

1
2
3
4 **Explicit Prediction of Hail in a Long-lasting Multi-cellular Convective System in Eastern**
5 **China Using Multi-moment Microphysics Schemes**

6
7 Liping Luo^{1,2}, Ming Xue^{1,2}, Kefeng Zhu¹ and Bowen Zhou¹

8
9 ¹ *Key Laboratory of Mesoscale Severe Weather/Ministry of Education and School of*
10 *Atmospheric Sciences, Nanjing University, Nanjing, 210023, China*

11 ² *Center for Analysis and Prediction of Storms and School of Meteorology,*
12 *University of Oklahoma, Norman Oklahoma, USA 73072*

13
14
15
16
17
18 Submitted to Journal of Atmospheric Sciences

19 October 2017

20 Revised February 2018

21
22
23 *Corresponding author address*

24
25 Ming Xue

26 Key Laboratory of Mesoscale Severe Weather/Ministry of Education
27 School of Atmospheric Sciences, Nanjing University, Nanjing, China, 210023, and

28 Center for Analysis and Prediction of Storms,

29 University of Oklahoma

30 Email: mxue@ou.edu

Abstract

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

During the afternoon of April 28, 2015, a multi-cellular convective system swept southward through much of Jiangsu Province, China over about seven hours, producing egg-sized hailstones on the ground. The hailstorm event is simulated using the Advanced Regional Prediction System (ARPS) at 1-km grid spacing. Different microphysics schemes are used predicting one, two, and three moments of the hydrometeor particle size distributions (PSDs). Simulated reflectivity and maximum estimated size of hail (MESH) derived from the simulations are verified against reflectivity observed by operational S-band Doppler radars and radar-derived MESH, respectively. Comparisons suggest that the general evolution of the hailstorm is better predicted by the three-moment scheme, and neighborhood-based MESH evaluation further confirms the advantage of three-moment scheme in hail size prediction.

Surface accumulated hail mass, number and hail distribution characteristics within simulated storms are examined across sensitivity experiments. Results suggest that multi-moment schemes produce more realistic hail distribution characteristics, with the three-moment scheme performing the best. Size-sorting is found to play a significant role in determining hail distribution within the storms. Detailed microphysical budget analyses are conducted for each experiment, and results indicate that the differences in hail growth processes among the experiments can be mainly ascribed to the different treatments of the shape parameter within different microphysics schemes. Both the differences in size sorting and hail growth processes contribute to the simulated hail distribution differences within storms and at the surface.

51 1. Introduction

52 Hailstorms are among the costliest natural disasters in China and many other countries;
53 hailstorms can cause severe injuries and extensive property damage. According to the Yearbooks of
54 Meteorological Disasters in China (e.g., 2013, 2014 and 2015), hail damages amount to billions of
55 U.S. dollars annually in China. Improving the prediction of hail, including the size and number of
56 hailstones, and the spatial and temporal coverage of hailfall, can help mitigate the impacts of
57 hailstorms through improved warnings. However, the prediction of hailstorms using operational
58 numerical weather prediction (NWP) models remains a challenge. The explicit prediction of hail at
59 the surface, including the spatial and temporal coverage of hailfall and the hail size distributions, is
60 even more challenging because of the complex microphysical as well as dynamic and
61 thermodynamic processes involved in hail production (Snook et al. 2016, Labriola et al. 2017).

62 Our general ability to forecast hail in operational and research settings is still limited (Moore
63 and Pino 1990; Brimelow et al. 2002; Guo and Huang 2002; Milbrandt and Yau 2006a, 2006b,
64 hereafter MY06a, MY06b; Brimelow and Reuter 2009; Luo et al. 2017, hereafter L17; Labriola et
65 al. 2017). Existing hail forecast methods include the following four types: i) hail diagnostics based
66 on observed soundings, ii) methods using a simple cloud model combined with a hail growth model
67 (e.g., HAILCAST) (Brimelow et al. 2002), iii) statistical and machine learning (ML) hail forecast
68 methods (e.g., random forests, gradient boosting trees, and linear regression), and iv) predictions
69 using convective-scale NWP models with sophisticated microphysics schemes.

70 For sounding-based hail diagnostic methods, the most important limitation is the lack of

71 timely soundings. HAILCAST addresses this issue by feeding prognostic model soundings into a
72 time-dependent hail growth model (Brimelow et al. 2002; Adams-Selin and Ziegler 2016). Gagne et
73 al. (2014) statistically validated the hail size and probability forecast skills of ML techniques and
74 HAILCAST, based on 12 hail days from May to June 2014 over the United States, and found that
75 the ML techniques produced smaller size errors compared to HAILCAST, and that both approaches,
76 especially HAILCAST, tended to overpredict the maximum hail size. ML methods also showed
77 temporal and spatial offsets with observed hailstorms (Gagne et al. 2014).

78 Since sophisticated, multi-moment microphysics parameterization (hereafter MP) schemes
79 in storm-scale NWP models are capable of predicting hydrometeor size distributions within realistic
80 environments, efforts to simulate and predict real hailstorms have been attempted using NWP
81 models in recent years (e.g., MY06a, b; Noppel et al. 2010; Snook et al. 2016; L17). For example,
82 MY06a conducted simulations of a supercell hailstorm that occurred in Canada, using the three-
83 moment Milbrandt and Yau (hereafter MY) scheme in a mesoscale NWP model. Comparisons with
84 radar observations indicated that the typical supercell structures such as the hook echo,
85 mesocyclone, and suspended overhang region were well reproduced, although the simulated
86 maximum hail size on the ground was underpredicted. Snook et al. (2016) is a more recent example,
87 which evaluated short-term ensemble forecasting of hail for a supercell storm over central
88 Oklahoma using a two-moment MP scheme. They noted that hail prediction might be improved by
89 using more advanced MP schemes, especially via better explicit prediction of the properties of
90 rimed ice.

91 In the bulk MP schemes typically used in NWP models, the PSD of each hydrometeor
92 category is assumed to have a Marshall-Palmer or Gamma distribution, and normally one, two, or
93 three moments of the distribution are predicted for each hydrometeor category. One-moment
94 schemes typically predict only the mixing ratios (Q) of various hydrometeors (e.g., Lin et al. 1983;
95 Kessler 1995). Two-moment schemes predict the mixing ratios (Q) and total number concentrations
96 (Nt) of all or some of the hydrometeors (e.g., Ferrier 1994; Walko et al. 1995; Meyers et al. 1997;
97 Thompson et al. 2004; Milbrandt and Yau 2005a, hereafter MY05a; Morrison et al. 2005; Morrison
98 and Gettelman 2008; Thompson et al. 2008). To make the Gamma distribution (involving three free
99 parameters) fully prognostic, Milbrandt and Yau (2005b) (hereafter MY05b) proposed a three-
100 moment scheme by adding a predictive equation for radar reflectivity factor (Z) of the hydrometeors
101 [related to the sixth moment] to their two-moment scheme.

102 However, many studies have noted that different bulk MP schemes often produce large
103 differences in various aspects of the simulated storms, including the storm structure, surface
104 accumulated precipitation, and cold pool (e.g., Gilmore et al. 2004; MY06b; Seifert et al. 2006;
105 Morrison et al. 2009; Dawson et al. 2010; Jung et al. 2010; Van Weverberg et al. 2012; Loftus and
106 Cotton 2014b; L17). Most of the prior studies have used idealized frameworks, in which the storm
107 environment is horizontally homogeneous, with MY06b, Snook et al. (2016), and L17 being the
108 exceptions. MY06b performed sensitivity experiments of a supercell hailstorm in Canada using
109 different MP schemes. They noted dramatic improvements in the storm structure and the predicted
110 precipitation when switching from a one-moment to a two-moment scheme. More recently, L17

111 investigated hail forecast skill using various MP schemes for a pulse-type hailstorm in eastern
112 China. They compared simulated total precipitation and maximum estimated size of hail (MESH)
113 (Witt et al. 1998) swaths against observations and found that the three-moment MY scheme
114 produced the best forecast.

115 Hail damage is not simply a function of maximum hail size; cumulative hail mass and
116 number concentration are also important factors (Changnon 1999; Gilmore et al. 2004). Therefore,
117 hail size distribution, cumulative mass, and number concentration predicted by various MP schemes
118 are also worth evaluating in addition to MESH. Moreover, to more robustly evaluate and document
119 the performances and behaviors of different MP schemes, and at the same time to achieve a better
120 understanding of hail production and growth processes, more studies using real data for diverse
121 types of hailstorms that may occur in different storm environments are still needed. Among the
122 existing real case studies that attempt explicit hail prediction to different degrees, MY06b and
123 Snook et al. (2016) dealt with supercell storms while L17 dealt with a pulse-type storm. To our
124 knowledge, there has not been a real-case study that focuses on the multi-cellular type of hailstorm
125 that may organize into mesoscale convective systems (MCSs).

126 In this study, explicit hail prediction of a long-lasting multi-cellular hailstorm event that
127 occurred on April 28, 2015 in eastern China is investigated. On that day, the multi-cellular
128 hailstorms swept southward through most of Jiangsu Province, China over about seven hours,
129 producing egg-sized hailstones with diameters of around 20-50 mm on the ground. As in L17, this
130 study employs the ARPS model (Xue et al. 2000, 2001, 2003). Simulation experiments are run at a

131 1 km horizontal grid spacing (instead of 3 km in L17) using one, two, and three-moment MY MP
132 schemes. The goals of this study are two-fold. First, the explicit hail forecast skill with different MP
133 schemes, in terms of surface hailstone size distribution, cumulative hail mass, and number
134 concentration, are evaluated. Secondly, combined with diagnostic analyses on microphysical terms,
135 the reasons for the hail forecast differences between various MP schemes are explored. Surface
136 accumulated precipitation, MESH, cumulative hail mass and number concentration are examined
137 and objective MESH evaluation is performed using the fractions skill score (FSS) “neighborhood”
138 technique (Ebert 2009).

139 The rest of this paper is organized as follows. In section 2, an overview of the 28 April 2015
140 multi-cellular hailstorm event is given. The ARPS model setup and hail forecast evaluation metrics
141 are described in sections 3 and 4, respectively. Section 5 compares the explicit hail predictions
142 using different MP schemes, and investigates the causes of such differences. Finally, a summary
143 and conclusions are presented in section 6.

144 **2. Case overview**

145 Hailstorms and other forms of severe weather events in the northeastern Asian Pacific
146 coastal regions including eastern and northeastern China are often associated with upper
147 troposphere cut-off lows (COLs) (Tao et al. 1980; Nieto et al. 2005, 2008; Zhang et al. 2008). Most
148 COLs occur over the northeastern part of China and often move southeastward to the coastal region
149 of Eastern China (Hu et al. 2010). The COLs can persist for several days and produce high
150 convective instability. The hailstorm studied here occurred during this type of synoptic situation,

151 where an upper-level COL swept from north to south across Jiangsu Province in eastern China,
152 producing severe hail. According to the Severe Weather Reports of the Chinese Meteorological
153 Administration (Fig. 1), parts of Shandong Province, north of Jiangsu, were first hit by intense small
154 size hailstones (<10 mm) in the morning (around 0900 LST). In the afternoon, multi-cellular
155 hailstorms formed along the west border of Jiangsu Province and produced a large number of egg-
156 sized hailstones on the ground with the maximum observed size being ~10 cm in Yizheng City,
157 Jiangsu Province. The hailfall from these storms extended through the western part of Jiangsu
158 Province from north to south, and hail fall was continuous for as long as seven hours from around
159 0700 through 1400 UTC. In addition, intense lightning and damaging surface winds ($\sim 23 \text{ m s}^{-1}$)
160 were also reported.

161 The multi-cellular hailstorms were observed by multiple operational S-band Doppler radars,
162 including those at Jinan, Xuzhou, Bengbu, Yancheng, Nanjing, and Nantong (see Fig. 1). Figure 2
163 presents the composite (column maximum) reflectivity from these radars. The lifespan of this long-
164 lasting multi-cellular hailstorm system can be characterized by two episodes. In the first episode, a
165 series of multi-cellular hailstorms was initiated and the storms intensified along the northwest
166 border of Jiangsu Province. The storms organized into a northwest-southeast line and moved
167 southeastward (Figs. 2a, b). In the second episode, from 1100 UTC onwards, as the hailstorms
168 moved southeastward the line gradually evolved into a bow-shaped echo, with the middle portion of
169 the line bulging eastward (Figs. 2c-f). By 1400 UTC, the apex of the bow had almost reached the
170 southeast corner of Jiangsu, while the southern portion of the bow was oriented from east to west

171 near the southern border of Jiangsu, with its western tail extending well into Anhui Province (Fig.
172 2f). The system started to weaken 1400 UTC. At all times shown, observed reflectivity exceeds 60
173 dBZ within some of the storm cells (Fig. 2).

174 The synoptic patterns associated with this event are shown in Fig. 3. At 0600 UTC (1400
175 LST) 28 April, the northeastern coastal regions of China were beneath a deep, positively-tilted,
176 semi-permanent upper-level East Asian trough (EAT). The East-Asia upper-tropospheric jet stream
177 (EAJS) was located at the southern periphery of the trough (Fig. 3a). The jet core over land was
178 located at $\sim 30^{\circ}\text{N}$ and 120°E , with a maximum wind exceeding 50 m s^{-1} (Fig. 3a). Jiangsu Province
179 (solid red line in Fig. 3a) was located ahead of the EAT and underneath the front-left (more so at
180 earlier times) exit region of the EAJS, where favorable positive vorticity advection from the EAT
181 and the upper-level divergence near the front-left EAJS exist region acted together to destabilize the
182 atmosphere. Moreover, as seen in Figs. 3b-d, from middle to low altitude, two COLs (denoted 'C')
183 were embedded within the EAT, with one over the eastern coast of China and the other over the
184 East China Sea.

185 Strong cold advection is found southeast of the western COL at the 500 hPa (Fig. 3b)
186 directly over Jiangsu Province. At the 850 hPa (Fig. 3b) and at the surface (Fig. 3d), a prominent
187 convergence line is present between the two cyclonic circulations; this convergence line is also the
188 convergence boundary between the warm (temperature $> 28^{\circ}\text{C}$) unstable (CAPE reaching 1500 J kg^{-1})
189 air mass from the southwest that is partly associated with the southern part of the western
190 cyclonic circulation, and the much colder air ($< 16^{\circ}\text{C}$) with no CAPE from the northeast that is part

191 of the eastern cyclone. The two low-level cyclones are responsible for setting up a strong
192 convergence zone and moderately high CAPE in the air south of the convergence line. The
193 convergence forcing, coupled with the destabilizing upper-level circulations, creates an environment
194 favorable for intense deep convection.

195 Because of the lack of observed soundings near the time of hailstorm initiation, a sounding
196 is extracted at 0600 UTC from the NCEP GFS $1^\circ \times 1^\circ$ Final Analysis (FNL) at (34°N, 117°E), which
197 is about 30 km southwest of Xuzhou (see Fig. 1). The hailstorm firstly initiates near Xuzhou
198 approximately one hour later at around 0700 UTC. The sounding (Fig. 4) has a CAPE of 1433 J kg^{-1}
199 and a convective inhibition of -5 J kg^{-1} . The situation is characterized by strong vertical wind shear,
200 with southeasterly winds below 850 hPa, and southwesterly to northwesterly winds within the
201 layers above. The bulk Richardson number is 31.1 and the 0-6 km vertical wind shear is $\sim 24.5 \text{ m s}^{-1}$;
202 these values are generally considered conducive for long-lasting severe convection (Weisman and
203 Klemp 1984). A capping inversion is present between 850 and 800 hPa; this inversion is sufficiently
204 weak that it can be overcome by strong low-level convergence while convection is generally
205 suppressed elsewhere, a situation favoring concentrated, intense deep convection. Above the
206 inversion, the air mass is dry and cold, which may be a result of the previously mentioned cold
207 advection. Many previous studies (e.g., Costa et al. 2001; L17) have noted that a dry mid-level layer
208 over a warm moist layer near the surface is favorable for larger hailstones reaching the surface due
209 to reduced melting of hail under such conditions. We note that there are uncertainties involved with
210 extracting a sounding from the FNL analysis dataset, however this sounding represents the best

211 available source of information about local environmental conditions. Overall, the sounding
212 indicates a conducive environment for deep convection, and a high likelihood for the production of
213 large hailstones.

214 3. Experiment setup

215 The Jiangsu hailstorm is simulated using the ARPS model (Xue et al. 2000, 2001, 2003).
216 ARPS is a three-dimensional, non-hydrostatic compressible model using generalized terrain-
217 following coordinates and was designed for regional to storm-scale atmospheric modeling and
218 prediction. All simulations are initialized at 0000 UTC on 28 April 2015 and are run for 16 hours.
219 The initial condition and boundary conditions at six-hour intervals are obtained from the NCEP
220 Final Analysis data at $1^\circ \times 1^\circ$ resolution.

