RMetS

Royal Meteorological Society

How many RCM ensemble members provide confidence in the impact of land-use land cover change?

Patrick Laux,[a]* Phuong N. B. Nguyen,[a,b] Johannes Cullmann,[c] Tan Phan Van[d] and
Harald Kunstmann[a,e]

[a] *Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany*
[b] *IHP/HWRP Secretariat, Federal Institute of Hydrology, Koblenz, Germany*
[c] *Climate and Water Department, World Meteorological Organization, Geneva, Switzerland*
[d] *Department of Meteorology, VNU Hanoi University of Science, Vietnam*
[e] *Institute of Geography, University of Augsburg, Germany*

ABSTRACT: Regional climate models (RCMs) include both terrestrial and atmospheric compartments and thereby allow studying land–atmosphere feedback, in particular, the impact of land-use land cover driven by biogeophysical processes on regional climate. In this study, a method is developed to separate the signals from the noise in RCM simulations of the effects of changes in land use, using perturbed initial boundary conditions (PICs). We want to know how many ensemble members are required to identify robust and statistically significant land-use land cover change (LULCC) effects from RCM LULCC studies. The method is applied to a case study of urbanization and deforestation, for which LULCC scenarios are implemented in the RCM Weather Research and Forecasting (WRF). Based on WRF ensemble simulations with PICs for 2010, the signal-to-noise ratio (SNR) is used to identify areas with pronounced effect of an LULCC or, rather, the parametrization of the land-use classes. While in the urbanization scenarios clear and statistically significant signals are found for air temperature and for both latent- and sensible heat (SNR values up to 24), the effects are less pronounced for precipitation, and for deforestation in general (SNR values < 1). For the case study of urbanization and precipitation, the impact of the ensemble size is studied in order to derive robust conclusions about the effects of LULCC on precipitation. We conclude that single RCM realizations of different land-use representations are not sufficient to derive LULCC-induced signals, particularly not for precipitation. Small ensemble sizes led to concluding there were significant LULCC-induced precipitation signals, but these disappeared when the ensemble size was increased. Our regional analysis suggests the need for ensemble sizes well above 10 for precipitation.

KEY WORDS   perturbed initial conditions; RCM ensemble simulations; signal-to-noise separation; urbanization; deforestation; Weather Research Forecasting (WRF); Vietnam; land-use land cover change

Received 22 February 2016; Revised 10 June 2016; Accepted 19 June 2016

1. Introduction

Nowadays, humans to some degree have already perturbed more than half of the global land surface. In the temperate regions of the Northern Hemisphere, e.g. croplands have been expanded mainly through clearing the natural forests and grasslands (e.g. Ellis *et al.*, 2010). Such conversions, hereinafter referred to as land-use land cover changes (LULCC), are known to provide a strong climate forcing through altering the exchange of energy, momentum, and water between the land surface and the atmosphere (Dirmeyer *et al.*, 2010; Pielke *et al.*, 2011), and the LULCC processes are more and more recognized by the climate modelling community as important factors. The Intergovernmental Panel of Climate Change (IPCC) included changes in land surface as forcing data in the CMIP5 climate simulations for the first time in the fifth Assessment Report (AR5). Globally, the radiative net effect is estimated to be relatively small compared to pre-industrial times, giving one possible reason why LULCC has often been omitted from climate models in previous IPCC assessments (Pielke *et al.* 2011).

There is a consensus in the literature about the processes of LULCC that affect the climate system:

- Biogeophysical processes, comprising the changes in the properties of the Earth's surface, such as surface roughness, albedo, the leaf area index (LAI), and vegetation stomata resistance, resulting in changes in the fluxes of energy, momentum and moisture.
- Biogeochemical processes, which increase the radiative forcing through changes of the composition of CO_2 and other greenhouse gases, such as CH_4 and N_2O.

However, the quantification of these processes remains uncertain, as it strongly depends on various factors, such as the degree of the conversion and the geographical location (e.g. Boisier *et al.*, 2012). Studies based on global circulation models (GCMs) generally agree that historical LULCC has predominately increased the albedo

* Correspondence to: P. Laux, Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Kreuzeckbahnstrasse 19, 82467 Garmisch-Partenkirchen, Germany. E-mail: patrick.laux@kit.edu

(clearings) leading to a cooling of the near surface air temperature in mid-latitudes, where most LULCC took place (e.g. Feddema *et al.*, 2005). Such a global negative radiative forcing is present in many studies (e.g. Betts, 2001; Betts *et al.*, 2007; Davin and de Noblet-Ducoudré, 2010), which may be enhanced or reduced by non-radiative processes of the alteration of surface fluxes such as the partitioning of net radiation fluxes between latent- and sensible heat flux.

This, in turn, can vary remarkably in space and time and can have large impacts on hydro-meteorological variables such as air temperature and precipitation (e.g. Davin *et al.*, 2007). As an example, deforestation in temperate regions may lead to evaporative cooling during spring and summer because crops often have higher evaporation rates than forests, assuming a sufficient water supply. At the same time, strong decreases in latent heat fluxes, leading to a net warming, are observed in lower latitudes. This is because their cropping systems are often less efficient than in temperate regions (Boisier *et al.*, 2012). Besides the effect of the partitioning between latent and sensible heat, LULCC may also change turbulent energy fluxes by altered roughness lengths. An overview of modelling and observational studies at global scales can be found by Betts *et al.* (2007) and Pitman *et al.* (2009).

To shed light on the impact of LULCC at coarse scales, the research initiative *Land-Use and Climate, Identification of robust Impacts* (LUCID) has been conducted to assess the robust global biogeophysical impacts of LULCC on climate from the pre-industrial era to the present day, using seven GCMs. In this project, Pitman *et al.* (2009) showed statistically significant impacts of LULCC on the simulated near-surface temperature and latent heat flux for the Northern Hemisphere. While most of the models agree on an LULCC-induced cooling effect, the magnitude varies considerably between the different GCMs. LULCC-induced latent heat flux responses not only differ in magnitude but also in sign. There is also less agreement for precipitation, and signals are found to be less significant over regions of LULCC (Seneviratne *et al.*, 2010). de Noblet-Ducoudré *et al.* (2012) explored the reasons for the LULCC-induced changes in Eurasia and North America; both are areas in which the land-use has changed considerably during the industrial age. By including also GHG concentrations and resulting changes in SST in their analyses, they found that LULCC has an impact on several variables of similar magnitude but of opposite sign on regional scales, to increased GHG and warmer oceans. It is found that the variability from LULCC is larger than that from the changes in GHG concentrations and SST within the GCM ensemble.

The impact of urbanization on climate has been analysed in numerous studies, such as e.g. the joint impact of urban expansion and climate change on heat stress (Argüeso *et al.*, 2015). A comprehensive overview of such studies is given by Arnfield (2003). Urban surfaces are largely composed of artificial buildings and roads, and are therefore clearly distinguished from natural surfaces in terms of their mechanical, radiative, thermal,

and hydraulic properties (Lee *et al.*, 2011). This leads finally to the so-called *Urban Heat Island* (UHI), through increased sensible heat fluxes, which can be several orders of magnitude higher than the latent energy flux. The magnitude of the partitioning, however, can strongly vary with the season and the geographic location (Mahmood and Pielke, 2014).

The LULCC forcing is expected to be remarkably larger at smaller scales than at global scales (e.g. Mahmood and Pielke, 2014), and even to vary in sign (Pielke *et al.*, 2011). The essential resources of food, water, energy, human health, and ecosystem function respond to the regional and local climate rather than to the global average, which requires detailed studies of regional LULCC–climate interactions (Rounsevell *et al.*, 2014).

As described by Pielke *et al.* (2011), LULCC can result in mesoscale and regional climate change if the areal coverage of the landscape conversion is large enough. A spatial heterogeneity of approximately 10–20 km has often been considered sufficient for creating mesoscale circulations under convective circulations, and smaller scales (2–5 km) are sufficient to trigger changes in the boundary layer conditions. Regional climate models (RCMs) coupled with land surface models (LSMs) representing the processes of land and atmosphere interactions are suitable tools to study the impact of LULCC on regional climate at such scales. In such coupled modelling systems, the accuracy of the land surface parameters can heavily influence the modelled land surface processes and atmospheric boundary layer characteristics, thereby affecting the model's performance (e.g. Hong *et al.*, 2009; Trail *et al.*, 2013). There are numerous studies using RCMs to study the impact of LULCC on a regional climate system (e.g. Moore *et al.*, 2010, 2012; Sertel *et al.*, 2010; Takahashi *et al.*, 2010). By using the Weather Research and Forecasting (WRF) model coupled to the Noah LSM, Cao *et al.* (2009) identified major deficiencies of using the prescribed land surface biophysical parameters to study the impact of LULCC on regional climate. These deficiencies centre around the limited spatial resolution of vegetation parameters (vegetation fraction, albedo, LAI), because they are derived from global scale remote sensing products, and thus are not suitable to adequately describe the landscape heterogeneity for high-resolution climate simulations, and the omission of the inter-annual changes of these vegetation characteristics. By comparing the differences between two simulations of two different years only, their results with updated and appropriate land surface properties revealed significant improvements compared to the default parameters used in the Noah LSM. Another weakness is its crude representation of urban areas. Recently, urban canopy models have been developed using more explicit representations of the urban morphology and parametrizations of the associated processes (e.g. Lee *et al.*, 2011), but if no detailed urban data are available, the traditional bulk urban parametrization is suggested (e.g. Loridan *et al.*, 2010; Chen *et al.*, 2011). An intercomparison study of different urban canopy models in a stand-alone mode is performed by Grimmond *et al.*

(2010), but similar studies using different urban canopy models coupled to RCMs are still missing.

In analogy to the uncertainties in the models in future climate projections (e.g. Giorgi and Bi, 2000; Haughton *et al.*, 2014), consisting of the uncertainties in the physics parameters of the model (e.g. the model setup and land-surface parameters), external forcing (e.g. radiation forcing) and internal model variability (e.g. initial and lateral boundary conditions), there is also strong evidence of high uncertainties involved in using climate models to study the impact of LULCC on climate systems. This has been demonstrated in the LUCID project, in which the issue of uncertainties derived from multiple GCMs and land surface schemes has been addressed (e.g. Boisier *et al.*, 2012). Besides the aforementioned uncertainties in GCM models, similar uncertainties are expected for RCMs, with the difference that RCMs apply fixed sea surface temperatures and lateral boundary conditions from the driving GCMs (e.g. Denis *et al.*, 2003), thereby having a reduced variability *per se*.

In our study we focus on the uncertainties deriving from the RCM internal model variability (Giorgi and Bi, 2000; Christensen *et al.*, 2001; Haughton *et al.*, 2014), which we try to quantify by using ensembles generated by perturbed initial boundary conditions (PICs). In theory, the uncertainty in the prediction of the model caused by PICs would be degraded gradually by increasing the number of lateral boundary condition updates (Wu *et al.*, 2005). In concert with limited computing resources, this is often used as a justification for applying single realizations with and without LULCC, just omitting a certain model spin-up period for the analyses (e.g. Sertel *et al.*, 2010; Takahashi *et al.*, 2010; Beltrán-Przekurat *et al.*, 2012).

Through own unpublished RCM pilot studies of the impact of LULCC on regional climates using PICs, we have found that the effects of PICs can be large. In the context of the research project *Land-use and Climate Change Interactions in Central Vietnam* (LUCCi), we have assessed the impact of PICs on the regional climate simulation results for the Vu Gia Thu Bon (VGTB) river basin in Central Vietnam while also examining the impact of LULCC. We used one-day lagged initializations, a simple perturbation technique commonly used in multi-year prediction studies (e.g. Müller *et al.*, 2012; Baehr and Piontek, 2014). To our knowledge, there is no framework in the literature of how the effect of an LULCC can be distinguished from the RCM's internal variability, as represented by the ensemble variance, from now on also referred to as noise. To put it another way, we want to know how many ensemble members are required to identify robust and statistically significant LULCC effects from LULCC studies.

To fill this gap, there is employed a method based on RCM ensemble simulations with PICs and suitable performance measures and tests. The method to study the effects of an LULCC in RCM simulations, irrespective of the study region, might be of high interest to the climate community. The procedure is demonstrated for two case studies in SE Asia, one on urbanization and the other on deforestation, representing recent and ongoing land-use changes in the region, with three scenarios considering different extents of areal conversions.

We hypothesize that depending on the spatial extent and the two different LULCC conversion scenarios, the considered climate variables at surface will change significantly. In analogy to climate change projections, we expect that we need multiple ensemble members to prove evidence for LULCC-induced changes.

This article is structured as follows: Section 2 outlines the applied methodologies and the data. Section 3 includes a description of the results and discussions. Concluding remarks are given in Section 4.

2. Data and methods

2.1. Land-use/land-cover data for 2010

The land-use land cover (LULC) data used in this study were obtained by preprocessing and classifying remote sensing data from 2010, which was obtained in the LUCCi project. The data include Spot 5 data in a 2.5-m spatial resolution, covering approximately 90% of the VGTB basin, Landsat ETM+ data (30 m spatial resolution) covering the entire basin, as well as road network vector data, FIPI forest map vector data, and river network vector data from 2010. After various preprocessing steps, such as geometric corrections, cloud and shadow removal, radiometric corrections, and gap-filling, an LU map was created based on a supervised classification algorithm. The classification has been checked using ground truthing activities in the VGTB basin, and the resulting map has been refined based on auxiliary information from local stakeholders. After that, the accuracy of the LULC data has been assessed, showing an overall accuracy of about 82% of correctly classified pixels according to the calculated confusion matrix. This LULC map of 2010 has a spatial resolution of 30×30 m and consists of 6 classes following the IPCC nomenclature, i.e. forest, cropland, grassland, water, settlements, and one additional class representing all other LULC classes.

To integrate the LUCCi LULC map 2010 of the VGTB basin with the WRF model, the LULC data are spatially aggregated to a 1-km resolution, and the dominant classes within the coarser 1-km grid are selected. This 1-km aggregated LULC map of 2010 (hereinafter referred to as LUCCi LULC map 2010) exhibits the following proportions of the classes: the areas for forest, grass land, and crop are 50, 18, and 26%, respectively, while urban land accounts for about 1% of the total area (Figure 1).

