ELSEVIER

Scopus

doi

Journal of Vibration Engineering

ISSN:1004-4523

Registered

- SCOPUS
- GOOGLE SCHOLAR
- DIGITAL OBJECT IDENTIFIER (DOI)
- IMPACT FACTOR 6.1

Our Website
www.jove.science

Estimation and Random Forest Prediction of Effective Earth-Radius Factor under Standard and Moderate Anomalies at Microwave Antenna Height in Port Elizabeth, South Africa

O. A. Layioye[1*], Y. B. Lawal[2], P. A. Owolawi[3], C. Tu[4], E. Van Wyk[5], and J. S. Ojo[6]

[1,2,3,4]*Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, Pretoria 0152, South Africa*

[5]*Faculty of Information and Communication Technology, Tshwane University of Technology, Pretoria 0152, South Africa*

[6]*Department of Physics, The Federal University of Technology Akure 340110, Nigeria*

ORCID: 0009-0003-7243-5994

Abstract: *The design and performance of terrestrial microwave and millimeter-wave links for 5G and beyond depend strongly on the variability of the effective Earth radius factor,* k*, which compactly represents the combined effects of atmospheric refraction and Earth curvature. Coastal regions, in particular, often deviate substantially from the standard* $k = 4/3$ *assumption because of strong temperature and humidity gradients that give rise to sub-refraction, super-refraction and ducting. This paper presents an eight-year (2016–2023) climatology and data-driven model of the k-factor at 65 m over Port Elizabeth, South Africa, using hourly meteorological data derived from reanalysis-based profiles. First, full-range statistics, probability density functions, cumulative distributions and regime frequencies are used to characterize the occurrence of standard, sub-refractive, super-refractive and ducting conditions. The distribution is found to be strongly heavy-tailed, with a small fraction of hours exhibiting extreme* k *values that are physically meaningful but numerically ill-conditioned for regression. Based on physical and design considerations, a modelling design range of* $0.3 \leq k \leq 3.0$ *is adopted, retaining over 88% of the data and corresponding to standard and moderate anomalous refraction. The overall mean refractivity gradient at 65 m is about –57 N-units* km^{-1}*, corresponding to an annual mean* $k \approx 1.22$*, indicative of predominantly super-refractive conditions. Seasonally, summer (DJF) exhibits the largest mean* k *(≈1.8) due to enhanced low-level moisture and stronger refractivity stratification, while winter (JJA) shows the lowest mean (≈0.4), consistent with drier, more stable boundary-layer conditions and more frequent strong-gradient events. Spring and autumn yield intermediate means (≈1.2–1.5), reflecting transitional thermal and moisture structures. Within this range, detailed diurnal, monthly and seasonal analyses show that Port Elizabeth is predominantly super-refractive, with enhanced anomalies around sunrise and during late autumn to winter. A Random Forest regression model driven by near-surface (10 m) and 65-m thermodynamic predictors achieves* $R^2 \approx 0.77$*, RMSE ≈ 0.21 and MAPE ≈ 6.8% on an independent test set. Feature-importance analysis reveals that vertical gradients of water-vapour pressure, relative humidity and potential temperature between the surface and 65 m are the dominant controls on* k *variability. The resulting model provides a practical tool for locally consistent k-factor estimation in Port Elizabeth, while the rare, extreme ducting regimes are reserved for subsequent classification-based analyses and explicit fade-margin studies at sub-millimeter wave and microwave frequencies.*

Keywords: Effective Earth-Radius factor, Refractivity Gradient, Tropospheric Propagation, Surface Ducting, Super-refraction, Terrestrial Microwave Link, Random Forest, 5G/6G Backhaul.

1. INTRODUCTION

Design and performance analysis of terrestrial microwave and millimetre-wave links for 5G and next generation 6G communication networks, depend critically on the state of the lower

* Corresponding Author

troposphere, particularly the behaviour of the radio refractive index and its vertical gradient near the Earth's surface [1]-[3]. The effective Earth-radius factor, k, is a key radio-meteorological parameter used in ITU-R P.530 to represent the combined influence of atmospheric refraction and Earth curvature on line-of-sight propagation [4]. It provides a convenient way to account for ray bending by replacing the true Earth with a fictitious sphere of radius kR_e, where R_e is the physical Earth radius. Under standard atmospheric conditions, the near-surface refractivity gradient is about -40 N-units km^{-1}, giving the widely used effective Earth-radius factor of $4/3$. However, coastal and tropical regions often exhibit pronounced departures from this standard state due to strong temperature and humidity gradients, leading to sub-refraction, super-refraction or ducting and, consequently, large variability in k [5]-[7]. In practical terms, low-k (sub-refractive) conditions shorten the radio horizon and increase diffraction loss, whereas high-k or ducting regimes extend the horizon, enhance over-the-horizon interference and can produce deep, multipath-induced fades on terrestrial links [1] & [8]. Accurate characterization of the effective Earth-radius factor is therefore essential for determining radio horizon distances, predicting diffraction fading, and specifying fade margins for high-capacity terrestrial backhaul and access links.

A substantial body of work has examined refractivity, refractivity gradients and related propagation parameters over different climatic zones. In West Africa, Adediji *et al*. (2011) analyzed the distribution of refractivity gradient and k-factor over Akure, reporting strong seasonal modulation associated with the West African monsoon [7]. Suleman *et al*. (2025) and Sheu *et al*. (2022) extended such analyses to other Nigerian locations, showing that coastal and humid stations experience more frequent super-refractive and ducting conditions and, hence, larger variability in effective Earth radius [9] & [10]. In Nigeria, Lawal and Omotoso (2023) used ERA5 reanalysis to estimate point refractivity gradient and geoclimatic factor at 70 m in Yenagoa, demonstrating that the coastal radio-climate is dominated by super-refraction and ducting events that significantly increase predicted fade depths [11].

Similar and more recent efforts have focused on mapping the k-factor over South Africa and other climatically diverse regions. For Southern Africa, Afullo and Odedina (2006) investigated k-factor distributions and diffraction fading, highlighting how non-standard refraction can substantially alter predicted fade margins over long terrestrial paths [12]. Nyete and Afullo (2013) modelled and mapped the seasonal distribution of effective Earth-radius factor across the country [13], while Odedina and Afullo (2007) applied spatial interpolation techniques to estimate geoclimatic factor and fade depth over Southern Africa [14]. Palmer and Baker (2004, 2021) used ground-based observations to predict both monthly cumulative distributions and long-term averages of the effective Earth-radius factor over South Africa, providing valuable large-scale design curves but without explicitly resolving diurnal and low-level thermodynamic controls on k [15] & [16]. More recently, Lawal *et al*. (2025) performed statistical estimation of point refractivity gradient and geoclimatic factor at microwave antenna height in Port Elizabeth, showing that the local coastal environment is strongly super-refractive and duct-prone at low levels [17]. Outside Africa, several studies have examined atmospheric ducts and refractive regimes in other coastal and maritime climates: Mentes and Kaymaz (2007) investigated surface duct conditions over Istanbul [18]; Zhou *et al*. (2022) analyzed the spatio-temporal distribution of lower atmospheric ducts over seas adjacent to China [19]; Jiang *et al*. (2022) studied mesoscale variability of surface ducts during Santa Ana wind episodes [20]; and Gunashékar (2006) reviewed trans-horizon propagation caused by evaporation ducts [1]. Advanced retrieval techniques, such as refractivity-from-clutter methods, have also been used to infer refractivity structure and ducting from marine radar echoes [2]. Collectively, these works confirm that coastal and maritime environments exhibit complex, highly

variable k-factor and ducting climatologies with strong implications for link design and interference management.

This study aims to highlight the importance of detailed radio climatological characterisation for both Earth–space and terrestrial links, and to underscore the need to extend similar, high-resolution analyses to effective Earth-radius factor in climatically sensitive coastal zones. Despite the contributions above, there is still limited work that: (i) quantifies the full diurnal, monthly and seasonal variability of k at microwave-antenna height over a South African coastal city using long-term, hourly data; and (ii) links this variability to physically interpretable, low-level thermodynamic predictors through modern machine-learning approaches. Therefore, several key radio-climate variables will be used to perform a machine learning based modeling of the k-factor over South Africa and assess its impact on South African satellite-link.

2. Uniqueness of the Present Study

Despite these advances, several gaps remain in the study of the effect of the k-factor over Earth–space and terrestrial links. First, most existing k-factor studies over Southern Africa have relied on radiosonde data with coarse temporal resolution and limited spatial coverage, which may not adequately capture coastal processes such as sea-breeze circulations, marine boundary-layer inversions and coastal upwelling. Second, the majority of works are purely statistical and do not leverage modern machine-learning techniques to model the non-linear relationships between meteorological variables, refractivity gradient and effective Earth radius. Machine-learning methods such as Random Forests (RF) have been successfully applied in atmospheric sciences for forecasting visibility, ducting conditions and other propagation-relevant parameters because they can handle complex interactions, non-linearities and mixed-scale predictors. However, their application to the direct prediction of effective Earth-radius factor or refractivity-gradient regimes for terrestrial microwave design remains largely unexplored, particularly in coastal South African environments.

This study addresses these gaps by developing a Random-Forest-based framework for modelling the effective Earth-radius factor using refractivity gradient at 65 m over Port Elizabeth (Gqeberha), a major coastal city on the south-eastern coast of South Africa. The location is characterized by a temperate maritime climate, persistent sea-breeze systems and frequent low-level inversions, making it a natural laboratory for investigating abnormal refraction and k-factor variability. The present research is unique because it: (i) focuses on a South African coastal city that has received limited attention compared to other coastal cities, (ii) explicitly applies refractivity gradient at an operational antenna height (65 m) to determine its corresponding effective earth radius factor, and (iii) integrates machine learning to provide predictive capability and feature-importance diagnostics for a range of k-factor that spans from standard to moderately anomalous conditions. The resulting k-factor climatology and predictive models are expected to support more reliable inter-terrestrial link design, refined fade-margin specifications and improved radio-climate characterization for current and future terrestrial microwave networks in South Africa.

3. Methodology and Computational Analysis

3.1 Research Location and Data Source

This study focuses on Port Elizabeth (Gqeberha), a coastal city on the south-eastern shoreline of South Africa at latitude and longitude of approximately 33.96°S, and 25.60°E respectively. It is situated along Algoa Bay in the Eastern Cape Province. The region experiences a temperate maritime climate with pronounced sea-breeze circulations, frequent low-level inversions and sharp humidity gradients, conditions that strongly

modulate near-surface radio refractivity and abnormal refraction events [17] & [21]. Port Elizabeth is a quad-season location with four distinct seasons which are summer (December, January and February), autumn (March, April and May), winter (June, July and August) and spring (September, October and November) [22]. These features make Port Elizabeth an ideal natural test site for effective Earth-radius factor (k) studies relevant to coastal terrestrial links. Eight years hourly meteorological data covering 2016–2023 were obtained from the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5). ERA5 provides globally consistent fields of surface (10 m) and pressure-level parameters on a 0.25° × 0.25° grid, including temperature, pressure, dew-point temperature, relative humidity and wind components [23]. The grid cell whose centre is closest to Port Elizabeth was extracted, and hourly data at the surface (10 m), 1000 hPa and 975 hPa were retained. These levels were selected to span the lowest 100 m above ground, allowing robust interpolation of meteorological variables to a nominal microwave antenna height of 65 m, in line with earlier refractivity-gradient and k-factor studies.