221 Two one-way nested grids at horizontal grid spacings of 3 and 1 km are used (Fig. 5). The 3-
222 km domain covers an area of $1200 \times 1200 \text{ km}^2$ and is centered at $(32.5^\circ\text{N}, 118.5^\circ\text{E})$. The 1-km
223 domain is $460 \times 460 \text{ km}^2$ in size and covers almost all of Jiangsu Province. Both domains have 53
224 vertical levels, which are stretched using a hyperbolic tangent function as described in Xue et al.
225 (1995), with vertical grid spacing varying from 50 m at the surface to nearly 1000 m at the model
226 top; the average vertical grid spacing is 500 m. The upper and lower boundaries are set as rigid
227 walls and a two-layer soil model is applied to facilitate the calculations of surface fluxes based on
228 the predicted surface temperature and soil moisture content. Sub-grid scale turbulent mixing is
229 parameterized using a 1.5-order turbulence kinetic energy (TKE) scheme and radiative processes
230 are parameterized via the NASA Goddard Space Flight Center long- and short-wave radiation

231 schemes. Fourth-order advection is used in both horizontal and vertical directions, and fourth-order
232 computational mixing is applied to suppress numerical noise. More details on the ARPS physics
233 schemes and their settings can be found in Xue et al. (2001, 2003), together with the references for
234 the parameterization schemes.

235 As noted by previous studies (e.g., Loftus et al. 2014a; Snook et al. 2016; L17), hail forecast
236 errors are closely tied to uncertainties within MP schemes. Herein, simulations are conducted using
237 the MY one-, two-, and three-moment schemes (MY05a, b). Morrison and Milbrandt (2010)
238 showed that even two very similar 2-moment MP schemes could produce distinct differences in
239 simulated storms due to differences in details of the schemes. For this reason, we choose to limit
240 ourselves to the comparison of MY schemes having the same treatment of microphysical processes
241 but predicting different number of moments or diagnosing one of the DSD parameters. This way,
242 focus is placed on the effects of the number of predicted moments on hail forecast. In the MY
243 schemes, six distinct hydrometeor categories, i.e., cloud water, cloud ice, rain, snow, graupel, and
244 hail, are included. The PSD of each hydrometeor is represented by a gamma distribution function,

$$245 \quad N_x(D) = N_{0x} D^{\alpha_x} \exp(-\lambda_x D) \quad (1)$$

246 where $N_x(D)$ is the total number concentration per unit volume of diameter D for hydrometeor
247 category x . α_x is the shape parameter, giving a measure of the spectral width. N_{0x} and λ_x are the
248 intercept and slope parameters, respectively.

249 The three-moment scheme is used as the control experiment, and three other experiments are
250 performed using the one- and two-moment MY schemes. Table 1 summarizes the key parameters of

251 all experiments. Two of the simulations called FixA and DiagA use variants of the MY two-
252 moment schemes with different treatment of the shape parameter. FixA sets the shape parameter to
253 a default constant value of 0 for all hydrometeor categories, while in DiagA the shape parameters of
254 hail and other hydrometeors are diagnosed from the mean-mass diameter of the corresponding
255 categories based on the Eqs. (12) and (13) of MY05a. The one-moment MY scheme is also tested
256 with intercept and shape parameters set to their default constant values. The above configurations
257 are similar to those of L17, except for the inclusion of the 1 km nested grid. Because 1-km grid is
258 expected to simulate the hailstorm better, we focus on results of 1-km grid in this paper. We also
259 examined the forecasts from the 3-km grid; the dominant cells simulated on the two grids are found
260 to be generally similar, although some differences exist in storm intensity at small scales.

261 **4. Evaluation metrics for hail prediction**

262 Three metrics for explicit hail prediction are used to evaluate hail forecast skill for the
263 various MP schemes within the sensitivity experiments. They are: maximum estimated hail size
264 (MESH) (Witt et al. 1998), maximum hail size (D_{max}) (MY06a), and surface accumulated hail
265 number concentration (SAHNC). In addition, an objective neighborhood-based evaluation
266 technique, the fractions skill score, is used to verify the simulated MESH against the radar-derived
267 counterpart. L17 examined the accumulated surface precipitation and MESH fields based on the
268 simulations of a pulse-hailstorm, but not the D_{max} and SAHNC. MESH and D_{max} were also
269 examined in Snook et al. (2016) for a supercell storm case.

270 *a. Maximum estimated hail size (MESH)*

271 As described in L17, the MESH algorithm uses a weighted integration of radar reflectivity
272 exceeding 40 dBZ above the melting level to obtain an estimate of the maximum size of hail
273 occurring at the surface. Following L17, reflectivity datasets from multiple radars are interpolated to
274 the model grid to derive MESH. Since the MESH algorithm was only configured for hail sizes
275 larger than 19 mm (Witt et al. 1998), and Cintineo et al. (2012) and L17 only evaluated MESH
276 down to the size of 21 and 19 mm respectively, MESH values below 20 mm are excluded in this
277 study. More details about the MESH algorithm can be found in L17.

278 We note that since the MESH algorithm relies entirely upon the weighted integration of
279 radar reflectivity exceeding 40dBZ above the 0°C level to estimate hail size at the surface, there
280 may exist some biases within the derived MESH swath (e.g., Cintineo et al. 2012; Ortega et al.
281 2009). Because no other high quality/high-resolution observation of hail size is available, herein we
282 choose to use the high-resolution radar-derived MESH for verification of the hail simulations.

283 *b. Maximum hail size (D_{max})*

284 The maximum hail size (D_{max}) (MY06a) is defined as the largest hail size for which the total
285 number concentration of hail particles greater than a diameter is equal to pre-specified total number
286 concentration, N_{THRE} . For example, if D_{max} is 40 mm, the total number concentration of hailstones
287 larger than 40 mm is N_{THRE} . The D_{max} parameter serves to identify the instantaneous presence of
288 large hail within the storm. Following MY06a, a threshold value of $N_{THRE} = 10^{-4} \text{ m}^{-3}$ is adopted here.

289 *c. Surface accumulated hail number concentration (SAHNC)*

290 Given that the accumulated hail number is also important for hail prediction, SAHNC is
291 proposed as a new parameter to estimate the surface accumulated number concentration of hail
292 larger than a particular size. The SAHNC parameter is not only useful for identifying the surface
293 accumulated hail size distribution, but also helpful in understanding storm evolution. SAHNC is
294 defined as an integration of the flux of large hail $R_h(D)$ at 60-second intervals during hailfall from
295 T_0 to T_1 ,

296
$$N_{Dh}(D) = \int_D^{\infty} N_h(D^*) dD^* \quad (2)$$

297
$$V_h(D) = \gamma a_h \times D^{b_h} \times e^{-f_h} \quad (3)$$

298
$$R_h(D) = N_h(D) \times V_h(D) \quad (4)$$

299
$$SHNAC(D) = \int_{T_0}^{T_1} R_h(D) dt \quad (5)$$

300 where $N_{Dh}(D)$ is the total number concentration of hail larger than diameter D , and the size
301 distribution of hail is described by gamma distribution function as equation (1). Terminal fall
302 velocity at the surface for a hailstone with diameter D is given by Eq. (3), where $\gamma = (\rho_0/\rho)^{1/2}$ is
303 the density correction factor with ρ_0 and ρ being the surface air and air density; a_h , b_h , and f_h are
304 set to be 206.89, 0.6384, and 0.0, respectively following Ferrier (1994).

305 *d. Neighborhood-based hail forecast evaluation*

306 As reviewed in Casati et al. (2010), evaluation of forecasts from high-resolution models has
307 been a subject of active research in recent years, and various evaluation metrics have been

308 developed. Objective evaluation of hail forecasts is still very challenging, partly due to the lack of
 309 high-quality, high-resolution hail observations (Snook et al. 2016). In this study, the fractions skill
 310 score (FSS) neighborhood technique (Roberts and Lean 2008), is applied to MESH fields derived
 311 from multiple radar observations and the simulations to examine hail size forecast skill using
 312 different MP schemes. Distinct from traditional point-by-point evaluation techniques, FSS
 313 compares fractional coverage of forecasts against that of observations within a neighborhood
 314 centered at each grid point. By varying the neighborhood size and the MESH threshold, scale-
 315 dependent forecast skill can be assessed. Following Roberts and Lean (2008), FSS is defined as

$$316 \quad FSS = 1 - \frac{FBS}{\frac{1}{N} \left[\sum_{i=1}^N \langle P_{F(i)} \rangle_s^2 + \sum_{i=1}^N \langle P_{O(i)} \rangle_s^2 \right]} \quad (6)$$

317 where FBS is the fractional Brier score, given as

$$318 \quad FBS = \frac{1}{N} \sum_{i=1}^N \left(\langle P_{F(i)} \rangle_s - \langle P_{O(i)} \rangle_s \right)^2 \quad (7)$$

319 In Eqs. (6) and (7), N is the total number of grid boxes in the predefined neighborhood
 320 (within a given radius); $P_{F(i)}$ and $P_{O(i)}$ are the fractional areas at the i^{th} neighborhood of forecast and
 321 observation, respectively. They are analogous to the probability that a given neighborhood contains
 322 values larger than the pre-specific threshold. Therefore, FSS compares fractional coverage over a
 323 neighborhood of given size, rather than values at each grid box, and FSS values range from 0 to 1.
 324 A score of 1 signifies a forecast perfectly matching the observation within a specific neighborhood
 325 for a given intensity threshold, while 0 signifies a complete mismatch. A forecast is considered to

326 be skillful when the FSS value exceeds FSS_{useful} , which is defined as (Roberts and Lean 2008),

327
$$FSS_{\text{useful}} = 0.5 + \frac{f_{\text{obs}}}{2} \quad (8)$$

328 where f_{obs} is an average of observed fraction within the entire domain. By calculating FSS at a
329 variety of spatial scales (neighborhood sizes) and MESH thresholds, one can determine how the
330 forecast skill varies with spatial scale and at which scale a forecast has useful skill for a given
331 MESH threshold.

332 5. Results

333 In this section, results of the simulations are presented. First, to validate the simulations of
334 the multi-cell hailstorm system, simulated composite (column-maximum) radar reflectivity is
335 compared with corresponding radar observations. Explicit hail forecast skills using various MP
336 schemes are then evaluated, in terms of MESH, surface accumulated solid water mass, and SAHNC.
337 Neighborhood-based FSSs for simulated MESH are calculated against radar-derived MESH. The
338 differences in hail distribution characteristics within storms simulated using different MP schemes
339 are also investigated. To understand the reasons behind the differences among various MP schemes
340 for hail prediction, microphysical budget analyses are performed.

341 *a. Simulated storm evolution*

342 As discussed in section 2, the lifespan of this multi-cellular hailstorm can be characterized
343 by two episodes between 0700 and 1400 UTC on April 28, 2015. Figures 6 and 7 show simulated
344 composite reflectivity fields from the experiments and radar observations from one time during each
345 episode, at 0900 and 1400 UTC, respectively. Comparisons with the radar observations (Figs. 6e, 7e)

346 indicate that the time and location of the hailstorm's initiation along the northwest border of Jiangsu
347 Province, as well the later organization into a large bow-shaped echo are well reproduced in the
348 simulations (Figs. 6a-d, 7a-d). The direction of movement of the simulated storms is also in general
349 agreement with observations.

350 Although the storms' evolution and motion are generally reproduced quite well in the
351 simulations, there exist significant storm intensity differences among the experiments using various
352 MP schemes. Experiment Single under-predicts the reflectivity magnitude, having few instances of
353 reflectivity exceeding 60 dBZ, and the stratiform precipitation region (<35 dBZ) in Single is larger
354 compared to simulations using multi-moment schemes (Figs. 6a, 7a). This result differs from those
355 of some previous studies (e.g., Morrison et al. 2009, Bryan and Morrison 2012, Baba and Takahashi
356 2014); their studies noted that the stratiform precipitation region in idealized two-dimensional
357 squall line simulations was smaller using a one-moment than a two-moment scheme and they
358 attributed it to decreased rain evaporation rates in the 2-moment schemes in the trailing stratiform
359 region. In our hailstorm case, it is believed that for the multi-moment scheme, the size-sorting
360 mechanism may allow larger sized hailstones to fall rapidly towards the ground, giving less time for
361 hail mass advection downwind of the updraft and hence the smaller stratiform precipitation region.
362 This may be one of the reasons at least.

363 The reflectivity magnitudes from the multi-moment schemes are more or less over-predicted
364 when compared with radar observations (Figs. 6b-e, 7b-e). Thus, storm intensities produced by the
365 multi-moment schemes are not too different based on radar reflectivity, while the hail prediction

366 skill of each scheme exhibits significant discrepancies (see Figs. 8-12). Moreover, in all cases, the
367 westward extension of the reflectivity towards Anhui Province is under-predicted, and the simulated
368 system exhibits slower southward movement (possibly due to uncertainties in the initial condition),
369 resulting in the entire system being displaced almost 80 km northward compared to the observations
370 by 1400 UTC.

371 *b. Hailstone forecast and evaluation*

372 As the main goal of this study is to evaluate the hail forecast skills of various MP schemes,
373 predicted hail size distribution features, including hailstone size, mass, and number concentration
374 will be examined in this section.

375 1) Radar-based hail forecast evaluation using MESH

376 Swaths of MESH derived from the forecasts and from radar observations between 0600 and
377 1600 UTC at 5-minute intervals are presented in Fig. 8. There is one primary, nearly continuous
378 MESH swath in the radar observations (Fig. 8e), with MESH values exceeding 40 mm in many
379 locations. Along the radar-indicated MESH swath, egg-sized hailstones (with diameters of 20-50
380 mm) were reported in several cities, including Yizheng, Changzhou, Wuxi, and Suzhou. Among
381 them, maximum MESH values exceed 80 mm over Yizheng, which coincides with the location of
382 largest hail reported during this storm—a report from Yizheng of hail over 100 mm in diameter.
383 Compared with radar-derived MESH, the MESH swath derived from experiment Single exhibits
384 smaller maximum hail sizes, with several scattered cores of MESH values of no more than 35 mm
385 predicted (Fig. 8a). In FixA and DiagA, which use two-moment schemes, MESH swaths exhibit

386 significant differences from observations. Maximum MESH in FixA is highly overestimated, the
387 MESH swath from this experiment exhibits large areas of MESH exceeding 70 mm (Fig. 8b); in
388 DiagA MESH is significantly underestimated, not exceeding 30 mm at any point (Fig. 8c). The
389 MESH swath of CNTL appears to match well with the observed swath, with the MESH values in
390 the range of 40-50 mm within narrow cores (Fig. 8d). A narrow core of high MESH values (over 80
391 mm) is present within the primary MESH swath in CNTL (Fig. 8d). This is generally consistent
392 with the maximum registered hail size in Yizheng, although the swath is displaced tens of
393 kilometers northward from the observations because of the overall northeastward position error of
394 the storms.

395 FSSs at different neighborhood radii and MESH thresholds, and the corresponding FSS_{useful}
396 values are presented in Fig. 9. The scale at which FSS exceeds FSS_{useful} can be considered the
397 “skillful scale” of forecast (Roberts and Lean 2008).

398 FSSs of FixA and CNTL with MESH thresholds of 30/40 mm, and Single with MESH
399 thresholds of 30, increase with increasing neighborhood radius, and higher scores are achieved at
400 lower MESH threshold (Fig. 9). For MESH thresholds of 30 and 40 mm (Figs. 9a, b), experiment
401 CNTL achieves useful skill for neighborhood radii of 47 and 50 km and larger, respectively. The
402 large neighborhood radius for useful skill is likely due in large part to the northeastward
403 displacement error of the model storms. The comparisons of FSSs from the experiments further
404 confirm that the three-moment scheme outperforms others in terms of hail forecast skill, especially
405 for hail exceeding 40 mm, as indicated by the notably higher FSSs at all spatial scales larger than 10

406 km (Fig. 9b). Experiment Single has no skill at any spatial scale for hailstones larger than 40 mm, as
407 indicated by the zero FSSs at all neighborhood radii (Fig. 9b). Similarly, DiagA shows no skill for
408 predicting hailstones larger than 30 mm (Fig. 9a).

409 2) Surface accumulated hail mass and number concentration

410 Since there are no in-situ hail-count observations, the simulated hail mass and number
411 concentration fields are compared among the experiments in light of the registered hail reports.
412 Surface accumulated hail mass fields from the experiments, calculated between 0600 and 1600
413 UTC throughout almost entire lifespan of the hailstorm are shown in Fig. 10. The corresponding
414 SAHNC, derived from the model output within the same period at 60-second intervals for hail
415 diameter thresholds of 30 and 40 mm, are shown in Figs. 11 and 12.

416 The storms in Single produce a northwest-southeast-oriented swath of hail mass, with peak
417 values of around 80 mm, mainly concentrated within a narrow band approximately 10 km in width
418 (Fig. 10a). In DiagA, the accumulated hail mass amounts are the smallest among all experiments,
419 with peak values below 40 mm (Fig. 10c). The width of the primary hail mass swath in FixA is the
420 widest (around 20 km) among the experiments. Peak values in FixA and CNTL are similar (~80
421 mm), and the hail mass predicted by CNTL appears to be more concentrated along a relatively
422 straight path (Fig. 10d).

423 Distributions of SAHNC from the experiments generally coincide with their hail mass
424 distributions. For example, SAHNC of Single with hail size larger than 30 mm (hereafter
425 SAHNC30) within the primary narrow hail mass band is around $10^4\sim 10^5$ m⁻² (Fig. 11a), and the

426 number concentration of hail larger than 40 mm (hereafter SAHNC40) is around 10^3 m^{-2} (Fig. 12a).
427 In contrast, SAHNC30 of DiagA is around $10^1 \sim 10^2 \text{ m}^{-2}$, approximately 1-2 orders of magnitudes
428 smaller than other experiments (Fig. 11c). Only a few small patches of SAHNC40 are presented in
429 Fig. 12c, which is also consistent with its MESH evaluation results that DiagA has no forecast skill
430 for hail larger than 40 mm. Although the magnitudes of SAHNC30/40 from FixA are similar to
431 those from CNTL, FixA produces more large hailstones at the surface than other experiments, and
432 predicts a swath almost twice the width of that in CNTL.

