Supplementary materials

Table S1 WRF Physical Options

Physical options	**Scheme**
Long wave radiation	RRTM Scheme
Short wave radiation	Dudhia Scheme
Surface layer	Monin-Obukhov Scheme
Boundary layer	Yonsei University Scheme (YSU)
Urban physics	Single Layer Urban Canopy Model (SLUCM)
Land Surface	Noah Land Surface Model (Noah LSM)
Land use	MODIS 30s and 3-category urban land use

Table S2 Three datasets used as initial/boundary forcing to WRF simulations

Dataset	Spatial/Temporal resolution	Vertical levels	Data source
ERA-interim	0.75 °×0.75 °/ 6h	37	ECMWF (Europe)
NCEP Final Analysis (FNL)	1 °×1 °/ 6h	26	NCEP (USA)
Japanese 55-year Reanalysis (JRA-55)	1.25 °×1.25 °/ 6h	37	JMA (Japan)

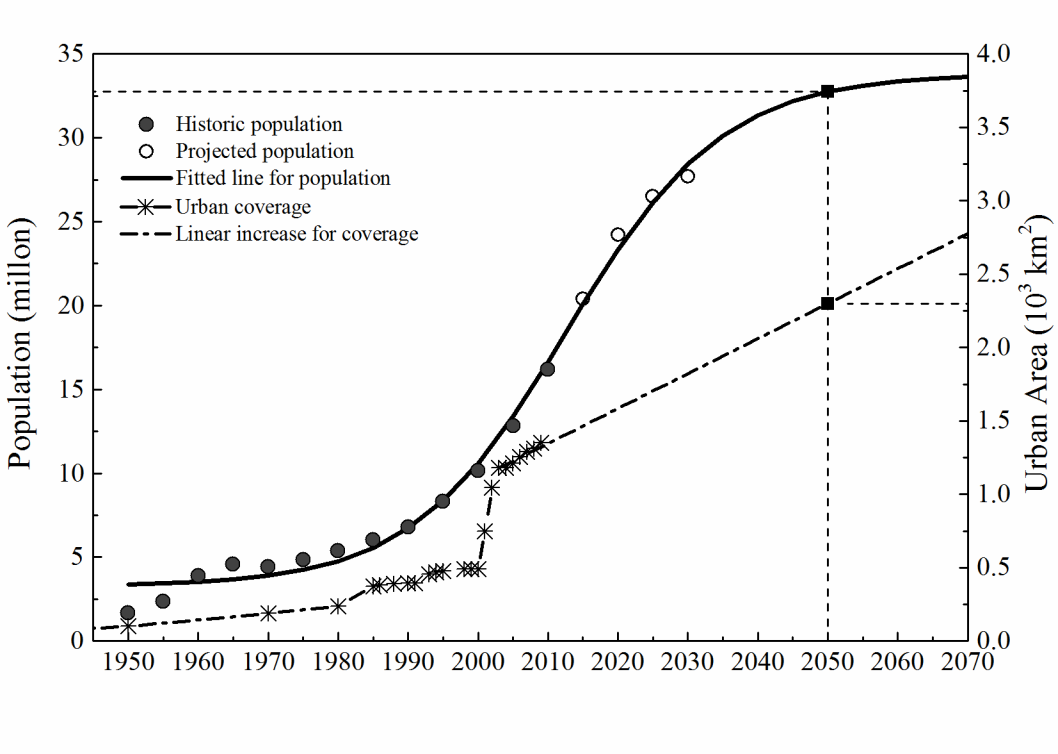

Figure S1 Temporal changes of urban population and urban coverage over Beijing metropolitan region from 1950s to 2050s. Black curve shows the fitted trend of urban population based on the logistic model.

