ITU-R Correspondence Group 3K1

COMMENTS ON THE SPHERICAL EARTH DIFFRACTION METHOD OF RECOMMENDATION ITU-R P.526-10

K H Craig

26 August 2008

Contents

1 Introduction 2

2 Background 2

3 The P.526 method 3

4 Comparison of P.526 and TN101 4

5 Comparison of P.526 and the residue series 9

6 Validity limits 13

7 P.526 nomograms 17

8 Conclusions and proposals 18

9 References 19

1 Introduction

ITU-R Document 3J/34 raised an issue concerning the sign of the 10log(X) term in Equation (17) of Section 3.1.1 of Recommendation ITU-R P.526-10. This equation is part of the numerical procedure for calculating the field strength due to diffraction over a spherical Earth for over-the-horizon paths. It was observed that changing the sign of this term for values of $X < 1$ appears to give a better fit to GRWAVE results in the line-of-sight region. It was further observed that the results obtained from the nomogram method in Section 3.1.2 appeared to give reasonable results well into the line-of-sight region. The question was raised as to whether Equation (17) may be in error.

In this document the derivation of the equations of Section 3.1.1.2 of P.526-10 are revisited. The equations are re-derived from the residue series and compared with other results in the literature. It is shown that Equation (17) is in fact correct. The problem is in the expectation that the method of Section 3.1.1 should be applicable in the line-of-sight region. It was possible to derive relatively simple formulae that define the limit of validity of the method, and it is proposed that this be added to P.526. Some discrepancies in Equation (16) defining β and Equation (18) defining the height-gain functions were found, and improved formulations are proposed.

2 Background

The method given for diffraction over a spherical Earth in Section 3.1 of P.526-10 has a heritage going back to at least 1986. The formulae of Section 3.1.1 and the nomograms of Section 3.1.2 exist in, and remain unmodified from, Recommendation 526-1/Report 715-2 (1986). This means, of course, that it was not possible to go back to the original input documents for a derivation of the method. On the other hand, Report 715-2 contains a reference section that gives some clues to the original open literature sources, although the specific equations in P.526 do not appear to come directly from any of these sources.

Section 3.1 of P.526-10 does make it clear that the method of Section 3.1.1 is based on “only the first term of the residue series”. It is further noted that “even near or at the horizon this approximation can be used with a maximum error of around 2 dB in most cases”.

The residue series method for calculating the effects of diffraction of radiowaves round the Earth’s surface is one of the canonical solutions of Maxwell’s equations, and was developed in the first half of the 20th Century by a number of people, including Watson, Van der Pol, Bremmer, Norton, Leontovich and Fock. The method allows the field strength to be calculated as a sum of contributions from discrete “modes”. The modes form a series, each mode having a higher attenuation rate with distance than the previous one. Thus at distances well beyond the horizon, only the lower order modes contribute significantly to the sum. Far enough beyond the horizon, it is possible to approximate the field by the contribution of a single (least attenuated) mode (the first term of the residue series). In this case, the mode “sum” becomes rather simple, and it is possible to factor the expression for the field strength into separate range dependent and height gain function terms. Indeed in some cases it is even possible to give analytical expressions for these terms. This is true for the simplest case of a linear vertical refractive index profile, and because this is a good approximation for the “average” atmosphere near the surface of the Earth, is widely used.

In practice, it turns out that a single mode can be a good approximation to the full mode sum even at distances quite close to the radio horizon. But this is not true at shorter distances, so a single mode approximation is not usable well within the line-of-sight region. However, the *full mode sum* is in principle *valid at all ranges*. Computer implementations of the mode method can be used to obtain valid results at line-of-sight ranges by including sufficient modes for convergence of the results. Computer implementations are also able to include more complicated refractive index profiles. Two examples are (a) the inclusion of an exponential refractive index profile, as in GRWAVE, and (b)

the inclusion of multi-linear refractive index profiles to model the effects of surface and elevated ducts: these programs can model the guided wave propagation in ducts and the diffracted "creeping" waves in a uniform framework. Unfortunately these more complex cases cannot be converted into simple analytical formulae.

Given the statement in Section 3.1 of P.526-10 that the formulae represent the first term of the residue series, a number of sources in the literature were referenced to check the P.526 equations. These included:

- Chapter 8 of NBS TN101 [1] contains a mixed analytical/graphical-based method from calculating the field strength due to diffraction over a spherical Earth, based on a residue mode series calculation. It was initially believed that the equations of P.526 were derived from this source
- The book edited by Kerr [2] is representative of a textbook treatment of the residue series, and the formulation turned out to be very close to that used in P.526. An implementation of the residue series for guided wave propagation written by the author in 1985 [3] based on the approach of Wait and Spies [4] was also used.

3 The P.526 method

The diffraction field strength, E, relative to the free-space field, E_0, is given by P.526-10 Equation (13):

$$20\log\frac{E}{E_0} = F(X) + G(Y_1) + G(Y_2) \tag{1}$$

The normalised range, X, and heights, Y_1 and Y_2, are given by P.526 Equations (14) and (15).

In self-consistent units:

$$X = \beta\left(\frac{\pi}{\lambda a_e^2}\right)^{1/3} d \quad \text{and} \quad Y = 2\beta\left(\frac{\pi^2}{\lambda^2 a_e}\right)^{1/3} h \tag{2}$$

In practical units;

$$X = 2.2\beta f^{1/3} a_e^{-2/3} d \quad \text{and} \quad Y = 0.0096\beta f^{2/3} a_e^{-1/3} h \tag{3}$$

where d is the path length (km), a_e is the equivalent Earth's radius (km), h is the antenna height (m) and f is the frequency (MHz).

The quantity β is given as a function of the surface admittance K of the ground:

$$\beta(K) = \frac{1 + 1.6K^2 + 0.75K^4}{1 + 4.5K^2 + 1.35K^4} \tag{4}$$

K can be calculated in terms of the frequency, polarisation, the effective Earth radius and the relative permittivity and conductivity of the ground, using formulae given in Section 3.1.1.1 of P.526-10. These formulae are well-established and straightforward, and there are no issues with them. Note that although K is referred to as the "surface admittance" in P.526, the surface admittance is a complex quantity, and K is actually the magnitude of the admittance. The phase of the admittance is ignored in P.526.

The function $F(X)$ depends only on the normalised distance X and is given analytically by P.526-10 Equation (17):

$$F(X) = 11 + 10\log X - 17.6X \tag{5}$$

The function $G(Y)$ depends on the normalised height Y and the surface admittance K and is given as a piecewise analytical formula in P.526-10 Equation (18). This formula has the appearance of a "curve fit".

ITU-R Document 3J/34 questions the sign of the 10log(X) term in eqn (5)[1]. Independent verification of this expression is sought in the following sections.

4 Comparison of P.526 and TN101

TN101 is the most accessible of the source documents as it gives a practical method as a combination of simple formulae and definitions based on plotted graphs. There are however three difficulties when comparing P.526 and TN101:

1. TN101 does not derive its equations from first principles, and the equations are given in "practical units" rather than in fundamental physical quantities. So while a comparison is possible, it is not always clear where the terms have come from.
2. The distance and height "normalisations" are different in the two sources. Indeed in TN101 antenna heights are defined in terms of their horizon distance, which leads to a square root relationship between the normalised heights in P.526 and TN101.
3. The same nomenclature is used with different meanings in the two sources, which can be confusing. Here we add subscripts "P526" and "TN101" where necessary to distinguish these.

Superficially, TN101 appears to have a similar structure to P.526. The equivalent to eqn (1) is given by TN101 Equation (8.1), here converted from attenuation to field strength and dropping the gaseous absorption term A_a which is not included in P.526:

$$20\log\frac{E}{E_0} = -G_{TN101}(x_0) + F_{TN101}(x_1) + F_{TN101}(x_2) + C_1(K,b) \tag{6}$$

Consider first the parameter $C_1(K,b)$. This is a function of K and b. It is clear that these are the magnitude and phase of the surface admittance, and are given graphically in TN101 in terms of the frequency, polarisation, the effective Earth radius and the relative permittivity and conductivity of the ground. The curves for K given in TN101 Figure 8.1 agree with those given in Figure 2 of P.526-10. It should be noted that for horizontal polarisation K is small (< 0.05) for all ground types at all frequencies, while for vertical polarisation, K is small (< 0.05) for frequencies above 20 MHz over land and above 300 MHz over sea.

