Titre: Title:	A diffusion process as a model for temperature variations
Auteurs: Authors:	Mario Lefebvre
Date:	2025
Type:	Article de revue / Article
Référence: Citation:	Lefebvre, M. (2025). A diffusion process as a model for temperature variations. Atti della Accademia Peloritana dei Pericolanti - Classe di Scienze Fisiche, Matematiche e Naturali, 103(S1), A8 (8 pages). https://doi.org/10.1478/aapp.103s1a8

DOI: 10.1478/AAPP.103S1A8

AAPP | Atti della Accademia Peloritana dei Pericolanti
Classe di Scienze Fisiche, Matematiche e Naturali
ISSN 1825-1242

Vol. 103, No. S1, A8 (2025)

A DIFFUSION PROCESS AS A MODEL FOR TEMPERATURE VARIATIONS

MARIO LEFEBVRE *

ABSTRACT. A geometric Brownian motion is proposed as a model for the variations of temperature from day to day. For short-term forecasts, the parameters of the geometric Brownian motion are estimated from the two most recent observations. Moreover, the same type of process is used to compute the risk of extreme heat during the summer. The model enables us to assess the potential effects of climate change on this risk. An application to Messina is presented.

1. Introduction

Let $\{W(t), t \geq 0\}$ be a Wiener process with infinitesimal parameters μ and σ^2. A geometric Brownian motion $\{X(t), t \geq 0\}$ is defined by

$$X(t) = e^{W(t)} \quad \text{for} \quad t \geq 0. \tag{1}$$

It follows that $X(t) > 0$ for any $t \geq 0$. For this reason, geometric Brownian motions are very important in financial mathematics, where they are used to model the variations of stock prices; (Ross 2011, see);

Since $W(t)$ has a Gaussian distribution with mean $W(0) + \mu t$ and variance $\sigma^2 t$ (see, for instance, Lefebvre 2007, p. 184); it follows that $X(t)$ has a lognormal distribution. Its probability density function is given by

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}\, x} \exp\left\{-\frac{[\ln(x/x_0) - \mu t]^2}{2\sigma^2 t}\right\} \quad \text{for} \quad x > 0, \tag{2}$$

where $x_0 = X(0) > 0$.

In this paper, we will try to use a geometric Brownian motion to model the variations of temperature from day to day. Because temperature has a periodic behaviour, it is not realistic to assume that the parameters μ and σ are fixed. In Section 2, for short-term forecasts, these parameters will be estimated from the two most recent observations. Then, in Section 3, we will use the model to compute the risk of extreme heat during the summer months. For this time period, the values of μ and σ will be assumed to be constant. Finally, our model will enable us to assess the potential effects of climate change on the risk of extreme heat.

To estimate the parameters μ and σ, the observed temperatures in the city of Messina in 2019 will be used. During that year, the minimum recorded temperature was 4 degrees Celsius. Therefore, a geometric Brownian motion could indeed be an appropriate model for the variations of temperature from day to day. In fact, if the minimum observed temperature, $x_{\min}$, had been negative, we could simply have defined $X^*(t) = X(t) - x_{\min} + 1$ to obtain a positive diffusion process $\{X^*(t), t \geq 0\}$. Actually, this definition could be used to study the temperature variations even if $x_{\min} > 0$.

2. Short-term temperature forecasts

In this section, although the aim of this paper is not to use a diffusion process to produce very precise short-term temperature forecasts, we want to show that a geometric Brownian motion can serve as a model for temperature variations. To do so, we will try to forecast the average temperature one day in advance, based on the observed temperatures up to the current day. To illustrate our technique, the observed temperatures for the city of Messina during the year 2019 will be used. These observed temperatures are available at the address https://rp5.ru/Weather_archive_in_Messina,_METAR. The temperatures are recorded every hour, from 0:55 to 23:55.

Remark. Unfortunately, the temperatures are recorded as integers. Therefore, they are either truncated or rounded, which can distort the statistical analysis that will be made.

