Isaac Newton and the astronomical refraction

Waldemar H. Lehn

Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg R3T 5V6, Canada

*lehn@ee.umanitoba.ca

Received 14 April 2008; revised 4 July 2008; accepted 4 July 2008;
posted 7 July 2008 (Doc. ID 94958); published 8 September 2008

In a short interval toward the end of 1694, Isaac Newton developed two mathematical models for the theory of the astronomical refraction and calculated two refraction tables, but did not publish his theory. Much effort has been expended, starting with Biot in 1836, in the attempt to identify the methods and equations that Newton used. In contrast to previous work, a closed form solution is identified for the refraction integral that reproduces the table for his first model (in which density decays linearly with elevation). The parameters of his second model, which includes the exponential variation of pressure in an isothermal atmosphere, have also been identified by reproducing his results. The implication is clear that in each case Newton had derived exactly the correct equations for the astronomical refraction; furthermore, he was the first to do so. © 2008 Optical Society of America

OCIS codes: 000.2850, 010.1290, 350.1260.

1. Astronomical Refraction before Newton

In the Western world, awareness of astronomical refraction extends back to the 4th Century BC [1]. However, the first person whose surviving work contains a clear atmospheric model was Ptolemy, a Greek astronomer who worked in Alexandria in the 2nd Century AD. His careful observations indicated that the apparent altitude of a star was slightly higher than its true one. His experiments on refraction in water and glass led him to conclude that air also refracted light and that the displacements of the stars were caused by this refraction. Accordingly, he visualized that the atmosphere consisted of a uniform spherical shell of air surrounding and concentric with the Earth (see Fig. 1). Outside of this shell was the ether that carried the stars in their orbits. Because the ether was far less dense than air, the refraction that would take place as a ray of light crossed the air–ether interface would have the effect of raising the apparent altitude of a star. Although Ptolemy did not try to measure the refraction, his model produced qualitatively correct results. This model endured for 15 centuries.

0003-6935/08/340H95-11$15.00/0

Astronomical measurements precise enough for the development of theories were first made by Tycho Brahe in the 16th Century. As his naked-eye instruments could achieve a resolution of 1 arc min, correction for astronomical refraction could no longer be avoided. Using his own careful measurements, Tycho became the first astronomer to construct a table of refractions [2], which he published in 1596 [3]. He did not, however, try to find a mathematical model to fit his data. This was done by his successor, Johannes Kepler, who developed a mathematical theory of refraction based on the Ptolemaic atmospheric model. In spite of much effort, the law of refraction that he deduced, while reasonably accurate, was not the correct one. Nevertheless, the table that he calculated holds up surprisingly well when compared with modern tables, at least for zenith distances less than 70°. Kepler was the first to calculate a table of refractions; he published it in 1604 [4].

Probably the last application of the Ptolemaic model was made by Gian Domenico Cassini in 1662 [5]. He applied the recently discovered Snell's law of refraction to the air–ether interface and adjusted his parameters to produce quite a good refraction table [6]. Its predictions were correct to within a few arcseconds [7] for zenith distances up to 70°. By this time the Ptolemaic model was just a convenient

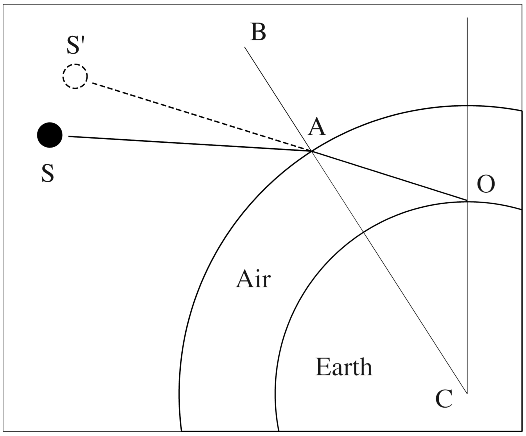

Fig. 1. Ptolemy's model of astronomical refraction. Light from a star S suffers a single refraction at A before it reaches the observer at O. The star is perceived at the higher position S', which lies along the straight backprojection of AO.

way of calculating; astronomers were already well aware that the true path of a light ray through the atmosphere was a curve.

In 1681, John Flamsteed, the Astronomer Royal of England, made exhaustive measurements of the astronomical refraction, to an unprecedented precision of 1 arc sec. He recognized the need for a theoretical model, but, being an experimentalist, he was not capable of developing one. Finally, in 1694, he wrote to Isaac Newton to request of him the creation of a mathematical model that would permit calculation of refraction in a consistent, theoretical way that would provide the mean corrections for all observations [8].

2. What Was Known about the Air in 1694?

In Newton's day, instruments for the precise measurement of atmospheric parameters were of relatively recent invention.

Measurement of air pressure was motivated by the study of the vacuum [9]. Experimentalists discovered that if a long tube filled with water was inverted into a pool of the same liquid, only a limited height (more or less always the same) of water remained in the tube. The space above the liquid was filled by the mysterious vacuum. It was soon realized that the weight of the air was supporting the column of liquid and that this weight was not constant. In 1644 in Florence, Torricelli invented the mercury barometer, a column of mercury with a calibrated scale. Atmospheric pressure could now be expressed as the length of this column. This invention reduced the size of the instrument to practical laboratory dimensions, for the height of the water column would be about 10 m, while the mercury column would be only 76 cm.

With this new instrument, experiments were soon undertaken to determine the nature of the atmosphere. The first experiment to see whether pressure varied with elevation was done in France in 1648. A qualitative result was observed, but as the true elevation difference was not known, a numerical result could not be obtained. One concrete fact was, however, accepted: that far enough away from the Earth, the pressure of the air would vanish. The question of the distribution of pressure with elevation remained. Robert Hooke was the first to solve this problem. He applied the law published by Robert Boyle in 1662, which stated an inverse proportion between the volume and the pressure of a gas. His discrete model considered a vertical column of air to be divided into 1000 parts, each having the same weight. Although Hooke was aware that gravity changed with elevation, he did not incorporate this into his model. Nor did he include any variations in temperature, because the laws relating temperature to pressure and volume had not yet been discovered. For this isothermal model, he obtained a correct result, which he published in 1665 [10]. In passing, he mentioned that one could allow the number of layers to become very large. Had he actually done this, and passed to the limit of an infinite number of layers, he would have obtained the exponential pressure variation that correctly represents the isothermal atmosphere. But calculus had not yet been discovered.