$C_1(K,b)$ is also defined graphically in Figure 8.4 of TN101. It has the limiting values $C_1(0,b) = 20.03$ and $C_1(\infty,b) = 20.94$, independent of the value of b. For K in the range 0.1 to 10, $C_1(K,b)$ varies between 16.5 and 21.5, depending on the value of b. Since P.526 does not take account of the phase of the surface admittance, and no such term as C_1 appears in the P.526 formula, eqn (1), it is assumed that a constant value of ~20dB (i.e. the small K value) should be assumed for $C_1(K,b)$ and this 20dB is absorbed somehow in the $G(.)$ and $F(.)$ functions.

The definition of the normalised distance, x_0, and heights, x_1 and x_2, requires quite a bit of manipulation to get them into a form that is comparable to the P.526 X and Y quantities. The relationship is:

$$x_0 = \Lambda X \quad \text{and} \quad x_{1,2} = \Lambda(\beta Y_{1,2})^{1/2} \tag{7}$$

[1] Here "eqn (N)" refers to equations in this document and "Equation (N)" refers to external documents.

where $$\Lambda = \left(\frac{8497^2 c}{\pi}\right)^{1/3} \frac{B(K,b)}{\beta(K)} \tag{8}$$

In this equation, the speed of light, c, is given in units of km.MHz (!). $B(K,b)$ is again defined graphically in TN101, Figure 8.3. $B(K,b)$ has the limiting values $B(0,b) = 1.607$ and $B(\infty,b) = 0.7003$, independent of the value of b. $B(K,b)$ is a strong function of K, but a weak function of b, and is (nearly) monotonically decreasing with K for fixed b. TN101 suggests that the limiting value $B(K,b) = 1.607$ for $K \rightarrow 0$ may be used for most cases of horizontal polarisation.

By inspection, it appeared that the P.526 $\beta(K)$ function of eqn (4) was proportional to $B(K,b)$ with $b = 0$. The proportionality constant must be 1.607 to satisfy the requirement that $B(0,b) = 1.607$ while $\beta(0) = 1$. This suggests that

$$B(K,b) = 1.607\beta(K) \tag{9}$$

Unfortunately this doesn't quite work, since the ratio $B(0,b)/B(\infty,b)$ from TN101 is 2.29, while the ratio $\beta(0)/\beta(\infty)$ is 1.8 from eqn (4). However, the identification of $B(K,b)$ in TN101 with $\beta(K)$ in P.526 is sufficiently compelling that, assuming that the stated $K = 0$ and ∞ limits given in TN101 are correct (and much of the rest of TN101 agrees with P.526 as will be shown below), we are inclined to deduce that the P.526 formula for $\beta(K)$ is in error. A small modification to eqn (4) corrects the situation:

$$\beta_{new}(K) = \frac{1 + 1.6K^2 + 0.67K^4}{1 + 4.5K^2 + 1.53K^4} \tag{10}$$

Differences between the expressions of eqns (4) and (10) are in any case only noticeable for vertical polarisation below 20 MHz over land or 300 MHz over sea; in all other cases, $\beta_{new}(K) \approx \beta(K) \approx 1$.

Eqn (10) leads to the satisfying result that the parameter Λ which relates the TN101 and P.526 normalised distance and height parameters in eqn (8) is a simple constant:

$$\Lambda = 305.75 \tag{11}$$

Finally we can compare the TN101 and P.526 distance and height gain functions. The TN101 function $G_{TN101}(x_0)$ is given by TN101 Equation (8.4):

$$G_{TN101}(x_0) = 0.05751x_0 - 10\log x_0 \tag{12}$$

Converting x_0 to X using eqns (7) and (11) gives:

$$G_{TN101}(x_0) = 17.58X - 10\log X - 24.85 \tag{13}$$

The coefficients of the linear and logarithmic terms in X agree well with the P.526 expression, eqn (5). This confirms that the linear and logarithmic terms should have the opposite sign. The constant term is different from the P.526 expression, but this is expected: we have already noted that the $C_1(K,b)$ term (a constant) needs to be absorbed into the $G(.)$ and $F(.)$ functions. We therefore have:

$$G_{TN101}(x_0) = -F_{P526}(X) - 13.85 \tag{14}$$

To make the P.526 and TN101 expressions for the diffraction field strength, eqns (1) and (6), agree, taking account of eqn (14) and a value of 20dB for $C_1(K,b)$ we postulate the equivalence:

$$F_{TN101}(x_{1,2}) = G_{P526}(Y_{1,2}) - 16.93 \tag{15}$$

The correctness of eqn (15) was confirmed by plotting curves of $F_{TN101}(x)$ derived from eqn (15) against x as follows:

1. For each value of x, calculate Y using eqns (7) and (11);
2. Calculate $G_{P526}(Y)$ using this value of Y and P.526-10 Equations (18);
3. Convert this value of $G_{P526}(Y)$ to $F_{TN101}(x)$ using eqn (15)

In the discussion of the results it will be necessary to refer to the $G_{P526}(Y)$ formulation of P.526-10 Equations (18), which is reproduced here:

$$G_{P526}(Y) \cong 17.6(Y-1.1)^{1/2} - 5\log(Y-1.1) - 8 \qquad \text{for } Y > 2 \qquad (16)$$

$$G_{P526}(Y) \cong 20\log(Y + 0.1Y^3) \qquad \text{for } 10K < Y < 2 \qquad (16a)$$

$$G_{P526}(Y) \cong 2 + 20\log K + 9\log(Y/K)\left[\log(Y/K)+1\right] \qquad \text{for } K/10 < Y < 10K \qquad (16b)$$

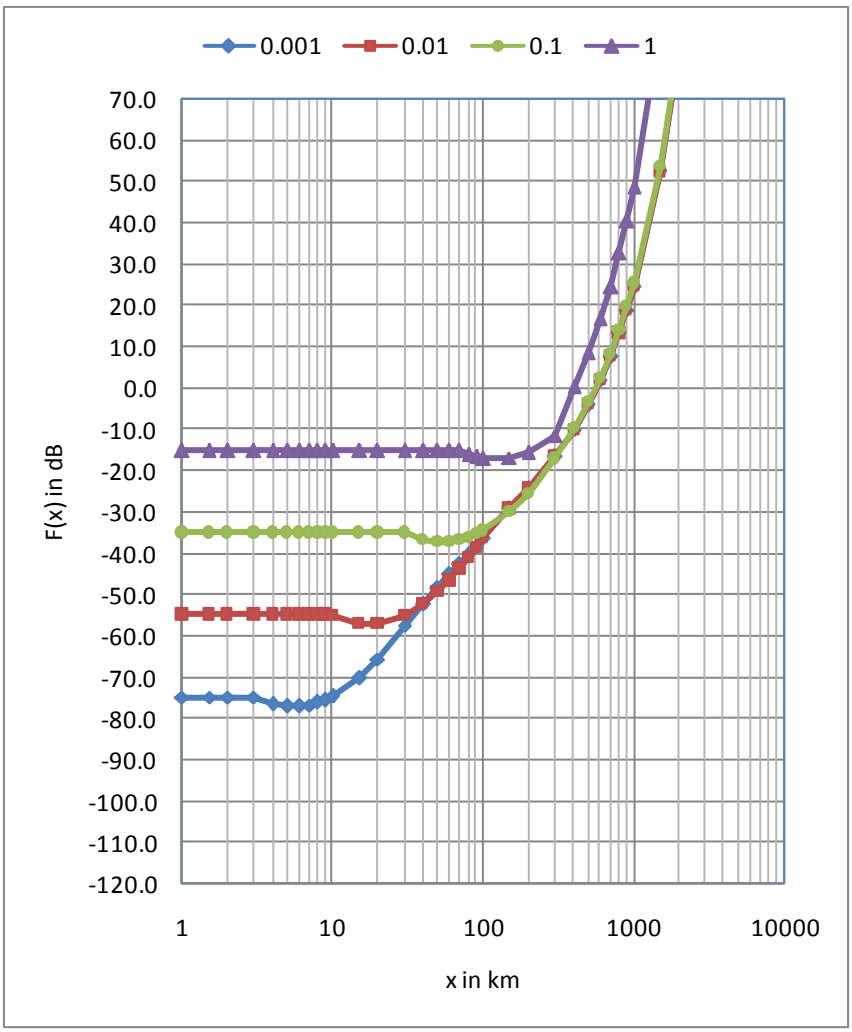
$$G_{P526}(Y) \cong 2 + 20\log K \qquad \text{for } Y < 10K \qquad (16c)$$