Our first task was to compute the average temperature on each day, from the hourly observations; there were on most days 24 observations, but also some missing data. The average temperature from January 1st to December 30th was equal to 19.096 degrees Celsius, with a standard deviation of 6.089 degrees. If we can indeed model the variations of daily temperatures as a geometric Brownian motion, then

$$Y(t) := \ln[X(t)] - \ln[X(t-1)] = W(t) - W(t-1) \quad \text{for} \quad t = 2, \ldots, 364 \tag{3}$$

should be independent observations of a Gaussian $\mathrm{N}(\mu, \sigma^2)$ distribution. This follows from the fact that the Wiener process $\{W(t), t \geq 0\}$ has independent and stationary increments. It is therefore a simple matter to estimate the parameters μ and σ^2 from the data. However, as mentioned above, it is not realistic to assume that μ and σ^2 are fixed.

To forecast the value of the average temperature on day t, denoted by $\widehat{T}(t)$, we used the average temperatures on days $t-1$ and $t-2$ to estimate μ and σ^2. Thus, we computed moving averages and variances:

$$\bar{x}(t-1) := \hat{\mu}(t-1) := \frac{Y(t-1) + Y(t-2)}{2} \tag{4}$$

and

$$s^2(t-1) := \hat{\sigma}^2(t-1) = Y^2(t-1) + Y^2(t-2) - 2\,[\bar{x}(t-1)]^2 \tag{5}$$

for $t = 3, \ldots, 364$.

Remark. We could of course use more than the two most recent days to estimate μ and σ^2. If we denote by $\bar{x}_k(t-1)$ and $s_k^2(t-1)$ the estimators based on the k most recent days, we would have

$$\bar{x}_k(t-1) := \frac{\sum_{j=1}^{k} Y(t-j)}{k} \tag{6}$$

and

$$s_k^2(t-1) := \frac{\sum_{j=1}^{k} Y^2(t-j) - k, [\bar{x}_k(t-1)]^2}{k-1} \tag{7}$$

for $t = 3, \ldots, 364$ and $k = 2, 3, \ldots$

Next, the forecasted average temperature on day t is given by the formula (see, Lefebvre 2007, p. 188);

$$\widehat{T}(t) = T(t-1) \exp\left\{\bar{x}(t-1) + \frac{1}{2} s^2(t-1)\right\}, \tag{8}$$

where $T(t-1)$ is the observed average temperature on day $t-1$.

Remark. The general formula for the forecasted average temperature r day(s) after $t-1$, based on the two most recent days, is

$$\widehat{T}_r(t) := T(t-1) \exp\left\{\left[\bar{x}(t-1) + \frac{1}{2} s^2(t-1)\right] r\right\}, \tag{9}$$

for $r = 1, 2, \ldots$ In theory, we can forecast the mean temperature for any period of time, for example for the afternoon hours of day $t + r$. However, the estimates of the parameters μ and σ will then have to be calculated considering the same period of time on the previous days.

In Figure 1, we present the observed average daily temperatures in Messina during the year 2019, together with the forecasts obtained from the geometric Brownian motion. We see that the forecasts follow quite closely the observations.

Moreover, the least squares regression line of $T(t)$ on $\widehat{T}(t)$ is

$$T(t) = 0.7097 + 0.9587\,\widehat{T}(t). \tag{10}$$

The coefficient of determination, R^2, is equal to 94.9%. Remember that R^2 gives here the proportion of the variation in the dependent variable $T(t)$ that can be explained by the independent variable $\widehat{T}(t)$. Thus, we may conclude that the proposed model for the temperature variations is indeed realistic.

Remarks. (i) If we use the three most recent days to estimate μ and σ^2, we find that

$$T(t) = 0.5735 + 0.9663\,\widehat{T}_3(t) \tag{11}$$

and that the coefficient of determination is $R^2 = 95.2\%$. Therefore, the results are slightly better.

