

Article

Modeling of Stochastic Wind Based on Operational Flight Data Using Karhunen–Loève Expansion Method

Xiaolong Wang ^{1,*} , Lukas Beller ¹, Claudia Czado ²  and Florian Holzapfel ¹ 

¹ Institute of Flight System Dynamics, Technical University of Munich, 85748 Garching, Germany; lukas.beller@tum.de (L.B.); florian.holzapfel@tum.de (F.H.)

² Applied Mathematical Statistics, Technical University of Munich, 85748 Garching, Germany; cczado@ma.tum.de

* Correspondence: xiaolong.wang@tum.de; Tel.: +49-89-289-16102

Received: 15 July 2020; Accepted: 15 August 2020; Published: 18 August 2020



Abstract: Wind has a significant influence on the operational flight safety. To quantify the influence of the wind characteristics, a wind series generator is required in simulations. This paper presents a method to model the stochastic wind based on operational flight data using the Karhunen–Loève expansion. The proposed wind model allows us to generate new realizations of wind series, which follow the original statistical characteristics. To improve the accuracy of this wind model, a vine copula is used in this paper to capture the high dimensional dependence among the random variables in the expansions. Besides, the proposed stochastic model based on the Karhunen–Loève expansion is compared with the well-known von Karman turbulence model based on the spectral representation in this paper. Modeling results of turbulence data validate that the Karhunen–Loève expansion and the spectral representation coincide in the stationary process. Furthermore, construction results of the non-stationary wind process from operational flights show that the generated wind series have a good match in the statistical characteristics with the raw data. The proposed stochastic wind model allows us to integrate the new wind series into the Monte Carlo Simulation for quantitative assessments.

Keywords: wind model; operational flight data; stochastic process; Karhunen–Loève Expansion; vine copula; spectral representation

1. Introduction

To improve the operational flight safety, an Acceptable Level of Safety Performance (ALoSP) [1] is defined to identify the safety level for airlines or airports. Wind has a significant influence on operational flight safety and should be considered in the quantitative assessments of the operational flight safety level. Mathematical wind models such as ‘1-cosine’ discrete wind gust model, low-level wind shear model, von Karman turbulence model, and Dryden turbulence model, are commonly used for flight simulation and flight control design [2–4]. European Aviation Safety Agency (EASA) also illustrated the models of mean wind, turbulence, and wind shear for airworthiness assessment [5]. However, the general wind models cannot represent the local statistical characteristics of the wind. Therefore, they are not suitable to quantify the operational risk for a specific airline or airport. Instead, wind estimated from operational flight data in the onboard Quick Access Recorder (QAR) [6,7], as a sufficient wind database, allows us to analyze the specific statistical characteristics of the real wind conditions.

A model-based predictive analysis framework to quantify the incident probability for a specific airline has been developed by the Institute of Flight System Dynamics [8,9]. Wind speed, as one of

the contributing factors to the incident, can be taken into account in this approach. The quantitative assessment of the incident probability is obtained via propagating the uncertainty of the contributing factors through the incident model. During the uncertainty propagation, each contributing factor remains constant in one evaluation of the incident model. As shown in the runway overrun case in the paper [8], the wind speed is constant in the incident model. However, the constant wind assumption is not suitable for all incidents, as the variation of wind speed and direction affects the flight performance and handling quality significantly. Therefore, wind series should be considered in the incident model. Although the traditional wind shear or turbulence model can be used in Monte Carlo simulation, they can not represent the statistical characteristics of the specific operational flight data. Thus, the objective of this paper is to build a stochastic wind model in order to generate new wind series, which follow the statistical characteristics of the real wind data and can be used in the incident model simulation.

Wind series can be assumed as a stochastic process. Conventional turbulence wind is modeled as a stationary stochastic process using the spectral representation (SR) methods [10]. However, such approaches can not deal with the non-stationary wind series. To obtain the realistic quantitative statement of the operational flight risk, a high fidelity model to describe the wind series is required. Markov processes are highly useful models for stochastic processes, but the process has to satisfy the Markov property [11]. Without any assumption about the process, a Karhunen–Loève (KL) expansion method is proposed in [12,13] to construct the observed non-stationary random phenomena, such as seismic ground motion and wave field. This method is also called functional principal component analysis [14], which are widely used in sensitivity analysis domain to identify the key parameters and also has a good performance in dimension reduction. In addition, some researchers show that the SR method and the KL expansion method coincide for the weakly stationary process [15,16]. In this paper, we apply the KL expansion method to model the encountered wind series during the operational flights. The coefficients in KL expansion are assumed to be independent in many applications [14]. However, it is not always true, especially if the stochastic process is not Gaussian process [12]. The pair dependence between every two coefficients in KL is considered in [13]. To better describe the statistical characteristics of the KL coefficients, a vine copula technique [17] is used to capture the high dimensional dependence among those coefficients instead of the pair dependence. With the integration of a vine copula into KL expansion, the better representation of KL coefficients would improve the accuracy of the constructed stochastic wind model. To compare with the conventional turbulence model used in aviation, the similarity and difference between the KL and the SR method for turbulence modeling are also discussed in this paper.

The paper starts with an explanation about the basic concepts of the KL expansion method and vine copula, followed by a comparison of SR and KL in terms of equations in Section 2. In Section 3, the modeling results of turbulence data and real headwind series data are presented and discussed. Besides, the modeling of extracted wind shear ramps is also implemented. Section 4 contains conclusions and possible improvements of the proposed wind construction model.

2. Modeling of Stochastic Process

Two main construction methods of the stochastic process are discussed in this section: KL expansion and SR methods. The KL expansion method is introduced to model the finite stochastic process. Then, a vine copula is brought in to construct the high dimensional dependence of the random variables in the KL expansion. The strategy that the vine copula describes the high dimensional dependence is explained using a three-dimensional case. Besides, we present the SR method for the stationary process, which is subsequently compared with the KL expansion.

2.1. Karhunen–Loève Expansion

Let $X(t)$ be a stochastic process at a given time interval $t \in [0, T]$. The autocorrelation and autocovariance functions are usually used to describe the random process. They are calculated as follows:

$$\begin{aligned} R(t_1, t_2) &= \mathbb{E}[X(t_1)X(t_2)], \\ \text{Cov}(t_1, t_2) &= \mathbb{E}[(X(t_1) - \mu_1)(X(t_2) - \mu_2)] = R(t_1, t_2) - \mu_1\mu_2, \end{aligned} \quad (1)$$

where $t_1, t_2 \in T$. \mathbb{E} is the expectation operator, and μ_i is the mean of the process $X(t)$ at given time t_i . For a zero-mean stochastic process, the autocorrelation and autocovariance are equivalent. As an arbitrary process can be transformed to a zero-mean process by subtracting its mean, without loss of generality, we can assume that $X(t)$ is a zero-mean process. Then, $X(t)$ can be reconstructed using the KL expansions [18]:

$$X(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \zeta_k \phi_k(t), \quad (2)$$

where the λ_k and the corresponding functions $\phi_k(t)$ are the solutions of the integral equation:

$$\int_D R(t, s) \phi(s) ds = \lambda \phi(s), \quad t, s \in D = [0, T]. \quad (3)$$

The corresponding random coefficients $\zeta_k, k = \{1, 2, \dots\}$ are computed using

$$\zeta_k = \frac{1}{\sqrt{\lambda_k}} \int_D X(t) \phi_k(t) dt. \quad (4)$$

By Mercer's theorem [18], the $R(t, s)$ can be expressed in terms of its eigenvalues and eigenvectors via eigendecomposition as follows:

$$R(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s), \quad (5)$$

where the obtained λ_k and $\phi_k(t)$ fulfill Equation (3). In addition, eigenvectors $\phi_k(t)$ are orthonormal and their inner product follows the equation:

$$\langle \phi_k(t), \phi_l(t) \rangle = \int_D \phi_k(t) \phi_l(t) dt = \delta_{kl}, \quad k, l \geq 1, \quad (6)$$

where δ_{kl} is the Kronecker delta function. In addition, $X(t)$ is a zero-mean process, the total variance of the $X(t)$ can be derived as below:

$$\int_D \text{Var}[X(t)] dt = \int_D \mathbb{E}[X^2(t)] dt = \int_D \mathbb{E}[(\sum_{k=1}^{\infty} \sqrt{\lambda_k} \zeta_k \phi_k(t))^2] dt = \sum_{k=1}^{\infty} \lambda_k. \quad (7)$$

The sum of all λ_k indicates the integrated variance of the process. In practice, Equation (2) is truncated at a certain order K . Therefore, the relative error ϵ of the truncated expansion can be calculated as below:

$$\epsilon = 1 - \frac{\sum_{k=1}^K \lambda_k}{\int_D R(t, t) dt} = 1 - \frac{\sum_{k=1}^K \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}. \quad (8)$$

$1 - \epsilon$ is called the cumulative variance ratio. As mentioned before, $X(t)$ is a process with zero mean. It means that $\mathbb{E}[X(t)] = 0$ for all $t \in D$. Considering an arbitrary process $Y(t)$, the truncated KL expansion series of an arbitrary process can be represented by adding its mean as below:

$$Y(t) \approx \mathbb{E}[Y(t)] + \sum_{k=1}^K \sqrt{\lambda_k} \zeta_k \phi_k(t), \quad (9)$$

where λ_k is sorted in descending order as $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$. As the eigenvectors ϕ_k are orthonormal and $Y(t) - \mathbb{E}[Y(t)]$ is zero mean, the realizations of ζ_k computed using Equation (4) satisfy

$$\mathbb{E}[\zeta_k] = 0, \text{Var}[\zeta_k] = 1, \mathbb{E}[\zeta_k \zeta_l] = \delta_{kl}, \quad (10)$$

which indicates that random variables ζ_k are mutually uncorrelated to each other, and distributed with zero mean and unit variance. In a particular case that the random process is a Gaussian process, all ζ_k are independent standard Gaussian distributed. For the non-Gaussian process, the dependence among ζ_k can be described by the joint distribution of ζ_k . New samples of ζ_k sampled from the distribution of ζ_k allow us to generate new realizations of the stochastic process using Equation (9).

