Article

Impacts of Spatial Zonation Schemes on Yield Potential Estimates at the Regional Scale

Hao Xu [†] , Fen Huang [†], Wenjun Zuo, Yongchao Tian, Yan Zhu , Weixing Cao * and Xiaohu Zhang *

National Engineering and Technology Center for Information Agriculture, Key Laboratory for Crop System Analysis and Decision Making, Ministry of Agriculture and Rural Affairs, PRC, Jiangsu Key Laboratory for Information Agriculture, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, 1 Weigang Road Nanjing, Nanjing 210095, China; haoxu1989@hotmail.com (H.X.); fenhuan@njau.edu.cn (F.H.); zoewenjuzuo@hotmail.com (W.Z.); yctian@njau.edu.cn (Y.T.); yanzhu@njau.edu.cn (Y.Z.)

* Correspondence: caow@njau.edu.cn (W.C.); zhangxiaohu@njau.edu.cn (X.Z.)

[†] The authors contributed equally to the work.

Received: 29 January 2020; Accepted: 28 April 2020; Published: 30 April 2020

Abstract: Simulations based on site-specific crop growth models have been widely used to obtain regional yield potential estimates for food security assessments at the regional scale. By dividing a region into nonoverlapping basic spatial units using appropriate zonation schemes, the data required to run a crop growth model can be reduced, thereby improving the simulation efficiency. In this study, we explored the impacts of different zonation schemes on estimating the regional yield potential of the Chinese winter wheat area to obtain the most appropriate spatial zonation scheme of weather sites therein. Our simulated results suggest that the upscaled site-specific yield potential is affected by the zonation scheme and by the spatial distribution of sites. As such, the distribution of a small number of sites significantly affected the simulated regional yield potential under different zonation schemes, and the zonation scheme based on sunshine duration clustering zones could effectively guarantee the simulation accuracy at the regional scale. Using the most influential environmental variable of crop growth models for clustering can get the better zonation scheme to upscale the site-specific simulation results. In contrast, a large number of sites had little effect on the regional yield potential simulation results under the different zonation schemes.

Keywords: Regional yield potential; crop growth model; spatial zonation; upscaling

1. Introduction

Yield potential refers to potential productivity, which is entirely determined by temperature and photosynthetically active radiation when nutrients, moisture, soils, cultivars, and other agricultural technological parameters are at optimum conditions [1,2]. Yield potential estimations at the regional scale can identify the variations in the upper yield limit, optimize the planting system, and improve the use efficiency of agroclimatic resources, thereby providing information for agricultural impact and risk assessment to support policymaking [1–5]. Additionally, the process-based crop growth model as a robust, generic, and cost-effective tool to simulate yields under a range of agricultural and climatic scenarios, in this context, crop growth models have been extensively used to estimate yield potentials over large areas. Currently, two main upscaling methods are employed for regional applications of crop growth models. The first upscaling method links a site-specific crop growth model with regional gridded data (rain, temperature, radiation, soil, etc.) and runs the crop growth model for each grid cell, a basic homogeneous spatial unit, to obtain the regional yields [4,6–8]. However, while this method can

reflect the spatial trends of the simulated yields, crop growth models are highly sensitive to changes in the spatial resolution of the data, and thus, it is necessary to preliminarily determine the appropriate spatial resolution [4,8]. Additionally, this approach requires interpolation of the meteorological data inside each grid cell, which introduces errors in the data and imposes large data storage requirements and computational costs [4,8,9]. The second upscaling method divides the study area into irregular basic spatial units that have the smallest spatial heterogeneities, and typical ecological sites within these units are then selected to adequately represent the area. Finally, the regional yield is calculated using the spatially weighted average of each site-specific simulation result [5,10–12]. The latter method has a high single-site simulation accuracy and superior computational efficiency because it uses meteorological, soil, cultivar, and management data acquired from a limited number of observation sites. Furthermore, considering the differences among and stability of regional agricultural production conditions, the zonation scheme is currently the most common method for upscaling the site-specific yield potential at the regional scale [1,9,10,13,14].

However, the second method also has some drawbacks; for example, different zonation schemes affect the results of upscaling the site-specific yield potential at the regional scale, i.e., the so-called modifiable areal unit problem (MAUP) that exists in the field of geography [15,16]. The spatial distribution of crops varies with climate, topography, and soil, and therefore, constitutes a complex, changing, and heterogeneous continuum with significant spatial heterogeneity that is affected by many human factors, such as policies. Therefore, theoretically, the regional yield simulated by a limited number of sites will introduce errors because of the spatial heterogeneity that exists in environmental data. An intuitive way to solve this problem is to increase the number of observation sites, but an excessive number of sites will increase the effort and financial costs and introduce redundant data [10]. Another method is to partition the study area based on prior knowledge or observation variables and thereby improve the representativeness of the site-specific simulation results for a region [9,10,17]. Most previous research that has used predetermined numbers and spatial layouts of observation sites has performed zonation based on meteorological factors and has used a variety of weight calculation methods (e.g., the planting area proportion, soil type proportion, and cropping system proportion) to upscale the yield potential simulation results of sites to the regional scale [5,9]. However, these methods are subject to the joint effects of the zonation scheme and the number and spatial distribution of sites; moreover, different zonation schemes will result in differences in the number and shape of basic spatial units. This directly affects the data upscaling results, and in regions with strong spatial heterogeneity, the number of sites must be sufficiently high to reduce the estimation errors at the regional scale [18]. In summary, based on the fact that the locations and number of sites are known in the existing data, previous studies have focused on directly upscaling the simulation results of sites to the regional scale by using a single zonation scheme without considering the impacts of different numbers of sites and different zonation schemes on the simulation results from site-specific crop growth models.