(ii) Moreover, the regression line for the forecasted average temperature two days after $t-1$, based on the two most recent days, is

$$T(t+1) = -0.2823 + 0.9646\,\widehat{T}(t) \tag{12}$$

and $R^2 = 93.9\%$. Hence, we can state that the technique works well at least up to two days in advance. For the forecasted value of $T(t+3)$, we obtain that $R^2 = 62.7\%$, which is still reasonable.

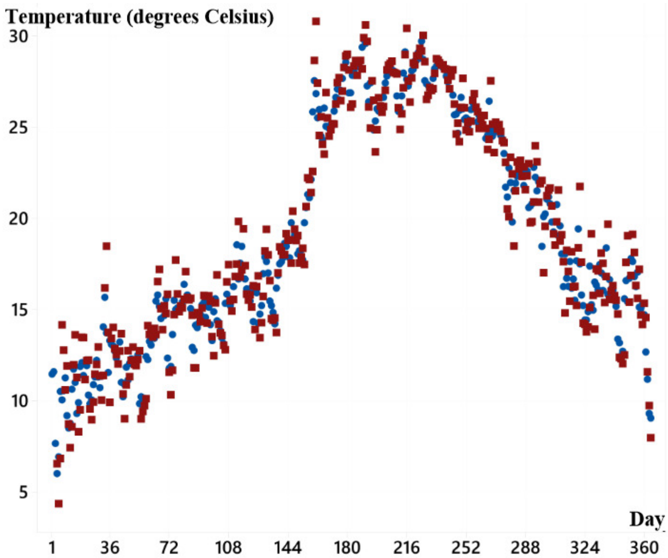

FIGURE 1. Observed (circles) and forecasted (squares) average daily temperatures in Messina during the year 2019.

In the next section, we will compute the risk of extreme heat during the summer in Messina, based on our model. We will also see the potential effects of climate change on this risk.

3. Computation of extreme heat risk

Let $M(t)$ denote the maximum observed temperature on day t. As in the previous section, we assume that $M(t) = e^{B(t)}$, where $\{B(t), t \geq 0\}$ is a Brownian motion process with infinitesimal parameters μ_M and σ_M^2, so that $\{M(t), t \geq 0\}$ is a geometric Brownian motion.

During the summer period, defined here to be from June 1st to September 30th, the parameters μ_M and σ_M^2 should be more or less stable. This is confirmed by Figure 2 in which the daily increments of the logarithms of the observed maximum temperatures are presented.

The point estimates of μ_M and σ_M are $\hat{\mu}_M = 0.00206$ and $\hat{\sigma}_M = 0.04735$.

If the daily increments indeed evolve like a geometric Brownian motion, then they should have a Gaussian distribution with parameters $\hat{\mu}_M$ and $\hat{\sigma}_M^2$. Graphically, we see in Figure 3 that the histogram of the daily increments is much too concentrated around the mean. It follows that a normality test performed with the statistical software program *Minitab* clearly rejects the normality assumption. However, one must remember that the temperatures were recorded as integers, which explains why the frequency of the interval around zero is much too large.

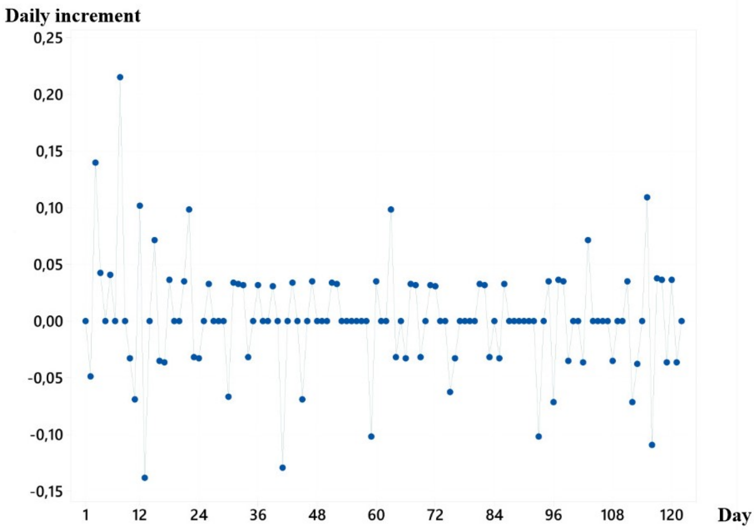

FIGURE 2. Daily increments of the logarithms of the observed maximum temperatures from June 1st to September 30th, 2019.