The P.526-derived $F_{TN101}(x)$ curves are shown in Figure 1 for four values of K (0.001, 0.01, 0.1 and 1). Comparing these with the curves taken from TN101 Figure 8.6 (reproduced here as Figure 2) we note:

1. The curves correspond very well, at least for the values of $K < 1$. A useful check is at very small x values (for example x = 1km), where eqn (16.c) applies. Here, the value of $G_{P526}(Y)$ depends only on K (20dB per decade of K, independent of the phase, b). The P.526 and TN101 intercept values at x = 1km agree very closely. This confirms the correctness of the constant of 16.93 postulated in eqn (15). (A close inspection of the TN101 curves of Figure 2 appears to show that the curve for $K \geq 10$ lies below the $K = 1$ curve, implying that the formula of eqn (16c) breaks down for very large K values. The reason for this is unknown, and it is possible that the TN101 labelling is in error. In any case, for horizontal polarisation and for all practical ground types at frequencies above 10MHz, K will be less than 1 and this discrepancy can be ignored.)
2. For mid-range values of x (10–100km), the “dip” in the TN101 curves depends on the phase of the surface admittance, p. The P.526-derived curves are phase-independent, but appear to follow the TN101 curves for a value of b somewhere between 0° and 90°. Within the accuracy that can be obtained when b is ignored (as it is in P.526), the agreement is good.
3. At the largest values of x (for example x = 1000km) the TN101 and P.526-derived curves agree for the smaller values of K, but deviate for $K = 1$.

FIGURE 1

$F_{TN101}(x)$ calculated using P.526 Equation (18) for 4 values of K



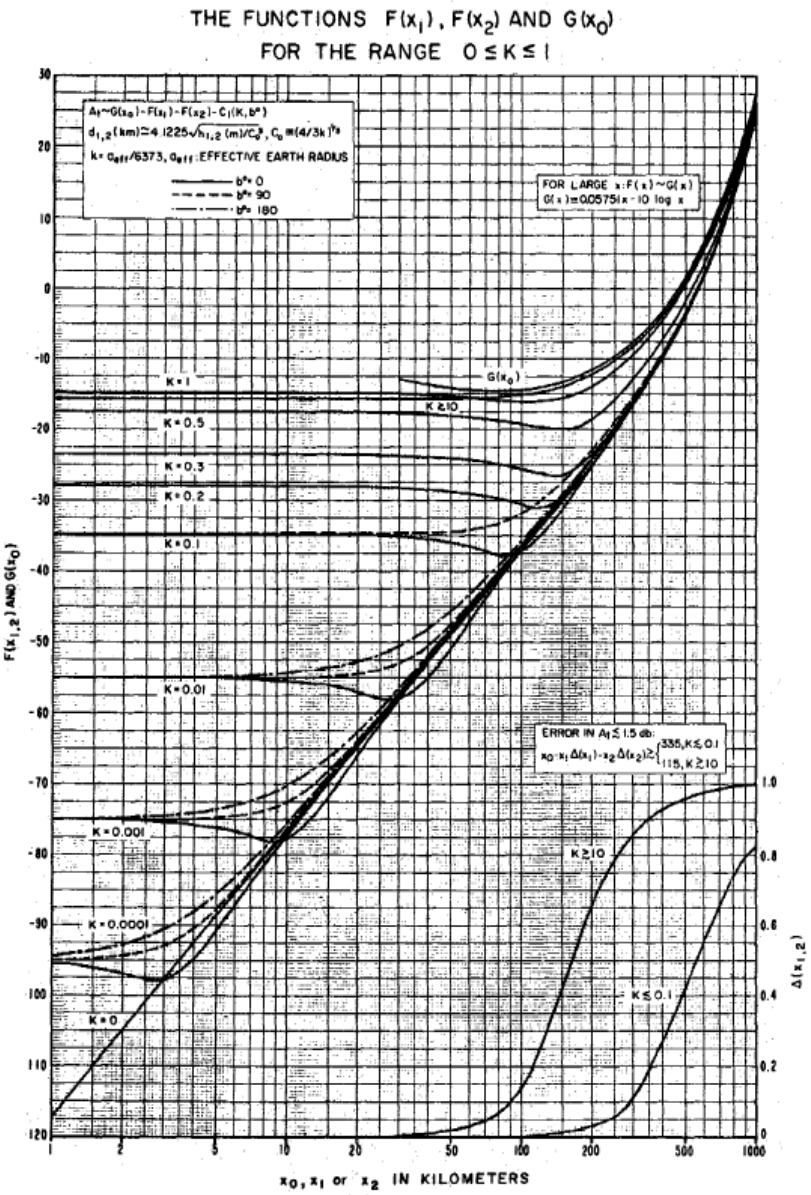
So apart from the small discrepancy for large x and K, the correspondence between the P.526 and TN101 formalisms described above appears to be correct. In particular, TN101 supports the formulation of P.526-10 Equation (17) with the positive sign in front of the logarithmic term.

It may appear that we have had to do rather a lot of work to confirm the validity of P.526-10 Equation (17). And in some ways, the derivation was not entirely satisfactory as a number of constants were involved that were not derived from first principles. To remedy this we will derive P.526-10 Equation (17) directly from the residue series in the next section. But the correspondence between the TN101 and P.526 formalisms is useful for another purpose. TN101 provides formulae for calculating the distance and height ranges over which the single mode residue series result is valid. Using the correspondence, we can translate these formulae into the P.526 formalism for incorporation into P.526.

Finally in this section, we address the large x and K discrepancy between P.526 and TN101. Figure 2 shows that, for large values of x, the height-gain curves of $F_{TN101}(x)$ become independent of K. Indeed it is noted on Figure 2 that for large x, $F_{TN101}(x) \sim G_{TN101}(x)$; that is, the "height gain" term is the same as the "distance" term! This is one advantage of the normalisation and height definition used in TN101, and is certainly not true of the P.526 distance and height gain terms.