The other important parameter was temperature. This quantity was measured by the spirit-in-glass thermometer, also known as the Florentine thermometer [11]. It was invented in Florence about 1650 and reached England in 1661. Great effort was spent on giving the inner diameter of the tube a constant value, in order that the column length of the working fluid be proportional to its volume. The device was calibrated in linear fashion, usually by means of two reproducible fixed points. However, in 1665 Robert Hooke proposed the use of the ice point and a reproducible degree size [12], a system that the Royal Society of London adopted for some decades. In 1680 the ice point and boiling point of water were proposed, but not put into universal practice until later. The quantities that this instrument measured were called "degrees of heat," assumed to be linearly related to the length of the fluid column. A further physical interpretation was not attempted.

With atmospheric temperature variation, however, no progress was made. No one had any idea of the lapse rate of temperature with elevation. Even had they known it, they did not possess the physical laws that would enable them to incorporate the idea. Nevertheless, Newton and others did know that heat reduced the density of air.

3. Newton's Conception of Optics in 1694

Snell's law of refraction was universally known, although the term "index of refraction" was not yet in use. Instead, scientists always referred to the "ratio of the sines" of incidence and refraction. This ratio was known to be a constant, independent of angle, when light moved from one medium to another [13].

Newton thought of light as a stream of small particles. In passing from one medium to another of different density, the particles would be attracted to the denser one by a refractive force normal to the interface. This force acted only in the immediate vicinity of the interface; its effective range was zero. Newton considered that the refractive force would instantaneously modify the square of a light particle's velocity by a fixed amount. In modern terms, the force would do a fixed amount of work, and modify the kinetic energy of any light particle by the same amount [14]. If the motion were into the denser medium, the change would be an increase. The component of velocity parallel to the interface would not be affected; the increase would appear only in the perpendicular component. In his *Principia* (1687), Newton proved that Snell's law followed from this entirely mechanical model. Further, he showed that the velocity of the light particles in the denser medium was higher, according to the ratio of sines [15].

Newton also tried to show, from experimental evidence, that the work done by the refractive force was proportional to the density. Although he had already studied this before 1694, he did not publish the conjecture until 1704 [16]. In any case, his model produced correct results for the atmosphere.

4. Newton's Approach to Astronomical Refraction

In 1836, just a year after Newton's correspondence with Flamsteed first appeared in print, Jean-Baptiste Biot was the first to publish (in great detail) his interpretation of Newton's mathematical analysis [17]. As will be done here, he endeavored to put equations and methods to Newton's words. In Biot's day, the effect of the refractive force would be written as $v^2 = c2 + \text{refractive power}$, where v is the speed of light within a medium, c is the speed of light in free space, and refractive power (in rather loose terminology) is proportional to the work done by the refractive force. The equation would be normalized with respect to c and rearranged to read as $v^2 - 1 = \text{refractive power}$. Since v differs from 1 by only a very small amount, the left-hand side is well represented by $2(v - 1)$, and thus $v - 1$ is proportional to the refractive power. In modern terms, we must consider Newton's v as proportional to our refractive index n, which leads one to conclude that the refractive power, as used by Biot and Newton, is proportional to what we today call the refractivity, $n - 1$.

Following his predecessors, Newton assumed that the atmosphere possessed spherical symmetry, concentric with the Earth. All points of equal density would have the same elevation. Thus the atmosphere could be imagined as a sequence of thin spherical shells, each having a fixed density. Refractive power was assumed to be proportional to density, a result that Hauksbee later demonstrated in 1708 [18]. Thus, $v^2 - 1$ would be proportional to density.

Because the refractive force was normal to the spherical shells, it would always point toward the center of the Earth. Hence it would be a central force, under whose influence the light particles would follow orbits that could be calculated from his theory of planetary motion. In particular, motion under any central force would conserve the angular momentum of the light particles. In the notation of Fig. 2, he would write $vr \sin\theta = K$, a constant. He named this quantity Kepler's constant, because it arose from Kepler's law of areas. Since in Newton's theory, $v = nc$, where n is our modern index of refraction and c is the speed of light in space, we could rewrite his equation as $nr \sin\theta = \text{constant}$. This is the well known ray invariant for light paths under spherical symmetry. The ray invariant thus follows directly from Newtonian mechanics.

5. Newton's First Model for Astronomical Refraction

On 17 November 1694 Newton sent Flamsteed a letter containing a new refraction table [19]. It was presented in three columns giving the expected refraction for Summer, Spring–Fall, and Winter. Newton had apparently used only two of Flamsteed's observations, all he needed to set the parameters of his model: these were the "horizontal refraction" (as the refraction for zero altitude was conventionally called) and that for the altitude of 3°. In this letter, Newton said absolutely nothing about his method of calculation.

After some prodding by Flamsteed, Newton finally sent a sketch (see Fig. 3) and a description of his mathematical model, dated 20 December 1694 [20]:

"Let AKL represent the globe of the earth, & suppose this globe is covered wth an Atmosphere of Air whose density decreases uniformly from ye earth upwards to the top wch is here represented by the circle MON. And let SO be a ray of light falling on ye top of this Atmosphere at O & in its passage from thence through ye Atmosphere to the spectator at A, continually refracted & bent in ye curve line OBA. From any point of this curve line B to ye center of the earth draw the right line BC cutting the surface of the earth in D & take CF a mean proportional between CB & CD & let AFG be ye Locus of the point F, that is

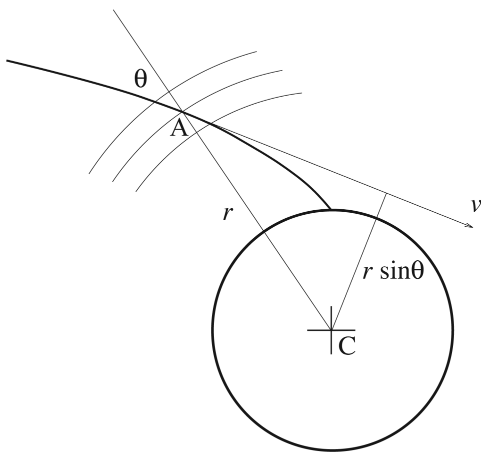

Fig. 2. Conservation of angular momentum. At any point A on the ray, r is the distance to the Earth's center, θ is the angle between the ray and the local vertical, and v is the velocity of the light particle. Its angular momentum relative to the center of force is $vr \sin\theta$.