In this study, the Chinese winter wheat production area was used, and various regional yield potential simulation scenarios were constructed under different zonation schemes. The influences of the number of sites and the zonation schemes on the regional wheat yield potential estimation results were investigated to determine the most suitable scheme for upscaling the site-specific yield potential. The results of this study provide a reference for the evaluation and prediction of the regional grain yield potential based on crop growth models.

2. Materials and Methods

2.1. Study Area

The main winter wheat production area in China (102°46′–122°11′E, 28°13′–41°10′N) was the focus of this study (Figure 1a). The study area contains 14 provinces and municipalities (Figure 1b). In the study area, there are 538 meteorological sites and 129 ecological observation sites that record the major growth stages, growth indices, and sowing, fertilization, and other management practices (Figure 1b).

China's winter wheat region includes a diversity of climate zones, including the moderate-temperate semiarid, moderate-temperate semihumid, warm-temperate semihumid, warm-temperate semiarid, northern subtropical humid, and mid-subtropical humid zones, each of which exhibits distinct temperature and precipitation conditions [19]. Among these zones, the annual average temperature varies from 9 to 15 °C, and the annual precipitation varies from 440 to 980 mm in the northern winter wheat region (including subregions X, VIII, and XI) and the Huang-Huai winter wheat region (including subregions I, IX, and V). In contrast, in the winter wheat region near the middle and lower reaches of the Yangtze River (including subregions VII and IV) and the southwestern winter wheat region (including subregions II, III, and VI), the annual average temperature and annual precipitation vary from 16 to 25 °C and generally more than 1000 mm, respectively [20]. The landforms of the Chinese winter wheat area include plains, hills, mountains, and basins, all of which display distinct elevation differences from east to west (the highest elevation is 5174 m, and the lowest elevation is −142 m below sea level).

Figure 1. Geographical locations of the winter wheat area in China and typical sites for calibration (a); the distribution of the winter wheat production region, weather sites, and all typical ecological sites (b). The production region includes the provinces of Shandong (I), Jiangsu (II), Anhui (III), Gansu (IV), Henan (V), Tianjin (VI), Sichuan (VII), Shaanxi (VIII), Shanxi (IX), Hubei (X), Hebei (XI), Chongqing (XII), Beijing (XIII), and Shanghai (XIV).

2.2. Method Roadmap

A modeling approach was used as the basis for simulating the wheat yield potential at 538 sites throughout the study area from 2000 to 2009 [21–30]. A total of 17 schemes were defined using spatial random sampling, and the number of sites per scheme ranged from 20 to 500 with an interval of 30. By combining six different zonation schemes, a total of 102 regional yield potential estimation scenarios were constructed. Then, the regional yield potential was calculated using the spatial weighted average method based on the site-specific simulation results [10,31]. By using the average of the simulation results for all 538 sites as the reference value, the relative errors in the yield potential and the reference values for different simulation scenarios were calculated to quantitatively evaluate the impact of each zonation scheme on the accuracy of the simulation results. The standard deviations of the simulated yields for the sites in the basic spatial units were calculated and used to evaluate the effects of different zonation schemes on the stability of the simulation results. Finally, an appropriate zonation scheme for estimating the regional yield potential was obtained based on the site-specific simulation results (Figure 2). Additionally, unlike previous studies that used multiyear mean simulated yields for analysis [8,9], we analyzed the simulated data every year from 2000 to 2009 to validate the proposed method and verify whether the conclusions are robust to within-year variability.

Figure 2. Method roadmap.

2.3. Data Description

The daily 538 weather site data needed for WheatGrow inputs were obtained from the Meteorological Data Sharing Network of the China Meteorological Administration (<http://www.nmic.cn>) for the period from 2000 to 2009 (Figure 1b). Because our simulation scenario was yield potential, the weather data needed for WheatGrow model included daily maximum temperature (Tmax), daily minimum temperature (Tmin), and sunshine duration (SSD). The Pohlert method was used to convert sunshine duration into solar radiation quantities [32]. The most popular cultivars in each subregion were used in the WheatGrow model. A total of 129 typical sites with historical observations of wheat phenology (sowing, heading, and maturity dates), grain yield, and management practices from 2000 to 2009 were obtained from the National Meteorological Center Library of China (Figure 1b). The sowing date was interpolated using the Thiessen polygons method for all sites [33].

2.4. WheatGrow Model Description and Validation

The WheatGrow model includes five modules, the apical and phenological development of wheat, photosynthesis and dry matter production, dry matter partitioning and organogenesis, yield and quality formation, and soil moisture and nutrition balance modules [21–27]. The WheatGrow model can simulate wheat growth and development conditions under three growth conditions: yield potential, water limitation, and nitrogen limitation [22,30]. The WheatGrow model has been validated in simulations of winter wheat at multiple ecological observation sites throughout the main winter wheat production area of China using field experiment data from different sowing dates, plant densities, and nitrogen fertilization strategies, and it is reported that the WheatGrow model displays good agreement between the predicted and observed values and can effectively capture the spatial variations of the yield at different regional scales [4,34,35].

From 129 sites, we selected 45 typical sites (Figure 1a) that had the same statistical yield time range from 2000 to 2009 for model validation using RMSE (Root Mean Squared Error) and NRMSE (Normalized Root Mean Squared Error). The smaller the RMSE, the better the WheatGrow model performance, and if the NRMSE value is less than 20%, showing the WheatGrow model can well reproduce the statistical yield.