3.2 Computation of Effective Earth Radius Factor

Hourly surface refractivity N_0 and refractivity at 65 m were computed from air pressure P (hPa), temperature T (K) and water-vapour pressure e (hPa) using equation (1) [6]:

$$N = 77.6\frac{P}{T} + 3.73 \times 10^5 \frac{e}{T^2} \quad (1)$$

where e was obtained from ERA5 dew-point temperature or relative humidity via a standard saturation vapour formulation. The hypsometric equation and hydrostatic balance were used to relate pressure levels to geometric height, and meteorological variables were vertically interpolated between the surface (10 m), 1000 hPa and 975 hPa to obtain temperature, pressure, humidity and refractivity at 65 m above ground level. The point near-surface refractivity gradient in the lowest 65 m, dN_1, was then approximated by a finite difference, dN_1 given by [6]:

$$dN_1 = \frac{N_{65} - N_S}{z_{65} - z_S} [\text{N-units km}^{-1}] \quad (2)$$

where N_s and N_{65} denote refractivity at the surface (10 m) and 65 m, respectively, and z_s and z_{65} denote altitude at the surface (10 m) and 65 m, respectively. Under the assumption of a spherically stratified atmosphere with small gradients, the effective Earth-radius factor k was obtained from the linearized relation used in ITU and related literature [6]:

$$k \approx (1 + \frac{1}{157} dN_1(\text{N km}^{-1}))^{-1}, \quad (3)$$

Equation (3) can be simply written as:

$$k \approx \left[1 + \frac{1}{157}\frac{dN}{dz}\right]^{-1} \quad (4)$$

Equation (4) links the near-surface refractivity gradient ($^{dN}/_{dz}$) to the curvature of radio rays relative to the physical Earth radius. Hourly values of k at 65 m were then used to construct diurnal, monthly, seasonal and annual statistics and to derive the 1-percentile "worst-case" values required for radio-climate analysis and machine-learning modelling.

3.3 Relationship between k-factor and Atmospheric Refraction Regimes

Atmospheric refraction regimes can be expressed either in terms of the vertical refractivity gradient dN/dz or equivalently via the effective Earth radius factor k. The approximate

relationship is given in equation (4), with dN/dz in N-units/km. Under standard refraction, $dN/dz \approx -40$ N/km and $k \approx 4/3$, which is widely used in ITU-R propagation models [6] & [24]. Sub-refraction corresponds to weakly negative or positive gradients (e.g. $dN/dz > -40$ N/km), giving $k < 4/3$, while super-refraction arises for more negative gradients (e.g. $dN/dz \approx -79$ N/km), leading to $k > 4/3$and an effectively larger Earth radius. For very strong negative gradients ($dN/dz \lesssim -157$ N/km), k can become very large or even negative, indicating ducting, where radio rays are trapped or strongly bent towards the surface. Climatological studies over African and coastal sites [7], [11], [25] & [26] confirm that transitions between sub-refraction, standard, super-refraction and ducting can therefore be compactly described in terms of characteristic k-factor ranges (typically $k < 0.7$: sub-refraction/ducting; $0.7 \lesssim k \lesssim 1.7$: standard to weakly anomalous; $k \gtrsim 1.7$: strong super-refraction), providing a convenient framework for classifying refractive conditions in radio-link design.

3.4 Variables for k-Factor Modelling

The linear correlation and Random Forest model were driven by physically motivated predictors that describe (i) near-surface meteorology, (ii) conditions at 65 m, (iii) vertical thermodynamic gradients between the surface and 65 m, (iv) water-vapour pressure and associated gradients, (v) potential temperature and static stability, and (vi) diurnal and annual time harmonics. Table 1 shows the summary of all predictor variables used for Random Forest k-factor modelling. All variables were either collected or computed from the hourly ERA5-derived and station-based fields over Port Elizabeth, South Africa.

Surface meteorology includes air temperature, dewpoint temperature, relative humidity, surface pressure and wind speed. Dewpoint depression (°C) is defined as:

$$T_{\mathrm{dd}} = T_0 - T_{d,0}, \tag{5}$$

where T_0 is surface air temperature and $T_{d,0}$ is surface dewpoint temperature.
At 65 m, analogous variables (air temperature, relative humidity and pressure) were derived from the interpolated profiles used to compute the refractivity gradient and k-factor. Vertical gradients were represented by simple level differences (65 m – surface) in temperature, relative humidity, pressure, potential temperature and water-vapour pressure. These differences capture the low-level thermal and moisture stratification that directly controls the refractivity gradient and hence the effective Earth radius factor.
Water-vapour pressure at each level (in hPa) was obtained from temperature and relative humidity as [6]:

$$e = \frac{\mathrm{RH}}{100}\, e_s(T), \tag{6}$$

with saturation vapour pressure given by:

$$e_s(T) = 6.112 \exp\left(\frac{17.62T}{243.12+T}\right), \tag{7}$$

where T is air temperature in °C.

Potential temperature (K) at each level was computed by [27]:

$$\theta = T_{\mathrm{K}}\left(\frac{p_0}{p}\right)^{R_d/c_p}, \tag{8}$$

where T_K is absolute temperature, p is pressure (hPa), $p_0 = 1000$ hPa, R_d is the gas constant for dry air and c_p is the specific heat at constant pressure. The difference $\Delta\theta_{65-10} = \theta_{65} - \theta_{10}$ serves as a low-level static-stability indicator: positive values indicate stable stratification, while negative values indicate convective or super-adiabatic conditions.

Finally, diurnal and annual cycles were encoded using sinusoidal time harmonics given as:

$$\text{Hour (sin)} = \sin\left(\frac{2\pi\ \text{hour}}{24}\right),\ \text{Hour (cos)} = \cos\left(\frac{2\pi\ \text{hour}}{24}\right), \quad (9)$$

$$\text{Day of Year (sin)} = \sin\left(\frac{2\pi\ \text{day}}{365}\right),\ \text{Day of Year (cos)} = \cos\left(\frac{2\pi\ \text{day}}{365}\right), \quad (10)$$

which provide smooth representations of daily and seasonal variability without introducing discontinuities at 23:00 or 31 December.

Table 1. Summary of predictor variables used for Random Forest k-factor modelling

Category	Variable name (used in text/plots)	Description	Units
Surface meteorology (0 m)	Air Temperature (Surface)	Screen-level dry-bulb air temperature	°C
	Dewpoint Temperature (Surface)	Screen-level dewpoint temperature	°C
	Dewpoint Depression	$T_0 - T_{d,0}$; measure of surface moisture deficit	°C
	Relative Humidity (Surface)	Screen-level relative humidity	%
	Pressure (Surface)	Surface atmospheric pressure	hPa
	Wind Speed (Surface)	Horizontal wind speed at screen level	ms^{-1}
Meteorology at 65 m	Air Temperature (65 m)	Air temperature at 65 m (absolute)	K
	Relative Humidity (65 m)	Relative humidity at 65 m	%
	Pressure (65 m)	Pressure at 65 m	hPa
Vertical thermodynamic differences (65 m – surface)	Δ Temperature (65–10 m)	Temperature difference between 65 m and surface	K
	Δ Relative Humidity (65–10 m)	Relative-humidity difference between 65 m and surface	%
	Δ Pressure (65–10 m)	Pressure difference between 65 m and surface	hPa
Water-vapour pressure & gradients	Water Vapour Pressure (Surface)	Water-vapour pressure computed from surface T_0, RH	hPa
	Water Vapour Pressure (65 m)	Water-vapour pressure at 65 m	hPa
	Δ Water Vapour Pressure (65–10 m)	Vapour-pressure difference between 65 m and surface	hPa
Potential temperature & static stability	Potential Temperature (Surface)	Surface potential temperature θ_0	K
	Potential Temperature (65 m)	Potential temperature at 65 m θ_{65}	K
	Δ Potential Temperature (65–10 m)	Stability indicator $\theta_{65} - \theta_0$	K

Time harmonics	Hour (sin)	$\sin(2\pi\,\text{hour}/24)$; diurnal phase	–
	Hour (cos)	$\cos(2\pi\,\text{hour}/24)$; diurnal phase	–
	Day of Year (sin)	$\sin(2\pi\,\text{day}/365)$; seasonal phase	–
	Day of Year (cos)	$\cos(2\pi\,\text{day}/365)$; seasonal phase	–
	Season (Summer: DJF) Season (Autumn: MAM) Season (Winter: JJA) Season (Spring: SON)	Four seasonal indices across one year-cycle	–

3.5 Application of Random Forest Algorithm

Random Forest (RF) is an ensemble of T decision trees $\{h_t(\mathrm{x})\}_{t=1}^{T}$, each trained on a bootstrap sample with random feature subsets. The RF algorithm regression equation for prediction is given as [28] & [29]:

$$\hat{y}(\mathrm{x}) = \frac{1}{T}\sum_{t=1}^{T} h_t(\mathrm{x}), \tag{11}$$

where $\hat{y}(\mathrm{x})$ is the predicted output (i.e. k) for input x (as indicated in Table 1) which is the input feature vector, T is the total number of trees in the forest, and $h_t(\mathbf{x})$ is the prediction of the t-th decision tree for input $\mathbf{x}$. Each tree minimizes node impurity (e.g., variance) when splitting.

The relationship between meteorological conditions and effective Earth-radius factor was modelled using RF regression algorithm. The target variable was the hourly k value at 65 m, while predictors included interpolated temperature, pressure, relative humidity and water-vapour pressure at surface and 65 m, and temporal descriptors such as hour of day, day of year and season. The full 2016–2023 data set was divided into training (75%) and testing (25%) subsets using a time-aware split to preserve temporal dependence. Model hyperparameters (number of trees, maximum depth, minimum samples per split and leaf) were tuned via grid search using k-fold cross-validation on the training set. Performance was evaluated on the test set using the coefficient of determination (R^2), root-mean-square error (RMSE) and mean absolute error (MAE).

Random Forest was chosen as the sole machine-learning model for three main reasons. First, RF can capture strongly non-linear and interactive relationships between predictors and k without requiring explicit functional assumptions, which is advantageous for complex coastal boundary-layer processes [30] & [31]. Second, RF is relatively robust to multicollinearity and non-Gaussian predictor distributions, and it handles large feature sets with minimal pre-processing, making it well suited to multi-level meteorological inputs [32] & [33]. Third, RF offers built-in measures of variable importance (e.g., permutation importance), which are valuable for physically interpreting which atmospheric variables most strongly influence the effective Earth radius factor, complementing earlier statistical analyses of refractivity gradient and geoclimatic factor [34] & [35]. Given these advantages, and to maintain methodological focus and interpretability, alternative algorithms (e.g., support vector regression, gradient boosting, and deep neural networks) were not pursued in this initial study but may be considered in future comparative work.

4. Results and Discussion

4.1 Overall Statistics of Effective Earth Radius Factor

Table 2 summarizes the bulk statistics of the refractivity gradient at 65 m and the corresponding effective Earth-radius factor, k, for Port Elizabeth over 2016–2023. The mean refractivity gradient is -57.5 N-units/km, with a median of -50.0 N-units/km. This is substantially more negative than the "standard atmosphere'' gradient of about -39 N-units/km, which corresponds to $k \approx 4/3$. The result indicates that, on average, the lower troposphere over Port Elizabeth is more strongly super-refractive than the ITU-R standard atmosphere, consistent with the coastal, maritime character of the site where cool, moist marine layers and strong temperature–humidity gradients are frequently observed. Similar tendencies towards more negative gradients and enhanced anomalous propagation have been reported for other tropical and subtropical coastal locations in Africa [11].

The refractivity-gradient distribution is moderately left-skewed (skewness = –1.09) and leptokurtic (kurtosis ≈ 4.56), indicating a heavier left tail than a Gaussian distribution. Physically, this means that episodes of very strong negative gradients (potentially associated with surface-based or elevated ducts) occur more often than would be expected under purely "normal'' conditions. These more negative gradients translate, through the standard relation given in equations (3) and (4) into larger deviations of the k-factor from its nominal value.

In contrast, the raw statistics of the k-factor in Table 2 reveal an extremely heavy-tailed distribution. While the median k is 1.43, close to the classical standard value of about 1.33 – 1.5 reported in several climatological studies [7], [26] & [36], the mean is only 1.22 because it is pulled down by very large-magnitude negative outliers. The minimum and maximum values ($k_{\min} \approx -1.43 \times 10^4$, $k_{\max} \approx 4.1 \times 10^3$) are far outside the range normally used in link-design handbooks and ITU-R recommendations [4]. This behaviour is reflected in the extreme skewness (–121) and kurtosis (~2.3×10⁴), implying that a very small number of hours correspond to "critical'' gradients for which the denominator in the k-factor expression approaches zero and the resulting k becomes very large in magnitude. Similar extreme-value behaviour has been noted in other k-factor climatologies based on point refractivity gradients [7] & [36].