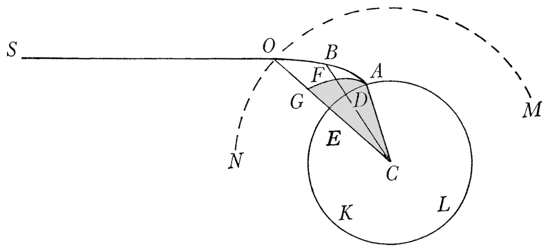

Fig. 3. Newton's diagram showing how to find the refraction; taken from [8], p. 61 (shading added).

the curve line in wch ye point F will be allways found: & if this curve line AFG cut the right line OC in G; the whole refraction of ye ray in passing from O to A will be proportional to the area AFGC & the refractions in passing through any part of that line OB or BA will be proportional to the corresponding part of the area GFCG or FACF. This Theorem is Geometrically demonstrable but the demonstration is too intricate to be set down in a Letter."

Flamsteed naturally wanted to know how one calculated the relevant areas. In his letter of 15 January 1694–1695 [21], Newton told him how: "The areas in that Theoreme I sent you are to be determined by the 5t Lemma of my Third Book of *Principia Math*. But ye calculation is intricate."

The lemma to which he referred explained how one could find the areas under curves by approximating them with parabolic arcs, the areas under which were well understood. This would indeed be very accurate for the smooth curves in question, although it would require an enormous effort. This point will be elaborated later.

In Newton's sketch, Fig. 3, shading has been added to highlight the area to which the refraction is supposed to be proportional. The distance CF to the curve is the geometric mean ("mean proportional") between the Earth's radius CD (R_E) and the distance to the ray CB (r). To facilitate the understanding of this model, one can express it in terms of modern calculus. Let f be the refraction, φ the polar angle at C, and l the distance CF. Then the refraction is to be proportional to the shaded area, or, in terms of infinitesimals, $\mathrm{d}f \propto l^2 \mathrm{d}\varphi/2$. Since $l = \sqrt{R_E r(\varphi)}$, one has $\mathrm{d}f \propto \frac{1}{2} R_E r(\varphi) \mathrm{d}\varphi$.

To carry out this integration, the ray path equation $r(\varphi)$ is required. This is directly available from the ray invariant $vr\sin\theta = K$, where one keeps in mind that Newton's light speed v is proportional to the modern refractive index n. From Fig. 4, one can see that $\mathrm{d}r/r\mathrm{d}\varphi = \cot\theta$. The ray invariant permits the elimination of θ through

$$\cot\theta = \frac{\sqrt{1-\sin^2\theta}}{\sin\theta} = \frac{\sqrt{1-K^2/v^2r^2}}{K/vr};$$

then the differential equation of the path is obtained:

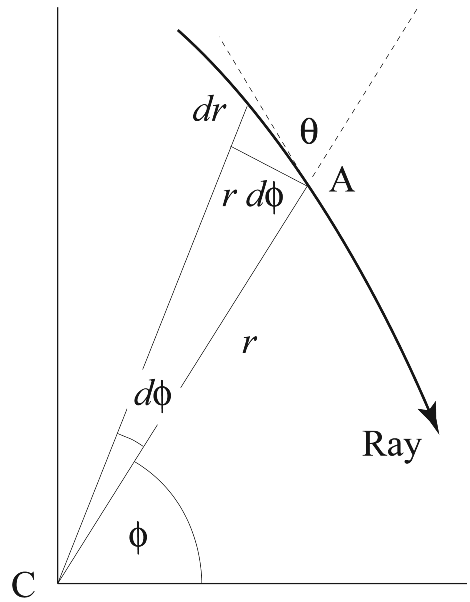

Fig. 4. Ray path geometry.

$$\mathrm{d}\varphi = \frac{1}{\sqrt{\frac{v^2r^2}{K^2}-1}}\frac{\mathrm{d}r}{r}. \tag{1}$$

Substitution of this into $\mathrm{d}f$ gives the following equation as the equivalent of Newton's model:

$$\mathrm{d}f \propto \frac{1}{\sqrt{\frac{v^2r^2}{K^2}-1}}\mathrm{d}r. \tag{2}$$

The modern form of the differential equation of refraction is [22]

$$\mathrm{d}f = \frac{-\frac{1}{n}\frac{\mathrm{d}n}{\mathrm{d}r}}{\sqrt{\frac{n^2r^2}{K^2}-1}}\mathrm{d}r, \tag{3}$$

where K is the constant of the ray invariant (Kepler's constant). Since $n-1$ is proportional to density, Newton's assumption of a constant density gradient makes $\mathrm{d}n/\mathrm{d}r$ constant. Then if the numerator term $1/n$ is replaced by unity (introducing an error of less than 3 parts in 10000) one obtains

$$\mathrm{d}f \propto \frac{1}{\sqrt{\frac{n^2r^2}{K^2}-1}}\mathrm{d}r, \tag{4}$$

exactly Newton's result. Clearly Newton fully understood the mathematics of astronomical refraction; and he was the first to do so [23].

To determine his constants of proportionality, Newton had two parameters to adjust: the depth of the atmosphere, h, and the refractive index at the surface, n_0 (here he would have used v_0, the velocity of light at the surface). For this he needed two data points. He selected the refraction for the apparent altitudes of 0° and 3°. Flamsteed's extract of 11 October 1694 [24] gave readings of 33′30″ and 14′00″, respectively. For some reason, Newton did not use exactly these values; his table implies that he settled on 33′20″ and 13′40″ to build the equinoctial table.

