Simulation of Atmosphere Temperature of the 20th Century in 50 US State Capitals with Random Walk

Shaomin Yan and Guang Wu*
National Engineering Research Center for Non-Food Biorefinery, State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Biomass Engineering Technology Research Center, Guangxi Key Laboratory of Bio-Refinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, 530007, China
*Corresponding author’s e-mail: hongguanglishibahao@gxas.cn

Abstract. Although humans have been recording the temperature for about 200 years, the temperature in the 20th century is more important because the trend of increase of temperature was observed. The increasing temperature has been the objective of many studies across natural and social sciences. The similarity between the temperature and the curve generated by the random walk model was noticed in 1991, since then, however, very few studies have been conducted along this research line due to the limitation of computational power. In this study, we use the random walk model to simulate the atmosphere temperature of the 20th century in 50 US state capitals in terms of converting the temperature into its simplest form with random walk simulation as well as the recorded temperature. The results show that the random walk model can simulate the simplest form of temperature and recorded temperature satisfactorily.

1. Introduction
Global warming and its devastating impacts have been the objective of many studies across natural and social sciences. Global warming even has impact on the pattern of occurrence of diseases [1], where the yearly fluctuation in temperature was noticed although the upward trend is clear. This fluctuation in temperature can be due to various trivial factors [2]. Nevertheless, temperature fluctuations should not be expected to limit to the fluctuations between years.

In past, the similarity between yearly temperature and the curve produced by random walk rather than fluctuation was noticed [3]. Due to the limitation of computational power as well as other unknown factors, few studies have been conducted along this line, i.e. simulation of temperature using a random walk model. Although the use of random walk to model the global temperature anomaly is very representative [4], many studies are needed to verify this analysis in individual cities.

Of individual cities, the cities in the US are important not only because the industrial sectors in the US are the most powerful and comprehensive but also because the US contributes greatly in global CO$_2$ emission as well as the US government’s attitude to the global warming including Paris Agreement changes frequently.

As a matter of fact, the temperature in the 20th is most important in wake of global warming because the global temperature entered a hiatus since 1998 [5]. Technically, it is only the global temperature is represented as temperature anomaly [6], whereas the temperature in individual cities is still reported using recorded temperature, which is familiar to people. This study aims to use the random walk model to simulate the atmosphere temperature of the 20th century in 50 US state capitals.
2. Materials and Methods

2.1. Assignment of Temperatures to 50 State Capitals
Fifty US state capitals are calibrated with their latitudes and longitudes. The temperatures in 0.5° by 0.5° latitude and longitude grid-box are obtained from the Oak Ridge National Laboratory [6].

2.2. Simplified Temperature
As the temperature is recorded over time, it can relate to the one-dimensional random walk [7], which starts at zero, and then can move up or down (±1) at each step, which composes a sequence by adding the value in previous step whereas whether or not it moves up or down is dependent upon the probability generated by a computer algorithm. In plain words, the one-dimensional random walk can be considered as a sequential result of tossing a fair coin, by which the head being up is considered as 1 and the tail being up is considered as –1, and then these values are added together to form a sequence over time. For simplicity, a recorded temperature can be defined either higher than that in the previous recorded time or lower than that in the previous recorded time, i.e. when the temperature at certain recording time is higher than its previous one, it is assigned with 1, otherwise with –1 (column 3, Table 1), and then these values are added as a random walk (column 4, Table 1).

Table 1. Construction of a simplified temperature walk and an integral random walk for temperature from 1901 to 1998 in Atlanta.

| Year | Average Temperature | Simplified Temperature | Temperature Walk | Generated Random Number | Random Step | Random Walk |
|------|---------------------|------------------------|------------------|-------------------------|-------------|-------------|
| 1901 | 15.0833             | 0                      | –0.8436          | 0                       |             |             |
| 1902 | 16.2833             | 1                      | 0.1314           | 1                       | 1           | 1           |
| 1903 | 15.6833             | –1                     | 0.2942           | 1                       | 2           | 2           |
| 1904 | 15.7750             | 1                      | 0.2737           | –1                      | 1           | 1           |
| 1905 | 15.9333             | 1                      | 0.4834           | 1                       | 2           | 2           |
| 1906 | 16.2917             | 1                      | 0.7477           | 1                       | 3           | 3           |
| 1907 | 16.8083             | 1                      | –0.2334          | –1                      | 2           | 2           |
| 1908 | 16.8583             | 1                      | 0.8341           | 1                       | 3           | 3           |
| 1909 | 16.5833             | –1                     | –0.1805          | –1                      | 2           | 2           |
| ...  | ...                 | ...                    | ...              | ...                     | ...         | ...         |
| 1993 | 17.4083             | 1                      | 0.1610           | 1                       | 12          | 12          |
| 1994 | 17.6667             | 1                      | –0.4318          | –1                      | 11          | 11          |
| 1995 | 17.1750             | –1                     | –0.6418          | –1                      | 10          | 10          |
| 1996 | 16.8583             | –1                     | –0.8581          | –1                      | 9           | 9           |
| 1997 | 16.1667             | –1                     | –0.0561          | 1                       | 10          | 10          |
| 1998 | 17.3833             | 1                      | –0.3421          | –1                      | 9           | 9           |