Table 2: Overall descriptive statistics of refractivity gradient and k-factor for Port Elizbeth at 65 m across 2016 to 2023

Variable	dN/dz at 65 m (N-units/km)	k-factor
Count	70124	70124
Mean	-57.465	1.217
Std	46.931	72.36
Min	-435.394	-14279.0
25%	-75.623	1.234
50%	-49.977	1.429
75%	-32.803	1.812
Max	276.939	4108.01
Skewness	-1.0854	-121.294
Kurtosis	4.5585	23333.74

The full-range probability density function in Figure 1 clearly illustrates this heavy-tailed structure. On the log-scaled ordinate, most hours cluster tightly around modest positive k-values near unity, while a few isolated bins appear far out on the negative and positive axes. These rare events correspond to strongly ducting or highly super-refractive situations where the ray curvature far exceeds or falls well below the Earth's curvature. Under such conditions, EM waves can be trapped in surface or elevated ducts, producing over-the-horizon propagation and severe departures from standard link-budget predictions.

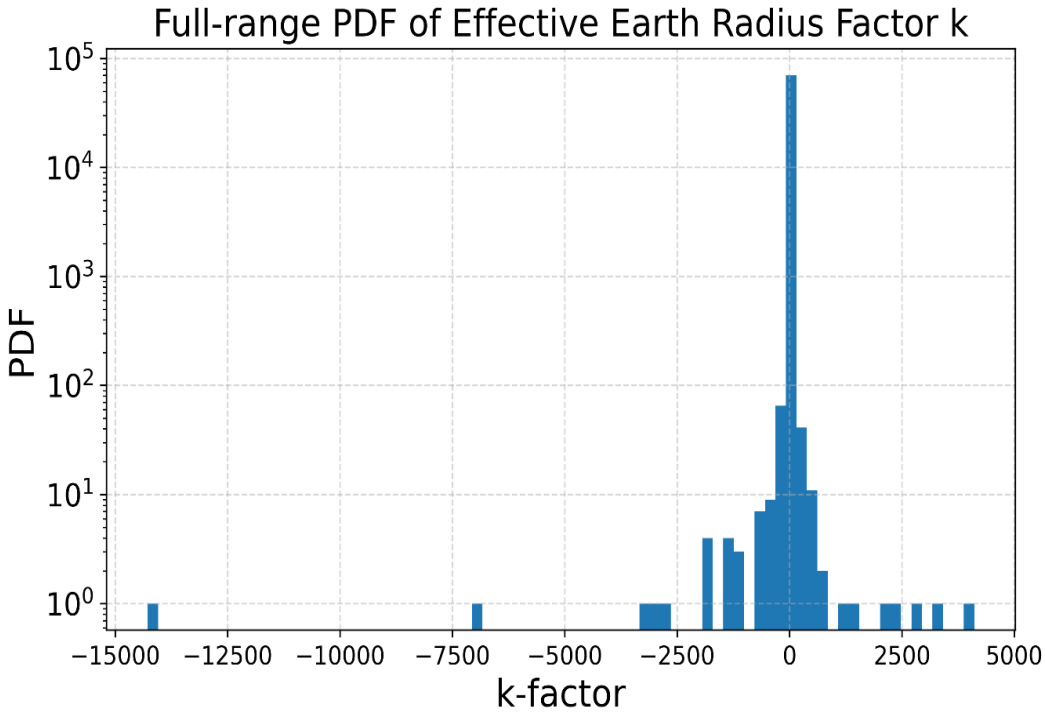

Figure 1: Full-range PDF of the Effective Earth Radius factor (k-factor) for Port Elizabeth over the years 2016 to 2023.

Table 3 and Figure 2 provide a complementary regime-based view by classifying each hour into ducting, sub-refraction, standard refraction, and super-refraction based on its k-factor. The analysis shows that ducting conditions occur in about 3.9% of hours, whereas super-refraction is the most frequent regime, accounting for almost half (≈49.4%) of the observations. Standard refraction is realized in about 25.6% of hours, and sub-refraction in roughly 21.0% of hours. This regime distribution confirms that Port Elizabeth is dominated by non-standard refractive conditions, with a marked bias towards super-refractive and ducting episodes. Such behaviour is typical of coastal and marine environments where strong low-level humidity and temperature inversions are common, especially under anticyclonic conditions or when cool marine air undercuts warmer continental air.

Table 3: k-factor refraction-regime frequency for Port Elizabeth over the entire period

Regime	**Count (Hours)**	**Percentage of Occurrence (%)**
Ducting	2719	3.88
Sub-refraction	14759	21.05
Standard	17977	25.64
Super-refraction	34669	49.44

From a propagation perspective, sustained super-refraction and occasional ducting are particularly important for terrestrial microwave links. Super-refractive layers bend radio waves more strongly towards the Earth, effectively increasing the apparent Earth radius (large positive k), extending radio-horizon distances, and potentially increasing interference between neighbouring cells or co-channel systems. Ducting (negative k) can trap energy within a narrow vertical layer, enabling over-the-horizon propagation but also causing deep fades and multipath effects when the duct intersects the link path. At the other extreme, sub-refractive conditions (very low k) bend rays away from the Earth, effectively reducing the radio horizon and enhancing diffraction fading; ITU-R P.530 explicitly links such low-k episodes to increased clearance losses and fade depths on terrestrial microwave paths [4].

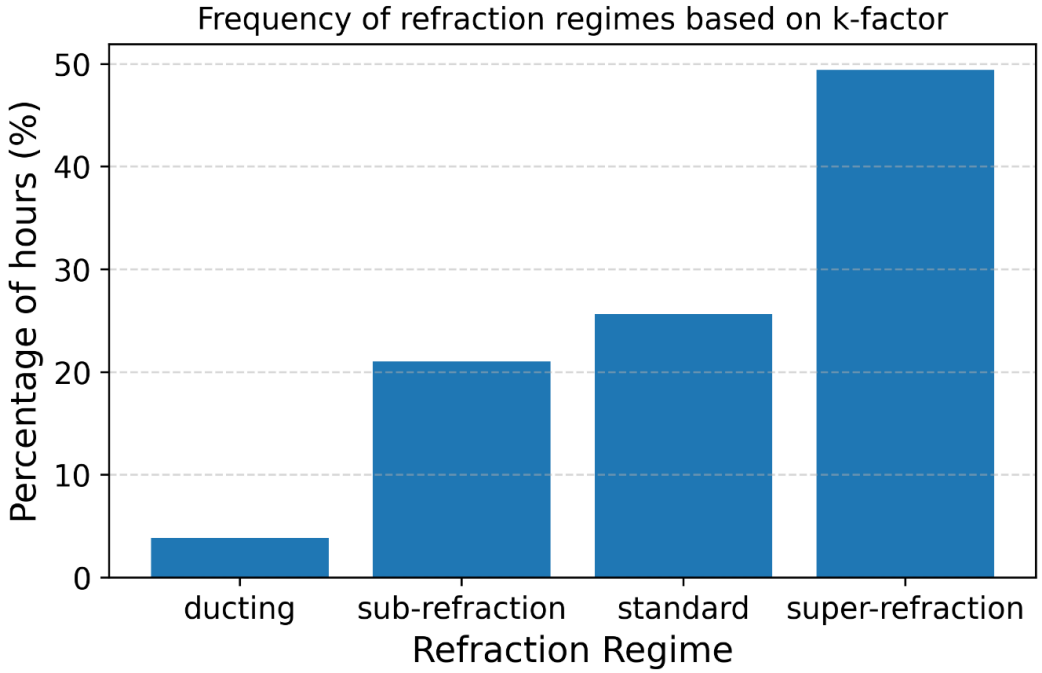

Figure 2: Histogram chart of the frequency of refraction regimes based on k-factor for Port Elizabeth over the years 2016 to 2023.

4.2 Design range selection for k-factor Prediction

The combination of Tables 2 to 3 and Figures 1 to 2 therefore provides the physical rationale for the modelling approach adopted in this study. The full-range k -factor distribution is dominated by a small fraction of extreme ducting and near-critical events which yield physically meaningful but numerically ill-conditioned k values. When these extremes are included in regression modelling, they act as high-leverage outliers that severely degrade the performance of Random Forest and other supervised-learning models. At the same time, Figures 1 and 2 show that the bulk of operationally relevant propagation, that is, standard to moderately anomalous refraction, occurs for k values clustered around order-unity magnitudes.

Figures 3 and 4 zoom into the central portion of the k-factor distribution ($-80 \leq k \leq 80$) and superimpose the proposed design bounds at $k = 0.3$ and $k = 3.0$. In Figure 3, the probability density function reveals a sharply peaked core centred slightly above $k = 1$, with rapidly decaying tails. The dashed vertical lines at $k = 0.3$ and $k = 3.0$ clearly bracket the high-density region where most hourly observations lie. Outside this interval

the PDF collapses towards zero, indicating that very low or very high k-values occur only rarely. The corresponding cumulative distribution in Figure 4 confirms this behaviour: the CDF rises very steeply between $k \approx 0.3$ and $k \approx 3$, indicating that the vast majority of hours are within this range, while the lower and upper tails contribute only a small fraction of the total probability mass.

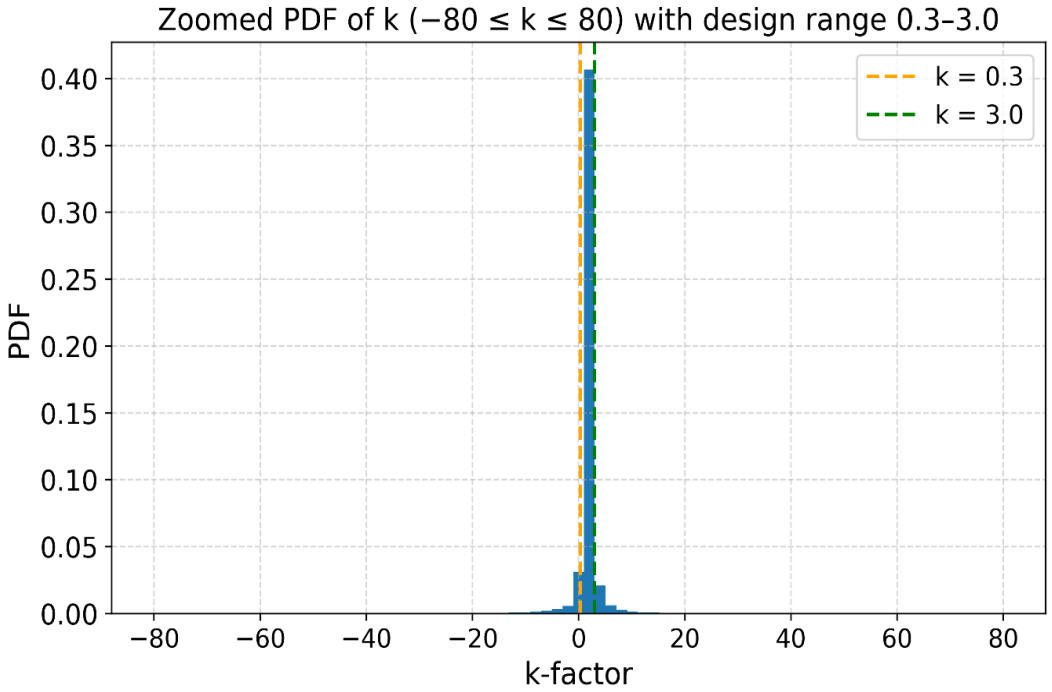

Figure 3: Probability Density Function plot showing k-factor (–80 ≤ k≤ 80) at 65 m with design range 0.3 to 3.0.