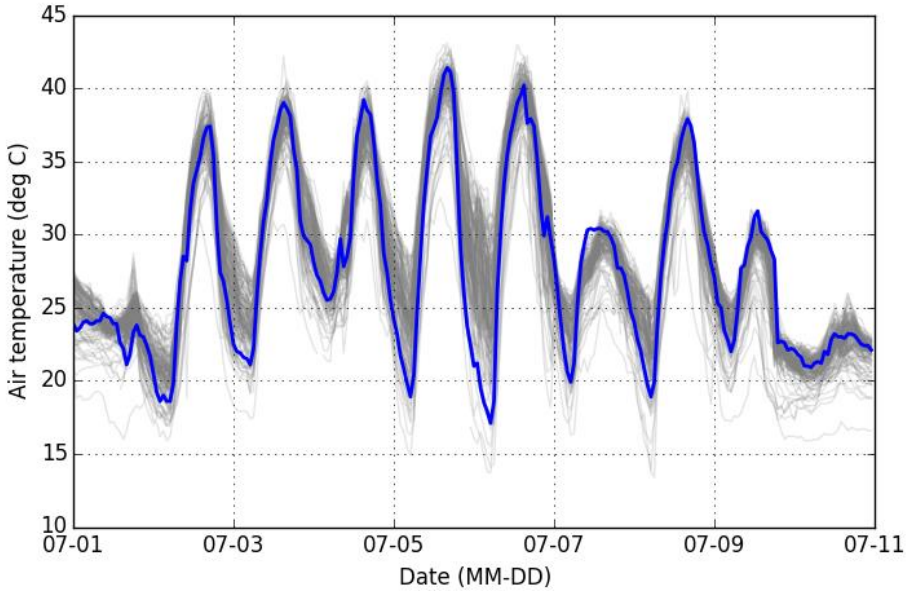

Figure S2 Time series of gauge-based 2-m air temperature (°C) for the 2010 heat wave case. Grey traces are based on all gauges over BMA, while blue is ensemble mean time series of hourly 2-m air temperature.

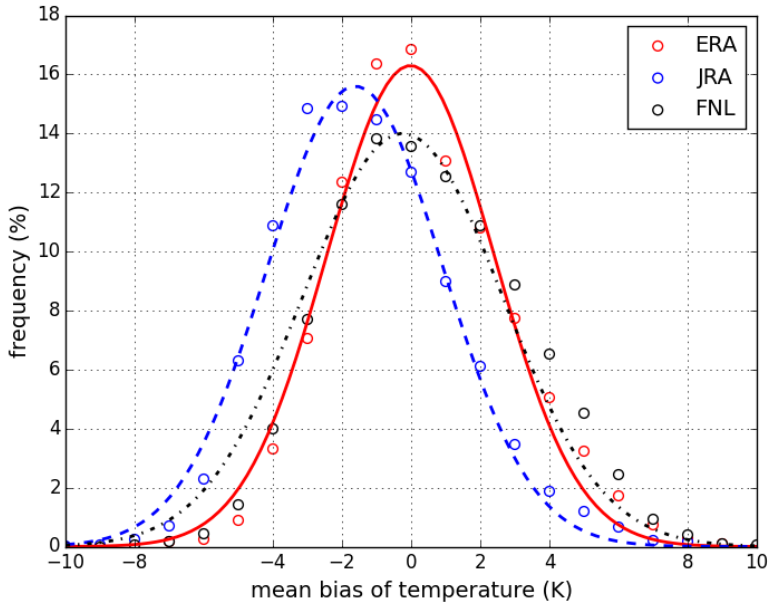

Figure S3 Distribution of model bias in 2-m air temperature for WRF simulations with three different initial/boundary conditions. Model bias is based on the difference of simulation and gauge-based observations (i.e. bias is model minus observation). Open circles represent frequency of model bias within each bin, while curves represent fitted lines assuming a normal distribution.