433 Forecasts of accumulated hail mass and number can also be cross-referenced with hail
434 reports and photographs from the event to infer their level of accuracy. In some areas, photographs
435 and reports indicate that the depth of surface accumulated hail exceeded 10 cm, which is more or
436 less consistent with the surface accumulated solid precipitation of CNTL (Fig. 10d); it produces a
437 concentrated hail mass band of over 8 cm in depth. If we assume hail to be spherical, and transfer
438 the reported hail depth and size to hail number accumulated at the surface, this corresponds to a
439 value of $10^3 \sim 10^4 \text{ m}^{-2}$ for hail larger than 4 cm. This is also better captured by CNTL than other
440 experiments, although there is still overestimation in some areas. Based on these inferences, the
441 surface accumulated hail mass and number predicted by CNTL appear to be more accurate
442 compared with other experiments.

443 3) Hailstone distribution characteristics within storms

444 Given the significant differences in the predicted surface hail size distributions among
445 various MP schemes, we next examine hail distribution properties within simulated storms.

446 Microphysical fields, including the hail mass content (Q_h), total hail number concentration (N_{th}),
447 maximum hail size (D_{max}), and reflectivity (Z) are examined at 1100 UTC, when the cells are
448 vigorous and well-developed (see Figs. 13 and 14). Vertical cross sections are taken from west to
449 east, passing through the primary hail mass core of the simulated cells; the rough location is
450 indicated by the thick black line in Fig. 2c. Hail size spectra at some typical points in sensitivity
451 experiments are also examined (Fig. 15).

452 The hail distribution characteristics within the storms from various experiments exhibit large
453 differences (Fig. 13). For example, FixA and DiagA produce copious hail mass aloft, with a peak
454 Q_h of 13 g m^{-3} (Fig. 13b, $x \sim 200\text{-}216 \text{ km}$, $z \sim 7\text{-}9 \text{ km}$; Fig. 13c, $x \sim 212\text{-}217 \text{ km}$, $z \sim 8\text{-}10 \text{ km}$), while
455 the peak Q_h in CNTL and Single is only about 7 g m^{-3} (Figs. 13a, d, $x \sim 198\text{-}202 \text{ km}$, $z \sim 4\text{-}6 \text{ km}$). N_{th}
456 fields from the two-moment simulations are generally similar to CNTL, all have larger N_{th} values
457 ($>10^3 \text{ m}^{-3}$) in the storm anvil region above the freezing level. The magnitudes of the peak N_{th} in
458 FixA are one to two orders higher than the peak N_{th} values in CNTL, especially in the rear part of
459 the cell (Fig. 13b, $x \sim 165\text{-}204 \text{ km}$, $z \sim 2\text{-}7 \text{ km}$; Fig. 13c, $x \sim 201\text{-}208 \text{ km}$, $z \sim 4\text{-}6 \text{ km}$). Similar
460 conclusions about the over-prediction of Q_h and N_{th} for two-moment schemes had also been made in
461 Loftus and Cotton (2014b) and MY06b in their hailstorm simulations. The over-prediction of the
462 moments for the two-moment scheme with a fixed shape parameter could be ascribed to the
463 excessive size sorting and inability to narrow the particle size spectrum with time (Milbrandt and
464 McTaggart-Cowan 2010). N_{th} diagnosed by Single is considerably different from the other runs; the
465 peak in N_{th} is much smaller ($<10^{2.5} \text{ m}^{-3}$). Since N_{th} in Single is a monotonic function of Q_h , the

466 peaks of N_{th} and Q_h are collocated.

467 The D_{max} and reflectivity of Single (Fig. 14a) are also monotonically related to Q_h , and their
468 peak cores are also collocated. Since only the mixing ratio is predicted in Single, all the diagnosed
469 moments sediment at the mass-weighted fall speed, precluding any size sorting. In experiments
470 using multi-moment schemes, D_{max} and reflectivity generally increase towards the surface, and the
471 high D_{max} and reflectivity columns are located almost directly below the corresponding Q_h cores
472 (Figs. 13, 14b-d), consistent with a size sorting process. To investigate the effect of size sorting on
473 hail distribution, additional sets of experiments with size sorting effect suppressed in the two and
474 three-moment schemes were conducted. In these experiments, size sorting for hydrometeor species
475 is disallowed by forcing all predicted moments to sediment at the mass-weighted fall speed. We
476 examined the microphysical fields from the experiments with size sorting disabled, and found that
477 they exhibited substantial similarities to experiment Single. The Q_h and N_{th} fields in these
478 experiments display a broader region of relatively weak gradients over most of the forward flank
479 above the melting layer, with smaller D_{max} values (not shown). This strongly suggests that the size
480 sorting effect plays an important role in controlling the hail distribution characteristics within the
481 storm. In multi-moment MY schemes, different moment-weighted terminal velocities enables size
482 sorting of particles, leading to more realistic hail distribution properties in the vertical.

483 Hail size spectra within the storms for each experiment at 1100 UTC, in the main updraft
484 region of each case (as indicated in Fig. 14), are plotted in Fig. 15. In FixA and Single, $\alpha_h = 0$, the
485 DSD curves (black and green in Fig. 15) are exponential. FixA suffers from excessive size sorting,

486 since Q_h sediments faster than N_{th} when α_h is fixed at 0 (as can be seen in Fig. 1 of MY05a),
487 consistent with the unrealistically large D_{max} at the low levels (Fig. 14b). In contrast, the diagnosed
488 α_h in DiagA is about 2.8 at this point (see Fig. 16) and size sorting is more limited. The hail size
489 spectrum appears to be artificially narrowed in DiagA compared to CNTL, causing a shift of
490 spectrum distribution towards smaller sizes. Figure 15b shows that in CNTL the hail size spectrum
491 becomes broader as height decreases (Fig. 15b), corresponding to decreasing slope parameter. In the
492 meanwhile, D_{max} increases quickly as the ground is approached (Fig. 14d).

493 Furthermore, given that shape parameter has significant effects on sedimentation and
494 microphysical growth rates (Milbrandt and McTaggart-Cowan 2010; Mansell 2010; Dawson et al.
495 2014) and size sorting can also affect size spectra of hydrometeors (MY05a), diagnostic analyses
496 are performed to assess the differences in α_h among the experiments. The horizontally- and
497 temporally-averaged α_h values within the storm, with the one standard deviation interval shaded,
498 are plotted for each experiment between 0600 and 1600 UTC in Fig. 16. The mean α_h of CNTL
499 decreases significantly from ~ 3.2 near surface to ~ 0.2 near the melting layer (approximately 4 km
500 above the surface). The decrease is almost linear with height up to approximately 2.5 km, and the
501 decrease continues above 4 km. This shape parameter profile agrees with previous studies (MY06a,
502 b), which noted that large α_h mainly occurred below 600 hPa, with near-zero values above 600 hPa.
503 The smaller α_h above the freezing level may partly result from creation of hail via freezing of
504 raindrops, which adds numerous small particles to the hail distribution (MY06b). Below the

505 freezing level, smaller hail particles tend to melt quickly, increasing α_h . In any case, the α_h profile
506 of CNTL indicates its variation with height from 0 to 5, suggesting that using a fixed α_h value is
507 inappropriate.

508 Although the diagnosed α_h in DiagA exhibits a vertical profile with right trend, it differs
509 quite significantly from that of CNTL. The diagnosed α_h does not decrease quickly with height and
510 maintains high values (around 2.9) at upper levels (Fig. 16). These result directly from the α_h
511 diagnostic formula used (see Eq. 12 in MY05a) that keeps the diagnosed value within a range of 2.8
512 to 4.5 for D_{mh} below 8 mm. This diagnostic relation was obtained using a one-dimensional model
513 where only the sedimentation process was considered (MY05b); it appears to be inaccurate
514 compared to values produced by the three-moment scheme.

515 c. *Microphysical budget analyses*

516 To gain additional physical insights regarding the differences between MP schemes, budget
517 analyses of microphysical processes within the simulated storms are performed. The hail production
518 terms are integrated over the entire 1-km simulation domain from surface to model top, according to
519 hail mixing ratio prediction equation for the MY scheme (see Eq. [A7] in MY05b). Figure 17 shows
520 the time series of total hail production tendency terms, including hail collection of rain (*colqrh*) and
521 cloud water (*colqch*), and hail melting to rain (*meltqh*), for the four experiments. Other processes,
522 including hail collection of snow or ice, are minimal (Heymsfield and Pflaum 1985), and are not
523 shown. Terms *colqrh*, *colqch* and *meltqh* in FixA are significantly larger than those in other
524 experiments between 0600 and 1500 UTC, with peak rates of approximate 3000, 8000 and -10200

525 kiloton/min at 0930 UTC, respectively. In other experiments, hail growth rates from these processes
526 are smaller, especially in DiagA and Single. Generally speaking, the main differences in hail growth
527 processes among the experiments match the differences in the predicted hail mass distributions
528 within storms and hail accumulation at surface, as discussed earlier.

529 Time-averaged vertical profiles of hail production terms *colqrh*, *colqch* and *meltqh* are
530 plotted in Fig. 18. The profiles are averaged horizontally over points within the hailstorm during its
531 lifespan between 0600 and 1500 UTC with data at 1-minute intervals. The profiles exhibit generally
532 similar vertical patterns for the four experiments, with larger values of *colqrh* and *meltqh* occurring
533 below ~4 km (around the height of freezing level), and *colqch* within the layer between ~2.5 and ~9
534 km. For experiments using multi-moment MP schemes, substantial differences are mainly located
535 below the melting layer. Compared to the peaks values of CNTL (485 kiloton/min for *colqrh* and -
536 840 kiloton/min for *meltqh*), FixA has much larger peak values of *colqrh* (696 kiloton/min) and
537 *meltqh* (-1212 kiloton/min). These larger peak values appear to reflect rapid growth of larger
538 hailstones in FixA through *colqrh* below the melting layer; the sensible heating from *colqrh* is offset
539 by the cooling of rapid melting of copious amounts of smaller hail, which may lead to more
540 excessive size sorting. In DiagA, as the hail size spectrum is artificially narrowed (see Fig. 15a),
541 lower collection efficiency of smaller hailstones leads to smaller peak *colqrh* (396 kiloton/min). The
542 smaller hailstones tend to melt more quickly (-922 kiloton/min), and have lower terminal velocities;
543 these effects combine to result in fewer hailstones accumulating at the surface (see Figs. 11c, 12c).

544 6. Summary and conclusions

545 This study evaluates the ability of different MP schemes within storm-scale NWP model
546 runs at a 1-km grid spacing to explicitly predict hail in a long-lasting multi-cellular hailstorm that
547 occurred in Jiangsu Province, China on 28 April, 2015. The hailstorm started within a strong low-
548 level convergence between two low-level cyclones that are underneath an upper-level cut-off low
549 within a deep coastal trough. The environment, associated with a weak warm-moist PBL-capping
550 inversion, featured strong 0-6 km wind shear, moderate CAPE, and a very low CIN. The mid-levels
551 were relatively dry. Such an environment is generally conducive to deep convection that tends to
552 produce large hailstones.

553 The simulations employed one-, two-, and three-moment MY MP schemes (MY05a, 05b).
554 Two variants of two-moment schemes were used, one in which the shape parameters of
555 hydrometeors were fixed at zero, and the other in which the shape parameters were diagnosed as a
556 monotonically increasing function of the mean mass-weighted diameter of hydrometeor particles.
557 Evaluations were performed against available observations, including radar reflectivity, radar-
558 derived maximum estimated hail size (MESH), and available severe weather reports. Furthermore,
559 neighborhood-based fraction skill scores (FSSs) were calculated for the simulated MESH fields for
560 objective evaluation.

561 Evaluations against observed radar reflectivity indicate that the time and location of the
562 hailstorm initiation and the later organization of storm cells into a large bow-echo are reasonably
563 reproduced by all experiments. Compared with radar observations, experiment FixA, which uses a
564 fixed shape parameter of zero, substantially over-predicts the magnitudes of reflectivity, and as a

565 result produces unrealistically high MESH values (with maxima exceeding 70 mm) compared with
566 radar-derived MESH (which has maxima of only 40-50 mm). In contrast, in Single and DiagA that
567 uses a single-moment and diagnostic shape parameter scheme, respectively, reflectivity and MESH
568 fields are under-predicted in both intensity and extent. CNTL using three-moment MY scheme
569 produces MESH swaths that agree more closely with radar-derived MESH swaths than other
570 experiments, and neighborhood-based MESH evaluations further show that the three-moment
571 scheme has notably higher fractional skill scores at all spatial scales compared to the other schemes,
572 especially for large hailstones.

573 Surface accumulated hail mass, number, and hail distribution characteristics within storms
574 are inter-compared among the experiments. Results suggest that FixA produces significant amounts
575 of large hail accumulated over a much wider swath than CNTL. For Single and DiagA, the peak
576 SAHNC values are about two orders of magnitude smaller than those of CNTL, especially in DiagA
577 where almost no hail larger than 40 mm is produced. Examinations of hail distributions within
578 storms indicate that since all the moments of a given hydrometeor type are monotonically related to
579 the mixing ratio (which is the only moment predicted) in Single, no size sorting can occur. For
580 multi-moment schemes, different moment-weighted terminal velocities allow for size sorting of
581 particles, making it possible to reproduce more realistic PSDs within the storm. However,
582 substantial differences in the hail size distributions are still present within storms simulated using
583 different multi-moment schemes. For example, FixA, which uses a two-moment scheme with a
584 fixed α_x value of zero, suffers from excessive size sorting which leads to an unrealistic shift in hail

585 DSD towards larger hailstones during sedimentation. On the other hand, the diagnostic α_x used in
586 DiagA is at least 2.8, resulting in a hail size spectrum that appears to be artificially narrowed
587 compared to CNTL; it causes a spectrum shift towards smaller hailstones and yields smaller D_{max}
588 and SAHNC. These results indicate that although excessive size sorting is more limited in DiagA,
589 the specific α_x diagnostic relation derived from sedimentation-only one-dimensional model appears
590 inaccurate. Therefore, more accurate diagnostic relations for α_x may need to be derived, using
591 perhaps output from full three-moment simulations. In fact, our preliminary results using this
592 approach are encouraging and more complete results will be reported in a separate paper.

593 Furthermore, budget analyses of hail production terms suggest that collection of rain and
594 cloud water by hail are dominant contributors to hail mass growth. The differences in hail growth
595 processes among different experiments are closely linked to the treatment of shape parameter in
596 different MP schemes, which further lead to the differences in the predicted surface accumulated
597 hail mass, SAHNC, and hail distribution within the simulated storms.

598 In the end, we note that there are many other possible configurations for one- or two-
599 moment MP schemes in terms of the choice of fixed or variable intercept and shape parameters,
600 which can be further evaluated in the future. We also note that due to the lack of reliable
601 observations of surface accumulated hailstones, evaluations of explicit hail prediction in this paper
602 carries a certain degree of uncertainty. In-situ observations of microphysical processes, as well as
603 hail size distributions are needed for more reliable evaluations.

604

605 *Acknowledgement.* This work was primarily supported by the National 973 Fundamental Research
606 Program of China (2013CB430103). Liping Luo was supported by Nanjing University for her
607 extended visit at CAPS, University of Oklahoma. The work was also supported by the National
608 Science Foundation of China (Grant No. 41405100), Foundation of China Meteorological
609 Administration special (Grant No. GYHY201506006). We gratefully acknowledge the High
610 Performance Center (HPCC) of Nanjing University for doing the numerical calculations in this
611 paper on its IBM Blade cluster system. The second author acknowledges the support of the National
612 Thousand Person Plan hosted at the Nanjing University, and the support of NSF grants AGS-
613 1261776 for the Severe Hail, Analysis, Representation, and Prediction (SHARP) project. Nathan
614 Snook and Jonathan Labriola are thanked for proofreading and improving the manuscript.
615 Suggestions and comments from three anonymous reviewers also improved our paper. The NCEP
616 FNL analysis data can be downloaded freely at <http://rda.ucar.edu/datasets/ds083.2/> The radar
617 dataset is provided by the Climate Data Center at National Meteorological Information Center of
618 China Meteorological Administration, and processed data used in this paper are available at
619 <https://pan.baidu.com/s/1kV0WrTP> (password: 293v).