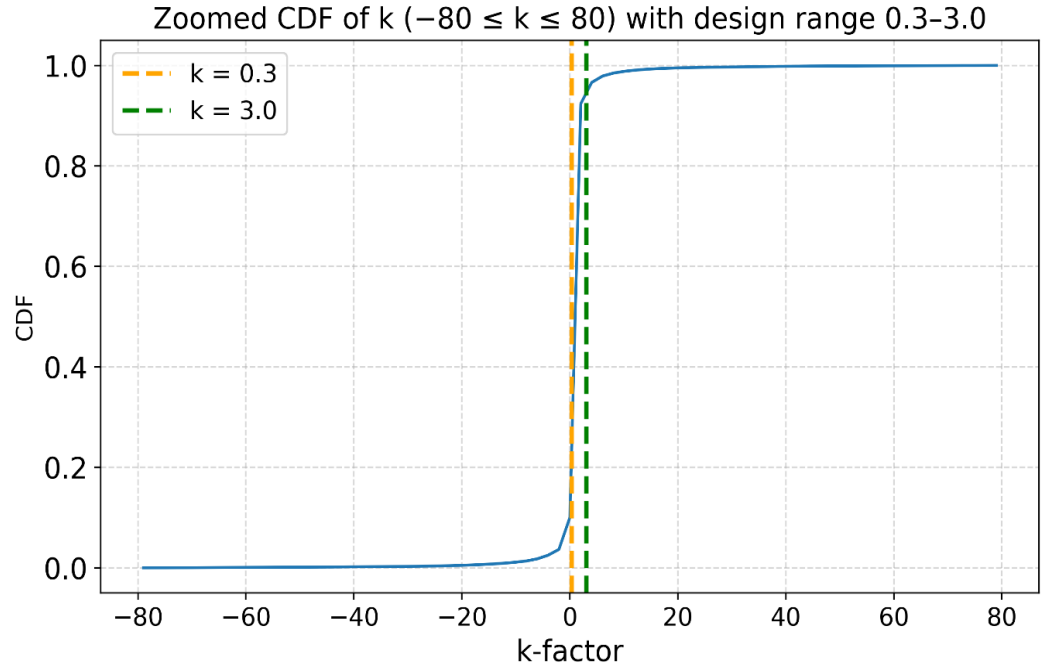

Figure 4: Cumulative Distribution Function plot showing k-factor (–80 ≤ k≤ 80) at 65 m with design range 0.3 to 3.0.

Table 4 and Figure 5 quantify this behaviour in terms of simple regime frequencies. Only about 3.9% of the hours exhibit $k < 0.3$, which correspond to very strong sub-refraction and ducting conditions; around 7.9% of the hours have $k > 3.0$, associated with pronounced super-refraction and highly extended radio horizons. In contrast, approximately 88.2% of all hours fall within $0.3 \leq k \leq 3.0$. Thus, the proposed design window retains almost nine out of every ten observations, while excluding only a small minority of extreme cases that are responsible for the pathological tails in the full-range statistics.

Table 4: Frequency of k-factor regimes relative to the selected design range at 65 m for Port Elizabeth (2016–2023).

Refraction Regime	Count (Hours)	Percentage (%)
k < 0.3	2719	3.88
$0.3 \leq k \leq 3.0$	61840	88.19
k > 3.0	5565	7.94

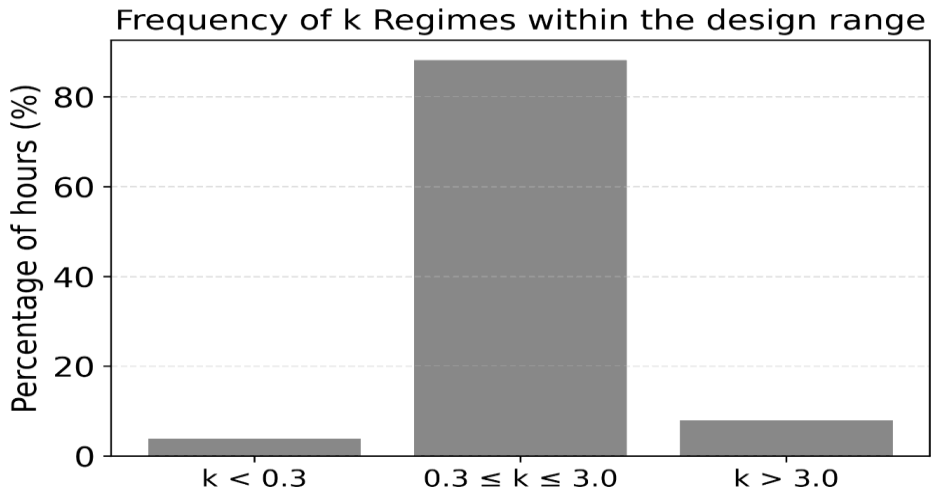

Figure 5: Frequency of the k-factor regimes within the selected design range of 0.3 to 3.0 at 65 m.

From a physical standpoint, the lower bound $k = 0.3$ removes the most severe sub-refractive and ducting episodes in which rays are strongly bent away from, or trapped near, the surface and the apparent Earth curvature become unrealistically large in magnitude. The upper bound $k = 3.0$ excludes hours with very strong super-refraction, where the effective Earth radius is several times the geometric radius and the radio horizon is pushed far beyond normal design expectations. These rare regimes are crucial for interference and over-the-horizon studies, but they are not representative of the typical operating conditions that dominate the long-term performance of a 5G inter-terrestrial link.

Statistically, restricting the regression to $0.3 \leq k \leq 3.0$ therefore achieves two objectives. First, it focuses the Random Forest model on the dense, well-sampled region of the distribution where the relationship between surface and 65-m meteorology and k is most stable and where accurate prediction is most beneficial for routine link engineering. Second, it prevents a small set of extreme points from dominating the loss function and feature-importance metrics, leading to more robust and interpretable models. The rarer, more extreme k-factor events ($k < 0.3$ or $k > 3.0$) are explicitly not discarded from the climatological analysis; instead, they are reserved for a follow-up classification-based study

in which their occurrence, physical drivers, and fade-margin implications can be treated using methods tailored to highly non-Gaussian, regime-switching behaviour.

Following the practice in previous k-factor climatologies and ITU-R line-of-sight design methods, which typically focus on moderate ranges such as $k \approx 0.7$–3 for reliability and clearance calculations, this study therefore concentrates the regression analysis on a design range of $0.3 \leq k \leq 3.0$. This range retains the vast majority of hourly samples (over 88% of the dataset) and encompasses standard and moderate anomalous conditions that are most relevant for routine terrestrial radio link design in Port Elizabeth.

4.3. Diurnal Variations of Effective Earth Radius Factor

Figure 6 presents the overall diurnal cycle of the mean effective Earth radius factor, k, at 65 m for the full 2016 to 2023 period. The mean k-factor remains predominantly super-refractive (i.e. $k > 1$) for most hours of the day, with mean values typically between about 1 and 2, consistent with the coastal–maritime setting of Port Elizabeth where moist marine air and shallow inversions frequently enhance refractivity gradients. The most pronounced departures from this background occur around local sunrise (≈06:00 to 08:00), when the mean curve exhibits a sharp negative excursion followed by a strong positive spike. These features are produced by a small number of early-morning ducting events with very negative refractivity gradients (low or even negative k) followed by rapidly strengthening super-refraction as the nocturnal boundary layer begins to erode. Similar sunrise "transition signatures'', with enhanced occurrence of ducts and strong super-refraction in the first hours after sunrise, have been reported in other tropical and subtropical sites [37], [38] & [39]. Although such extremes are rare in an absolute sense, they strongly distort the hourly mean, which is why the regression modelling later in the study is restricted to the more stable design range $0.3 \leq k \leq 3.0$.

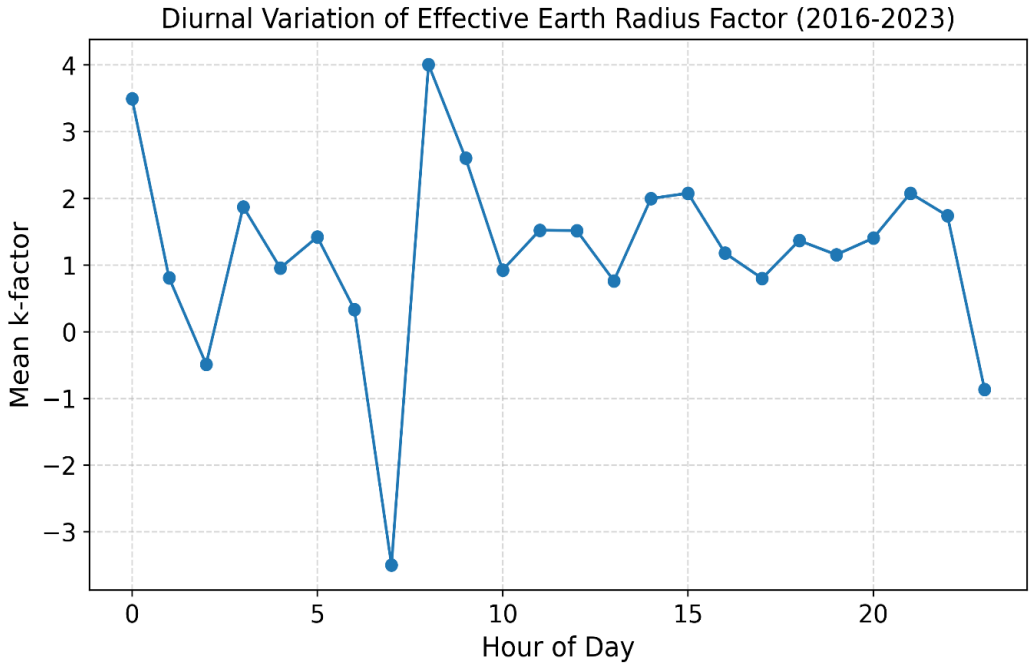

Figure 6: Overall diurnal variations of the mean Effective Earth Radius k-factor at 65 m for Port Elizabeth across the entire period under study.

The seasonal breakdown in Figure 7 highlights how this diurnal behaviour is modulated by synoptic-scale forcing. In all seasons, nighttime and early-morning hours (00:00–06:00) are generally more super-refractive than the afternoon, reflecting the formation of stable, moist layers close to the surface under radiative cooling. Winter months (JJA) show the most dramatic anomalies: at around 07:00 the mean k-factor plunges to strongly negative values, indicating frequent strong ducting during the morning transition under clear, high-pressure conditions typical of South African coastal winters. Autumn season (MAM) exhibits a contrasting peak with very large positive mean k near 08:00, suggesting occasional intense super-refraction when cool marine air undercuts warmer continental air during sea-breeze onset. Spring season (SON) is comparatively benign, with a smoother diurnal cycle and mean k-values remaining mostly between 1 and 2, indicative of weaker inversions and stronger daytime mixing. In general, summer months (DJF) relatively display a stable diurnal cycle. Overall, these seasonal patterns are consistent with previous studies showing that coastal stations tend to experience stronger and more frequent anomalous refractive conditions in the cool, dry season than in the warm, convective season [11].

While Figures 6 and 7 are based on the full k-factor range and therefore reflect the influence of extremes on the mean, Figure 8 focuses explicitly on the design interval $0.3 \leq k \leq 3.0$ through seasonal cumulative probability distribution curves (CPDCs). All four curves lie within a relatively narrow band, confirming that the central portion of the k-distribution is broadly similar across seasons and dominated by moderately super-refractive conditions. The 50-percent (median) k-values lie near 1.4–1.6 in all seasons, comparable to the 1.4–1.6 range reported for other coastal and low-latitude sites [7] & [38]. The lower decile of the distribution (≈10% of time) typically falls between $k \approx 1.0$ and 1.2, while the upper decile (≈90% of time) is around $k \approx 2.0$–2.2, indicating that, within the design range, the atmosphere dwells most of the time in mildly to strongly super-refractive states rather than near the standard $k = 4/3$ value.

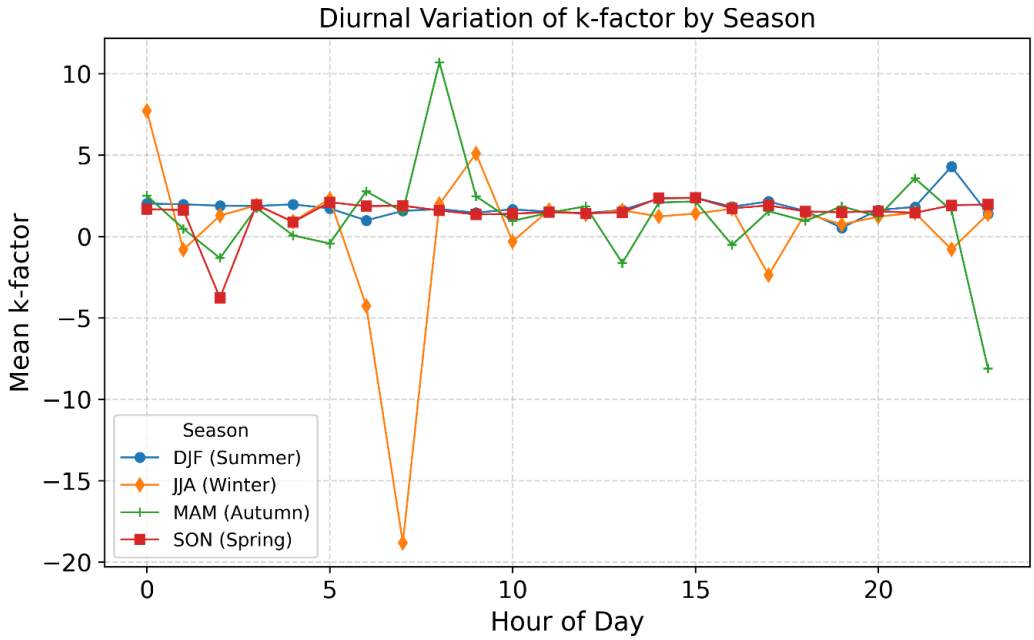

Figure 7: Diurnal variations of the mean k-factor by season for Port Elizabeth.

