, g) and intensities (b, d, f, h) for different types of precipitation for 2004 – 2017. (a) and (b) are for total precipitation, (c) and (d) are for MCS precipitation, (e) and (f) are for IDC precipitation, and (g) and (h) are for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

464
465 **Figure 6.** Distributions of the fractions of different types of precipitation (MCS, IDC, NC).
466 Here, precipitation refers to annual mean values for 2004 – 2017. We exclude hourly data with
467 precipitation ≤ 1 mm h$^{-1}$ in the calculation.

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

473 *3.2.2 Seasonal spatial distributions of different types of precipitation*

474    Figures 8, A6, and A7 display the mean seasonal distributions of precipitation amounts,

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

476    2017. The MCS precipitation center migrates northwards from Arkansas in spring to northern

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

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

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

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

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

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

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

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

485    intensity in summer (5.2 mm h$^{-1}$) is also larger than that in spring (4.1 mm h$^{-1}$) (Figure A7). The

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

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

488    2019). As a result, spring MCSs are larger and longer-lasting, and they produce more rainfall

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

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

491    amounts are generally > 35% over the Northern and Southern Great Plains in spring and

492    summer and can reach up to over 70% within the MCS precipitation center (Figures A6a –

493    A6b). The results are roughly consistent with Fritsch et al. (1986), which showed that MCSs

494    accounted for about 30% – 70% of the warm-season (April-September) precipitation over much

495    of the region between the Rocky Mountains and the Mississippi River. The results are also

496    consistent with Haberlie and Ashley (2019) showing MCS precipitation fractions generally >

497    30% with a peak > 60% over the Great Plains between May and August. Due to the low

498    precipitation amounts of IDC and NC, the fractions of MCS precipitation amounts in autumn

499    and winter are also large, showing over 50% within the MCS precipitation center (Figures A6c

500    – A6d).

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

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

amounts (Figures 8i – 8l and A6e – A6h). Winter has the least IDC precipitation. Areas of high

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

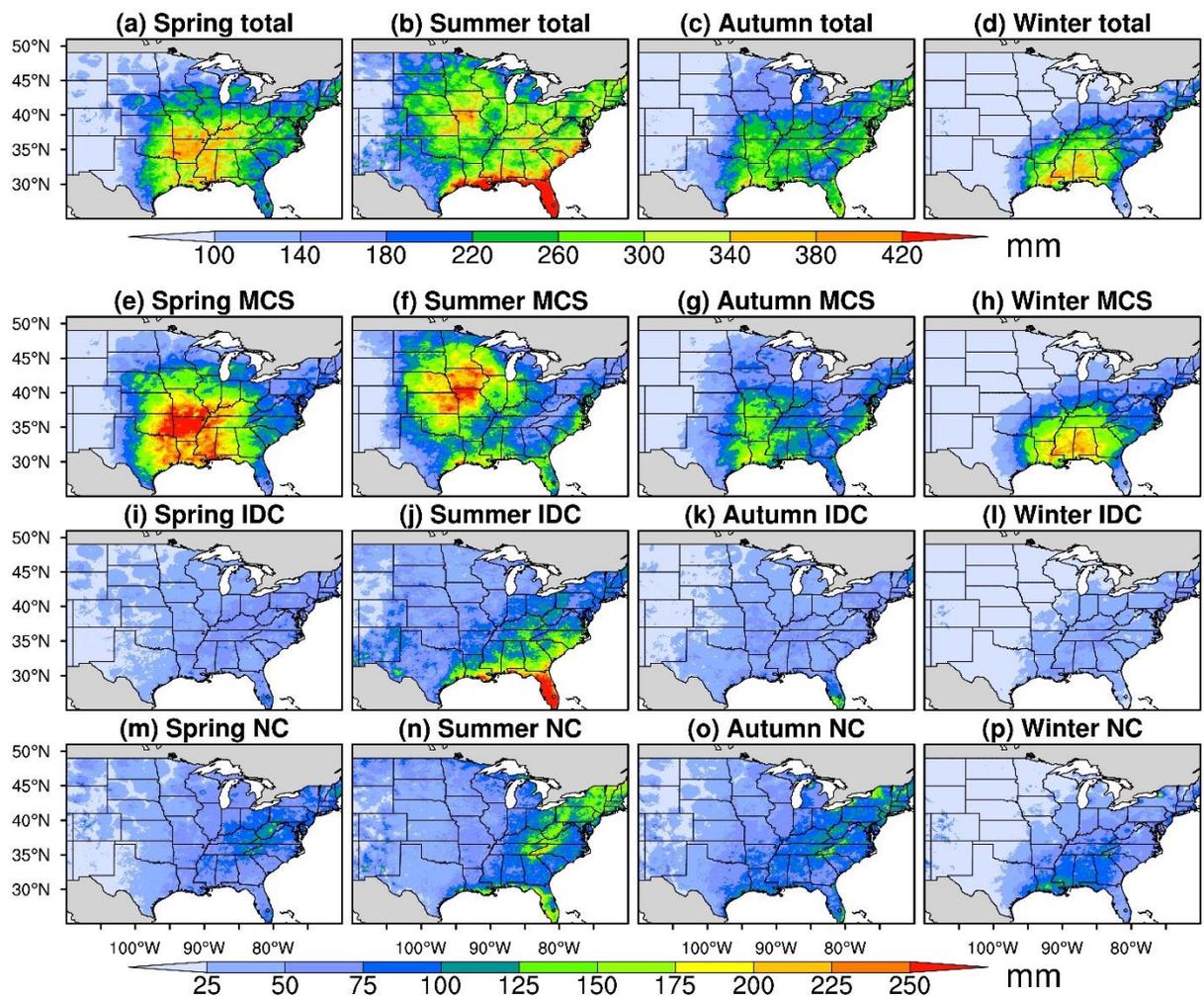
local conditions such as moisture availability, local solar radiation, and land-atmosphere

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

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

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

is smaller than that of winter (32%) (Figures A6i – A6l). Winter NC precipitation center occurs

in the SE coastal areas (Figure 8p).

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

518      The precipitation intensities of all three types peak in summer and reach minimums in

519 winter (Figure A7). In each season, precipitation intensities in the south are larger than those in

520 the north except for MCS precipitation intensities in summer, which maximize in Oklahoma.

521 We summarize the mean seasonal precipitation amounts and intensities of different types of

522 precipitation over the 4 climate regions of Figure 1 in Table A4.

523 *3.2.3 Diurnal cycles of different types of precipitation*

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

525 IDC, and NC in the NGP, SGP, SE, and NE, respectively. Generally, MCS precipitation peaks

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

527 SGP to summer in the NGP reflects the northward migration of the MCS precipitation center in

528 the Great Plains (Figures 8e and 8f).

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

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

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

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

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

534 center first appears in the western SE regions (Alabama, Mississippi, and Louisiana) in April,

535 and then moves northwards to Arkansas in May. In summer, the SE MCS precipitation diurnal

536 cycles are more like those of IDC (Figures 9g and 9h), peaking in the late afternoon and much

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

538    variations between the Great Plains and SE were also identified by Haberlie and Ashley (2019).

539    We find that most summer MCS precipitation over the SE occurs near the coastal areas (Figure

540    8f), far from the MCS precipitation center in northern Missouri and Iowa, suggesting either a

541    different MCS genesis mechanism in the SE from those in the SGP and NGP (Feng et al., 2019)

542    or long-duration deep convective systems showing MCS characteristics (Geerts, 1998). In

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

544    precipitation in September shows mixed features of summer and autumn with peaks both in the

545    morning and the afternoon. In winter months, the diurnal cycle of the SE MCS precipitation

546    shifts from the autumn feature to the spring feature, with peaks shifting from the morning to the

547    afternoon. The distinct diurnal cycles of SE MCS precipitation in different seasons in Figure 9g

548    are roughly consistent with the corresponding seasonal diurnal variations of MCS occurrence

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

550    central time between the initiation and decay of the MCS.

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

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

553    of local instability driven by the solar forcing on IDC development. NC precipitation (Figures

554    9c, 9f, 9i, and 9l) shows some diurnal cycle characteristics similar to IDC precipitation. It may

555    be caused by the limitation of the temporal resolution of the datasets used in the FLEXTRKR

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

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

558    analysis time. These weak IDC could be aliased to NC precipitation, therefore showing some

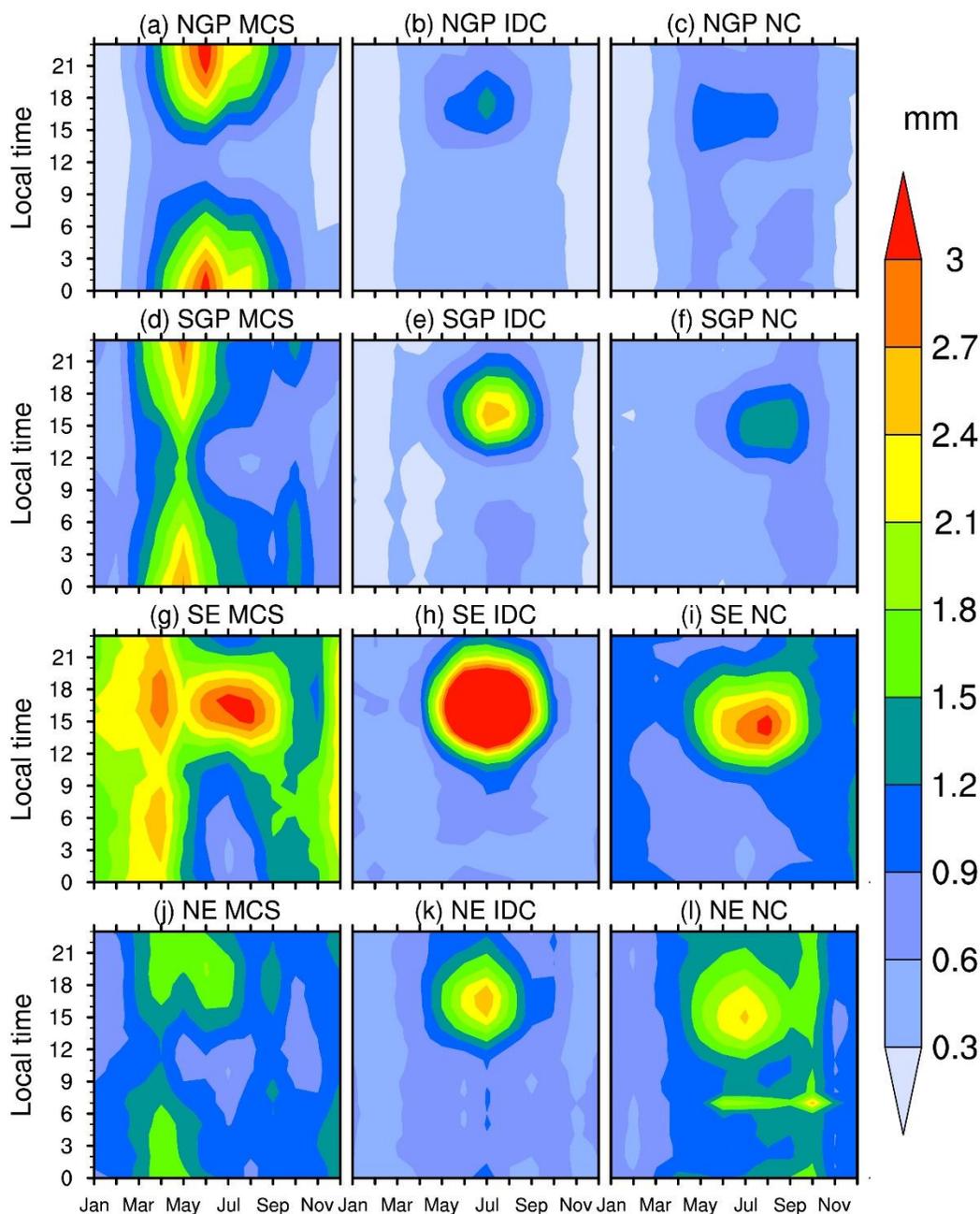
559    similar diurnal cycles as IDC. Another possible reason is that the FLEXTRKR algorithm may

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

561    precipitation exhibits some IDC characteristics.