2.2. Land-use/land-cover change scenarios

To assess the sensitivity of LULCC over this region, LULCC scenarios were produced to examine the impacts of LULCC on the regional climate. In the LUCCi LULC map 2010, the cropland in a 20-, 14-, and 9-km radius around Da Nang station is replaced by urban land (referred to from now on as LULC-urbanization) and used for regional climate simulations. The area of the LULC converted from cropland to urban can be seen in Figure 2(a)–(c). The forest in a 20-, 14-, and 9-km radius
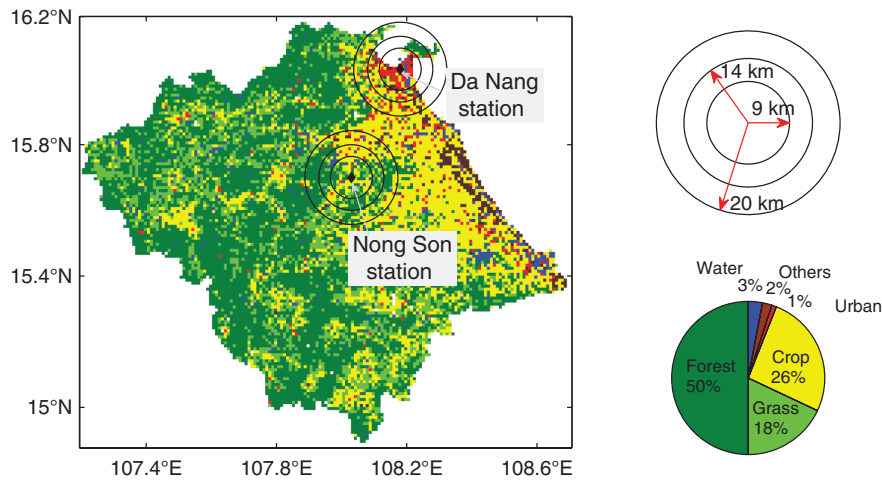
Figure 1. Land-use information of the Vu Gia Thu Bon (VGTB) river basin in Central Vietnam and the two selected locations for artificial land-use conversion experiments.
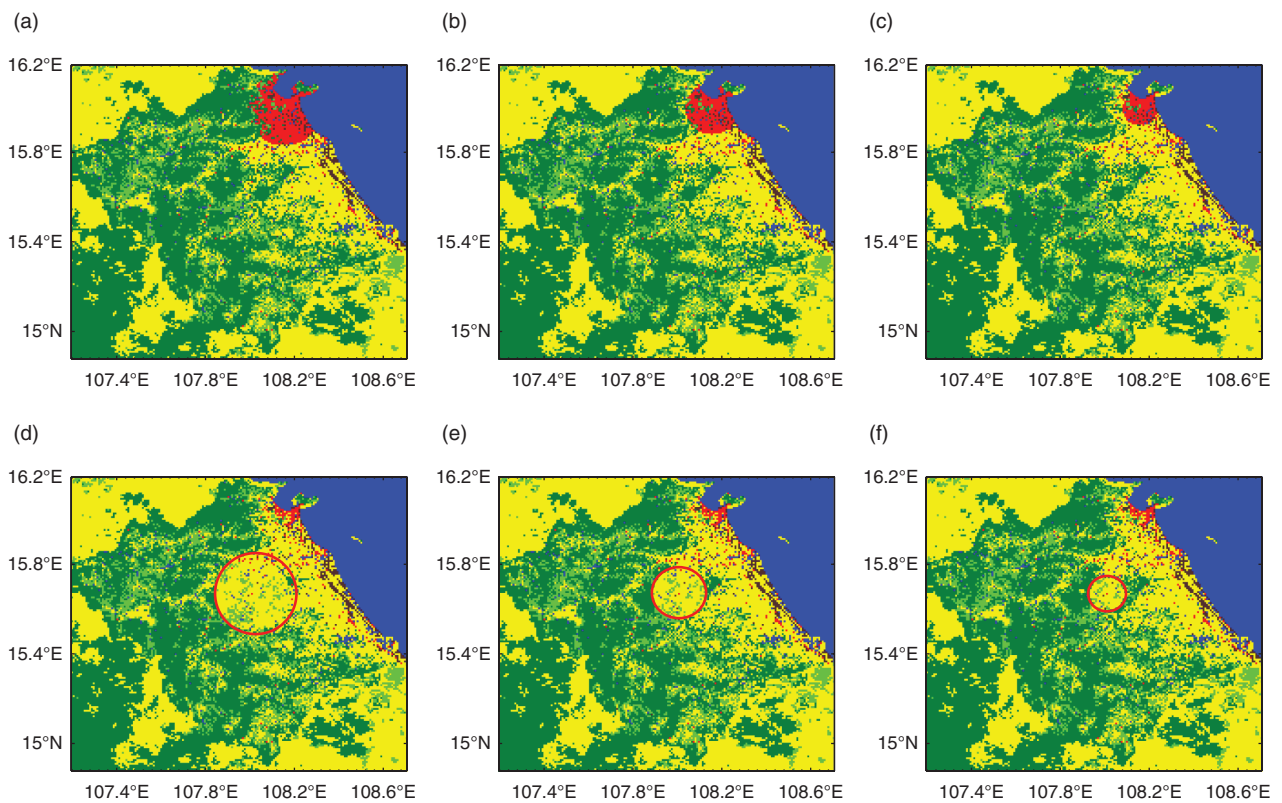

Figure 2. Artificially converted land-use land cover in VGTB basin. (a–c) LULC-urbanization, (d–f) LULC-deforestation. From the left to the right are the LULC maps with different degrees of conversion from high to low, i.e. conversion in a radius of 20, 14, and 9 km around Da Nang station (a–c) and Nong Son (d–f).

around Nong Son station is replaced by cropland, shown in Figure 2(d)−(f).

2.3. Regional climate information

2.3.1. Observation data

The availability of hydro-meteorological data in the VGTB basin is limited. Overall, the measurement network is relatively sparse: the mountainous slopes in the western area of the basin are ungauged, and the recorded time series are not

complete (Souvignet *et al.*, 2014). The observational data used in this study include monthly surface temperature and precipitation for 2010 for Da Nang, and precipitation only for Nong Son station. The data were complete for 2010, and used to analyse the accuracy of the WRF simulations.

2.3.2. High-resolution regional climate simulations

WRF simulations (version WRFv3.6.1) were performed to derive high-resolution climate information for the year
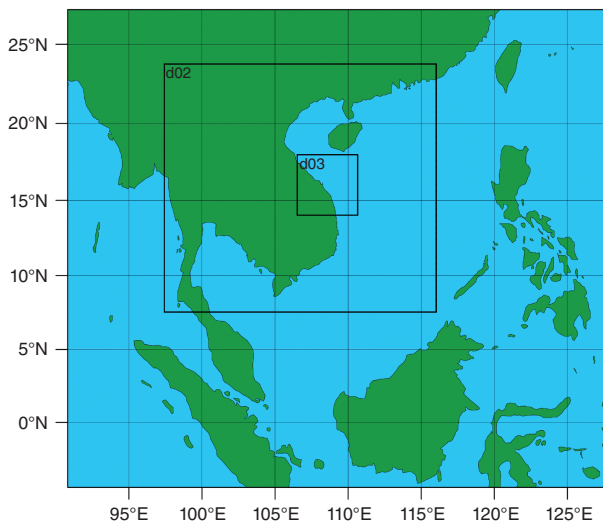
Figure 3. Domain setup used for the WRF climate simulations: Domain 1 (D01) has 90×80 grid points with a resolution of 45 km, domain 2 (D02) includes 136×124 grid points with a resolution of 15 km, and domain 3 (D03) has 87×81 grid points with a resolution of 5 km.

2010. The ERA-interim reanalysis data were used for the boundary conditions (updated every 6 h) to drive the WRF model. A nested approach was used to downscale the climate information to horizontal grid resolutions of 45, 15, and 5 km. The main domain covers Southern Asia from 5°S to 27°N latitude and 90°E to 130°E longitude (Figure 3). The first nested domain covers South-Eastern Asia and parts of the South China Sea. The second nested domain focuses on the study area, the VGTB basin (Laux *et al.*, 2013).

As it has already been demonstrated to be a suitable parametrization for this region in a previous study (Laux *et al.*, 2013, 2014), we used the cumulus parametrization scheme from Betts–Miller–Janjic, the microphysics from the WSM 3-class simple ice scheme, the boundary layer from the YSU scheme, and the land surface scheme from the Noah land-surface model (LSM).

The Noah LSM (Chen and Dudhia, 2001) provides the bottom boundary conditions for the WRF model. Therefore, it calculates the turbulence exchanges of momentum, mass, and energy between the surface and the overlying atmosphere. The land surface at each grid cell is represented by LULC and soil, consisting of 24 LULC categories (following the USGS classification) and 16 types of soil texture. Each LULC category is characterized by physical and aerodynamic parameters, such as surface roughness length and displacement height, albedo, emissivity, green vegetation fraction, and LAI. Each soil texture type is characterized by parameters such as the soil heat conductivity and diffusivity, the maximum soil moisture content, and the wilting point soil moisture (Lee *et al.*, 2011). WRF updates the albedo on a monthly basis based on the vegetation fraction. A detailed description of the Noah LSM is given in Chen and Dudhia (2001).

In this study, we used the bulk urban parametrization, which is based on the concept that the physical parametrization for an urban patch is identical to that for vegetation types (big leaf concept, see Liu *et al.*, 2006). In recent versions of WRF, more advanced urban canopy model options are available, e.g. the *Single-Layer Urban Canopy Model* (SLUCM) (e.g. Kusaka *et al.*, 2001; Chen *et al.*, 2011) and the *Multi-layer, Building Environment Model* (BEM). To achieve an improved urban atmospheric numerical modelling, more detailed data, such as, e.g. urban vegetation and urban surface morphology, as well as explicit parametrizations of the urban physical processes are required (e.g. Chen *et al.*, 2011; Lee *et al.*, 2011; Deng *et al.*, 2013), which was not possible in the context of this study.

2.4.   Experimental design

We performed a series of regional climate simulations by implementing the different LULCC scenarios described above in WRF. To investigate the impact of LULCC on the regional climate, we first had to separate the signal from the noise which inherently exists in climate simulations. For this reason, we carried out ensemble WRF simulations consisting of 15 ensemble members for each LULCC experiment, i.e. the LULC-2010, LULC-urbanization and LULC-deforestation, by perturbing the initial boundary conditions of the WRF simulations. This was done using different initialization dates, shifted by 1 day (Figure 4). So as not to affect the statistics of the climate simulations by the slightly varying initialization dates of the WRF, a spin-up time of 3 months was used. The perturbed simulations were initialized from October 2009 and run to the end of December 2010. Only the results for 2010 are used in the analyses. Based on the variance of these ensemble simulations, the amount of noise can be estimated and used to identify the signals within each LULCC scenario. The results of the WRF model derived from the LULC-2010 (based on the LUCCi LULC map 2010) experiment is referred to as control. The LULC-urbanization scenarios with conversion radii of 20, 14, and 9 km are referred to as Urban-20, Urban-14, and Urban-09, respectively. The LULC-deforestation scenarios with radii of 20, 14, and 9 km are referred to as Deforest-20, Deforest-14, and Deforest-09, respectively.

The signal-to-noise ratio (SNR) is calculated to estimate the effect that can be attached to the WRF LULCC simulation results, by estimating the amount of noise in the ensemble. Note that apart from the usual meaning in statistics or time series analysis, here the 'signal' is defined as the mean value of an ensembles, and the 'noise' is defined as the sum of the corresponding standard deviations. The effect of an LULCC scenario can thus be quantified as follows:

$$SNR = \frac{|\mu_A - \mu_B|}{\sigma_A + \sigma_B} \qquad (1)$$

where *A* denotes the WRF control simulations (based on the LUCCi LULC-2010 map) and *B* denotes the simulations with the LULCC scenarios. μ_A and μ_B are the mean, and σ_A and σ_B are the standard deviation of the 15 perturbed runs of the two experiments, respectively.
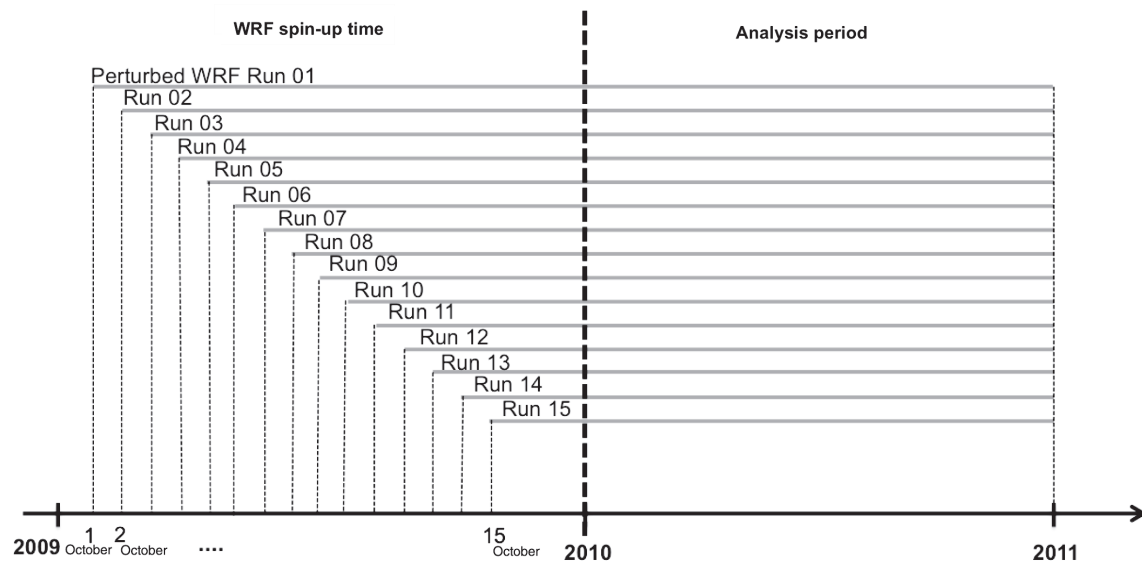
Figure 4. Ensemble approach of WRF simulations consisting of 15-member ensembles for each LULCC experiment. WRF simulations are driven by perturbed initial boundary conditions (PICs), which has been achieved by using different initialization dates.