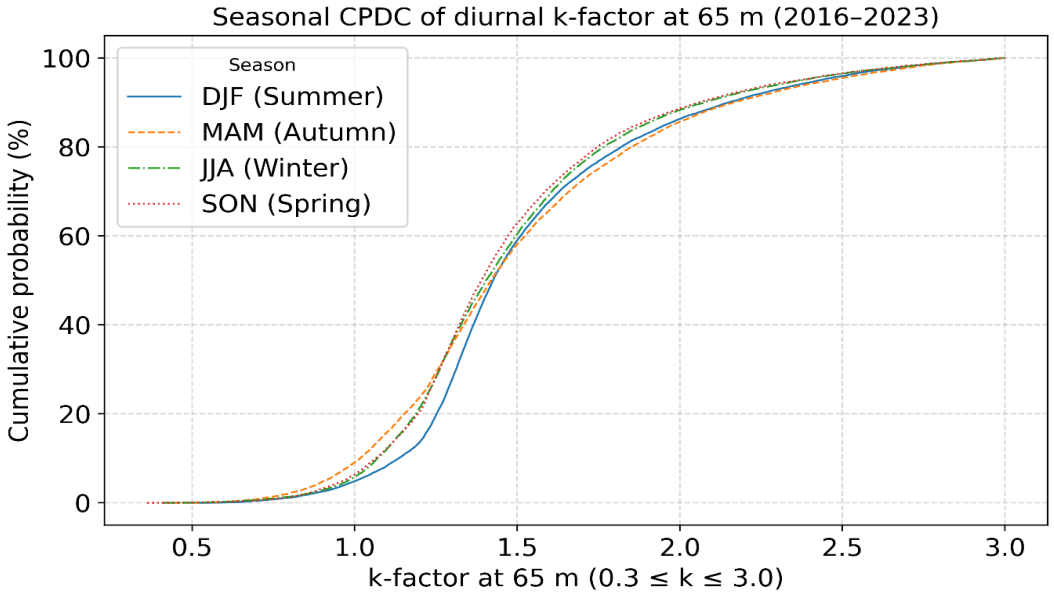

Figure 8: Seasonal Cumulative Probability Distribution Curve of diurnal Variations of k-factor at 65 m over the design range 0.3 to 3.0, for Port Elizabeth across the entire period under study.

In general, Figures 6 to 8 show that: (i) Port Elizabeth exhibits a pronounced diurnal cycle in k, with the most anomalous conditions occurring around sunrise; (ii) winter and, to a lesser extent, autumn are the seasons most prone to extreme ducting and intense super-refraction; and (iii) when attention is restricted to $0.3 \leq k \leq 3.0$, the effective Earth radius factor remains within a relatively well-behaved super-refractive band in all seasons. These results reinforce the choice of the 0.3 to 3.0 design range for Random Forest modelling in this study. It captures the dominant diurnal and seasonal modulation of k that controls routine link performance, while leaving the rarer, more severe ducting regimes, most prominent in winter sunrise hours, for a dedicated classification-based analysis in subsequent work.

4.4 Monthly and Seasonal Variations of Effective Earth Radius Factor

4.4.1 Monthly Variations of Effective Earth Radius Factor

Table 5 and Figures 9 to 11 summarize the monthly behaviour of the k-factor at 65 m over Port Elizabeth for 2016–2023 using the full data range. The monthly means shown in Table 5 and Figure 10 generally lie between about 0.2 and 2.6, indicating that, on average, the troposphere is super-refractive all year, in agreement with the overall statistics discussed earlier. However, the very large standard deviations and extreme values in Table 5 reveal that each month contains a mixture of benign and highly anomalous refractive conditions.

The heatmap in Figure 9 makes this contrast especially clear. The row corresponding to the mean shows that March, November and December have the largest monthly mean k-values (≈2–2.6), while May and July exhibit much lower means (0.2 and –0.92, respectively). Physically, the summer to autumn months (DJF–MAM) are characterized by warmer, more humid boundary layers and frequent sea-breeze intrusions, which tend to increase near-surface refractivity and favour super-refraction. By contrast, late autumn and winter (May–July) are dominated by cooler and often drier continental air masses with stronger nocturnal inversions; these can generate both strong ducts and sub-refractive conditions, resulting in

lower or even negative monthly mean k-values despite the predominance of super-refractive hours.

The rows for standard deviation, skewness and kurtosis in Figure 9 and Table 5 highlight how strongly the monthly distributions are influenced by a small number of extremes. The month of July, for example, has a standard deviation of 197, a minimum of about -1.43×10^4 and a maximum exceeding 3.3×10^3, with strongly negative skewness (–63) and very large kurtosis ($\sim4.6\times10^3$). Similar but less intense behaviour is seen in March, May, August and September. These statistics indicate that most July hours fall within a relatively narrow, mildly super-refractive band, but a few hours with very large-magnitude k-values dominate the higher moments. Such rare events correspond to near-critical refractivity gradients in which the denominator of the standard k-factor formula becomes very small, yielding extremely high positive or negative k. This pattern is consistent with previous k-factor climatologies in coastal and tropical environments, where extreme ducting or super-refractive layers are infrequent but have a disproportionate statistical impact.

Table 5: Overall monthly descriptive statistics of k-factor for Port Elizabeth across the years 2016 to 2023

Month	Count	Mean	Std	Min	25%	50%	75%	Max	Skewness	Kurtosis
Jan	5948	1.92	7.28	-75.37	1.28	1.43	1.80	460.21	45.41	2718.00
Feb	5424	1.88	33.06	-1104.07	1.30	1.48	1.89	2033.07	36.13	2875.38
Mar	5952	2.57	58.00	-1421.53	1.27	1.47	1.87	4108.01	57.52	4284.96
Apr	5760	0.92	31.68	-1781.89	1.17	1.41	1.84	606.95	-33.54	1874.76
May	5952	0.20	110.61	-7024.73	1.14	1.43	1.93	2738.86	-44.63	2918.70
Jun	5760	1.33	25.45	-1106.31	1.19	1.43	1.89	696.18	-16.62	901.30
Jul	5952	-0.92	197.13	-14278.9	1.18	1.43	1.84	3366.12	-62.98	4645.95
Aug	5952	0.75	47.50	-2740.04	1.21	1.39	1.73	1297.47	-38.66	2319.43
Sep	5760	1.18	42.09	-3125.00	1.21	1.38	1.69	342.38	-70.98	5288.55
Oct	5952	1.26	20.42	-1264.36	1.23	1.42	1.80	174.31	-46.66	2664.34
Nov	5760	1.97	11.96	-125.09	1.26	1.42	1.73	557.98	36.03	1519.18
Dec	5952	1.63	7.27	-234.03	1.29	1.44	1.80	321.87	1.36	1057.18

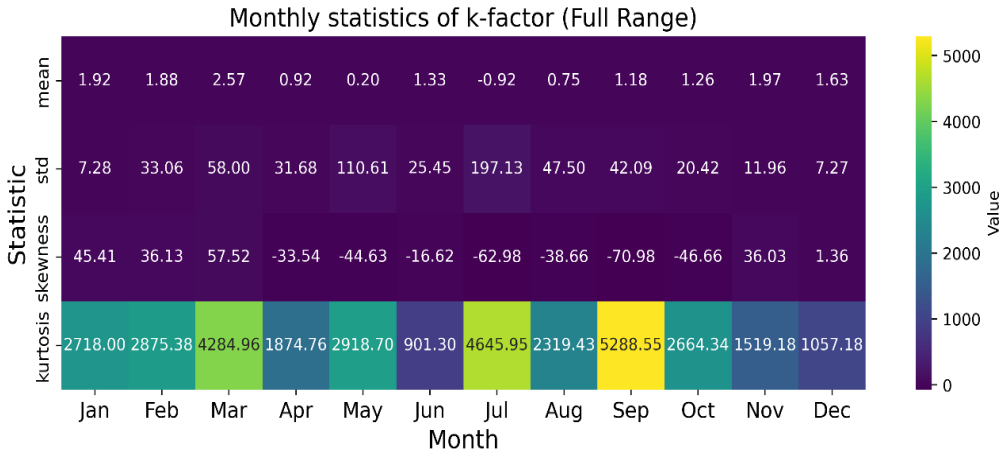

Figure 9: Monthly statistics heatmap of k-factor at 65 m for Port Elizabeth over the full range of data across 2016 to 2023.

Figure 10 shows the monthly mean k-factor as a bar chart and mirrors the heatmap's first row. The highest mean values occur in March and November–December, suggesting that late summer and early summer are the periods when effective Earth radius is largest, and

radio horizons are most extended on average. The lowest means occur in May to July, implying that 5G terrestrial links may be more vulnerable to sub-refractive fades and rapid changes in path clearance during late autumn and mid-winter.

Figure 11 presents the monthly CPDCs on a logarithmic probability axis for the full k-factor range. For all months, the CDFs rise steeply near modest positive k -values, indicating that the bulk of each monthly distribution is concentrated in a narrow band around $k \approx 1$–2. However, several months, particularly May, July and September, show small but distinct probability "tails'' extending to very large negative or positive k. On the log scale, these tails correspond to probabilities well below 1% but with large magnitudes on the abscissa. This confirms that extreme ducting and super-refraction are climatologically rare but present in most months, and that their inclusion in regression modelling would introduce strong non-Gaussian behaviour and numerical instability. Similar heavy-tailed monthly CPDCs have been reported in other subtropical coastal locations.

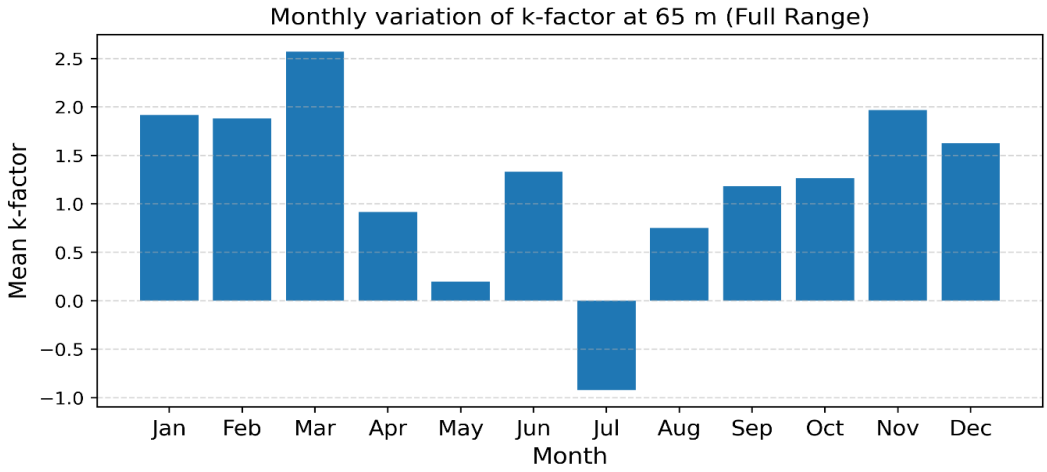

Figure 10: Monthly variation of k-factor at 65 m for Port Elizabeth over the full range of data across 2016 to 2023.