2.3. Integral Random Walk
The simple way to generate a random walk is to use a certain algorithm to generate random numbers, for example, the Monte Carlo algorithm. Usually, the generated random numbers are centred around zero, so a random number can also be defined either 1 when it is larger than its previous random number or –1 when it is smaller than its previous one (column 6, Table 1). These values are added together and become an integral random walk (column 7, Table 1).

2.4. Searching for Seed in Monte Carlo Algorithm
To generate random numbers, whose addition is most similar to the addition of simplified or real recorded temperatures, it is necessary to find a seed in Monte Carlo algorithm to generate such a series of random numbers. To the best of our knowledge, there is no simple way to find such a seed, and thus, the brutal force method [8] is employed to try each seed to generate a series of random numbers to generate a random walk and compare with either simplified or recorded temperature. The seed, whose
random walk has a minimal difference with recorded temperature, is the optimal seed. The SigmaPlot [9] is used to generate random numbers with different seeds through searching of one million of seeds.

2.5. Decimal Random Walk
Because the random walk comes from tossing of double-sided coin, accordingly we broaden this concept into tossing of dice, which can be not only six-sided but as many as we need. Thus we can use this imaginary multi-sided dice to fit the real temperature.

2.6. Comparison
The sum of least squares in difference between simplified temperature and integral random walk, and between recorded temperature and decimal random temperature is used to evaluate the simulation.

3. Results and Discussions
Table 1 details how to construct a simplified temperature walk and an integral random walk. The starting point is 1901 (cell 2, column 1), and the average temperature for Atlanta was 15.0833 C (cell 2, column 2). As 1901 is the starting point, so there is no temperature step (cell 2, column 3), and the simplified temperature walk starts from zero (cell 2, column 4). A particular seed, which is suitable for this dataset, is 0.56738, produces the first random number of –0.8436 (cell 2, column 5). Similarly, the starting point has no random step (cell 2, column 6) and the starting point for random walk is zero (cell 2, column 7). For 1902 (cell 3, column 1), the average temperature was 16.2833 (cell 3, column 2), which is higher than the average temperature for 1901 (cell 2, column 2), so the temperature step is positive (cell 3, column 3), and the simplified temperature walk is therefore 1 (0 + 1, cell 3, column 4). The second generated random number is 0.1314 (cell 3, column 5), which is larger than –0.8436 (cell 2, column 5), so the random step is positive (cell 3, column 6), and the integral random walk is therefore 1 (0 + 1, cell 3, column 7). In this manner, the temperature and random walks can be constructed for following years and the rest US state capitals.

| Year | Average Temperature | Difference | Generated Random Number | Random Walk |
|------|---------------------|------------|-------------------------|-------------|
| 1901 | 15.0833             |            |                         | 15.0833     |
| 1902 | 16.2833             | 1.2        | -0.34334                | 14.73999    |
| 1903 | 15.6833             | -0.6       | -0.51194                | 14.22805    |
| 1904 | 15.7750             | 0.0917     | 0.73951                 | 15.46755    |
| 1905 | 15.9333             | 0.1583     | 0.57299                 | 15.54054    |
| 1906 | 16.2917             | 0.3584     | 0.52793                 | 16.08474    |
| 1907 | 16.8083             | 0.5166     | 0.34801                 | 16.41648    |
| 1908 | 16.8583             | 0.05       | 0.56151                 | 16.97799    |
| 1909 | 16.5833             | -0.275     | -0.10783                | 16.87016    |
| …    | …                   | …          | …                       | …           |
| 1993 | 17.4083             | 0.2584     | 0.40907                 | 17.88184    |
| 1994 | 17.6667             | -0.4917    | -0.61418                | 17.26767    |
| 1995 | 17.1750             | -0.3167    | -0.30631                | 16.96136    |
| 1996 | 16.8583             | -0.6916    | -0.40652                | 16.55484    |
| 1997 | 16.1667             | 1.2166     | -0.65107                | 15.90377    |
| 1998 | 17.3833             | 0.2584     | -0.25473                | 15.64904    |

Table 2. Construction of decimal random walk for simulation of temperature in Atlanta.