Here it was attempted to reproduce this table by using modern methods, starting with Eq. (3). Let $A = nr/K$ and $r = R_E + z$, where z is elevation above the Earth's surface. The linear density profile is $\rho = \rho_0(1 - z/h)$, where ρ_0 is the density at the surface. The refractive index n is related to density $n = 1 + \varepsilon\rho$, where ε has the numerical value of 226×10^{-6} in SI units. Now, as was done to derive Newton's result, the numerator term $1/n$ is replaced by 1. Then A may be written as

$$A = \frac{(1+\varepsilon\rho)\left(1+\frac{z}{R_E}\right)}{n_0 \sin\theta_1}, \tag{5}$$

where the ray invariant has been expressed in terms of surface parameters n_0, R_E, and the zenith angle of the ray at the observer, θ_1: $K = n_0 R_E \sin\theta_1$. Because both $\varepsilon\rho$ and z/R_E are much less than 1, the expansion of the numerator suffers little loss of accuracy when the product of these terms is neglected. Next replace ρ with its linear form, to obtain A as a linear function of z:

$$A = \frac{1 + Bz/n_0}{\sin\theta_1}, \tag{6}$$

where

$$B = \left(\frac{1}{R_E} - \frac{\varepsilon\rho_0}{h}\right).$$

The numerator of the refraction integral becomes a constant:

$$\frac{dn}{dz} = -\frac{\varepsilon\rho_0}{h},$$

and Eq. (3) becomes

$$f = \frac{\varepsilon\rho_0}{h}\int_0^h \frac{dz}{\sqrt{A^2-1}}, \tag{7}$$

which can be evaluated in closed form. The result is a logarithmic function:

$$f = \frac{\varepsilon\rho_0}{Bh}\sin\theta_1 \ln\left(\frac{1+Bh+\sqrt{(1+Bh)^2-\sin^2\theta_1}}{1+\sqrt{1-\sin^2\theta_1}}\right). \tag{8}$$

Note that only one first-order approximation for A was necessary to get to this point.

An additional first-order approximation leads to a simpler form. In the linear term of A, the coefficient of z is very small, so that a good approximation to A^2 is the linear function

$$A^2 = \frac{1+2Bz/n_0}{\sin^2\theta_1}.$$

The integral is now very easy, and the result is

$$f = \frac{\varepsilon\rho_0 \sin\theta_1}{Bh/n_0}\left(\sqrt{\cos^2\theta_1 + 2Bh/n_0} - \cos\theta_1\right). \tag{9}$$

Calculations made with this form agree exactly (to the arcsecond) with those made with the logarithmic form. Therefore this form will be used in the following work. It is worth noting that, for the atmosphere of linear density variation, the troublesome singularity that usually arises at $\theta_1 = 90°$ simply does not exist.

The refraction equation contains two parameters that must be set from experimental data. They are (keeping in mind that $n_0 = 1 + \varepsilon\rho_0$) the quantities $\varepsilon\rho_0$ and h. The first is the refractivity of air at the surface, and the second is the height of the finite atmosphere. Two observations suffice to determine them, as Newton informed Flamsteed [25].

Let the two reference observations be (θ_1, f_1) and (θ_2, f_2). When substituted into Eq. (9), they yield two conditions that must be solved simultaneously. If the two n_0 terms are taken equal to one exactly (the relative error in so doing being negligible, as previously stated), then one can explicitly solve for the unknown parameters. The solutions are

$$\varepsilon\rho_0 = \frac{f_1 f_2(f_2 \sin\theta_1 \cos\theta_1 - f_1 \sin\theta_2 \cos\theta_2)}{f_2^2 \sin^2\theta_1 - f_1^2 \sin^2\theta_2}, \tag{10}$$

$$\frac{h}{R_E} = Bh + \varepsilon\rho_0, \tag{11}$$

where

$$Bh = \frac{2(\varepsilon\rho_0)^2 \sin^2\theta_1 - 2f_1\varepsilon\rho_0 \sin\theta_1 \cos\theta_1}{f_1^2}. \tag{12}$$

Newton's table contained 3 columns of refractions: for summer, equinox, and winter. Consider the equinoctial table, and use the two reference observations that Newton says he used, namely, a refraction of 33′20″ at 90° zenith angle and one of 13′40″ at 87° zenith angle. Then one obtains $\varepsilon\rho_0 = 250.6 \times 10^{-6}$ and $h = 10105$ m. Now a table of refractions can be calculated, an extract of which is given in Table 1. The agreement at altitudes 0° and 3° is forced, but the remaining refractions deviate significantly from Newton's values. It seems clear that he could not have obtained his table in this way.

A remarkable thing happens if the first data point (the horizontal refraction) is excluded. If the adjustable parameters are recalculated by using, for

Table 1. Refractions Calculated on the Basis of Altitudes 0° and 3°

Apparent Elevation	Calculated Refraction	Newton's Table	Difference
0°	33′20″	33′20″	0″
1°	23′55″	23′12″	43″
2°	17′42″	17′29″	13″
3°	13′40″	13′40″	0″
4°	10′59″	11′04″	–5″
5°	9′06″	9′13″	–7″

example, altitudes 1° and 3°, one gets $\varepsilon\rho_0 = 256.63 \times 10^{-6}$ and $h = 11600\,\mathrm{m}$. Then a calculation of the refraction at *integer* values of apparent altitude from 1° on (see Table 2) produces an almost exact fit to Newton's table [26]. A slight optimization of the parameters, to $\varepsilon\rho_0 = 256.75 \times 10^{-6}$ and $h = 11620\,\mathrm{m}$, reduces the errors at the half-angles to two instances of 1 arc sec each. Such agreement, unlikely to happen by chance, suggests that he did not use the horizontal refraction as one of his reference points. Searching through Flamsteed's data, I have not been able to identify exactly which two reference points he did use, but these results strongly imply that Newton's mathematical model was very like the one described by Eq. (9) above.

6. Newton's First Refraction Table

A few remarks on Newton's first refraction table would be in order. The refractions of summer and winter are in the ratio of 8 : 9, and the equinoctial refraction is their exact arithmetic mean. Now, to anyone doing these calculations by hand, it would be reasonable to calculate only one column and then find the other two by proportions. Biot proposed that Newton did this, calculating the winter table first [27], then finding the others by proportion. But the better numerical agreement obtained by starting with the equinoctial table suggests that Newton may have calculated it first, after which he could multiply by 16/17 and 18/17, respectively, to get the summer and winter values.

Further, Biot, still adhering to the particle model of light, suggested that Newton used orbital mechanics to determine the path of a ray through the atmosphere. Biot considered the effect of the central refractive force, which would be constant in the case of the uniform density gradient. Because orbits for such a force were extremely difficult to calculate, he suggested that Newton would have replaced it by an inverse square law force, for which the orbits were known. He argued that this was valid because an inverse square law would vary only slightly over the shallow depth of the atmosphere, and thus would be a good approximation to the constant central force.