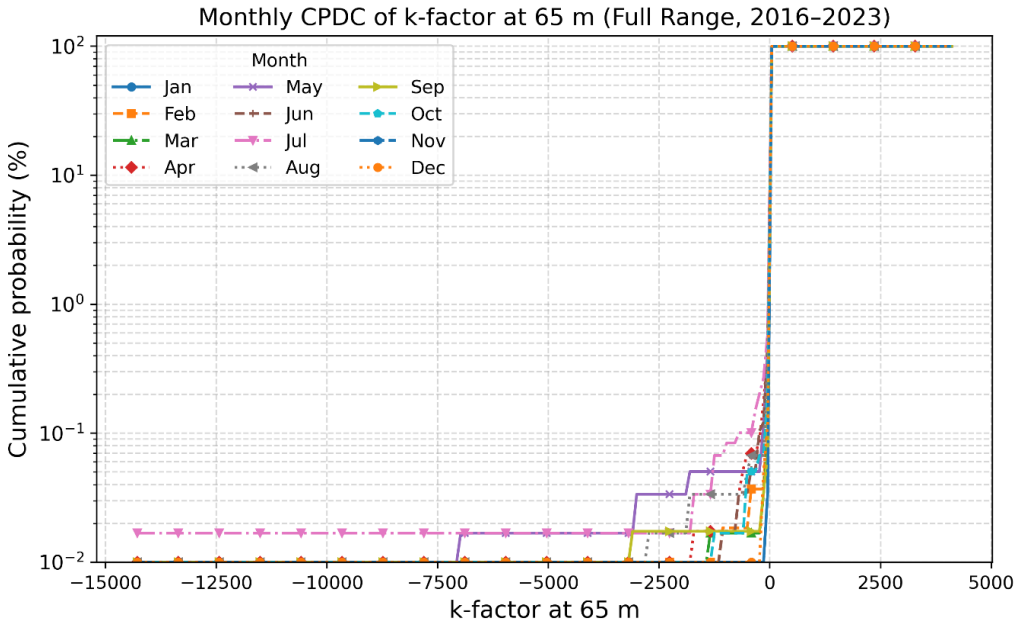

Figure 11: Monthly Cumulative Probability Distribution Curve of k-factor at 65 m for Port Elizabeth over the full range of data across 2016 to 2023.

Overall, the monthly analysis reinforces two key conclusions. First, Port Elizabeth exhibits super-refractive conditions in all months, with late summer and early summer being the most favourable for extended radio horizons. Second, extreme k-factor events occur in many months but with very low frequency, which justifies treating them separately (via classification modelling) while focusing regression on the more densely populated design range $0.3 \leq k \leq 3.0$ in this work.

4.4.2 Seasonal and Annual trends of Effective Earth Radius Factor

Figure 12 summarizes the seasonal variation of the mean k-factor using the full range of data. The seasonal means are all greater than unity, confirming that Port Elizabeth is super-refractive on average in every season. Summer (DJF) has the highest mean k (~1.8), followed by spring (SON, ~1.5) and autumn (MAM, ~1.2), while winter (JJA) exhibits the lowest seasonal mean (~0.37). This seasonal ordering is consistent with the monthly statistics, where during summer and spring, enhanced low-level moisture and frequent marine inversions over the Indian Ocean promote super-refractive gradients, increasing the effective Earth radius. In winter, the prevalence of high-pressure systems, clear skies and strong nocturnal cooling over land favours both ducting and sub-refractive conditions, which reduce the seasonal mean despite the frequent occurrence of super-refraction at night and around sunrise.

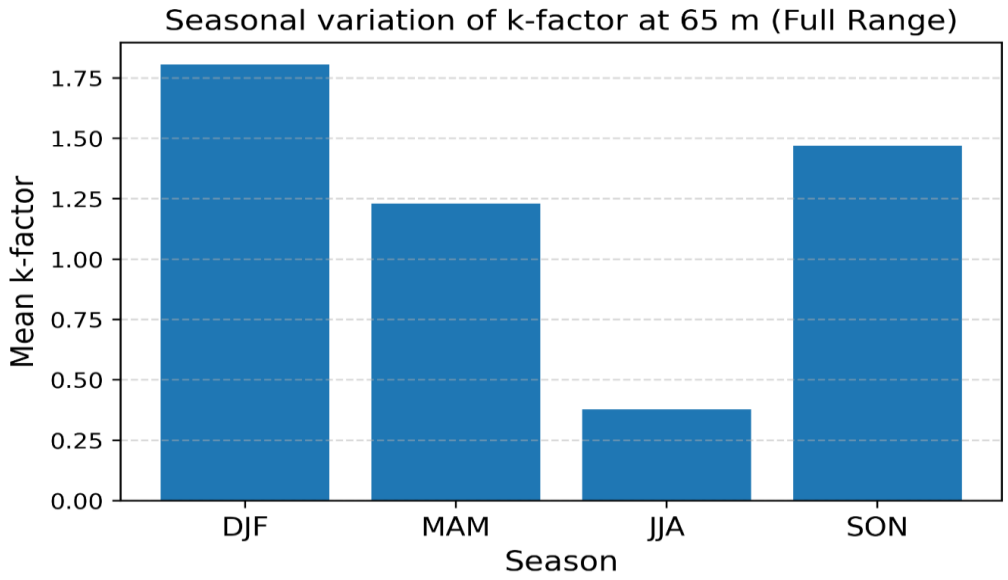

Figure 12: Seasonal Variation of k-factor at 65 m for Port Elizabeth over the full range of data across 2016 to 2023.

The annual trend in Figure 13 shows that the mean k-factor varies from year to year but remains generally super-refractive. After a slightly negative mean in 2016, the annual average rises to about 1.5–1.7 in 2017–2019, dips again in 2020, and then returns to values above 1.6 in 2021–2023. This interannual variability likely reflects a combination of changes in large-scale circulation patterns (e.g., El Niño–Southern Oscillation phase, regional pressure anomalies) and local mesoscale processes (e.g., sea-breeze characteristics, frequency of coastal inversions) that modulate the frequency and intensity of anomalous refractive conditions [18], [19] & [20]. Although a detailed attribution is beyond the scope of this study, the absence of any long-term downward trend suggests that super-refractive conditions, and thus extended radio horizons and potential interference paths, are a persistent feature of the local climate over the study period.

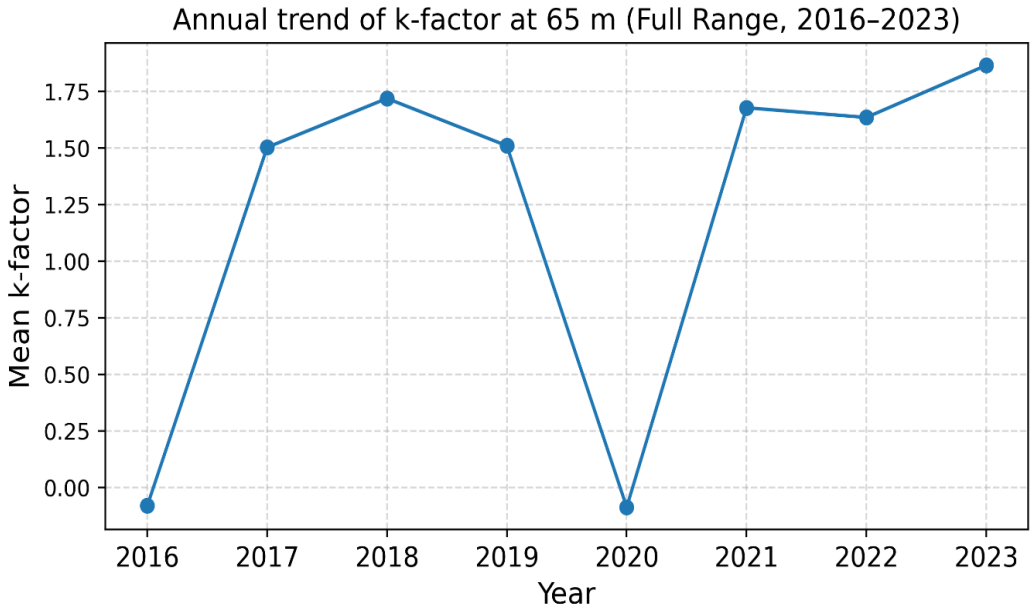

Figure 13: Annual trend of k-factor at 65 m for Port Elizabeth over the full range of data across 2016 to 2023.

From an engineering standpoint, the seasonal and annual statistics in Figures 12 and 13 emphasize that link-design margins for 30-GHz terrestrial systems in Port Elizabeth cannot be based solely on a fixed "standard'' $k = 4/3$ value. Instead, link budgets and clearance analyses should account for the strong seasonal modulation of k, with particular attention to winter and late autumn, when the mean k is lowest and the probability of extreme ducting or sub-refraction is highest. ITU-R P.530 [4] and related studies on k-factor and refractivity gradients similarly recommend using locally derived statistics rather than a single global standard when assessing path clearance and fade risk. In summary, the monthly, seasonal and annual analyses demonstrate that Port Elizabeth is a super-refractive coastal site with pronounced temporal structure: strong super-refraction in summer and spring, reduced mean k and enhanced extremes in winter, and modest interannual variability. These features buttress the design-range selection adopted for the Random Forest modelling in this study.

4.5 Linear Correlation of k-factor to the atmospheric input variables

Figure 14 shows the Pearson correlation matrix between the k-factor (within the design range $0.3 \leq k \leq 3.0$) and all atmospheric input variables defined in the methodology (surface and 65-m temperature, dewpoint temperature, relative humidity, pressure, wind speed, their vertical differences, water-vapour pressure and potential temperature, together with diurnal and annual harmonics and seasonal indicators).

Overall, the linear correlations between k and any single predictor are weak to at most moderate, with the largest absolute coefficients associated with vertical gradients in humidity, water-vapour pressure and potential temperature between the surface and 65 m, while single-level quantities (e.g. surface temperature or pressure alone) show only small correlations with k. This is physically consistent with the dependence of refractivity gradient on vertical moisture and temperature structure rather than on absolute values at one level. The relatively modest linear coefficients also confirm that no single variable can explain the observed k-factor variability, thereby justifying the use of multivariate, non-

linear Random Forest modelling in the next subsection rather than relying on simple linear regression with one or two predictors.

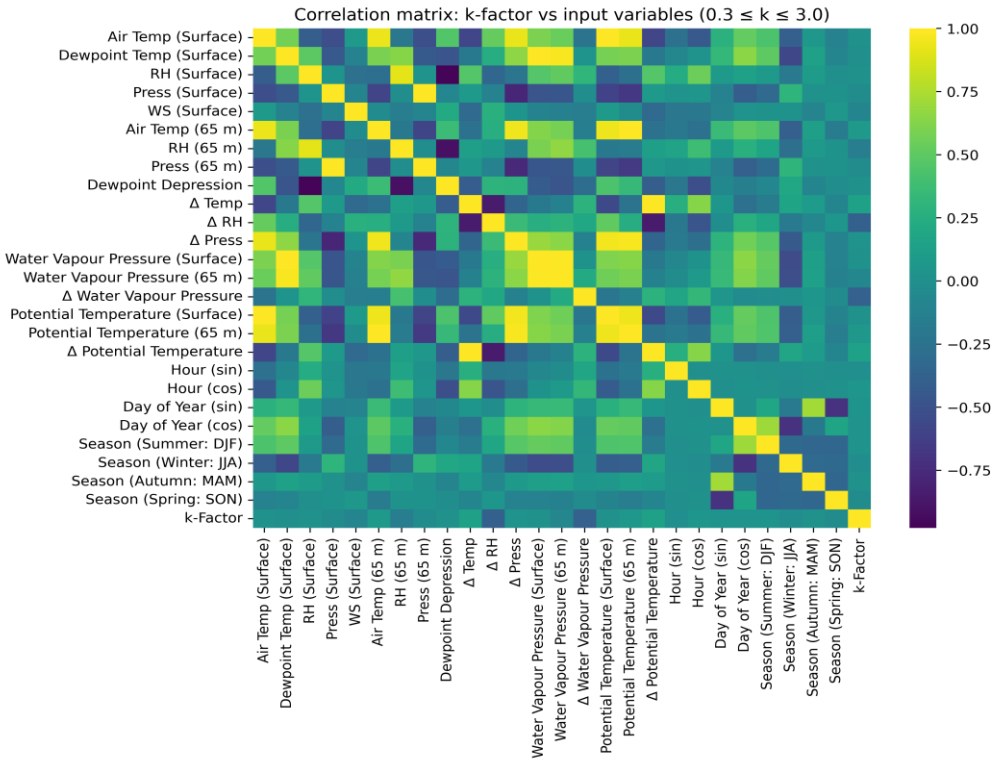

Figure 14: Linear correlation matrix of k-factor against the input variables under the design range 0.3 to 3.0.