FIGURE 2

$F_{TN101}(x)$ taken from TN101 Figure 8.6 for several values of K

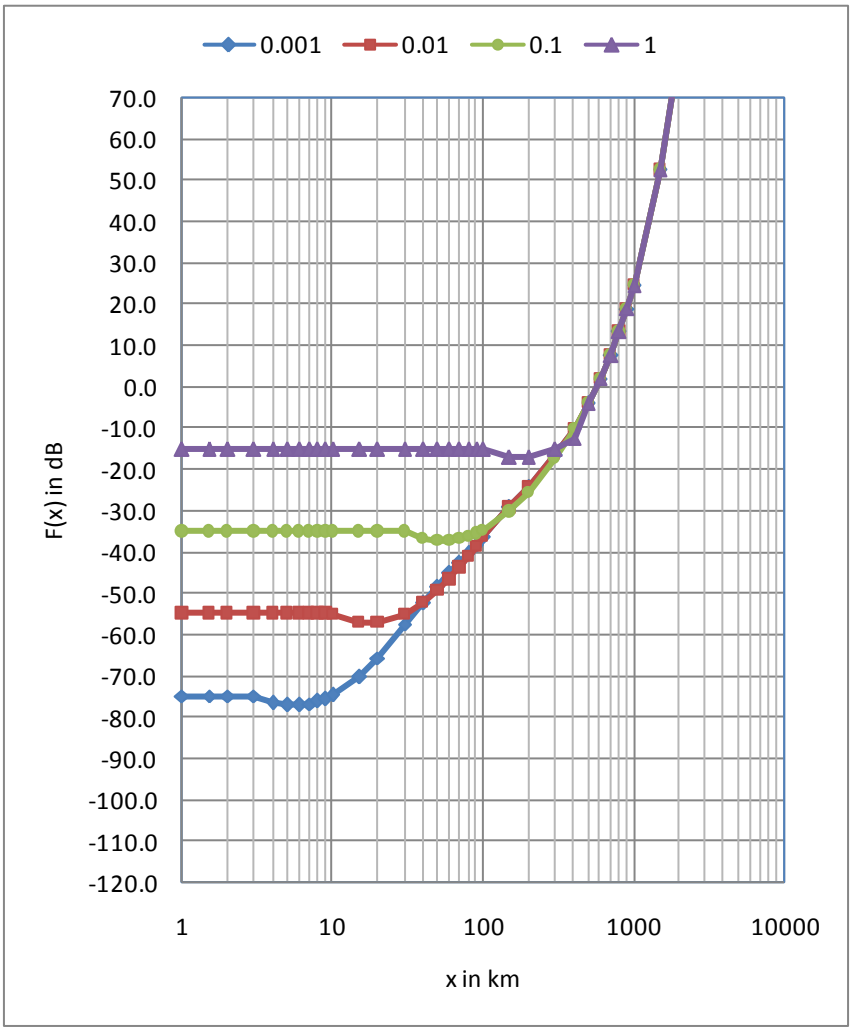


However, we should still expect the P.526-derived $F_{TN101}(x)$ curves of Figure 1 to be independent of K at large x, but this is not what is seen for $K = 1$. Eqn (16) shows that $G_{P526}(Y)$ is independent of K for $Y > 2$. But an implicit K-dependence comes in through the transformation from x to Y. Eqn (7) shows that this involves β which depends on K. For all values of $K \leq 0.1$, $1 \geq \beta > 0.97$, so the K-dependence is negligible. But $\beta = 0.49$ for $K = 1$ which significantly changes the value of Y used in eqn (16) for a given value of x. This explains the higher curve for $K = 1$.

Fortunately there is a simple solution to this problem. Eqn (7) shows that if we use the quantity βY in eqns (16)-(16c) instead of Y, then the K-independence of $F_{TN101}(x)$ for large x will be restored. This result is shown in Figure 3. It is proposed that Equations (18)-(18c) of P.526 be amended along these lines.

FIGURE 3

$F_{TN101}(x)$ calculated using modified P.526 Equation (18) for 4 values of K



In [5] Longley and Rice provide fits to their TN101 Figure 8.6 height gain curves. On the face of it, these (formulated in terms of x and K, and with different expressions in different domains) do not appear to be the same as the P.526 formulation. It may be worth examining these further.

5 Comparison of P.526 and the residue series

In this Section we show how the P.526-10 distance term can be derived directly from the residue series. As has been said, this is a canonical electromagnetic wave propagation problem, and the full derivation is not given. It is standard textbook material, and requires an unnecessary amount of mathematics for an ab initio derivation from Maxwell's equations. It turned out that the derivation given in Chapter 2 of the book by Kerr [2] used conventions and normalisations that are close to the formalism of P.526.

Kerr Equation (2.427a) gives the residue sum for the diffractive field strength, following a significant amount of mathematical reduction:

$$\left|\frac{E}{E_0}\right| = 2\sqrt{\pi X}\left|\sum_{m=1}^{\infty} e^{iA_m X} U_m(Z_1) U_m(Z_2)\right| \tag{17}$$

Using Kerr Equations (2.347), (2.351), (2.355), (2.358) and (2.359) X and Z can be shown to be:

$$X = \left(\frac{\pi n_0}{\lambda a_e^2}\right)^{1/3} d \quad \text{and} \quad Z = 2\left(\frac{\pi^2 n_0^2}{\lambda^2 a_e}\right)^{1/3} h \tag{18}$$

Comparison with eqn (2) shows that the Kerr X is the same at the P.526 X, and the Kerr Z is the same as the P.526 Y, apart from the terms n_0 (the refractive index at the surface) and the P.526 β (a function of the surface admittance). These differences relate to different formulations of the atmosphere and the boundary conditions in the two approaches (the Kerr formalism assumes a linear refractive index profile and a perfectly conducting boundary for simplicity). In this section, the analysis will not be sufficiently detailed for the electrical properties of the surface to matter, so for present purposes we can assume that the Kerr and P.526 distance and height normalisations agree.

In eqn (17), the diffraction field strength relative to free space is given by a simple $X^{1/2}$ dependence times a mode sum. For each mode, the $\exp(iA_m X)$ term tells us how the mode propagates with distance. The A_m are the roots of the mode equation which depends on the refractive index profile and boundary conditions. The A_m are complex numbers, so each mode has a different phase velocity and attenuation rate. For a single mode, the $U_m(Z)$ terms can be interpreted as height gain functions. But it is important to note that the summation over modes means that it is *not* possible to factor the field strength formula into separate distance and height gain terms.

In practice, the values of the imaginary part of A_m increases rapidly with m. For X sufficiently large, only a few modes will contribute significantly to the mode sum. The formulae in P.526 and TN101 correspond to the case when all modes, apart from the first, can be ignored. In this case, eqn (17) *does* factor into separate distance and height gain functions:

$$\left|\frac{E}{E_0}\right| = 2\sqrt{\pi X}\left|e^{iA_1 X}\right| \times \left|U_1(Z_1)\right| \times \left|U_1(Z_2)\right| \tag{19}$$

For the linear atmosphere, it can be shown that the roots of the mode equation, A_m, can be written in terms of the zeros, ζ_m, of the Airy function, $Ai(z)$, as:

$$A_m = \left|\zeta_m\right| e^{2\pi i/3} + \frac{1}{Hp} \tag{20}$$

The distance term in eqn (19) then becomes:

$$f_{mode}(X) = 2\sqrt{\pi X}\, e^{-\left|\zeta_m\right|\cos(\pi/6)X}\, e^{X/\left|Hp\right|} \tag{21}$$

or in decibels:

$$F_{mode}(X) = 20\log\left|\frac{E}{E_0}\right| = 10.992 + 10\log X - \left[7.5222\left|\zeta_m\right| - \frac{8.6859}{\left|Hp\right|}\right] X \tag{22}$$

The Hp term is a second order correction that depends on frequency and the electrical properties of the ground. It is normally negligible for horizontal polarisation, but can contribute for vertical polarisation.

The zeros, ζ_m, of the Airy function lie along the negative real axis and are tabulated in Abramowitz and Stegun [6] Table 10.3, or can be calculated using an asymptotic formula:

$$\zeta_m = -f\left[\tfrac{3}{2}\left(m-\tfrac{1}{4}\right)\pi\right] \tag{23a}$$

where
$$f(z) = z^{2/3}\left[1+\frac{5}{48z^2}-\frac{5}{36z^4}+\frac{77125}{82944z^6}-\cdots\right] \tag{23b}$$

Retaining only the zeroth order term in the square brackets of eqn (23b) (i.e. the constant 1), the value of ζ_1 is already accurate to better than 1%, and including the first z^2 correction reduces this to 0.06%. The calculated values of ζ_m are of course even more accurate as *m* increases.

The first few zeros, ζ_m, are given in Table 1.

TABLE 1

First 5 zeros, ζ_m, of the Airy function

m	ζ_m
1	-2.3381
2	-4.0879
3	-5.5206
4	-6.7867
5	-7.9441

Substituting the value of ζ_1 into eqn (22) and ignoring the *Hp* correction term gives:

$$F_{mode(m=1)}(X) = 10.992 + 10\log X - 17.588X \tag{24}$$

This agrees with the P.526 distance term, eqn (5), confirming the sign of the 10log(*X*) term. Further confirmation came from an implementation of the residue series for guided wave propagation written by the author in 1985 [3] based on the approach of Wait and Spies [4]. This gave the same expression for the lowest order diffractive mode.