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

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

564     warm season (Figure A8).

565
566     **Figure 9.** Monthly mean diurnal cycles of precipitation amounts from MCSs (a, d, g, j), IDC (b,
567     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
568     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) algorithm biases, bad gauge data removal inconsistency among different RFCs, multisensory

591     processing algorithm inconsistency among different RFCs, and mosaicking border

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

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

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

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

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

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

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

599     4,800 $km^2$) with intense precipitation (> 40 mm $h^{-1}$) (Figure A9), which is physically not

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

601     precipitation."

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

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

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

605     IDC based on any single dataset. And to consider the potential error of ERA5 melting level

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

607     algorithm (Table A2).

608     4.2 The impact of missing data

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

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

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

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

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

614    (1.8%). All other years have no more than 57 missing hours. During the link procedure of the

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

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

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

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

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

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

621    2017 in Figure A10. The fractions are over 97% for all grid cells of the US in the domain. Most

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

623    the data product.

624        Figure A11 shows the fractions of available Gridrad reflectivity data from 2004 to 2017

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

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

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

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

629    90% between 2 and 9 km and 80% between 10 and 12 km. The completeness of the Gridrad

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

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

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

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

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

635    4.3 Temporal resolution limitation of the source datasets

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

637    possible aliasing from IDC precipitation. Some weak IDC events are so short that the hourly

638    data cannot properly capture their occurrence, especially for Gridrad $Z_H$, which only includes

639    reflectivities within $\pm$ 3.8 minutes of each hour. We calculate the cumulative distribution

640    functions of PF-based lifetimes for MCS and IDC events and their associated precipitation in

641    the data product for 2004 – 2017, as shown in Figure 10. About 75% of IDC events have a PF-

642    based lifetime of 1 hour. Therefore, it is almost certain that we miss some IDC events shorter

643    than 1 hour in the data product. Here we give an estimate of the probability $p$ that a given IDC

644    event with a convective signal duration of $x$ minutes is detected by radar, as expressed below:

645    $$p = \frac{2 \times 3.8}{60 - x} \tag{1}$$

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

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

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

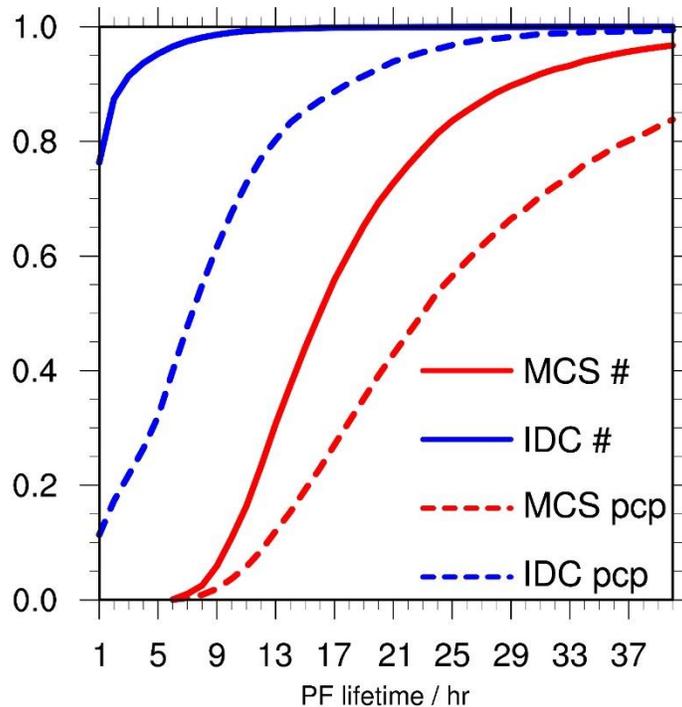
649    the distribution of IDC convective signals with durations less than 1 hour using the currently

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

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

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

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

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

655
656 **Figure 10.** Cumulative distribution functions of PF-based lifetimes for MCS and IDC events
657 and their associated precipitation in the data product domain for 2004 – 2017. The red solid line
658 is for the number of MCSs, the red dash line for MCS associated precipitation, the blue solid
659 line for the number of IDC events, and the blue dash line for IDC associated precipitation.

660 4.4 The impact of MCS and IDC definition criteria

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

662 2019; Geerts et al., 2017; Haberlie and Ashley, 2019; Pinto et al., 2015; Prein et al., 2017).

663 Here, we briefly examine the impact of different MCS/IDC definition criteria on the data

664 product. We change the definition of MCSs to relax the CCS and PF size and duration

665 thresholds. Specifically, the second and third criteria listed in Section 2.2.2 are modified as

666 follows: 2) CCS areas associated with the track surpass 40,000 $km^2$ for more than 4 continuous

667 hours; 3) PF major axis length exceeding 80 km and intense convective cell areas $\geq$ 16 $km^2$

668 exist for more than 3 consecutive hours. And we also require that each merge/split-track

669 associated with MCS/IDC events must have a CCS-based lifetime of no more than 3 hours. We

670 keep the definition of IDC the same as described in Section 2.2.2, which is a limit for IDC that

671 we can identify based on the source datasets.

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

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

674    criteria (Tables 1 and A5). The annual number of MCSs identified in 2004 – 2017 increases

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

676    151 in autumn, and 37 to 62 in winter. As PF-based lifetimes of MCS/IDC events in summer

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

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

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

680    IDC events is smaller than the increase in the number of MCSs. Annual mean MCS

681    precipitation east of the Rocky Mountains increases from 313 mm to 353 mm, while IDC

682    precipitation decreases from 170 mm to 130 mm. The fraction of MCS precipitation only

683    increases by 6% (from 45% to 51%), compared to the almost doubling of MCS number (from

684    454 to 857), suggesting the MCS definition in the original data product is capable of capturing

685    most of the important MCSs with heavy precipitation. Similar to MCS numbers, summer has

686    the most increase in MCS precipitation amount, from 100 mm to 119 mm. And annual mean

687    MCS and IDC precipitation intensities decrease slightly as MCS precipitation intensities are

688    somewhat larger than IDC in most regions (Tables A3, A4, A6, and A7). We summarize the

689    regional precipitation statistics of the NGP, SGP, SE, and NE based on the new definition in

690    Tables A6 and A7.

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

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

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

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

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

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

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

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

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

700 characteristics of MCS and IDC events.

701 4.5 Recommendations for the usage of the MCS/IDC data product

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

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

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

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

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

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

708 border and over the Rocky Mountains (Figure A11), we recommend the usage of the data

709 product within the CONUS east of the Rocky Mountains to alleviate the impact of the

710 termination of MCS/IDC tracks due to poor radar coverage and missing radar data beyond their

711 maximum scan range.

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

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

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

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

716 examining a specific MCS/IDC track if there are too many missing data (precipitation, $T_b$, or

717 $Z_H$) along the track. Users planning to apply the data product for a specific case study should

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

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

720     https://rda.ucar.edu/datasets/ds841.0/ (Table A1).

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

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

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

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

725     on atmospheric circulation, environmental conditions associated with the initiation and

726     evolution of MCS and IDC events, and MCS/IDC associated weather hazards.

727 **5 Data availability**

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

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

730     original format of the data files is NetCDF-4, and we archive them as compressed files for each

731     year so that the data product is easily accessible. The user guide contains a brief explanation

732     about the approach to develop the data product and a detailed description of the data file content

733     to help users understand the data product.

734 **6 Conclusions**

735     Here we present a unified high-resolution (4 km, hourly) data product that describes the

736     spatiotemporal characteristics of MCS and IDC events from 2004 to 2017 east of the Rocky

737     Mountains over the CONUS. We produce the data product by applying an updated FLEXTRKR

738     algorithm to the NCEP/CPP L3 4 km Global Merged IR V1 $T_b$ dataset, ERA5 melting level

739     heights, the 3-D Gridrad radar reflectivity dataset, and the Stage IV precipitation dataset.

740     Climatological features of the MCS and IDC events from the data product are compared, with a

741 focus on their precipitation characteristics. Consistent with our definitions of MCSs and IDC in

742 the FLEXTRKR algorithm, we find that MCSs have much broader spatial coverage and longer

743 duration than IDC events. While there are many more frequent IDC occurrences than MCSs,

744 the mean convective intensities of IDC events are comparable to those of MCSs. MCS and IDC

745 events both contribute significantly to precipitation east of the Rocky Mountains but with

746 distinct spatiotemporal variabilities. MCS precipitation affects most regions of the eastern US in

747 all seasons, especially in spring and summer. The MCS precipitation center migrates

748 northwards from Arkansas in spring to northern Missouri and Iowa in summer, followed by a

749 southward migration to Louisiana in autumn, and finally to Mississippi and Alabama in the

750 Southeast in winter. IDC precipitation mostly concentrates in the Southeast in summer. IDC

751 precipitation shows a significant diurnal cycle in summer months with a peak around 16:00 –

752 17:00 Local Time over all regions east of the Rocky Mountains. In contrast, MCS precipitation

753 peaks during nighttime in spring and summer for most regions except for the Southeast, where

754 MCS precipitation peaks in the late afternoon in summer, similar to IDC precipitation. Lastly,

755 we analyze the potential uncertainties of the data product and the sensitivity of the dataset to

756 MCS definitions and give our recommendations for the usage of the data product. The data

757 product will be useful for investigating the atmospheric environments and physical processes

758 associated with convective systems, quantifying the impacts of convection on hydrology,

759 atmospheric chemistry, severe weather hazards, and other aspects of the energy, water, and

760 biogeochemical cycles, and improving the representation of convective processes in weather

761 and climate models.