2.5. Design of the Zonation Schemes and the Regional Yield Potential Simulation Scenarios

This study was designed for six basic zonation schemes: the administrative zones of provinces (AZP), based on provincial administrative boundaries (Figure 3a) [30]; the agricultural comprehensive zones (ACZ), based on the natural environment, regional agricultural functions, and spatial agricultural characteristics (Figure 3b) [36]; the geomorphology zones (GZ), based on the elevation and terrain relief (Figure 3c) [37]; the ecological zones (EZ), based on the geographic environment, natural conditions, tillage system, cultivar type, production level, cultivation characteristics, and the occurrence of pests and diseases associated with wheat planting (Figure 3d) [38]; the temperature clustering zones (TCZ), based on the spatial clustering of the 10-year average temperatures in the study area (Figure 4e); and the sunshine duration clustering zones (SDCZ), based on the spatial clustering of the 10-year average

sunshine durations (Figure 3f). The Scikit-learn module was adopted for spatial clustering using the k-means clustering algorithm, and the minimum means distortion index was used to determine the optimum number of zones [39]. Moreover, to ensure the integrity of the basic spatial units so that sites belonging to the same spatial unit are spatially connected, this study merged small and isolated spatial units [10,31]. Among them, the ACZ has the greatest number (18) of basic spatial units (Table 1).

Figure 3. The six zonation schemes: the administrative zones of provinces (AZP) (a), the agricultural comprehensive zones (ACZ) (b), the geomorphology zones (GZ) (c), the ecological zones (EZ) (d), the temperature clustering zones (TCZ) (e), and the sunshine duration clustering zones (SDCZ) (f). The spatial distributions of different numbers of sites are also shown; for example, the number of sites was 110 in the top panels (a–f), and the number of sites was 200 in the bottom panels (g–l).

Figure 4. Validation of the yield data from 2000 to 2009.

Table 1. Characteristics of each zonation scheme: the administrative zones of provinces (AZP), the agricultural comprehensive zones (ACZ), the geomorphology zones (GZ), the ecological zones (EZ), the temperature clustering zones (TCZ), and the sunshine duration clustering zones (SDCZ).

Zonation Scheme	Number	Maximum Area (km ²)	Minimum Area (km ²)	Mean Area (km ²)
AZP	11	313,747.8	105,505.6	204,452.7
ACZ	18	361,497.3	30,636.7	124,943.3
GZ	12	396,268.4	32,847.2	187,415
EZ	11	622,829.2	41,791.6	204,452.7
TCZ	9	648,279.9	91,468.6	249,886.7
SDCZ	9	431,735.3	64,543.5	249,886.7

Furthermore, to reduce the impact of random sampling on the results of this study, spatial random sampling was performed 1000 times for each scheme, and the average of the 1000 simulation results for each simulation scenario was used as the final simulated regional yield potential.

2.6. Upscaling the Yield Potential of the Sites and Performing Uncertainty Analysis

2.6.1. Calculation of the Regional Yield Potential Using the Spatial Weighted Average Method

This study used the spatial weighted average method to obtain the regional yield potential $\hat{Y}$ (Equation (1)) [10,40], where m is the total number of basic spatial units in the region, i is the i th basic spatial unit, $\bar{y}_i$ is the average value of the simulated yield potential of the sites in the i th basic spatial unit, and ω_i is the weight contribution of $\bar{y}_i$ to the regional yield potential $\hat{Y}$ (Equation (2)).

$$\hat{Y} = \sum_{i=1}^m (\omega_i \bar{y}_i) \quad (1)$$

$$\omega_i = \frac{N_i}{N} \quad (2)$$

$$n_i = \frac{N_i \cdot n}{N} \quad (3)$$

In Equation (2), N_i represents the number of sites in the i th basic spatial unit, N is the total number of sites in the study area, n_i is the number of selected sites in i th basic spatial unit, and n is the number of selected sites.

2.6.2. Uncertainty of the Regional Yield Potential Estimation

This study used the average value of the simulated yield potential for all sites in the study area $\bar{Y}$ as the reference value (Equation (4)). The relative errors δ (Equation (5)) and the standard deviations σ (Equation (6)) were further calculated to quantitatively evaluate the accuracy and stability, respectively, of the different zonation schemes when obtaining the regional yield potential. Values of the relative error δ that are closer to 0 indicate a higher accuracy for a zonation scheme [41,42], while smaller values of the standard deviation σ indicate that a zonation scheme has greater stability [17,31,42].

$$\bar{Y} = \frac{\sum_{j=1}^N y_j}{N} \quad (4)$$

$$\delta = \frac{|\hat{Y} - \bar{Y}|}{\bar{Y}} \quad (5)$$

$$\sigma = \frac{\sum_{i=1}^m \sqrt{\frac{\sum_{k=1}^{N_i} (y_i^k - \bar{y}_i)^2}{N_i - 1}}}{m} \quad (6)$$

In Equation (4), y_j represents the simulation result of the yield potential at site j in the study area. In Equation (6), y_i^k represents the simulated yield potential of site k in the i th basic spatial unit.

3. Results

The WheatGrow model validation results show that the RMSE between the simulated and observed yields was approximately 1000 kg/ha for 2000 to 2009 (Figure 4). Except that the NRMSE in 2002 and 2003 was slightly higher than 20%, the NRMSE in all other years was within 20%. The WheatGrow model has a good performance.