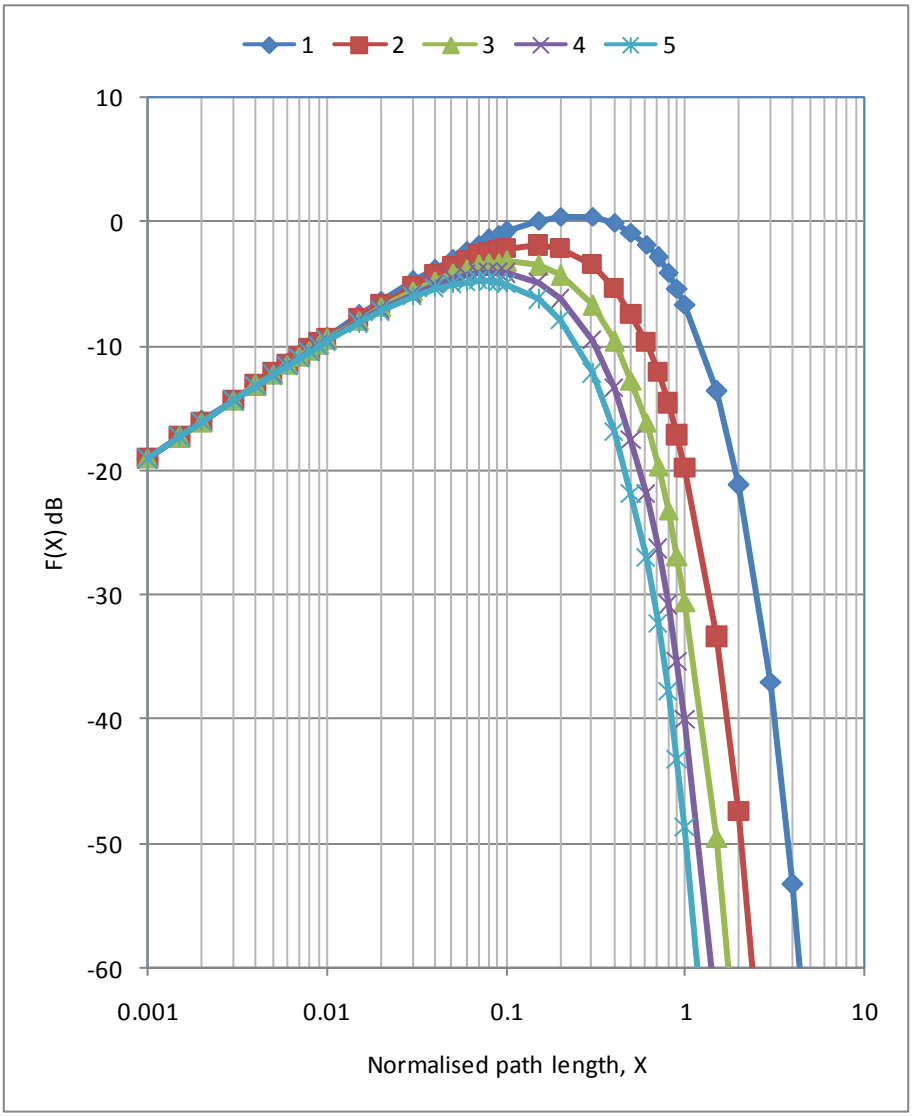
It is worth noting the effect of the *Hp* correction term in eqn (22) which can be non-negligible for vertical polarisation. The sign of the correction term is such that the correction *reduces* the coefficient of the linear *X* term; that is, it *reduces* the loss with range compared to horizontal polarisation. As an example, for a vertically polarised wave propagating over sea water at 100MHz, the correction to $F_{mode}(X)$ due to the *Hp* correction is about +1.1*X*; that is, a vertically polarised wave would experience ~0.03 dB/km lower attenuation than a horizontally polarised wave at 100 MHz. This effect is ignored in P.526.

Figure 4 shows the distance term corresponding to eqn (24) for the first 5 modes. The attenuation rates for these (i.e. the coefficient of the linear *X* term) are 17.6, 30.8, 41.5, 51.1 and 59.8 dB/km, respectively. Of course if other than the first mode is included, then the mode *sum* of eqn (17) must be used, rather than the individual modes. But Figure 4 gives a good indication of when the higher modes can be neglected. For example, at *X* = 2 the second mode has been attenuated by > 25dB more than the second mode, and so its contribution can probably be ignored. At *X* = 1 the difference has reduced to ~13dB. By *X* = 0.1 the difference is less than 2dB, and a single mode assumption would give wrong results. Indeed at the shorter distances, the number of modes required for the mode sum to converge increases very rapidly and becomes computationally intractable. So in practice, a residue series approach is not appropriate well inside the line-of-sight region. Note that the mode sum of eqn (17) involves phases as well as amplitudes. It is the phasor interference between the various modes that can generate patterns at the shorter distances which give rise to the

line-of-sight interference lobing. It follows that it is not possible to generate interference lobing in a single mode approximation, and is another reason why the method of P.526 cannot be used well into the line-of-sight region.

FIGURE 4

Distance term $F_{mode}(X)$ for the first 5 modes



The height gain functions $U_m(Z)$ for a single mode in eqn (17) can easily be expressed as ratios of Airy functions and their derivative. Thus

$$G_{mode}(Z) \equiv 20\log|U_m(Z)| = 20\log\left|\frac{Ai(Z+A_m)e^{\pi i/3}}{Ai'(A_m e^{\pi i/3})}\right| \tag{25}$$

The Airy functions can be calculated numerically, or approximated by Taylor and asymptotic series. The TN101 height gain curves $F_{TN101}(x)$ curves of Figure 2 were presumably calculated numerically in this way. A "curve-fit" to these curves such as that given in P.526-10 appears to be adequate for the purpose, so there is not much to be gained by pursuing a more theoretically-based analytical approximation.

6 Validity limits

Figure 4 gives some insight into the minimum distance for which the single mode assumption of P.526 and TN101 is valid. But it is possible to express the validity limits in some reasonably simple formulae. In TN101 a simple prescription was given for determining the distance and antenna height combinations that would produce an estimate of the diffraction field strength to better than 1.5dB. This was based on expressions of the form:

$$x_0 - x_1\Delta(x_1) - x_2\Delta(x_2) > x_{\lim} \tag{26}$$

where the constant x_{lim} and the function $\Delta(x)$ was defined for the low and high K limits (0 and ∞). $\Delta(x)$ was defined graphically (and in fact the $\Delta(x)$ curves are visible at the bottom-right of Figure 2). For values between these two limits, linear interpolation between these curves and the two values of x_{lim} was used.

Here we convert the TN101 formulae into the formalism of P.526 and instead of graphical lookup have fitted tanh(.) functions to the $\Delta(x)$ curves. This provides relatively simple analytical expressions for estimating the validity limits.

$$X - (\beta Y_1)\Delta(Y_1, K) - (\beta Y_2)\Delta(Y_2, K) > X_{\lim}(K) \tag{27a}$$

where
$$X_{\lim}(K) = 1.096 - 1.280\left(1 - \beta(K)\right) \tag{27b}$$

and
$$\Delta(Y, K) = \Delta(Y,0) + 1.779\left(1 - \beta(K)\right)\left[\Delta(Y,\infty) - \Delta(Y,0)\right] \tag{27c}$$

$\Delta(Y,0)$ and $\Delta(Y,\infty)$ are given by (note that the different signs in front of the constant 0.255 in the two equations is intentional):

$$\Delta(Y,0) = 0.5\left[1 + \tanh\left(\frac{0.5\log_{10}(\beta Y) - 0.255}{0.3}\right)\right] \tag{27d}$$

$$\Delta(Y,\infty) = 0.5\left[1 + \tanh\left(\frac{0.5\log_{10}(\beta Y) + 0.255}{0.25}\right)\right] \tag{27e}$$

For given antenna heights h_1 and h_2, Eqn (27a) can be used to calculate the minimum distance d_{min} at which the diffraction field strength is accurate to 1.5dB. Examples are given in Figures 5–9. These show path loss calculated using the full P.526-10 model, with markers to show (a) the validity limit, d_{min}; (b) the distance at which the path is just line-of-sight; (c) the distance at which the normalised distance, X, = 1. In all these cases, the parameters were kept constant except for frequency, which is 3MHz, 30MHz, 300MHz, 3GHz and 30GHz respectively.