# Appendix A

**Table A1.** Summary of source datasets used to develop the MCS/IDC data product

| Dataset name | NCEP/CPP L3 half-hourly 4 km Global Merged IR | Three-dimensional Gridded NEXRAD Radar (Gridrad) | NCEP Stage IV precipitation | ERA5 melting level |
|---|---|---|---|---|
| Dataset version | V 1 | V 3.1 | V 1.0 | |
| DOI | 10.5067/P4HZB9N27EKU | 10.5065/D6NK3CR7 | 10.5065/D69Z93M3 | 10.24381/cds.adbb2d47 |
| URL | https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_1/summary | https://rda.ucar.edu/datasets/ds841.0/ | https://rda.ucar.edu/datasets/ds507.5/ | https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview |
| Last access | Dec 28, 2019 | Jan 2, 2020 | Dec 28, 2019 | Jan 24, 2020 |
| Initial spatial resolution | Horizontal: ~ 4 km | Horizontal: 0.02° Vertical: 1 km | Horizontal: ~ 4 km | Horizontal: 0.25° |
| Initial temporal resolution | 0.5 hours | 1 hour | 1 hour | 1 hour |

Table A2. The classification criteria of the Storm Labeling in Three Dimensions (SL3D) algorithm in this study

| SL3D category | Criteria |
| --- | --- |
| convective | $Z_H^1 = 25$ dBZ echo-top height $\geq 10$ km; or $Z_H \geq 45$ dBZ above $(Z_{melt}^2 + 1)$ km; or $Z_H$ peakedness[3] exceeding thresholds[4] in at least 30% of the echo column between surface and 9 km. After the above filtering, exclude isolated convective grid points. Finally, grid points that have $Z_{Hmax}^5 \geq 25$ dBZ and are immediately adjacent to other convective grid points are classified as convective. |
| precipitating stratiform | $Z_H \geq 20$ dBZ at 3 km; or $Z_H \geq 10$ dBZ at 1 km or 2 km |
| non-precipitating stratiform | no echo or $Z_H < 20$ dBZ at 3 km, and echo presents above 3 km. If no echo at 3 km − 5 km, but echo presents above 5 km, classified as an anvil. |
| anvil | No echo at 3 km − 5 km, but echo presents above 5 km |
| convective updraft | convective grid points satisfy: (1) $Z_{Hmax} \geq 40$ dBZ, and (2) $\dfrac{\partial Z_H}{\partial z} \geq 8$ dBZ km$^{-1}$ with echoes in at least six of eight horizontally adjacent grid volumes presents between the surface and 7 km. |

[1] $Z_H$: logarithmic radar reflectivity.

[2] $Z_{melt}$: melting level height. If temperatures at different vertical levels within a grid column are all below zero, there is no melting level. In this situation, we set $Z_{melt} = -2$.

[3] Peakedness is the difference between the $Z_H$ of the grid point being evaluated and the median $Z_H$ of a horizontal 12-km radius around the point.

[4] $threshold = \max\left(4.0\,dBZ, 10.0 - \dfrac{Z_H^2}{337.5}\,dBZ\right)$.

[5] $Z_{Hmax}$ denotes column max reflectivity.

773
774 **Table A3.** Annual mean precipitation amounts and intensities for different types of precipitation in different regions of the US for 2004 – 2017

| | Precipitation amount / mm | | | | Precipitation intensity / mm h$^{-1}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | MCS | IDC | NC | Total | MCS | IDC | NC |
| NGP | 515 | 254 | 116 | 145 | 3.3 | 4.3 | 3.3 | 2.4 |
| SGP | 613 | 308 | 149 | 156 | 4.1 | 5.2 | 4.4 | 2.9 |
| SE | 1,156 | 526 | 303 | 327 | 4.5 | 5.2 | 5.3 | 3.3 |
| NE | 889 | 324 | 228 | 337 | 3.2 | 3.7 | 3.6 | 2.6 |

775

776
777 **Table A4.** Annual mean seasonal precipitation amounts and intensities for different types of precipitation in different regions of the US for 2004 – 2017

| | | Precipitation amount / mm | | | | Precipitation intensity / mm h$^{-1}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Total | MCS | IDC | NC | Total | MCS | IDC | NC |
| NGP | spring | 150 | 78 | 31 | 40 | 2.9 | 3.6 | 2.8 | 2.2 |
| | summer | 214 | 117 | 47 | 50 | 4.2 | 5.0 | 4.5 | 3.0 |
| | autumn | 109 | 43 | 27 | 39 | 2.9 | 3.9 | 3.1 | 2.3 |
| | winter | 42 | 15 | 11 | 15 | 1.9 | 2.4 | 1.9 | 1.7 |
| SGP | spring | 176 | 119 | 27 | 30 | 4.2 | 5.2 | 3.9 | 2.9 |
| | summer | 200 | 83 | 71 | 47 | 4.7 | 5.5 | 5.3 | 3.2 |
| | autumn | 150 | 62 | 36 | 52 | 4.1 | 5.3 | 4.6 | 3.0 |
| | winter | 87 | 44 | 16 | 27 | 2.8 | 3.6 | 2.6 | 2.2 |
| SE | spring | 275 | 157 | 52 | 66 | 4.6 | 5.3 | 4.8 | 3.3 |
| | summer | 367 | 112 | 156 | 99 | 5.2 | 5.7 | 6.1 | 3.7 |
| | autumn | 249 | 109 | 55 | 85 | 4.6 | 5.4 | 5.5 | 3.5 |
| | winter | 265 | 147 | 40 | 78 | 3.8 | 4.7 | 3.7 | 2.8 |
| NE | spring | 230 | 97 | 56 | 78 | 2.9 | 3.5 | 3.2 | 2.4 |
| | summer | 276 | 80 | 85 | 111 | 4.2 | 4.9 | 5.0 | 3.3 |
| | autumn | 218 | 75 | 49 | 94 | 3.2 | 3.8 | 3.6 | 2.6 |
| | winter | 165 | 72 | 39 | 55 | 2.4 | 2.9 | 2.4 | 2.1 |

778

779
780

**Table A5.** Annual and seasonal mean characteristics of MCS and IDC events in the data product domain for 2004 – 2017 by using the new MCS definition[1]

| | MCS | | | | | IDC | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | annual | spring | summer | autumn | winter | annual | spring | summer | autumn | winter |
| CCS-based lifetime / hour | 17.1 | 17.6 | 16.0 | 18.2 | 20.0 | 2.0 | 2.1 | 2.0 | 2.0 | 2.6 |
| CCS area / km$^2$ | 135,541 | 172,517 | 93,828 | 139,837 | 295,931 | 6,657 | 9,379 | 4,314 | 6,352 | 21,484 |
| CCS major axis length / km | 579 | 667 | 475 | 615 | 935 | 99 | 117 | 85 | 99 | 173 |
| PF-based lifetime / hour | 15.0 | 15.6 | 14.1 | 15.8 | 17.1 | 1.6 | 1.6 | 1.6 | 1.6 | 1.8 |
| Major axis length of the largest PF / km | 321 | 357 | 264 | 357 | 518 | 63 | 69 | 55 | 68 | 93 |
| PF convective area / km$^2$ | 6,119 | 6,468 | 6,091 | 5,897 | 5,697 | 477 | 496 | 463 | 487 | 520 |
| PF stratiform area / km$^2$ | 28,570 | 34,718 | 17,997 | 34,607 | 67,902 | 1,205 | 1,559 | 774 | 1,517 | 3,113 |
| PF mean convective rain rate / mm h$^{-1}$ | 4.5 | 4.0 | 4.8 | 4.6 | 3.9 | 4.1 | 3.4 | 4.5 | 4.3 | 3.0 |
| PF mean stratiform rain rate /mm h$^{-1}$ | 2.7 | 2.4 | 2.8 | 2.7 | 2.3 | 2.8 | 2.5 | 3.0 | 2.9 | 2.3 |
| PF mean convective 20-dBZ echo-top height / km | 6.6 | 6.2 | 7.2 | 6.1 | 5.0 | 6.5 | 6.1 | 7.0 | 6.2 | 5.0 |
| Area of the largest CCF / km$^2$ | 2,094 | 2,081 | 2,317 | 1,754 | 1,392 | 339 | 355 | 333 | 337 | 347 |
| Major axis length of the largest CCF / km | 95 | 96 | 99 | 88 | 82 | 29 | 30 | 28 | 29 | 30 |

[1] Refer to Section 4.4 for the new MCS definition.