Table 2 details how to construct a decimal random walk for simulation of temperature. The starting year is 1901 (cell 2, column 1), and its average temperature is 15.0833 C (cell 2, column 2). For this starting point, there is no difference (cell 3, column 3), no random number is generated (cell 2, column 4) and the random walk begins from 15.0833 (cell 3, column 5). For 1902 (cell 3, column 1), the average temperature is 16.2833 (cell 3, column 2), the difference in average temperature between 1901 and 1902 is 1.2 (16.2833 – 15.0833, cell 3, column 3). A particular seed, which is suitable for this dataset, is 4.38130, which generates –0.34334 (cell 3, column 4) and this step of random walk is...
14.73999 \ (15.0833 + \ (-0.34334), \ cell \ 3, \ column \ 5). \ In \ this \ way, \ the \ temperature \ difference \ and \ random \ walk \ can \ be \ constructed \ for \ following \ years \ as \ well \ as \ the \ rest \ US \ state \ capitals.

Figure 1 shows the results based on the approaches in Tables 1 and 2 for simplified temperature walk (left panel) and the decimal random walk (right panel) in seven US state capitals at different locations. On the left panel, the simplified temperature walk can continuously take several steps either
up or down, which would be the trend for recorded temperature changes as the temperature increased over years. Although the simplified temperature walk is the conversation from recorded temperature, it can indeed be simulated by an integral random walk no matter of whether the simplified temperature walk goes up or down. The simulations were done for all 50 US state capitals (columns 4 and 5 in Table 2). It would suggest that the temperature change could be attributed to a random mechanism. On the right panel, the decimal random walk is used to simulate the recorded temperature. As can be seen, the decimal random walk goes well along the general trend of recorded temperature. In particular, the random walk produces a fluctuated temperature over time, which is very different from the simulation using deterministic models. Table 3 details the simulation results for 50 US state capitals from 1901 to 1998. In this table, the most important is the sum of least squares, which can compare with the results obtained from other climate models in simulation of temperatures in these cities.

Table 3. Seeds and sum of least squares (SLS) in difference between simplified temperature and integral random walk, and between recorded temperature and decimal random walk temperature in 50 US state capitals from 1901 to 1998.