There are a number of things to note. The order of the three markers is different at the different frequencies. It is not difficult to demonstrate that all possible distance orderings of these markers is possible. From eqn (3) the relationship between distance and normalised distance depends only on frequency (and very weakly on the ground constants through β, and effective Earth radius), while the line-of-sight distance depends only on antenna heights (and weakly on effective Earth radius). The validity limit depends on all parameters.

It is clear that the X = 1 distance is not a good indicator of the limit of validity of the residue series. This is particularly noticeable at the lowest and highest frequencies. At 3MHz (Figure 5) X = 1 at 300km while the P.526 model is valid down nearly to the line-of-sight distance at 30km. On the other hand, at 30GHz (Figure 9) X = 1 at 6km which is well below the validity and line-of-sight distance at 30km. This raises doubts about the simple strategy suggested in ITU-R Document 3J/34 for dealing with the line-of-sight region using the P.526-10 distance term of eqn (5), namely to switch the sign of the 10log(X) term at the value X = 1.

FIGURE 5

Path loss at 3MHz, antenna heights = 17 and 10m, sea, vertical pol

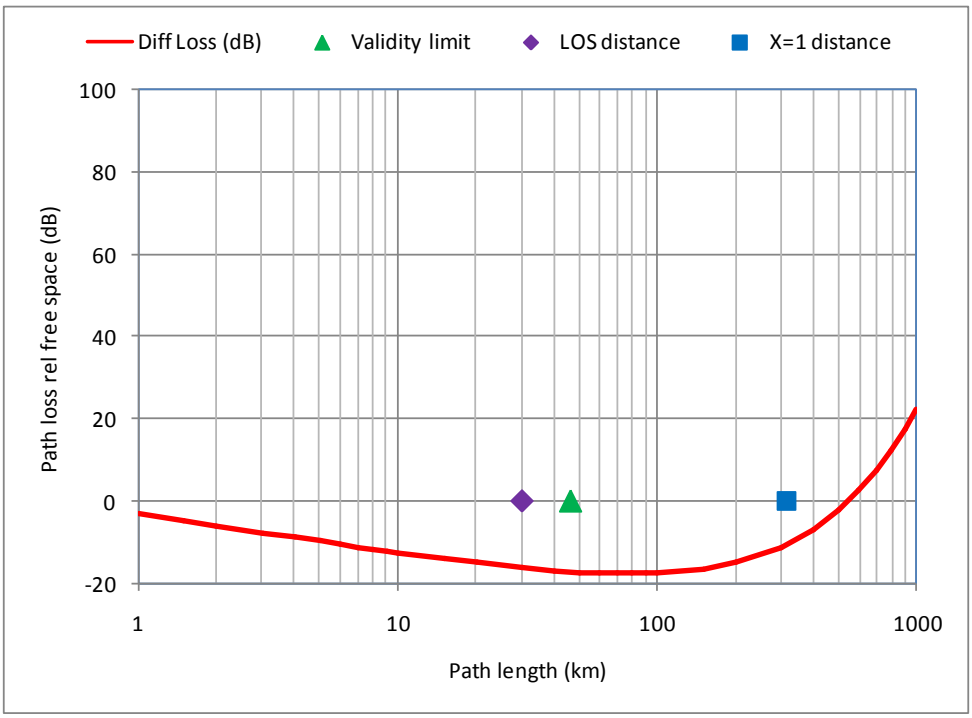


FIGURE 6

Path loss at 30MHz, antenna heights = 17 and 10m, sea, vertical pol

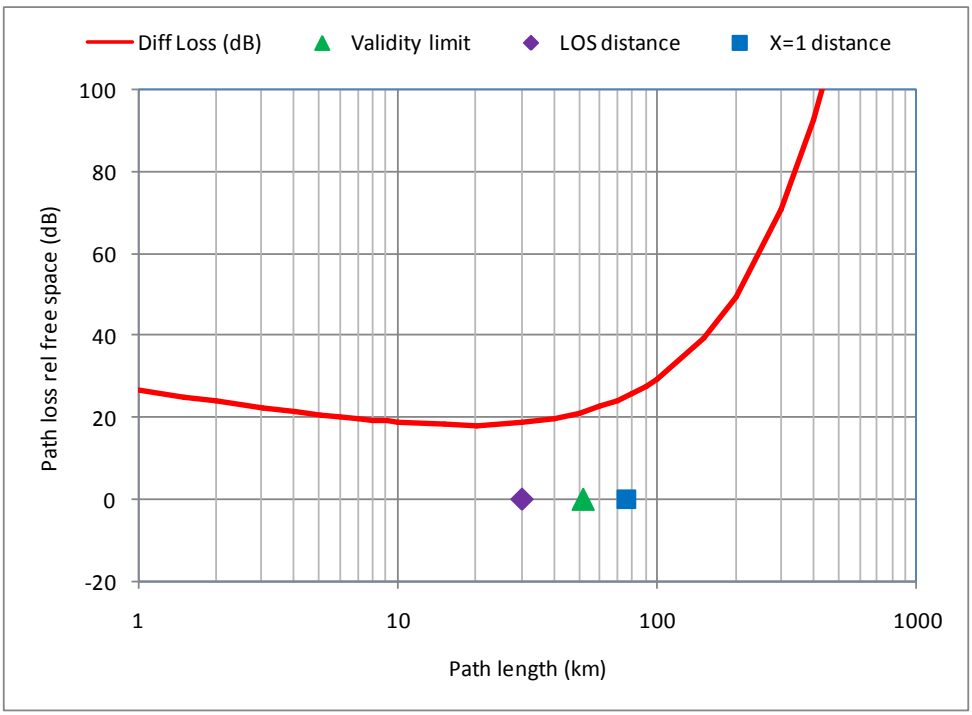


FIGURE 7

Path loss at 300MHz, antenna heights = 17 and 10m, sea, vertical pol