4.6 Random Forest Prediction and Modeling of Effective Earth Radius Factor

Table 6 summarizes the performance of the Random Forest (RF) regression model applied to the design range $0.3 \leq k \leq 3.0$ using the physically based predictors described in the methodology. The model attains a coefficient of determination $R^2 = 0.77$ on the independent test set, with RMSE ≈ 0.21 and MAE ≈ 0.12 in k-units, corresponding to a mean absolute percentage error of about 6.8%. The cross-validation statistics (mean CV $R^2 \approx 0.75$ with a small standard deviation ≈0.04) indicate that the model skill is stable across different time folds and not the result of overfitting.

Table 6: Performance metric analysis of the Random Forest prediction of k-factor within the design range $(0.3 \leq k \leq 3.0)$ at 65 m

Metric	Value
R^2	0.7661
RMSE	0.2120
MAE	0.1155
MAPE	6.8317
CV of mean R^2	0.750309
CV of mean R^2	0.041907

The scatter plot in Figure 15 further illustrates the model's skill. Most test points cluster tightly around the 1:1 line, especially in the core range $k \approx 1.0–2.2$, showing that the RF

successfully captures the bulk of k-factor variability driven by low-level moisture and temperature structure. There is some widening of the scatter at the highest and lowest k-values within the design range, with a tendency to slightly underestimate the largest peaks and overestimate the smallest troughs, which is typical of ensemble tree models in the presence of skewed target distributions. Nevertheless, the residual spread remains modest in absolute terms, confirming that the design-range restriction has effectively removed the most problematic extremes while preserving physically meaningful variability for regression.

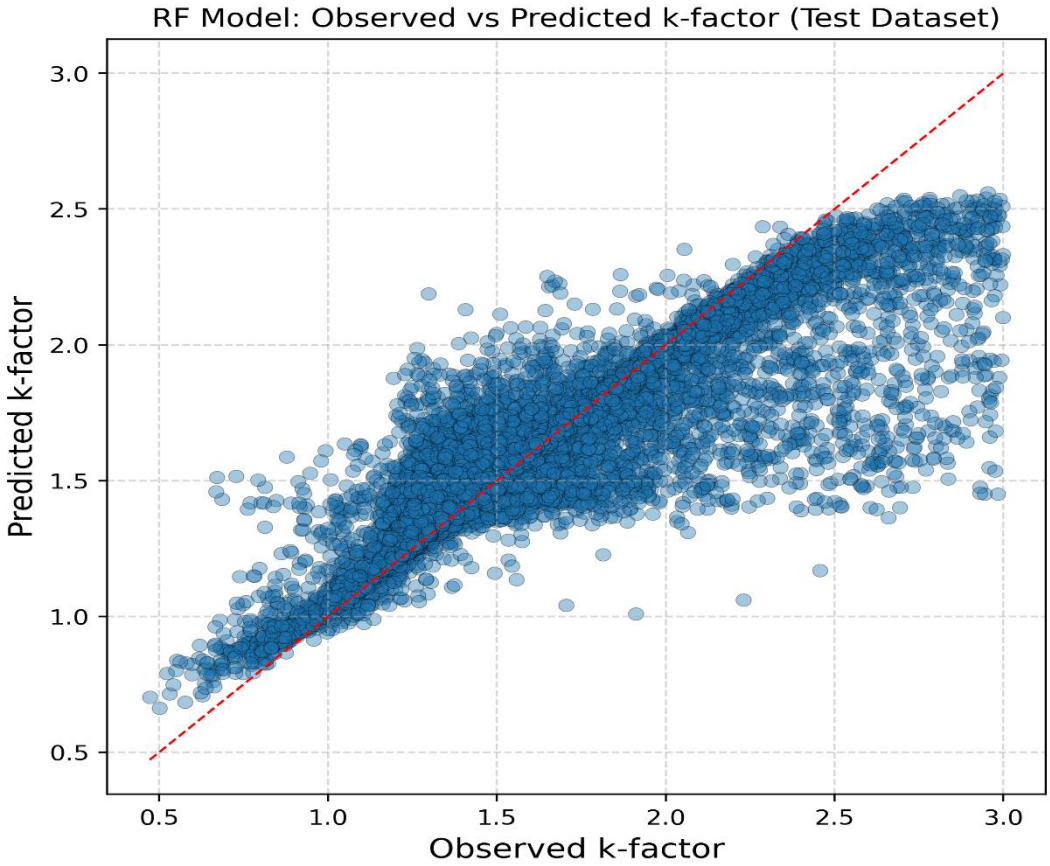

Figure 15: Random Forest regression plot of the observed vs RF-predicted k-factor within the design range $(0.3 \leq k \leq 3.0)$ at 65 m for Port Elizabeth, over the test dataset (2022-2023).

The time-series comparison for December 2023 in Figure 16 provides a more intuitive view of model behaviour on operational time scales. The RF-predicted k-factor closely tracks the observed hourly fluctuations throughout the month, reproducing both gradual multi-day changes and rapid intra-day oscillations associated with diurnal heating, sea-breeze passages and frontal activity. Peaks and dips in the observed series are generally well captured, with only minor smoothing of the sharpest transitions. This indicates that, given only surface and 65 m meteorological data, the RF model can reliably anticipate day-to-day and hour-to-hour variations in effective Earth radius factor within the standard–moderate anomaly regime.

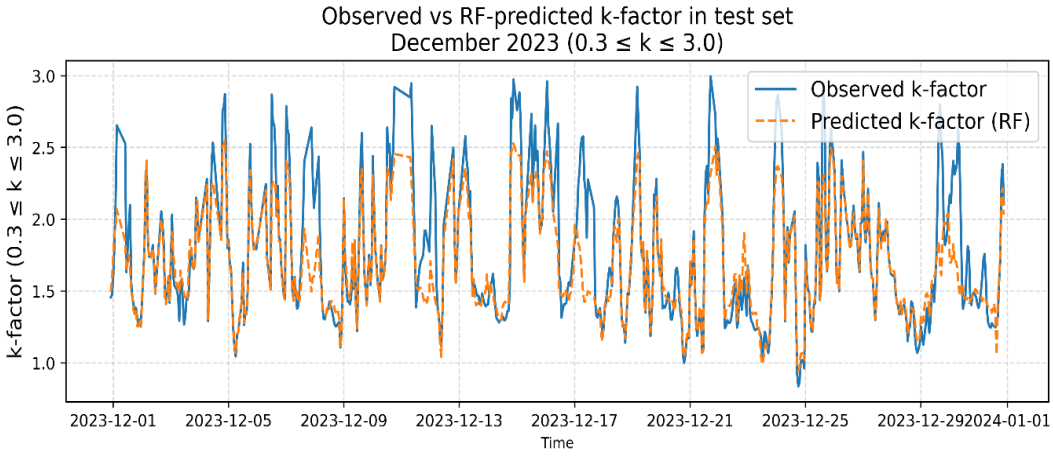

Figure 16: Time series prediction plot of observed vs RF-predicted k-factor within the design range $(0.3 \leq k \leq 3.0)$ at 65 m for Port Elizabeth, over December 2023.

In summary, the RF results show that: (i) physically motivated low-level predictors are sufficient to explain roughly three-quarters of the variance of k in the design range; (ii) the model generalizes well in time; and (iii) the remaining errors are likely attributable to unresolved vertical structure above 65 m, mesoscale dynamics and measurement noise. These limitations motivate the more advanced classification and higher-altitude feature augmentation proposed for the companion papers, but they do not detract from the usefulness of the present RF model as a practical tool for k-factor estimation and link-design studies over Port Elizabeth.

4.2. Feature Importance of Input Variables

Table 7 and Figure 17 show that the Random Forest model is primarily controlled by vertical moisture and humidity structure between the surface and 65 m. The most influential predictors are:

- Change in Water Vapour Pressure (65–0 m) with importance ≈ 0.32 and
- Change in Relative Humidity (65–0 m)having importance ≈ 0.23,

together explaining more than half of the total importance. This is fully consistent with the refractivity expression $N \propto P/T + e/T^2$, where large vertical gradients of water vapour e and relative humidity generate strong refractivity gradients and therefore large excursions in the k-factor.

The static pressure terms at 65 m and at the surface (≈0.07 each) form the next tier, reflecting the contribution of dry-air stratification to the total refractivity gradient. Vertical gradients of potential temperature and air temperature (Δθ and ΔT, respectively) also rank highly (≈0.05 each), highlighting the role of thermal stability in modulating k, especially during nocturnal inversions and early-morning transition periods.

Single-level quantities such as RH (surface and 65 m), dewpoint depression, water-vapour pressure at each level, and wind speed each carry modest but non-negligible importance (≈0.01–0.03), indicating that they refine the prediction but do not control it in isolation. Harmonic time descriptors (sine and cosine of the hour and day-of-year) and the four seasonal variables contribute very little individually (<0.01), implying that once the instantaneous thermodynamic profile is known, explicit calendar information adds only marginal extra magnitude. Overall, the feature-importance pattern confirms that k-factor

variability in the design range is governed chiefly by low-level moisture and stability gradients rather than by absolute surface conditions or season alone, in agreement with previous k-factor and refractivity-gradient studies.

Table 7: Feature importance analysis of each input variable of the Random Forest prediction

Feature	Importance
Δ Water Vapour Pressure	0.3193
Δ Relative Humidity	0.2290
Pressure (65 m)	0.0720
Pressure (Surface)	0.0714
Δ Potential Temperature	0.0553
Δ Temperature	0.0537
RH (65 m)	0.0294
Δ Pressure	0.0164
Dewpoint Depression	0.0154
RH (Surface)	0.0150
Potential Temperature (65 m)	0.0149
Potential Temperature (Surface)	0.0131
Air Temp (65 m)	0.0128
WS (Surface)	0.0127
Air Temp (Surface)	0.0122
Water Vapour Pressure (65 m)	0.0106
Day of Year (cos)	0.0092
Water Vapour Pressure (Surface)	0.0083
Dewpoint Temp (Surface)	0.0077
Hour (cos)	0.0076
Day of Year (sin)	0.0071
Hour (sin)	0.0035
Season (Summer: DJF)	0.0012
Season (Autumn: MAM)	0.0010
Season (Winter: JJA)	0.0006
Season (Spring: SON)	0.0005

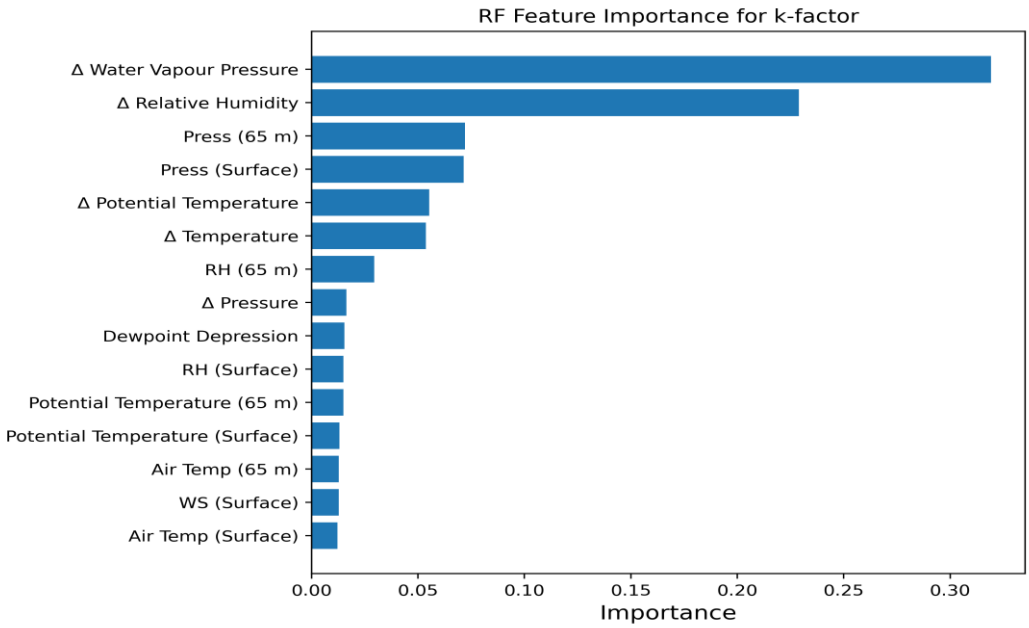

Figure 17: Random Forest tree model plot of the feature importance for the estimated k-factor within the design range $(0.3 \leq k \leq 3.0)$ at 65 m.