2.2. Vine Copula

The integration of a vine copula into KL expansion is proposed to deal with the dependent ζ_k . A brief introduction related to the copula is presented. For details, the readers are referred to [17,19]. Based on the theorem of Sklar [20], the joint distribution F of the dependent ζ_k can be decomposed by the marginal cumulative distribution functions (CDF) F_k and a copula structure C as below:

$$F(\zeta_1, \zeta_2, \dots, \zeta_K) = C(F_1(\zeta_1), F_2(\zeta_2), \dots, F_K(\zeta_K)). \quad (11)$$

As F_k is the CDF of ζ_k , the realization of F_k is uniformly distributed. C describes the dependence among ζ_k and eliminates the impact of the marginal distributions. The probability density f is computed as below:

$$f(\zeta_1, \zeta_2, \dots, \zeta_K) = c(F_1(\zeta_1), F_2(\zeta_2), \dots, F_K(\zeta_K)) \prod_{k=1}^K f_k(\zeta_k). \quad (12)$$

As the truncated order K of the KL expansion might go to 20 or more for highly accurate reconstruction, the copula density function c becomes too complex. Based on the Rosenblatt transformation [21] and the conditioning arguments [17], the vine copula approach allows us to construct a K -dimensional copula density function using $\frac{K(K-1)}{2}$ bivariate copula densities. The maximum likelihood method is applied to estimate the bivariate copula densities based on the given data. An example of a three-dimensional probability distribution constructed from three bivariate copula densities is presented in Appendix A. Since the construction of high dimensional copula in Equation (A3) is not unique, [22] describes a sequential strategy to obtain the most suitable construction structure. After the vine copula dependence model of ζ_k is obtained, dependent samples of the ζ_k can be generated using the vine copula. Based on the KL expansion and vine copula methods, the procedure of the stochastic process reconstruction and regeneration is summarized in Algorithm 1.

2.3. Spectral Representation

The SR method is often applied to deal with a stationary stochastic process. SR describes the features of the process in the frequency domain using the power spectral density (PSD) function $P(\omega)$. $P(\omega)$ can be computed via the Fourier transformation of the autocovariance function $R_X(t + \tau, t)$ of the process [23] as below:

$$P(\omega) = \int_{-\infty}^{\infty} R_X(t + \tau, t) e^{-i\omega\tau} d\tau, \quad (13)$$

where ω is the angular frequency and τ is the time lag between two time points. $R_X(t + \tau, t) = R_X(\tau)$ for the stationary process, and it can be obtained based on the inverse Fourier transformation as below:

$$R_X(t + \tau, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega\tau} d\omega. \quad (14)$$

Algorithm 1 Procedure of the stochastic process reconstruction and regeneration

- 1: Collection of the measured realizations of the stochastic process Y
- 2: Center the data Y using $X(t) = Y(t) - \mathbb{E}[Y(t)]$
- 3: Calculate the autocovariance $R(t, s)$ of X using Equation (1)
- 4: Obtain the eigenfunctions ϕ_i and eigenvalues λ_i via the eigendecomposition of $R(t, s)$ using Equation (5)
- 5: Calculate the cumulative variance ratio $1 - \epsilon$ using Equation (8) and set the truncated order K depending on the required $1 - \epsilon$.
- 6: Calculate the coefficients ζ_k , where $k = 1, 2, \dots, K$ using Equation (4)
- 7: **if** Reconstruct the original process **then**
- 8: Calculate the Y_{rec} using the computed ζ_k based on the Equation (9)
- 9: **end if**
- 10: **if** Regenerate the new process **then**
- 11: Estimate the marginal distributions of ζ_k
- 12: **if** Independent Sampling **then**
- 13: Get samples of ζ_k via sampling individually from the marginal distributions
- 14: **else if** Copula Sampling **then**
- 15: Estimate the vine copula of ζ_k .
- 16: Get samples of ζ_k via the copula sampling
- 17: **end if**
- 18: Calculate the Y_{new} using the sampled ζ_k based on the Equation (9)
- 19: **end if**

For a finite stationary stochastic process with zero-mean $X(t)$, $t \in [0, T]$, it can be represented using the exponential form of the Fourier series expansion as below:

$$X(t) = \sum_{k=-\infty}^{\infty} \mathcal{F}_k e^{ik\Delta\omega t} = \sum_{k=-\infty}^{\infty} |\mathcal{F}_k| e^{i\psi_k} e^{i\omega_k t}, \quad (15)$$

where $\omega_k = k\Delta\omega = k \cdot 2\pi/T$ and \mathcal{F}_k is the Fourier coefficients. $|\mathcal{F}_k|$ and ψ_k are the frequency amplitude and corresponding phase, respectively. \mathcal{F}_k can be calculated via the Fourier inverse transformation as below:

$$\mathcal{F}_k = \frac{1}{T} \int_0^T X(t) e^{-i\omega_k t} dt. \quad (16)$$

The average power of the process P is defined as

$$P = \frac{1}{T} \int_0^T X^2(t) dt = \frac{1}{T} \int_0^T \left(\sum_{k=-\infty}^{\infty} \mathcal{F}_k e^{i\omega_k t} \right)^2 dt. \quad (17)$$

Since the inner product of the exponential terms in Equation (17) satisfies

$$\left\langle \frac{e^{i\omega_k t}}{\sqrt{T}}, \frac{e^{i\omega_l t}}{\sqrt{T}} \right\rangle = \int_0^T \frac{e^{i\omega_k t}}{\sqrt{T}} \frac{e^{-i\omega_l t}}{\sqrt{T}} dt = \delta_{kl}, \quad (18)$$

the average power P can be rewritten as below:

$$P = \sum_{k=-\infty}^{\infty} |\mathcal{F}_k|^2 = \sum_{k=-\infty}^{\infty} P_k, \quad (19)$$

where P_k is the power for the specific ω_k . Substitute $|\mathcal{F}_k|$ using $\sqrt{P_k}$ in Equation (15):

$$X(t) = \sum_{k=-\infty}^{\infty} \sqrt{P_k} e^{i\psi_k} e^{i\omega_k t} = \sum_{k=-\infty}^{\infty} \sqrt{P_k T} e^{i\psi_k} \frac{e^{i\omega_k t}}{\sqrt{T}}. \quad (20)$$

Therefore, $R_X(\tau)$ in Equation (14) can be rewritten in the Fourier series expansion:

$$R_X(\tau) = \sum_{k=-\infty}^{\infty} P_k e^{i\omega_k \tau} = \sum_{k=-\infty}^{\infty} P_k e^{i\omega_k(t-s)} = \sum_{k=-\infty}^{\infty} P_k T \left\langle \frac{e^{i\omega_k t}}{\sqrt{T}}, \frac{e^{i\omega_k s}}{\sqrt{T}} \right\rangle, \quad (21)$$

where $t, s \in T$. For an infinite stationary process, $P(\omega)$ can be obtained by taking the limit of $P_k T$ when $T \rightarrow \infty$ as follows:

$$P(\omega) = \lim_{\Delta\omega_k \rightarrow 0} \frac{2\pi P_k}{\Delta\omega_k} = \lim_{T \rightarrow \infty} P_k T, \text{ for } \omega_k \in [\omega_{k-1}, \omega_k], \quad (22)$$

where $\Delta\omega_k = \omega_k - \omega_{k-1}$. Compare the representation equation Equation (15) with the KL expansion equation Equation (2), and the autocovariance equation Equation (21) with Equation (5), respectively. It shows that $\lambda_k = P_k T$ and $\phi_k = e^{i\omega_k t} / \sqrt{T}$ are the solution of the integral equation Equation (3). It reveals that the SR and the KL expansion coincide in the construction of a finite stationary process. For the infinite stationary process, it is worth mentioning that $P(\omega)$ is continuous and has two sides $\omega < 0$ and $\omega > 0$. However, the λ_k in KL expansion method is discrete and the corresponding ω_k is always larger than 0.

3. Results and Discussion of Wind Process Modeling

This section starts with a turbulence model construction. Wind data generated from a well-known von Karman turbulence model is used to compare the KL expansion and the SR method. Then, the wind series during the operational flight is reconstructed and regenerated using the KL expansion method. The original headwind data is estimated based on 849 flights of a B747-8F that landed on the same runway. A preprocessing of the data is implemented using the Rauch-Tung-Striebel smoother smoother to improve the data quality. For the details of the data smoothing, readers are referred to reference given in [24]. Modeling of headwind data during the final approach is presented, followed by another case study of the horizontal wind shear ramps construction.

3.1. Turbulence

The conventional wind turbulence model in aviation is described using the PSD function of the SR method. For an aircraft flying at speed V through a frozen turbulence field, PSD functions $\Phi(\Omega)$ of the von Karman model [10] in three directions are shown as below:

$$\begin{aligned} \Phi_u(\Omega) &= \sigma_u^2 \frac{2L_u}{\pi} \frac{1}{[1 + (1.339L_u\Omega)^2]^{5/6}}, & (\text{Longitudinal}) \\ \Phi_v(\Omega) &= \sigma_v^2 \frac{L_v}{\pi} \frac{1 + \frac{8}{3}(1.339L_v\Omega)^2}{[1 + (1.339L_v\Omega)^2]^{11/6}}, & (\text{Lateral}) \\ \Phi_w(\Omega) &= \sigma_w^2 \frac{L_w}{\pi} \frac{1 + \frac{8}{3}(1.339L_w\Omega)^2}{[1 + (1.339L_w\Omega)^2]^{11/6}}, & (\text{Vertical}) \end{aligned} \quad (23)$$

where u , v and w refer to the longitudinal, lateral and vertical turbulence, respectively. σ_u , σ_v , and σ_w are the intensity measures of the turbulence. L_u , L_v , and L_w (in feet) are the length scales of the turbulence. Ω is the spatial frequency, which is equal to the angular frequency ω divided by V . In general, ω is in the range of $[0, 100\pi]$ [10]. With respect to ω , the PSD $S(\omega)$ is computed as follows:

$$S(\omega) = \frac{\Phi(\Omega)}{V} = \frac{\Phi(\omega/V)}{V}. \quad (24)$$

When the altitude of aircraft h (in feet) is less than 1000 feet, the L and σ can be calculated as below:

$$L_w = h, \quad (25)$$

$$L_u = L_v = \frac{h}{(0.177 + 0.000823h)^{1.2}},$$

$$\sigma_w = 0.1V_{W20}, \quad (26)$$

$$\sigma_u = \sigma_v = \frac{\sigma_w}{(0.177 + 0.000823h)^{0.4}},$$

where V_{W20} is the wind speed measured at 20 feet (6.10 m). Given a certain flight condition, for example, $h = 600$ feet (182.88 m), $V_{W20} = 15$ knots (7.72 m/s), and $V = 140$ knots (72.02 m/s), the PSD of the turbulence model in three directions is shown in Figure 1. 2000 time series of 256 s in 16 Hz are generated for each direction using the Equation (23) in the same flight condition. One example of the generated time series is shown using the blue dash line in Figure 2. The KL expansion method is implemented based on the generated 2000 turbulence series. Compared with the PSD, eigenvalues with the corresponding index in KL expansion are also shown in Figure 1.

