

Variable Daily Air Temperature Model for Analysis and Design

G. Danko^{1,2}, C. Lu¹

¹Mackay School of Earth Science and Engineering, University of Nevada, Reno, Reno, NV, USA

²Research Institute of Applied Earth Sciences, University of Miskolc, Miskolc, Hungary

Email: danko@unr.edu

How to cite this paper: Danko, G. and Lu, C. (2018) Variable Daily Air Temperature Model for Analysis and Design. *Applied Mathematics*, 9, 1015-1038.
<https://doi.org/10.4236/am.2018.98069>

Received: July 3, 2018

Accepted: August 27, 2018

Published: August 30, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

An analytical model, $T_A(t)$, for the observed outside air temperature change, $T_a(t)$, with time is developed using two components: one for the variation caused by the Earth's movement, plus any other quasi-stationary thermodynamic effects due to industrialization; and one for the random variation caused by stochastic and/or chaotic, local environmental changes. The first component, $T_R(t)$, describes a regular trend, expressed by periodic functions of time and constants unchanged with time. The second component, T_S , is a random, stochastic variation. For the observed outside air temperature, the analytical model of $T_A(t) = T_R(t) + T_S$ is such as to give a statistically best approximation for the observed time period with $\|T_a(t) - T_A(t)\| = \min$. Several versions for the $T_R(t)$ functions are defined and tested in the study for an example location for 20 years. The best model for $T_R(t)$ is found as a linear function with time plus a variable-coefficient Fourier series with linearly changing amplitude with time. It is found that the final analytical temperature, $T_A(t)$, can be used not only to represent the historical daily mean temperature but also to predict the future daily mean temperature at the given location. The upper and lower boundaries give safety limits for the temperature prediction. The stochastic component identified in the model is stable and stationary. The method of model identification for $T_A(t)$ can be used for determining input temperature functions for supporting engineering design; or for an unbiased scientific inquiry of temperature change with time in climate studies.

Keywords

Air Temperature Model, Deterministic and Stochastic Changes, Temperature Trends

1. Introduction

Climate is one of the main elements of the natural environment. Temperature has a direct impact on atmospheric stability, evaporation, precipitation, and many other conditions of life [1]. Temperature change affects living conditions, agriculture, industry, tourism and engineering designs. Atmospheric temperature is needed for planning and forecasting both in the average and the variation components of temperature in order to prevent hazards to life as well as finances [2].

Based on the long-term trends study of maximum, minimum and mean annual air temperature, e.g., in the northwest Himalayan region during the twentieth century, increasing trends are seen both in the mean and the diurnal range of temperature. The daily maximum temperatures have increased more rapidly than the decrease in the low temperatures in the last century resulting in a risen mean temperature of about 1.6°C [3]. An extrapolation for the next century would give an increase about 3°C to 5°C in the global mean annual temperature [4]. A more exact forward prediction approach may give different result. It is difficult, however, to analyze slowly-changing trends in the outside temperature as they are buried within large-amplitude, harmonic, cyclic variations as well as stochastic and/or chaotic changes.

The observed, outside air temperature, $T_a(t, \tau)$, includes both seasonal components represented by t (days of the year) and τ (the hours in day t). When tabulated, $T_a(t, \tau)$ is a matrix with t rows (the total number of days) and τ columns (24, the number of hours in a day). The outside air temperature has a regular, seasonal variation for the daily average temperature during every year, (more exactly, every four years as a true time period) and an hourly variation during any given day for the hourly mean temperature. These two components are regular and periodic in nature caused by the earth movement in the solar system. Other regular changes superimposed to that of the known movement of the Earth may also be present, such as caused by the heat balance of the globe by industrialization.

The goal of the paper is to separate the observed outside air temperature into variation caused by the Earth's movement, plus any other quasi-stationary thermodynamic effects, and random variation caused by stochastic and/or chaotic, local environmental changes. It's necessary to separate the hourly temperature variations from that of the seasonal first. For describing the seasonal temperature variation, the daily average temperature $T_a(t)$ must be defined. The $T_a(t)$ may be defined as the integral mean value:

$$T_a(t) = \frac{1}{T} \int_0^T T_a(t, \tau) d\tau \quad (1)$$

In (1) t denotes days, τ is the hours in a day, $\tau \in [0, T]$ and $T = 24$. If hourly average temperature is available, (1) gives the accurate value for the daily average temperature meaning the thermodynamic energy of the air. If the daily mean temperature is obtained from weather service, the data may be given

pre-calculated from the Standard Model that provides the average of the daily maximum and minimum air temperatures, $T_a(t) \cong [T_{\max}(t) + T_{\min}(t)]/2$. The Standard Model, reviewed by Bilbao *et al.* [5], may not be accurate if the temperature change is not symmetrical between day and night [6]. Nevertheless, the convenience in using the Standard Model instead of using the hourly temperature data for every day and using the expression (1) may overwhelm the concerns in the accuracy of the mean, daily temperature.

Two components of the air temperature are distinguished in the present paper for modeling daily mean air temperature variation, $T_a(t)$, with time. The first component, $T_R(t)$, describes a regular trend, expressed by functions of time and constants unchanged with time over which the model is defined. The regular trend is defined as stationary for a long period of time, characteristic to a given physical location governed by deterministic causes such as the Earth's movement in the solar system. The second component, T_S , is a random, stochastic variation around the regular trend. The T_S component is caused by the stochastic and/or chaotic process in the atmosphere, defined as difference between the observed outside air temperature, $T_a(t)$, and the temperature from the stationary trend model as $T_S = T_a - T_R$. The daily mean value of the outside temperature at any given day is the sum of the regular trend component, T_R , and a stochastic variation part, T_S :

$$T_a(t) = T_R(t) + T_S \quad (2)$$

Note that the stochastic component is stationary and irrespective of the seasonal variation, a simplification for model formulation. However, the stochastic temperature variation in some part of the year may be more disturbed than in another, raising the possibility for improvement of the assumption used in the current work, a task left for the interested reader.

The analytic function for $T_R(t)$ must be the best fit to the measured outside temperature data for a given location. The concept of Fourier's series approximation [7] is employed for constructing a model for a mainly periodic ambient air temperature, T_R . Joseph Fourier, a French mathematician and physicist introduced in 1807 the approximation of any function $f(x)$ over a finite interval with an infinite sum of sine and cosine functions as $g(x) = a_0 + \sum_{i=1}^{\infty} [a_i \sin(ix) + b_i \cos(ix)]$ such that $\|f(x) - g(x)\|^2 = \min$, where a_0, a_i, b_i are unknown constants. An equivalent formulation may be used as $g(x) = a_0 + \sum_{i=1}^{\infty} a_i \sin(ix + b_i)$ for brevity. Instead of an infinite series, $i \in [0, n]$ is used for the approximation of a mainly periodic function, eliminating the terms multiplied by the insignificant amplitudes $a_i, i > n$.

Applying the concept for T_R with a mean temperature, T_m , and a harmonic variation component, $\sum A_{\omega_i} \sin(\omega_i t + \alpha_i)$, where A_{ω_i} is the amplitude of the harmonic variation component of $f_i(t) = \sin(\omega_i t + \alpha_i)$. The amplitude, A_{ω_i} , may be a linear function of time in some models, $A_{\omega_i}(t) = d_{1,i} + d_{2,i}t$.

There are various choices to model the T_R component, listed as M1 through M5. The M1-type model is a linear function. It has the least coefficients and can

be used to describe the yearly mean temperature change. However, it does not have the ability to reflect any periodic temperature variation. The M2-type model is the general Fourier series function. It assumes that the yearly mean temperature and amplitudes for the pre-selected, finite number of frequencies are constant. It might be accurate for a short period of time, such as one year. The problem with the M2-type model is that it cannot reflect the long-term average, the maximum and the minimum temperature changes with time as in Bhutiya-ni's study [3]. The M3-type model is an updated function from the M2-type model. It has the variable yearly mean temperature with time. However, it still does not have the ability to reflect the maximum and minimum temperature changes with time. The M4-type model is an improved form over the M2-type model in terms of allowing the variation of maximum and minimum temperatures, but still using the constant yearly mean temperature. The M5-type model combines the advantages of both the M3-type and M4-type of models.

Therefore, the five different models tested are as follows:

M1. Variable mean temperature

$$T_R(t) = T_m + c \cdot t \quad (3)$$

M2. Constant mean temperature and constant amplitude series:

$$T_R(t) = T_m + \sum A_{\omega_i} f(t) \quad (4)$$

M3. Variable mean temperature and constant amplitude series:

$$T_R(t) = T_m(t) + \sum A_{\omega_i} f(t) \quad (5)$$

M4. Constant mean temperature and variable amplitude series:

$$T_R(t) = T_m + \sum A_{\omega_i}(t) f(t) \quad (6)$$

M5. Variable mean temperature and variable amplitude series:

$$T_R(t) = T_m(t) + \sum A_{\omega_i}(t) f(t) \quad (7)$$

The M1-type model is used for comparison with other models for the yearly mean temperature variation evaluation. Bhutiya-ni [3] studied the average, maximum and minimum temperature trends for 100 years and found them all changing with time. Therefore, the M2-type model with a stationary, constant mean temperature and constant amplitudes is inferior for practical application and not used for comparisons. The M4-type model is not recommended nor studied for brevity as the M5-type model gives better result for the same effort. Therefore, for comparison purpose, only the M1, M3 and M5 model types are used in model testing.

The first task is to depress the stochastic temperature variation in order to find the statistically most significant trend for a base stationary temperature model. Second, the stochastic or chaotic deviations must be defined to match the observed temperature. Therefore, the designers or analysts may conduct their studies or design safely without missing the expected, maximum or minimum temperature values for the study time period. **Figure 1** shows the components of establishing an analytical temperature model.

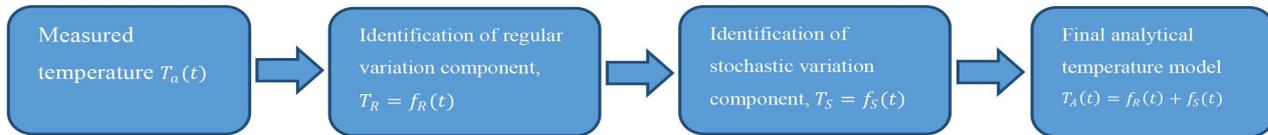


Figure 1. Components of establishing an analytical temperature model.

2. Measured Temperature Data

Daily mean temperature measurements are used in the study from a middle-west location in North America for 20 years. The data, $T_a(t)$, are taken for 7305 days from 04/01/1996 to 03/31/2016, downloaded from <https://www.wunderground.com>, and plotted in **Figures 2(a)**. **Figure 2(b)** and **Figure 2(c)** are enlargement of **Figure 2(a)** for year 1 and year 20, respectively. Although the regular trend, $T_R(t)$, is to be discussed in Point 3 of the paper, it is also shown from the M5-type model in **Figure 2(a)**. **Figure 2(b)** and **Figure 2(c)** for illustrating the model concept. The daily mean temperature data, $T_a(t)$, is taken from the Standard Model method in the study as a practical compromise. Three different ways of using the measured data are tried for comparison and to find the best way of data processing.

2.1. Single-Year Temperature Cycle Evaluation

In the first usage of the data, $T_a(t)$, the measured mean daily temperatures for 20 years are divided into twenty sets for single year from 1 to 20 to be able to analyze the model validity for model type M1 and M3. Individual yearly data, $T_{a,y}(t)$, $y \in [1, 20]$, $t \in [1, 365]$ are grouped for 15 regular years and $t \in [1, 366]$ for 5 leap years.