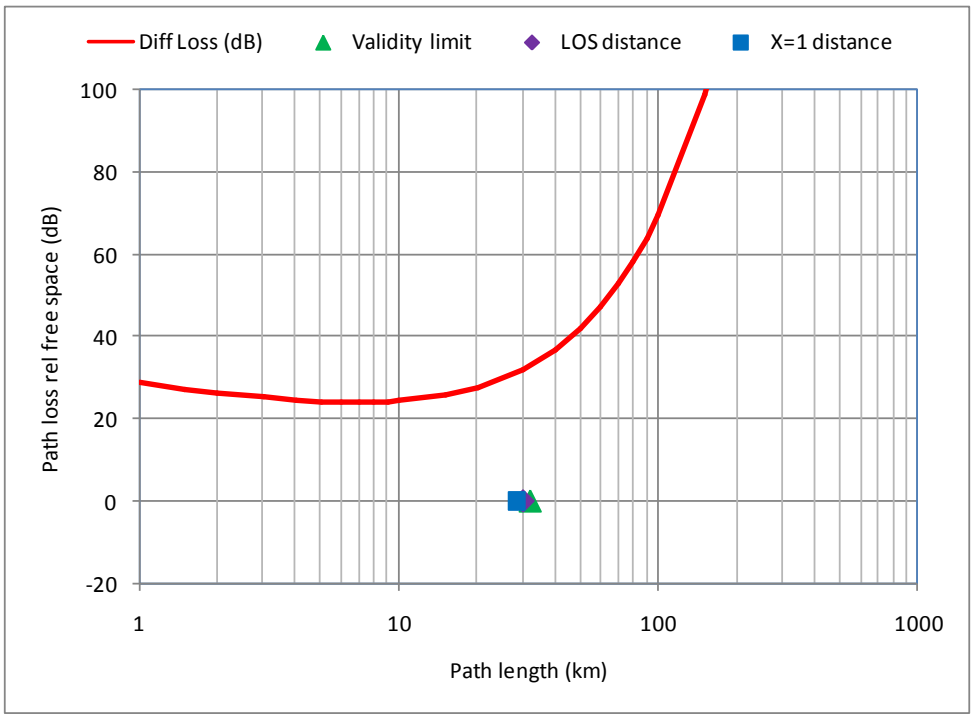


FIGURE 8

Path loss at 3GHz, antenna heights = 17 and 10m, sea, vertical pol

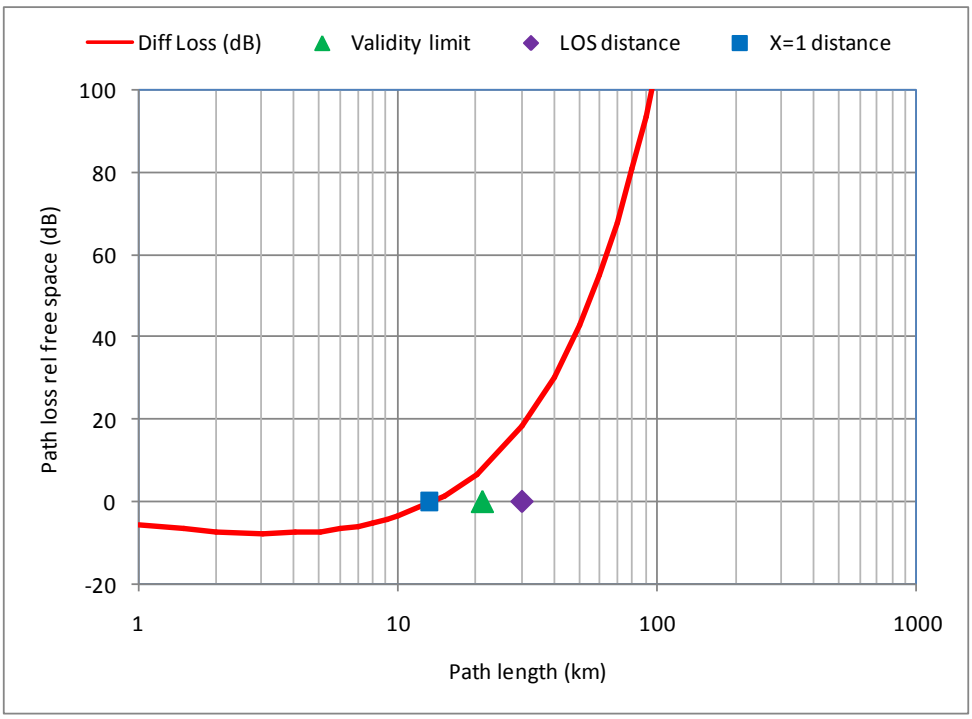
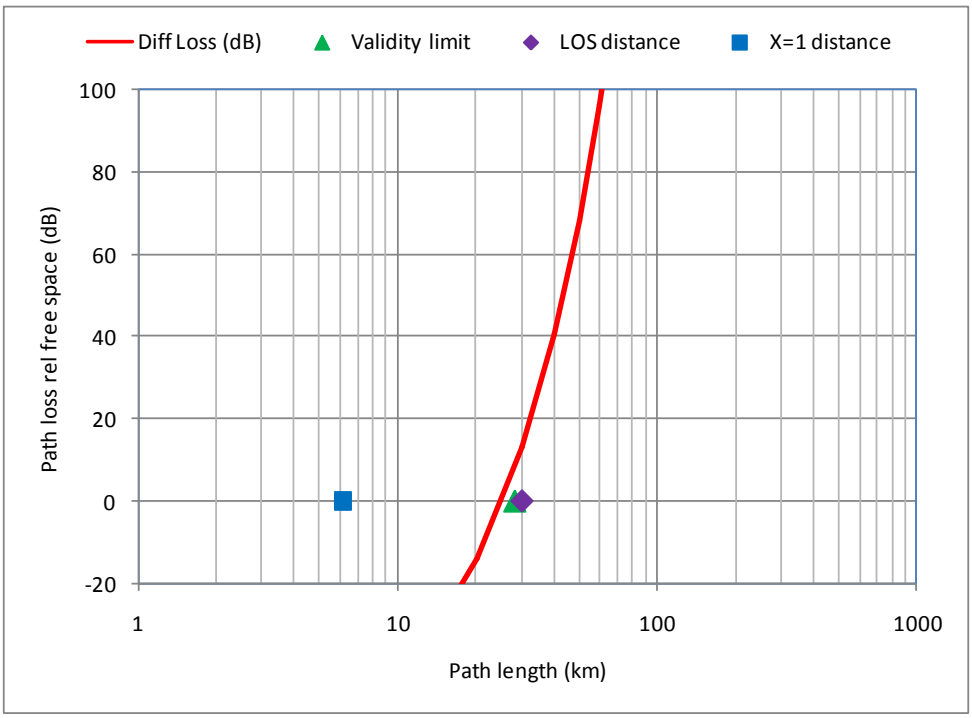


FIGURE 9

Path loss at 30GHz, antenna heights = 17 and 10m, sea, vertical pol

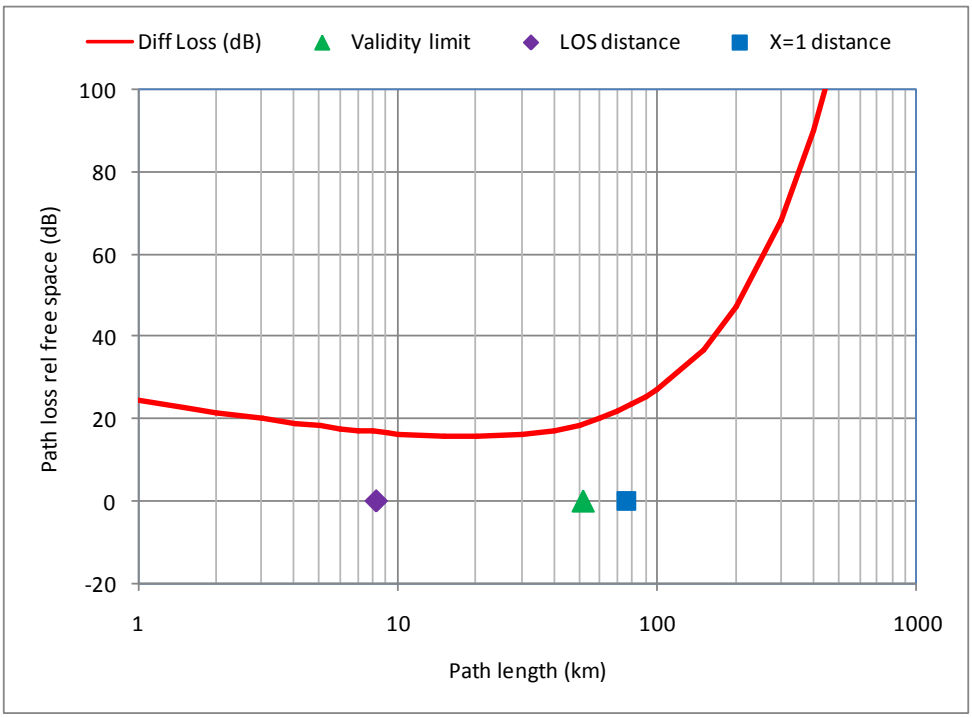


Considering the value of the predicted path loss at the line-of-sight distance, note that at the lowest frequencies (3 and 30MHz) the line-of-sight distance is less than the minimum valid range of the model, so the prediction should not be trusted. At the higher frequencies, the prediction is within the validity of the model. The predictions give a positive path loss relative to free space at the line-of-sight distance, the magnitude of the loss decreasing with frequency. Note that the losses are greater than 6dB which would be predicted by a single knife-edge Fresnel model. This is expected, as diffraction losses at grazing incidence over a rounded obstacle are higher than over a knife-edge.

In Figures 5–9 the limit of validity was always quite close to the line-of-sight distance. That this is not always true is illustrated in Figure 10 which is the same as Figure 6 (30MHz) but with both antennas at 1m. Here the validity limit (and indeed, the path loss curve) changes little, but of course the line-of-sight distance is much less, at only 8km. In this case, the model is only valid far into the shadow region.