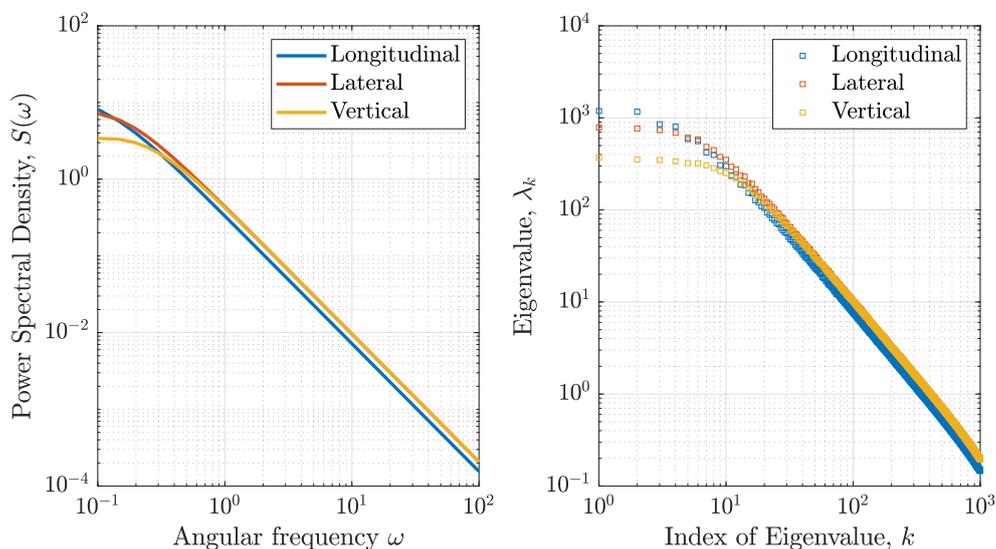


Figure 1. Comparison of the SR and KL construction for turbulence data in a certain flight condition.

Figure 1 shows that the shape of PSD and eigenvalues in the KL expansion are similar. As mentioned in Section 2.3, the index k of the KL expansion terms corresponds to a certain angular frequency ω_k and the eigenvalue λ_k in the KL expansion is equivalent to the PSD $P(\omega)$. The PSD of the turbulence model decreases with an increase of ω . Meanwhile, the eigenvalues λ_k in the KL expansion are sorted. Thereby, the shape of PSD and eigenvalues is almost similar. This result validates that the KL expansion method coincides with the SR method for the stationary stochastic process.

Two truncated orders $K = 30$ and $K = 100$ are selected in the KL expansion of turbulence. One example of KL reconstructed results is shown in Figure 2. It shows that the reconstructed process and the series generated from the PSD match very well. The reconstructed accuracy is improved with the increase of the truncated order. As the high order term is corresponding to the high frequency in this case, higher frequencies are neglected in the lower order truncated KL reconstruction. To quantify the influence of the KL truncated order K , the cumulative variance ratio $1 - \epsilon$ of the KL expansions in the case of longitudinal turbulence is calculated using Equation (8) and also shown in Figure 3a. It shows that the reconstructed turbulence with $K = 100$ can achieve 84.6% of variance. When K increases to 200, 92.6% of variance is achieved. Furthermore, the PSDs of the 2000 reconstructed longitudinal turbulence series using KL expansions are computed and shown in Figure 3b. The ‘theoretical’ black line is computed directly based on the von Karman model using Equation (23). The ‘raw data’ indicates the PSD are computed based on the turbulence data generated from the von Karman model.

The ' $K = 200, 100, 30$ ' shows the mean of PSD computed based on the KL reconstructed turbulence data with difference truncated orders. It also shows that the accuracy of PSD obtained from reconstructed data will be improved with the increasing order K . This property is also indicated by the cumulative variance ratio. Nevertheless, the KL expansion with sufficient orders based on the turbulence data is consistent with the conventional von Karman model. This approach allows us to construct the turbulence wind process.

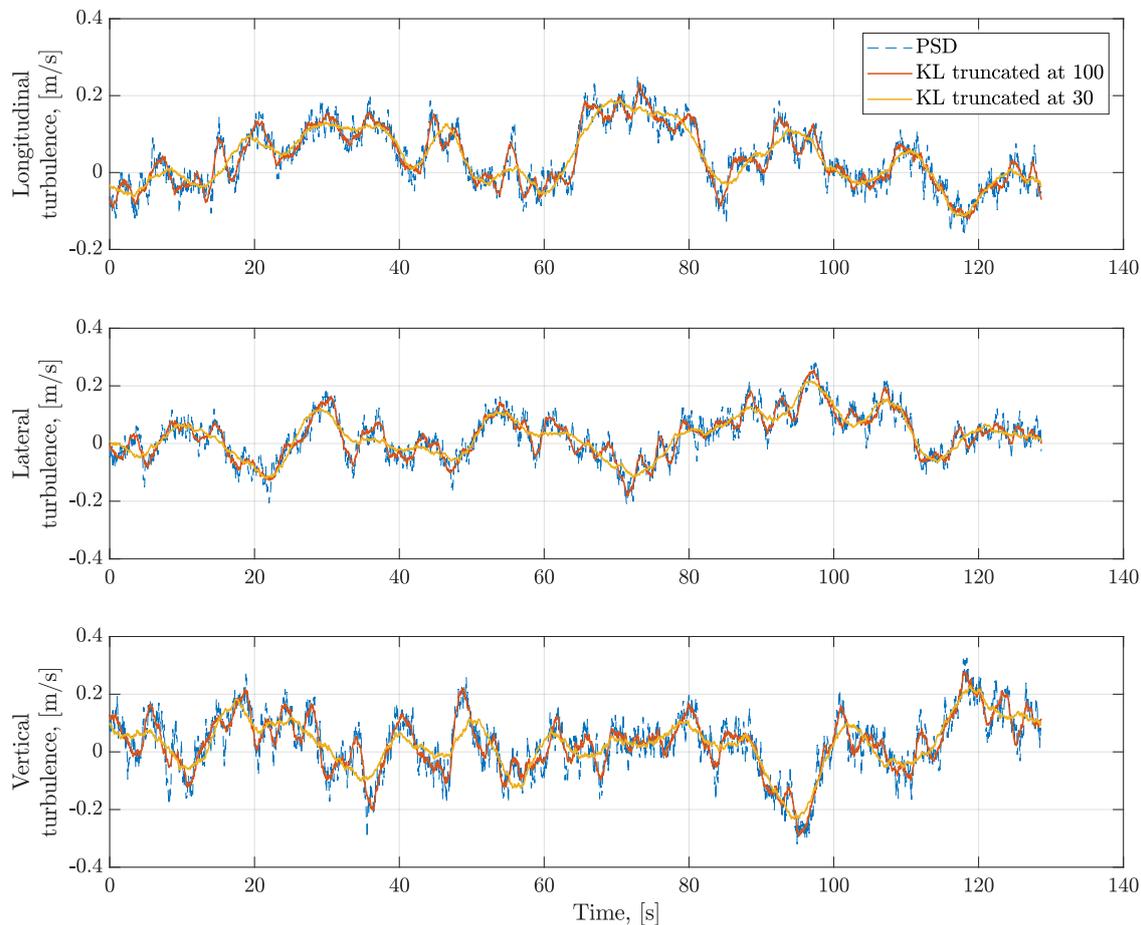


Figure 2. Comparison between the generated turbulence based on the PSD and the reconstructed turbulence using the KL expansion.

3.2. Headwind

Headwind has a significant influence on landing performance. A case of headwind process construction is demonstrated in this section. The headwind from 1000 feet (304.80 m) to 50 feet (15.24 m) above ground level (AGL) is estimated from the operational flight data in the final approach phase. 849 series of headwind (in knots) are resampled from the smoothed headwind based on the altitude (in feet) shown in Figure 4a. All wind series are homogenized after resampling.

The KL expansion method is implemented based on the headwind series of 849 flights. Figure 4b shows the relationship between the cumulative variance ratio and the truncated order K . $K = 20$ is selected in this case, and more than 99.9% variance of the process is obtained. Eigenfunctions ϕ_k of the autocorrelation matrix are shown in Figure 4c. According to the Equation (9), three arbitrary reconstructed wind series using the calculated ζ_k are illustrated and compared with raw wind series data in Figure 4d. Although only 20 terms KL functions are used in the construction, the reconstructed wind series and raw wind series match very well, which indicates that truncated KL expansion can be used to reconstruct the headwind series.