Values of the SNR that are greater than 1 indicate comparatively higher signals (compared to the noise) within the ensemble simulations.

2.5. Tests of statistical significance

High SNR values suggest an impact of an LULCC on the climate. To strengthen the evidence, the results for each scenario are tested for statistical significance. The standard t-test is not applicable in this case due to the fact that the data are serially correlated and non-Gaussian. To take these issues into account, the following two tests haven been used: by applying two different tests:

1. A t-test, to check whether H_0 (the mean values obtained by the control and the LULCC-induced simulations are from the same distribution) can be rejected, or whether H_1 (the mean values come from different distributions) has to be accepted. As the variance is reduced in serially dependent data, the test is modified using a variance inflating factor (Wilks, 2006). Serial dependence is found for all variables under consideration, but a first-order autoregressive model was found to be suitable to remove the serial dependence (Figure S1, Supporting Information).
2. A bootstrap test if the data are not symmetrically distributed around the mean (Gaussian assumption) and the t-test is not applicable. The Gaussian assumption is found to be violated for monthly precipitation, e.g. the November precipitation analysed in this study. One reason for that is the small sample size of the monthly precipitation values.

3. Results and discussion

In this paragraph, two different case studies, i.e. urbanization scenarios of Da Nang (case study I) and deforestation scenarios of Nong Son (case study II) are described.

Table 1. Mean and standard deviation of annual air surface temperature (T2) [°C], latent heat flux (LHF) [W m^{-2}], sensible heat flux (SHF) [W m^{-2}], and precipitation (P) [mm] for the control simulations, and the scenarios Urban-20, Urban-14, and Urban-09.

Variable	Mean (standard deviation)			
	Control	Urban-20	Urban-14	Urban-09
T2	25.8 (0.05)	27.4 (0.05)	26.9 (0.12)	26.9 (0.04)
LHF	74.7 (1.9)	4.3 (0.2)	13.9 (2.6)	14.6 (0.3)
SHF	52.4 (1.7)	107 (1.5)	100.9 (2.4)	100.3 (1.3)
P	3701 (455)	4086 (263)	3985 (290)	4028 (375)

Both case studies analyse the signals in the different meteorological surface variables induced by the different extents of the LULCC, in the sense of the different radii around the station in which the land-cover is changed, based on the values of SNRs of the WRF ensemble simulations.

3.1. The control simulations: ensemble WRF simulations using the LUCCi LU 2010 map

First, the spread of the ensembles of the control simulations is analysed for all meteorological variables at both locations (Da Nang and Nong Son). In order to obtain more robust statistics, the variables of the WRF simulations are averaged over 9 grid-points next to the station. For the surface air temperature and the latent and sensible heat flux, it is found that the spread of the ensembles is low. The mean and standard deviations of the 15-member ensembles can be found in Tables 1 and 2 (left column). For precipitation, however, the spread is much higher, which may be the result of the influence of particular members.

We use the ensemble of the control simulations to (1) identify ensemble members which may drastically

Table 2. Mean and standard deviation of annual surface temperature (T2) [°C], latent heat flux (LHF) [W m^{-2}], sensible heat flux (SHF) [W m^{-2}], and precipitation (P) [mm] for the WRF control, Deforest-20, Deforest-14, and Deforest-09 scenario simulations.

Variable	Mean (standard deviation)			
	Control	Deforest-20	Deforest-14	Deforest-09
T2	25 (0.04)	25 (0.04)	25 (0.03)	25 (0.05)
LHF	98.8 (2.1)	101.3 (1.2)	100.7 (1.2)	100.5 (2.3)
SHF	27.3 (1.5)	25.3 (1.0)	25.4 (1.1)	25.2 (1.3)
P	4780 (458)	4732 (489)	4849 (443)	4908 (472)

increase the ensemble variability, and (2) quantify the minimum number of ensemble simulations needed to derive robust signals. The impact of adding certain members to the ensemble is expressed in terms of the mean absolute error (MAE) compared to the observations, as shown in Table 3 for Da Nang and Table 4 for Nong Son. The table for Da Nang, e.g. reads as follows: the three-member case leads to an MAE of 96 mm, consisting of the members 14, 8, and 3. It is found that the MAE increases until five members are included, after which it remains stable until it sharply increases for the 14- and the 15-member solutions. Thus, for the control simulations, 5 members are found to be necessary to yield robust statistics, 13 members are needed to obtain acceptable results, and 15 members will lead to a high variability within the ensemble. Something similar can be said for Nong Son. Here, three members could be identified (ensemble members: 4, 10, and 2).

Figure 5 demonstrates the impact of the ensemble size on both the ensemble mean values as well as the ensemble variances of the precipitation at the Da Nang station. Figure 6 shows the results for Nong Son station. Tables 5 and 6 indicate the mean values and the standard deviations for the different ensemble sizes, respectively.

Note that as omitting selected members can increase the variance, the use of a single realization or of an RCM ensemble consisting of only a few members may lead to different conclusions than those resulting from the use of the full ensemble.

3.2. Case study I: ensemble WRF simulations using the urbanization scenarios of Da Nang

3.2.1. *Expected impact on the annual cycle of the meteorological variables*

After analysing the control simulations based on the LUCCi land-use map 2010, the impact of the urbanization scenario will now be analysed. Table 1 shows the mean values and the standard deviations for the Urban-20, Urban-14, and Urban-09 artificial land-use conversion experiments. The mean and standard deviation of 15 perturbed runs in each experiment are calculated for the annual surface temperature and the latent and sensible heat flux. In general, the standard deviations of the meteorological variables are lower than their mean values, thus indicating signals which are much greater than the noise in the WRF simulations.

As expected, there are remarkable differences in the statistics of the different meteorological variables between the control simulations and the urbanization experiments, whereas the differences within the urbanization experiments are relatively high between the Urban-20 and the other two scenarios. The mean values of Urban-14 and Urban-09 are in the same range, whereas the standard deviation is markedly increased in the Urban-14 experiments. The mean of the surface temperature for the Urban-20 is increased by 2.4 °C, and increased by 1.1 °C in the Urban-14 and Urban-09 experiments (compared to the control). The mean of the latent heat flux is remarkably decreased, ranging from 74.7 for the control to 4.3, 13.9, and 14.6 W m^{-2} for Urban-20, Urban-14, and Urban-09, respectively. In the opposite direction to the results of the latent heat flux, the mean values of the sensible heat flux are nearly doubled in the experiments of the urbanization scenario. The results of the surface meteorological variables might be biased in terms of the magnitudes of the signals due to inaccurate representation of the UHI, as, e.g. shown in the study of Lee *et al.* (2011), resulting from the applied bulk urban parametrization in the Noah-LSM. In contrast to Lee *et al.* (2011), Giannaros *et al.* (2013) concluded that, overall, the WRF/Noah modelling system (using the bulk urban parametrization) was suitable for representing the major features of the UHI for Athens

Table 3. MAE values [mm] of the ensemble, including all ensemble members from the left to the horizontal position of the MAE value (Da Nang station).

MAE [mm]	77	88	96	112	121	121	126	128	137	143	144	160	168	185	218
Ensemble member	14	8	3	5	12	15	11	4	7	2	9	1	13	6	10

The ensemble members are labelled according to the starting day of the simulation in October 2009.

Table 4. MAE values [mm] of the ensemble, including all ensemble members from the left to the horizontal position of the MAE value (Nong Son station).

MAE [mm]	105	112	118	119	137	137	151	155	158	165	168	175	190	214	241
Ensemble member	5	14	15	13	3	7	12	6	8	1	11	9	4	10	2

The ensemble members are labelled according to the starting day of the simulation in October 2009.
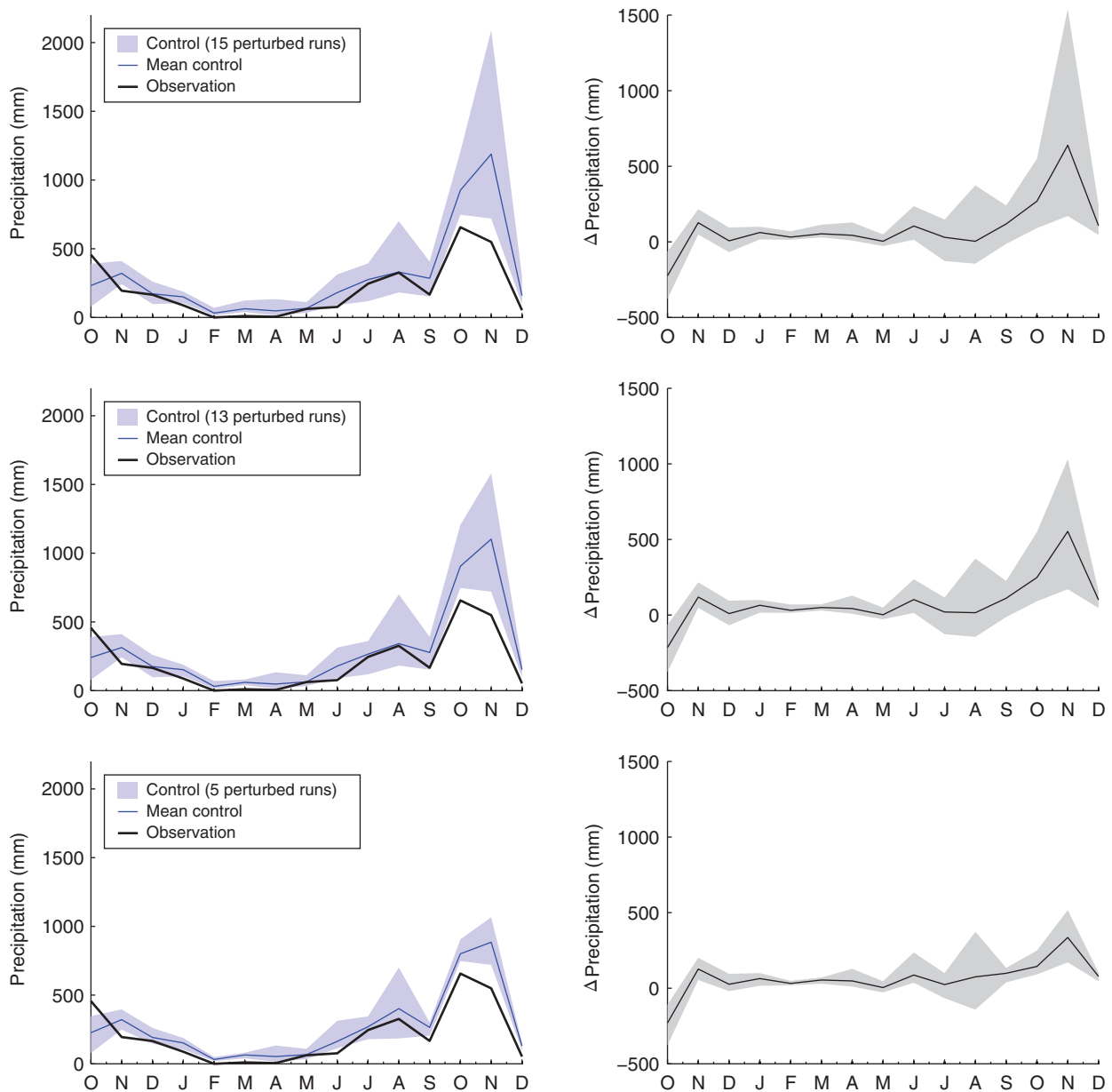
Figure 5. Range of simulated monthly precipitation (control ensemble simulations) using a 15-member solution (top), a 13-member solution (middle), and a 5-member solution (bottom), and the respective differences (control *minus* observation) at the Da Nang station (right) for the period October 2009–December 2010.

(Greece). However, a thorough validation was not possible, due to missing observations in their study (Giannaros *et al.*, 2013).

The temporal variations between the urbanization scenarios and the control simulations are depicted in Figure 7. Compared to the observations, the seasonal cycle of surface air temperature of the control simulations are in good agreement with the available observation data, giving evidence that the LUCCi land-use map 2010 as well as the applied WRF physical parametrization schemes are appropriate. The surface temperatures of all urbanization experiments are consistently higher than those of the control simulations throughout the entire year 2010. The alterations of the surface temperature could possibly be explained by two effects: the effect

of the radiative process of the decreased albedo and the non-radiative process of the increased roughness length when converting from cropland to urban. A decrease in the albedo usually leads to a warming of the surface by the increased solar radiation. Besides this radiative effect, an increase of the roughness length may increase the magnitude of the turbulent energy fluxes, thus decreasing the surface temperature (Boisier *et al.*, 2012).