781
782

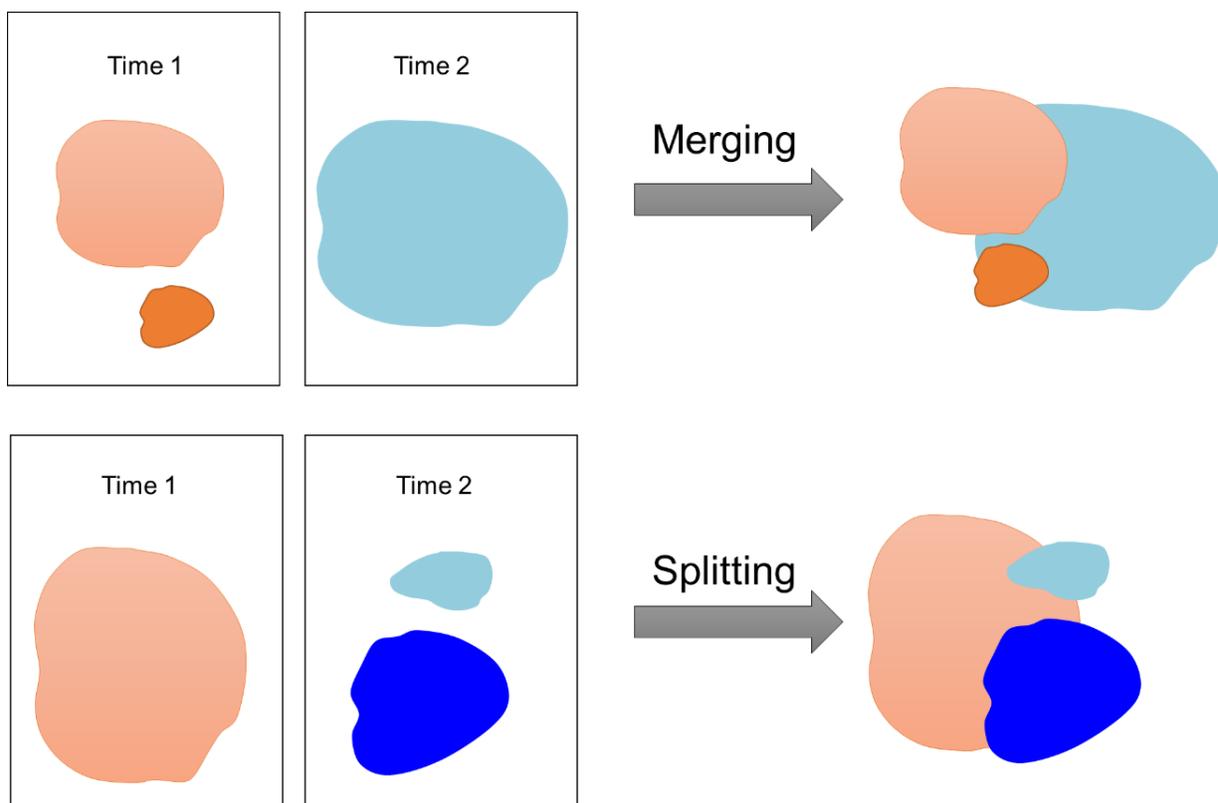
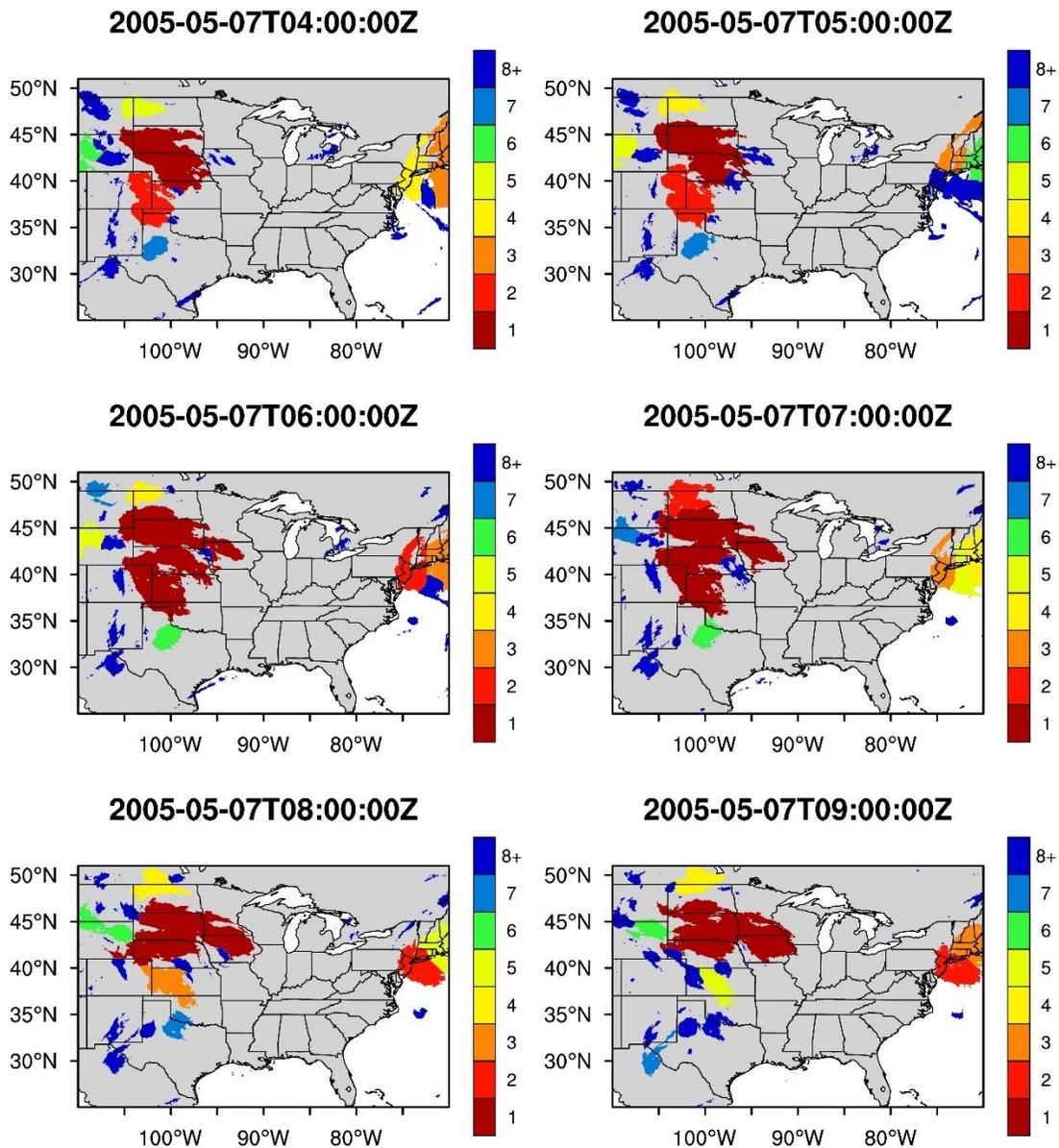
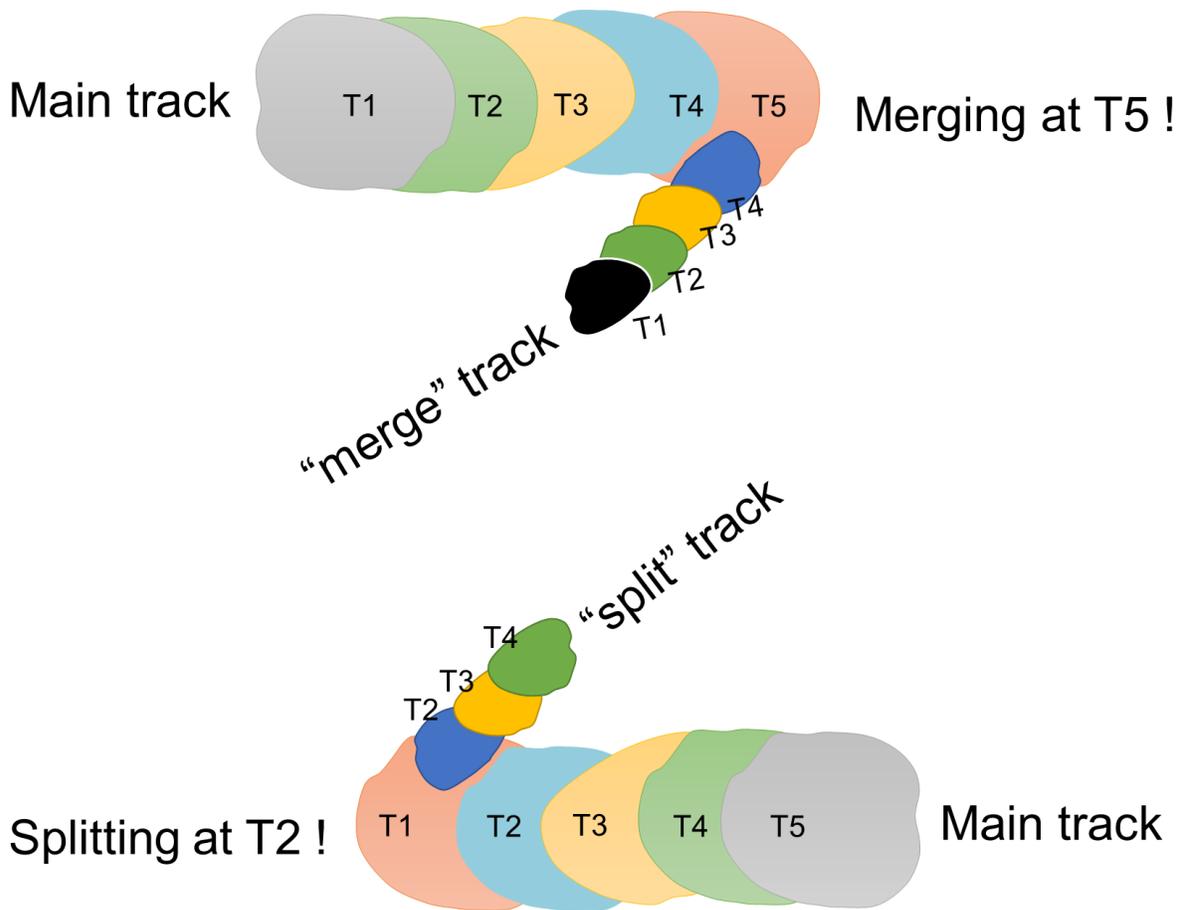
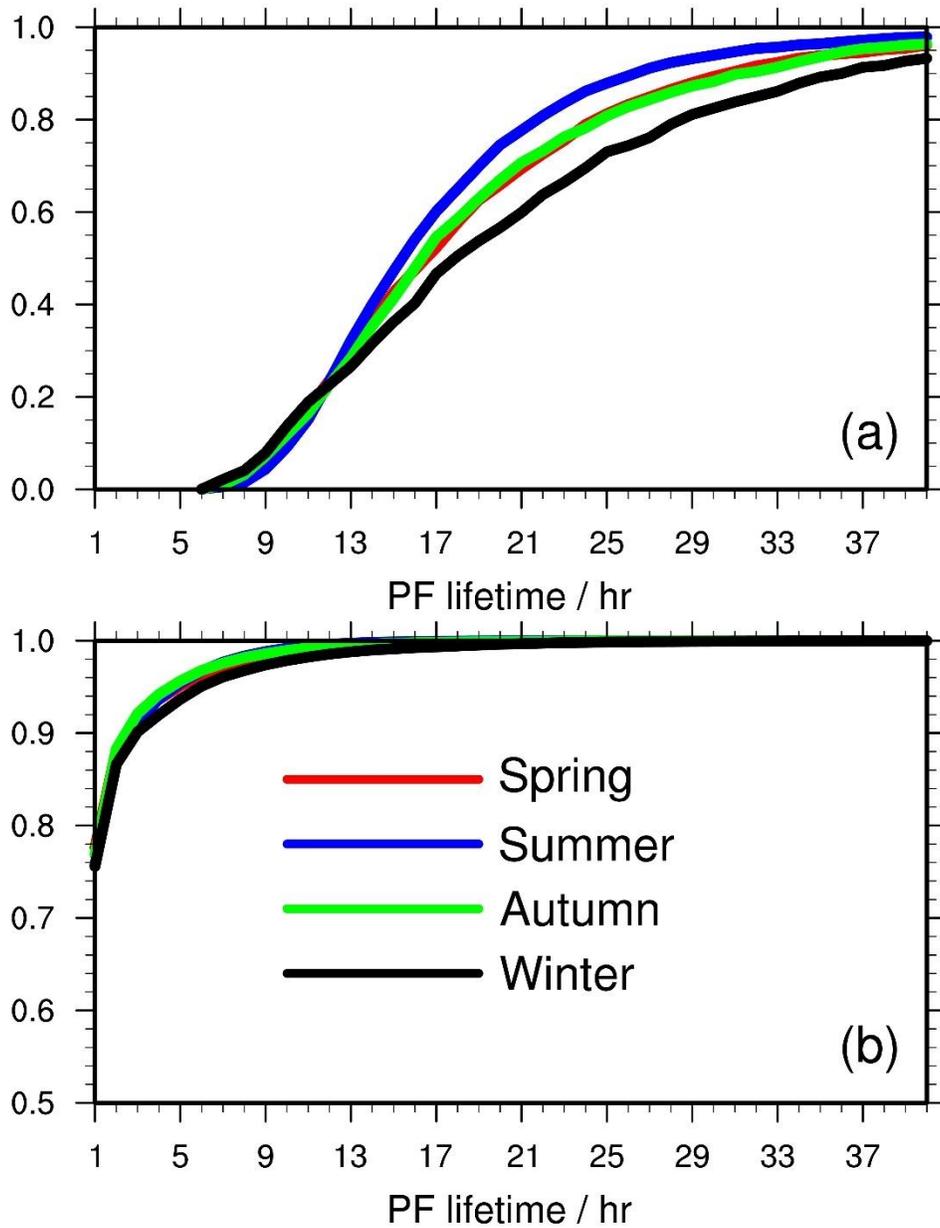
783
784 **Table A6.** Annual mean precipitation amounts and intensities for different types of precipitation in different regions of the US for 2004 – 2017 by using the new MCS definition

|  | Precipitation amount / mm | | | | Precipitation intensity / mm h$^{-1}$ | | | |
|---|---|---|---|---|---|---|---|---|
|  | Total | MCS | IDC | NC | Total | MCS | IDC | NC |
| NGP | 515 | 280 | 89 | 145 | 3.3 | 4.2 | 3.2 | 2.4 |
| SGP | 613 | 344 | 113 | 156 | 4.1 | 5.1 | 4.4 | 2.9 |
| SE | 1,156 | 602 | 227 | 327 | 4.5 | 5.3 | 5.3 | 3.3 |
| NE | 889 | 371 | 181 | 337 | 3.2 | 3.7 | 3.5 | 2.6 |