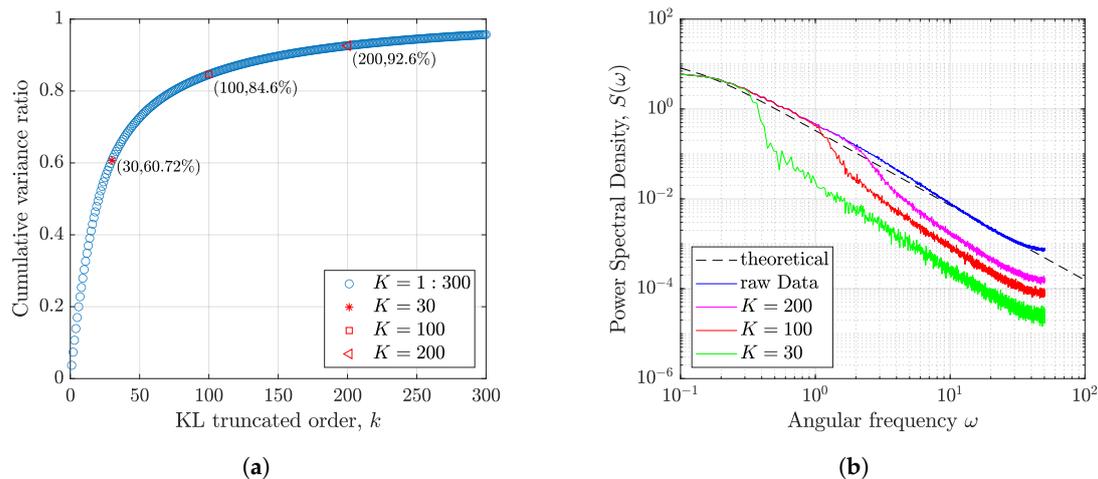


Figure 3. KL expansion results of longitudinal turbulence. (a) Cumulative variance ratio of KL expansion; (b) PSD comparison of reconstructed turbulence.

New samples of the ζ_k are required to generate the new wind series. The marginal distribution of each ζ_k is studied. Box plot of ζ_k in Figure 5a shows the computed ζ_k are distributed close to zero mean and unit variance, which is also derived in Equation (10). However, Figure 5a also reveals that the median of ζ_k is not always zero. The empirical CDF of ζ_k and standard Gaussian distribution are compared in Figure 5b. All blue dash lines are related to the empirical CDF of ζ_k . It indicates that some of ζ_k are almost following the standard Gaussian distribution. However, some deviations of CDF do exist in this case. Furthermore, marginal distributions for all ζ_k are estimated using maximum likelihood methods from the distribution candidates, such as 'Gaussian', 'Student-t', 'Generalized Extreme Value', 't Location-Scale', 'Logistic' distributions, and so on. Estimated probability densities of ζ_k are compared in Figure 5c, and the corresponding distribution parameters are shown in Table A1 of Appendix B. Deviations in PDF between the estimated distributions and Gaussian distribution are obvious. Instead of Gaussian distribution, the estimated marginal distributions of ζ_k are used to generate the realizations of ζ_k , which are subsequently used for the new wind series production based on the Equation (9).

As mentioned before, dependence among ζ_k might exist for non-Gaussian distributed ζ_k . A vine copula approach is applied to capture the dependence of the obtained ζ_k . To analyze the pairwise dependence between two ζ_k , the empirical contour plots using the copula data are shown in Figure A1 of Appendix C. The copula data has three scales: X scale denoted by ζ_k^X , U scale denoted by ζ_k^U , and Z scale denoted by ζ_k^Z . ζ_k^X is the same with the raw data ζ_k . ζ_k^U is obtained via transforming all samples of ζ_k into the uniform space as follows:

$$\zeta_k^U = F_k(\zeta_k), \quad (27)$$

where F_k is the CDF of the marginal distribution of ζ_k . ζ_k^Z is used in the normalized contour plots, and it is obtained via transforming all samples of ζ_k into the standard Gaussian space using:

$$\zeta_k^Z = \Psi^{-1}(F_k(\zeta_k)), \quad (28)$$

where Ψ is the CDF of the standard Gaussian distribution. Figure A1 shows that most dependence between two ζ_k is small, but some pair dependence exist, for example, the dependence between ζ_{10} and ζ_{16} . A vine copula with the estimated marginal distributions is used in this case to capture the dependence. The parametric and nonparametric bivariate copula are used separately and normalized contour plots on the Z scale are shown in Figure A2a,b, respectively. The parametric bivariate copula is using the defined copula function to fit the data, and the unknown parameters in the copula functions are estimated [17]. In contrast, the nonparametric bivariate copula is using the empirical function for two variables, which is the same as the empirical CDF for one variable. New realizations of ζ_k can be generated via independent sampling, parametric vine copula sampling, and nonparametric

vine copula sampling. To visualize the difference of generated samples, sample plots between ξ_{10} and ξ_{16} are shown in Figure 6. The raw data (Figure 6a), independent samples (Figure 6b), parametric (Figure 6c) and nonparametric copula samples (Figure 6d) in Z scale are illustrated, respectively. The dependence between ξ_{10} and ξ_{16} is evident in this case. Compared with the individually sampling from the marginal distributions, samples generated using the vine copula capture the dependence of the raw data and thus give a better representation of the raw data.

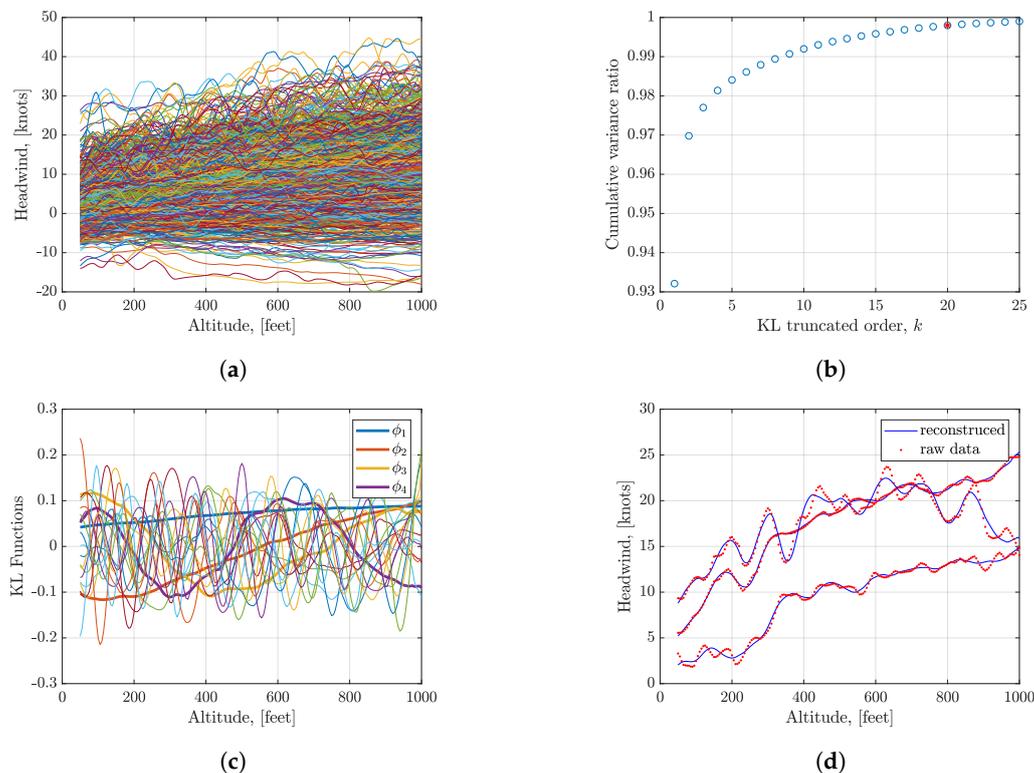


Figure 4. Headwind series and KL expansion results. (a) Headwind of 849 flights; (b) Cumulative variance ratio of KL expansion; (c) Eigenfunctions of KL expansion; (d) Reconstruction of three arbitrary wind series.

The statistical characteristics of the generated wind series are analyzed to validate the construction results. 5000 samples of ξ_k are generated using the independent sampling and the vine copula sampling. Then the corresponding 5000 wind series are generated for independent samples and dependent samples, respectively. The empirical CDFs and the kernel fitted PDFs of headwind speed at four different altitudes: 1000 feet (304.80 m), 600 feet (182.88 m), 300 feet (91.44 m), and 50 feet (15.24 m) are shown in Figures 7 and 8. It indicates that the generated wind series obtain a similar statistical distribution at a certain altitude compared with the raw data. Moreover, there is no significant difference in the headwind distribution between the independent and dependent sampling in 1000 feet, 600 feet, 300 feet. However, headwind distribution densities at 50 feet using the parametric and nonparametric copula sampling are closer to the raw data than the independent sampling.

To further analyze the results, the first four order sample moments of the observed wind data along with the altitude are calculated and shown in Figure 9. It shows that the constructed series have a good match at the mean and standard deviation with the raw data. For the independent sampling, the skewness is close to 0, and the kurtosis is always close to 3 for all altitudes. It is close to the moments of Gaussian distribution. In terms of the skewness and kurtosis, results show that the generated series can not exactly capture the skewness and kurtosis. However, the statistical moments using the parametric and nonparametric copula are closer to the raw statistical moments than the independent sampling, especially at the low altitude. Since there is no significant deviation

in the numerical value of the skewness and kurtosis, the generated headwind series can be used in simulations for the quantitative assessment.