The spatial distributions of the yield potential tended to be consistent for all sites throughout the winter wheat-producing area of China (Figure 5). The lower values were mainly distributed in the Sichuan Basin subregion (III) with relatively poor light and temperature conditions, for which the minimum and maximum 10-year average values were only 3209 kg/ha and 4612 kg/ha, respectively. This subregion was followed by the subregion comprising the mountainous and hilly areas of

southern Shaanxi Province and western Hubei Province (VI), the Yangtze-Huai Plain subregion (VII), and the riverside and lakeside subregion (IV). Higher values were found in the eastern subregions of the study area, including the Huang-Huai Plain subregion (I), the Jiaodong hilly subregion (V), the Yunnan-Guizhou Plateau subregion (II), and the subregion comprising the gully area of the Loess Plateau (X), among which the minimum and maximum 10-year average values were 6100 kg/ha and 11,700 kg/ha, respectively.

Figure 5. The simulated yield potential for each site in (a) 2000, (b) 2001, (c) 2002, (d) 2003, (e) 2004, (f) 2005, (g) 2006, (h) 2007, (i) 2008, and (j) 2009 and (k) the 10-year average of each site-specific simulated yield potential.

The number of sites had a significant impact on the regional yield potential estimation results under the six different zonation schemes. Based on the simulation results from all scenarios, when the number of sites was small, the regional yield potential was overestimated (Figure 6), the relative error was high (Figure 7), and the standard deviation was low (Figure 8). As the number of sites increased, the simulation results tended to become more accurate and stable (Figure 6).

Figure 6. Regional yield potential ($\hat{Y}$) under different zonation schemes from 2000 to 2009. The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the regional yield potential estimation under the different zonation schemes.

As the number of sites increased to 200, the differences among the simulated yield potential under the different zonation schemes gradually declined to less than 100 kg/ha; additionally, the relative errors of the simulated yield potential under the different zonation schemes decreased, and their differences gradually decreased to approximately 0.01; moreover, the differences among the standard deviations of the simulated yield potential under the different zonation schemes were stable at approximately 300 kg/ha (Figure 8a–j). Therefore, when more than 200 sites selected in the study area, the regional yield potential simulation difference between zonation schemes became smaller, and the different zonation schemes did not have a significant impact on the regional yield potential estimation.

Figure 7. Relative error (δ) under different zonation schemes from 2000 to 2009. The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the calculated relative error under the different zonation schemes.

Figure 8. Standard deviation (σ) of the site-specific yield potential under different zonation schemes from 2000 to 2009. The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the calculated standard deviation under the different zonation schemes.

However, the zonation schemes significantly affected the simulated regional yield potential when the number of sites was small (Figure 9a–j, red). When the number of sites was only 20, the simulation results from the different zonation schemes varied considerably. Because the sites' spatial distribution is not uniform when the sites are selected in proportion, the n_i (Equation (4)) may be zero, because there are few sites in the basic spatial units, and the selected sites are concentrated in the basic spatial unit with high yield potential; thus, the simulated regional yield potential of the ACZ was the largest. The simulated regional yield potential of the SDCZ was the smallest. Moreover, the relative error of the simulated regional yield potential of the SDCZ remained the lowest (Figure 10a–j), and the standard deviations of the simulated regional yield potential were small and less than 400 kg/ha under all the different zonation schemes (Figure 11a–j, red).

With an increase in the number of sites, the simulated regional yield potential under the various zonation schemes all showed varying degrees of reduction, while the differences among the simulated regional yield potential caused by different zonation schemes significantly decreased (Figure 9a–j, green). When the number of sites increased from 20 to 200, the average values of the simulated regional yield potential were similar under the different zones and converged (Figure 9a–j, green), and the relative errors of the simulated regional yield potential were less than 0.1 under all six zonation schemes (Figure 10a–j, green). The accuracy of the simulation results was relatively high under all

zonation schemes; however, the standard deviation of the simulated regional yield potential under each zonation scheme gradually increased, and the stability of the simulation weakened. At this point, the standard deviations of the simulated regional yield potential of the ACZ and SDCZ were the smallest, suggesting that these two zonation schemes are superior to the other zonation schemes (Figure 11a–j).

Figure 9. Regional yield potential ($\hat{Y}$) under different zonation schemes from 2000 to 2009 for 20 (red) and 200 (green) sites within a scheme. The zonation schemes included the agricultural comprehensive zones (ACZ), sunshine duration clustering zones (SDCZ), temperature clustering zones (TDZ), ecological zones (EZ), geomorphology zones (GZ), and administrative zones of provinces (AZP). The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the calculated relative error by using 1000 iterations of spatial random sampling under each number of sites.

Figure 10. Relative errors (δ) of simulated regional yield potential under different zonation schemes from 2000 to 2009 for 20 (red) and 200 (green) sites within a scheme. The zonation schemes included the agricultural comprehensive zones (ACZ), sunshine duration clustering zones (SDCZ), temperature clustering zones (TDZ), ecological zones (EZ), geomorphology zones (GZ), and administrative zones of provinces (AZP). The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the calculated relative error by using 1000 iterations of spatial random sampling under each number of sites.

Figure 11. Standard deviations (σ) of the simulated regional yield potential under different zonation schemes from 2000 to 2009 for 20 (red) and 200 (green) sites within a scheme. The zonation schemes included the agricultural comprehensive zones (ACZ), sunshine duration clustering zones (SDCZ), temperature clustering zones (TDZ), ecological zones (EZ), geomorphology zones (GZ), and administrative zones of provinces (AZP). The upper, middle, and lower lines of the box represent the upper quartiles, mean, and lower quartiles, respectively, of the calculated relative error by using 1000 iterations of spatial random sampling under each number of sites.