Let

$$p(s,y;x) := P\left[\max_{0<t\leq s} M(t) \geq y \mid M(0) = x\right], \tag{13}$$

where $x < y$. That is, $p(s,y;x)$ denotes the probability that the observed maximum temperature will be at least equal to y in the time interval $(0,s]$, where s is in days, given that it is equal to $x < y$ at the initial time. We can write that

$$p(s,y;x) = \rho(s,\ln(y);\ln(x)) := P\left[\max_{0<t\leq s} B(t) \geq \ln(y) \mid B(0) = \ln(x)\right], \tag{14}$$

with $\ln(x) < \ln(y)$. Moreover, we have

$$P\left[\max_{0<t\leq s} B(t) \geq \ln(y) \mid B(0) = \ln(x)\right] = P[\tau(\ln(x),\ln(y)) \leq s], \tag{15}$$

where $\tau(\ln(x),\ln(y))$ is the *first-passage time* defined by

$$\tau(\ln(x),\ln(y)) = \inf\{t > 0 : B(t) \geq \ln(y) \mid B(0) = \ln(x)\}. \tag{16}$$

The function $\rho(s,\ln(y);\ln(x))$ can be computed explicitly (Ross 2014, see);

$$\rho(s,\ln(y);\ln(x)) = \int_0^s \frac{\ln(y)-\ln(x)}{\sqrt{2\pi}\sigma_M t^{3/2}} \exp\left\{-\frac{(\ln(y)-\ln(x)-\mu_M t)^2}{2\sigma_M^2 t}\right\} dt. \tag{17}$$

In Table 1, we give the value of $p(s,y;x) = \rho(s,\ln(y);\ln(x))$ for various values of s, x and y.

Thus, if we assume that the observed maximum temperature on June 30th was equal to 30 degrees Celsius, the probability that a maximum of at least 40 degrees will be observed within the next five days is almost negligible. However, the probability that this event will

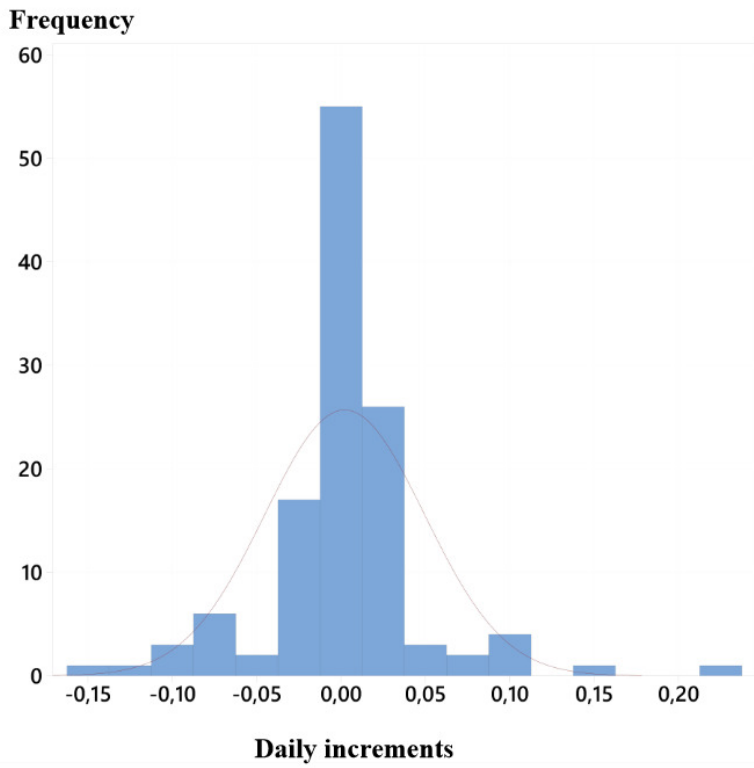

FIGURE 3. Histogram of the daily increments of the logarithms of the observed maximum temperatures from June 1st to September 30th, 2019.