620

621 **References**

- 622** R.D. Adams-Selin and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth
623 model within WRF. *Mon. Wea. Rev.*, **144**, 4919-4939.
- 624** Baba, Y., and K. Takahashi, 2014: Dependency of stratiform precipitation on a two-moment cloud
625 microphysical scheme in mid-latitude squall line. *Atmos. Res.*, **138**, 394-413.
- 626** Brimelow, J. C., and G. W. Reuter, 2009: Explicit Forecasts of Hail Occurrence and Expected Hail
627 Size Using the GEM–HAILCAST System with a Rainfall Filter. *Wea. Forecasting*, **24**, 935-
628 945.
- 629** Brimelow, J. C., G. W. Reuter, and E. R. Poolman, 2002: Modeling Maximum Hail Size in Alberta
630 Thunderstorms. *Wea. Forecasting*, **17**, 1048-1062.
- 631** Bryan, G. H., and H. Morrison, 2012: Sensitivity of a Simulated Squall Line to Horizontal
632 Resolution and Parameterization of Microphysics. *Mon. Wea. Rev.*, **140**, 202-225.
- 633** Casati, D. B., and Coauthors, 2010: Forecast verification: current status and future directions.
634 *Meteor. Appl.*, **15**, 3-18.
- 635** Changnon, S. A., 1999: Data and Approaches for Determining Hail Risk in the Contiguous United
636 States. *J. Appl. Meteor.*, **38**, 1730-1739.
- 637** China Meteorological Administration, 2013: *Yearbook of Meteorological Disasters in China*, pp.
638 31-43, China Meteorological Press, Beijing, China. (in Chinese)

- 639 China Meteorological Administration, 2014: *Yearbook of Meteorological Disasters in China*, pp.
640 37-50, China Meteorological Press, Beijing, China. (in Chinese)
- 641 China Meteorological Administration, 2015: *Yearbook of Meteorological Disasters in China*, pp.
642 28-43, China Meteorological Press, Beijing, China. (in Chinese)
- 643 Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective
644 high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**,
645 1235-1248.
- 646 Costa, S., P. Mezzasalma, V. Levizzani, P. P. Alberoni, and S. Nanni, 2001: Deep convection over
647 Northern Italy: Synoptic and thermodynamic analysis. *Atmos. Res.*, **56**, 73-88.
- 648 Dawson, D. T., M. Xue, J. A. Milbrandt, and M. Yau, 2010: Comparison of Evaporation and Cold
649 Pool Development between Single-Moment and Multimoment Bulk Microphysics Schemes
650 in Idealized Simulations of Tornadic Thunderstorms. *Mon. Wea. Rev.*, **138**, 1152-1171.
- 651 Dawson, D. T., E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level
652 Z DR signatures in supercell forward flanks: The role of size sorting and melting of hail. *J.*
653 *Atmos. Sci.*, **71**, 276-299.
- 654 Ebert, E. E., 2009: Neighborhood Verification: A Strategy for Rewarding Close Forecasts. *Wea.*
655 *Forecasting*, **24**, 1498-1510.
- 656 Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I:
657 Description. *J. Atmos. Sci.*, **51**, 249-280.
- 658 Gagne, D. J., I. A. McGovern, and X. M (2014), Machine learning enhancement of storm scale

659 ensemble precipitation forecasts, *Wea. Forecasting*, **29**, 1024-1043.

660 Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation uncertainty due to variations
661 in precipitation particle parameters within a simple microphysics scheme. *Mon. Wea. Rev.*,
662 **132**, 2610-2627.

663 Guo, X., and M. Huang, 2002: Hail formation and growth in a 3D cloud model with hail-bin
664 microphysics. *Atmos. Res.*, **63**, 59-99.

665 Heymsfield, A. J., and J. C. Pflaum, 1985: A Quantitative Assessment of the Accuracy of
666 Techniques for Calculating Graupel Growth. *J. Atmos. Sci.*, **42**, 2264-2274.

667 Hu, K., R. Lu, and D. Wang, 2010: Seasonal climatology of cut-off lows and associated
668 precipitation patterns over Northeast China. *Meteor. Atmos. Phys.*, **106**, 37-48.

669 Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell
670 storm using a two-moment bulk microphysics scheme. *J. Appl. Meteor. Climatol.*, **49**, 146-
671 163.

672 Kessler, E., 1995: On the continuity and distribution of water substance in atmospheric circulations.
673 *Atmos. Res.*, **38**, 109-145.

674 Labriola, J., N. Snook, Y. Jung, B. Putnam, and M. Xue, 2017: Ensemble Hail Prediction for the
675 Storms of 10 May 2010 in South-Central Oklahoma Using Single- and Double-Moment
676 Microphysical Schemes. *Mon. Wea. Rev.*, **145**, 4911-4936.

677 Lin, Y. L., R. D. Farley, and H. D. Orville, 1983: Bulk Parameterization of the Snow Field in a

678 Cloud Model. *J. Appl. Meteor.*, **22**, 1065-1092.

679 Loftus, A., and W. Cotton, 2014b: A triple-moment hail bulk microphysics scheme. Part II:
680 Verification and comparison with two-moment bulk microphysics. *Atmos. Res.*, **150**, 97-128.

681 Loftus, A. M., W. R. Cotton, and G. G. Carrió, 2014a: A triple-moment hail bulk microphysics
682 scheme. Part I: Description and initial evaluation. *Atmos. Res.*, **149**, 35-57.

683 Luo, L., M. Xue, K. Zhu, and B. Zhou, 2017: Explicit prediction of hail using multimoment
684 microphysics schemes for a hailstorm of 19 March 2014 in eastern China. *J. Geophys. Res.*
685 *Atmos*, **122**, 7560-7581.

686 Mansell, E. R., 2010: On Sedimentation and Advection in Multimoment Bulk Microphysics. *J.*
687 *Atmos. Sci.*, **67**, 3084-3094.

688 Meyers, M. P., R. L. Walko, J. Y. Harrington, and W. R. Cotton, 1997: New RAMS cloud
689 microphysics parameterization. Part II: The two-moment scheme. *Atmos. Res.*, **45**, 3-39.

690 Milbrandt, J., and M. Yau, 2005a: A multimoment bulk microphysics parameterization. A
691 multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral
692 shape parameter. *J. Atmos. Sci.*, **62**, 3051-3064.

693 ———, 2005b: Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**,
694 3065-3081.

695 ———, 2006a: A multimoment bulk microphysics parameterization. Part III: Control simulation of a
696 hailstorm. *J. Atmos. Sci.*, **63**, 3114-3136.

697 ———, 2006b: A multimoment bulk microphysics parameterization. Part IV: Sensitivity experiments.

698 *J. Atmos. Sci.*, **63**, 3137-3159.

699 Milbrandt, J. A., and R. McTaggart-Cowan, 2010: Sedimentation-Induced Errors in Bulk
700 Microphysics Schemes. *J. Atmos. Sci.*, **67**, 3931-3948.

701 Moore, J. T., and J. P. Pino, 1990: An interactive method for estimating maximum hailstone size
702 from forecast soundings. *Wea. Forecasting*, **5**, 508-525.

703 Morrison, H., and A. Gettelman, 2008: A New Two-Moment Bulk Stratiform Cloud Microphysics
704 Scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and
705 Numerical Tests. *J. Climate*, **21**, 3642-3659.

706 Morrison, H., J. A. Curry, and V. I. Khvorostyanov, 2005: A New Double-Moment Microphysics
707 Parameterization for Application in Cloud and Climate Models. Part I: Description. *J. Atmos.*
708 *Sci.*, **62**, 1665-1677.

709 Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of Cloud Microphysics on the
710 Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of
711 One- and Two-Moment Schemes. *Mon. Wea. Rev.*, **137**, 991-1007.

712 Morrison, H., and J. Milbrandt, 2010: Comparison of Two-Moment Bulk Microphysics Schemes in
713 Idealized Supercell Thunderstorm Simulations. *Mon. Wea. Rev.*, **139**, 1103-1130.

714 Nieto, R., and Coauthors, 2005: Climatological Features of Cutoff Low Systems in the Northern
715 Hemisphere. *J. Climate*, **18**, 3085-3103.

716 Nieto, R., M. Sprenger, H. Wernli, R. M. Trigo, and L. Gimeno, 2008: Identification and
717 climatology of cut-off lows near the tropopause. *Ann. N. Y. Acad. Sci.*, **1146**, 256-290.

- 718 Noppel, H., U. Blahak, A. Seifert, and K. D. Beheng, 2010: Simulations of a hailstorm and the
719 impact of CCN using an advanced two-moment cloud microphysical scheme. *Atmos. Res.*,
720 **96**, 286-301.
- 721 Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. C. Kolodziej, and J. J.
722 Gourley, 2009: The severe hazards analysis and verification experiment. *Bull. Ameri. Meteor.*
723 *Soc.*, **90**, 1519.
- 724 Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from
725 high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78-97.
- 726 Seifert, A., A. Khain, A. Pokrovsky, and K. D. Beheng, 2006: A comparison of spectral bin and two-
727 moment bulk mixed-phase cloud microphysics. *Atmos. Res.*, **80**, 46-66.
- 728 Snook, N., Y. Jung, J. Brotzge, B. Putnam, and M. Xue, 2016: Prediction and Ensemble Forecast
729 Verification of Hail in the Supercell Storms of 20 May 2013. *Wea. Forecasting*, **31**, 811-825.
- 730 Tao, S. Y., 1980: *Heavy Rain in China*, 224 pp, *Science Press*, Beijing, China. (in Chinese)
- 731 Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation
732 using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis.
733 *Mon. Wea. Rev.*, **132**, 519-542.
- 734 Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter
735 Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a
736 New Snow Parameterization. *Mon. Wea. Rev.*, **136**, 5095-5115.
- 737 Van Weverberg, K., A. M. Vogelmann, H. Morrison, and J. A. Milbrandt, 2012: Sensitivity of

738 Idealized Squall-Line Simulations to the Level of Complexity Used in Two-Moment Bulk
739 Microphysics Schemes. *Mon. Wea. Rev.*, **140**, 1883-1907.

740 Walko, R. L., W. R. Cotton, M. Meyers, and J. Harrington, 1995: New RAMS cloud microphysics
741 parameterization part I: the single-moment scheme. *Atmos. Res.*, **38**, 29-62.

742 Weisman, M. L., and J. B. Klemp, 1984: The structure and classification of numerically simulated
743 convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479-2498.

744 Witt, A., M. D. Eilts, G. J. Stumpf, J. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998: An
745 enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286-303.

746 Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System
747 (ARPS) – A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II:
748 Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161-193.

749 Xue, M., K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: Advanced Regional
750 Prediction System (ARPS) version 4.0 user's guide. *Center for Analysis and Prediction of*
751 *Storms, University of Oklahoma* [Available at <http://www.caps.ou.edu/ARPS>], 380pp.

752 Xue, M., D. Wang, J. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional
753 Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation.
754 *Meteor. Atmos. Phys.*, **82**, 139-170.

755 Xue, M., and Coauthors, 2001: The Advanced Regional Prediction System (ARPS) – A multi-scale
756 nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and
757 applications. *Meteor. Atmos. Phys.*, **76**, 143-165.

758 Zhang, C., Q. Zhang, and Y. Wang, 2008: Climatology of Hail in China: 1961-2005. *J. Appl. Meteor.*

759 *Climatol.*, **47**, 795-804.

760

761

762

763

TABLE 1. List of 3-km and 1-km simulations

Experiment	Description
CNTL	Three-moment; full version of scheme
DiagA	Two-moment; diagnosis of relation for $\alpha_x = f(D_{mx}), x \in (r, i, s, g, h)$
FixA	Two-moment; fixed $\alpha_x = 0$
Single	One-moment; $\alpha_x = 0, N_{ic} = 1 \times 10^8 m^{-3}, N_{0r} = 8 \times 10^6 m^{-4}$ $N_{0s} = 3 \times 10^6 m^{-4}, N_{0g} = 4 \times 10^5 m^{-4}, N_{0h} = 4 \times 10^4 m^{-4}$

764

765

766 List of figures

767 Fig. 1. Twenty-four-hour reports of severe weather in eastern China, starting from 2715 UTC 28
768 April 2015. Provinces of Shandong and Jiangsu are labeled in red. The open black circles
769 indicate the locations of six operational S-band radars at Jinan, Xuzhou, Bengbu, Yancheng,
770 Nanjing, and Nantong. The large gray dashed circles denote the 230-km range ring for each
771 radar. The red “×” denotes the location of the extracted sounding shown in Fig. 4. The base
772 map is obtained from Severe Weather Report Maps of Chinese Meteorological Society.

773 Fig. 2. Composite (column-maximum) reflectivity fields of operational radars from 0900 to 1400
774 UTC at 1-h interval. Provinces of Shandong, Jiangsu, and Anhui are labeled in black. The
775 thick straight line in (c) marks the rough location of vertical cross sections presented in Fig.
776 13 and 14.

777 Fig. 3. Synoptic features of (a) 200 hPa, (b) 500 hPa, (c) 850 hPa, and (d) 1000 hPa at 0600 UTC 28
778 April 2015, showing wind barbs (one full barb denotes 2.5 m s^{-1}), temperature (magenta
779 dashed contours, with interval of 4°C), geopotential height (solid black contours, gpm). The
780 shadings in (a) and (b) denote the horizontal wind speed (m s^{-1}), and that in (d) denotes
781 convective available energy (J kg^{-1}). The bold solid brown lines in (a) and (b), (c) and (d)
782 indicate trough lines and shear lines, respectively. Jiangsu Province is outlined by red solid

783 line in (a), and the blue capital C denotes the COLs at 500 hPa and cyclonic circulations at
784 850 and 1000 hPa. The maps are drawn from NCEP GFS final analysis (FNL) data.

785 Fig. 4. Skew-T plot of a sounding extracted from NCEP FNL analysis data at (34°N, 117°E) at 0600
786 UTC 28 April 2015 near Xuzhou.

787 Fig. 5. Model domains of 3 km (outer black box) and 1 km (inner red box) grid spacings for the
788 hailstorm simulations. Terrain elevation (m) is plotted in color. Cities of Jinan, Changzhou,
789 Wuxi, and Suzhou where hailstones are reported are labeled in black. The city of Yizheng is
790 labeled in magenta, where maximum hail size is registered of around 10 cm. Provinces of
791 Shandong, Jiangsu, and Anhui are labeled in red.

792 Fig. 6. Composite (column-maximum) reflectivity fields from simulations using different
793 microphysics schemes and operational radar observations, namely (a) Single, (b) FixA, (c)
794 DiagA, (d) CNTL, and (e) radar observations at 0900 UTC 28 April 2015.

795 Fig. 7. As Fig. 6, but at 1400 UTC 28 April 2015.

796 Fig. 8. Simulated MESH with one-, two- and three-moment microphysics schemes, namely (a)
797 Single, (b) FixA, (c) DiagA, (d) CNTL, and (e) observed MESH derived from WSR-98D
798 radar observations. The MESH fields are created as a composite from observed/simulated
799 data between 0600 and 1600 UTC at 5-minute intervals. Locations of the cities of
800 Changzhou, Wuxi, and Suzhou, where hail was reported are represented by the black “×”.

801 The city of Yizheng is labeled in magenta, where maximum hail size is registered of around
802 10 cm.

803 Fig. 9. Fraction skill scores of MESH from 0600 to 1600 UTC for the experiments, i.e., Single,
804 FixA, DiagA, and CNTL, with MESH exceeding (a) 30 mm and (b) 40 mm, over
805 neighborhood radii of 10, 30, 50, and 60 km. FSS_{useful} value is also shown.

806 Fig. 10. Surface accumulated solid precipitation in water-equivalent depth (mm) between 0600 and
807 1600 UTC from simulations (a) Single, (b) FixA, (c) DiagA, and (d) CNTL.

808 Fig. 11. Surface accumulated hail number concentration (SAHNC) between 0600 and 1600 UTC at
809 60-second intervals on 28 April 2015, derived from simulations (a) Single, (b) FixA, (c)
810 DiagA, and (d) CNTL, with hail diameter larger than 30 mm. SAHNC is in base-10
811 logarithmic scale, i.e., $\log_{10}(\text{SAHNC})$ (m^{-2}).

812 Fig. 12. As Fig. 11, but for hail diameter exceeding 40 mm.

813 Fig. 13. Vertical cross sections of the mass content (black contours; g m^{-3}) and total number
814 concentration of hail in base-10 logarithmic scale, i.e., $\log_{10}(N_{th})$ (shaded; m^{-3}), through the
815 hail mass cores of the primary hail-producing storm A at 1100 UTC (Fig. 2c), for
816 experiments of (a) Single, (b) FixA, (c) DiagA, and (d) CNTL. The location of the vertical
817 slice is marked by a solid black line in Fig. 2c, with a slight north-south shift in each
818 experiment so as to cut through the maximum hail mass core. The contour of 0°C in magenta
819 color is also overlaid.

820 Fig. 14. As Fig. 13, but for D_{max} (shaded; mm). The reflectivity contours of 30, 40, 50, and 60 dBZ
821 (black) and temperature contours of 0°C (magenta) are also overlaid. The black circles in the
822 panels refer to the approximate locations of the hail size distributions for the sensitivity
823 experiments in Fig. 15.