785

786 **Table A7.** Annual mean seasonal precipitation amounts and intensities for different types of
787 precipitation in different regions of the US for 2004 – 2017 by using the new MCS definition

| | | Precipitation amount / mm | | | | Precipitation intensity / mm h$^{-1}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Total | MCS | IDC | NC | Total | MCS | IDC | NC |
| NGP | spring | 150 | 83 | 26 | 41 | 2.9 | 3.5 | 2.8 | 2.2 |
| | summer | 214 | 130 | 34 | 50 | 4.2 | 5.0 | 4.5 | 3.0 |
| | autumn | 109 | 50 | 20 | 39 | 2.9 | 3.8 | 3.0 | 2.3 |
| | winter | 42 | 17 | 9 | 16 | 1.9 | 2.4 | 1.9 | 1.7 |
| SGP | spring | 176 | 126 | 20 | 30 | 4.2 | 5.0 | 3.9 | 2.9 |
| | summer | 200 | 102 | 51 | 47 | 4.7 | 5.5 | 5.2 | 3.2 |
| | autumn | 150 | 70 | 28 | 52 | 4.1 | 5.2 | 4.5 | 3.0 |
| | winter | 87 | 47 | 13 | 27 | 2.8 | 3.5 | 2.6 | 2.2 |
| SE | spring | 275 | 170 | 39 | 66 | 4.6 | 5.2 | 4.8 | 3.3 |
| | summer | 367 | 153 | 115 | 99 | 5.2 | 5.8 | 6.1 | 3.7 |
| | autumn | 249 | 122 | 42 | 85 | 4.6 | 5.4 | 5.5 | 3.5 |
| | winter | 265 | 156 | 31 | 78 | 3.8 | 4.6 | 3.7 | 2.8 |
| NE | spring | 230 | 108 | 44 | 78 | 2.9 | 3.5 | 3.1 | 2.4 |
| | summer | 276 | 99 | 66 | 111 | 4.2 | 4.9 | 5.0 | 3.3 |
| | autumn | 218 | 85 | 39 | 94 | 3.2 | 3.8 | 3.5 | 2.6 |
| | winter | 165 | 79 | 31 | 55 | 2.4 | 2.9 | 2.3 | 2.1 |

788

789
790   **Figure A1.** Schematic of CCS merging and splitting.

791
**Figure A2.** An example of CCS merging and splitting from 2005-05-07T4:00:00Z –
T9:00:00Z. Cloud 1 and Cloud 2 at 5:00:00Z merged into Cloud 1 at 6:00:00Z. And Cloud 1 at
7:00:00Z at least split to Cloud 1 and Cloud 3 at 8:00:00Z.
795

796
797    **Figure A3.** Schematic of "merge" tracks and "split" tracks.

798
**Figure A4.** Seasonal cumulative distribution functions of PF-based lifetimes for (a) MCSs and
800 (b) IDC in the data product domain for 2004 – 2017. Red lines denote spring, blue lines denote
801 summer, green lines denote autumn, and black lines denote winter.
802

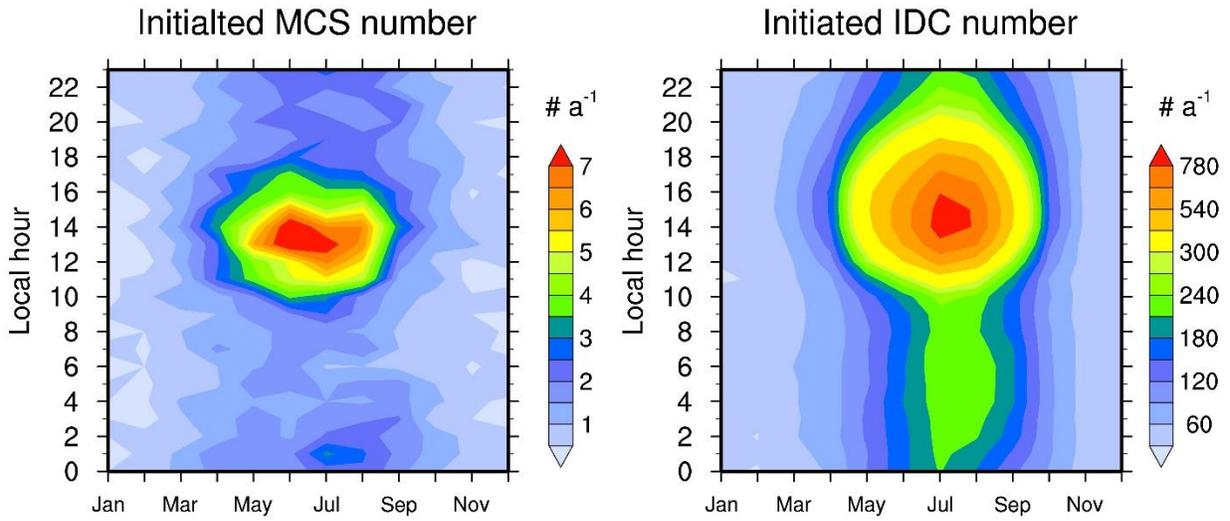



**Figure A5.** Annual mean monthly diurnal cycles of initiated MCS (left panel) and IDC (right panel) numbers in the data product domain for 2004 – 2017. Here, we define that an MCS or IDC event initiates when the first PF appears. Therefore, we can derive the initiated time of all MCS and IDC events, which is the basis of this figure. For example, on average, more than 7 MCSs initiated at 14:00 Local Time every June between 2004 and 2017.

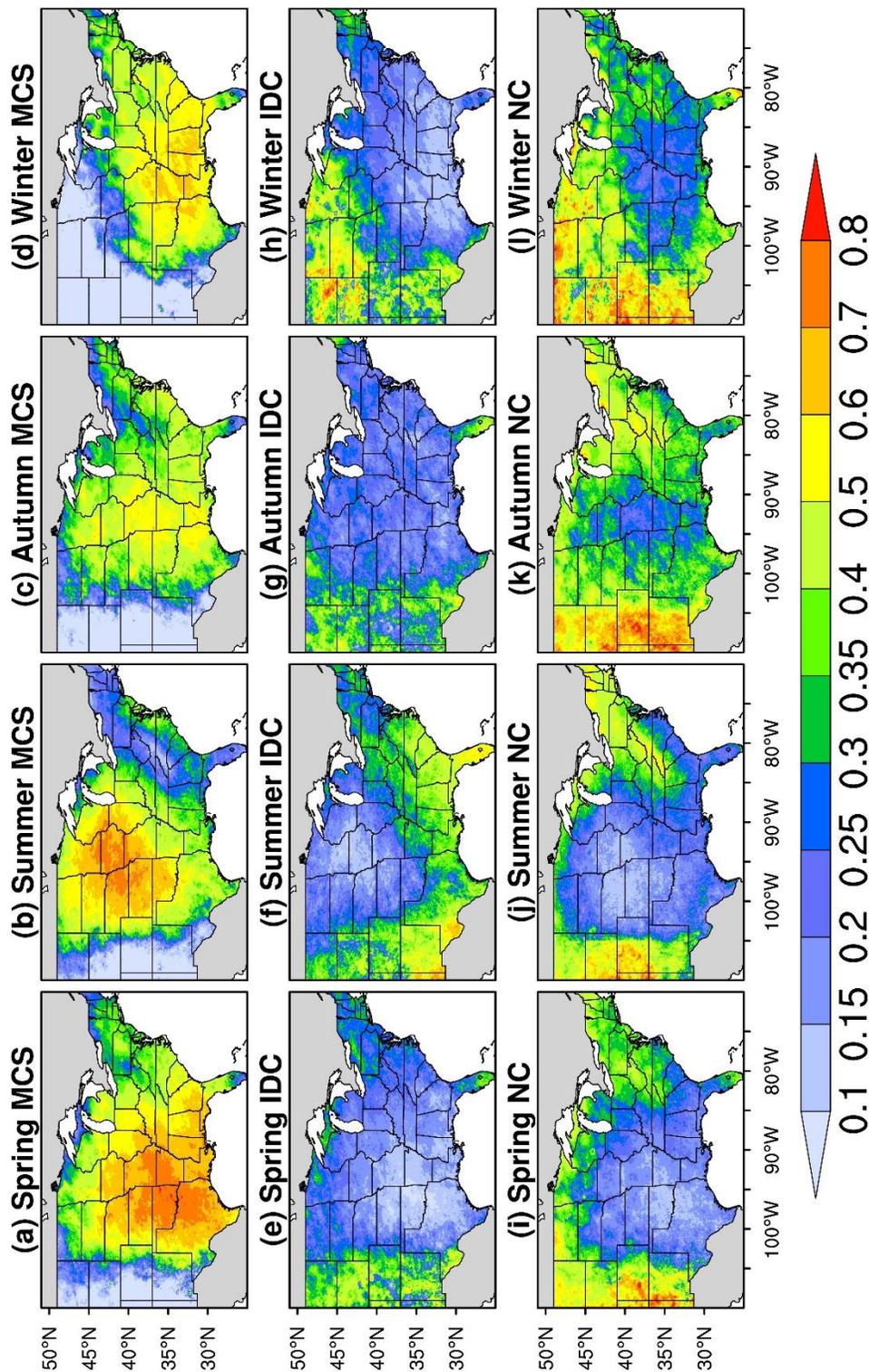



**Figure A6.** Distributions of the fractions of different types of precipitation in each season. Here, precipitation refers to annual mean seasonal amounts for 2004 – 2017. We exclude hourly data with precipitation ≤ 1 mm h⁻¹ in the calculation. The first row is for total precipitation, the second for MCS precipitation, the third for IDC precipitation, and the fourth for NC precipitation. The first column shows spring precipitation, the second for summer, the third for autumn, and the fourth for winter.