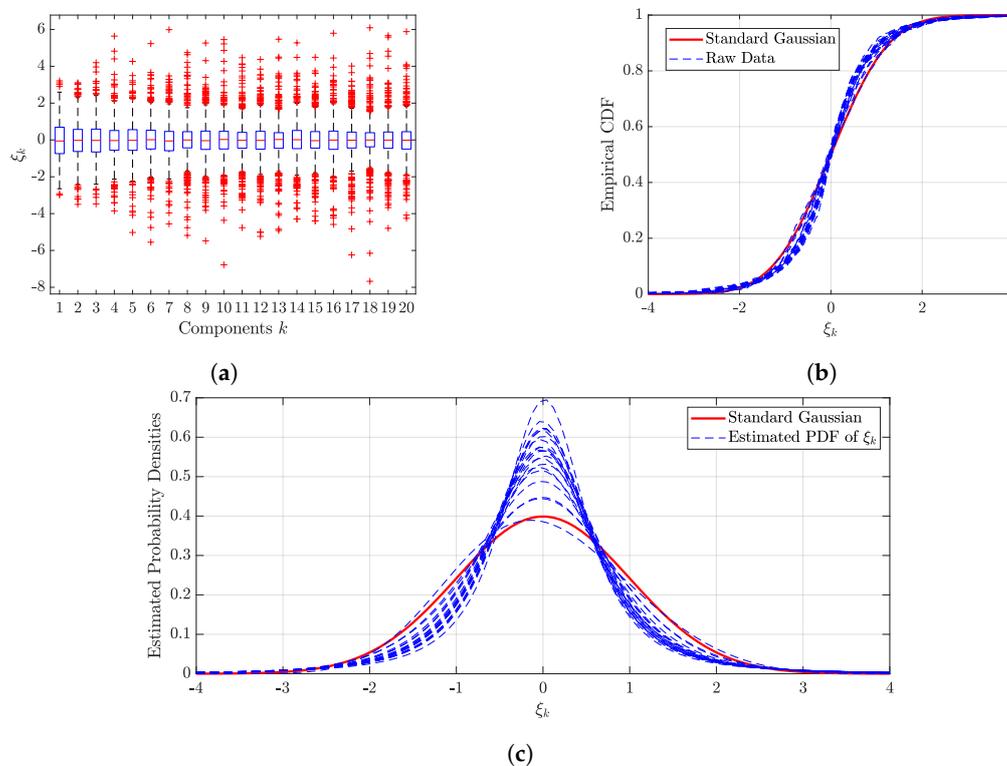


Figure 5. Marginal distributions of ζ_k . (a) Box plot of ζ_k ; (b) Empirical CDF of ζ_k ; (c) Estimated PDF of ζ_k .

3.3. Windshear

A safety-critical wind series, low-level wind shear, has been recognized as a severe hazard for flight safety. To analyze and quantify the occurrence probability and severity of the wind shear, the significant change of headwind V_W with the change of the altitude h , so-called horizontal wind shear ramps, are extracted from the flight data. The algorithm is to detect the extreme incremental values of ΔV_W within a certain interval Γ , which is defined as a certain height deviation in this case. A headwind increment for a moving window with a given width of $\Delta H = \Gamma/2$ is defined as below:

$$\Delta V_W = \max V_W(h : h + \Delta H) - V_W(h). \quad (29)$$

A headwind series of one flight is illustrated in Figure 10. The headwind speed incremental threshold ΔV_W^* and the detected height window ΔH^* are predefined. As shown in Figure 10, the incremental threshold ΔV_W^* is set to 5 knots (2.57 m/s) and the width of the rolling window ΔH is set to 100 feet. Two wind shear ramps with $\Delta V_W \geq 5$ knots and $\Gamma = 2\Delta H = 200$ feet are extracted from this flight.

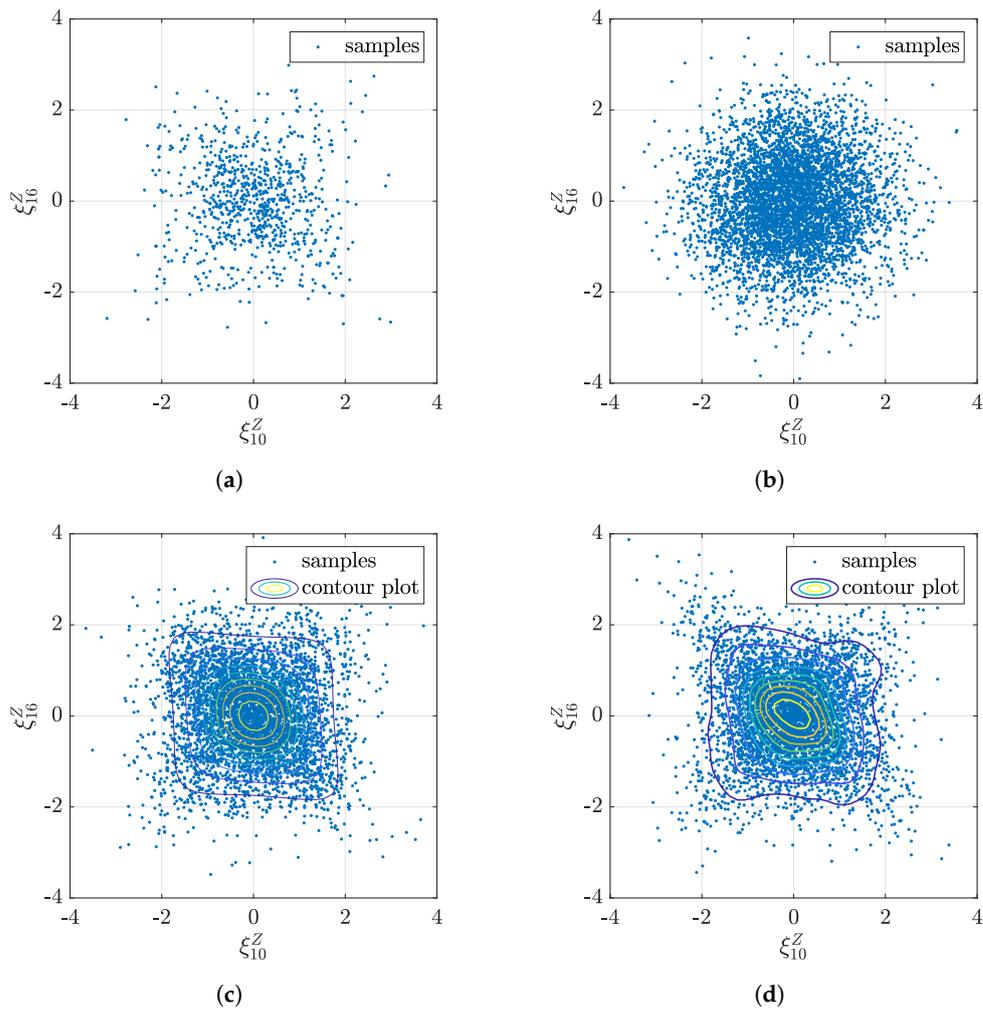


Figure 6. Different sample sets of ξ_{10} and ξ_{16} in Z scale (standard Gaussian space). (a) Raw samples; (b) Independent sampling; (c) Parametric copula sampling; (d) Nonparametric copula sampling.

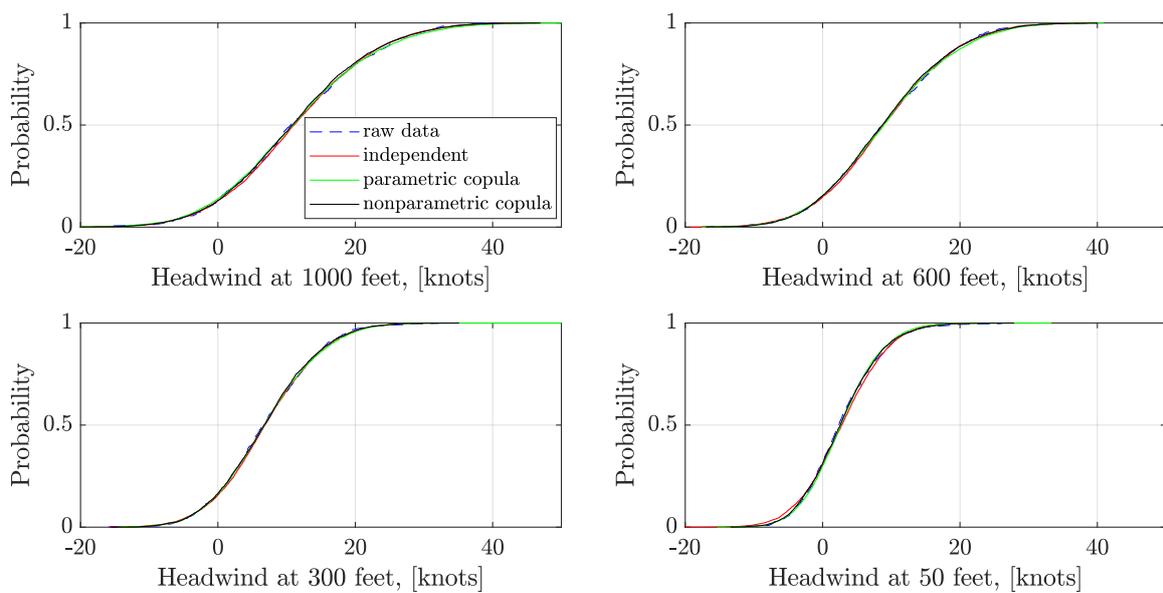


Figure 7. Comparison of the empirical CDFs of headwind at a certain altitude.