2.2. Four-Year Temperature Cycle Evaluation

In the second usage of the data, $T_a(t)$, the 20 years measured mean daily temperatures are divided into 4-year period sets, giving 5 groups as 1), 2), 3), 4) and 5). The justification of employing four years as the true solar time period is that the yearly time period for regular years is distorted by the deficiency of 0.25 days while the time period in leap years is longer by 0.75 days and affecting the averaging. The 4-year time sequence of 1461 days is considered as the repeating time period for the T_S stationary temperature component in the model. Therefore, properties of temperature (average, etc.) also must be considered distinguished when evaluating for the yearly time period.

1) For year 1 - 4:

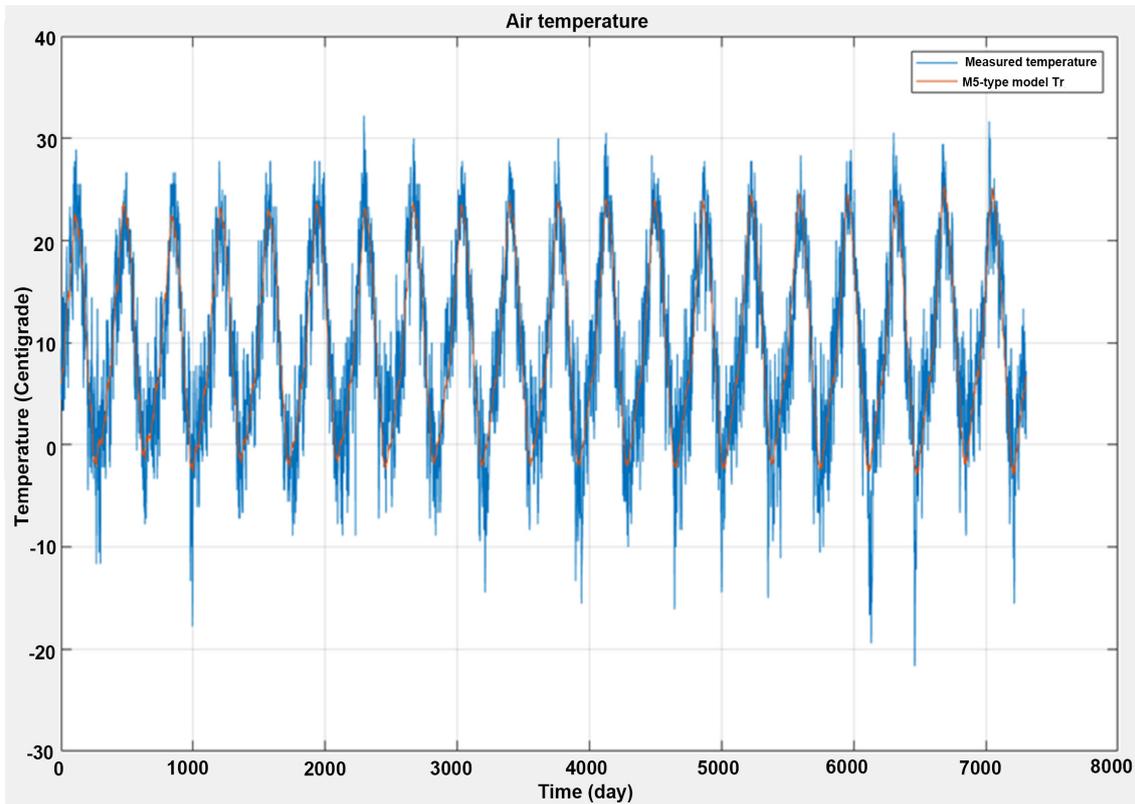
$$T_{a,a} = T_a(t), \quad t \in [1, 1461] \quad (8)$$

2) For year 5 - 8:

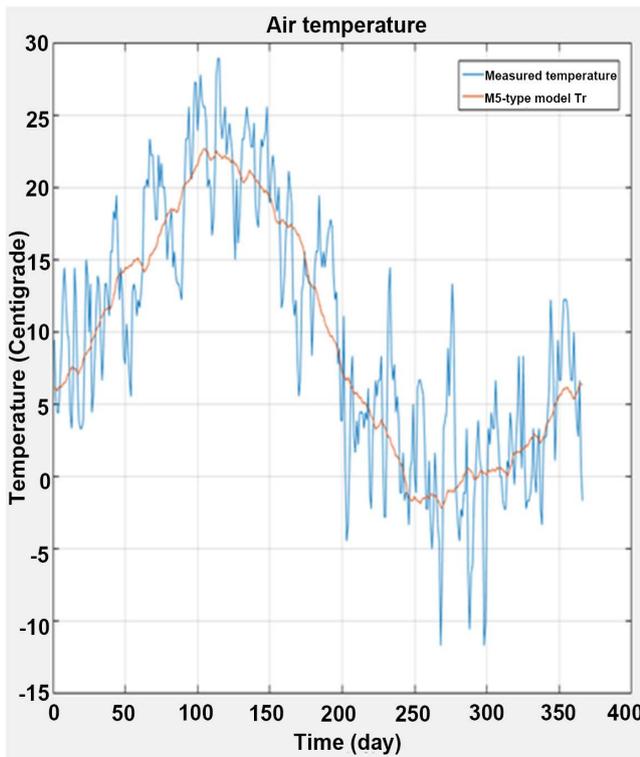
$$T_{a,b} = T_a(t), \quad t \in [1462, 2922] \quad (9)$$

3) For year 9 - 12:

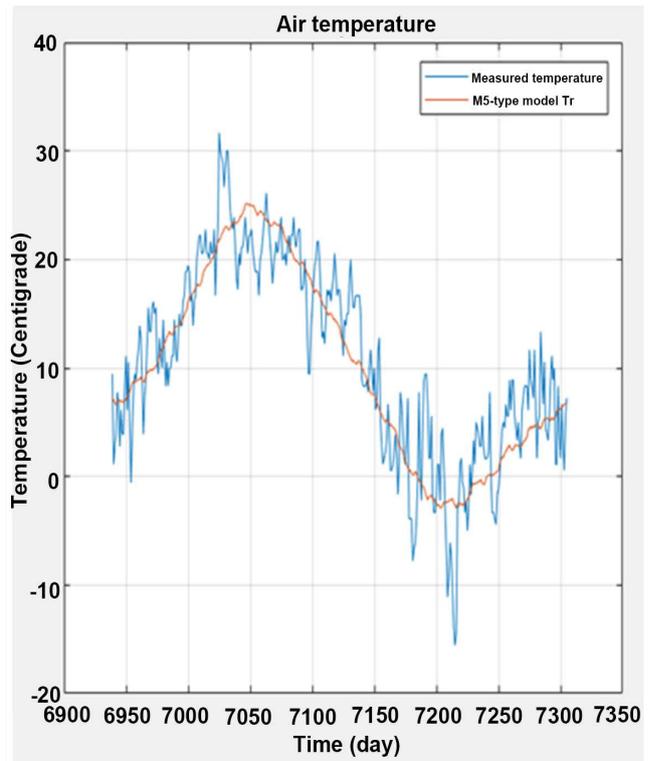
$$T_{a,c} = T_a(t), \quad t \in [2923, 4383] \quad (10)$$



(a)



(b)



(c)

Figure 2. (a) Measured 20 years daily mean temperature at middle west of North America; (b) Zoom in **Figure 2(a)** for the first year; (c) Zoom in **Figure 2(a)** for the last year.

4) For year 13 - 16:

$$T_{a,d} = T_a(t), \quad t \in [4384, 5844] \quad (11)$$

5) For year 17 - 20:

$$T_{a,e} = T_a(t), \quad t \in [5846, 7305] \quad (12)$$

2.3. Continuous, Repeating Four-Year Temperature Cycles Evaluation

In the third usage of the data, $T_a(t)$, the total 20 years 7,305 data, $T_a(t)$, $t \in [1, 7305]$, are used for testing and establishing properties of $T_R(t)$ in model type M4 and M5.

3. Determination of T_R

To determine the model coefficients, the obvious choice is to use the Least-squares (LSQ) fit method. The LSQ method optimally fits the data to a given function with unknown constant parameters in such a way that the root-mean-square of the error between model and measured data is minimalized. The expected, fitted equations represent the significant, regular temperature trend, T_R . Determining the significant part of the data out of a noisy observation may also be done by filtering or neural network. The advantage of using the LSQ method is to be able to define function T_R in advance, whereas signal processing does not give an analytical form of such a function [8].

For supporting the ways of using measured data, different fitting function may be defined with a number of unknown parameters to be determined by the LSQ fitting algorithm. Five different fitting functions may be considered for M1, M3 and M5 model types as follows.

3.1. T_R Function for the Entire 20 Years for M1-Type Model

The LSQ linear function is:

$$T_{m,t} = T_m + c \cdot t \quad (13)$$

3.2. T_R Function for M3-Type Model

For single year data in the M2-type model, a regular year of 365 days and a leap year of 366 days must be distinguished. The LSQ function for regular year is written as:

$$T_{R,y}(t) = T_m + c \times t + A_1 \times \sin(2\pi \times t \times a_1 + b_1) + A_2 \times \sin(2\pi \times t \times a_2 + b_2) + A_3 \times \sin(2\pi \times t \times a_3 + b_3) + \dots + A_{15} \times \sin(2\pi \times t \times a_{15} + b_{15}) \quad (14)$$

where $a_k = \frac{1}{365} \times \left\{ \left\{ 2^{i-1} \right\}, 3 \times \left\{ 2^{j-1} \right\} \right\}$; for $k \in [1, 15]$, $k = i + j$; $i = i \in [1, 8]$; $j \in [1, 7]$, all fixed frequency components. Unknown parameters are T_m , c , A_1 through A_{15} , and b_1 through b_{15} .

The LSQ function for leap year is derived from (14) by changing 365 days to 366 days as:

$$T_{R,y}(t) = T_m + c \times t + A_1 \times \sin(2\pi \times t \times a_1 + b_1) + A_2 \times \sin(2\pi \times t \times a_2 + b_2) + A_3 \times \sin(2\pi \times t \times a_3 + b_3) + \dots + A_{15} \times \sin(2\pi \times t \times a_{15} + b_{15}) \quad (15)$$

where $a_k = \frac{1}{365} \times \{ \{2^{i-1}\}, 3 \times \{2^{j-1}\} \}$; for $k \in [1, 15]$, $k = i + j$; $i = i \in [1, 8]$; $j \in [1, 7]$, all fixed frequency components. Unknown parameters are T_m , c , A_1 through A_{15} , and b_1 through b_{15} .

The LSQ function for 4 years data and 20 years data will add the four-year and two-year period frequencies, also will change the one-year period to 365.25 days, the function is written as:

$$T_R(t) = T_m + c \times t + A_1 \times \sin(2\pi \times t \times a_1 + b_1) + A_2 \times \sin(2\pi \times t \times a_2 + b_2) + A_3 \times \sin(2\pi \times t \times a_3 + b_3) + \dots + A_{17} \times \sin(2\pi \times t \times a_{17} + b_{17}) \quad (16)$$

where $a_k = \frac{1}{365.25} \times \{ \{2^{i-3}\}, 3 \times \{2^{j-1}\} \}$; for $k \in [1, 17]$, $k = i + j$; $i = i \in [1, 10]$; $j \in [1, 7]$, all fixed frequency components. Unknown parameters are T_m , c , A_1 through A_{17} and b_1 through b_{17} .