4. Discussion

The use of high-quality site data to estimate the regional yield potential through spatial zonation has great significance for identifying the variation patterns for the upper limit of crop yields, optimizing cropping systems, and improving the efficiency of agroclimatic resource utilization [8,11,43]. The environmental variables across our study areas where spatial heterogeneity and autocorrelation are ubiquitous can cause the simulated yields to show a spatial continuity, i.e., data at two nearby locations are on average more similar than data at two widely spaced locations [10]. For example, the simulated yield potential in the Sichuan Basin subregion (III) was relatively low and exhibited spatial agglomeration (Figure 5). By implementing a zonation scheme, the spatial heterogeneity of the simulated yields can be significantly reduced for each basic spatial unit, and the representativeness of the site-specific simulated yields can be improved [11,44]. To upscale the simulated yield potential of sites to the regional scale by using spatial zonation, previous studies have mostly focused on the use of a single existing zonation scheme, such as climate zonation or agricultural zonation; however, existing zonation schemes may have negative effects on the regional estimation results because of inappropriate prior information used for the zonation, and as a consequence, homogeneous simulation results were not obtained within the irregular basic spatial units [10]. Moreover, as the number of sites increased, the yield potential and relative error decreased, and the standard deviation increased (Figures 6–8), but previous studies did not investigate the effects of different numbers of sites on the regional yield potential simulation results [5,11,12]. Our results provide a reference for the design of zonation schemes and the selection of site numbers for estimating the regional yield potential in different regions.

Due to the spatial heterogeneity in environmental data, estimating the regional yield potential based on site data is affected by both the zonation scheme and the number of sites. In particular, when the number of sites is limited, if the simulated yield potential of a single site is excessively high or low, the simulated single-site potential will significantly impact the estimated regional yield potential. Hence, the use of a suitable zonation scheme for estimating the regional yield potential can yield relatively reliable results [17,45,46]. However, environmental data exhibit spatial continuity. Except for the clear boundary between ocean and land, there is no universal zonation scheme [41].

The comparison among the six zonation schemes considered in this study shows that the regional yield potential simulation is sensitive to the zonation scheme. Because the spatial heterogeneity existed in site-specific yield potential, the standard deviation between different zonation schemes varies greatly when the number of sites was 200 (Figure 11, green). Compared with ACZ, the SDCZ can guaranteed homogeneity of the site-specific yield potential even with a small number of basic spatial units, especially in 2002, 2007, and 2009 (Figure 11, green). Furthermore, even with only 20 sites, the SDCZ can keep the relative error within 0.2 from 2000 to 2009 (Figure 10). The SDCZ scheme can reduce both the number of sites needed and the amount of data required by the crop growth model under the premise of ensuring the simulation accuracy (Figure 10a–j, red). Therefore, the SDCZ is the most suitable zonation scheme. Increasing the number of sites can improve the regional yield potential estimation accuracy and reduce the variability in the estimated regional yield potential under different zonation schemes [11,46]. Therefore, as the number of sites increases to 200, the relative error falls below 0.1 (Figure 10a–j, green), and the relative errors of the different zonation schemes gradually decrease (Figure 7a–j). Although the ACZ scheme has the strongest stability and certain advantages, the effects of the differences among the zonation schemes on the simulated regional yield potential can be neglected.

In addition, the simulation results of the process-based crop growth model are jointly determined by the model structure, the input data and the process parameters. The input data are the main cause of the spatial variation in the regional simulation results from the crop growth model [47]. Therefore, through a sensitivity analysis of the crop growth model, we can identify the input parameters that have significant influences on the simulation results of the model, thereby improving the reliability of the model prediction [47,48]. By correlation testing, this study found that the simulated yield potential was negatively correlated with the temperature variables and significantly positively correlated with the average sunshine duration (Table 2). This finding is similar to those of previous studies that used the AGROC, APSIM-Wheat, and CoupModel models [32,49–51]. Because the simulated scenario of this study is the yield potential, due to gradually increasing sunshine duration, higher amounts of solar radiation were intercepted, and higher amounts of dry matter biomass were produced; however, the rising temperature can shorten the grain-filling period and may cause a reduction in yield [52]. Therefore, when a small number of sites were used, the SDCZ scheme could still obtain relatively good regional simulation results (Figure 10). This means that uncertainties could be introduced by different zonation schemes when upscaling site-specific simulation results to the regional scale. If a sufficient data set is available, then the most appropriate zonation scheme should be determined to guarantee the simulation accuracy at the regional scale regardless of which crop model is used. Accordingly, by dividing the area of interest into irregular basic spatial units with the most sensitive variables of a crop growth model, a relatively high regional estimation accuracy can be achieved with a small number of sites, thereby improving the availability of data at a low cost.

However, in addition to meteorological factors, the spatial variations in crop yield are also affected by the spatial heterogeneity of the soil factors, cultivars, and management measures [53]. Since the object of this study is the potential productivity as entirely determined by light and temperature conditions, only the influences of temperature and sunlight among the meteorological factors were considered. In future analyses of the potential rainfed productivity and the potential productivity determined by nitrogen content, the correlations of potential productivity with various other factors related to the soil, precipitation, and fertilization need to be analyzed, and different crop growth models need to be compared to conclusively determine the optimum zonation scheme [10]. Moreover, over a long period of time, climate change can cause zone instability. For example, due to the redistribution of light, heat, water, and soil resources caused by climate warming, the suitable planting boundary in China will move to the north, and the clustering results of weather data will also change [54]. The uncertainty caused by the period was not taken into account in this study.