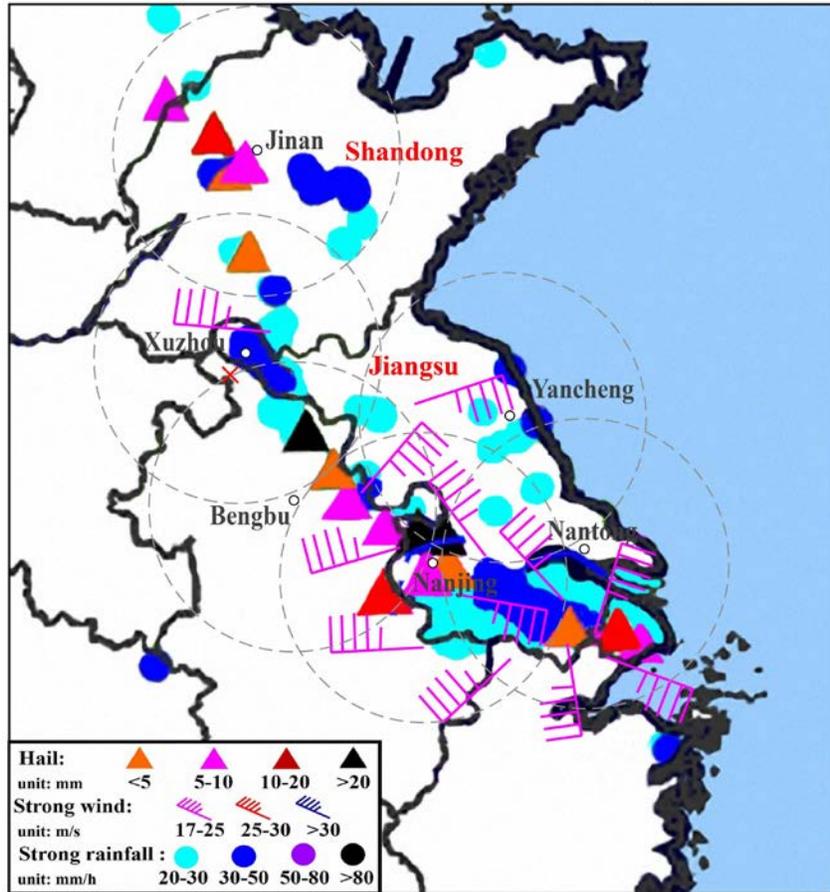
824 Fig. 15. Hail size distributions at 1100 UTC (a) from Single, FixA, DiagA, and CNTL at $z=251$ m
825 corresponding to the locations indicated in each panel of Fig. 14, and (b) from various levels
826 of CNTL, within layers between 25 m and 7 km corresponding to the black dot locations in
827 Fig. 14d.

828 Fig. 16. Profiles of the shape parameters of experiments of Single, FixA, DiagA, and CNTL. The
829 profiles are horizontal averages within the simulated hailstorms over the 1-km domain of the
830 simulations between 0600 and 1600 UTC at 1-min interval. The shaded areas denote the one
831 standard deviation from the average profiles.

832 Fig. 17. Time series of microphysical tendency terms in the hail mass equation for experiments
833 Single, FixA, DiagA, and CNTL, between 0600 and 1500. The line color denotes the
834 experiments while line pattern denotes microphysical processes responsible for hail growth.
835 The microphysical processes include hail collection of cloud water (colqch), hail collection
836 of rain (colqrh), and hail melting to rain (meltqh). Hail production from other processes,
837 including hail collection of snow or ice, are minimal and are not plotted.

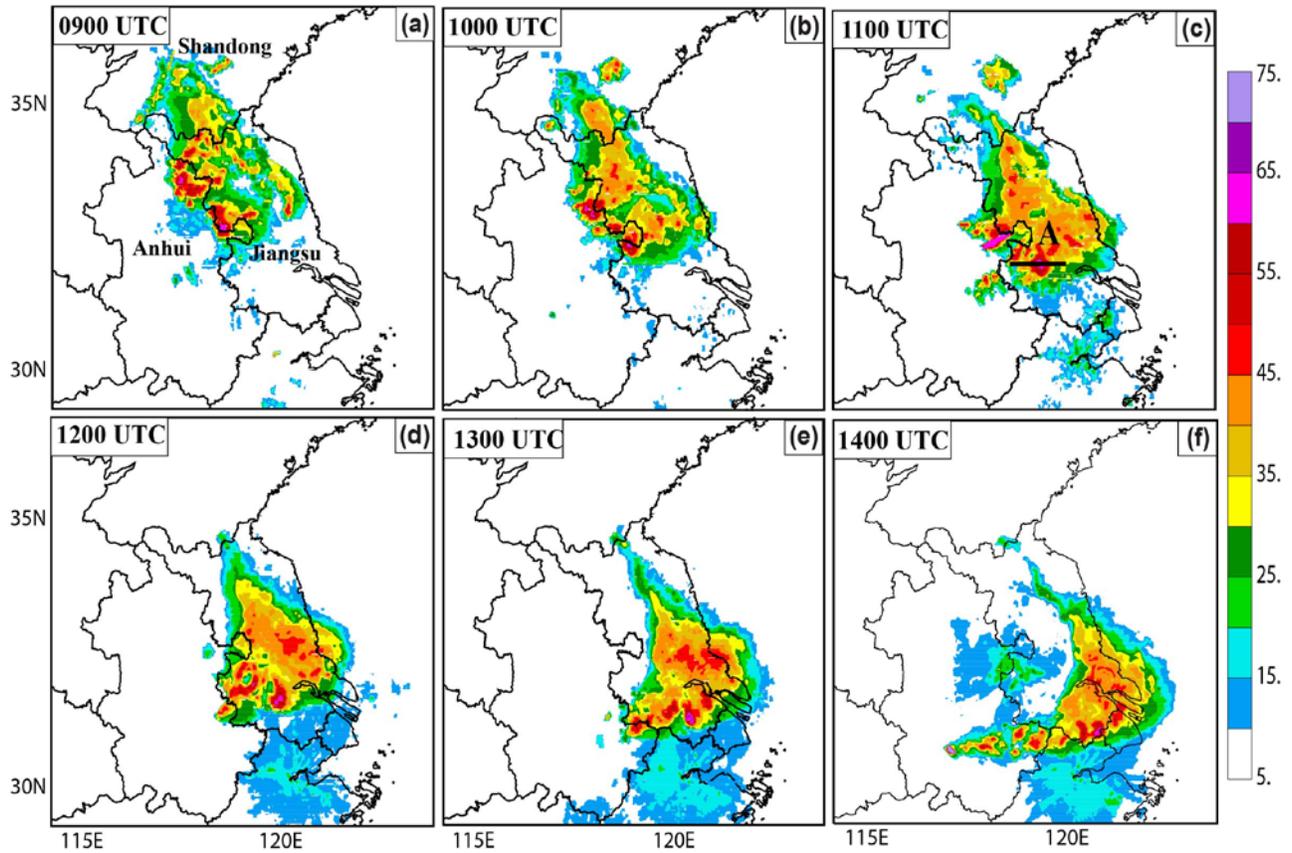
838 Fig. 18. Vertical profiles of the mean hail production rates for hail collection of cloud (colqch), hail
839 collection of rain (colqrh), and hail melting to rain (meltqh) for experiments Single, FixA,
840 DiagA, and CNTL. The profiles are averaged horizontally over points within the hailstorm
841 between 0600 and 1500 UTC at 1-min intervals on April 28 2015.
842

National Strong Convective Weather Observation
(accumulated from 2715~2815 UTC)



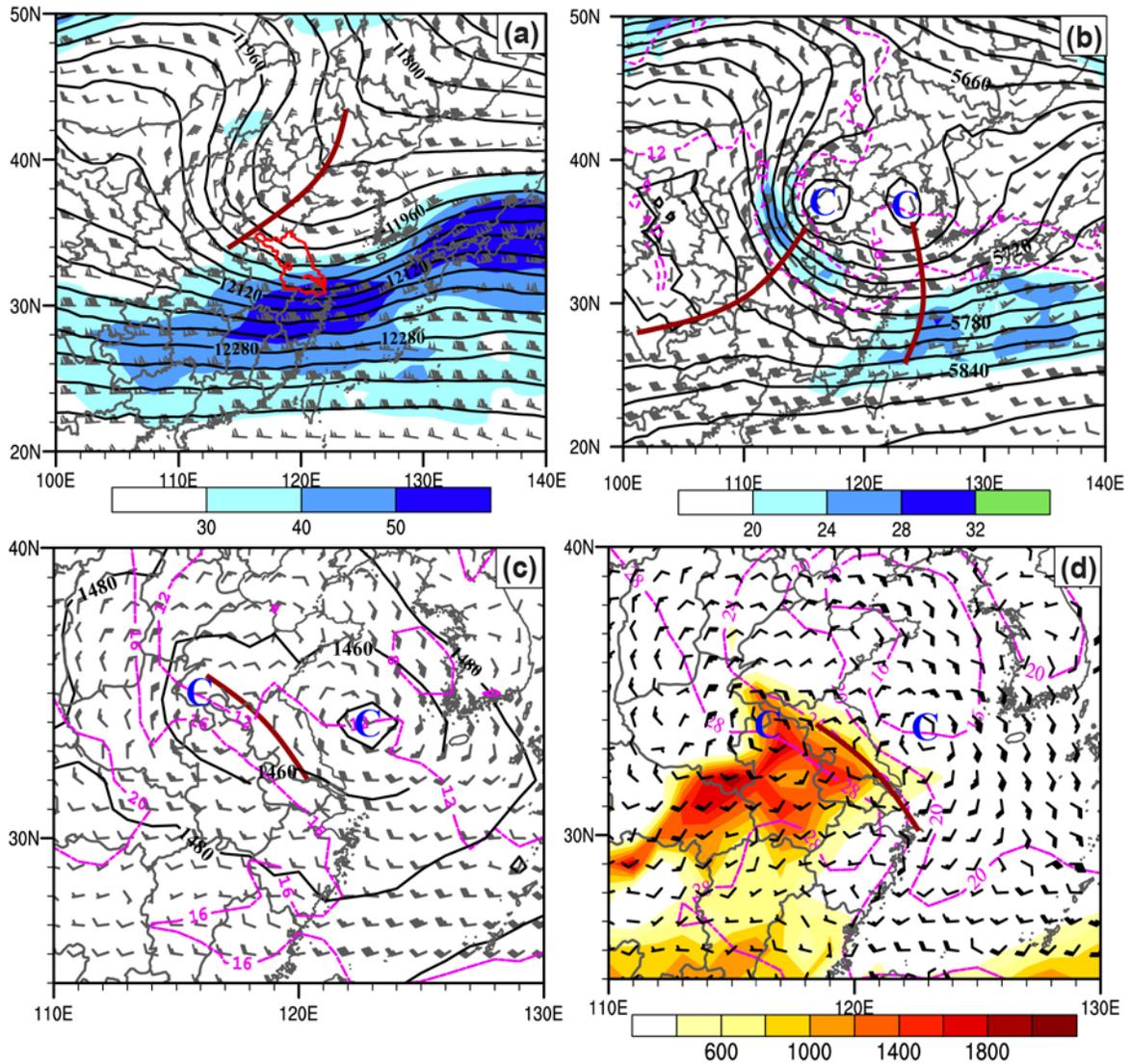
843

844 Fig. 1. Twenty-four-hour reports of severe weather in eastern China, starting from 2715 UTC 28
 845 April 2015. Provinces of Shandong and Jiangsu are labeled in red. The open black circles indicate
 846 the locations of six operational S-band radars at Jinan, Xuzhou, Bengbu, Yancheng, Nanjing, and
 847 Nantong. The large gray dashed circles denote the 230-km range ring for each radar. The red “×”
 848 denotes the location of the extracted sounding shown in Fig. 4. The base map is obtained from
 849 Severe Weather Report Maps of Chinese Meteorological Society.

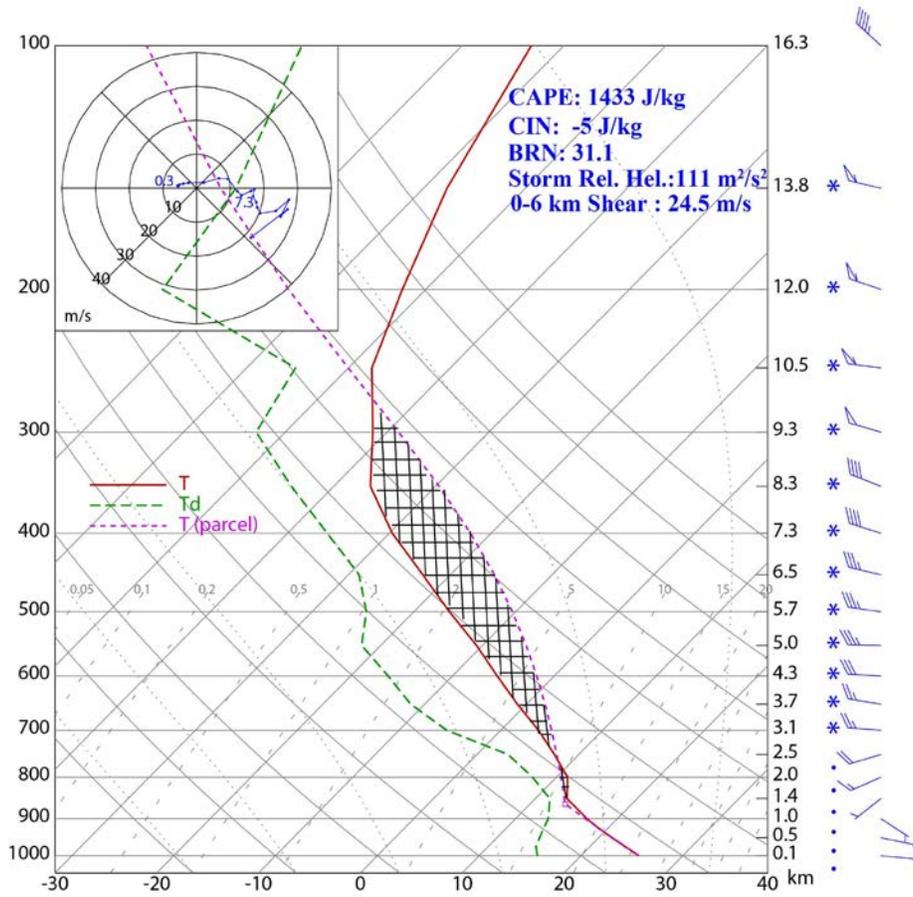


850

851 Fig. 2. Composite (column-maximum) reflectivity fields of operational radars from 0900 to 1400
 852 UTC at 1-h interval. Provinces of Shandong, Jiangsu, and Anhui are labeled in black. The thick
 853 straight line in (c) marks the rough location of vertical cross sections presented in Fig. 13 and 14.
 854

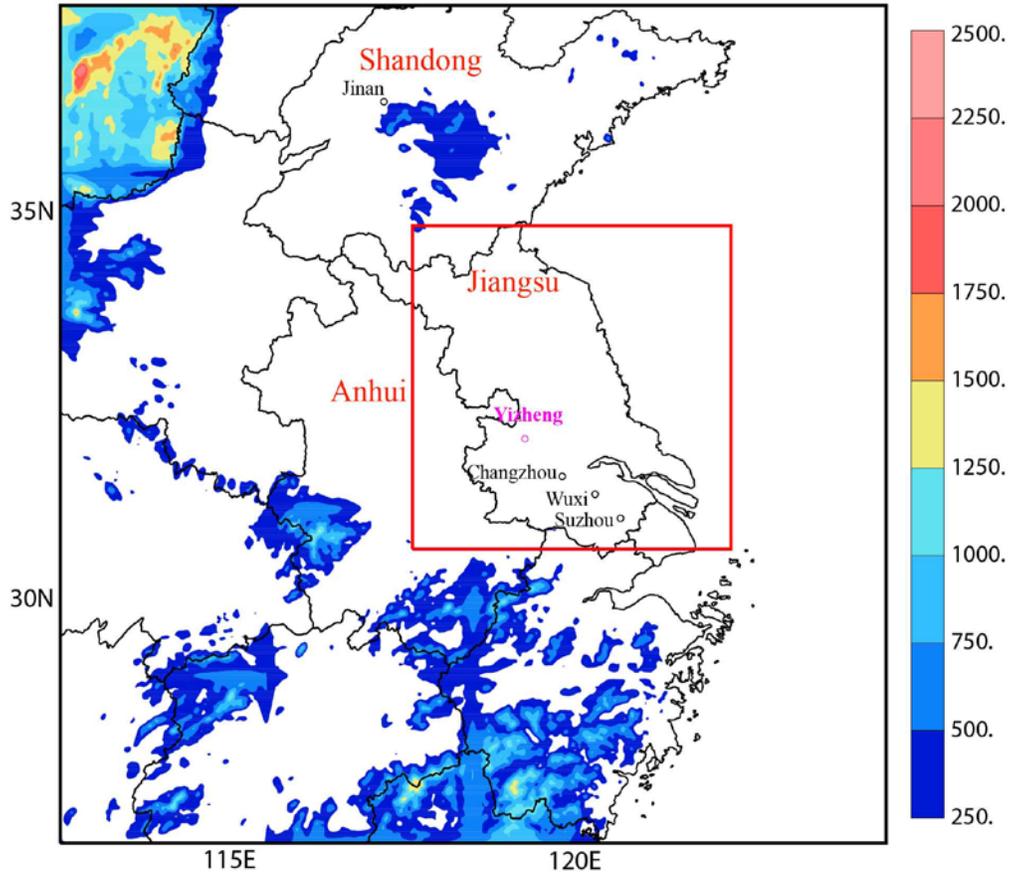


855
 856 Fig. 3. Synoptic features of (a) 200 hPa, (b) 500 hPa, (c) 850 hPa, and (d) 1000 hPa at 0600 UTC 28
 857 April 2015, showing wind barbs (one full barb denotes 2.5 m s^{-1}), temperature (magenta dashed
 858 contours, with interval of 4°C), geopotential height (solid black contours, gpm). The shadings in (a)
 859 and (b) denote the horizontal wind speed (m s^{-1}), and that in (d) denotes convective available energy
 860 (J kg^{-1}). The bold solid brown lines in (a) and (b), (c) and (d) indicate trough lines and shear lines,
 861 respectively. Jiangsu Province is outlined by red solid line in (a), and the blue capital C denotes the
 862 COLs at 500 hPa and cyclonic circulations at 850 and 1000 hPa. The maps are drawn from NCEP
 863 GFS final analysis (FNL) data.