3.3. T_R Function for M5 Type Model

With the assumption that the amplitudes may also vary with time, a modified LSQ function over (16) is established as:

$$T_R(t) = T_m + c \times t + (A_{1c} + A_{1v} \times t) \times \sin(2\pi \times t \times a_1 + b_1) + (A_{2c} + A_{2v} \times t) \times \sin(2\pi \times t \times a_2 + b_2) + \dots + (A_{17c} + A_{17v} \times t) \times \sin(2\pi \times t \times a_{17} + b_{17}) \quad (17)$$

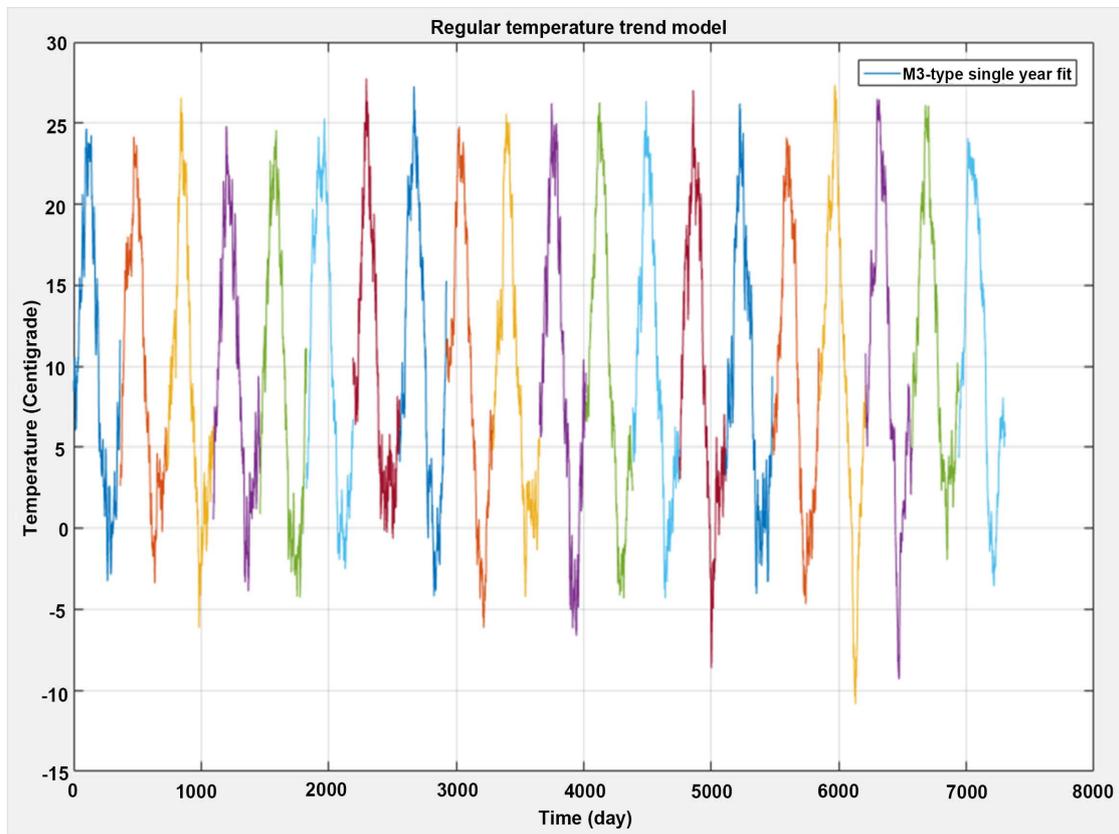
where $a_k = \frac{1}{365.25} \times \{ \{2^{i-3}\}, 3 \times \{2^{j-1}\} \}$; for $k \in [1, 17]$, $k = i + j$; $i = i \in [1, 10]$; $j \in [1, 7]$, all fixed frequency components. Unknown parameters are T_m , c , A_{1c} through A_{17c} , A_{1v} through A_{17v} , and b_1 through b_{17} .

4. Evaluation of the Statistically Most Significant, Mean Temperature Trends

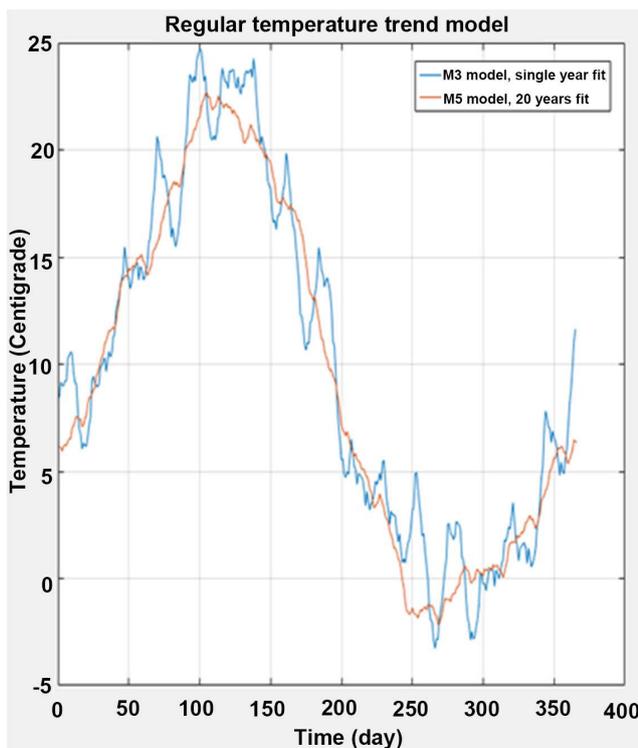
The LSQ fitting method is used to determine the mean temperature trend, $T_R(t)$, of the measured data, $T_a(t)$. The LSQ method provides the statistically most significant result for $T_R(t)$ as a regular, deterministic trend, depressing the random variation component of temperature around $T_R(t)$ with assumed, normal distribution as a noise due to stochastic or chaotic causes.

First, the best LSQ fit is determined on all single year data separately. The parameters of function (14) are applied for the regular years, and (15) is applied for the leap years. The fitted results are shown in **Figure 3(a)**, **Figure 3(b)** and **Figure 3(c)**. The mean, maximum and minimum values from the fitted result are shown in **Figure 4**.

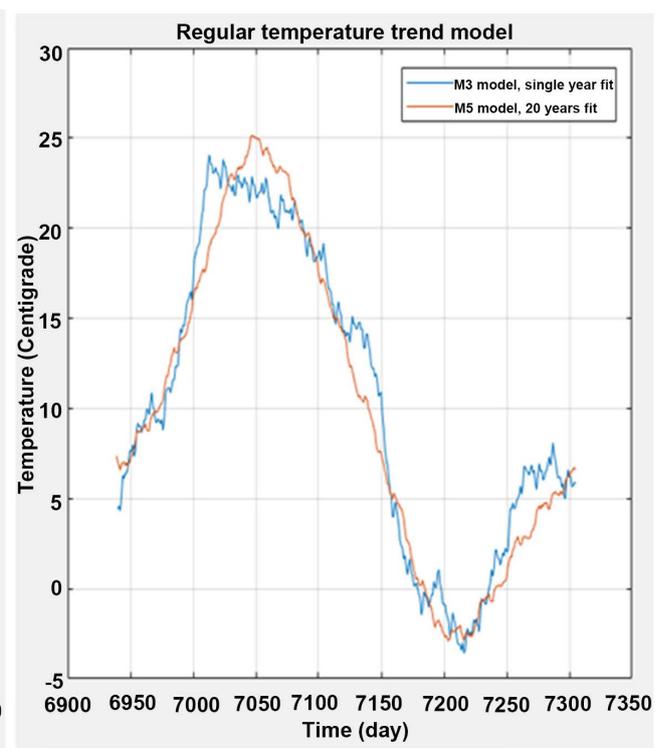
Second, the best LSQ fit is found using function (16) for five 4-year temperature data sets separately. The fitted results are shown in **Figure 5(a)**, **Figure 5(b)** and **Figure 5(c)** the mean, maximum and minimum values from the fitted



(a)



(b)



(c)

Figure 3. (a) LSQ fitting curves for each year of 20 years; (b) Zoom in **Figure 3(a)** for the first year; (c) Zoom in **Figure 3(a)** for the last year.

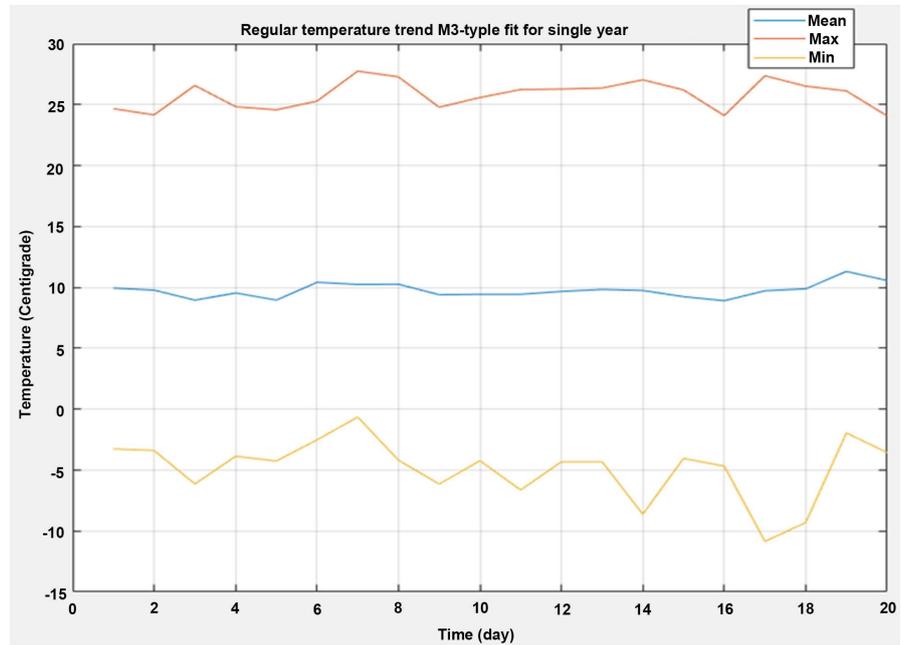


Figure 4. The mean, maximum and minimum values for LSQ fitting results for each year.

result are shown in **Figure 6**. Parameters for the fitted functions $T_{a,a}(t), \dots, T_{a,e}(t)$, are listed in **Tables 1-5**, respectively.

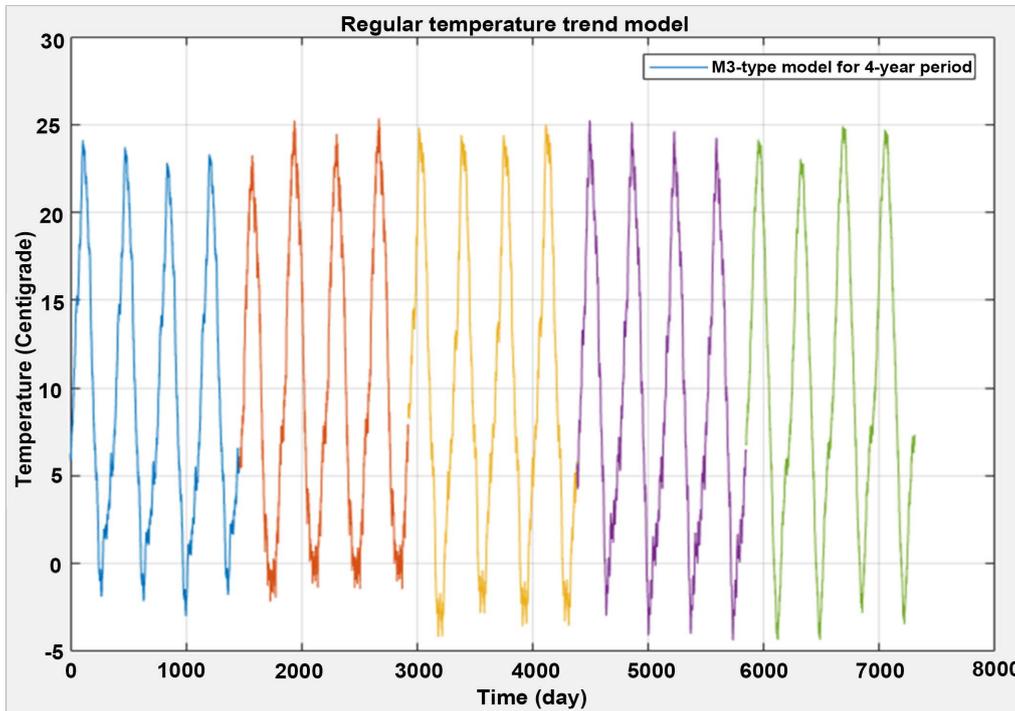
Third, the best LSQ fit is determined using the M3-type function (16) for all 20 years data, $T_a(t)$, together. The fitted function is depicted in **Figure 7(a)**, **Figure 7(b)** and **Figure 7(c)**. The parameters of the fitted function are listed in **Table 6**. The best LSQ fit is also determined using the improved, M5-type function (17), applied also for all 20 years data, $T_a(t)$. The fitted results are depicted in **Figure 8(a)**, **Figure 8(b)** and **Figure 8(c)**; and the parameters of the fitted function are listed in **Table 7**.