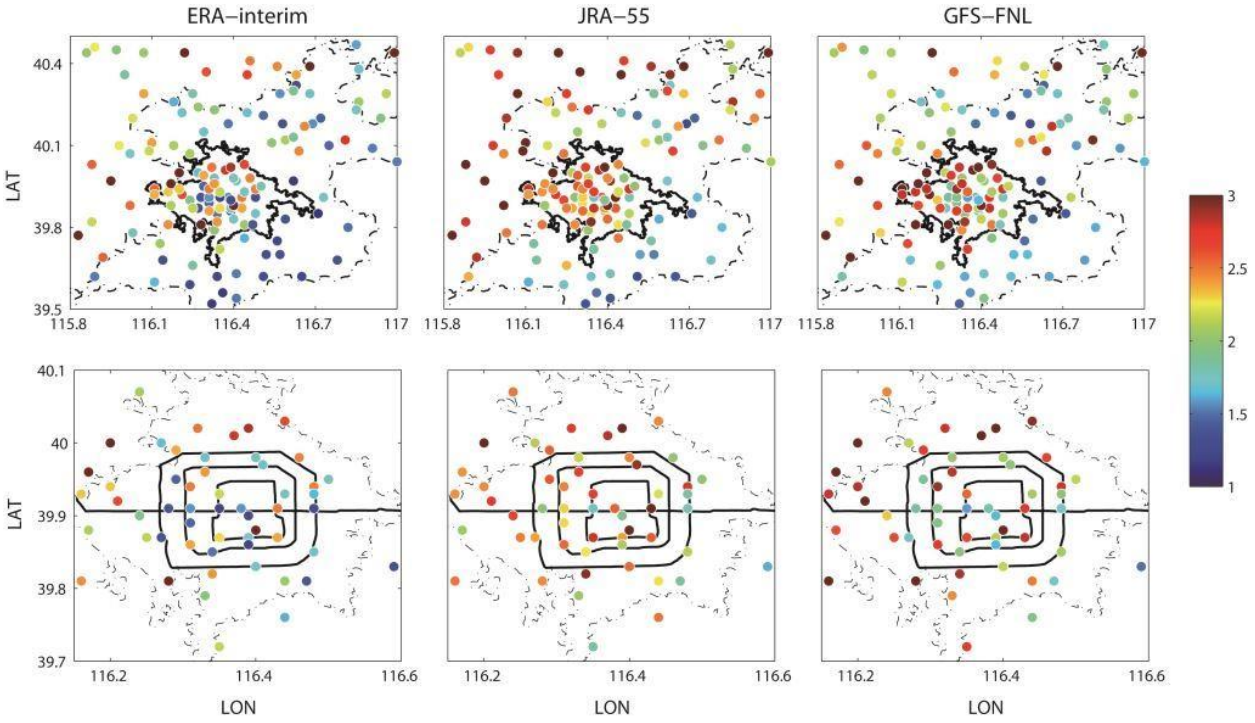

Figure S4 Spatial distribution of absolute mean bias of 2-m air temperature (°C), temporal average over the peak heat wave period. The upper panels show the entire BMA (black line denotes boundary of urban core region, while the dashed line represents boundary of plain region). The lower panels focus on urban core region. Black lines show main roads ("ring roads") in Beijing.