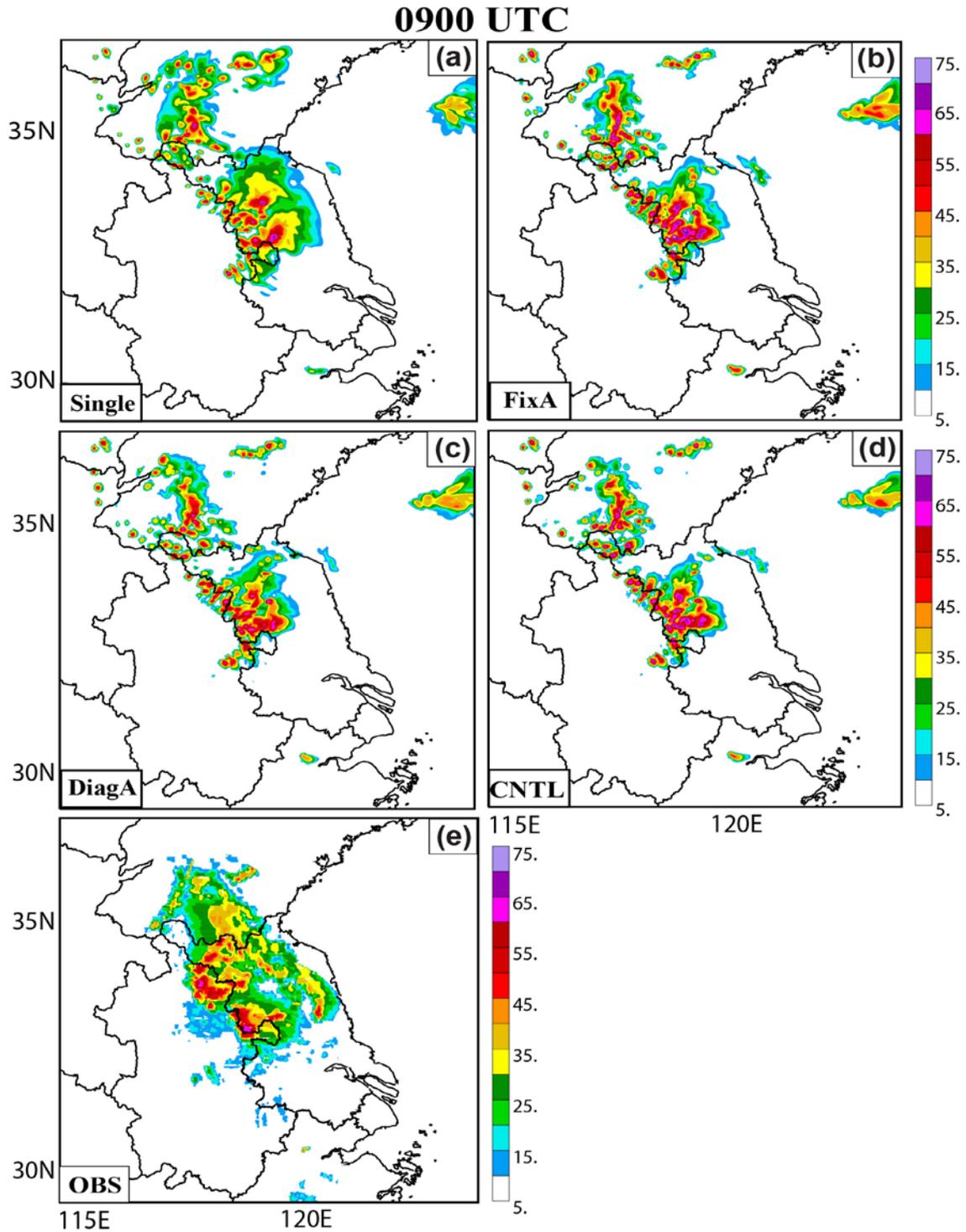
864

865 Fig. 4. Skew-T plot of a sounding extracted from NCEP FNL analysis data at (34°N, 117°E) at 0600
 866 UTC 28 April 2015 near Xuzhou.



867

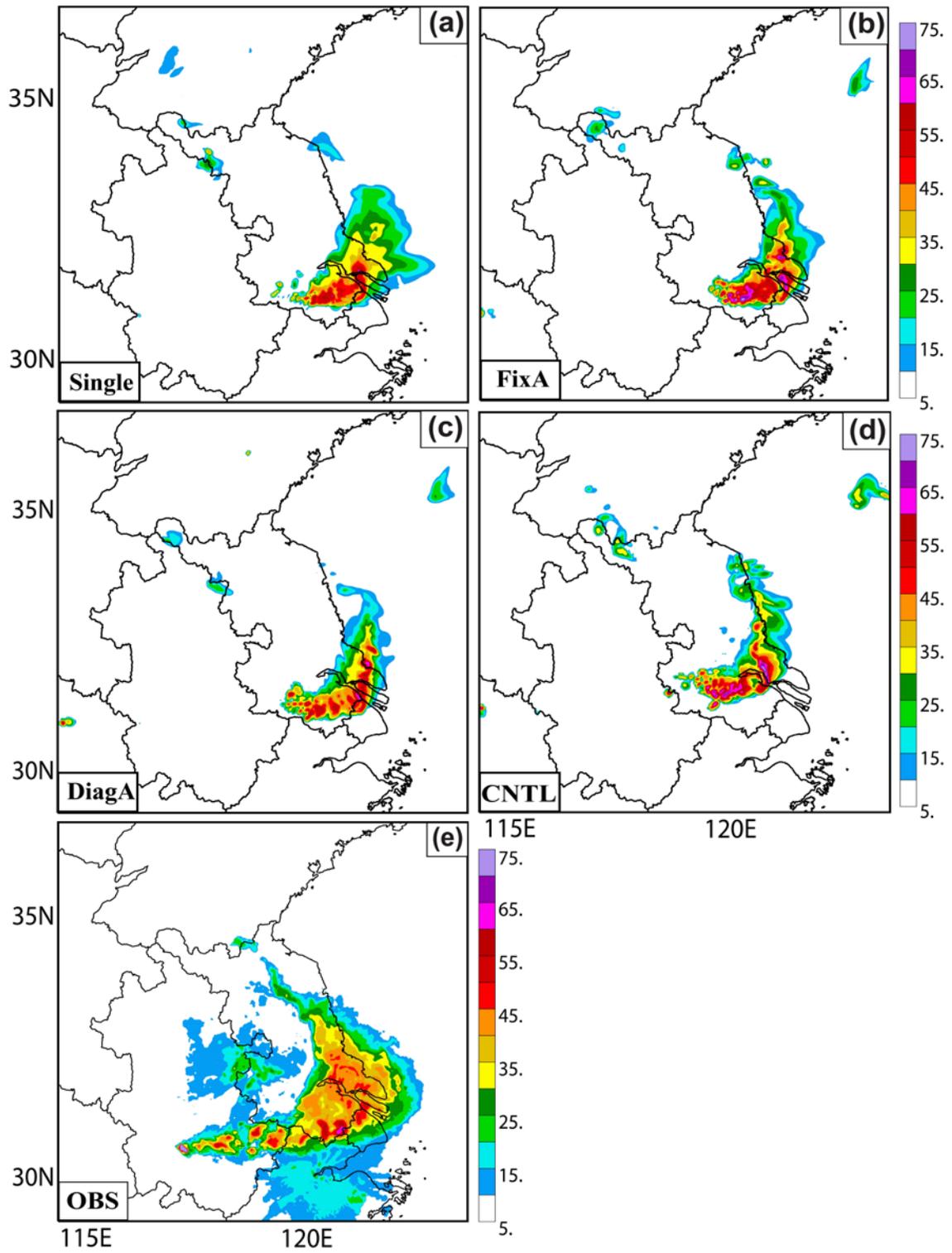
868 Fig. 5. Model domains of 3 km (outer black box) and 1 km (inner red box) grid spacings for the
 869 hailstorm simulations. Terrain elevation (m) is plotted in color. Cities of Jinan, Changzhou, Wuxi,
 870 and Suzhou where hailstones are reported are labeled in black. The city of Yizheng is labeled in
 871 magenta, where maximum hail size is registered of around 10 cm. Provinces of Shandong, Jiangsu,
 872 and Anhui are labeled in red.



873

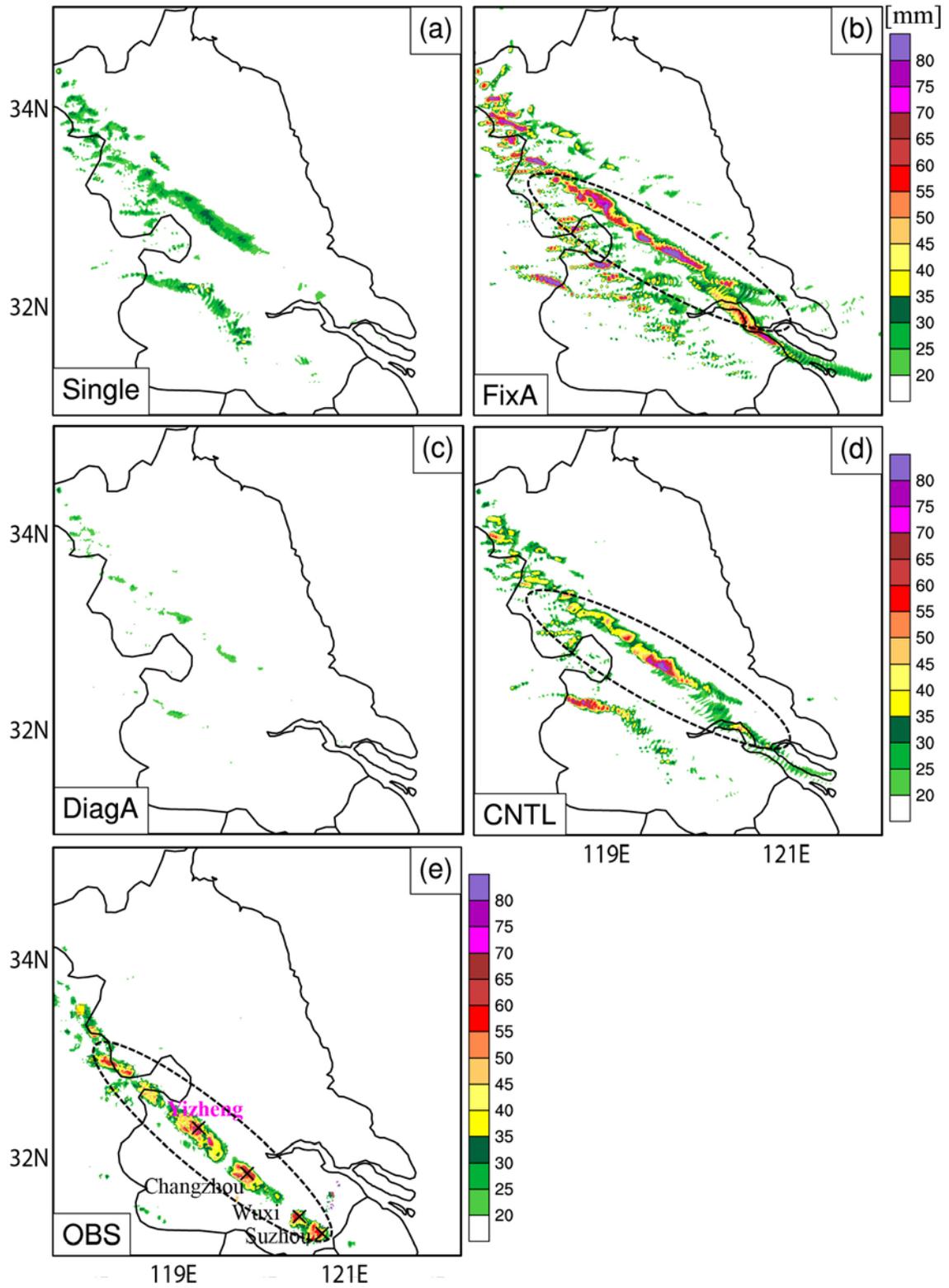
874 Fig. 6. Composite (column-maximum) reflectivity fields from simulations using different
 875 microphysics schemes and operational radar observations, namely (a) Single, (b) FixA,
 876 (c) DiagA, (d) CNTL, and (e) radar observations at 0900 UTC 28 April 2015.

1400 UTC



877

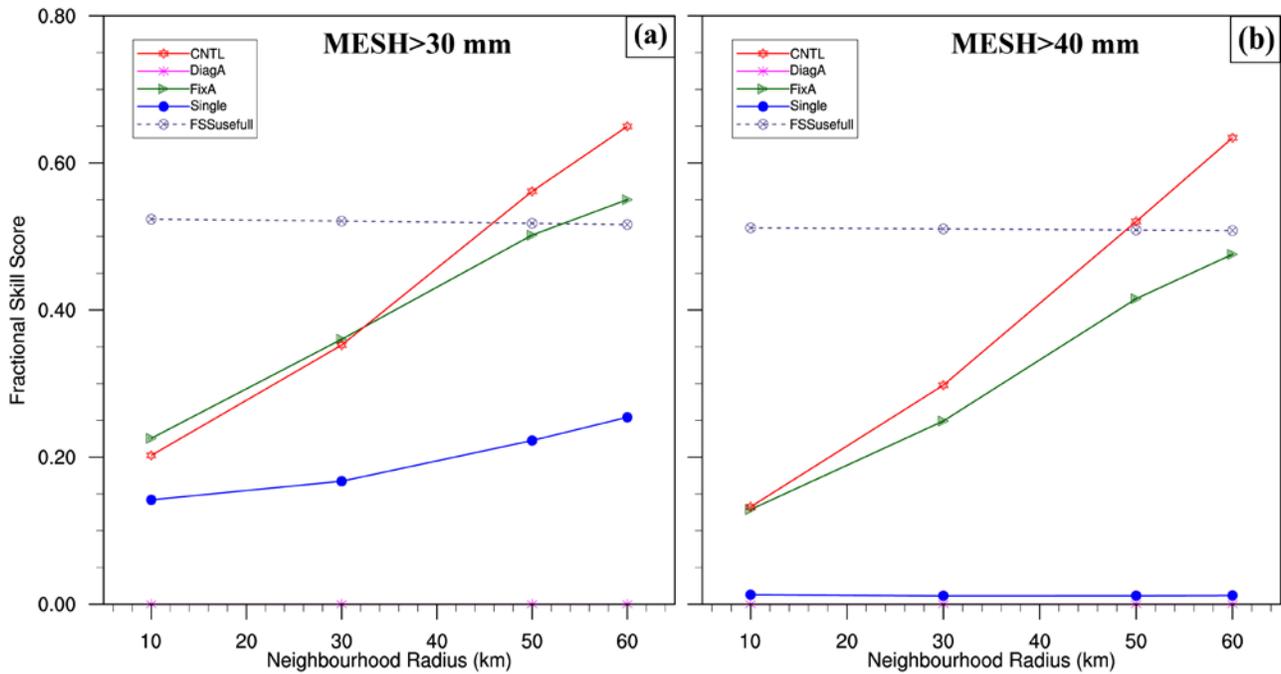
878 Fig. 7. As Fig. 6, but at 1400 UTC 28 April 2015.