5. Evaluation of the Stochastic Variation and Final Analytical Temperature Model

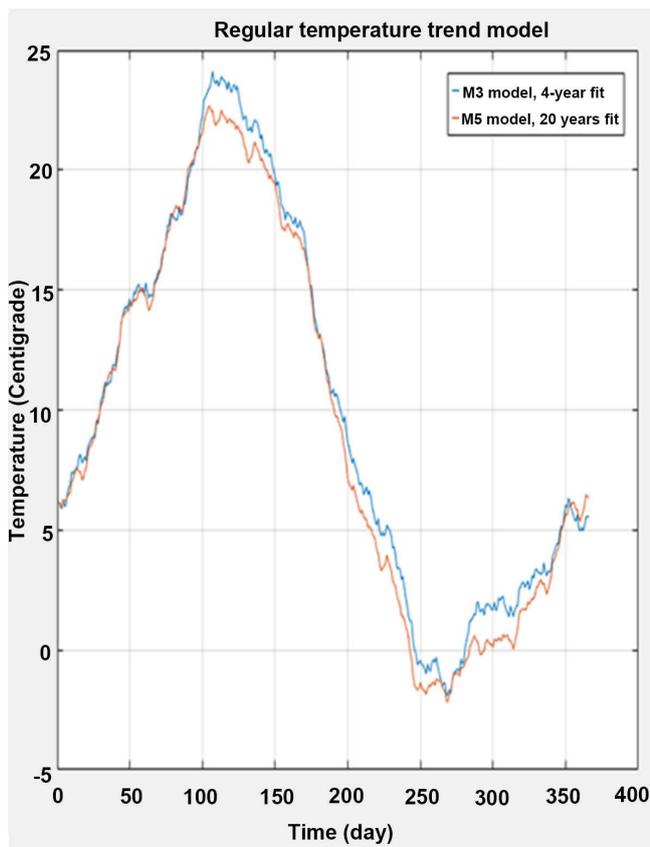
The stochastic variation must be defined by subtracting the statistic periodic function from the measured data. First, the stochastic variation, $T_S(t)$, is defined by the difference between deterministic function result, $T_R(t)$ and measured data $T_a(t)$ as:

$$T_S(t) = T_a(t) - T_R(t) \tag{18}$$

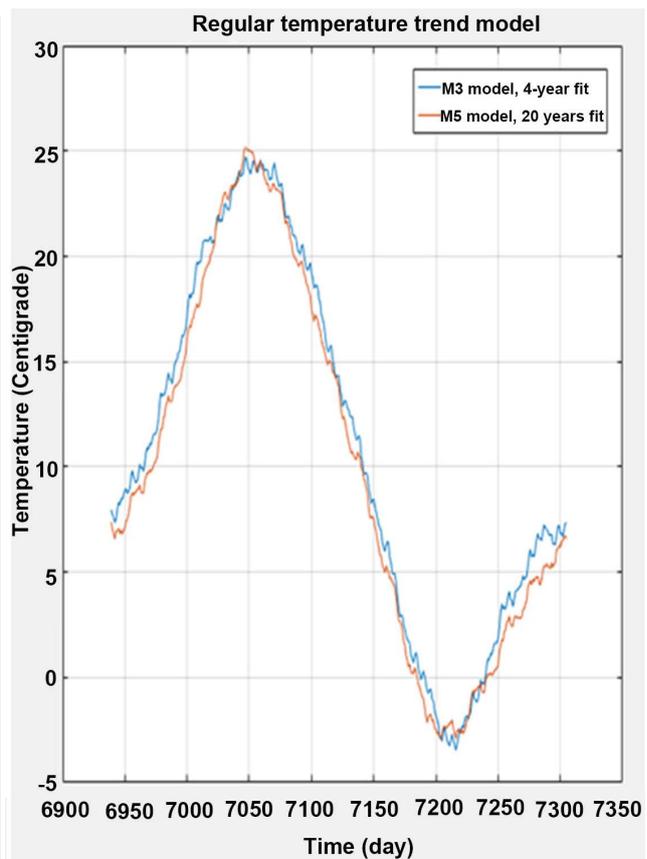
The data of (18) is depicted in **Figure 9(a)**, **Figure 9(b)** and **Figure 9(c)**. The density analysis on the (18) data and the MATLAB normal distribution fit of the data is shown in **Figure 10**. From the fit, a mean value $mu = -6.33 \times 10^{-8} \approx 0$, and standard deviation of $\sigma = 4.0012$ are obtained, where $mu = \sum T_S(t)/n$, $\sigma = \sqrt{\frac{1}{n} \sum (T_S(t) - mu)^2}$ and n is the number of days, t . Using $mu = 0.00$ and $\sigma = 4.0012$, a normally-distributed random noise series is generated to represent $T_S(t)$. The $NormRandom(mu, \sigma, t)$ function is used in Matlab that



(a)



(b)



(c)

Figure 5. (a) LSQ fitting curves for five 4-year sets; (b) Zoom in **Figure 5(a)** for the first year; (c) Zoom in **Figure 5(a)** for the last year.

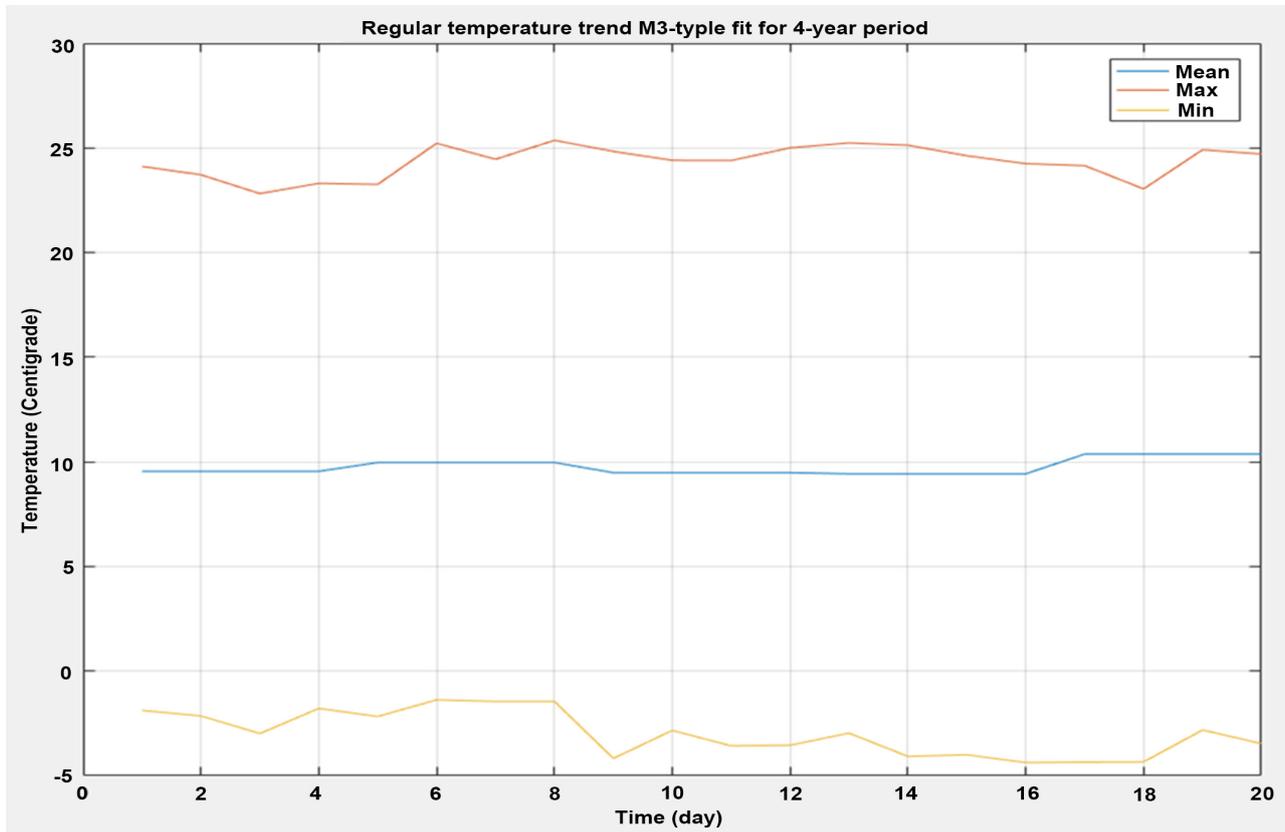


Figure 6. The mean, maximum and minimum values for LSQ fitting results for 4-year sets.

Table 1. Parameters for year 1 - 4 data, the LSQ fitted function (16) result, $T_{a,a}$.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
0.731	-2.48×10^{-4}	0.527	0.982	0.282	2.468	10.616	-0.343	2.281	3.248
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
0.257	2.759	0.584	-0.816	0.798	2.985	0.072	-0.525	0.399	4.295
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.285	-0.807	0.181	-17.310	0.154	1.797	0.117	0.950	0.092	3.044
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.080	1.639	0.142	4.827	0.043	1.059				

Table 2. Parameters for year 5 - 8 data, the LSQ fitted function (16) result, $T_{a,b}$.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
9.208	1.038×10^{-3}	0.377	-1.097	0.529	3.677	11.764	-0.306	1.836	3.450
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
0.231	2.328	0.723	2.227	0.520	2.349	0.158	0.536	0.496	3.325
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.287	0.278	0.116	2.199	0.234	-17.177	0.472	3.009	0.221	3.941
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.178	-0.255	0.124	3.226	0.025	10.918				

Table 3. Parameters for year 9 - 12 data, the LSQ fitted function (16) result, $T_{a,c}$.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
10.735	-1.715×10^{-3}	1.028	3.156	0.756	2.810	11.843	-0.268	1.931	3.252
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
1.239	2.354	0.431	-0.535	0.766	-1.907	0.558	1.876	0.381	-0.040
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.372	1.158	0.195	3.605	0.341	-0.427	0.073	4.363	0.157	1.924
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.167	2.334	0.036	4.023	0.044	4.296				

Table 4. Parameters for year 13 - 16 data, the LSQ fitted function (16) result, $T_{a,d}$.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
9.242	2.539×10^{-4}	0.645	0.197	0.427	-0.565	11.355	-0.357	2.566	3.116
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
0.543	3.556	0.790	0.205	0.723	3.246	0.252	2.302	0.482	3.649
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.417	2.377	0.592	1.153	0.175	2.700	0.153	2.907	0.084	-7.709
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.077	2.546	0.046	0.520	0.100	3.479				

Table 5. Parameters for year 17 - 20 data, the LSQ fitted function (16) result, $T_{a,e}$.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
10.289	1.073×10^{-4}	0.883	3.050	0.372	0.298	11.798	-0.288	2.138	2.958
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
1.234	3.131	0.412	10.828	0.110	1.318	0.206	4.330	0.308	-0.624
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.081	2.438	0.126	2.924	0.125	4.272	0.147	7.643	0.243	2.356
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.038	1.193	0.080	3.917	0.063	2.760				

Table 6. Parameters for year 1 - 20 data, the LSQ fitted function (16) result.