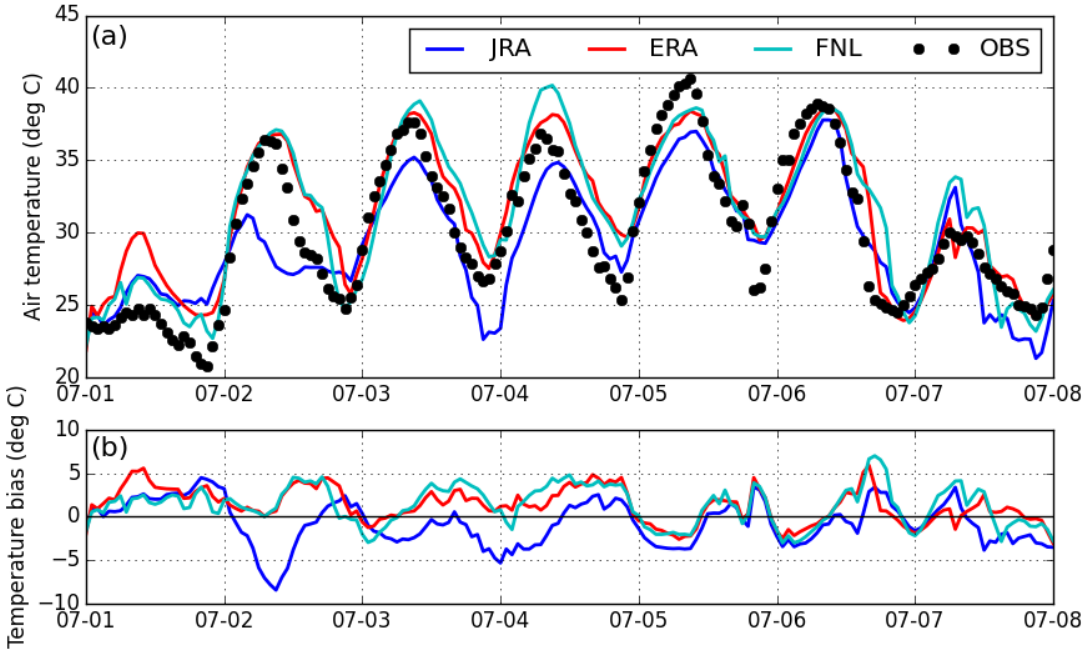

Figure S5 Time series of (a) 2-m air temperature (°C) averaged over urban core region (downtown area) and (b) temperature bias (°C) between different simulations and the observations.

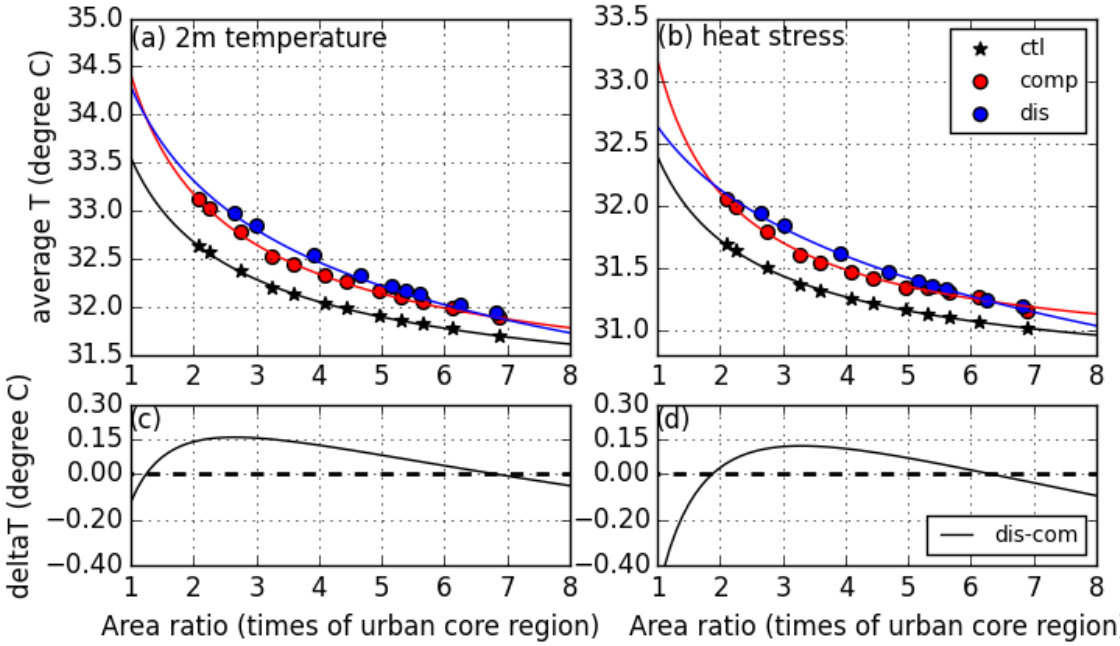

Figure S6 Horizontal spatially averaged (a) 2-m air temperature and (b) heat stress index over and beyond the extent of urban coverage under current climate condition. X-axis is the horizontal extent. The differences between compact- and dispersed-city scenario are shown in (c) and (d) for 2-m temperature and heat stress index, respectively. Dots are based on results from WRF simulations. Lines shown in (a) and (b) are fitted using power-law functions.