Table 2. Analysis of the correlations between simulated yield potential and the following environmental variables using the *Spearman* correlation test of 538 sites: the maximum average temperature (AveTmax), minimum average temperature (AveTmin), average temperature (AveTEM), average sunshine duration (AveSSD) (during the entire growth period in each year), and site elevation (DEM). ** and * represent significance at the 1% and 5% levels, respectively.

Year	Meteorological and Topographic Factors				
	AveTmax	AveTmin	AveTEM	AveSSD	DEM
2000	−0.707 **	−0.699 **	−0.715 **	0.831 **	−0.189 **
2001	−0.703 **	−0.687 **	−0.705 **	0.83 **	−0.214 **
2002	−0.7 **	−0.696 **	−0.71 **	0.818 **	−0.07
2003	−0.673 **	−0.655 **	−0.674 **	0.815 **	−0.174 **
2004	−0.714 **	−0.682 **	−0.706 **	0.852 **	−0.248 **
2005	−0.761 **	−0.743 **	−0.762 **	0.871 **	−0.138 **
2006	−0.764 **	−0.754 **	−0.771 **	0.877 **	−0.144 **
2007	−0.722 **	−0.713 **	−0.728 **	0.841 **	−0.099 *
2008	−0.741 **	−0.716 **	−0.738 **	0.863 **	−0.24 **
2009	−0.703 **	−0.687 **	−0.705 **	0.819 **	−0.089 *

5. Conclusions

In this study, the Chinese winter wheat production region was used as the study area, and regional yield potential simulation scenarios were established under different zonation schemes using different numbers of sites for site-specific model simulations. Data from 2000 to 2009 were used to study the uncertainties associated with the upscaling of the site-specific yield potential using the spatial weighted average method. We used the average value of the simulated yield potential for all the sites in the study area as the reference value, and the results of an uncertainty analysis among the different zonation schemes suggested that the zonation scheme and distribution of sites affect the regional yield potential simulation. As such, a small number of sites significantly affected the simulated regional yield potential under different zonation schemes, and the zonation scheme based on sunshine duration clustering zones (SDCZ) could effectively guarantee the simulation accuracy at the regional scale. Thus, using the most influential environmental variable of crop growth models for clustering can get a better zonation scheme to upscale the site-specific simulation results. However, a large number of sites resulted in little difference among the regional yield potential simulation results under the different zonation schemes.

Author Contributions: Conceptualization, H.X. and X.Z.; funding acquisition, W.C. and X.Z.; Methodology, Y.Z. and X.Z.; Software, H.X. and W.Z.; Visualization, W.Z. and F.H.; Writing—original draft, H.X. and F.H.; Writing—review and editing, Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by grants from the National Key R&D Program of China (No. 2016YFD0300607) and National Natural Science Foundation of China (No. 41961124008).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Van Wart, J.; Kersebaum, K.C.; Peng, S.; Milner, M.; Cassman, K.G. Estimating crop yield potential at regional to national scales. *Field Crops Res.* **2013**, *143*, 34–43. [\[CrossRef\]](#)
2. Cedrez, C.B.; Hijmans, R.J. Methods for spatial prediction of crop yield potential. *Agron. J.* **2018**, *110*, 2322–2330. [\[CrossRef\]](#)
3. Monjardino, M.; Hochman, Z.; Horan, H. Yield potential determines Australian wheat growers' capacity to close yield gaps while mitigating economic risk. *Agron. Sustain. Dev.* **2019**, *39*, 49. [\[CrossRef\]](#)
4. Zhang, X.; Xu, H.; Jiang, L.; Zhao, J.; Zuo, W.; Qiu, X.; Tian, Y.; Cao, W.; Zhu, Y. Selection of appropriate spatial resolution for the meteorological data for regional winter wheat potential productivity simulation in China based on wheatgrow model. *Agriculture* **2018**, *8*, 198. [\[CrossRef\]](#)