T_m	c	A_1	a_1	A_2	a_2	A_3	a_3	A_4	a_4
9.489	7.367×10^{-5}	0.095	2.478	0.092	3.072	11.489	-0.311	2.113	3.194
A_5	a_5	A_6	a_6	A_7	a_7	A_8	a_8	A_9	a_9
0.622	2.836	0.222	-0.292	0.418	3.237	0.122	1.909	0.197	4.259
A_{10}	a_{10}	A_{11}	a_{11}	A_{12}	a_{12}	A_{13}	a_{13}	A_{14}	a_{14}
0.123	1.122	0.153	1.723	0.041	1.358	0.121	2.682	0.093	2.957
A_{15}	a_{15}	A_{16}	a_{16}	A_{17}	a_{17}				
0.047	-4.723	0.052	4.109	0.031	3.388				

Table 7. Parameters for year 1 - 20 data, the LSQ fitted function (17) result.

T_m	c		A_{1c}	A_{1v}	b_1	A_{2c}	A_{2v}	b_2
9.477	7.367×10^{-5}		0.392	-1.254×10^{-4}	0.066	0.611	-1.409×10^{-4}	3.017
A_{3c}	A_{3v}	b_3	A_{4c}	A_{4v}	b_4	A_{5c}	A_{5v}	b_5
11.099	1.066×10^{-4}	-0.312	1.953	4.541×10^{-5}	3.196	0.020	1.635×10^{-4}	2.989
A_{6c}	A_{6v}	b_6	A_{7c}	A_{7v}	b_7	A_{8c}	A_{8v}	b_8
0.163	1.524×10^{-5}	-0.244	0.731	-8.600×10^{-5}	3.187	0.216	-2.764×10^{-5}	1.628
A_{9c}	A_{9v}	b_9	A_{10c}	A_{10v}	b_{10}	A_{11c}	A_{11v}	b_{11}
0.281	-2.284×10^{-5}	4.153	0.311	-9.252×10^{-5}	-0.670	0.134	4.996×10^{-6}	-4.559
A_{12c}	A_{12v}	b_{12}	A_{13c}	A_{13v}	b_{13}	A_{14c}	A_{14v}	b_{14}
0.248	-5.649×10^{-5}	1.295	0.142	-5.864×10^{-5}	2.677	0.049	1.188×10^{-5}	2.816
A_{15c}	A_{15v}	b_{15}	A_{16c}	A_{16v}	b_{16}	A_{17c}	A_{17v}	b_{17}
0.014	8.843×10^{-6}	1.729	0.099	-1.302×10^{-5}	4.222	0.029	-1.647×10^{-5}	0.078

generates t number of random values from the normal distribution with a mean value mu , and standard deviation value σ . Applying it to $T_s(t)$, it gives:

$$T_s(t) = NormRandom(mu, \sigma, t) \tag{19}$$

For $T_A(t)$, (19) is added to (17):

$$T_A(t) = T_R(t) + NormRandom(mu, \sigma, t) \tag{20}$$

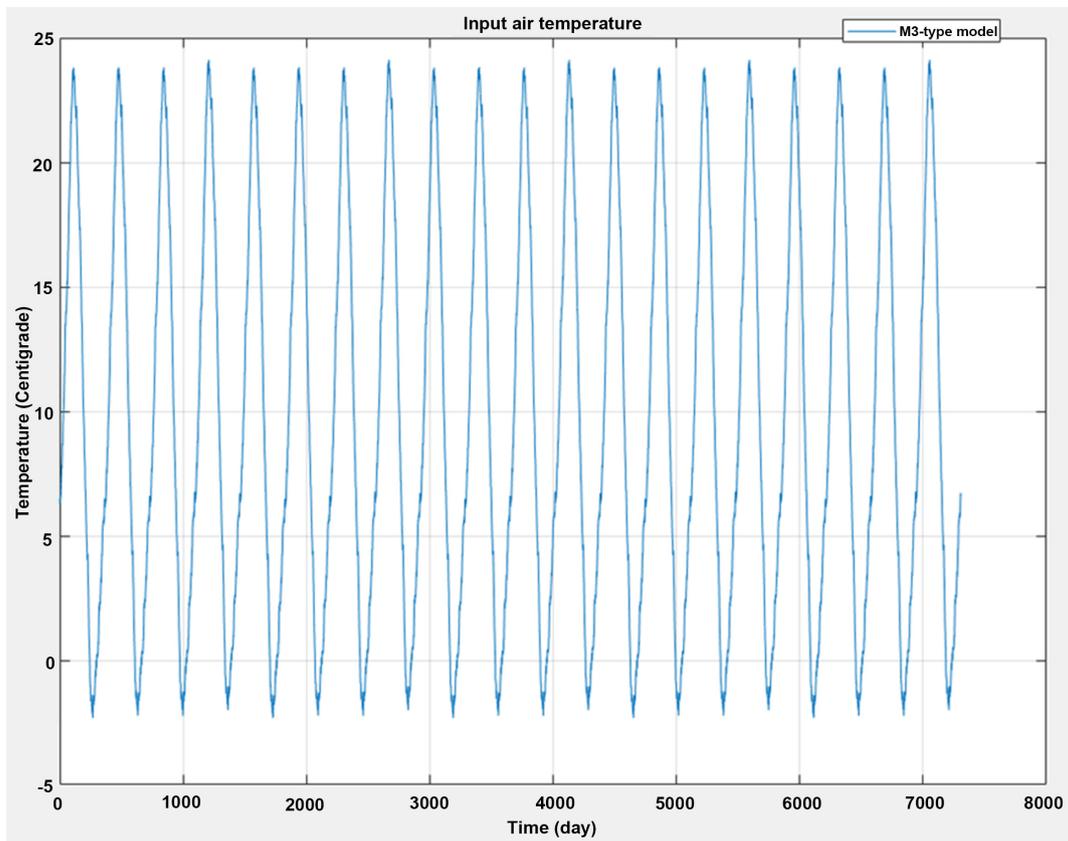
However, (20) is not an analytical function since it includes an algorithm. To overcome this and understanding that the daily variation for random causes is a sample of $T_s(t)$, the maximum and minimum values can be generated with a 99 per cent confidence by a fluctuating temperature with a 2-day cycle time:

$$T_A(t) = T_R(t) + 3\sigma(-1)^t \tag{21}$$

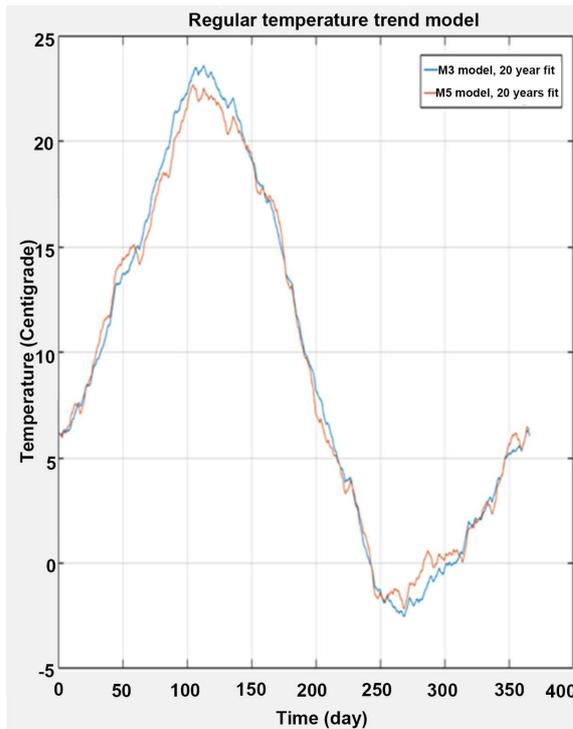
Substituting the preferred model in (17), the final analytical temperature model, $T_A(t)$ is:

$$\begin{aligned}
 T_A(t) = & T_m + c \times t + (A_{1c} + A_{1v} \times t) \times \sin(2\pi \times t \times a_1 + b_1) \\
 & + (A_{2c} + A_{2v} \times t) \times \sin(2\pi \times t \times a_2 + b_2) + \dots \\
 & + (A_{17c} + A_{17v} \times t) \times \sin(2\pi \times t \times a_{17} + b_{17}) + 3\sigma(-1)^t
 \end{aligned} \tag{22}$$

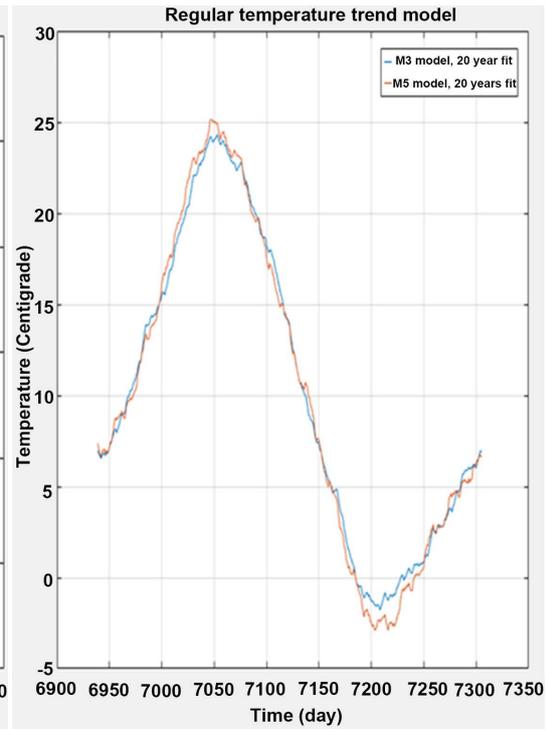
Comparison between simulated temperature, $T_A(t)$, from (20) and measured data, $T_a(t)$, is show in **Figure 11(a)**, **Figure 11(b)** and **Figure 11(c)**. Using an uncertainty interval of $3\sigma = 12.0036$ in the example, safe limits of maximum and minimum temperature are show in **Figure 12(a)**, **Figure 12(b)** and **Figure 12(c)**. A smaller or wider uncertainty interval may also be selected based on the risk to be taken regarding the missing number of days to be in the expected temperature range in the outside air temperature.