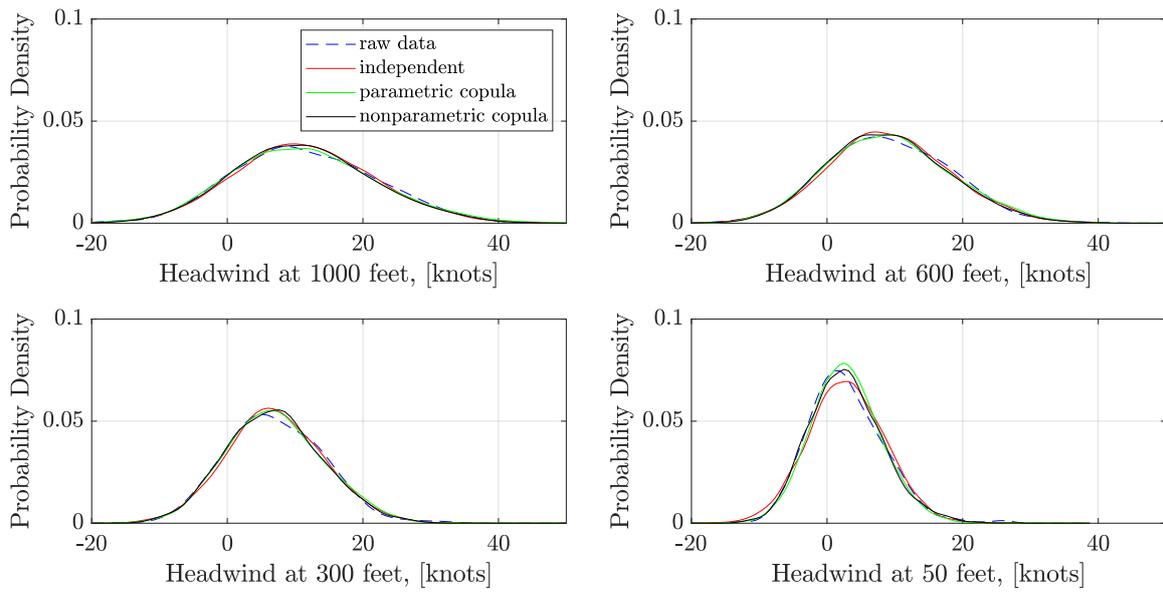


Figure 8. Comparison of the estimated PDFs of headwind at a certain altitude.

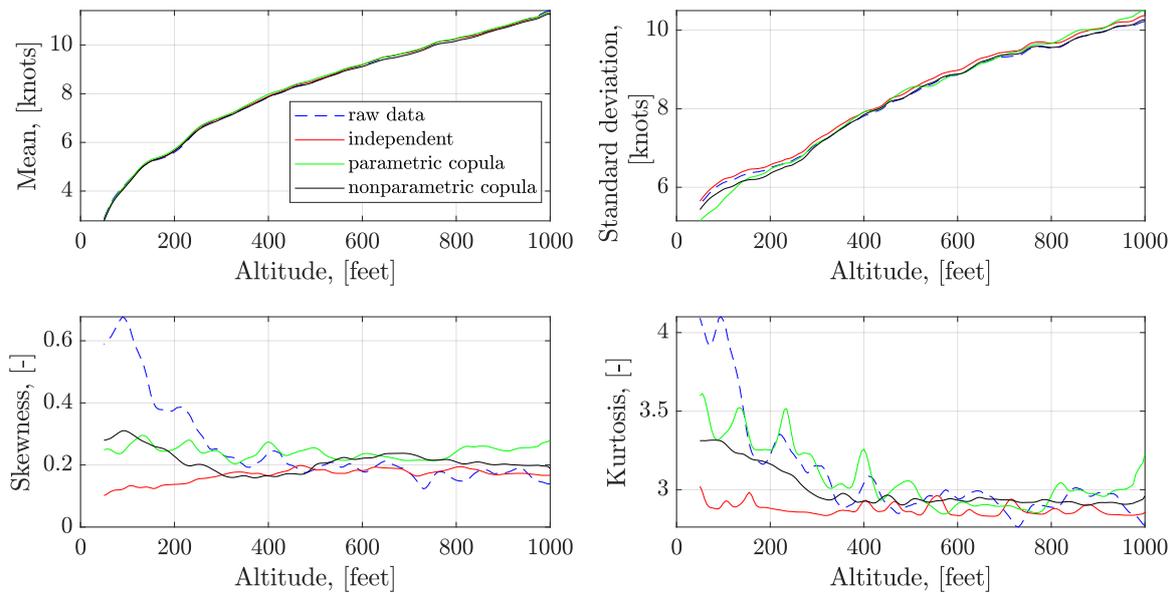


Figure 9. Statistical moments comparison of headwind along with the altitude.

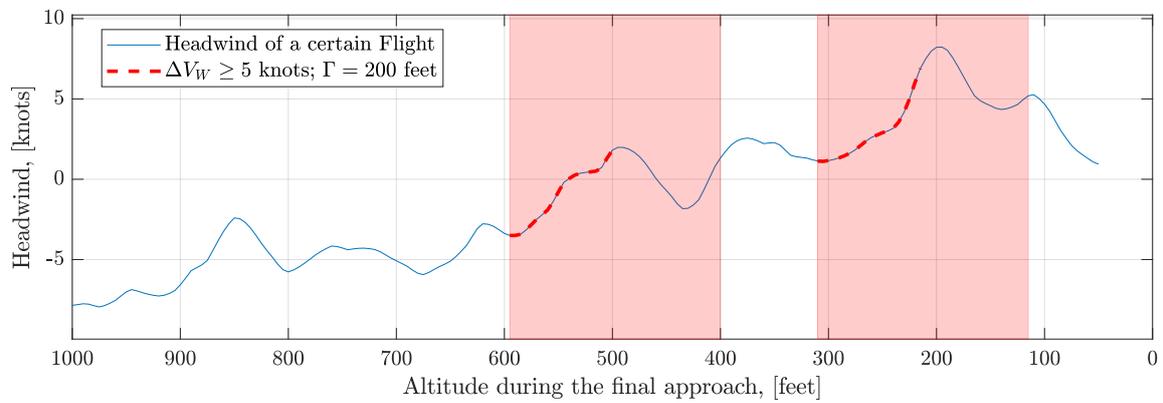


Figure 10. Detection of wind shear ramps in one flight.

The detection algorithm is applied to the 849 flights. The incremental threshold of ΔV_W is set to 5 knots, and the search window is set to 50 feet, 100 feet, and 200 feet, respectively. The mean and standard deviation of the extracted wind shear ramps series and the corresponding number of extracted series are visualized in Figure 11. It shows that the features of the mean and standard deviation depend on the width of the search window.

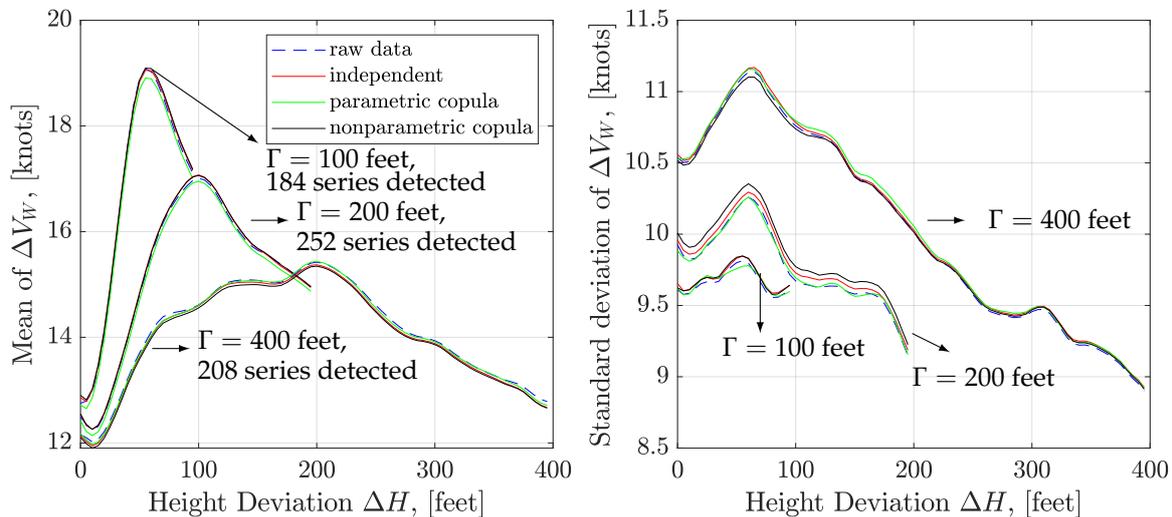


Figure 11. Mean and standard deviation of the extracted and generated wind shear ramps.

Given the detected series of wind shear ramps, the KL expansion method allows us to construct new series and enrich the database of the wind shear ramps. The KL truncated order $K = 10$ is used in the construction of wind shear ramps. 5000 new series of $\Delta V_W \geq 5$ knots are generated via the independent sampling, parametric and nonparametric copula sampling, respectively. As shown in Figure 11, the mean value of the raw data and generated series match very well, and there are only small biases in the standard deviation. In the case of the $\Gamma = 200$ feet, the parametric copula sampling has a better match with the raw data than others. The comparison results show that KL constructed model can be used to produce the new wind shear ramps series with similar statistical characteristics.

To further validate if the generated series can be used for the analysis of the wind shear intensity, a metric of $\max \Delta V_W / \Delta h$ related to the wind shear intensity is defined using the maximum change of headwind with the altitude as below:

$$\frac{\max \Delta V_W}{\Delta h} = \frac{\max V_W(h) - \min V_W(h)}{h_{V_W, \max} - h_{V_W, \min}}, \quad h \in [0, \Gamma]. \quad (30)$$

The metric $\max \Delta V_W / \Delta h$ is computed using the raw data and generated series data, respectively. The probability densities of the metric in the three cases are compared. The PDFs of the metrics are estimated using the kernel density functions. As illustrated in Figure 12, the estimated PDFs based on the raw data and generated new series have a good match for the case $\Gamma = 200$ feet. However, the small deviation exists for the case $\Gamma = 100$ feet, and $\Gamma = 400$ feet. Since the numbers of the detected wind shear ramps are limited, increasing the raw data set might improve the results. Nevertheless, the larger difference in PDFs among the three cases can be captured using the generated series. The constructed results also show the same statistical characteristics in wind shear intensity with the raw data. The KL expansion model of wind shear ramps can be used for the analysis of the wind shear intensity.