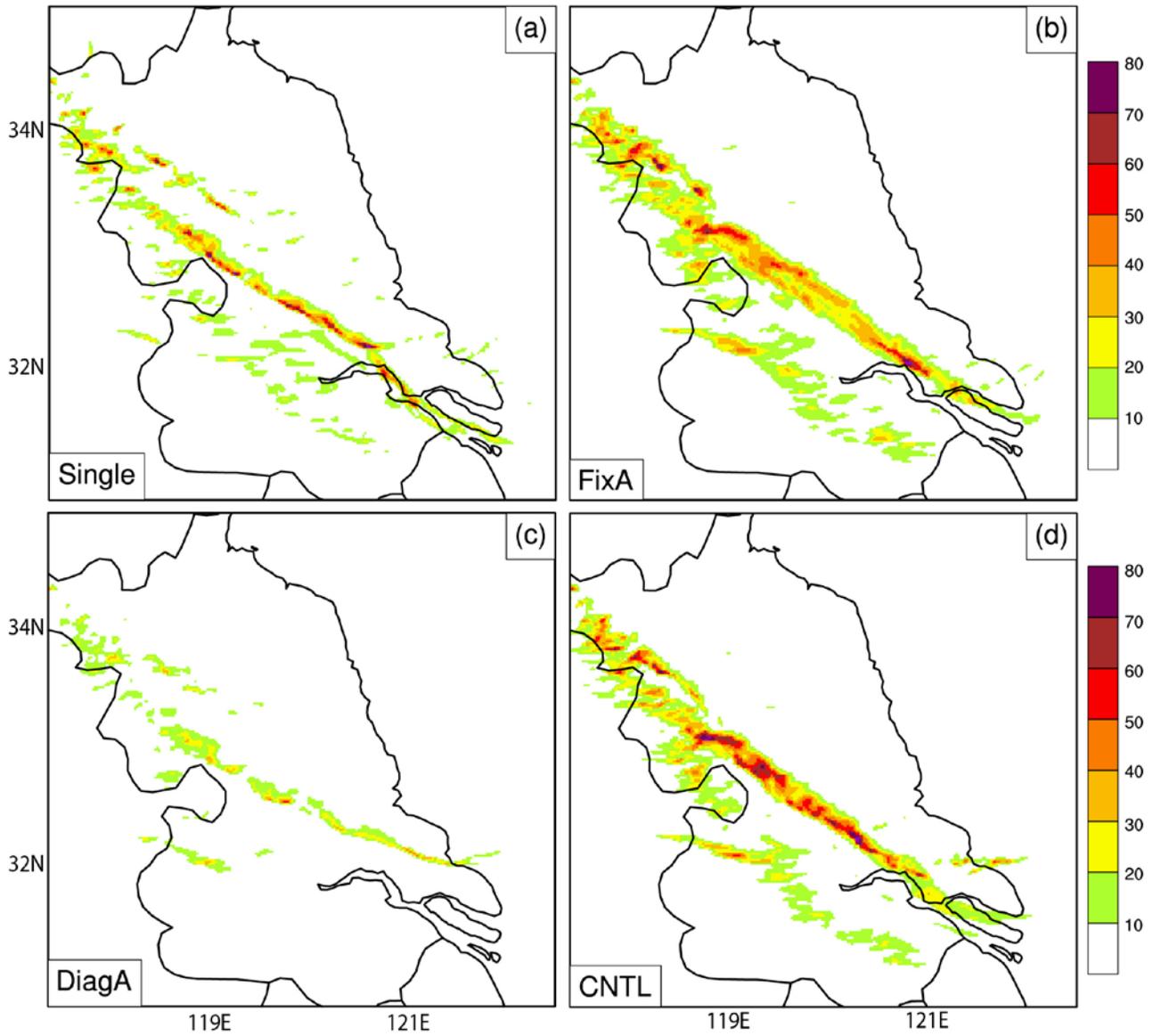
879
880

Fig. 8. Simulated MESH with one-, two- and three-moment microphysics schemes, namely (a)

881 Single, (b) FixA, (c) DiagA, (d) CNTL, and (e) observed MESH derived from WSR-98D radar
 882 observations. The MESH fields are created as a composite from observed/simulated data between
 883 0600 and 1600 UTC at 5-minute intervals. Locations of the cities of Changzhou, Wuxi, and Suzhou,
 884 where hail was reported are represented by the black “x”. The city of Yizheng is labeled in magenta,
 885 where maximum hail size is registered of around 10 cm.
 886

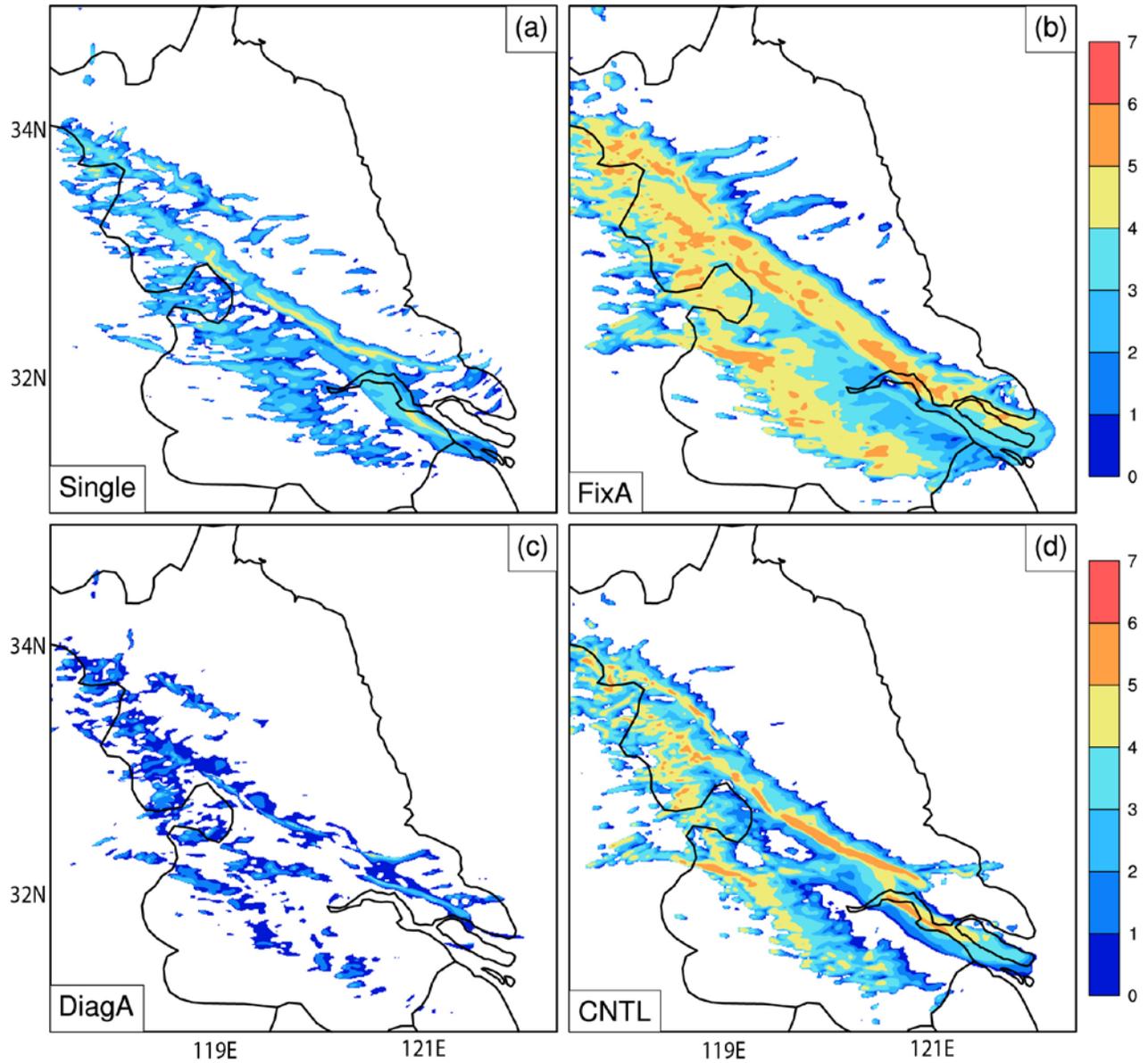


887
 888 Fig. 9. Fraction skill scores of MESH from 0600 to 1600 UTC for the experiments, i.e., Single,
 889 FixA, DiagA, and CNTL, with MESH exceeding (a) 30 mm and (b) 40 mm, over neighborhood
 890 radii of 10, 30, 50, and 60 km. FSS_{useful} value is also shown.



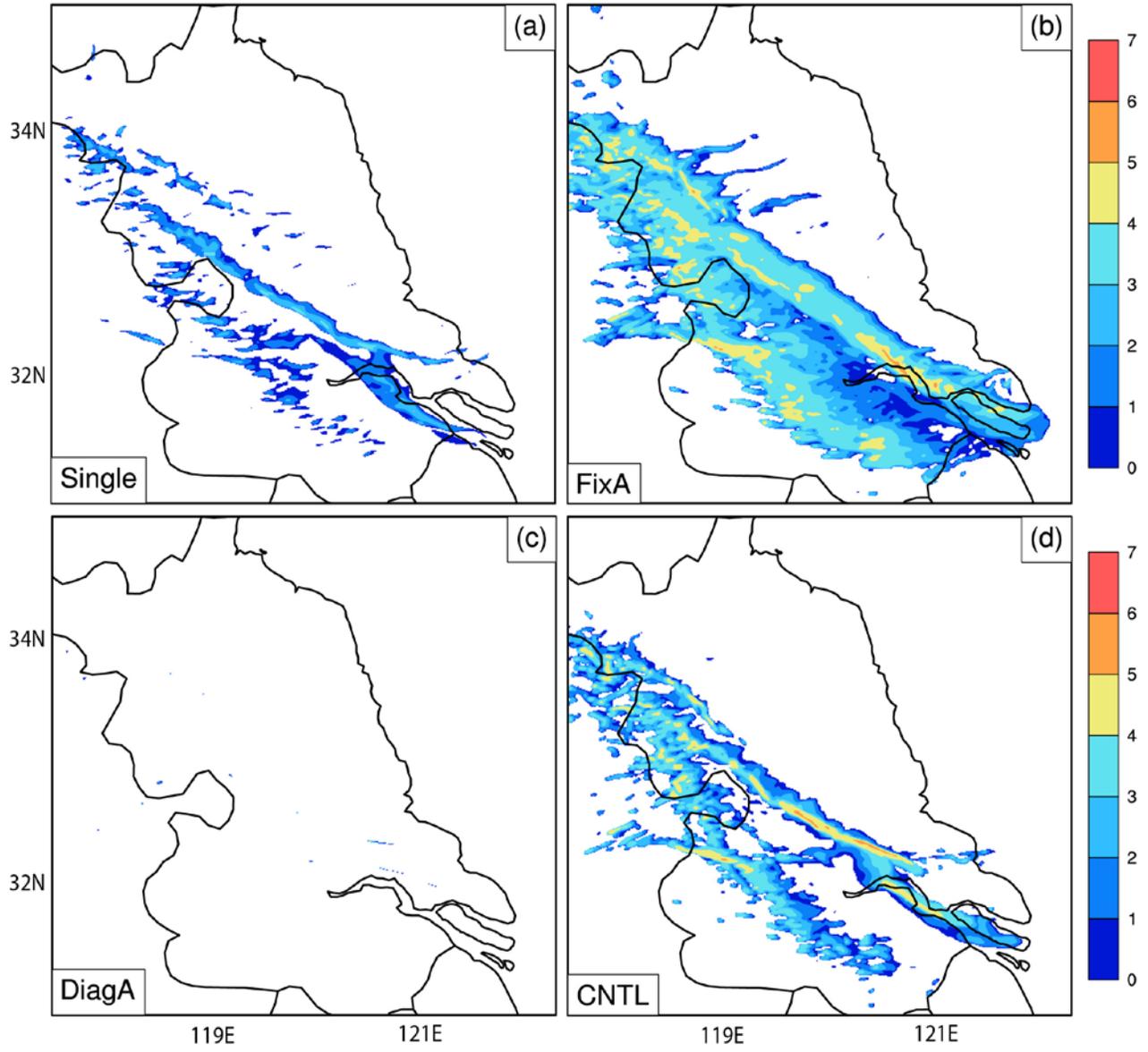
891

892 Fig. 10. Surface accumulated solid precipitation in water-equivalent depth (mm) between 0600 and
 893 1600 UTC from simulations (a) Single, (b) FixA, (c) DiagA, and (d) CNTL.



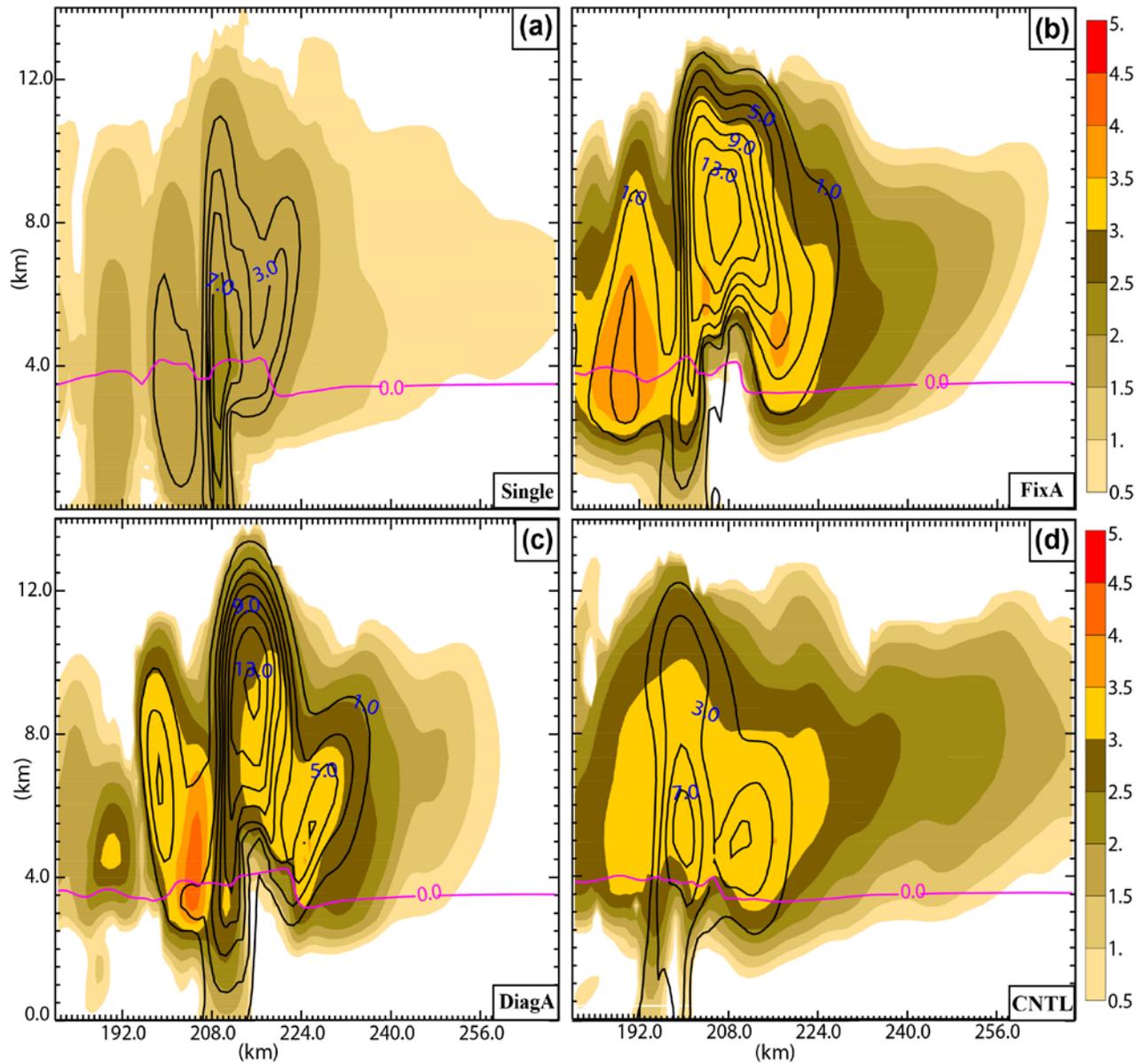
894

895 Fig. 11. Surface accumulated hail number concentration (SAHNC) between 0600 and 1600 UTC at
 896 60-second intervals on 28 April 2015, derived from simulations (a) Single, (b) FixA, (c) DiagA, and
 897 (d) CNTL, with hail diameter larger than 30 mm. SAHNC is in base-10 logarithmic scale, i.e.,
 898 $\log_{10}(\text{SAHNC})$ (m^{-2}).