(a)

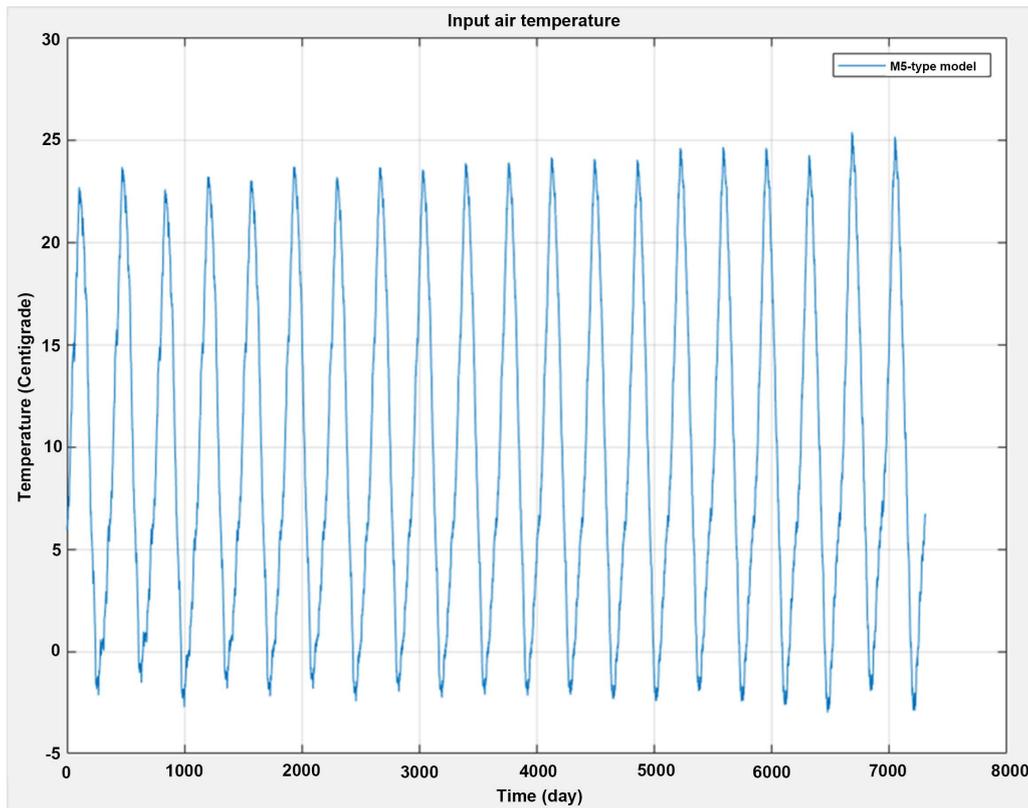


(b)

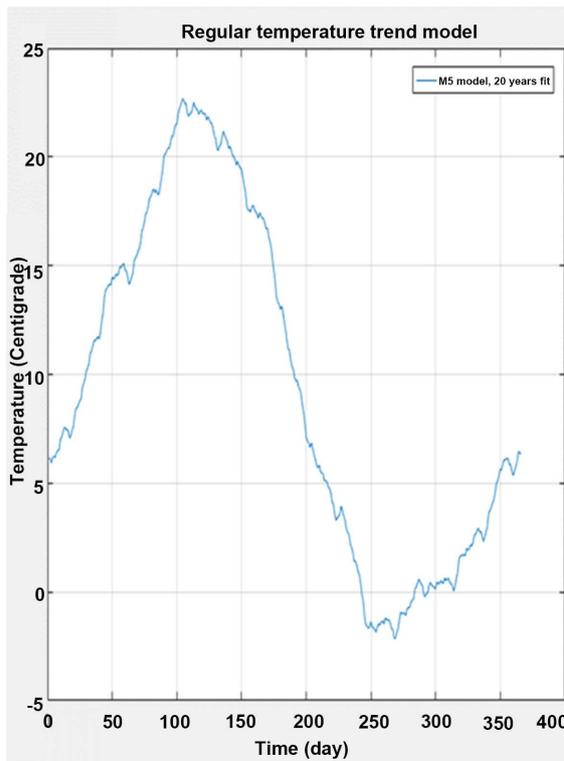


(c)

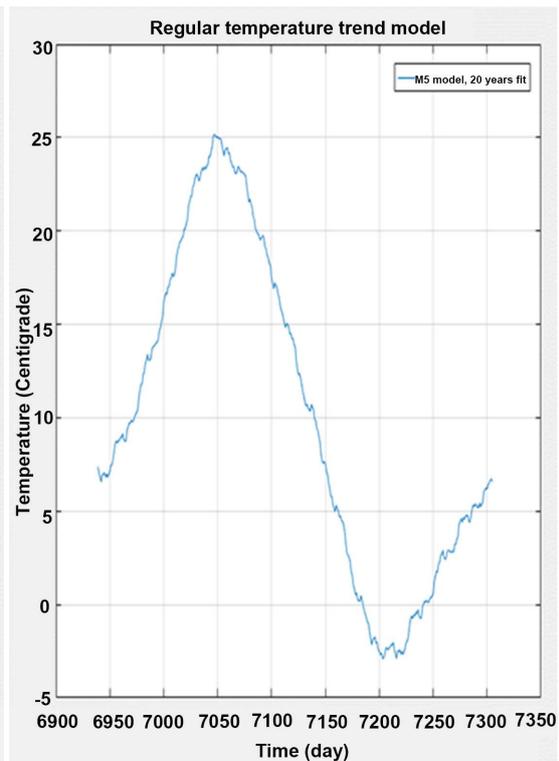
Figure 7. (a) LSQ function (16) fitting curve for total data; (b) Zoom in **Figure 7(a)** for the first year; (c) Zoom in **Figure 7(a)** for the last year.



(a)

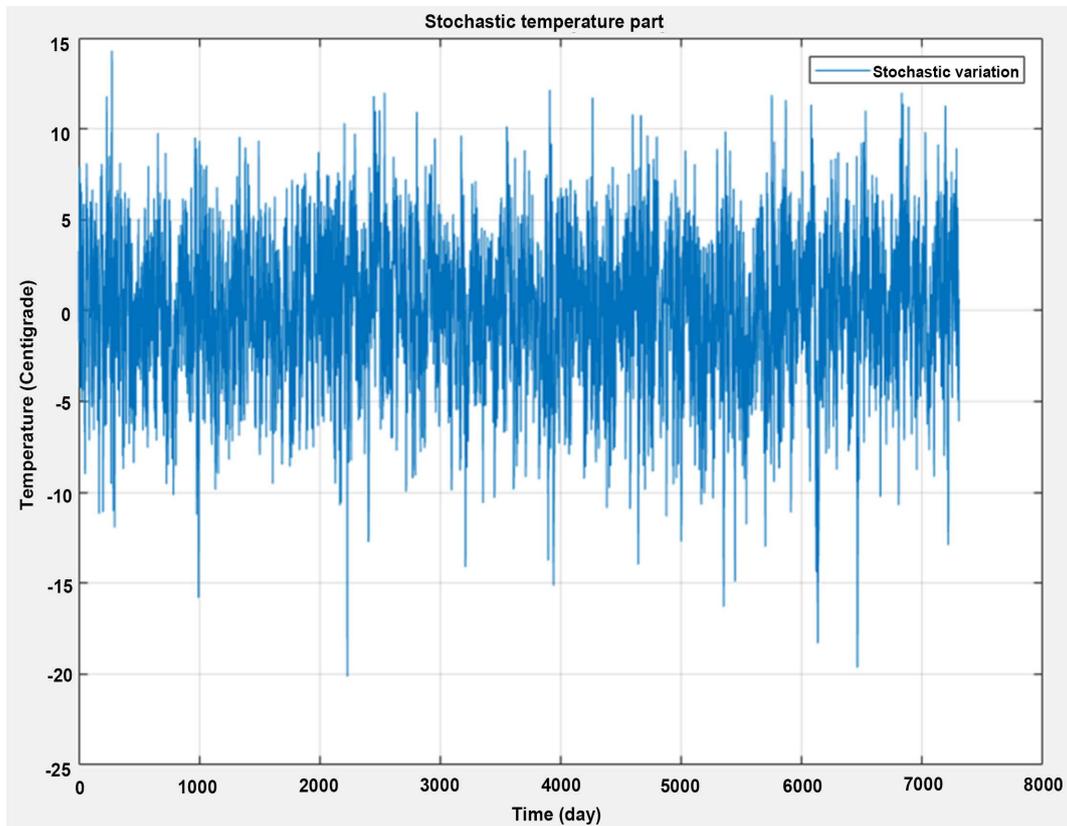


(b)

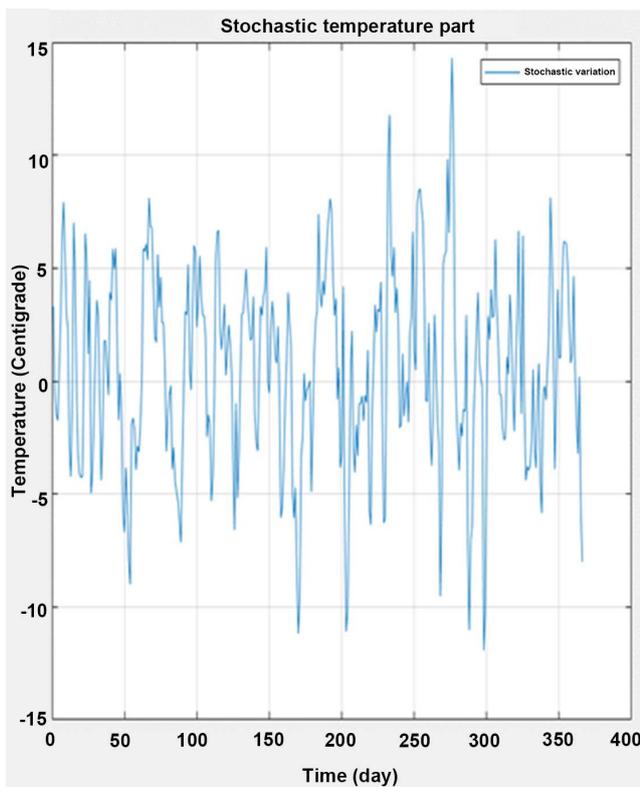


(c)

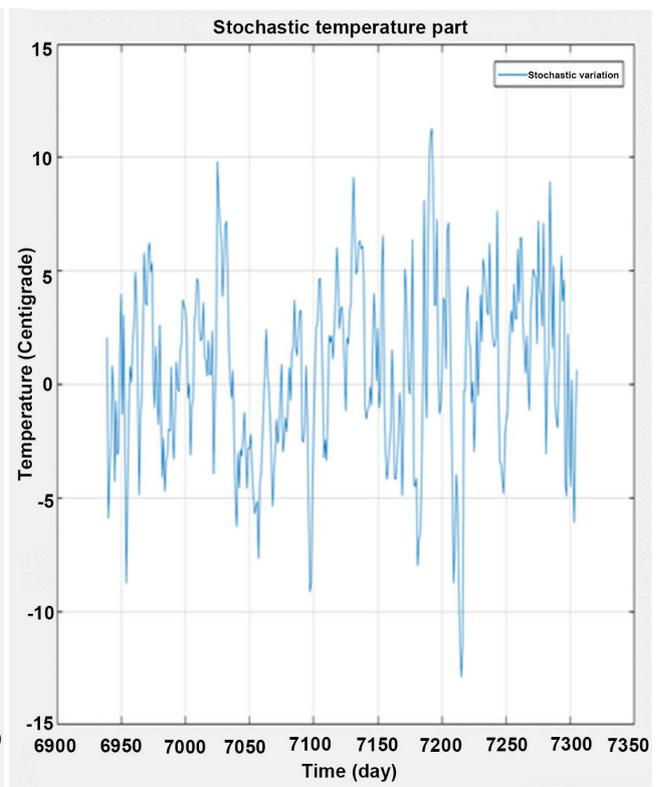
Figure 8. (a) LSQ function (17) fitting curve for total data; (b) Zoom in **Figure 8(a)** for the first year; (c) Zoom in **Figure 8(a)** for the last year.



(a)



(b)



(c)

Figure 9. (a) Equation 18 result; (b) Zoom in **Figure 9(a)** for the first year; (c) Zoom in **Figure 9(a)** for the last year.

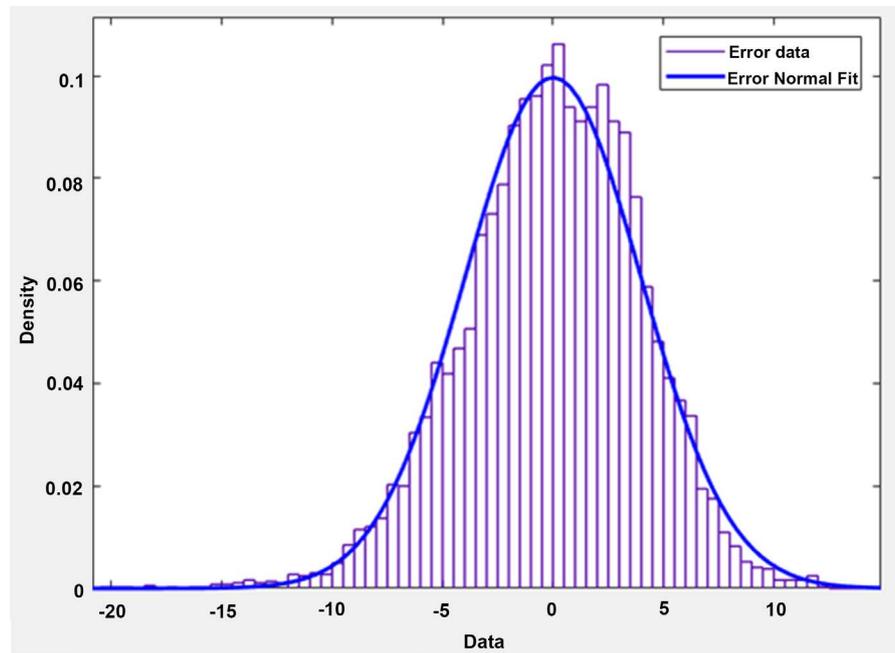


Figure 10. Data Equation (18) result density analysis and normal distribution fit.