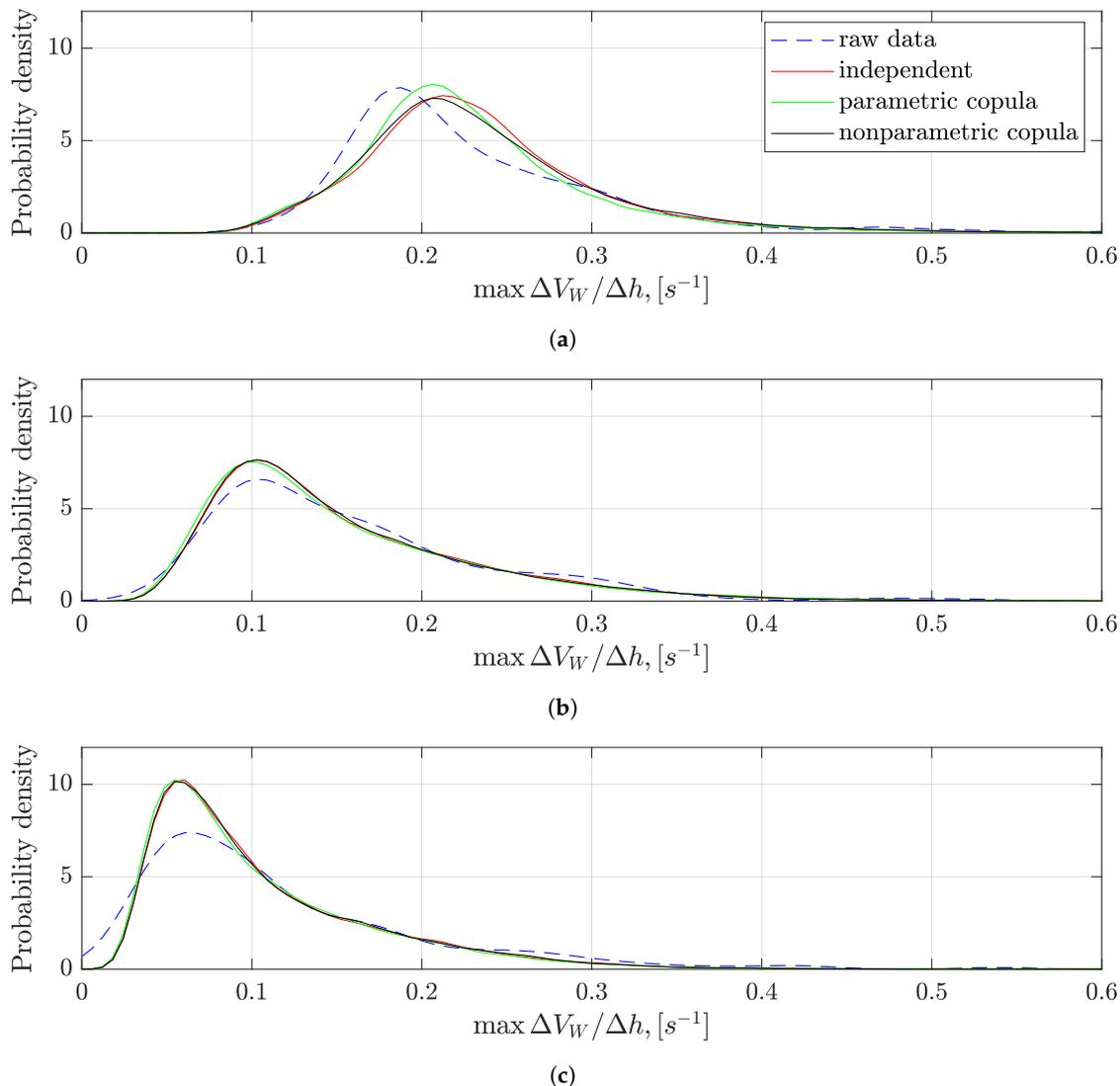


Figure 12. The estimated PDFs of the change rate of headwind along with the altitude. (a) The change rate of headwind for $\Delta V_W \geq 5$ knots, $\Gamma = 100$ feet; (b) The change rate of headwind for $\Delta V_W \geq 5$ knots, $\Gamma = 200$ feet; (c) The change rate of headwind for $\Delta V_W \geq 5$ knots, $\Gamma = 400$ feet.

4. Conclusions

This paper presents an algorithm to construct a stochastic wind model based on the operational flight data using the KL expansion method with the vine copula dependence model. The generated wind series based on the KL construction model has the same statistical characteristics as the raw data. For the stationary wind process like turbulence, the similarity of the KL expansion and SR method is pointed out in terms of construction equations, and the construction results of the turbulence reveal that two construction methods coincide. For the non-stationary wind process, KL construction results of headwind and horizontal wind shear ramps validate that the generated series follow the statistical characteristics of the raw data. Therefore the KL expansion model of wind series can be used in simulations for the quantitative analysis.

The vine copula sampling allows us to capture the dependence among the original random coefficients ζ_k . Series generated via vine copula sampling reduce the deviation from higher order moments of the raw series compared to independent sampling. For further study, more flight data is required to improve accuracy. Additionally, the constructed wind model will be integrated into the incident model simulation settings to quantify the influence of the wind series on operational flight safety.

Author Contributions: Conceptualization and methodology, X.W., L.B., and C.C.; software, X.W.; validation, X.W.; formal analysis, X.W.; investigation, X.W. and L.B.; writing—original draft preparation, X.W.; writing—review and editing, X.W., L.B., and C.C.; visualization, X.W.; supervision, F.H. and C.C.; funding acquisition, F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Technical University of Munich.

Acknowledgments: The authors want to acknowledge the Technical University of Munich (TUM) for supporting the publication in the framework of the Open Access Publishing Program.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses; in the writing of the manuscript, or in the decision to publish the result.

Abbreviations

The following abbreviations are used in this manuscript:

KL	Karhunen–Loève
SR	Spectral representation
PSD	Power spectral density

Appendix A. Three Dimensional Joint Probability Density Function Decomposition

According to the definition of conditional probability, the joint probability density f of three random variables X_1, X_2, X_3 can be reconstructed as below:

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f_1(x_1). \quad (\text{A1})$$

According to the Equation (12), $f_{3|12}(x_3|x_1, x_2)$ and $f_{2|1}(x_2|x_1)$ can be written as below:

$$\begin{aligned} f_{2|1}(x_2|x_1) &= \frac{f_{12}(x_1, x_2)}{f_1(x_1)} \\ &= c_{12}(F_1(x_1), F_2(x_2))f_2(x_2), \\ f_{3|12}(x_3|x_1, x_2) &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))f_{3|2}(x_3|x_2) \\ &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))c_{23}(F_2(x_2), F_3(x_3))f_3(x_3), \end{aligned} \quad (\text{A2})$$

where $c_{ij;k}$ is the copula density between the random variables X_i and X_j for a given X_k . After substitution of $f_{3|12}(x_3|x_1, x_2)$ and $f_{2|1}(x_2|x_1)$ in Equation (A1), the joint probability density f is represented by three bivariate copula densities and three marginal probability densities as below:

$$\begin{aligned} f(x_1, x_2, x_3) &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times c_{23}(F_2(x_2), F_3(x_3)) \times c_{12}(F_1(x_1), F_2(x_2)) \\ &\times f_1(x_1) \times f_2(x_2) \times f_3(x_3). \end{aligned} \quad (\text{A3})$$

Appendix B. Marginal Distributions of ζ_k in Headwind Modeling

The fitted marginal distributions can be used to generate new realizations of ζ_k individually for independent sampling. In addition, to study dependence among several variables using a copula approach, we also have to first find appropriate marginal distributions for each variable involved and then transform it using the fitted marginal distributions to the uniform copula space. This separates the marginal behavior from the underlying dependence structure, which is then captured by the copula.

Table A1. Estimated marginal distributions of ζ_k in headwind modeling.

ζ_k	Distribution	Parameters
ζ_1	Generalized Extreme Value	$[k, \sigma, \mu] = [-0.2220, 0.9704, -0.3804]$
ζ_2	Logistic	$[\mu, \sigma] = [-0.0118, 0.5587]$
ζ_3	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0222, 0.8706, 8.1359]$
ζ_4	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0263, 0.7279, 3.9941]$
ζ_5	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0158, 0.7763, 4.8160]$
ζ_6	t Location-Scale	$[\mu, \sigma, \nu] = [0.0141, 0.7007, 3.6527]$
ζ_7	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0501, 0.7124, 3.7186]$
ζ_8	t Location-Scale	$[\mu, \sigma, \nu] = [0.0060, 0.6046, 2.7169]$
ζ_9	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0234, 0.6777, 3.3137]$
ζ_{10}	t Location-Scale	$[\mu, \sigma, \nu] = [0.0051, 0.6529, 3.2108]$
ζ_{11}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0087, 0.5859, 2.4526]$
ζ_{12}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0262, 0.6376, 2.9137]$
ζ_{13}	t Location-Scale	$[\mu, \sigma, \nu] = [0.0075, 0.5823, 2.4866]$
ζ_{14}	t Location-Scale	$[\mu, \sigma, \nu] = [0.0231, 0.6695, 3.2488]$
ζ_{15}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0345, 0.6110, 2.6175]$
ζ_{16}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0016, 0.6387, 2.9972]$
ζ_{17}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0095, 0.5603, 2.3219]$
ζ_{18}	t Location-Scale	$[\mu, \sigma, \nu] = [0.0147, 0.5131, 2.1901]$
ζ_{19}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0003, 0.5795, 2.4952]$
ζ_{20}	t Location-Scale	$[\mu, \sigma, \nu] = [-0.0211, 0.6468, 2.9694]$

Appendix C. Dependence of ζ_k in Headwind Modeling

In Figure A1, marginal histograms of ζ_k^U are shown in the diagonal. The number from 1 to 20 denotes the corresponding variable ζ_k . All marginal histograms of ζ_k^U are close to uniform, which indicates that the estimated marginal distributions are appropriate. The upper triangle shows the sample plots of the copula data ζ_k^U , and the number marked in red color is the estimated Kendall's τ between two ζ_k^U . Empirical contour plots of the normalized copula data ζ_k^Z are shown in the lower triangle.