The conversion to urban area may lead to significant alterations of the partitioning of the turbulent energy fluxes. This reduces the latent heat flux, which will increase the sensible heat flux due to conservation of energy (Figure 8). This leads to a net warming in our simulations and is evident in all urbanization experiments. The largest decrease of latent heat flux is found in the
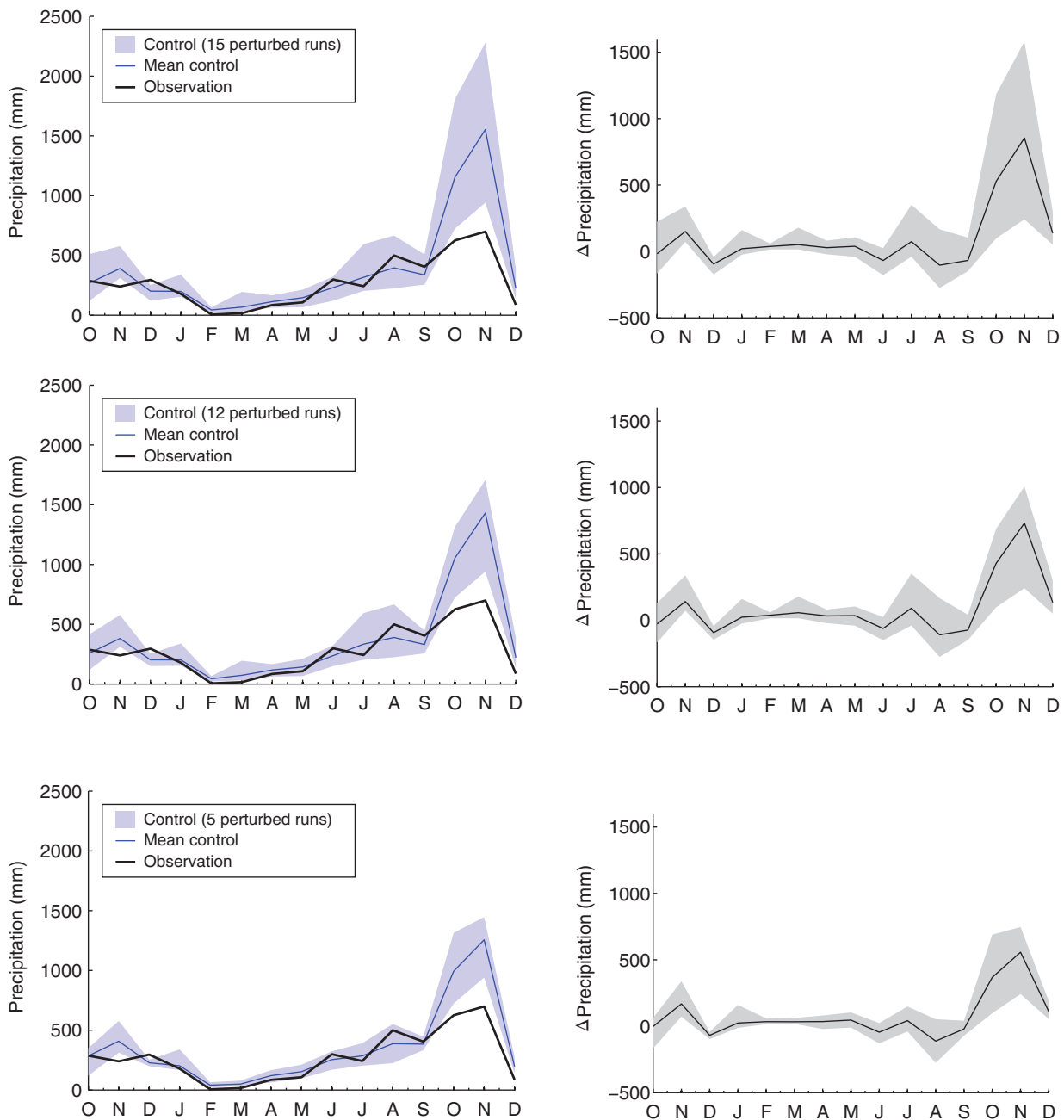
Figure 6. Range of simulated monthly precipitation (control ensemble simulations) using a 15-member solution (top), a 12-member solution (middle), and a 5-member solution (bottom), and the respective differences (control *minus* observation) at the Nong Son station (right) for the period October 2009–December 2010.

Table 5. Mean and standard deviation of precipitation [mm] for the control simulations, and the scenarios Urban-20, Urban-14, and Urban-09.

| | Mean (standard deviation) | | | |
	Control	Urban-20	Urban-14	Urban-09
5 perturbed runs	3280 (327)	4147 (305)	4113 (324)	4110 (320)
13 perturbed runs	3491 (291)	4036 (294)	4023 (306)	4021 (302)
15 perturbed runs	3701 (455)	4086 (263)	3985 (290)	4028 (375)

Table 6. Mean and standard deviation of precipitation [mm] for the control simulations, and the scenarios Deforest-20, Deforest-14, and Deforest-09.

| | Mean (standard deviation) | | | |
	Control	Deforest-20	Deforest-14	Deforest-09
5 perturbed runs	4323 (344)	4560 (548)	4548 (369)	4970 (553)
12 perturbed runs	4579 (380)	4765 (547)	4804 (454)	4845 (531)
15 perturbed runs	4780 (458)	4732 (489)	4849 (443)	4908 (472)

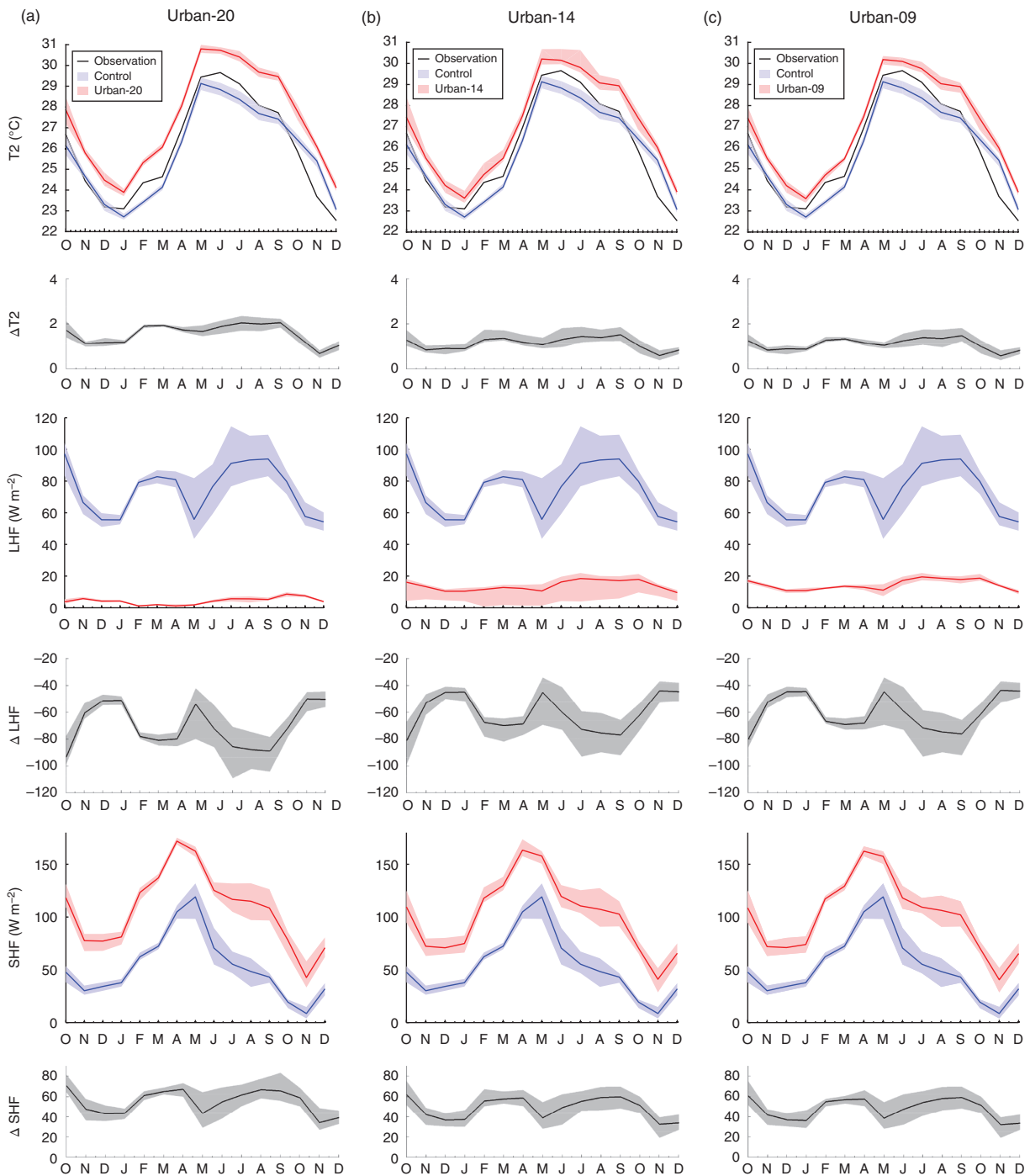Figure 7. Monthly surface air temperature (T2), latent heat flux (LHF), and sensible heat flux (SHF) (from top to bottom) for the three different urbanization scenarios averaged over the next 9 grid cells around Da Nang station. The envelopes for the control (blue) are based on the observed LUCCi LU map 2010, the three different urbanization experiments [(a) Urban-20, (b) Urban-14, and (c) Urban-09] are derived from ensemble WRF simulations using 15-member ensembles with different initialization conditions (red). The subfigures illustrate the range of differences between the LUC scenarios and the control simulations. The WRF simulations were run from October 2009–December 2010.

Urban-20 experiment. The latent heat flux in the control experiments demonstrates the effects of cropland in the VGTB, which peaks both in March and September and plunges in May, the driest month of the year 2010. In the urbanization experiments, however, the peaks and the plunge are reduced. In turn, there is no clear annual cycle of latent heat fluxes visible in the urbanization experiments. The figure illustrates the great variation among the

Urban-20 and both the Urban-14 and Urban-09 experiments. The net turbulent heat fluxes are lower than the control run. From previous studies, however, it is expected that increased roughness lengths lead to increased turbulent energy fluxes, and thus contribute to reducing the aforementioned effects of the net warming at the surface (e.g. Davin and de Noblet-Ducoudré, 2010; Boisier *et al.*, 2012).
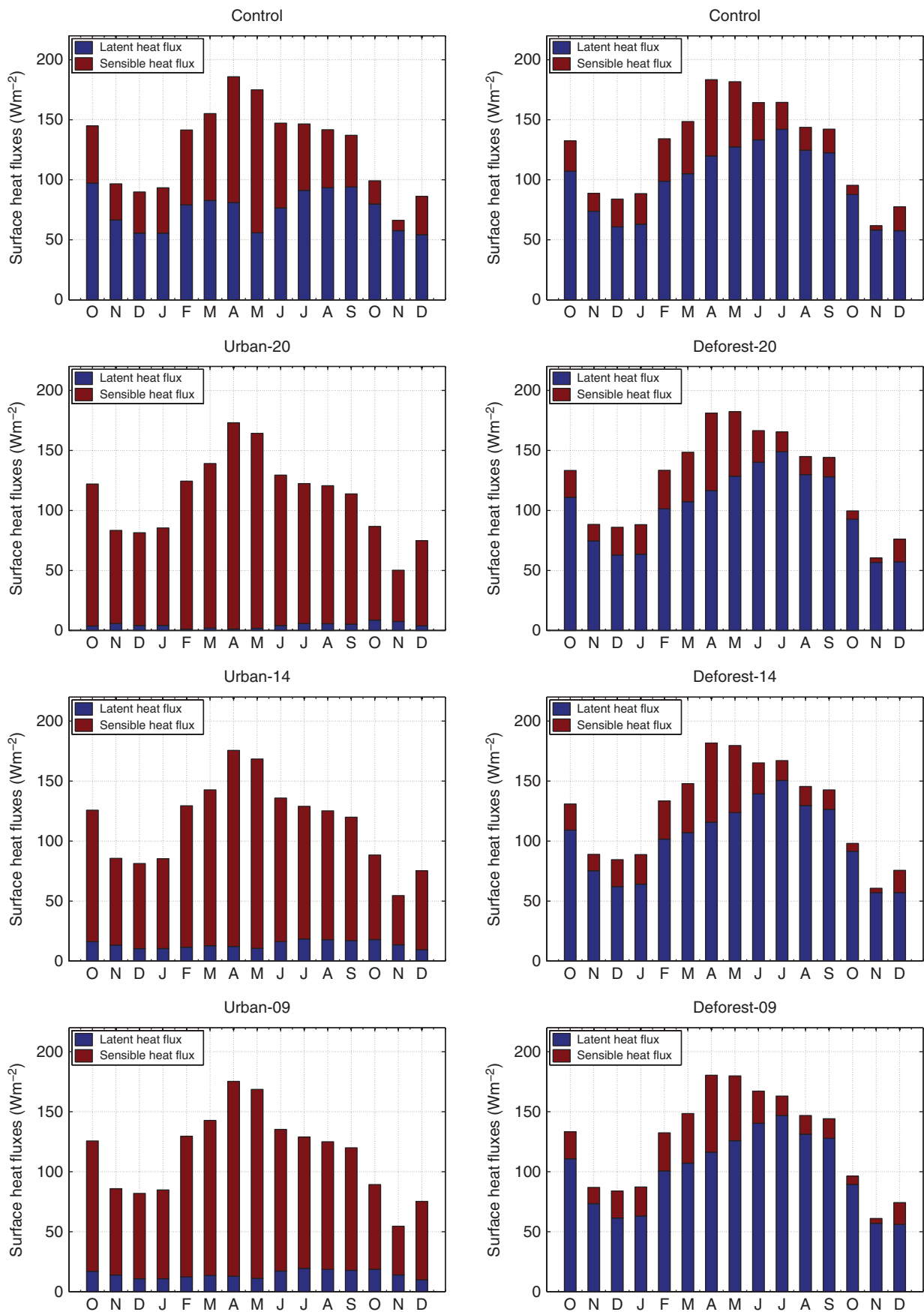
Figure 8. Partition between LHF and SHF for urbanization scenarios at Da Nang station (left column) and deforestation scenarios at Nong Son station (right column) based on the ensemble mean of the 15 PICs.
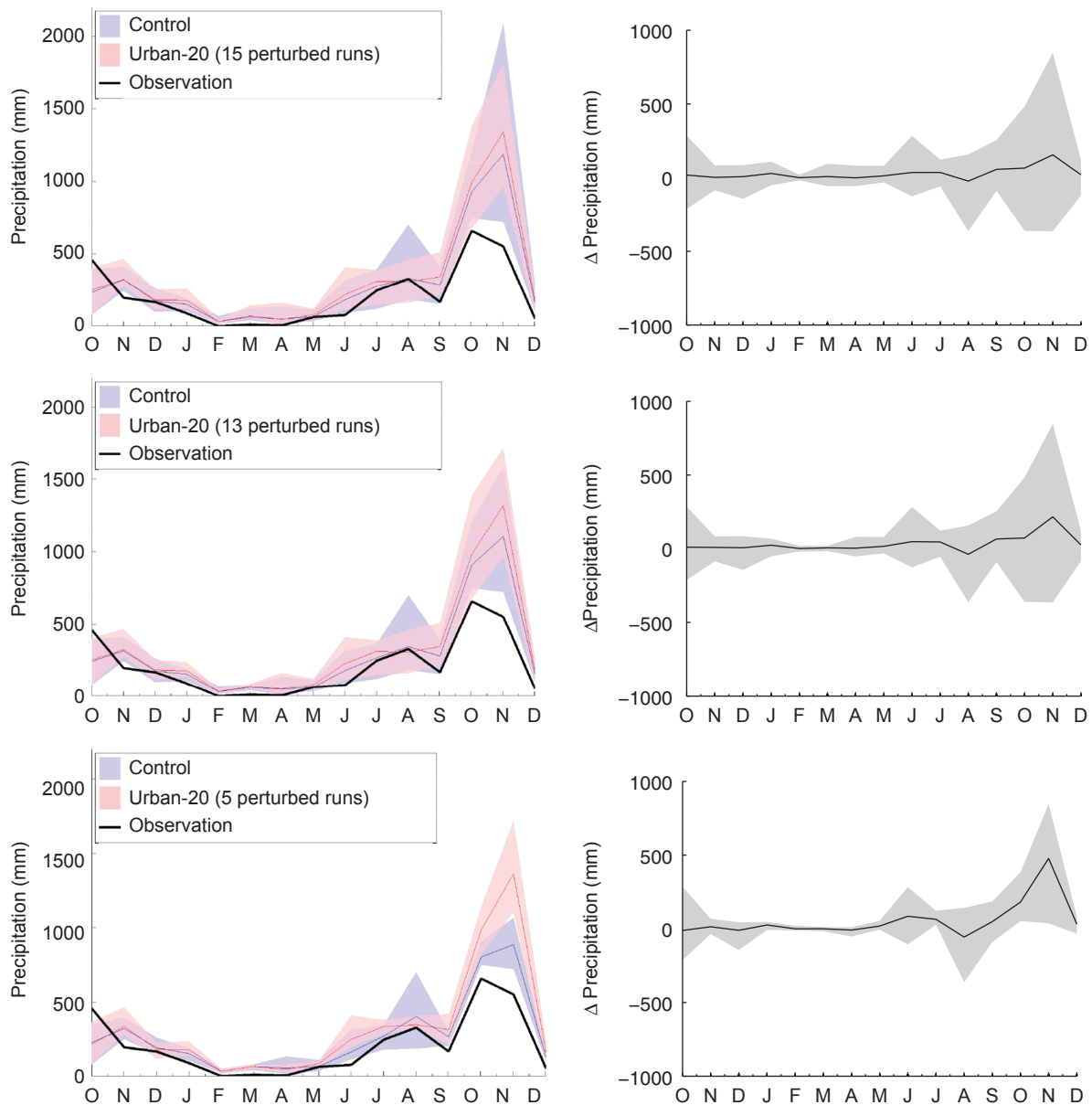
Figure 9. Monthly total precipitation [mm] for the Urban-20 scenario and the control simulations at Da Nang station in 2010 using 15-, 13-, and 5-member ensembles. The surrounding 9 grid cells of Da Nang are used (left). Differences between the scenario and the control simulations (Urban-20 *minus* control) for the period October 2009–December 2010.