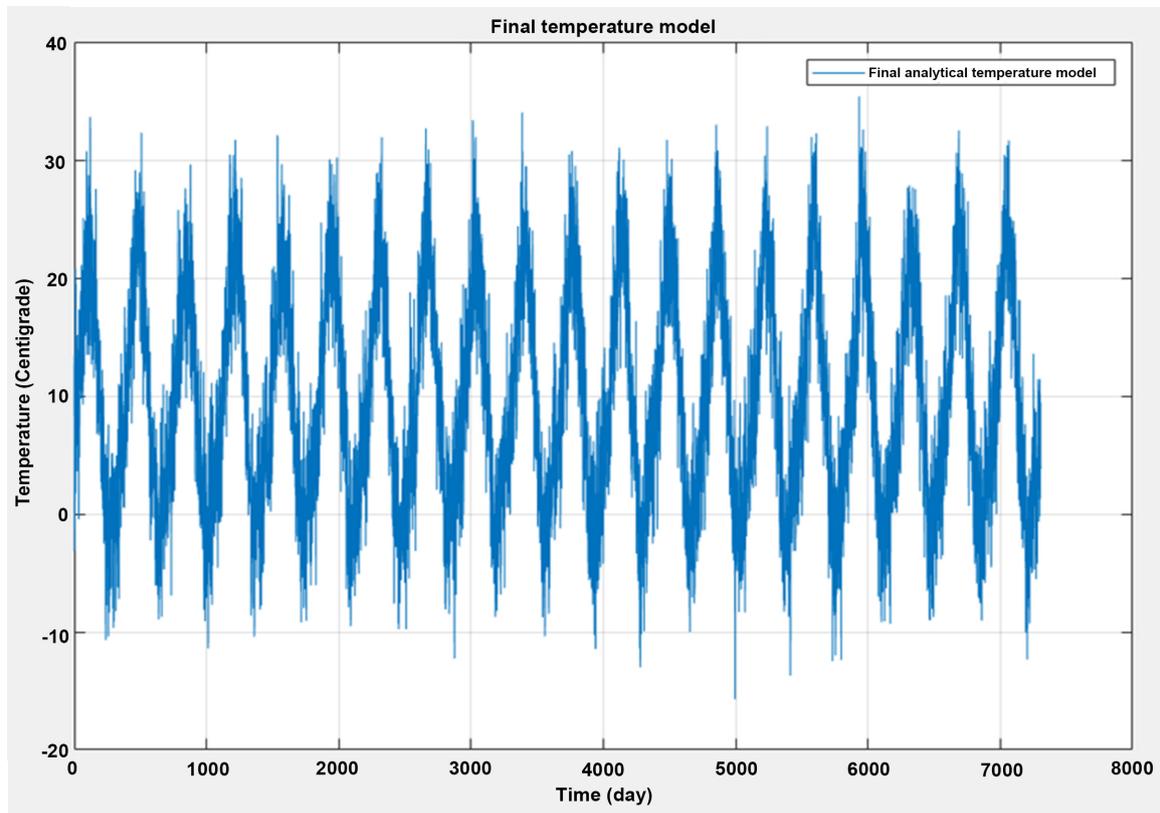
6. Discussion of the Results

6.1. Discussion of the Results for the Air Temperature Variation as a Regular Trend

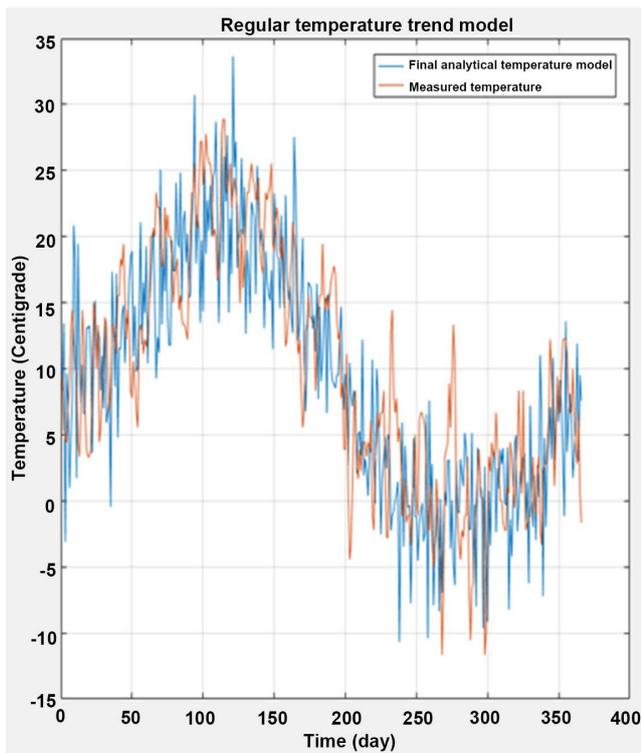
For comparison purposes, the linear regression function (13) for the entire 20-year data set is applied to various, fitted model results; or original, unprocessed data. The fitted model results for the shorter time periods represent the significant part of the repeated trends whereas the noise is intentionally depressed in the LSQ norm sense. Therefore, a linear regression evaluation for the 20-year long time period is assumed to evaluate the most significant, time-average of the linear change in the magnitude of $T_a(t)$. Common expectation dictates that the linear regression evaluation for the 20-year long time period of the original data may provide an un-biased result for the linear change in the magnitude of $T_a(t)$. The following studied are completed for fitting a longer-time linear regression to model results, $T_R(t)$, of shorter time periods:

- a) Yearly mean temperatures, $T_R(t)$, (for 15 regular and 5 leap years) evaluated from fitted function to single-years data, $T_{a,y}(t)$ with M1-type model;
- b) Yearly mean temperatures, $T_R(t)$, evaluated from fitted function data to 4-year data sets, $T_{a,a}(t), \dots, T_{a,e}(t)$ with M1-type model;
- c) Yearly mean temperatures, $T_R(t)$, from fitted function to continuous 20 years data, $T_a(t)$ with M1-type model;
- d) 20 years measured data, $T_a(t)$, used unprocessed.

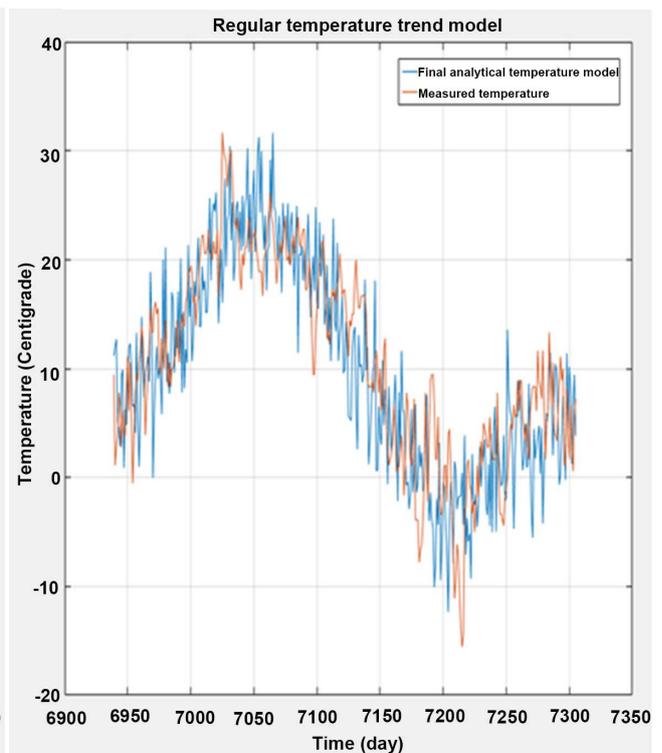
The results from the evaluation are listed in **Table 8**. As shown, the linear trends for the mean value, T_m , and the slope, c , are very similar for cases a) and b) with b) giving a slightly lower RMS error variation than those of a). The reason lays in the fact that the true time period for the variation of the outside



(a)

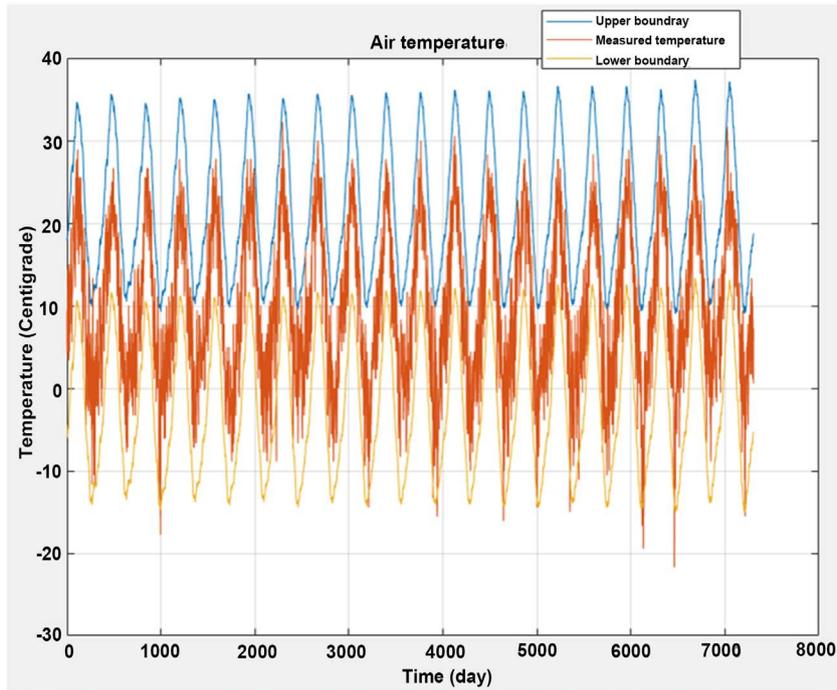


(b)

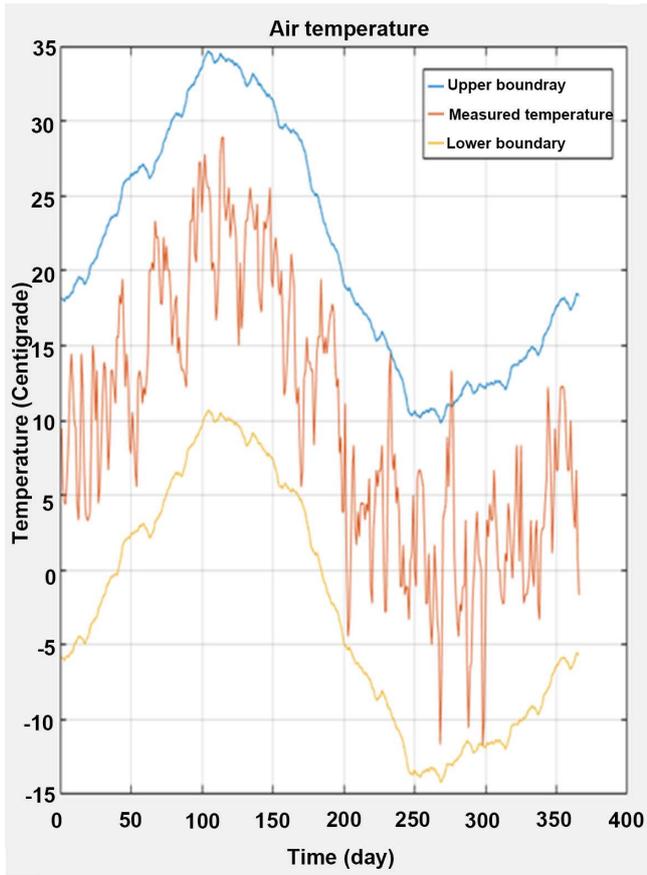


(c)