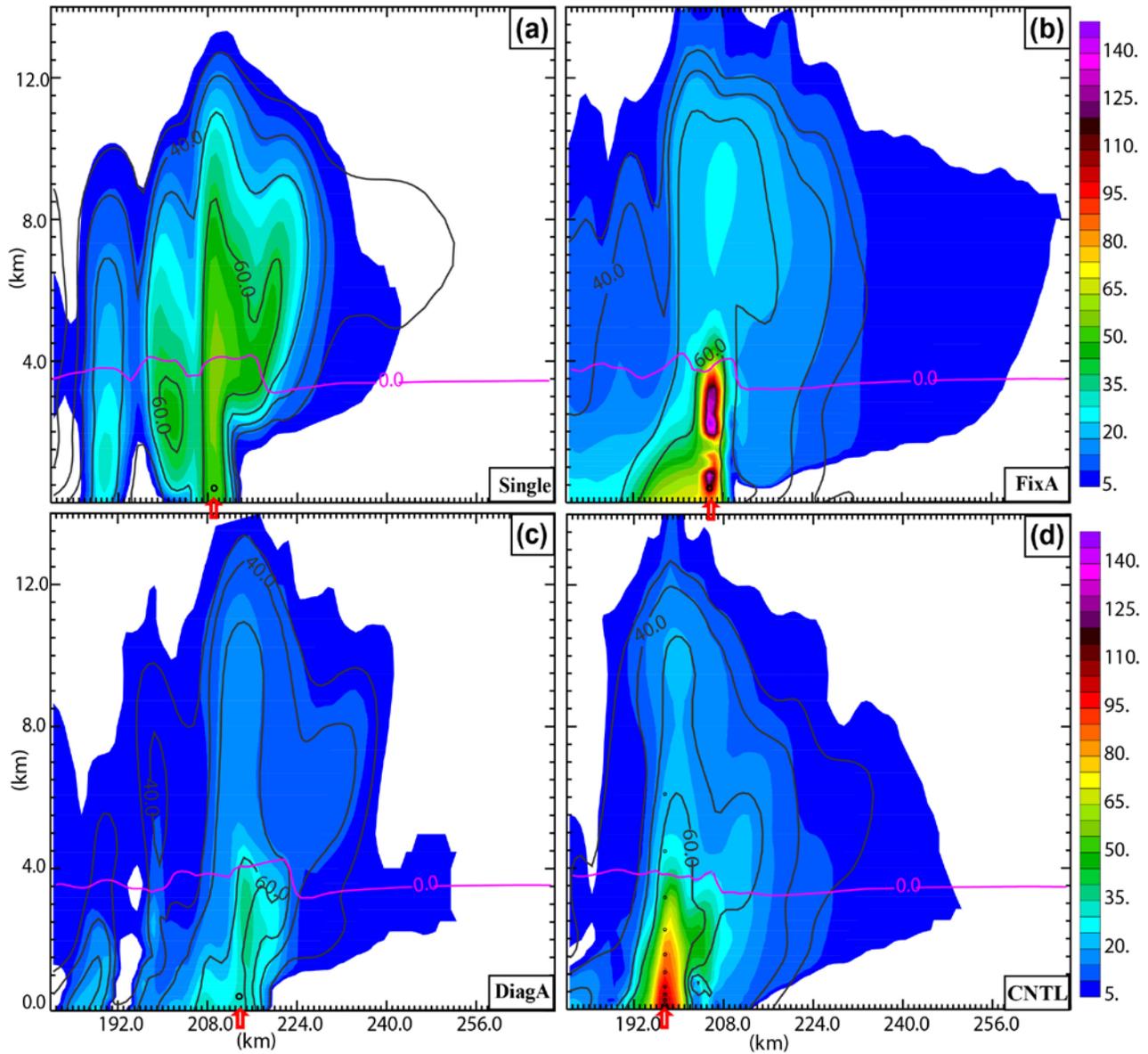
899

900 Fig. 12. As Fig. 11, but for hail diameter exceeding 40 mm.



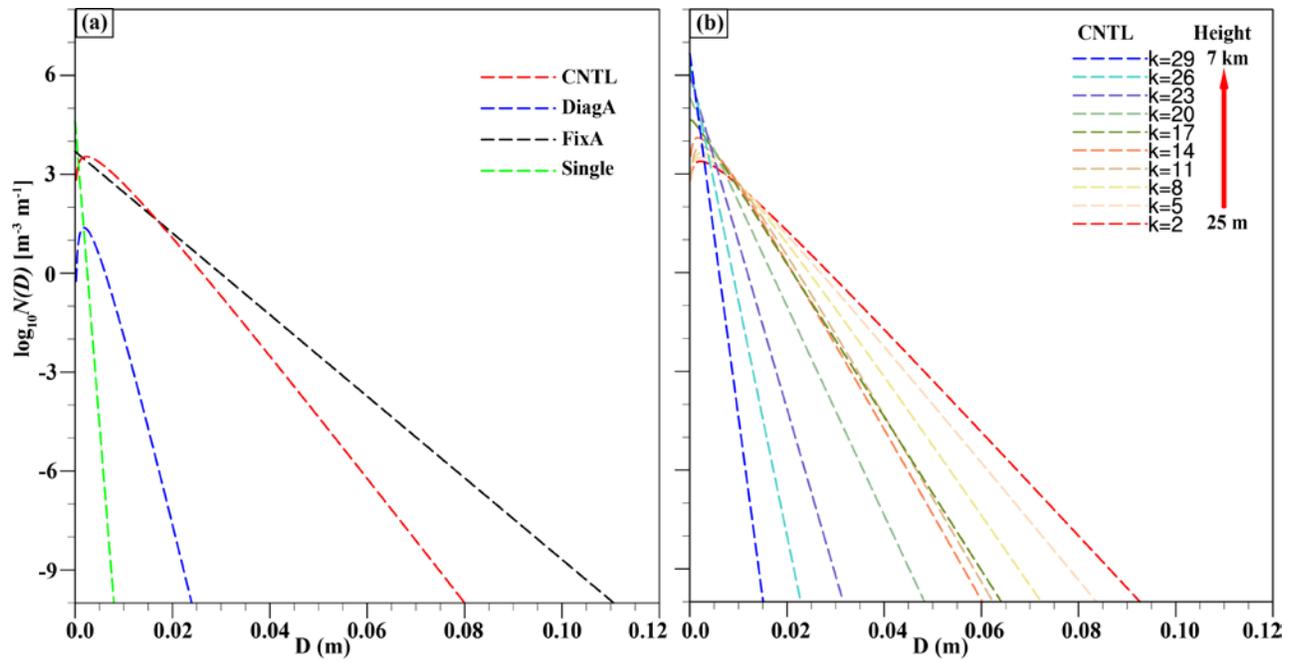
901

902 Fig. 13. Vertical cross sections of the mass content (black contours; $g\ m^{-3}$) and total number
 903 concentration of hail in base-10 logarithmic scale, i.e., $\log_{10}(N_{th})$ (shaded; m^{-3}), through the hail
 904 mass cores of the primary hail-producing storm A at 1100 UTC (Fig. 2c), for experiments of (a)
 905 Single, (b) FixA, (c) DiagA, and (d) CNTL. The location of the vertical slice is marked by a solid
 906 black line in Fig. 2c, with a slight north-south shift in each experiment so as to cut through the
 907 maximum hail mass core. The contour of $0^{\circ}C$ in magenta color is also overlaid.



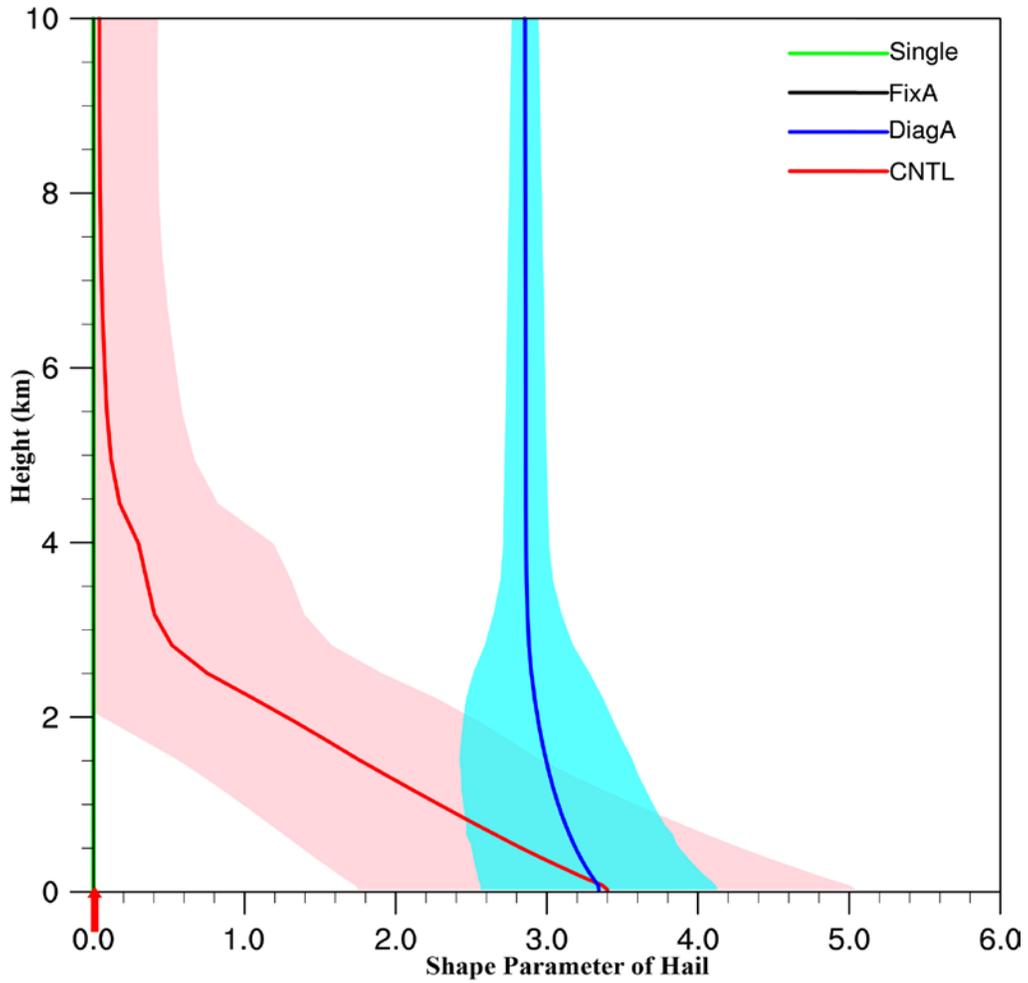
908

909 Fig. 14. As Fig. 13, but for D_{max} (shaded; mm). The reflectivity contours of 30, 40, 50, and 60 dBZ
 910 (black) and temperature contours of 0°C (magenta) are also overlaid. The black circles in the panels
 911 refer to the approximate locations of the hail size distributions for the sensitivity experiments in Fig.
 912 15.



913

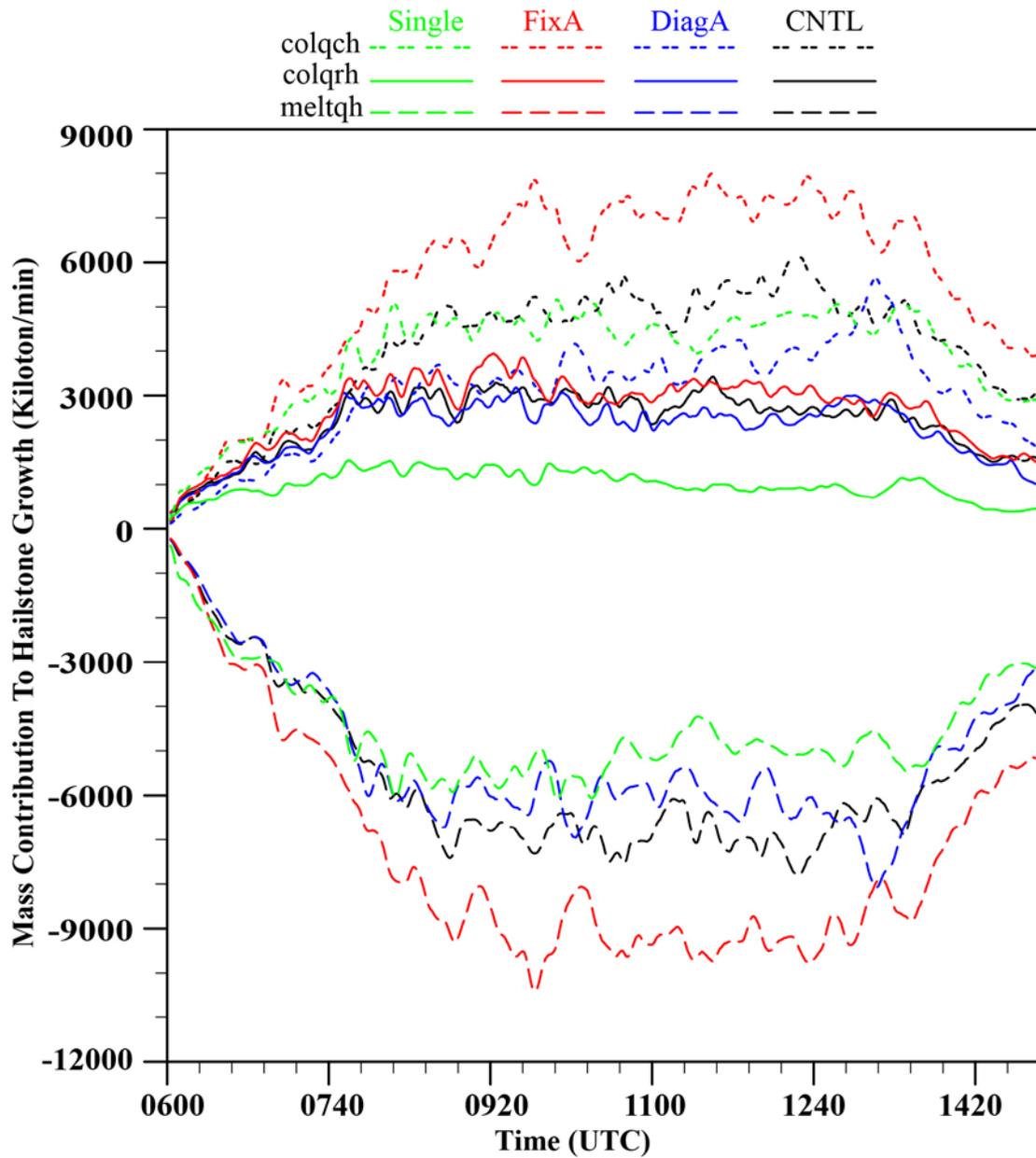
914 Fig. 15. Hail size distributions at 1100 UTC (a) from Single, FixA, DiagA, and CNTL at $z=251$ m
 915 corresponding to the locations indicated in each panel of Fig. 14, and (b) from various levels of
 916 CNTL, within layers between 25 m and 7 km corresponding to the black dot locations in Fig. 14d.



917

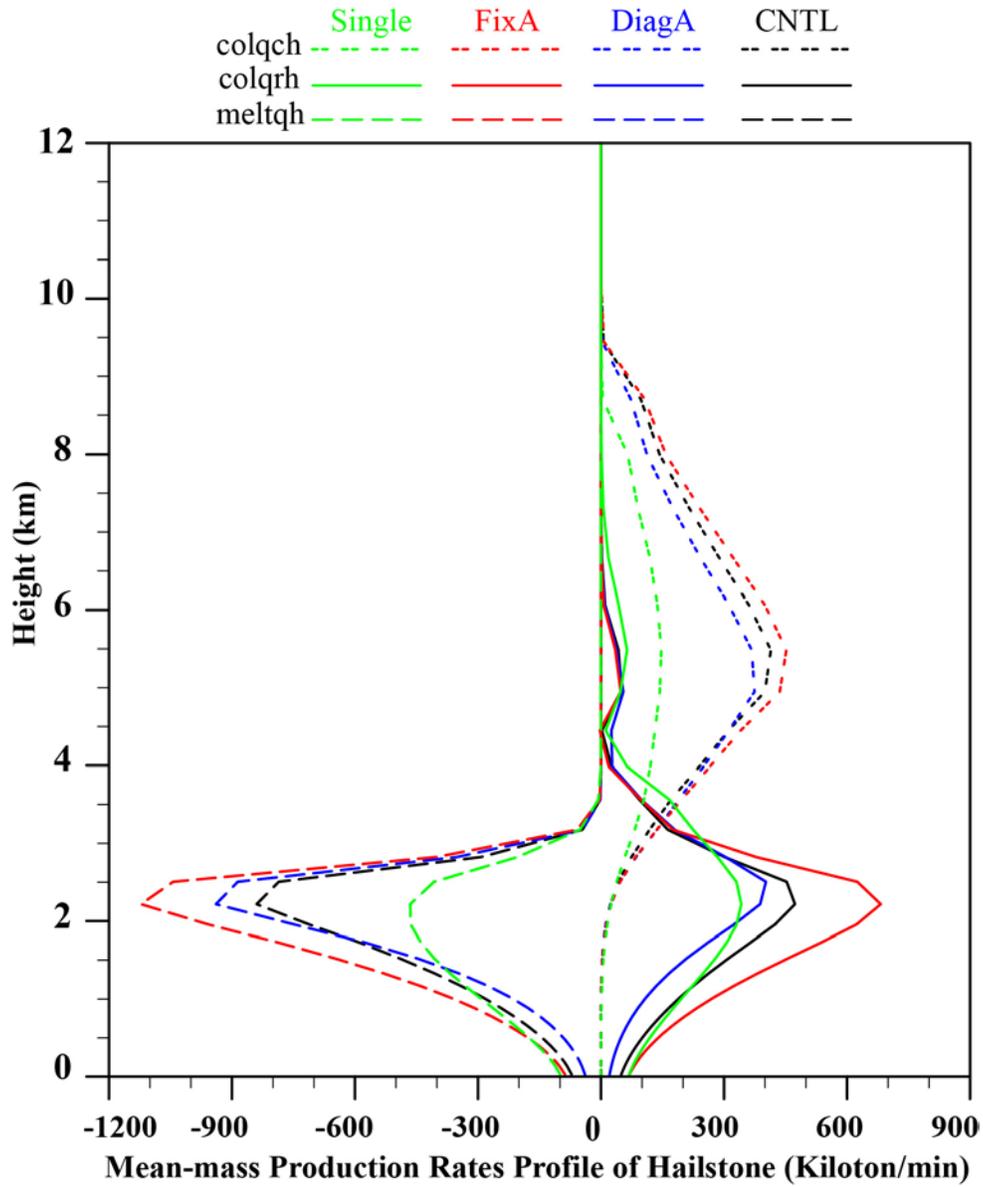
918 Fig. 16. Profiles of the shape parameters of experiments of Single, FixA, DiagA, and CNTL. The
 919 profiles are horizontal averages within the simulated hailstorms over the 1-km domain of the
 920 simulations between 0600 and 1600 UTC at 1-min interval. The shaded areas denote the one
 921 standard deviation from the average profiles.

922



923

924 Fig. 17. Time series of microphysical tendency terms in the hail mass equation for experiments
 925 Single, FixA, DiagA, and CNTL between 0600 and 1500 UTC. The line color denotes the
 926 experiments while line pattern denotes microphysical processes responsible for hail growth. The
 927 microphysical processes include hail collection of cloud water (colqch), hail collection of rain
 928 (colqrh), and hail melting to rain (meltqh). Hail production from other processes, including hail
 929 collection of snow or ice, are minimal and are not plotted.



930

931 Fig. 18. Vertical profiles of the mean hail production rates for hail collection of cloud (colqch), hail
 932 collection of rain (colqrh), and hail melting to rain (meltqh) for experiments Single, FixA, DiagA,
 933 and CNTL. The profiles are averaged horizontally over points within the hailstorm, between 0600
 934 and 1500 UTC at 1-min intervals on April 28 2015.