Although the standard deviations are greater during the months from May to September, the envelopes of the ensembles for the urbanization and the control do not overlap at any time in 2010 for the surface air temperature or the latent and sensible heat fluxes, i.e. their median values are much greater than their standard deviations, indicating again the clear signal induced by the artificial LULCC throughout the entire year.

It is well known that urbanization and other LULCCs alter the surface energy balance and may also affect the thermodynamic processes in the boundary layer of the atmosphere (Mahmood and Pielke, 2014), potentially altering directly or indirectly the precipitation signal. Relatively high standard deviations in the precipitation series (resulting in lower SNR values compared to those of, e.g. the temperature) may hinder clear

inferences about the impact of an LULCC on precipitation. Therefore, the impact of the ensemble member size is demonstrated in the following. At least five members are needed to obtain robust estimates (converging standard deviation), but if the ensemble members 6 and 10 are included, the standard deviation of the ensemble increases from about 280 mm to about 440 mm (Figure S2). The precipitation time series for the urbanization scenario Urban-20, depending on the number of ensemble members, are shown in Figure 9. Note that the ensemble size of the control simulations is changed accordingly, i.e. the same members are omitted from the ensemble. It is expected from Figure 9 that the 15-member solution would not disclose any signal during the rainy season (October to December) induced by urbanization, due to the high variability within both ensembles (control and
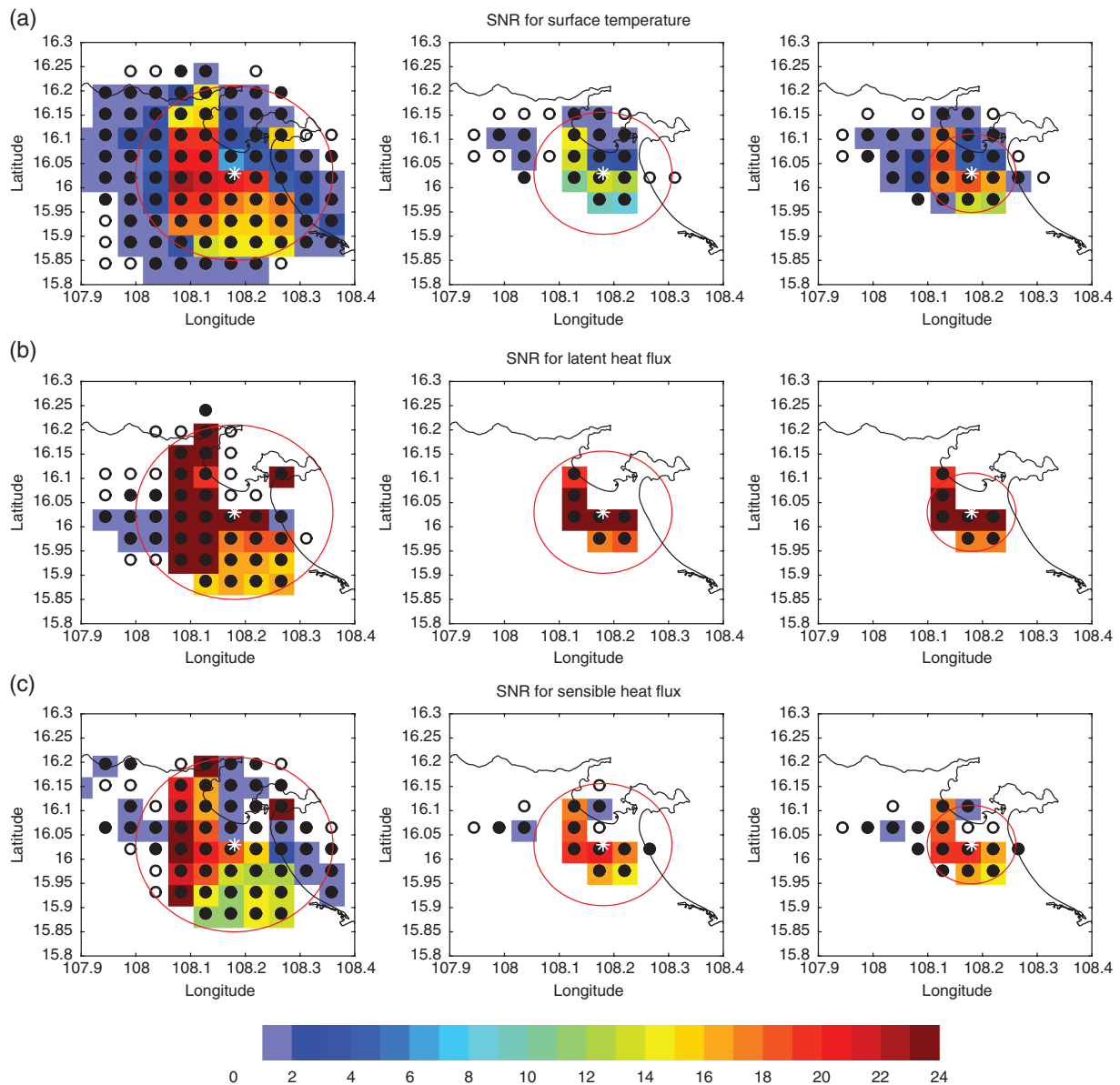
Figure 10. SNR values [.] of the different urbanization scenarios (Urban-20, Urban-14, and Urban-09, from left to the right) for (a) surface temperature, (b) latent heat flux, and (c) sensible heat flux based on the full RCM ensemble. The stippled grid cells show the statistically significant grid cells in which the *t*-test rejects the null hypothesis of having the same mean values for the scenarios and the control simulations at $\alpha = 0.05$ (empty circles) and $\alpha = 0.01$ (filled circles).

Urban-20). Reducing the ensemble size to 13, however, already provides the potential to identify a signal in the time series, whereas the five-member solution indicates high discriminative power because the envelopes do not overlap during the rainy season. The case of the intentionally pre-selected five members demonstrates the need to be cautious with LULCC-induced inferences, particularly for precipitation.

Apart from the intentionally selected ensemble, in the following we show how to estimate the critical number of ensemble members to be used to avoid such misinterpretations without having any *a priori* knowledge of the ensemble variance induced by each single ensemble member. For demonstration purposes, we focus our analysis on November precipitation because the intentionally

reduced 5-member solution suggests a high potential for erroneous conclusions for this month.

We applied a bootstrap test of 1000 samples to check how many arbitrarily chosen ensemble members are required to come to robust conclusions, i.e. conclusions resembling the statistics of the 15-members ensemble, the population of our sample.

The bootstrap sampling distribution of SNR values for November precipitation using 3 and 15 members is shown for Da Nang, demonstrating that for higher ensemble sizes the mean of the sampled distribution moves closer to the empirical value obtained from the control. This means that the null hypothesis (SNR values come from the same distribution) cannot be rejected even for very conservative assumptions about the level of significance α, whereas for

small ensemble sizes, the probability is very high that H_0 is rejected, even if a very high α is assumed (Figure S3).

Another simulation has been performed using the SLUCM parametrization for the urban class. The results for temperature and precipitation at Da Nang station can be found in Figure S4. It can be seen that SLUCM reduces the signal for temperature compared to the bulk urban parametrization, but there is still a significant increase compared to the control. For precipitation, it can be observed that SLUCM reduces the effect for November precipitation, i.e. the simulation is also closer to the control. However, there is no possibility of validating the performance of both parametrizations in this case.

3.2.2. Expected impact on the spatial extent of the meteorological variables

After analysing the impact of urbanization on the annual cycle in the region where the LULC is changed, the spatial impacts are analysed in the following, i.e. the question is addressed of whether the urbanization around Da Nang only has an impact in the direct vicinity of Da Nang or whether consequences in remote regions have to be expected.

For this reason, the SNR values are calculated for every grid cell in domain 3 of the WRF simulations. Figure 10 shows the SNR values for the different urbanization experiments for the surface temperature and the latent and sensible heat fluxes using a 15-member solution, which is appropriate because clear signals can be expected from a visual inspection of the time series in the previous section (Section 3.2). The values can range between 0 and approximately 24 for the urbanization scenarios. The SNR gradually changes, depending on the degree of urbanization (i.e. the radii of land-use conversion around Da Nang station). Only little difference is found between the Urban-14 and Urban-9 experiments, whereas the Urban-20 experiments shows more pronounced signals. For surface air temperature and sensible heat flux, the magnitude of the SNR is smaller (on average between 1 and 2), but more statistically significant grid points are found beyond the radius of LULCC compared to the latent head flux. This is particularly true for the Urban-14 and the Urban-09 scenarios. In general, H_0 can only be rejected within the region of the LULC conversion and grid cells in the vicinity only: no remote effects can be observed. The identified signals can be explained by changes of the surface energy balance, which is mainly triggered by the increased albedo through urbanization.

For precipitation, and based on the full ensemble, relatively higher November precipitation SNR values, up to 0.6, are found for remote regions, e.g. for the southwestern part of the VGTB basin (Figure 11). There, precipitation is expected to be decreased by about 400 mm compared to the November precipitation of the control, whereas rainfall increases of up to 250 mm can be expected in the Northeast (the region of Da Nang) and East coastal region, and also outside the basin. However, no coherent spatial patterns can be identified for which the null hypothesis is rejected at $\alpha = 99$ and 95.
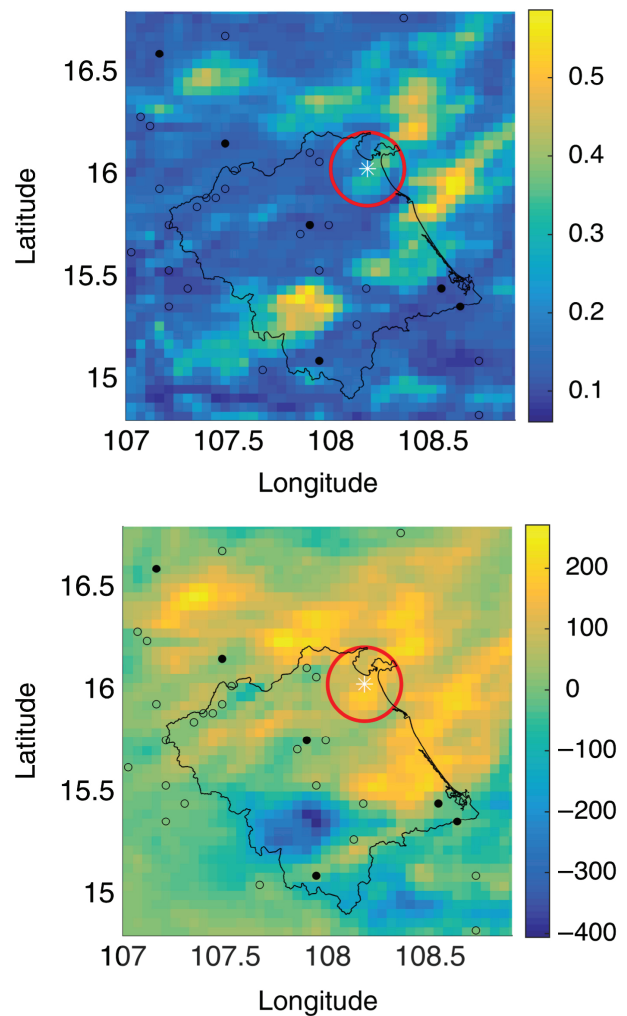

Figure 11. SNR values [.] of November precipitation for the urbanization scenario Urban-20 (top) and the simulated impacts for precipitation (Urban-20 *minus* control) (bottom), both based on the full ensemble. The stippled grid cells show the statistically significant grid cells in which the bootstrap test rejects the null hypothesis of having the same mean values for the scenario and the control simulations at $\alpha = 0.05$ (empty circles) and $\alpha = 0.01$ (filled circles).