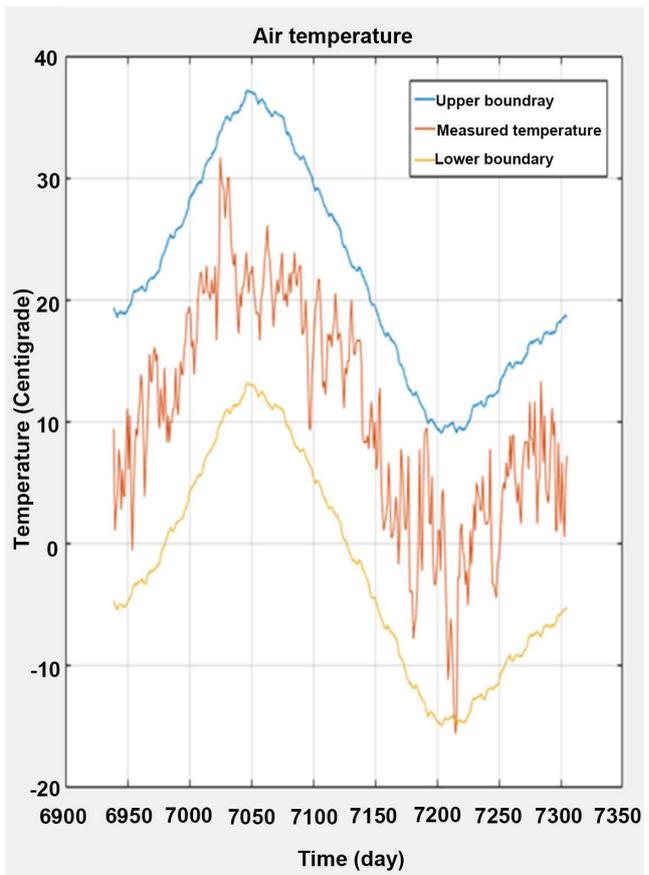
Figure 11. (a) Final analytical temperature model for 20 years; (b) Zoom in **Figure 11(a)** for the first year; (c) Zoom in **Figure 11(a)** for the last year.



(a)



(b)



(c)

Figure 12. (a) Upper boundary, lower boundary and measured temperature; (b) Zoom in **Figure 12(a)** for the first year; (c) Zoom in **Figure 12(a)** for the last year.

Table 8. Linear temperature variation trends evaluated from different models and processes.

Data and Model Type	T_m	c	RMS
a) Yearly mean temperature evaluated from fitted function data, $T_R(t)$, using single-years data, $T_{a,y}(t)$ and M1-type model	9.456	7.88×10^{-5}	0.603
b) Yearly mean temperature evaluated from fitted function data, $T_R(t)$, using 4-year sets, $T_{a,a}(t), \dots, T_{a,e}(t)$ and M1-type model	9.481	7.23×10^{-5}	0.390
c) Yearly mean temperature from fitted function data, $T_R(t)$, using continuous 20 years data, $T_a(t)$ and M1-type model	9.456	7.88×10^{-5}	0.603
d) 20 years measured data, $T_a(t)$, unprocessed; M1-type model	9.929	-4.66×10^{-5}	(9.206)
e) 20 years measured data, $T_a(t)$, unprocessed; M3-type model	9.489	7.37×10^{-5}	(4.026)
f) 20 years measured data, $T_a(t)$, unprocessed; M5-type model	9.477	7.42×10^{-5}	(4.001)
g) 20 years model output data, $T_R(t)$ from M5-type model	9.929	-4.66×10^{-5}	(8.291)

temperature is 365.25 days, giving a rounding error with a weight of $-0.25/4$ days for the regular years and of $+0.75/4$ day for the leap year in the single-year model fits. The model fit to the 4-year time periods does not have the rounding error problem and, therefore, a smoother fit is expected. Indeed, the RMS error of 0.39 is lower for case b) than value of 0.603 for case a).

The results in case c) is identical to those of case a) for obvious reason of using the same linear regression repeated two times sequentially, the second time obtaining zero RMS value. The result for case d) is very different from those in cases a) through c). Why does a 20-year long data set gives an average decrease of temperature change negative that would translate to “global cooling” as opposed to “global warming” for the example location? The answer is the wrong-type function choice for the most significant variation trend, $T_R(t)$, being a linear function with time. This exercise highlights the importance of the selection for the shape of $T_R(t)$. If a form as inadequate as a linear function is selected for $T_R(t)$ for estimating the periodic nature of the outside temperature, the coefficients of the function cannot be trusted even for the general slope, as demonstrated with case d).

Two more choices are also studied for comparison for evaluating the linear trend which the models already include as the mean value, T_m , and the slope, c . Due to these built-in components, no additional, linear regression fit is needed for determining the values of T_m and c :

e) 20 years measured data, $T_a(t)$, unprocessed; M3-type model;

f) 20 years measured data, $T_a(t)$, unprocessed; M5-type model.

The results from the evaluation are listed in **Table 8**. As shown, the linear trends for the mean value, T_m , and the slope, c , in both cases e) and f) are very similar for case b) giving nearly the same average temperature, T_m , but a slightly lower value for the slope c . The RMS error of fit, shown in parentheses are irrelevant for comparison as these model fits use the measurement data directly without pre-processing as $T_R(t)$ with an assumed model first. The preferred solution for trend analysis is the M5-type model that has a built-in, linear am-

plitude variation component with time in case f). The preferred solution is to directly apply the most relevant model type to raw measurement data for capturing the characteristics of the regular temperature change with time.

Re-fitting another linear regression model to the model output data, $T_R(t)$, from the M5-type model in case f) for re-capturing the mean value, T_m , and the slope, c , does not give back the same values as those built in the best-fit model, shown in case g) in **Table 8**. The reason for the mismatch is the inadequate function type of a linear variation attempting to evaluate the more complex interactions of periodic and monotonic components in $T_R(t)$. However, the general proof of this mismatch is left to the reader.

6.2. Discussion of the Results for the Random Component of Air Temperature Variation

The air temperature model component for the description of the random part due to stochastic or chaotic causes is simplified to be time-independent. The stochastic component, T_R , satisfies the zero mean value and zero slope with time. No attempt has been made to vary the magnitude of randomness with the seasons. Refinement for this component is left for the interested reader. The observed histogram for the example shows a close-to normal distribution, allowing to estimate the error limit for daily mean temperature fluctuations from the standard deviation, σ , obtained from model identification.

6.3. Discussion of the Complete Temperature Model for Daily Air Temperature Variation

The complete temperature model is given in (21) and (22). The model predicts the daily average temperature variation with time as well as the expected the maximum and minimum temperatures due to stochastic process components. The comparison between measured data and model prediction with $\pm 3\sigma$ amplitude around the $T_R(t)$ function from the M5-type model in (22) is illustrated in **Figure 12(a)**. The graphs in the zoom-in **Figure 12(b)** and **Figure 12(c)** convincingly show that the measured values are almost always remain between the modeled maximum and minimum values.

7. Conclusions

- ◆ Analytical functional forms and their numerical algorithms are presented for representing the measured time-variable outside air temperature, $T_a(t)$ for engineering design and analysis of the human environment. The algorithms for $T_R(t)$ and T_S are easy to use for processing the available data sets, $T_a(t)$, at any physical location from the weather service, typically using several tens of thousands of measured values. In the final functional form of the outside air temperature function, $T_A(t)$, only a few dozens of constants are needed.
- ◆ The final analytical temperature, $T_A(t)$, can be used not only to represent the historical temperature data but also to predict the future temperature

variations at any given location from which the input data is used from measurements. The upper and lower boundaries may be used for safe temperature prediction.

- ♦ The regular component of temperature change with time, $T_r(t)$, in the M5-type model is described by a linear function plus a time-variable Fourier series to represent the long term linear change both in mean temperature and amplitude. Only 53 constants are needed, obtainable from the presented method, to represent the outside mean air temperature at any day of the year as long as need over decades of time.
- ♦ The confidence interval for the stochastic variation may be selected by the user via the multiplication factor of the standard deviation of the model match between measured, time-variable outside air temperature, $T_a(t)$ and the regular component in the analytical mode, $T_r(t)$.
- ♦ The stochastic component used in the final model, $T_A(t)$, is stable and stationary. The variability of the stochastic component over the season of the year may be considered in a future study, but presently is omitted for simplicity.
- ♦ The study shows that the prediction of temperature trends such as for cooling or warming in the future can only be evaluated using an M5-type model fit to the data. The trend-setting components, such as the annual change of the mean temperature or the variation of the amplitude change with time of the periodic components can only be evaluated with a model which has these components built into the structure of the model.
- ♦ The minimum, adequate time period for building an outside air temperature model is 4 years, the periodic cycle time of the solar environment. It is recommended to use a multiple of the 4-year periods for model-building (e.g., the $5 \times 4 = 20$ years period in present study) preferably for as long a time period as data are available.

Acknowledgements

A research grant from National Institute of Occupational Safety and Health (NIOSH) is gratefully recognized. The research was thankfully supported by the GINOP-2.3.2-15-2016-00010 “Development of enhanced engineering methods with the aim at utilization of subterranean energy resources” project of the Research Institute of Applied Earth Sciences of the University of Miskolc in the framework of the Széchenyi 2020 Plan, funded by the European Union, co-financed by the European Structural and Investment Funds.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Erinc, S. (1957) Applied Climatology and Climatic Conditions of Turkey. Istanbul

Technical University, Hydrogeology Institute, Istanbul.

- [2] Ustaoglu, B., Cigizoglu, H.K. and Karaca, M. (2008) Forecast of Daily Mean, Maximum and Minimum Temperature Time Series by Three Artificial Neural Network Methods. *Meteorological Applications*, **15**, 431-445. <https://doi.org/10.1002/met.83>
- [3] Bhutiyani, M., Kale, V.S. and Pawar, N.J. (2007) Long-Term Trends in maximum, Minimum and Mean Annual Air Temperatures across the Northwestern Himalaya during the Twentieth Century. *Climatic Change*, **85**, 159-177. <https://doi.org/10.1007/s10584-006-9196-1>
- [4] Wodon, Q., Liverani, A. and Joseph, G. (2014) Climate Change and Migration, Evidence from the Middle East and North Africa. World Bank, Washington, DC.
- [5] Bilbao, J., Miguel, A. and Kambezidis, H. (2002) Air Temperature Model Evaluation in the North Mediterranean Belt Area. *Journal of Applied Meteorology*, **41**, 872-884. [https://doi.org/10.1175/1520-0450\(2002\)041<0872:ATMEIT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2002)041<0872:ATMEIT>2.0.CO;2)
- [6] Aguiar, R. (1997) Climatic Synthetic Series for the Mediterranean Belt. Final CLIMED Project Report.
- [7] Komzsik, L. (2017) Approximation Techniques for Engineers. 2nd Edition, CRC Press, Boca Raton. <https://doi.org/10.1201/9781315205007>
- [8] Zapranis, A. and Alexandridis, A. (2007) Weather Derivatives Pricing: Modeling the Seasonal Residual Variance of an Ornstein-Uhlenbeck Temperature Process with Neural Networks. *Neurocomputing*, **73**, 37-48. <https://doi.org/10.1016/j.neucom.2009.01.018>