810
811
812
813
814
815

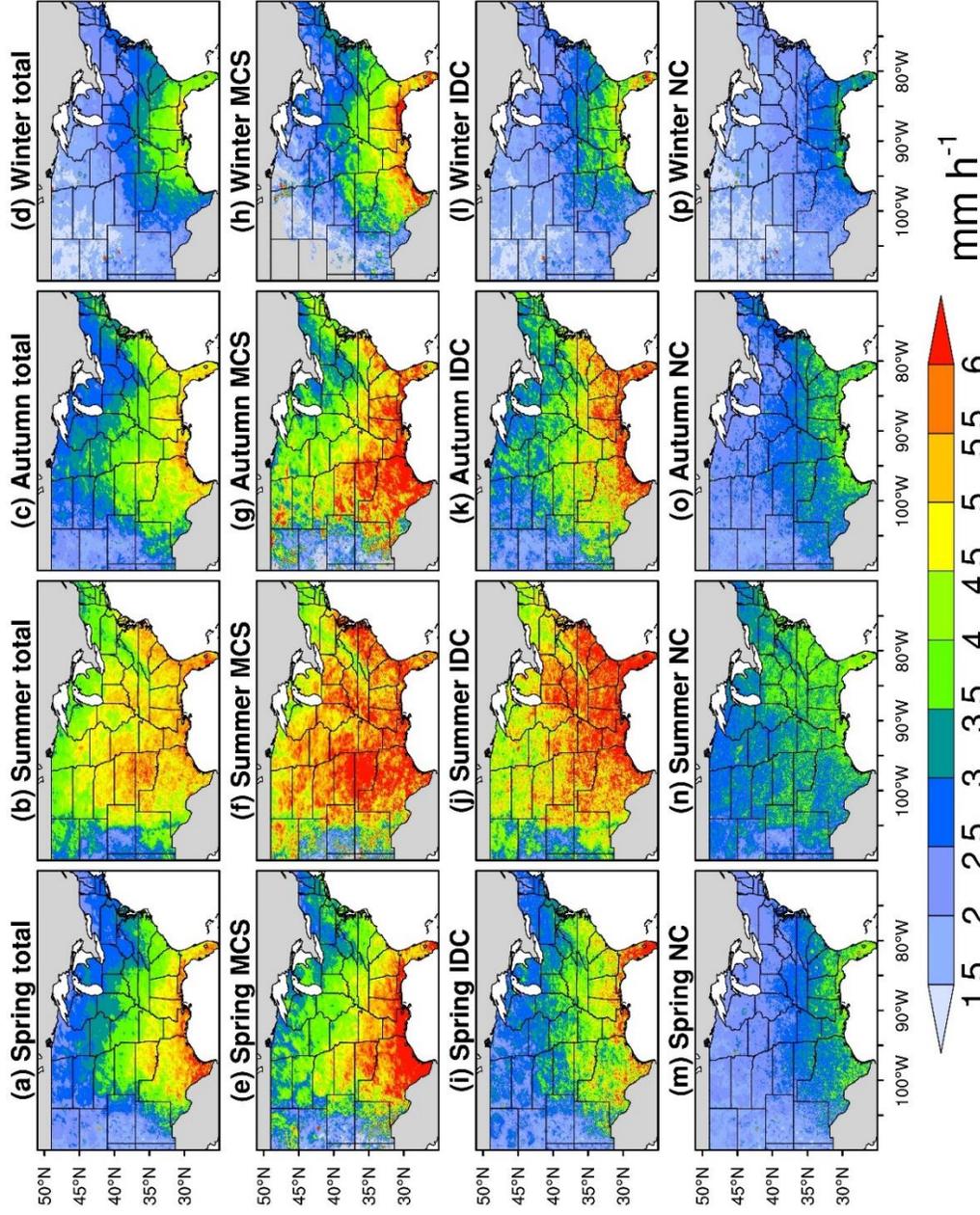
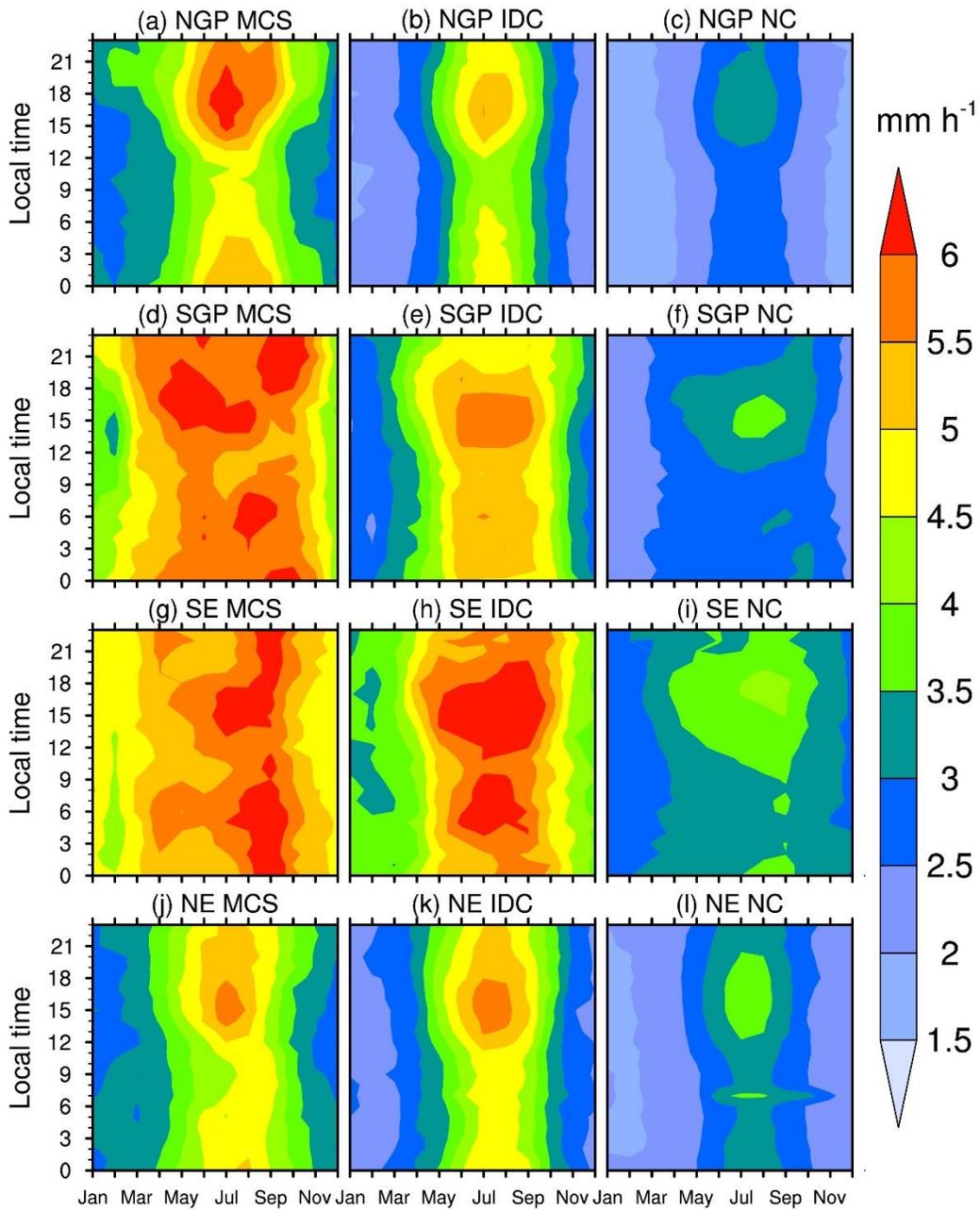
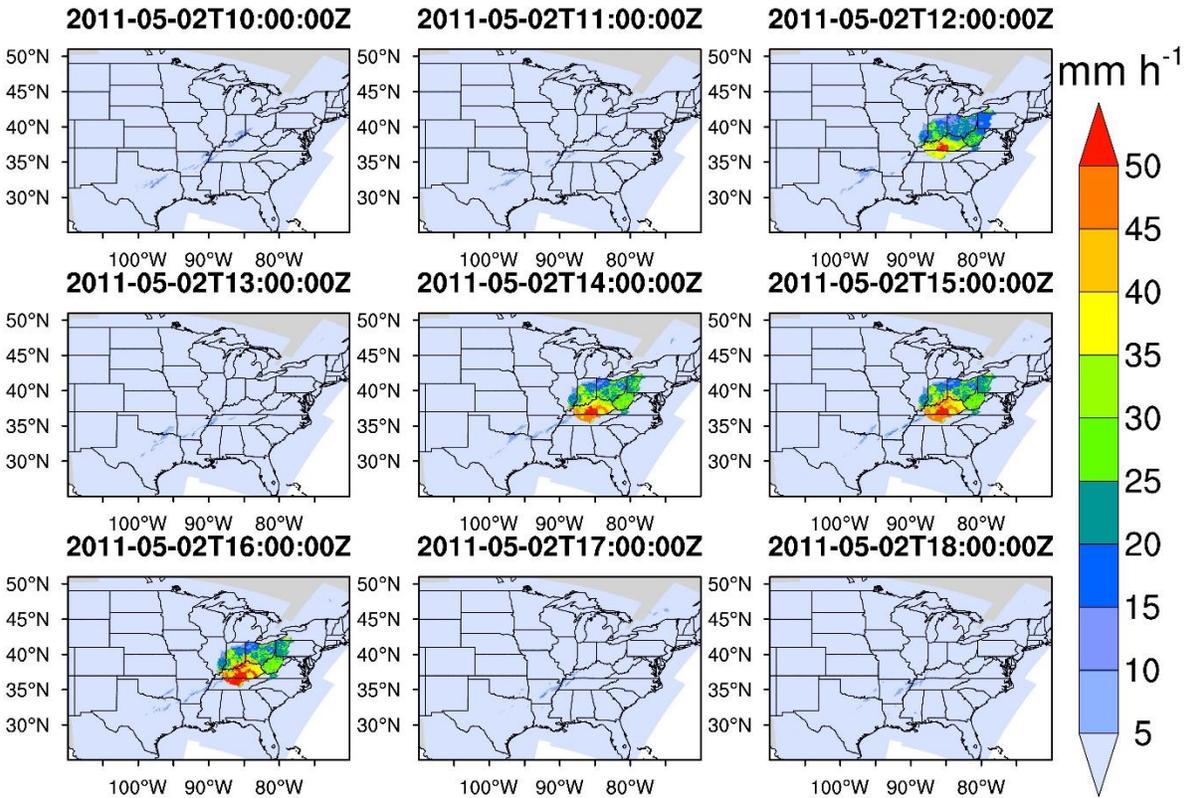
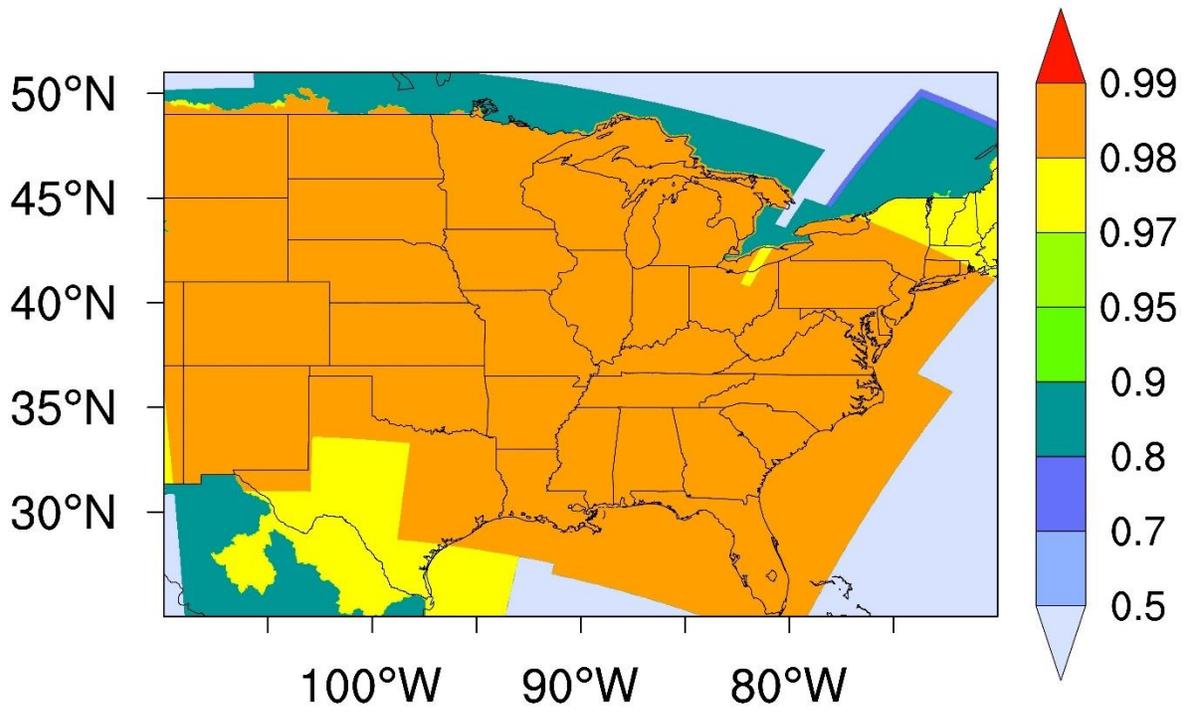
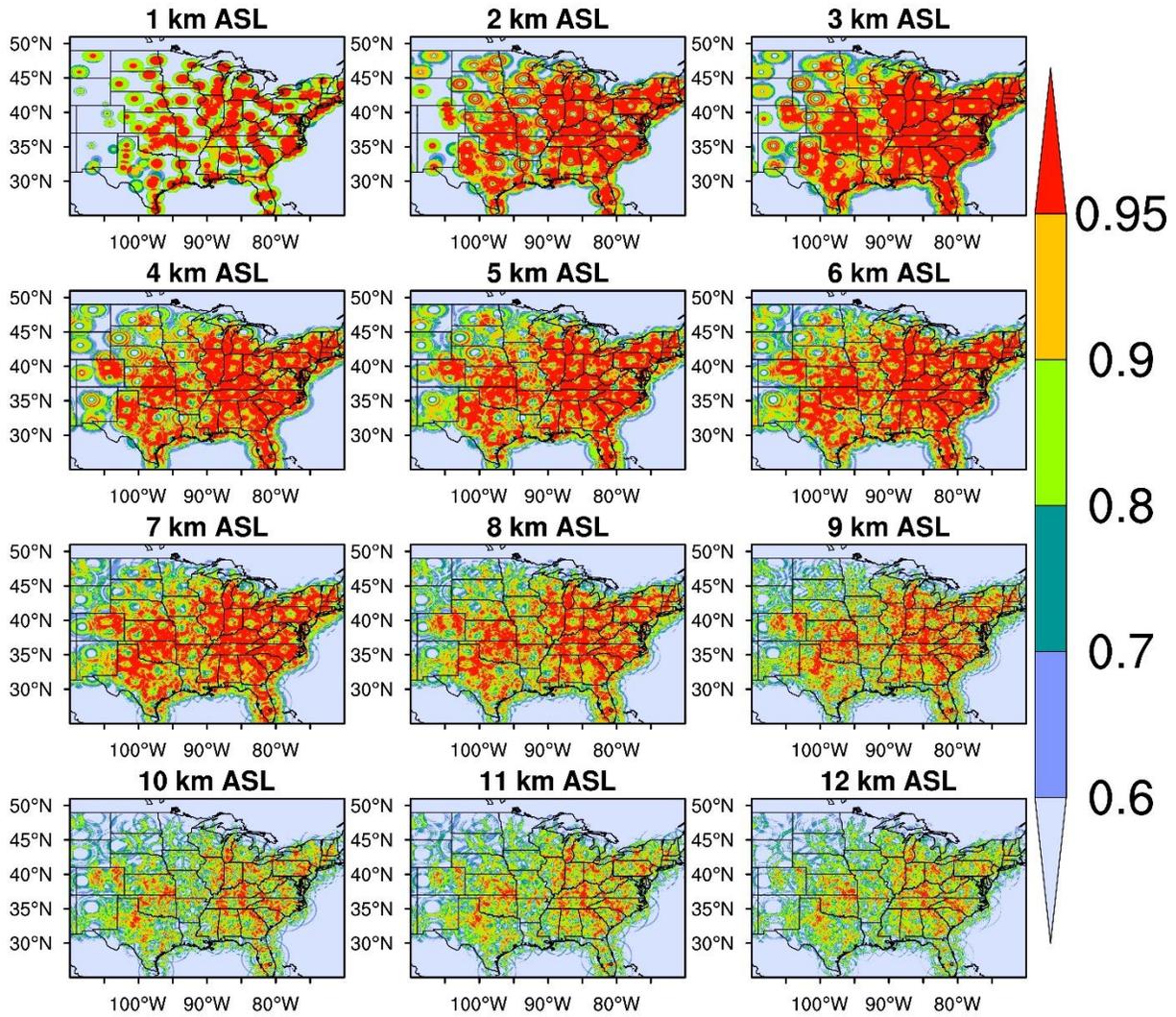