It is found that the spatial results strongly depend on the number of ensemble members, not only for the magnitude of the SNR values but also for the statistically significant patterns (Figure 12). Based on only two (arbitrarily sampled) ensemble members, one would reject H_0 for large parts of the domain, although H_0 would still hold true for large parts of the northern VGTB basin using 5 and 11 members. Using 12 or more members displays only few statistically significant grid cells in regions outside of the radius of urbanization, thus suggesting only weak spatial impacts.

From the analysis above, we have learned about the influence of the ensemble size on our results for the SNR, not only in terms of their magnitude, but their statistical significance as well. Although it is not good scientific practice to anticipate the decision of a statistical test, we would conjecture that the level of significance α to be prescribed to reject H_0 should depend on the ensemble size. In this case, this procedure might be well justified since we know
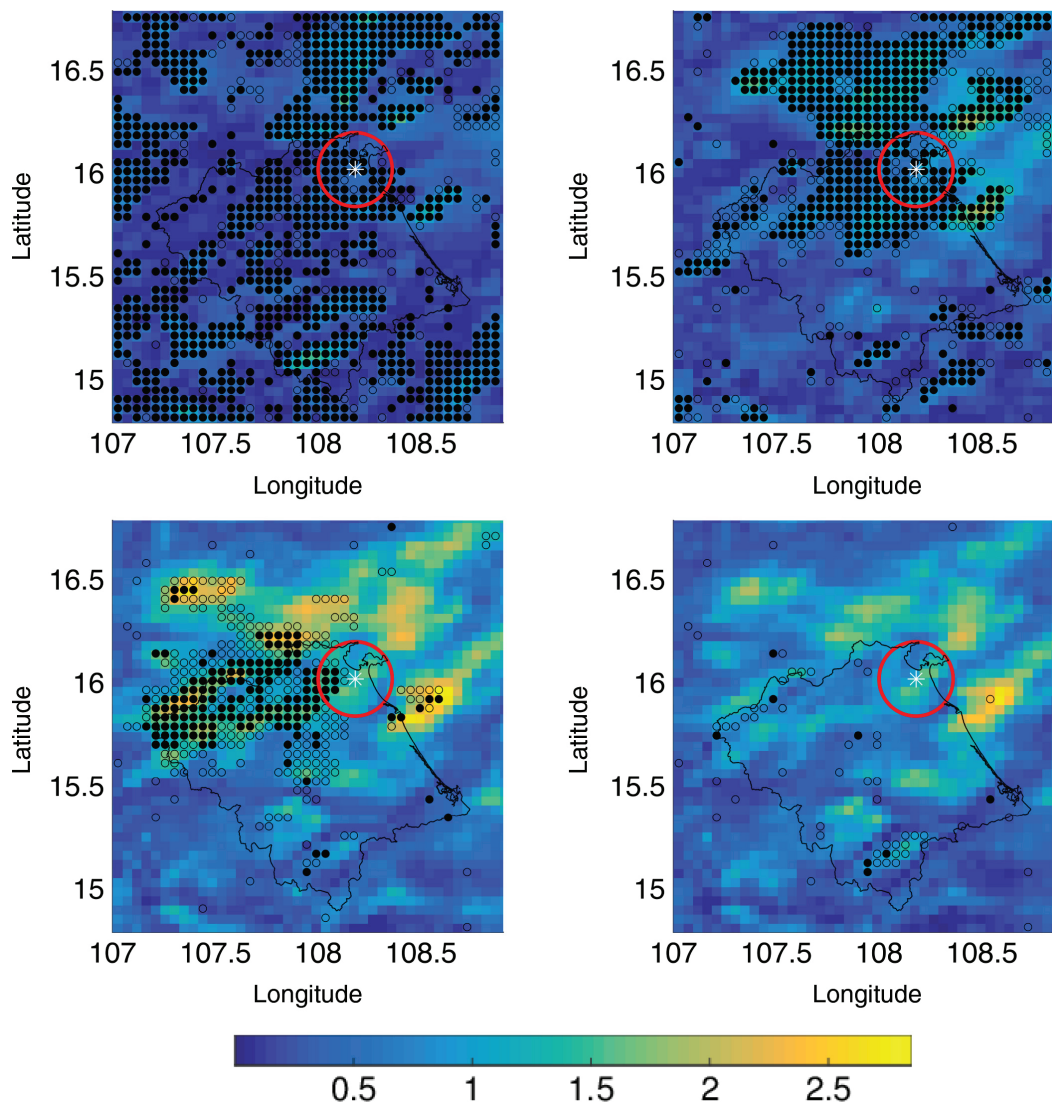
Figure 12. SNR values [.] of November precipitation as obtained for selected ensemble sizes of 2, 5, 11, and 12 members. Bootstrap tests are performed to test the statistical significance of rejecting H_0, at $\alpha = 0.05$ (empty circles) and $\alpha = 0.01$ (filled circles).

the test decision from the ensemble population (15 members), i.e. the rejection of the null hypothesis. The objective of this procedure is to identify the number of ensemble members needed to come to a robust decision about the impact of an LULCC. For the grid cell corresponding to Da Nang, a 6- and 7-member solution suggests already a rejection of H_0. However, this number is not found to be robust. Finally, at least 12 members are required to robustly reject H_0 (Figure S5). Note that this does not necessarily mean that H_1 can be accepted without any doubts. High α values only cast doubt on the validity of H_0.

This approach is applied to the whole domain (see Figure 13). It can be seen that the ensemble size needed to reject H_0 may differ remarkably across the domain. For large parts of the VGTB basin, an ensemble size of 12–14 members is required. Exceptions are the eastern and western regions of the VGTB basin, where H_0 can be robustly rejected for ensemble sizes of 2–6.

Figure S6 shows the potential impact of urbanization on the VGTB basin using two different parametrizations for urban. Heterogeneous and partly very strong differences in the annual precipitation can be observed, mainly in the southern part of the basin, whereas the differences are relatively small and homogeneous for air temperature within the converted area and remote areas. It must be noted that only one simulation was performed to estimate the impact of the different urbanization schemes, thus it does not take into account the internal model variability in this case.

3.3. Case study II: ensemble WRF simulations using deforestation scenarios of Nong Son

3.3.1. Expected impact on the annual cycle of the meteorological variables

In contrast to the pronounced signals identified in the urbanization experiments, the comparison between the deforestation scenarios and the control simulations shows no clear differences in surface air temperature or latent and sensible heat fluxes (Table 2). Figure 14 shows the time
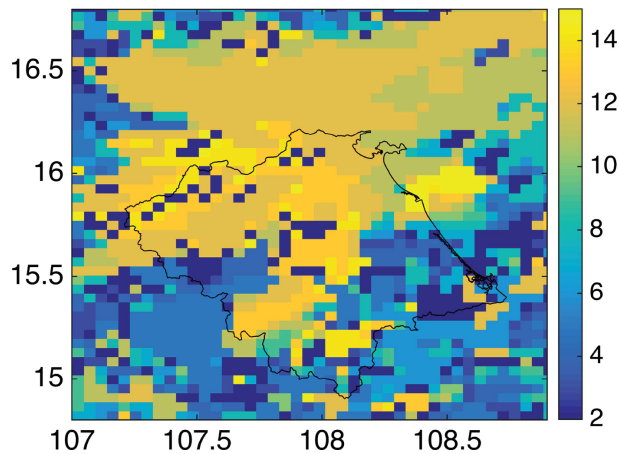
Figure 13. Minimum number of ensemble members required to robustly reject the H_0 hypothesis (SNR values of November precipitation from equal distribution). More information and the example for the grid cell corresponding to Da Nang is given in Figure S5.

series of the monthly surface air temperature, latent heat flux, and sensible heat flux for the three different deforestation scenarios, averaged over the next 9 grid cells around Nong Son station. For the surface air temperature, there is nearly no visible difference between the deforestation scenarios and the control simulations, no matter which degree of deforestation is considered. The mean differences range between +0.1 and −0.1 °C. For the latent heat flux, differences between the control simulations and the deforestation ensembles can be found for June and July, 2010. In the deforestation simulations, the mean values are increased and the variability is decreased. For Deforest-09, however, the variability of the ensemble is slightly higher compared to the control ensemble. For the sensible heat flux, the deforestation ensembles show decreased mean values for the period June to August 2010. There are no significant changes in the partition of surface heat fluxes for the deforestation scenarios (see Figure 8, right column). From previous studies, one would expect a remarkable change, either a reduction of the latent heat flux, leading to a net warming, as frequently observed in tropical regions (e.g. Lawrence and Chase, 2010), or an evaporative cooling during the season when the crops have higher evaporation rates, as observed in temperate regions (e.g. Baldocchi *et al.*, 1997).

In general, the temporal illustrations suggest that the deforestation ensembles do not offer significant changes (signal) in the simulated meteorological variables.

Similar to the urbanization scenarios, the impact of the ensemble size on the variance is investigated for precipitation (Figure 15). Intentionally selected 5 members up to 12 members reveal robust variances before they increase markedly (Figure S7).

3.3.2. Expected impact on the spatial extent of the meteorological variables

Figure 16 illustrates the impact of deforestation (Deforest-20 scenario) in a spatial context. Apart from

the urbanization scenarios, deforestation does not clearly affect the surface air temperature or the latent and sensible heat fluxes of many grid cells: neither in nor outside the domain where forest is converted to crop land. The smallest SNR values are identified for air temperature (up to 0.5), and range from 0 to 1 for all variables. The signal is restricted to the converted area exclusively, and fewer grid cells are found to reveal a statistically significant signal compared to the urbanization scenarios. No pronounced signals for precipitation are found in the deforestation scenarios.

4. Summary and conclusions

The sensitivity of the effects of LULCCs on different meteorological variables at regional scales has been analysed by implementing different LULCC scenarios in the regional climate model WRF. For this purpose, urbanization and deforestation scenarios of different degrees (in circles with varying radii of 20, 14, and 9 km) have been used, and RCM simulations conducted for the year 2010. An ensemble approach using PICs was followed to estimate the noise in RCM simulation. The signal induced by an LULCC is then quantified by calculating the SNR and tested for statistical significance.

For our study region in Central Vietnam, the urbanization experiments (conversion from cropland to urban) revealed stronger impacts compared to deforestation (conversion from forest to cropland). The surface temperatures of the urbanization experiments are constantly higher than those of the control simulations throughout the entire year 2010. The mean of the latent heat flux decreases remarkably with the urbanization parametrization applied, ranging from 74.7 for the control to 4.3, 13.9, and 14.6 W m^{-2} for Urban-20, Urban-14 and Urban-09, respectively, i.e. the partitioning of the surface heat fluxes changed tremendously. In the opposite direction to the results of the latent heat flux, the mean values of the sensible heat fluxes are nearly doubled in the experiments of the urbanization scenario.

In contrast to the pronounced signals identified in the urbanization experiments, the deforestation scenarios show no clear differences in surface air temperature or the latent and sensible heat fluxes from those of the control simulations. Only marginal differences in the partitioning of the surface heat fluxes into sensible and latent heat fluxes are observed, which is possibly related to the very broad representation of the different land-use classes and the improper parametrization of the vegetation parameters in the Noah LSM. In addition, the effect induced by deforestation is found to be less spread in space compared to the urbanization scenario.