5. Morell, F.J.; Yang, H.S.; Cassman, K.G.; Van Wart, J.; Elmore, R.W.; Licht, M.; Coulter, J.A.; Ciampitti, I.A.; Pittelkow, C.M.; Brouder, S.M.; et al. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? *Field Crops Res.* **2016**, *192*, 1–12. [\[CrossRef\]](#)
6. Tveito, O.E.; Bjørndal, I.; Skjelvåg, A.O.; Aune, B. A GIS-based agro-ecological decision system based on gridded climatology. *Meteorol. Appl.* **2010**, *12*, 57–68. [\[CrossRef\]](#)
7. Elliott, J.W. The Global Gridded Crop Model Intercomparison (GGCMI). *Geosci. Model Dev. Discuss.* **2015**, *7*, 4383–4427. [\[CrossRef\]](#)
8. Zhao, G.; Siebert, S.; Enders, A.; Rezaei, E.E.; Yan, C.; Ewert, F. Demand for multi-scale weather data for regional crop modeling. *Agric. For. Meteorol.* **2015**, *200*, 156–171. [\[CrossRef\]](#)
9. Van Bussel, L.; Ewert, F.; Zhao, G.; Hoffmann, H.; Enders, A.; Wallach, D.; Asseng, S.; Baigorria, G.A.; Basso, B.; Biernath, C.; et al. Spatial sampling of weather data for regional crop yield simulations. *Agric. For. Meteorol.* **2016**, *220*, 101–115. [\[CrossRef\]](#)
10. Zhao, G.; Hoffmann, H.; Yeluripati, J.; Xenia, S.; Nendel, C.; Coucheney, E.; Kuhnert, M.; Tao, F.; Constantin, J.; Raynal, H.; et al. Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. *Environ. Model. Softw.* **2016**, *80*, 100–112. [\[CrossRef\]](#)
11. Van Bussel, L.; Grassini, P.; Van Wart, J.; Wolf, J.; Claessens, L.; Yang, H.; Boogaard, H.; De Groot, H.; Saito, K.; Cassman, K.G.; et al. From field to atlas: Upscaling of location-specific yield gap estimates. *Field Crops Res.* **2015**, *177*, 98–108. [\[CrossRef\]](#)
12. Van Wart, J.; Van Bussel, L.; Wolf, J.; Licker, R.; Grassini, P.; Nelson, A.; Boogaard, H.; Gerber, J.S.; Mueller, N.D.; Claessens, L.; et al. Use of agro-climatic zones to upscale simulated crop yield potential. *Field Crops Res.* **2013**, *143*, 44–55. [\[CrossRef\]](#)
13. Rötter, R.; Palosuo, T.; Kersebaum, K.-C.; Valadez, C.E.A.; Bindi, M.; Ewert, F.; Ferrise, R.; Hlavinka, P.; Moriondo, M.; Nendel, C.; et al. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. *Field Crops Res.* **2012**, *133*, 23–36. [\[CrossRef\]](#)
14. Van Ittersum, M.K.; Cassman, K.G.; Grassini, P.; Wolf, J.; Titttonell, P.; Hochman, Z. Yield gap analysis with local to global relevance—A review. *Field Crops Res.* **2013**, *143*, 4–17. [\[CrossRef\]](#)
15. Dark, S.J.; Bram, D. The Modifiable Areal Unit Problem (MAUP) in Physical Geography. *Prog. Phys. Geogr.* **2007**, *31*, 471–479. [\[CrossRef\]](#)
16. Ruddell, D.; Wentz, E.A. Multi-tasking: Scale in geography. *Geogr. Compass* **2009**, *3*, 681–697. [\[CrossRef\]](#)
17. Simbahan, G.C.; Dobermann, A. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma* **2006**, *133*, 345–362. [\[CrossRef\]](#)
18. Sun, P.; Congalton, R.G.; Pan, Y. Using a simulation analysis to evaluate the impact of crop mapping error on crop area estimation from stratified sampling. *Int. J. Digit. Earth* **2019**, *12*, 1046–1066. [\[CrossRef\]](#)
19. Zheng, J.; Yin, Y.; Li, B. A new scheme for climate regionalization in China. *Acta Geogr. Sin.* **2010**, *65*, 3–12.
20. Zhao, G. Study on Chinese wheat planting regionalization (II). *J. Triticeae Crops* **2010**, *30*, 1140–1147.
21. Liu, B.; Liu, L.; Asseng, S.; Zou, X.; Li, J.; Cao, W.; Zhu, Y. Modelling the effects of heat stress on post-heading durations in wheat: A comparison of temperature response routines. *Agric. For. Meteorol.* **2016**, *222*, 45–58. [\[CrossRef\]](#)
22. Yan, M.; Cao, W.; Luo, W.; Jiang, H. A mechanistic model of phasic and phenological development of wheat. I. Assumption and description of the model. *Chin. J. Appl. Ecol.* **2000**, *11*, 355.
23. Cao, W.; Moss, D.N. Modelling phasic development in wheat: A conceptual integration of physiological components. *J. Agric. Sci.* **1997**, *129*, 163–172. [\[CrossRef\]](#)
24. Liu, T.; Cao, W.; Luo, W.; Wang, S.; Yin, J. A simulation model of photosynthetic production and dry matter accumulation in wheat. *Acta Triticeae Crops* **2001**, *21*, 26–30.
25. Liu, T.; Cao, W.; Luo, W.; Wang, S.; Guo, W.; Zou, W.; Zhou, Q. Quantitative simulation on dry matter partitioning dynamic in wheat organs. *J. Triticeae Crops* **2001**, *21*, 25–31.
26. Pan, J.; Zhu, Y.; Jiang, D.; Dai, T.; Li, Y.; Cao, W. Modeling plant nitrogen uptake and grain nitrogen accumulation in wheat. *Field Crops Res.* **2006**, *97*, 322–336. [\[CrossRef\]](#)
27. Pan, J.; Zhu, Y.; Cao, W. Modeling plant carbon flow and grain starch accumulation in wheat. *Field Crops Res.* **2007**, *101*, 276–284. [\[CrossRef\]](#)
28. Hu, J.; Cao, W.; Jiang, D.; Luo, W. Quantification of water stress factor for crop growth simulation I. Effects of drought and waterlogging stress on photosynthesis, transpiration and dry matter partitioning in winter wheat. *Acta Agron. Sin.* **2004**, *30*, 315–320.