In spite of his conviction that Newton knew the differential equation for refraction, Biot accepted Newton's instructions on how to do the calculations. The business of repeated parabolic fits would be an enormous amount of work. Given the agreement of Newton's table with the mathematical model proposed here, it seems reasonable to believe that Newton himself had worked out a closed form solution for the refraction. His message on how to do the calculation was for Flamsteed's benefit alone; someone like Flamsteed, who knew no calculus, could do it in no other way.

Biot was unable to get a good agreement between his model and Newton's table. This is understandable because he accepted Newton's statement about using the horizontal refraction as a reference point. The foregoing analysis strongly suggests that Newton invented the first value in his table, in order that it would agree more closely with observations. After all, no one was capable of proving him wrong.

Suppose one assumes that Newton artificially selected the refraction for 0°, and that he calculated all of the remaining integer altitudes using a formula like Eq. (9). Rather than calculating the half-angle values directly, he could have used interpolation [28]. Newton's tables are indeed consistent with a hyperbolic interpolation. For example, such an interpolation based on the data for altitudes 1°, 2°, and 3° reduces the 1″ difference at 1°30′ to zero. A problem remains for altitude 0°30′, however. The closest fit to this point is provided by a parabolic interpolation based on 0°, 1°, 2°, which leaves a difference of only 2″; but in general parabolic interpolations work poorly on Newton's data. There is no particular reason to suppose that Newton used the parabolic form here, while using hyperbolic forms elsewhere. More likely he invented the refraction at 0°30″ as well.

On the other hand, interpolation saves little time if a closed form solution is available. It would be quite reasonable to assume that Newton calculated all of

Table 2. Portion of Newton's Tabula Refractionum[a]

Altitudo apparens ° ′	Refractio aestiva ′ ″	Refractio verna et autumnalis ′ ″	Refractio hyberna ′ ″	Calculated Values for Spring–Fall ′ ″	Difference ″
0.00	31.30	33.20	35.10	31.32	−108
0.30	26.06	27.45	29.24	27.00	−45
1.00	21.50	23.12	24.34	23.12	0
1.30	18.51	20. 2	21.13	20. 3	1
2.00	16.27	17.29	18.31	17.29	0
2.30	14.31	15.23	16.15	15.23	0
3.00	12.52	13.40	14.28	13.40	0
3.30	11.32	12.15	12.58	12.15	0
4.00	10.25	11. 4	11.43	11. 4	0
4.30	9.29	10. 5	10.41	10. 4	−1
5.00	8.40	9.13	9.46	9.13	0
6.00	7.24	7.52	8.20	7.52	0
7.00	6.27	6.51	7.15	6.51	0
8.00	5.42	6. 3	6.24	6. 3	0
9.00	5. 5	5.24	5.43	5.24	0
10.00	4.36	4.53	5.10	4.53	0
11.00	4.11	4.27	4.43	4.27	0
12.00	3.51	4. 5	4.19	4. 5	0
13.00	3.33	3.46	3.39	3.46	0
14.00	3.18	3.30	3.42	3.30	0
15.00	3. 4	3.16	3.28	3.15	−1

[a] From [8], p. 49. The calculated values are found from Eq. (7).

his tabulated values (except the first two) by using his closed form. The 1″ discrepancies at 1°30′ and 4°30′, possibly due to his round-off procedure, are too small to be significant. An alternate possibility is transcription error. According to Scott [29], the handwritten original of the table is in poor condition and hard to read. It is thus easy to imagine minor errors creeping in when the original was converted to print [30]. Clearly such a point could be resolved by examining the original table.

Although Newton did not publish his solution for the linear variation of density, the model appeared to take on a life of its own. Thus in 1743, Thomas Simpson produced an analysis of the same problem, unaware that Newton had already solved it and discarded the result [31]. Simpson expressed the refraction in a simple, elegant formula, whose two parameters were set to match the measured refraction for altitudes 0° and 30°. A few decades later, Tobias Mayer worked on the same problem, and in 1770 published a refraction formula that could be shown to follow from Simpson's. Finally, James Bradley recast the formula into a slightly different form, made small adjustments to the parameters, and in 1798 published a refraction table that came into wide and prolonged use. Both Mayer and Bradley included correction tables for the effects of local temperature and pressure. In spite of the fact that none of these efforts was based on a correct atmospheric model, the linear density model became the first practical replacement for Cassini's tables. An excellent synopsis of the 18th Century research is given by Bruhns [23].

7. Newton's Second Model for Astronomical Refraction

On reflection Newton recognized the flaw in his first model: that its constant density gradient would produce the same refraction at the top of the atmosphere as at the bottom. Further, he was still worried about not being able to match the 0° and 3° values (entirely understandable if one accepts that the 0° value was not actually calculated) [32]. So he started over again.

This time he used a more realistic density profile. He assumed that density was proportional to pressure, a relation well known to his contemporaries Robert Boyle and Robert Hooke. Then he could apply his analysis of the spherically symmetric static atmosphere in an inverse square law gravitational field, which he had previously published in his *Principia* [33]: "… if the distances [from the centre] be taken in harmonic progression, the densities of the fluid at those distances will be in a geometrical progression." (Interestingly, though Newton was familiar with Robert Hooke's earlier and much simpler proof, he neither used nor acknowledged it [10]).