The degree of conversion, assessed as the radius of change (in km) around a location, does affect the results. A change through a radius of 20 km leads to the most pronounced changes (in both case studies), while radii of 14 and 9 lead to similar results as each other. This might be affected by the horizontal resolution of the
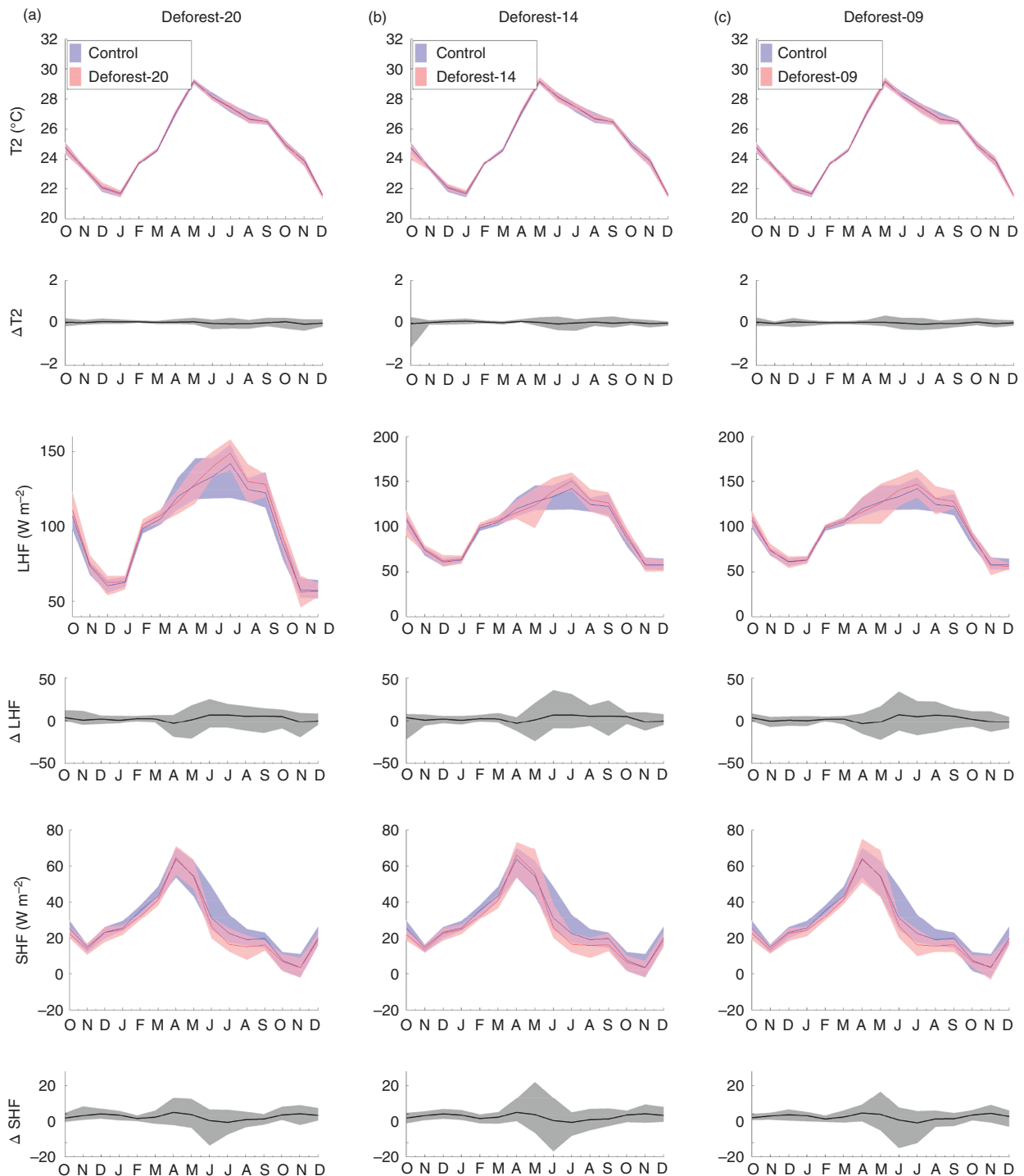
Figure 14. Monthly surface air temperature, latent heat flux, and sensible heat flux (from top to bottom) for the three different deforestation scenarios averaged over the next 9 grid cells around Nong Son station. The envelopes for the control (blue) are based on the observed LUCCi LU map 2010, the three different deforestation scenarios [(a) Deforest-20, (b) Deforest-14, and (c) Deforest-09] are derived from ensemble WRF simulations using 15 ensemble members with different initialization conditions (red). The subfigures illustrate the range of differences between the LUC scenarios and the control simulations. The WRF simulations were run from October 2009–December 2010.

WRF in the inner domain (5 km), and might be additionally related to the critical areal conversion to affect the mesoscale circulation (Pielke *et al.*, 2011). More research on this is, however, necessary to confirm these speculations.

For precipitation, although some spatial patterns are visible in the ensemble, SNR values smaller than unity suggest no or only little evidence of changes due to urbanization. Based on the conducted statistical bootstrap tests, the hypothesis that the control simulation and the LULCC simulations are from the same population cannot be rejected at $\alpha = 0.95$, if the whole ensemble (variance) is considered. Moreover, precipitation is found to be very sensitive to the urbanization scheme applied.
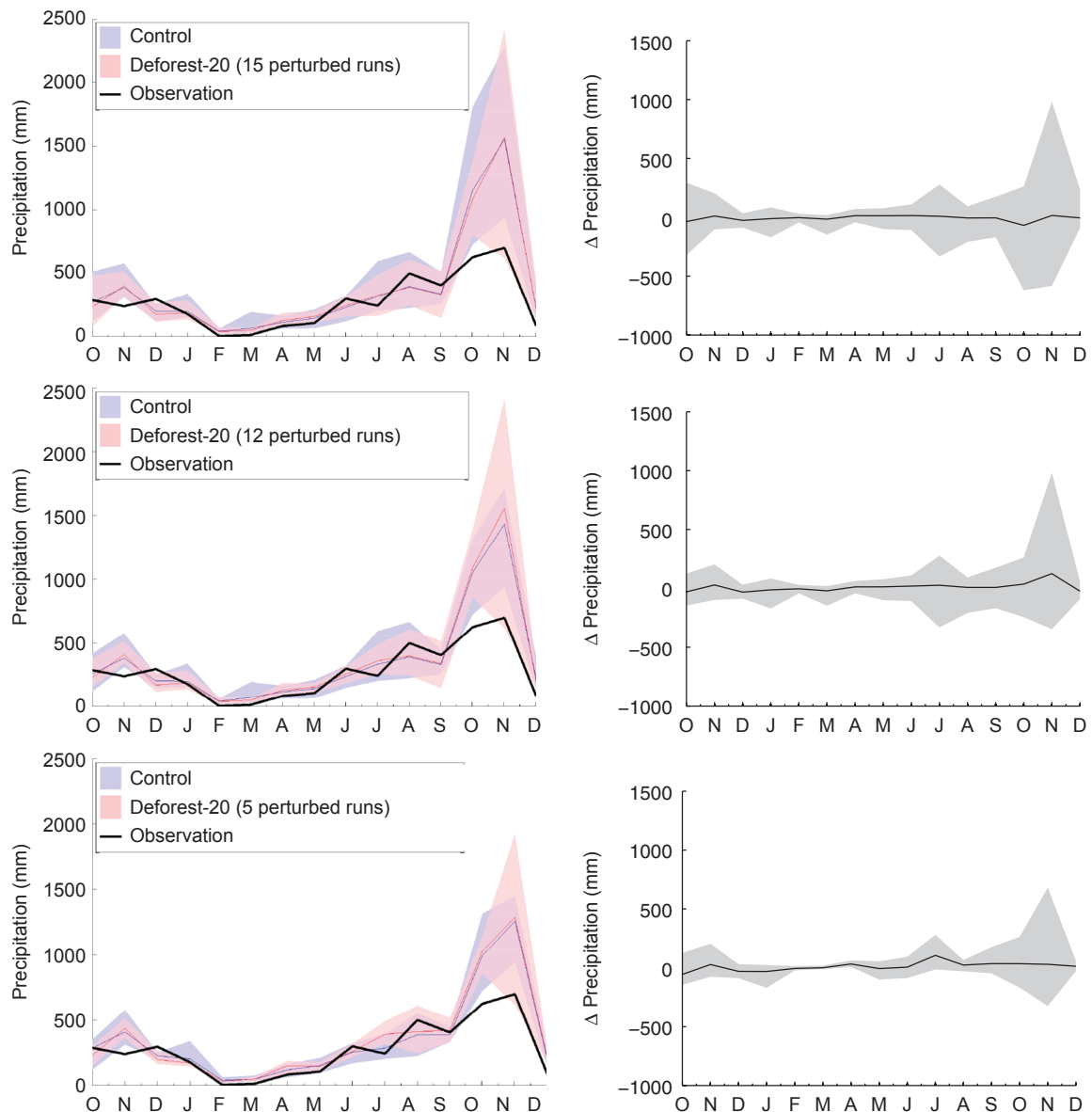
Figure 15. Monthly total precipitation [mm] for the Deforest-20 scenario and the control simulations at Nong Son station in 2010 using 15-, 12-, and 5-member ensembles. The surrounding 9 grid cells of Nong Son are used (left). Differences between the scenario and the control simulations (Deforest-20 minus control).

Based on the results obtained in this study, we draw the following conclusions:

- Ensemble simulations based on perturbed initial conditions in combination with the statistical performance measures and tests presented in this study provide a strategy to attribute and quantify RCM-based LULCC-induced impacts on the regional climate.
- For the urbanization scenarios in Central Vietnam, the impact on the simulated air temperature and the latent and sensible heat fluxes at surface are large, and can be clearly attributed to the LULCC parametrization. Only a few ensemble members are required to derive robust effects of LULCC. This is not the case for the deforestation scenarios.
- For precipitation, no clear effects can be distinguished from the internal model variability. This is indicated

by the small magnitudes of the SNR and confirmed by the applied statistical tests. If small ensembles are considered, the chances for erroneous conclusions about the effects of LULCC on precipitation patterns are increased. Although the results may depend on specific factors, such as the region of interest and specific RCM settings (e.g. land-use class parametrizations), we recommend conducting RCM ensemble simulations to study the internal model variability so as to extract robust effects. This is crucial for climate impact studies, in particular if precipitation is considered.

5. Limitations and future research needs

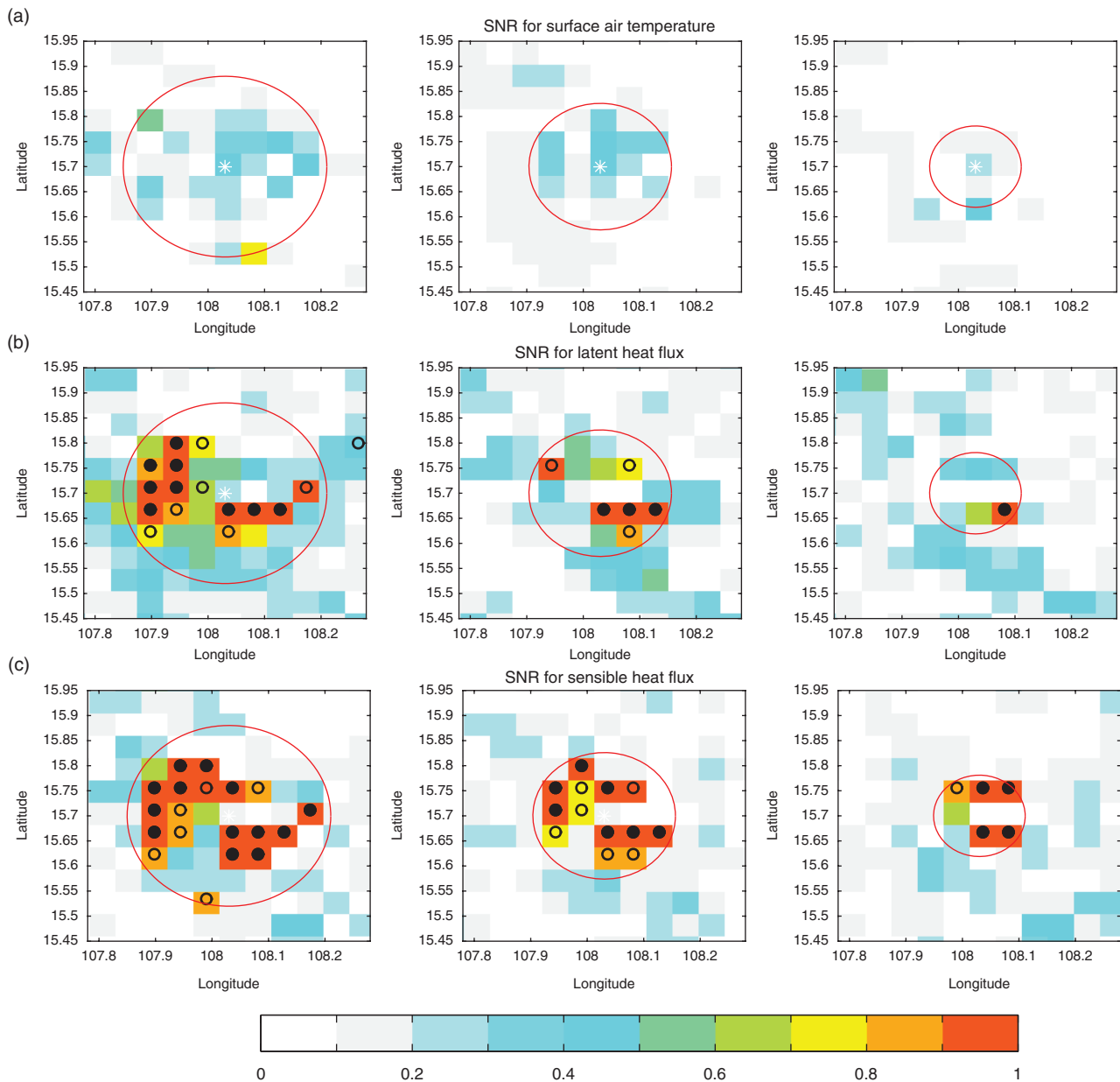Several aspects of our analysis can be improved in future work, summarized as follows:

Figure 16. SNR values [.] of the different urbanization scenarios (Deforest-20, Deforest-14 and Deforest-09, from left to right) for (a) surface air temperature, (b) latent heat flux, and (c) sensible heat flux. More information about the deforestation scenarios can be found in Section 2.4. The stippled grid cells show the statistically significant grid cells in which the *t*-test rejects the null hypothesis of having the same mean values for the LULCC scenarios and the control simulations at $\alpha = 0.05$ (empty circles) and $\alpha = 0.01$ (filled circles).

Only one RCM model driven with the boundary conditions of one GCM has been used. It is suspected that various RCM results need to be analysed to obtain more robust conclusions about LULCC-induced impacts. The same holds true for the applied LSM. Therefore, the implementation of various LSMs, including a prior offline validation, is suggested. High-resolution remote sensing products can be used to better parametrize the LSM schemes. Within-ensemble variability is suppressed because SSTs and boundary conditions are prescribed by the GCM. Other perturbation methods should therefore be implemented in the ensemble generation.

Acknowledgements

Supporting information

The following supporting information is available as part of the online article:

Figure S1. Lag-1 (top) and lag-2 autocorrelation (bottom), exemplary shown for the monthly mean temperatures of the Control simulation for the year 2000. It can be seen that a first first-order autoregressive model is found to be suitable to remove the serial dependence across the VGTB basin appropriately. The lag-1 autocorrelation coefficients are used to correct for the variances estimated in the *t*-test for each variable and grid cell, respectively. Similar autocorrelation coefficients and patterns are found for the LULCC-induced WRF simulations. Autocorrelation coefficients were found to be generally weaker for other variables.

Figure S2. Standard deviation of annual precipitation [mm] at the Da Nang station for the urbanization scenario Urban-20 depending on different ensemble sizes for the year 2010. The three-member solution, e.g. comprises the members 14, 8, and 3 (see Table 3 for more information).