29. Zhuang, H.Y.; Cao, W.X.; Jiang, S.X.; Wang, Z.G. Simulation on nitrogen uptake and partitioning in crops. *Syst. Sci. Compr. Stud. Agric.* **2004**, *20*, 5–8.
30. Lv, Z.; Liu, X.; Tang, L.; Liu, L.; Cao, W.; Zhu, Y. Estimation of ecotype-specific cultivar parameters in a wheat phenology model and uncertainty analysis. *Agric. For. Meteorol.* **2016**, *221*, 219–229. [[CrossRef](#)]
31. Brus, D.J.; Spätjens, L.E.E.M.; De Gruijter, J.J. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. *Geoderma* **1999**, *89*, 129–148. [[CrossRef](#)]
32. Pohlert, T. Use of empirical global radiation models for maize growth simulation. *Agric. For. Meteorol.* **2004**, *126*, 47–58. [[CrossRef](#)]
33. Brassel, K.E.; Reif, D. A procedure to generate Thiessen polygons. *Geogr. Anal.* **1979**, *11*, 289–303. [[CrossRef](#)]
34. Lv, Z.; Liu, X.; Cao, W.; Zhu, Y. Climate change impacts on regional winter wheat production in main wheat production regions of China. *Agric. For. Meteorol.* **2013**, *171*, 234–248. [[CrossRef](#)]
35. Huang, Y.; Zhu, Y.; Wang, H.; Yao, X.; Cao, W.; Hannaway, D.B.; Tian, Y. Predicting winter wheat growth based on integrating remote sensing and crop growth modeling techniques. *Acta Ecol. Sin.* **2011**, *31*, 1073–1084.
36. Liu, Y.; Zhang, Z.; Wang, J. Regional differentiation and comprehensive regionalization scheme of modern agriculture in China. *Acta Geogr. Sin.* **2018**, *2*, 203–218.
37. Li, B.Y.; Pan, B.; Cheng, W.; Han, J.; Qi, D.; Zhu, C. Research on geomorphological regionalization of China. *Acta Geogr. Sin.* **2013**, *68*, 291–306.
38. Shanbao, J. *Chinese Wheat*; China Agricultural Press: Beijing, China, 1996.
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2013**, *12*, 2825–2830.
40. Boschetti, L.; Stehman, S.V.; Roy, D.P. A stratified random sampling design in space and time for regional to global scale burned area product validation. *Remote Sens. Environ.* **2016**, *186*, 465–478. [[CrossRef](#)]
41. De Gruijter, J.; Brus, D.J.; Bierkens, M.F.; Kotters, M. *Sampling for Natural Resource Monitoring*; Springer: Berlin/Heidelberg, Germany, 2006.
42. Wang, J.; Haining, R.; Cao, Z. Sample surveying to estimate the mean of a heterogeneous surface: Reducing the error variance through zoning. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 523–543. [[CrossRef](#)]
43. Baron, C.; Sultan, B.; Balme, M.; Sarr, B.; Traore, S.; Lebel, T.; Janicot, S.; Dingkuhn, M. From GCM grid cell to agricultural plot: Scale issues affecting modelling of climate impact. *Philos. Trans. R. Soc. B Biol. Sci.* **2005**, *360*, 2095–2108. [[CrossRef](#)] [[PubMed](#)]
44. Alexandridis, T.K.; Katagis, T.; Gitis, I.Z.; Silleos, N.G.; Eskridge, K.M.; Gritzas, G. Investigation of aggregation effects in vegetation condition monitoring at a national scale. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 507–521. [[CrossRef](#)]
45. Van Groenigen, J.W. The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma* **2000**, *97*, 223–236. [[CrossRef](#)]
46. Fortin, M.J.; Drapeau, P.; Legendre, P. Spatial autocorrelation and sampling design in plant ecology. *Vegetation* **1989**, *83*, 209–222. [[CrossRef](#)]
47. Wisz, M.S.; Hijmans, R.J.; Li, J.; Peterson, A.T.; Graham, C.H.; Guisan, A.; NCEAS. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **2008**, *14*, 763–773. [[CrossRef](#)]
48. Varella, H.; Guérif, M.; Buis, S. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environ. Model. Softw.* **2010**, *25*, 310–319. [[CrossRef](#)]
49. Zhao, G.; Bryan, B.A.; Song, X. Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters. *Ecol. Model.* **2014**, *279*, 1–11. [[CrossRef](#)]
50. Sadras, V.O.; Monzon, J.P. Modelled wheat phenology captures rising temperature trends: Shortened time to flowering and maturity in Australia and Argentina. *Field Crops Res.* **2006**, *99*, 136–146. [[CrossRef](#)]
51. Asseng, S.; Jamieson, P.D.; Kimball, B.; Pinter, P.; Sayre, K.; Bowden, J.W.; Howden, S.M. Simulated wheat growth affected by rising temperature, increased water deficit and elevated atmospheric CO₂. *Field Crops Res.* **2004**, *85*, 85–102. [[CrossRef](#)]
52. Liu, L.; Wallach, D.; Li, J.; Liu, B.; Zhang, L.; Tang, L.; Zhang, Y.; Qiu, X.; Cao, W.; Zhu, Y. Uncertainty in wheat phenology simulation induced by cultivar parameterization under climate warming. *Eur. J. Agron.* **2018**, *94*, 46–53. [[CrossRef](#)]

53. Grassini, P.; Van Bussel, L.; Van Wart, J.; Wolf, J.; Claessens, L.; Yang, H.; Boogaard, H.; De Groot, H.; Van Ittersum, M.; Cassman, K.G. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Res.* **2015**, *177*, 49–63. [[CrossRef](#)]
54. Gao, J.; Liu, Y. Climate warming and land use change in Heilongjiang Province, Northeast China. *Appl. Geogr.* **2011**, *31*, 482. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).