| State       | Capital     | Integral Random Walk | Decimal Random Walk |
|-------------|-------------|----------------------|---------------------|
| Alabama     | Montgomery  | 0.01753, 96          | 0.92965, 45.32      |
| Alaska      | Juneau      | 1.68924, 152         | 0.96195, 92.05      |
| Arizona     | Phoenix     | 0.81221, 138         | 1.02646, 44.23      |
| Arkansas    | Little Rock | 3.40995, 156         | 0.47055, 46.02      |
| California  | Sacramento  | 0.07284, 160         | 1.20460, 41.65      |
| Colorado    | Denver      | 0.77993, 152         | 5.17076, 70.67      |
| Connecticut | Hartford    | 6.81387, 112         | 0.29972, 59.60      |
| Delaware    | Dover       | 0.59641, 128         | 3.33953, 53.41      |
| Florida     | Tallahassee | 0.46018, 140         | 2.10318, 47.50      |
| Georgia     | Atlanta     | 2.51881, 156         | 4.37599, 50.40      |
| Hawaii      | Honolulu    | 3.92830, 171         | 5.42073, 14.69      |
| Idaho       | Boise       | 3.25632, 140         | 1.74445, 80.19      |
| Illinois    | Springfield | 1.80909, 115         | 1.03233, 81.28      |
| Indiana     | Indianapolis| 0.67101, 112         | 6.87268, 83.40      |
| Iowa        | Des Moines  | 0.24257, 144         | 1.03233, 103.75     |
| Kansas      | Topeka      | 1.76253, 159         | 1.03233, 71.69      |
| Kentucky    | Frankfort   | 1.00920, 113         | 1.53686, 69.47      |
| Louisiana   | Baton Rouge | 6.38266, 147         | 0.04349, 44.29      |
| Maine       | Augusta     | 0.27741, 107         | 4.26952, 60.20      |
| Maryland    | Annapolis   | 1.28499, 128         | 5.14355, 47.79      |
| Massachusetts| Boston     | 0.29101, 116         | 0.29972, 51.06      |
| Michigan    | Lansing     | 0.01965, 120         | 7.50539, 90.22      |
| Minnesota   | Saint Paul  | 0.35895, 152         | 0.36478, 109.72     |
| Mississippi | Jackson     | 2.08091, 140         | 0.32423, 53.18      |
| Missouri    | Jefferson City | 2.03635, 142   | 2.73722, 73.82      |
| Montana     | Helena      | 2.54924, 132         | 2.21326, 111.39     |
| Nebraska    | Lincoln     | 0.36806, 152         | 2.73722, 99.21      |
| Nevada      | Carson City | 1.57222, 152         | 0.84321, 44.24      |
| New Hampshire| Concord    | 4.17823, 120         | 9.93686, 61.93      |
| New Jersey  | Trenton     | 0.03355, 152         | 1.07158, 56.42      |
| New Mexico  | Santa Fe    | 1.61915, 163         | 0.78631, 34.95      |
| New York    | Albany      | 4.35626, 133         | 0.53868, 73.97      |
| North Carolina | Raleigh  | 2.58263, 132         | 0.67411, 48.42      |
| North Dakota| Bismarck    | 1.29533, 140         | 1.70767, 131.48     |
| Ohio        | Columbus    | 0.30987, 150         | 0.32423, 81.20      |
| Oklahoma    | Oklahoma City | 8.85332, 152   | 1.27221, 62.06      |
| Oregon      | Salem       | 2.40359, 144         | 2.25410, 44.95      |
| Pennsylvania| Harrisburg  | 0.08013, 129         | 7.29458, 58.79      |
| Rhode Island| Providence  | 3.27201, 128         | 2.20070, 49.12      |
| South Carolina | Columbia | 0.39266, 136         | 0.04349, 38.15      |
| South Dakota| Pierre      | 2.10012, 148         | 0.70169, 105.90     |
| Tennessee   | Nashville   | 2.65617, 139         | 0.32423, 60.41      |
Texas Austin 3.47021 120 1.03292 50.24
Utah Salt Lake City 0.85287 134 0.28539 70.22
Vermont Montpelier 0.14196 120 0.91978 70.56
Virginia Richmond 0.38121 124 0.32423 47.96
Washington Olympia 1.15451 124 0.03475 33.93
West Virginia Charleston 2.40603 146 0.04349 63.54
Wisconsin Madison 7.52922 120 0.36478 93.67
Wyoming Cheyenne 0.23926 130 1.17243 65.81

Theoretically, a perfect simulation is an event with small probability of occurrence. For example, the chance of complete simulation of simplified temperature is $0.5^{-98}$, which could not be easy to reach. Furthermore, the chance for simulation of recorded temperature would be far smaller than $0.5^{-98}$ because the decimal random walk has more options than integral random walk. From the view of parsimony in model parameters, the random walk model is very simple, which dramatically and sharply reduces the uncertainty in comparison with the current climate models [10].

4. Conclusion
This study demonstrates that the random walk model can simulate the atmosphere temperatures in 50 US state capitals. In plain words, a good seed using Monte Carlo mechanism can generate a series of random numbers, whose addition is similar to temperature change in these capitals for the 20th century.

Acknowledgments
This study was supported by Key Project of Guangxi Scientific Research and Technology Development Plan (AB17195034).

References
[1] Yan, S., Wu, G. (2019) Impact of temperature on influenza status based on global surveillance and temperature data. Biomed. Environm. Sci., 32: 554-557.
[2] Williams, P.D. (2005) Modelling climate change: the role of unresolved processes. Phil Trans R Soc A, 363: 2931-2946.
[3] Gordon, A.H. (1994) Global warming as a manifestation of a random walk. J. Climate, 4: 589-597.
[4] Yan, S, Wu, G. (2010) Modeling of global temperature change from 1850 to 2009 using random walk. Guangxi Sci., 17: 148–150.
[5] Knutson, T.R, Zhang, R, Horowitz, L.W. (2016) Prospects for a prolonged slowdown in global warming in the early 21st century. Nature Commun., 7: 13676.
[6] New, M, Hulme, M., Jones, P. (2000) Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. J. Climate, 13: 2217-2238.
[7] Wikimedia Foundation, Inc. 2020. Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Random_walk.
[8] Levitin, A. (2003) Introduction to the Design and Analysis of Algorithms. 1st ed., New Jersey: Pearson Education.
[9] SPSS Inc. 1986-2001. SigmaPlot 2002 for Windows Version 8.02.
[10] Cawley, G.C., Janacek, G.J, Haylock, M.R, Dorling, S.R. (2007) Predictive uncertainty in environmental modelling. Neural Network, 20: 537-549.