Figure S3. Probability density functions (PDF) obtained by 1000 bootstrap realizations from the LULCC-induced simulations (Urbanization-20) of November precipitation at Da Nang for 15 ensemble members (black line) and 3 ensemble members (blue line). The red vertical line illustrates the empirical SNR value from the 15 members of the Control simulations, i.e. no LULCC, and serves as reference value for statistical significance tests. This value is close to the mean value of the bootstrap PDF of the 15-member simulations, leading to an acceptance of *H0* (similar distributions), whereas it is not covered by the sampled distribution of the three members. In this case, *H0* is rejected, suggestion different distributions and a LULCC-induced SNR value. Please note that the estimated PDFs are based on the applied smoothing algorithm (Bowman AW, Azzalini A. 1997. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press Inc.: New York), leading to the artefact of negative values as e.g. visible for the lower tail of the PDF distribution of the 15-member bootstrap realizations.

Figure S4. Monthly total precipitation amount (top) and monthly mean surface air temperature (bottom) from October 2009 to December 2010 for the Urban-20 scenario at Da Nang station, using a single non-perturbed WRF simulation based on the bulk urban parameterization and the Single-Layer Urban Canopy Model (SLUCM). The surrounding 9 grid cells of Da Nang are used. It is shown that precipitation is reduced in November 2010 for SLUCM.

Figure S5. Level of significance α (in %) at which the *H0* hypothesis (SNR values of November precipitation from equal distribution) is rejected as a function of the number of ensemble members, shown for Da Nang station. It shows that the probability to reject *H0* is low for six and seven ensemble members, increases again to probabilities $\alpha > 80\%$, and finally decreases again to low α values. Statistically robust ensemble sizes to reject *H0* are values ≥ 12. Please note that this does not necessarily mean that *H1* can be accepted without any doubts. High α values only cast doubt on the validity of *H0*.

Figure S6. Difference of annual precipitation amount (top) and annual mean surface air temperature (bottom) between non-perturbed WRF simulation using bulk urban parameterization and non-perturbed WRF simulation using Single-Layer Urban Canopy Model (SLUCM), i.e. bulk urban *minus* SLUCM. It can be seen that the parameterization heavily impacts on the precipitation with highest differences in remote regions, whereas the differences for temperature are locally restricted to the region of LULCC.

Figure S7. Standard deviation of annual precipitation [mm] at the Nong Son station for the deforestation scenario Deforest-20 depending on different ensemble sizes for the year 2010. The three-member solution, e.g. comprises the members 5, 14, and 15 (see Table 4 for more information).

References

Argüeso D, Evans JP, Pitman AJ, Di Luca A. 2015. Effects of city expansion on heat stress under climate change conditions. *PLoS One* **10**(2): e0117066, doi: 10.1371/journal.pone.0117066.

Arnfield a J. 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **23**(1): 1–26, doi: 10.1002/joc.859.

Baehr J, Piontek R. 2014. Ensemble initialization of the oceanic component of a coupled model through bred vectors at seasonal-to-interannual timescales. *Geosci. Model Dev.* **7**(1): 453–461, doi: 10.5194/gmd-7-453-2014.

Baldocchi DD, Vogel CA, Hall B. 1997. Seasonal variation of energy and water vapor exchange rates above and below a boreal jack pine forest canopy. *J. Geophys. Res.* **102**: 28939–28951, doi: 10.1029/96JD03325.

Beltrán-Przekurat A, Pielke RA Sr, Eastman JL, Coughenour MB. 2012. Modelling the effects of land-use/land-cover changes on the near-surface atmosphere in southern South America. *Int. J. Climatol.* **32**(8): 1206–1225.

Betts RA. 2001. Biogeophysical impacts of land use on present-day climate: near-surface temperature change and radiative forcing. *Atmos. Sci. Lett.* **2**(1–4): 39–51, doi: 10.1006/asle.2001.0023.

Betts RA, Falloon PD, Goldewijk KK, Ramankutty N. 2007. Biogeophysical effects of land use on climate: model simulations of radiative forcing and large-scale temperature change. *Agric. For. Meteorol.* **142**(2–4): 216–233, doi: 10.1016/j.agrformet.2006.08.021.

Boisier JP, de Noblet-Ducoudré N, Pitman AJ, Cruz FT, Delire C, van den Hurk BJJM, van der Molen MK, Müller C, Voldoire A. 2012. Attributing the impacts of land-cover changes in temperate regions on surface temperature and heat fluxes to specific causes: results from the first LUCID set of simulations. *J. Geophys. Res. Atmos.* **117**(D12): D12116, doi: 10.1029/2011JD017106.

Cao Q, Yu D, Georgescu M, Han Z, Wu J. 2009. Impacts of land use and land cover change on regional climate: case study in Northern China. *Environ. Res. Lett.* **11**(12): 13945, doi: 10.1088/1748-9326/10/12/124025.

Chen F, Dudhia J. 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. *Mon. Weather Rev.* **129**(4): 569–585.

Chen F, Kusaka H, Bornstein R, Ching J, Grimmond CSB, Grossman-Clarke S, Loridan T, Manning KW, Martilli A, Miao S, Sailor D, Salamanca FP, Taha H, Tewari M, Wang X, Wyszogrodzki AA, Zhang C. 2011. The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *Int. J. Climatol.* **31**: 273–288, doi: 10.1002/joc.2158.

Christensen OB, Gaertner MA, Prego JA, Polcher J. 2001. Internal variability of regional climate models. *Clim. Dyn.* **17**(11): 875–887.

Davin EL, de Noblet-Ducoudré N. 2010. Climatic impact of global-scale deforestation: radiative versus nonradiative processes. *J. Clim.* **23**(1): 97–112, doi: 10.1175/2009JCLI3102.1.

Davin EL, de Noblet-Ducoudré N, Friedlingstein P. 2007. Impact of land cover change on surface climate: relevance of the radiative forcing concept. *Geophys. Res. Lett.* **34**(13): 1–5, doi: 10.1029/2007GL029678.

Deng, X, Zhao C, Yan H. 2013. Systematic modeling of impacts of land use and land cover changes on regional climate: a review. *Adv. Meteorol.*, **2013**: 317678, doi:10.1155/2013/317678.

Denis B, Laprise R, Caya D. 2003. Sensitivity of a regional climate model to the resolution of the lateral boundary conditions. *Clim. Dyn.* **20**: 107–126, doi: 10.1007/s00382-002-0264-6.

Dirmeyer PA, Niyogi D, de Noblet-Ducoudré N, Dickinson RE, Snyder PK. 2010. Impacts of land use change on climate. *Int. J. Climatol.* **30**: 10–1002.

Ellis EC, Goldewijk KK, Siebert S, Lightman D, Ramankutty N. 2010. Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* **19**(5): 589–606, doi: 10.1111/j.1466-8238.2010.00540.x.

Feddema JJ, Oleson KW, Bonan GB, Mearns LO, Buja LE, Meehl GA, Washington WM. 2005. The importance of land-cover change in simulating future climates. *Science* **310**(5754): 1674–1678.

Giannaros TM, Melas D, Daglis IA, Keramitsoglou I, Kourtidis K. 2013. Numerical study of the urban heat island over Athens (Greece) with the WRF model. *Atmos. Environ.* **73**: 103–111, doi: 10.1016/j.atmosenv.2013.02.055.

Giorgi F, Bi X. 2000. A study of internal variability of a regional climate model. *J. Geophys. Res. Atmos.* **105**(D24): 29503–29521.

Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik J-J, Belcher SE, Bohnenstengel SI, Calmet I, Chen F, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Shashua-Bar L, Steeneveld G-J, Trombou M, Voogt J, Young D, Zhang N. 2010. The international urban energy balance models comparison project: first results from phase 1. *J. Appl. Meteorol. Climatol.* **49**(6): 1268–1292, doi: 10.1175/2010JAMC2354.1.

Haughton N, Abramowitz G, Pitman A, Phipps SJ. 2014. On the generation of climate model ensembles. *Clim. Dyn.* **43**: 2297–2308, doi: 10.1007/s00382-014-2054-3.

Hong S, Lakshmi V, Small EE, Chen F, Tewari M, Manning KW. 2009. Effects of vegetation and soil moisture on the simulated land surface processes from the coupled WRF/Noah model. *J. Geophys. Res. Atmos.* **114**(18): 1–13, doi: 10.1029/2008JD011249.

Kusaka H, Kondo H, Kikegawa Y, Kimura F. 2001. A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Bound.-Layer Meteorol.* **101**(3): 329–358.

Laux P, Lorenz C, Thuc T, Ribbe L, Kunstmann H. 2013. Setting up regional climate simulations for Southeast Asia. In *High Performance Computing in Science and Engineering '12*, *Conference Proceedings*, Springer: Stuttgart, Germany, 391–406.

Laux P, Thuc T, Kunstmann H. 2014. High resolution climate change information for the Lower Mekong River Basin of Southeast Asia. In *High Performance Computing in Science and Engineering '13*, *Conference Proceedings*, Springer: Stuttgart, Germany, 543–551.

Lawrence PJ, Chase TN. 2010. Investigating the climate impacts of global land cover change in the community climate system model. *Int. J. Climatol.* **30**(13): 2066–2087.

Lee SH, Kim SW, Angevine WM, Bianco L, McKeen SA, Senff CJ, Trainer M, Tucker SC, Zamora RJ. 2011. Evaluation of urban surface parameterizations in the WRF model using measurements during the Texas Air Quality Study 2006 field campaign. *Atmos. Chem. Phys.* **11**(5): 2127–2143.

Liu Y, Chen F, Warner T, Basara J. 2006. Verification of a mesoscale data assimilation and forecasting system for the Oklahoma city area during the joint urban 2003 field project. *J. Appl. Meteorol.* **45**: 912–929.

Loridan T, Grimmond CSB, Grossman-Clarke S, Chen F, Tewari M, Manning K, Martilli A, Kusaka H, Best M. 2010. Trade-offs and responsiveness of the single-layer urban canopy parametrization in WRF: an offline evaluation using the MOSCEM optimization algorithm and field observations. *Q. J. R. Meteorol. Soc.* **136**(649): 997–1019, doi: 10.1002/qj.614.

Mahmood R, Pielke R. 2014. Land cover changes and their biogeophysical effects on climate. *Int. J. Climatol.* **34**: 929–953.

Moore N, Torbick N, Lofgren B, Wang J, Pijanowski B, Andresen J, Kim D-Y, Olson J. 2010. Adapting MODIS-derived LAI and fractional cover into the RAMS in East Africa. *Int. J. Climatol.* **30**(13): 1954–1969.

Moore N, Alagarswamy G, Pijanowski B, Thornton P, Lofgren B, Olson J, Andresen J, Yanda P, Qi JG. 2012. East African food security as influenced by future climate change and land use change at local to regional scales. *Clim. Change* **110**(3–4): 823–844, doi: 10.1007/s10584-011-0116-7.

Müller WA, Baehr J, Haak H, Jungclaus HJ, Kröger J, Matei D, Notz D, Pohlmann H, von Storch JS, Marotzke J. 2012. Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophys. Res. Lett.* **39**(22): 1–7, doi: 10.1029/2012GL053326.

de Noblet-Ducoudré N, Boisier J-P, Pitman A, Bonan GB, Brovkin V, Cruz F, Delire C, Gayler V, van den Hurk BJJM, Lawrence PJ, van der Molen MK, Müller C, Reick CH, Strengers BJ, Voldoire A. 2012. Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: results from the first set of LUCID experiments. *J. Clim.* **25**(9): 3261–3281, doi: 10.1175/JCLI-D-11-00338.1.

Pielke RA, Pitman A, Niyogi D, Mahmood R, McAlpine C, Hossain F, Goldewijk KK, Nair U, Betts R, Fall S, Reichstein M, Kabat P, de Noblet N. 2011. Land use/land cover changes and climate: modeling analysis and observational evidence. *Wiley Interdiscip. Rev. Clim. Chang.* **2**(6): 828–850, doi: 10.1002/wcc.144.

Pitman AJ, de Noblet-Ducoudré N, Cruz FT, Davin EL, Bonan GB, Brovkin V, Claussen M, Delire C, Ganzeveld L, Gayler V, van den Hurk BJJM, Lawrence PJ, van der Molen MK, Müller C, Reick CH, Seneviratne SI, Strengers BJ, Voldoire A. 2009. Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. *Geophys. Res. Lett.* **36**(14): 1–6, doi: 10.1029/2009GL039076.

Rounsevell MDA, Arneth A, Alexander P, Brown DG, de Noblet-Ducoudré N, Ellis E, Finnigan J, Galvin K, Grigg N, Harman I, Lennox J, Magliocca N, Parker D, O'Neill BC, Verburg PH, Young O. 2014. Towards decision-based global land use models for improved understanding of the Earth system. *Earth Syst. Dyn.* **5**: 117–137, doi: 10.5194/esd-5-117-2014.

Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ. 2010. Investigating soil moisture-climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**(3–4): 125–161, doi: 10.1016/j.earscirev.2010.02.004.

Sertel E, Robock A, Ormeci C. 2010. Impacts of land cover data quality on regional climate simulations. *Int. J. Climatol.* **30**(13): 1942–1953.

Souvignet M, Laux P, Freer J, Cloke H, Thinh DQ, Thuc T, Cullmann J, Nauditt A, Flügel WA, Kunstmann H, Ribbe L. 2014. Recent climatic trends and linkages to river discharge in Central Vietnam. *Hydrol. Process.* **28**: 1587–1601, doi: 10.1002/hyp.9693.

Takahashi HG, Yoshikane T, Hara M, Takata K, Yasunari T. 2010. High-resolution modelling of the potential impact of land surface conditions on regional climate over Indochina associated with the diurnal precipitation cycle. *Int. J. Climatol.* **30**: 2004–2020, doi: 10.1002/joc.2119.

Trail M, Tsimpidi AP, Liu P, Tsigaridis K, Hu Y, Nenes A, Stone B, Russell AG. 2013. Potential impact of land use change on future regional climate in the Southeastern U.S.: Reforestation and crop land conversion. *J. Geophys. Res. Atmos.* **118**(20): 11577–11588, doi: 10.1002/2013JD020356.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Academic Press, 467 pp.

Wu W, Lynch AH, Rivers A. 2005. Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions. *J. Clim.* **18**(7): 917–933.