Newton's proposition is exactly what emerges if one models the isothermal atmosphere in an inverse square law gravitational field. The relation between pressure and density is given by $\mathrm{d}p/\mathrm{d}r = -g\rho$, where $g(r)$ is the acceleration of gravity. This integrates to the law of pressure variation,

$$p = p_0 \exp(-R_E/H) \exp(R_E^2/rH), \qquad (13)$$

where p_0 is the pressure at the surface and H is the scale height of the atmosphere. If r is replaced by $R_E + z$, and z is considered to be much less than R_E, i.e., gravity is taken to be constant, this reduces to the familiar exponential variation of pressure and density with elevation [34]:

$$p = p_0 \exp(-z/H). \qquad (14)$$

The exponential model entered the public domain two decades after Newton completed his work. Brook Taylor was actually the first to publish it [35], and for some time he carried the credit for priority in its discovery. But because of the work begun by Simpson, nearly a century passed before the exponential model reemerged. Kramp's work on the subject (1799) was considered excellent [36], but it was soon superceded by Laplace's elegant analysis in his widely read *Mécanique Céleste* of 1805 [37]. In 1825, Thomas Young summarized the history of the exponential model [38], but because this predated the publication of Newton's correspondence, Young was unable to say much about Newton's contribution. As previously mentioned, Biot finally had access to Newton's letters in 1836. He closely followed the methods of Laplace, who had available to him all of the gas laws as well as a good estimate of the temperature lapse rate. Among modern authors, Whiteside provided an interpretation of how Newton solved both of his refraction models [39], and Whiteside discussed in detail the relevant correspondence between Newton and Flamsteed [40].

The modern form shows what kind of integral Newton now had to evaluate. Recall Eq. (3), here written in integral form in terms of z:

$$f = \int_0^\infty -\frac{1}{n}\frac{\mathrm{d}n}{\mathrm{d}z}\frac{1}{\sqrt{A^2-1}}\mathrm{d}z, \qquad (15)$$

where $A = [n(R_E + z)]/K$.

As before, Newton correctly took the refractive power $(n-1)$ as proportional to density, and density as proportional to pressure. For the constant-gravity approximation, the equation is $n - 1 \propto p_0 \exp(-z/H)$. Then the terms in the integrand become

$$\frac{\mathrm{d}n}{\mathrm{d}z} = -\frac{\varepsilon\rho_0}{H}\exp\left(-\frac{z}{H}\right), \qquad (16)$$

$$A = \frac{\left(1 + \varepsilon\rho_0 \exp(-z/H)\right)(1 + z/R_E)}{n_0 \sin\theta_1}. \qquad (17)$$

Again, two parameters must be set to match the observational data. They are $\varepsilon\rho_0$ (as before) and H (this time the scale height rather than the absolute height of the atmosphere). However, the refraction integral (15) is now impossible to evaluate in closed form.

From a few statements in Newton's correspondence, Biot concluded that he fitted segments of the integrand by parabolic arcs, the area under which was known. While extremely tedious, this should be quite accurate because the integrand is smooth and slowly varying. The one exception is a problem that arises at zero altitude. In this case the integrand becomes singular at the lower limit of $z = 0$. However the integral remains finite, and the calculation can be handled (see the Appendix).

In his analysis, Biot assumed that Newton's two reference points were 0° and 45°. Although Newton's correspondence provides no evidence that he actually used these points, one could follow this line of thought for a moment. For the higher altitudes, say 45° and up, the refraction can be well estimated by an approximation expressed in terms of zenith angle θ_1 (see the Appendix):

$$f \approx \varepsilon\rho_0 \tan\theta_1. \tag{18}$$

This equation is independent of the scale height H, and thus permits solution for $\varepsilon\rho_0$ based on a single measurement [41].

If Newton had done this, he would be left with a single parameter (H) to set. Given that H appears mostly within exponential functions, this process would be extremely laborious. Speculation on how he might have done it will not be attempted.

Biot identified $\varepsilon\rho_0 = 262.5068 \times 10^{-6}$ and $H = 8597.78$ m as his best fit to Newton's table. He carried out the numerical integration to the elevation at which the density in Newton's model would be 1% of the surface value (39, 594 m). Unfortunately, his calculations did not agree well with Newton's table.

Biot did not consider that his result might be sensitive to his chosen upper limit of integration and made no attempt to deduce the height at which Newton terminated his integration. However, when this third parameter (the upper limit h) was allowed to vary, an improved fit to Newton's table became possible. The optimum solution, which minimized the sum of the absolute errors, was $\varepsilon\rho_0 = 267.7 \times 10^{-6}$ and $H = 8725$ m, with a reduced integration height of 29, 200 m.

Table 3 summarizes the results of the calculations. The first two columns list every second entry in Newton's table, up to an altitude of 10°. Column 3 is the refraction that arises from the optimum three parameters, calculated by means of MATLAB numerical integration. The difference between this calculation and Newton's tabulation is given in column 4, in units of arcseconds. A calculation for Biot's values, based on the parameters that he identified as his best solution, appears in column 5; Biot himself did not give a tabulation. The differences in column 6 are the deviations between Biot's and Newton's values.

Table 3. Portion of Newton's Second Refraction Table[a]

Altitudo apparens ° ′	Refractio ′ ″	Calculated Values ′ ″	Diff. ″	Biot's values ′ ″	Diff. ″
0. 0	33.45	33.45	0	33.33	−12
0.30	27.35	27.35	0	27.25	−10
1. 0	23. 7	23. 5	−2	22.57	−10
1.30	19.46	19.41	−5	19.36	−10
2. 0	17. 8	17. 4	−4	17. 0	−8
3. 0	13.20	13.20	0	13.18	−2
4. 0	10.48	10.50	2	10.50	2
5. 0	9. 2	9. 5	3	9. 5	3
6. 0	7.45	7.47	2	7.48	3
7. 0	6.47	6.47	0	6.48	1
8. 0	6. 0	6. 1	1	6. 2	2
9. 0	5.22	5.23	1	5.24	2
10. 0	4.52	4.52	0	4.53	1

[a] From [8], p. 95.

The three parameter solution produces a distinct improvement over that of Biot. At the altitudes of 0° and 3°, which Newton considered to be important reference points, the agreement is exact. At the integer altitudes, the maximum discrepancy is 4″; most are 2″ or less.

Integer angles are again emphasized here because it seems unlikely that Newton would have calculated any integrals for the fractional angles. When a hand calculation of such complexity is discussed, it is essential to consider the possibility of interpolation [42]. Typically one would calculate only a few points, say at integer altitudes, and interpolate the intermediate points. To investigate this hypothesis, one should first see whether Newton's table is consistent with it. Three anchor points are necessary to fit a hyperbola of form $y = a + [b/(x - c)]$ to the data. If these points are chosen fairly close together, e.g., adjacent integers at the low altitudes, then almost any combination of anchor points produces an interpolation that matches Newton's table very well, with discrepancies not exceeding 1 arc sec. For example, with interpolations based on the set of altitudes [0, 1, 2], [2, 3, 4], [4, 5, 6], [6, 8, 10], and [10, 15, 20], he would have to do only 11 integrations to get the first 35 points in his table. For this selection, 8 of the interpolated points deviate only 1″ from his table, while the remaining 16 agree exactly. When the fact that Newton was doing numerical integration by hand calculation is taken into account, a discrepancy of 1″ must be considered to be entirely negligible. In other words, Newton's table is consistent with the interpolation hypothesis.