5. Conclusion

This study has developed an eight-year (2016–2023) climatology and Random Forest model of the effective Earth radius factor, k, at 65 m over Port Elizabeth using hourly surface and 65-m meteorological observations. The full-range statistics show that the site is super-refractive on average all year, but that a small number of hours are characterized by very large positive or negative k values associated with intense ducting or near-critical refractivity gradients. These extremes inflate skewness and kurtosis in almost every month and season, making full-range regression modelling numerically unstable and physically difficult to generalize. To obtain a robust predictive model for routine 5G or next generation 6G terrestrial link design, the analysis therefore focused on a design range $0.3 \leq k \leq 3.0$, which retains more than 88% of the data and corresponds to standard and moderate anomalous refraction. Within this range, diurnal and seasonal analyses revealed that mildly to strongly super-refractive conditions dominate, with enhanced anomalies around sunrise and during late autumn–winter, in agreement with previous k-factor and refractivity-gradient studies over coastal and tropical regions.

Using physically motivated predictors derived from near-surface thermodynamic structure, the Random Forest model achieved $R^2 \approx 0.77$ on an independent test set, with RMSE ≈ 0.21 and MAPE ≈ 6.8%. Feature-importance analysis demonstrated that vertical gradients of water-vapour pressure, relative humidity and potential temperature between the surface and 65 m are the dominant controls on k variability, while single-level quantities and seasonal indicators play secondary roles. These results confirm that low-level moisture and stability are key drivers of effective Earth radius factor over Port Elizabeth and that machine-learning models based solely on routinely available near-surface measurements can provide accurate, hour-by-hour estimates of k within the most operationally relevant range.

Future work will extend this framework by incorporating higher-altitude profile information, performing detailed fade-margin analysis at sub-millimeter wave or microwave frequencies, and applying dedicated classification models to the rare but critical ducting regimes excluded from the present regression.

Acknowledgments

The authors wish to appreciate Tshwane University of Technology for the supports provided during the course of this project.

REFERENCES

[1] S. D. Gunashékar, "Transhorizon radiowave propagation due to evaporation ducts", Resonance, vol. 11, no. 1, (2006), pp. 51–62. https://doi.org/10.1007/BF02834828.
[2] C. Tepecik and H. Erol, "Atmospheric refractivity estimation from radar sea clutter using refractivity from clutter technique", Radio Science, vol. 55, no. 5, (2020), e2019RS006911. https://doi.org/10.1029/2019RS006911.
[3] A. A. Willoughby, M. M. Tanko, H. D. Musa, and M. S. Abdullahi, "Estimation of some radio propagation parameters using radiosonde measurements over Nigeria", Journal of the Nigerian Society of Physical Sciences, vol. 5, (2023), p. 875. https://doi.org/10.46481/jnsps.2023.0875.
[4] ITU-R, "Recommendation ITU-R P.530-17: Propagation data and prediction methods required for the design of terrestrial line-of-sight systems", International Telecommunication Union Radiocommunication Sector, (2017).
[5] B. R. Bean and E. J. Dutton, Radio Meteorology, vol. 92, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., (1966).
[6] ITU-R, "Recommendation ITU-R P.453-14: The radio refractive index: Its formula and refractivity data", International Telecommunication Union Radiocommunication Sector, (2019).
[7] A. T. Adediji, M. O. Ajewole, and S. E. Falodun, "Distribution of radio refractivity gradient and effective earth radius factor (k-factor) over Akure, South-Western Nigeria", Journal of Atmospheric and Solar-Terrestrial Physics, vol. 73, no. 16, (2011), pp. 2300–2304.
[8] T. J. O. Afullo and P. K. Odedina, "Effective earth radius factor characterization for line of sight paths in Botswana", In *2004 IEEE Africon. 7th Africon Conference in Africa (IEEE Cat. No. 04CH37590)*, vol. 1, (2004), pp. 227-231.
[9] K. O. Suleman et al., "Investigation of diurnal and seasonal variation of radio refractivity and field strength over Ogbomoso, South Western Nigeria", Proceedings of URSI GASS 2025, (2025).
[10] A. L. Sheu, J. S. Ojo, and A. M. Aibinu, "Investigation of the vertical profile radio refractivity gradient and effective Earth-radius factor (k-factor) in transmission link over Oyo, Nigeria", Tanzania Journal of Science, vol. 48, no. 1, (2022), pp. 1–13.
[11] Y. B. Lawal and E. T. Omotoso, "Investigation of point refractivity gradient and geoclimatic factor at 70 m altitude in Yenagoa, Nigeria", Journal of the Nigerian Society of Physical Sciences, vol. 5, (2023), p. 1081. https://doi.org/10.46481/jnsps.2023.1081.
[12] T. J. O. Afullo and P. K. Odedina, "On the k-factor distribution and diffraction fading for Southern Africa", Southern African Institute of Electrical Engineers Africa Research Journal, vol. 97, no. 2, (2006), pp. 172-181.
[13] A. M. Nyete and T. J. O. Afullo, "Seasonal distribution modeling and mapping of the effective Earth radius factor for microwave link design in South Africa", Progress in Electromagnetics Research B, vol. 51, (2013), pp. 1–32. https://doi.org/10.2528/PIERB13030406.
[14] P. K. Odedina and T. J. O. Afullo, "Use of spatial interpolation technique for determination of geoclimatic factor and fade depth calculation in Southern Africa", IEEE Africon, vol. 51, (2007), pp. 1–5.
[15] A. J. Palmer and D. C. Baker, "Predicting the monthly cumulative distribution of the effective Earth radius factor for South Africa", IEEE Africon, (2004), pp. 1007–1010. https://doi.org/10.1109/AFRICON.2004.1406787.
[16] A. J. Palmer and D. C. Baker, "Predicting the long-term average of the effective Earth radius factor for South Africa using ground-based observations", SAIEE Africa Research Journal, vol. 97, no. 2, (2021), pp. 182–185. https://doi.org/10.23919/SAIEE.2021.9443731.
[17] Y. B. Lawal, O. A. Layioye, P. A. Owolawi, C. Tu, E. Van Wyk, and J. S. Ojo, "Statistical Estimation of Point Refractivity Gradient and Geoclimatic Factor at Microwave Antenna Height in Port Elizabeth, South Africa", Journal of Vibration Engineering, vol. 25, no. 11, (2025), pp. 1–16. https://doi.org/20.14118/J0ve/V25.I11/251114.
[18] S. Mentes and Z. Kaymaz, "Investigation of surface duct conditions over Istanbul, Turkey", Journal of Applied Meteorology and Climatology, vol. 46, no. 3, (2007), pp. 318–337. https://doi.org/10.1175/JAM2452.1.
[19] Y. Zhou, Y. Liu, J. Qiao, J. Li, and C. Zhou, "Statistical analysis of the spatiotemporal distribution of lower atmospheric ducts over the seas adjacent to China", Remote Sensing, vol. 14, no. 19, (2022), p. 4864. https://doi.org/10.3390/rs14194864.
[20] Q. Jiang, Q. Wang, and S. Gaberšek, "Mesoscale variability of surface ducts during Santa Ana wind episodes", Journal of Geophysical Research: Atmospheres, vol. 127, no. 9, (2022), e2022JD036698. https://doi.org/10.1029/2022JD036698.

[21] Port Elizabeth, "Understanding the Impact of Port Elizabeth's Coastal Winds", Port Elizabeth website, (2025). Available at: https://portelizabeth.com/weather/understanding-the-impact-of-port-elizabeths-coastal-winds (accessed 2 September 2025).
[22] A. C. Kruger and M. P. Nxumalo, "Historical rainfall trends in South Africa: 1921–2015", Water SA, vol. 43, (2017), pp. 285–297. https://doi.org/10.4314/wsa.v43i2.12.
[23] Copernicus Climate Change Service (C3S), "ERA5 hourly data on pressure levels from 1940 to present", (2023). https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels.
[24] ITU-R, "Recommendation ITU-R P.530-18: Propagation data and prediction methods required for the design of terrestrial line-of-sight systems", International Telecommunication Union, (2021).
[25] D. O. Akpootu, M. Idris, I. Nouhou, M. I. Iliyasu, and A. O. Aina, "Estimation and investigation of variability of tropospheric radio refractivity and radio field strength over Accra, Ghana", Journal of Atmospheric and Earth Science, vol. 5, (2021), p. 026.
[26] K. O. Suleman, S. B. Suleiman, L. A. Sunmonu, and A. L. Sheu, "Estimation of refractivity gradients and effective earth radius factor (K-factor) in the lowest 100 m of the atmosphere over Ogbomoso, southwestern Nigeria", Journal of Engineering, Technology and Industrial Applications, vol. 10, no. 2, (2024), pp. 1–10.
[27] J. A. Curry, "THERMODYNAMICS: Saturated Adiabatic Processes", in Encyclopedia of Atmospheric Sciences (Second Edition), Academic Press, (2015), pp. 398–401. https://doi.org/10.1016/B978-0-12-382225-3.00406-0.
[28] L. Breiman, "Random forests", Machine Learning, vol. 45, no. 1, (2001), pp. 5–32.
[29] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests", Pattern Recognition Letters, vol. 31, no. 14, (2010), pp. 2225–2236.
[30] D. J. Lieske, M. S. Schmid, and M. Mahoney, "Ensembles of ensembles: Combining the predictions from multiple machine learning methods", in Machine Learning for Ecology and Sustainable Natural Resource Management, Springer, Cham, (2018), pp. 109–121. https://doi.org/10.1007/978-3-319-96978-7_5.
[31] A. Krzemińska and T. Miller, "Leveraging random forest techniques for enhanced microbiological analysis: A machine learning approach to investigating microbial communities and their interactions", InterConf, vol. 32, no. 151, (2023), pp. 386–398. https://doi.org/10.51582/interconf.19-20.04.2023.040.
[32] N. Gunduz and E. Fokoué, "Robust classification of high dimension low sample size data", *arXiv: Applications*, arXiv:1501.00592 [stat.ML], (2015). Available at: https://arxiv.org/abs/1501.00592.
[33] B. F. F. Huang and P. C. Boutros, "The parameter sensitivity of random forests", BMC Bioinformatics, vol. 17, no. 1, (2016), p. 331. https://doi.org/10.1186/S12859-016-1228-X.
[34] S. J. Winham, C. L. Colby, R. R. Freimuth, X. Wang, M. de Andrade, M. Huebner, and J. M. Biernacka, "SNP interaction detection with random forests in high-dimensional genetic data", BMC Bioinformatics, vol. 13, no. 1, (2012), p. 164. https://doi.org/10.1186/1471-2105-13-164.
[35] J. Paul and P. Dupont, "Statistically interpretable importance indices for Random Forests", Technical report, Université catholique de Louvain, (2014). Available at: https://dial.uclouvain.be/pr/boreal/object/boreal:147350.
[36] Y. B. Lawal and A. G. Ashidi, "Statistical estimation of effective Earth radius factor over Lagos using radiosonde data", International Journal of Multidisciplinary and Current Research, vol. 5, (2017), pp. 343–348.
[37] A. T. Adediji and M. O. Ajewole, "Vertical profile of radio refractivity gradient in Akure South-West Nigeria", Progress in Electromagnetics Research C, vol. 4, (2008), pp. 157–168.
[38] S. P. Venkiteshwaran, "Diurnal variation of the effective Earth's radius factor (k) over India", Indian Journal of Radio & Space Physics, vol. 1, no. 1, (1972), pp. 1–6.
[39] S. L. Usman et al., "Exploring vertical gradients of radio refractivity and effective Earth radius factor over Nigerian stations", Science World Journal, vol. 19, no. 1, (2024), pp. 40–49.