**Figure A7.** Distributions of annual mean seasonal precipitation intensities for different types of precipitation for 2004 – 2017. The first row is for total precipitation, the second for MCS precipitation, the third for IDC precipitation, and the fourth for NC precipitation. The first column shows spring precipitation, the second for summer, the third for autumn, and the fourth for winter. We exclude hourly data with precipitation ≤ 1 mm h$^{-1}$ in the calculation.

821
822 **Figure A8.** Monthly mean diurnal cycles of precipitation intensities for MCSs (a, d, g, j), IDC
823 (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)
824 during 2004 – 2017.
825

826
827 **Figure A9.** An example of Stage IV erroneous precipitation. Stage IV shows a large area of
828 intense precipitation suddenly appearing at 2011-05-02T12:00:00Z, which then unexpectedly
829 disappears at 13:00:00Z, comes back abruptly at 14:00:00Z, and finally goes away immediately
830 at 17:00:00Z.
831

832
833 **Figure A10.** Distribution of the fraction of valid Stage IV precipitation data for 2004 – 2017.
834 Here, "valid" means that precipitation data are available and reasonable. The erroneous
835 precipitation discussed in Section 4.1 is unreasonable and invalid.
836

837
838 **Figure A11.** Distributions of the fractions of available radar reflectivity data for 2004 – 2017 at
839 different vertical levels. As long as radars scan a grid cell, we think of it as "available" even
840 though there is no echo.

## Author contributions

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

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

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

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

reviewed the manuscript.

## Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgments

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

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

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

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

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

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

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

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

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

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

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

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

862    3D Gridrad dataset is from https://rda.ucar.edu/datasets/ds841.0/ (last access: Jan 2, 2020). We

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

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

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

866

# References

867 **References**

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

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

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

879 Bigelbach, B., Mullendore, G., and Starzec, M.: Differences in deep convective transport characteristics
880 between quasi‐isolated strong convection and mesoscale convective systems using seasonal WRF
881 simulations, J. Geophys. Res.-Atmos., 119, 11,445-411,455, https://doi.org/10.1002/2014JD021875,
882 2014.

883 Brooks, H. E., Doswell III, C. A., and Kay, M. P.: Climatological estimates of local daily tornado
884 probability for the United States, Weather and Forecasting, 18, 626-640, https://doi.org/10.1175/1520-
885 0434(2003)018<0626:CEOLDT>2.0.CO;2, 2003.

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

888 Changnon, S. A.: Damaging thunderstorm activity in the United States, Bulletin of the American
889 Meteorological Society, 82, 597-608, https://doi.org/10.1175/1520-
890 0477(2001)082<0597:DTAITU>2.3.CO;2, 2001a.

891 Changnon, S. A.: Thunderstorm rainfall in the conterminous United States, Bulletin of the American
892 Meteorological Society, 82, 1925-1940, https://doi.org/10.1175/1520-
893 0477(2001)082<1925:TRITCU>2.3.CO;2, 2001b.

894 Choi, S., Joiner, J., Choi, Y., Duncan, B., Vasilkov, A., Krotkov, N., and Bucsela, E.: First estimates of
895 global free-tropospheric $NO_2$ abundances derived using a cloud-slicing technique applied to satellite
896 observations from the Aura Ozone Monitoring Instrument (OMI), Atmos. Chem. Phys, 14, 10,565-
897 510,588, https://doi.org/10.5194/acp-14-10565-2014, 2014.

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

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

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

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

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

912     Doswell III, C. A., Brooks, H. E., and Maddox, R. A.: Flash flood forecasting: An ingredients-based
913     methodology, Weather and Forecasting, 11, 560-581, https://doi.org/10.1175/1520-
914     0434(1996)011<0560:FFFAIB>2.0.CO;2, 1996.

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

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

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

924     Feng, Z., Leung, L. R., Houze Jr, R. A., Hagos, S., Hardin, J., Yang, Q., Han, B., and Fan, J.: Structure
925     and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection‐
926     permitting simulations over the United States, Journal of Advances in Modeling Earth Systems, 10, 1470-
927     1494, https://doi.org/10.1029/2018MS001305, 2018.

928     Feng, Z., Houze Jr, R. A., Leung, L. R., Song, F., Hardin, J. C., Wang, J., Gustafson Jr, W. I., and
929     Homeyer, C. R.: Spatiotemporal characteristics and large-scale environments of mesoscale convective
930     systems east of the Rocky Mountains, J. Clim., 32, 7303-7328, https://doi.org/10.1175/JCLI-D-19-
931     0137.1, 2019.

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

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

937     Fritsch, J. M., Kane, R. J., and Chelius, C. R.: The contribution of mesoscale convective weather systems
938     to the warm-season precipitation in the United States, Journal of Applied Meteorology and Climatology,
939     25, 1333-1345, https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2, 1986.

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

942   Geerts, B.: Mesoscale convective systems in the southeast United States during 1994–95: A survey,
943   Weather and Forecasting, 13, 860-869, https://doi.org/10.1175/1520-
944   0434(1998)013<0860:MCSITS>2.0.CO;2, 1998.