This postulated interpolation is the reason that the calculations presented here should be compared with the table only at integer altitudes (and perhaps not even all of those). In any case the agreement is so good that it is hard to imagine that Newton could have used an integral different from the one proposed in Eqs. (15)–(17).

It might be of interest to calculate the refraction by using numerical values available to Newton. In his

Opticks, Newton listed the refractive index of air as 3201/3200, and its specific gravity as 0.0012, without giving the temperature or pressure [43]. The standard pressure at the time was commonly taken as 30 in. of mercury. Temperature will be ignored, as not being relevant to Newton's model. From these one can find the scale height $H = 8630$ m and the refractivity $\varepsilon\rho_0 = 312.5 \times 10^{-6}$. These values are very different from the ones given above that match Newton's table, and, as expected, the refractions calculated with them do not agree at all with the table. The discrepancies are of the order of hundreds of arcseconds. Newton clearly did not use these values, even though they were available to him. As postulated previously, he must have set his parameters simply to match the measured refraction data.

Newton was never fully satisfied with his results and never published them. He knew that the temperature distribution in the air was an important factor [44]. He could not derive it himself from mechanical principles, as he had done for pressure, nor did he have access to any scientific observations about it. It may be for these reasons that his *Opticks* (in 1704) contained only a very limited discussion of astronomical refraction, for the flat-Earth case. Nevertheless, the second table was published by Edmund Halley in 1721 [45].

8. Conclusions

When one considers that a 17th Century hand-calculated numerical integration is being compared with that of a modern MATLAB program, the agreement is actually quite remarkable. It suggests that the integrals Newton solved to build his tables were rather similar to the ones presented here. The conclusion appears inescapable that Isaac Newton fully understood the refraction of the isothermal atmosphere and that he was the first to produce a correct mathematical model for the phenomenon.

Appendix

The refraction integral has a singularity at its lower limit when $\theta_1 = 90°$. Consider Newton's isothermal, spherically symmetric atmosphere in a constant gravitational field, as described by Eqs. (15)–(17). Combining these gives the full expression of the integral:

$$f = \int_0^\infty \frac{\varepsilon\rho_0(1+\varepsilon\rho_0)\sin\theta_1\ \exp(-z/H)\mathrm{d}z}{nH\sqrt{[1+\varepsilon\rho_0\ \exp(-z/H)]^2[1+z/R_E]^2 - [1+\varepsilon\rho_0]^2\sin^2\theta_1}}. \tag{A1}$$

For the horizontal ray, when $\theta_1 = 90°$, the denominator becomes zero when $z \to 0$. However, in this limit the contents of the square root can be replaced by their first-order expansions; for example, replace $\exp(-z/H)$ with $(1 - z/H)$, etc. The root will then contain a linear function of z, which enables the integral to be carried out down to the lower limit. The total horizontal refraction can then be found by integrating the approximate expression between limits $(0, \delta)$ and adding to that the full integral (A1) evaluated between limits (δ, ∞). The parameter δ must be relatively small; however, good results are obtained even if it is as high as 500 m.

At the other end of the scale, namely, small zenith angles, the refraction integral can be greatly simplified, with little loss of accuracy. In the refraction integral, Eq. (A1), because $\varepsilon\rho_0$ and z/R_E are much smaller than 1, the elements in the denominator may be replaced by their first-order expansions to yield

$$f = \int_0^\infty \frac{\varepsilon\rho_0(1+\varepsilon\rho_0)\sin\theta_1\ \exp(-z/H)\mathrm{d}z}{nH\sqrt{\cos^2\theta_1 + 2\varepsilon\rho_0\ \exp(-z/H) + 2z/R_E - 2\varepsilon\rho_0\ \sin^2\theta_1}}. \tag{A2}$$

With Newton, again take $n = 1$, and cancel it with the surface refractive index $1 + \varepsilon\rho_0$. Now divide out the $\cos^2$ term to get

$$f = \int_0^\infty \frac{\varepsilon\rho_0\tan\theta_1\ \exp(-z/H)\mathrm{d}z}{H\sqrt{1 + \frac{2\varepsilon\rho_0\ \exp(-z/H)}{\cos^2\theta_1} + \frac{2z}{R_E\cos^2\theta_1} - 2\varepsilon\rho_0\tan^2\theta_1}}. \tag{A3}$$

Expand the root in first order terms and move it to the numerator:

$$f=\int_0^\infty \frac{\varepsilon\rho_0}{H}\tan\theta_1\,\exp(-z/H)\left[1-\frac{\varepsilon\rho_0\exp(-z/H)}{\cos^2\theta_1}-\frac{z}{R_E\cos^2\theta_1}+\varepsilon\rho_0\tan^2\theta_1\right]\mathrm{d}z. \tag{A4}$$

This integrates to

$$f=\varepsilon\rho_0\,\tan\theta_1\left[1+\varepsilon\rho_0\left(\tan^2\theta_1-\frac{1}{2\cos^2\theta_1}\right)-\frac{H}{R_E\cos^2\theta_1}\right]. \tag{A5}$$

To obtain Newton's resolution of 1 arc sec, the term in the brackets can safely be considered to be unity. For example, when $\theta_1=45°$, the factor is 0.9973. For this case Newton's table contains a refraction of 54″, and ignoring a factor of 0.9973 introduces a deviation far less than 1″.

Biot extracted the following values from Newton's data: $\varepsilon\rho_0=0.0002625068$ and $H/R_E=0.001350536$. He used them to construct a slightly different formula for small zenith angles:

$$f=54.0652''\,\tan\theta_1-0.065988''\,\tan^3\theta_1. \tag{A6}$$

At $\theta_1=45°$, Eqs (A5) and (A6) produce exactly the same result of 54.00″. But given that Biot's $\tan^3\theta_1$ correction term is completely negligible in the light of 1″ precision, a very simple form is left,

$$f=\varepsilon\rho_0\,\tan\theta_1, \tag{A7}$$

which with Biot's value for $\varepsilon\rho_0$ reproduces Newton's table for all zenith angles less than 45°.