FIGURE 10

Path loss at 30MHz, antenna heights = 1 and 1m, sea, vertical pol



7 P.526 nomograms

Section 3.1.2 of P.526-10 contains a set of nomograms for calculating the field strength for diffraction round the spherical Earth. The text notes that the nomogram method is "under the same approximation condition (the first term of the residue series is dominant)" as in the rest of Section 3.1. The calculation is separated into distance and height gain terms:

$$20\log\frac{E}{E_0} = F(d) + H(h_1) + H(h_2) \tag{28}$$

The functions $F(d)$ and $H(h)$ are calculated by means of nomograms. The similarity of eqn (28) to eqn (1) suggests the identification of the functions $F(d)$ and $H(h)$ of P.526-10 Section 3.1.2 with the functions $F(X)$ and $G(Y)$ of P.526-10 Section 3.1.1, although this is not actually stated.

In ITU-R Document 3J/34 an example was presented (Document 3J/34 Figure 2) which showed that the nomogram $F(d)$ gave an excellent match to $F(X)$ for values of $X > 1$, but that it disagreed for values of $X < 1$. In the $X < 1$ regime, $F(d)$ was shown to match better to a modified version of $F(X)$ where the sign of the $10\log(X)$ term in eqn (5) was changed from plus to minus. In further examples of Document 3J/34 it was shown that the overall diffraction loss predicted by the nomogram for $X < 1$ was a reasonable match to the GRWAVE results.

As has been shown above, the P.526-10 Section 3.1.1 method is certainly *not valid* for distances much less than line-of-sight (although it has also been shown that the line-of-sight distance is not the same as the point at which $X = 1$). But why does the nomogram method appear to work at line-of-sight ranges?

P.526 is rather inconsistent on this point. In Note 1 of Section 3.1.2, P.526-10 does state that if the nomogram method "gives a value above the free-space field, the method is invalid". So formally,

the validity range of the nomogram method is consistent with that of the numerical method. On the other hand, the nomograms for calculating the distance term $F(d)$ do give field strengths that are up to 20dB above free space, and Document 3J/34 has shown that the overall prediction method can give above line-of-sight predictions that are consistent with GRWAVE.

From ITU-R Recommendation 526-1/Report 715-2 (1986) the source of the nomograms appears to be Boithias [7]. It has not been possible to obtain this book to check whether the nomograms are sourced directly from it, and whether there is an explanation of the values in the line-of-sight region. The nomograms predate GRWAVE, but there were of course other computer implementations of the residue series available. Ground wave curves existed in Recommendation 368 before GRWAVE, the main contribution of GRWAVE being the modelling of an exponential, rather than a linear, atmosphere. So it is possible that the "line-of-sight" extension of the nomograms were obtained from numerical results or from the Rec 368 ground wave curves.

This doesn't provide a definitive explanation of why the P.526 nomograms appear to give reasonable results in the line-of-sight region. Consequently it is felt that the comment in P.525-10, that the nomogram results are invalid when the predicted field strength is above the free space level, should be retained. Indeed it may be prudent to strengthen this caveat by stating that the results are invalid if the value of the *distance function* itself is greater than zero. This would be compatible with the assumptions underlying the formulae of P.526-10 Section 3.1.2. If the nomogram is retained in its current form, an explanation for the difference between the nomogram and the formulae of Section 3.1.2 should be found and made explicit in P.526.

8 Conclusions and proposals

This document has considered a number of issues raised in ITU-R Document 3J/34. In particular the issue of whether P.526-10 Equation (17) is correct was answered in the affirmative. This equation was re-derived from formulations given in TN101, and also from the basic equations of mode theory. The criticism in Document 3J/34 of the predictions of P.526-10 Equation (17) within the line-of-sight region is unfounded, since the method of P.526-10 Section 3.1 is only meant to be used "at long distances over the horizon".

This conclusion does not necessarily invalidate a proposal made in Document 3J/34 that a modified form of Equation (17) be used for values of the normalised distance parameter $X < 1$. However, it has been shown that the quantity X is probably not the correct one to use as a switch between the long range and short range versions of Equation (17).

In reaching these conclusions, a number of issues concerning P.526-10 Section 3.1.1.2 were addressed, and consequently some changes to P.526 are proposed. The proposed modifications are given in the Annex, covering the following:

(a) A minor change to P.526-10 Equations (14a) and (15a) to increase the significance of the numerical constants 2.2 and 9.6 in the definition of X and Y. In this study, the low accuracy of these numbers caused some difficulty in getting agreement with other sources, and it seems unnecessary to round to only 2 significant figures.

(b) A correction to the β function of P.526-10 Equation (16);

(c) A correction to the definition of the $G(Y)$ height gain term of P.526-10 Equation (18);

(d) The addition of formulae that define the validity of the model;

9 References

[1] Rice P.L., Longley A.G., Norton K.A. and Barsis A.P. "Transmission loss predictions for tropospheric communications circuits", National Bureau of Standards, Technical Note 101 Vol 1, rev January 1, 1967

[2] Kerr .D.E. (ed) "Propagation of short radio waves", Chapter 2, Peter Peregrinus Ltd, 1987

[3] Craig K.H., "Roots of the mode equation for propagation in an elevated duct", Fourth International Conference on Antennas and Propagation (ICAP 85), U.K., *IEE Conf. Pub.* **248**, pp 274–278, 1985

[4] Wait J.R. and Spies K.P. "Internal guiding of microwaves by an elevated tropospheric layer", Radio Science, **4**, pp 319-326, 1969

[5] Longley A.G. and Rice P.L. "Prediction of tropospheric radio transmission loss over irregular terrain, a computer method–1968", Annex 3, ESSA Research Laboratories, ERL79-ITS67, US Government Printing Office, Washington D.C., 1968

[6] Abramowitz M. and Stegun I.A. "Handbook of mathematical functions", Chapter 10, Dover Publications, 1972

[7] Boithias L. "Propagation des ondes radioélectrique dans l'environnement terrestre", 2nd edition, Dunod, Paris, 1984

Annex

Proposed revision to Recommendation ITU-R P.526-10

In Section 3.1.1.2 make the following changes:

(a) Replace Equations (14a) and (15a) with the following:

"
$$X = 2.188\beta f^{1/3} a_e^{-2/3} d \qquad (14a)$$

$$Y = 9.575\times 10^{-3}\beta f^{2/3} a_e^{-1/3} h \qquad (15a)"$$

(b) Replace Equation (16) with the following:

"
$$\beta = \frac{1+1.6K^2+0.67K^4}{1+4.5K^2+1.53K^4} \qquad (16)"$$

(c) Replace Equations (18) to (18c) inclusive with the following:

"
$$G(Y) \cong 17.6(B-1.1)^{1/2} - 5\log(B-1.1) - 8 \qquad \text{for } B > 2 \qquad (18)$$

For $B < 2$ the value of $G(Y)$ is a function of the value of K computed in § 3.1.1.1:

$$G(Y) \cong 20\log(B+0.1B^3) \qquad \text{for } 10K < B < 2 \qquad (18a)$$

$$G(Y) \cong 2 + 20\log K + 9\log(B/K)\left[\log(B/K)+1\right] \qquad \text{for } K/10 < B < 10K \qquad (18b)$$

$$G(Y) \cong 2 + 20 \log K \qquad \text{for } B < 10K \tag{18c}$$

where

$$B = \beta Y \tag{18d}$$

The accuracy of the diffracted field strength given by equation (13) is limited by the approximation inherent in only using the first term of the residue series. Equation (13) is accurate to better than 2 dB for values of X, Y_1 and Y_2 that are constrained by the formula:

$$X - (\beta Y_1)\Delta(Y_1, K) - (\beta Y_2)\Delta(Y_2, K) > X_{\lim} \tag{19}$$

where

$$X_{\lim} = 1.096 - 1.280\,(1 - \beta) \tag{19a}$$

$$\Delta(Y, K) = \Delta(Y, 0) + 1.779\,(1 - \beta)\left[\Delta(Y, \infty) - \Delta(Y, 0)\right] \tag{19b}$$

$\Delta(Y,0)$ and $\Delta(Y,\infty)$ are given by:

$$\Delta(Y, 0) = 0.5\left[1 + \tanh\left(\frac{0.5 \log(\beta Y) - 0.255}{0.3}\right)\right] \tag{19c}$$

$$\Delta(Y, \infty) = 0.5\left[1 + \tanh\left(\frac{0.5 \log(\beta Y) + 0.255}{0.25}\right)\right] \tag{19d}$$

(d) Renumber subsequent equations in P.526.