945   Geerts, B., Parsons, D., Ziegler, C. L., Weckwerth, T. M., Biggerstaff, M. I., Clark, R. D., Coniglio, M.
946   C., Demoz, B. B., Ferrare, R. A., and Gallus Jr, W. A.: The 2015 plains elevated convection at night field
947   project, Bulletin of the American Meteorological Society, 98, 767-786, https://doi.org/10.1175/BAMS-D-
948   15-00257.1, 2017.

949   Giangrande, S. E., Krause, J. M., and Ryzhkov, A. V.: Automatic designation of the melting layer with a
950   polarimetric prototype of the WSR-88D radar, Journal of Applied Meteorology and Climatology, 47,
951   1354-1364, https://doi.org/10.1175/2007JAMC1634.1, 2008.

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

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

957   Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B., and Lawrimore, J. H.:
958   Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from
959   in situ observations, Journal of hydrometeorology, 5, 64-85, https://doi.org/10.1175/1525-
960   7541(2004)005<0064:CCOTHC>2.0.CO;2, 2004.

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

963   Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J. M., Nicolas, J., Radu, R., Schepers, D.,
964   Simmons, A., and Soci, C.: Global reanalysis: goodbye ERA-Interim, hello ERA5, in: ECMWF Newsl,
965   159, 17-24, https://doi.org/10.21957/vf291hehd7, 2019.

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

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

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

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

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

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

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

986 Kalinga, O. A., and Gan, T. Y.: Estimation of rainfall from infrared‐microwave satellite data for basin‐
987 scale hydrologic modelling, Hydrological processes, 24, 2068-2086, https://doi.org/10.1002/hyp.7626,
988 2010.

989 Koehler, T. L.: Cloud-to-Ground Lightning Flash Density and Thunderstorm Day Distributions over the
990 Contiguous United States Derived from NLDN Measurements: 1993–2018, Monthly Weather Review,
991 148, 313-332, https://doi.org/10.1175/MWR-D-19-0211.1, 2020.

992 Li, J., Feng, Z., Qian, Y., and Leung, L. R.: MCSs and IDC in the US for 2004–2017, available at
993 https://data.pnnl.gov/dataset/13218, http://dx.doi.org/10.25584/1632005, 2020 (last access: Jun 18,
994 2020).

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

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

1000 Lopez, P.: Direct 4D-Var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF,
1001 Monthly Weather Review, 139, 2098-2116, https://doi.org/10.1175/2010MWR3565.1, 2011.

1002 Machado, L., Rossow, W., Guedes, R., and Walker, A.: Life cycle variations of mesoscale convective
1003 systems over the Americas, Monthly Weather Review, 126, 1630-1654, https://doi.org/10.1175/1520-
1004 0493(1998)126<1630:LCVOMC>2.0.CO;2, 1998.

1005 Motew, M., Booth, E. G., Carpenter, S. R., Chen, X., and Kucharik, C. J.: The synergistic effect of
1006 manure supply and extreme precipitation on surface water quality, Environmental Research Letters, 13,
1007 044016, https://doi.org/10.1088/1748-9326/aaade6, 2018.

1008 Nearing, M., Pruski, F., and O'neal, M.: Expected climate change impacts on soil erosion rates: a review,
1009 Journal of soil and water conservation, 59, 43-50, 2004.

1010 Nelson, B. R., Prat, O. P., Seo, D.-J., and Habib, E.: Assessment and implications of NCEP Stage IV
1011 quantitative precipitation estimates for product intercomparisons, Weather and Forecasting, 31, 371-394,
1012 https://doi.org/10.1175/WAF-D-14-00112.1, 2016.

1013 Piani, C., Durran, D., Alexander, M., and Holton, J.: A numerical study of three-dimensional gravity
1014 waves triggered by deep tropical convection and their role in the dynamics of the QBO, Journal of the
1015 atmospheric sciences, 57, 3689-3702, https://doi.org/10.1175/1520-
1016 0469(2000)057<3689:ANSOTD>2.0.CO;2, 2000.

1017 Pinto, J. O., Grim, J. A., and Steiner, M.: Assessment of the High-Resolution Rapid Refresh model's
1018 ability to predict mesoscale convective systems using object-based evaluation, Weather and Forecasting,
1019 30, 892-913, https://doi.org/10.1175/WAF-D-14-00118.1, 2015.

1020 Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., and Clark, M.: Simulating
1021 North American mesoscale convective systems with a convection-permitting climate model, Climate
1022 Dynamics, 1-16, https://doi.org/10.1007/s00382-017-3993-2, 2017.

1023 Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., and Bloomfield, J.: Increased crop damage in the
1024 US from excess precipitation under climate change, Global Environ. Change, 12, 197-202,
1025 https://doi.org/10.1016/S0959-3780(02)00008-0, 2002.

1026 Rowe, A. K., Rutledge, S. A., and Lang, T. J.: Investigation of microphysical processes occurring in
1027 isolated convection during NAME, Monthly weather review, 139, 424-443,
1028 https://doi.org/10.1175/2010MWR3494.1, 2011.

1029 Rowe, A. K., Rutledge, S. A., and Lang, T. J.: Investigation of microphysical processes occurring in
1030 organized convection during NAME, Monthly weather review, 140, 2168-2187,
1031 https://doi.org/10.1175/MWR-D-11-00124.1, 2012.

1032 Seeley, J. T., and Romps, D. M.: The effect of global warming on severe thunderstorms in the United
1033 States, J. Clim., 28, 2443-2458, https://doi.org/10.1175/JCLI-D-14-00382.1, 2015.

1034 Sieglaff, J. M., Hartung, D. C., Feltz, W. F., Cronce, L. M., and Lakshmanan, V.: A satellite-based
1035 convective cloud object tracking and multipurpose data fusion tool with application to developing
1036 convection, Journal of Atmospheric and Oceanic Technology, 30, 510-525,
1037 https://doi.org/10.1175/JTECH-D-12-00114.1, 2013.

1038 Smalley, M., L'Ecuyer, T., Lebsock, M., and Haynes, J.: A comparison of precipitation occurrence from
1039 the NCEP Stage IV QPE product and the CloudSat Cloud Profiling Radar, Journal of hydrometeorology,
1040 15, 444-458, https://doi.org/10.1175/JHM-D-13-048.1, 2014.

1041 Song, F., Feng, Z., Leung, L. R., Houze Jr, R. A., Wang, J., Hardin, J., and Homeyer, C. R.: Contrasting
1042 spring and summer large-scale environments associated with mesoscale convective systems over the US
1043 Great Plains, J. Clim., 32, 6749-6767, https://doi.org/10.1175/JCLI-D-18-0839.1, 2019.

1044 Starzec, M., Homeyer, C. R., and Mullendore, G. L.: Storm labeling in three dimensions (SL3D): A
1045 volumetric radar echo and dual-polarization updraft classification algorithm, Monthly Weather Review,
1046 145, 1127-1145, https://doi.org/10.1175/MWR-D-16-0089.1, 2017.

1047 Steiner, M., Houze Jr, R. A., and Yuter, S. E.: Climatological characterization of three-dimensional storm
1048 structure from operational radar and rain gauge data, Journal of Applied Meteorology, 34, 1978-2007,
1049 https://doi.org/10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2, 1995.

1050 Stensrud, D. J.: Effects of persistent, midlatitude mesoscale regions of convection on the large-scale
1051 environment during the warm season, Journal of the atmospheric sciences, 53, 3503-3527,
1052 https://doi.org/10.1175/1520-0469(1996)053<3503:EOPMMR>2.0.CO;2, 1996.

1053 Stensrud, D. J.: Upscale effects of deep convection during the North American monsoon, Journal of the
1054 atmospheric sciences, 70, 2681-2695, https://doi.org/10.1175/JAS-D-13-063.1, 2013.

1055 Taszarek, M., Allen, J. T., Púčik, T., Hoogewind, K. A., and Brooks, H. E.: Severe convective storms
1056 across Europe and the United States. Part II: ERA5 environments associated with lightning, large hail,
1057 severe wind, and tornadoes, J. Clim., 33, 10263-10286, https://doi.org/10.1175/JCLI-D-20-0346.1, 2020.

1058 Thompson, A. M., Tao, W.-K., Pickering, K. E., Scala, J. R., and Simpson, J.: Tropical deep convection
1059 and ozone formation, Bulletin of the American Meteorological Society, 78, 1043-1054,
1060 https://doi.org/10.1175/1520-0477(1997)078<1043:TDCAOF>2.0.CO;2, 1997.

1061 Tian, B., Held, I. M., Lau, N. C., and Soden, B. J.: Diurnal cycle of summertime deep convection over
1062 North America: A satellite perspective, J. Geophys. Res.-Atmos., 110,
1063 https://doi.org/10.1029/2004JD005275, 2005.

1064 Tippett, M. K., Allen, J. T., Gensini, V. A., and Brooks, H. E.: Climate and hazardous convective
1065 weather, Current Climate Change Reports, 1, 60-73, https://doi.org/10.1007/s40641-015-0006-6, 2015.

1066 Twohy, C. H., Clement, C. F., Gandrud, B. W., Weinheimer, A. J., Campos, T. L., Baumgardner, D.,
1067 Brune, W. H., Faloona, I., Sachse, G. W., and Vay, S. A.: Deep convection as a source of new particles in
1068 the midlatitude upper troposphere, J. Geophys. Res.-Atmos., 107, AAC 6-1-AAC 6-10,
1069 https://doi.org/10.1029/2001JD000323, 2002.

1070 Walker, J. R., MacKenzie Jr, W. M., Mecikalski, J. R., and Jewett, C. P.: An enhanced geostationary
1071 satellite–based convective initiation algorithm for 0–2-h nowcasting with object tracking, Journal of
1072 Applied Meteorology and Climatology, 51, 1931-1949, https://doi.org/10.1175/JAMC-D-11-0246.1,
1073 2012.

1074 Wang, P. K.: Moisture plumes above thunderstorm anvils and their contributions to cross‐tropopause
1075 transport of water vapor in midlatitudes, J. Geophys. Res.-Atmos., 108,
1076 https://doi.org/10.1029/2002JD002581, 2003.

1077 Yang, Q., Houze Jr, R. A., Leung, L. R., and Feng, Z.: Environments of long‐lived mesoscale
1078 convective systems over the central United States in convection permitting climate simulations, J.
1079 Geophys. Res.-Atmos., 122, 13,288-213,307, https://doi.org/10.1002/2017JD027033, 2017.

1080 Yuan, H., McGinley, J. A., Schultz, P. J., Anderson, C. J., and Lu, C.: Short-range precipitation forecasts
1081 from time-lagged multimodel ensembles during the HMT-West-2006 campaign, Journal of
1082 Hydrometeorology, 9, 477-491, https://doi.org/10.1175/2007JHM879.1, 2008.

1083 Zhang, K., Randel, W. J., and Fu, R.: Relationships between outgoing longwave radiation and diabatic
1084 heating in reanalyses, Climate Dynamics, 49, 2911-2929, https://doi.org/10.1007/s00382-016-3501-0,
1085 2017.

1086