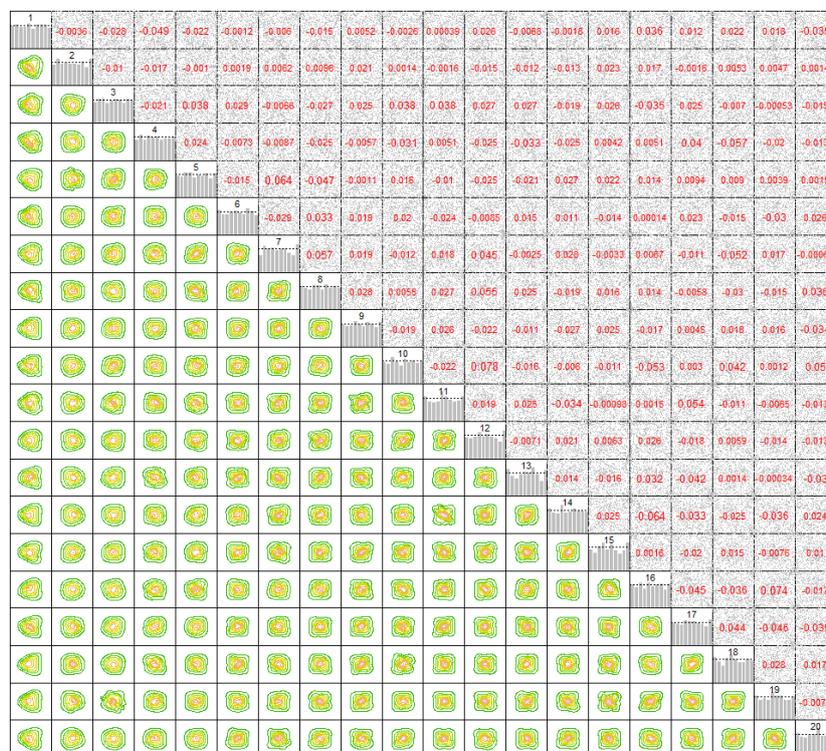


Figure A1. Empirical normalized pairwise contour plots in the headwind modeling (generated using R package “VineCopula” [25]).

As the truncated order of the KL expansion is set to 20 in the case of headwind construction, 20 dependent random variables ζ_k are obtained. Therefore, $\frac{20 \cdot (20-1)}{2} = 190$ bivariate copula densities are estimated. Normalized contour plots of the first three kinds of conditional copula density are shown in Figure A2. The parametric copula density and nonparametric bivariate copula density functions are used, respectively. The label above the contour plot indicates the related random variables ζ_k . For example, the label of first top left copula '2,5;3,15' indicates $c_{2,5;3,15}$, which is the copula density between the random variables ζ_2 and ζ_5 for the given ζ_3 and ζ_{15} . The concentric circles in the contour plots indicate that the two variables are independent, which is obvious in many contour plots of Figure A2b. Otherwise, dependence exists. Several dependence structures are apparent, especially in unconditional bivariate copula densities, as shown in the last row of Figure A2a,b.

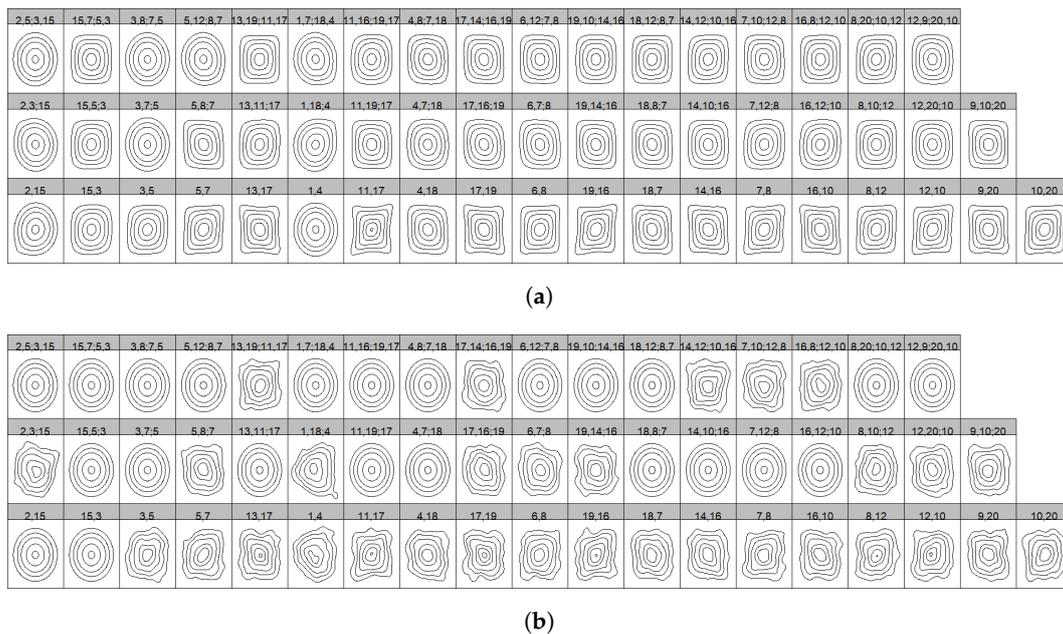


Figure A2. Normalized contour plots of bivariate copulas in headwind modeling (generated using R package 'rvinecopulib' [26]). (a) Contour plots using the parametric bivariate copula; (b) Contour plots using the nonparametric bivariate copula.

References

1. International Civil Aviation Organization. Safety Management Manual (Doc 9859). 2013. Available online: <https://www.skybrary.aero/bookshelf/books/644.pdf> (accessed on 5 July 2020).
2. Hoblit, F.M. *Gust Loads on Aircraft: Concepts and Applications*; AIAA Education Series; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 1988. [CrossRef]
3. Frost, W.; Camp, D.W.; Wang, S.T. *Wind Shear Modeling for Aircraft Hazard Definition*; National Aeronautics and Space Administration Marshall Space Flight Center: Huntsville, AL, USA, 1978.
4. Tatom, F.B.; Smith, S.R.; Fichtl, G.H.; Campbell, C.W. Simulation of atmospheric turbulent gusts and gust gradients. *J. Aircr.* **1982**, *19*, 264–271. [CrossRef]
5. European Aviation Safety Agency. Easy Access Rules for All Weather Operations (CS-AWO). 2018. Available online: <https://www.easa.europa.eu/sites/default/files/dfu/Easy%20Access%20Rules%20for%20CS-AWO.pdf> (accessed on 5 July 2020).
6. Haverdings, H.; Chan, P.W. Quick access recorder data analysis software for windshear and turbulence studies. *J. Aircr.* **2010**, *47*, 1443–1447. [CrossRef]
7. Sembiring, J.; Drees, L.; Holzapfel, F. Extracting Unmeasured Parameters Based on Quick Access Recorder Data Using Parameter-Estimation Method. In Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Boston, MA, USA, 19–22 August 2013. [CrossRef]

8. Drees, L.; Holzapfel, F. Determining and quantifying hazard chains and their contribution to incident probabilities in flight operation. In Proceedings of the AIAA Modeling and Simulation Technologies Conference, Minneapolis, MN, USA, 13–16 August 2012. [CrossRef]
9. Drees, L.; Sembiring, J.; Höhdorf, L.; Wang, C.; Holzapfel, F. *IATA Safety Report 2013, Section 9: Predictive Analysis*; International Air Transport Association: Montreal, Geneva, 2014.
10. Moorhouse, D.; Woodcock, R. *US Military Specification MIL-F-8785C*; US Department of Defense: Arlington, VA, USA, 1980.
11. Dehling, H.G.; Gottschalk, T.; Hoffmann, A.C. *Stochastic Modelling in Process Technology*; Elsevier: Amsterdam, The Netherlands, 2007; pp. 29–46.
12. Poirion, F.; Zentner, I. Stochastic model construction of observed random phenomena. *Probabilistic Eng. Mech.* **2014**, *36*, 63–71. [CrossRef]
13. Poirion, F.; Benoit, M. Stochastic Model Construction of Correlated Nonlinear Wave Fields. In Proceedings of the ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering, St. John's, NL, Canada, 31 May–5 June 2015. [CrossRef]
14. Ramsay, J.O.; Silvermann, B.W. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005. [CrossRef]
15. Grigoriu, M. Evaluation of Karhunen–Loève, spectral, and sampling representations for stochastic processes. *J. Eng. Mech.* **2006**, *132*, 179–189. [CrossRef]
16. Chen, E.J.; Ding, L.; Liu, Y.; Ma, X.; Skibniewski, M.J. On spectral representation method and Karhunen–Loève expansion in modeling construction material properties. *Arch. Civil Mech. Eng.* **2018**, *18*, 768–783. [CrossRef]
17. Czado, C. *Analyzing Dependent Data with Vine Copulas*; Springer International Publishing: New York, NY, USA, 2019. [CrossRef]
18. Le Maître, O.; Knio, O.M. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 18–26. [CrossRef]
19. Joe, H.; Kurowicka, D. *Dependence Modeling: Vine Copula Handbook*; World Scientific: Singapore, 2011. [CrossRef]
20. Sklar, A.; SKLAR, A.; Sklar, C. Fonctions de répartition à dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
21. Rosenblatt, M. Remarks on a multivariate transformation. *Ann. Math. Stat.* **1952**, *23*, 470–472. [CrossRef]
22. Dissmann, J.; Brechmann, E.C.; Czado, C.; Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Comput. Stat. Data Anal.* **2013**, *59*, 52–69. [CrossRef]
23. Ricker, D.W. *Echo Signal Processing*; Springer: New York, NY, USA, 2012; Volume 725. [CrossRef]
24. Höhdorf, L.; Siegel, J.; Sembiring, J.; Koppitz, P.; Holzapfel, F. Reconstruction of Aircraft States During Landing Based on Quick Access Recorder Data. *AIAA J. Guid. Control Dyn.* **2017**, *40*, 1–6. [CrossRef]
25. Thomas, N. Statistical Inference of Vine Copulas: Package ‘VineCopula’. 2019. Available online: <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf> (accessed on 5 July 2020).
26. Thomas, N. High Performance Algorithms for Vine Copula Modeling. 2019. Available online: <https://cran.r-project.org/web/packages/rvinecopulib/rvinecopulib.pdf> (accessed on 5 July 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).