References

1. W. H. Lehn and S. Y. van der Werf, "Atmospheric refraction: a history," Appl. Opt. **44**, 5624–5636 (2005).
2. R. Grant, *History of Physical Astronomy* [H.G. Bohn, 1852 (Johnson Reprint, 1966)], p. 321.
3. Kepler reproduced Tycho's table in 1604. J. Kepler, *Optics: Paralipomena to Witelo, & Optical Part of Astronomy*, W. H. Donahue, trans. (Green Lion, 2000), p. 135. The resolution displayed in the table (down to 5 arc sec in places) can come only from interpolation.
4. Kepler's refraction table is calculated to a resolution of 1 arc sec; [3], p. 138.
5. [2], p. 322.
6. The numerical values in Cassini's model can be identified as the following: depth of the atmosphere, 3880 m; refractivity of the air, $n_0-1=284.1\times10^{-6}$.
7. J. B. J. Delambre, *Histoire de l'Astronomie Moderne* [Courcier, 1821 (Johnson Reprint, 1969)], Vol II, p. 725.
8. J. F. Scott, ed., *The Correspondence of Isaac Newton* (Cambridge U. Press, 1967), Vol. IV, Letter 474, 11 October 1694.
9. W. E. K. Middleton, *The History of the Barometer* (Johns Hopkins, 1964).
10. R. Hooke, *Micrographia* (Martyn and Allestry, 1665), p. 227. Newton's extensive notes on this book are in A. R. Hall and M. B. Hall, eds., *Unpublished Papers of Isaac Newton* (Cambridge U. Press, 1962).
11. W. E. K. Middleton, *A History of the Thermometer and its Use in Meteorology* (Johns Hopkins, 1966).
12. [10], p. 38.
13. I. Newton, *Opticks*, 4th ed. [William Innys, 1730 (Dover, 1952)], Proposition 10.
14. J. Ivory, "The Bakerian lecture: on the theory of the astronomical refractions," Philos. Trans. R. Soc. London **128**, 169–229 (1838).
15. I. Newton, *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his System of the World*, F. Cajori, ed. (U. of California Press, 1960), Book I, Propositions 94 and 95.
16. [13], Proposition 10. Here Newton used the term "refractive power."
17. J.-B. Biot, "Analyse des Tables de Réfraction construites par Newton, avec l'indication des procédés numériques par lesquels il a pu les calculer," J. Savants 735–754 (1836). His source for Newton's correspondence was F. Baily, *An Account of the Revd. John Flamsteed, the first Astronomer-Royal* (Clowes 1835).
18. Hauksbee presented the experiment in 1708, but published it later, in F. Hauksbee, *Physico-mechanical Experiments on Various Subjects*, 2nd ed. (J. Senex, 1719).
19. [8], Letter 480, 17 November 1694.
20. [8], Letter 485, 20 December 1694.
21. [8], Letter 487, 15 January 1694/5. In Newton's day, the New Year began on March 25; he would have thought of his date as January 1694.
22. J. M. Pernter and F. Exner, *Meteorologische Optik*, 2nd ed. (Braumüller, 1922), p. 71.
23. C. Bruhns, *Die astronomische Strahlenbrechung in ihrer historischen Entwickelung* (Voigt & Günther, 1861), p. 48. Biot ([17], p. 744), also had no doubts that Newton knew the differential equation for the refraction.
24. [8], Letter 474, 11 October 1694.
25. [8], Letter 478, 1 November 1694: "Neither need you send me your larger synopsis of ye refractions."
26. The calculated value at altitude 15° misses the mark by only 1/6 arc sec.
27. [17], p. 746.
28. Interpolation would have been essential if Newton had actually done the calculation in the manner he described to Flamsteed.
29. [8], Editor's footnote to Letter 480, 17 November 1694.
30. Newton was usually very careful about how he wrote his numbers. This can be observed, for example, in D. T. Whiteside, ed., *The Mathematical Papers of Isaac Newton* VI (Cambridge U. Press, 1974), plates 420 and 430. But occasionally, when he would write one number over another, the result was hard to read.
31. T. Simpson, "Determination of the astronomical refractions," in *Mathematical Dissertations on a Variety of Physical and Analytical Subjects* (T. Woodward, 1743). Simpson knew about the exponential model, but considered his linear model to be more accurate.
32. [8], Letter 494, 16 February 1694/5.
33. [15], Book II, Proposition 22.
34. In the isothermal atmosphere, the scale height H also appears in the relation between density and pressure: $\rho=p/g_0H$ (where g_0 is the acceleration of gravity at the Earth's surface).
35. B. Taylor, *Methodus Incrementorum Directa & Inversa* (Londini, 1715).

36. C. Kramp, *Analyse des réfractions astronomiques et terrestres* (Dannbach, 1799).
37. P. S. Laplace, *Celestial Mechanics*, N. Bowditch, trans. (Chelsea, 1966), Vol. IV, pp. 438–540.
38. G. Peacock, ed., *Miscellaneous Works of the Late Thomas Young* [J. Murray, 1855 (Johnson Reprint, 1972)], Vol. II, pp. 52–74.
39. [30], pp. 431–434.
40. D. T. Whiteside, "Kepler, Newton and Flamsteed on refraction through a 'regular aire': the mathematical and the practical," Centaurus **24**, 288–315 (1980).
41. This simple expression is the first term of a power series for refraction (Oriani's Theorem, 1788). Oriani showed that up to a zenith angle of 70°, the refraction was independent of the structure assumed for the atmosphere. This explains why Cassini's model was so successful over this range. See [23], pp. 70–74.
42. Whiteside remarks on the necessity of interpolation. See [30], pp. 433–434.
43. The full equation for the density of dry air is $\rho = \beta p/T$, where $\beta = 3.484 \times 10^{-3}$ in SI units. At STP, the density is $1.2923\,\mathrm{kg\,m^{-3}}$.
44. [8], Letter 494, 16 February 1694/5.
45. E. Halley, "Some remarks on the allowances to be made in astronomical observations for the refraction of the air," Philos. Trans. R. Soc. London **31**, 169–172 (1720–1721).