TABLE 1. Probability $p(s,y;x)$ for various values of s, x and y

$p(s,y;x)$	0.0085	0.3527	0.2336	0.6859
s	5	31	5	31
x	30	30	35	35
y	40	40	40	40

occur within the month of July is equal to more than 1/3. If the maximum on June 30th was rather equal to 35 degrees, then the previous probabilities jump respectively to almost 1/4 and to more than 2/3.

Finally, to see the potential effects of climate change on the temperature in Messina, we computed the probabilities given in Table 1 in the case when the parameters μ_M and σ_M are both increased first by 5% and then by 10%. The results are presented respectively in Table 2 and Table 3. Although still small, the probability $p(5,40;30)$ is more than doubled when both μ_M and σ_M are increased by 10%, whereas $p(31,40;30)$ soars by almost 21%.

TABLE 2. Probability $p(s,y;x)$ for various values of s, x and y when both μ_M and σ_M are increased by 5%

$p(s,y;x)$	0.0124	0.3782	0.2575	0.7011
s	5	31	5	31
x	30	30	35	35
y	40	40	40	40

TABLE 3. Probability $p(s,y;x)$ for various values of s, x and y when both μ_M and σ_M are increased by 10%

$p(s,y;x)$	0.0173	0.4065	0.2820	0.7184
s	5	31	5	31
x	30	30	35	35
y	40	40	40	40

TABLE 4. Probability $p(s,y;x)$ for various values of s, x and y when only μ_M is increased by 10%

$p(s,y;x)$	0.0088	0.3609	0.2384	0.6930
s	5	31	5	31
x	30	30	35	35
y	40	40	40	40

TABLE 5. Probability $p(s,y;x)$ for various values of s, x and y when only σ_M is increased by 10%

$p(s,y;x)$	0.0168	0.3946	0.2778	0.7089
s	5	31	5	31
x	30	30	35	35
y	40	40	40	40

The results when only μ_M (respectively σ_M) is increased by 10% are given in Table 4 (respectively Table 5). We can see that it is the parameter σ_M that has the greatest influence on the various probabilities. The same holds true when only μ_M or σ_M is increased by 5%.

4. Conclusion

In this note, we saw that, at least for the city of Messina, daily temperature variations could be modelled as a geometric Brownian motion with time-varying infinitesimal parameters μ and σ. Then, in Section 3, the model was used to first calculate the risk of extreme heat during the summer months in Messina. Finally, this risk was also evaluated when the parameters of the geometric Brownian motion are increased by 5 and 10 percent.

For a short enough time period, the parameters μ and σ can be considered as constants. As a follow-up to this work, we could add jumps according to a Poisson process to the model to explain the potential heat or cold waves.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada. The author also wishes to thank the anonymous reviewers of this paper for their constructive comments.

References

Lefebvre, M. (2007). *Applied Stochastic Processes*. New York, NY: Springer. DOI: 10.1007/978-0-387-48976-6.

Ross, S. M. (2011). *An Elementary Introduction to Mathematical Finance*. 3rd ed. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9780511921483.

Ross, S. M. (2014). *Introduction to Probability Models*. 11th ed. San Diego, CA: Academic Press. DOI: 10.1016/B978-0-12-407948-9.00012-8.

---

* Polytechnique Montréal,
Department of Mathematics and Industrial Engineering,
C.P. 6079, Succursale Centre-ville, Montréal, Canada H3C 3A7

Email: mlefebvre@polymtl.ca

Paper contributed to the international conference on
“Atmospheric Monitoring, Modeling and Simulation”, held in Messina, Italy (2-3 December 2019)
under the patronage of the *Accademia Peloritana dei Pericolanti*

Manuscript received 21 July 2022; published online 1 October 2025