Complexity

Control and Optimization of Complex Agriculture Systems Facing New Risks

Lead Guest Editor: François Pérès
Guest Editors: R. Willem Vervoort, Bertrand Jouve, and Ronald van Nooijen

Control and Optimization of Complex Agriculture Systems Facing New Risks

Complexity

Control and Optimization of Complex Agriculture Systems Facing New Risks

Lead Guest Editor: François Pérès
Guest Editors: R. Willem Vervoort, Bertrand Jouve, and Ronald van Nooijen

Contents

Hindawi
Complexity
Volume 2021, Article ID 8848424, 13 pages
https://doi.org/10.1155/2021/8848424

WILEY | Hindawi

Research Article

Can Global Economic Policy Uncertainty Drive the Interdependence of Agricultural Commodity Prices? Evidence from Partial Wavelet Coherence Analysis

Siaw Frimpong,[1] **Emmanuel N. Gyamfi**,[2] **Zangina Ishaq**,[1] **Samuel Kwaku Agyei**,[1] **Daniel Agyapong**,[1] **and Anokye M. Adam**[1]

[1]*Department of Finance, School of Business, University of Cape Coast, Cape Coast, Ghana*
[2]*GIMPA Business School, Accra, Ghana*

Correspondence should be addressed to Anokye M. Adam; aadam@ucc.edu.gh

Received 6 August 2020; Revised 10 May 2021; Accepted 26 May 2021; Published 9 June 2021

Academic Editor: François Pérés

This paper employed wavelet coherence and partial wavelet coherence to investigate the time-frequency effect of global economic policy uncertainty on the comovement of five agricultural commodities such as maize, oat, rice, soybean, and wheat using monthly data from January 1997 to December 2019. In general, we observed heterogeneity in comovement structures of the agricultural commodities market at different time-frequency scales which are profound at high frequencies from the bivariate wavelet coherence. The partial wavelet coherence analysis shows that global economic policy uncertainty is a driver of agricultural commodity market connectedness. This implies that extreme changes in economic policy uncertainty have the tendency to influence commodity price comovement. This poses risk to the stability of the agricultural commodities market, which requires the policymaker's intervention to protect against the spillover risk contagion effect in uncertain times.

1. Introduction

The surge in price and price volatility of agricultural commodities, especially food prices, has attracted the attention of academics, policymakers, investors, farmers, and consumers to this market because of its immediate impact on food security around the world, particularly low-income food-deficit countries. The prices of agricultural commodities have experienced long-term and sharp fluctuations since the year 2000. The prices of major agricultural commodities, from 2006, have generally exhibited upward trends with a sharp fluctuation in 2013 and 2014. These behaviours have attributed to external factors such as macroeconomic uncertainties, agricultural production, financial crises, large and persistent demand, biofuels demand, different stock market phases, and climate warming [1–6]. For example, the outbreak of COVID-19 in 2019 has overturned the stagnation in food prices after its downward trend in 2015-2016. The Food and Agricultural Organisation Food Price Index surged to its highest level since 2014.

Moreover, the financialization in commodity markets has changed the dependency structure of agricultural commodity markets [7, 8]. Consequently, the traditional description of commodities in general as an asset class that reliably delivers returns with low correlation to the stock market has changed [9–13]. It instructive to note that agricultural commodity financialization has increased the comovement and volatility spillover within its market and with the traditional asset classes, limiting its diversification benefits [14–17]. The interest of academics and market participants has therefore been drawn to the level of comovement and predictors of commodity prices and volatility spillover [7, 17, 18]. The comovement or otherwise of these commodities provides important information to portfolio investors on diversification opportunities and policymakers on policy interventions to mitigate price fluctuations on the world poorest people.

Several studies have examined the drivers of agricultural commodities, and different external factors have been identified which include crude oil prices, stock market phases, exchange rate, and production indices [14]. The interdependency structure of agricultural commodity prices has been studied and observed with varying degrees of interdependence [1, 17]. For example, Živkov et al. [7] analysed the comovement structure of five agricultural commodities (corn, wheat, soybean, rice, and oats) using wavelet coherence and phase difference and observed low coherence at shorter time-horizons among the commodities, while periodic coherence at longer time-horizons was found. In the work of Amrouk et al. [17], a multivariate Copula-DCC-GARCH model was employed to investigate the price dynamics among maize, wheat, soybeans, cotton, coffee, cocoa, and sugar futures prices. The results show that the intensity of interaction varies considerably over the sample time but is generally positive and stronger during the period 2007–2012 associated with high commodity prices and financial market stress. A recent study by Yuan [19] showed that agricultural commodity markets of corn, wheat, soybeans, soya oil, cotton, and oats tend to crash (boom) together during extreme events using the Copula-GARCH model.

These strands of findings bring to the fore the fundamental question of what drives commodity price interdependence. The 2007–2011 general boom in international commodity prices has not only been cited as the cause of agricultural commodity hikes but also increases the interdependency [5, 18–20]. Similarly, Wang et al. [6] linked episodes of crisis such as adverse weather, export bans, financial crisis, and depreciation of dollars to the changing behaviour of agricultural commodities (food prices). These have heightened the interest of uncertainty on the behaviour of agricultural commodities.

The theoretical evidence of Keynes [21] and Working [22] identified term structure, hedging pressure, and risk factors as the main drivers of commodity return behaviour. Economic policy uncertainty represents information about the future state of the economy with regards to regulation, fiscal policy, and monetary policy. Pastor and Veronesi [23] theoretical model showed that government economic policy uncertain will lead to both correlations in stock prices and increases in volatilities. By analogy, we expect that global economic policy uncertainty will lead to similar asset price comovements in global financial markets. Frankel [24] showed the importance of monetary policy for commodity prices. As an element of economic policy, Frankel [24] argument on monetary policy also supported the links between economic policy uncertain and commodity prices.

While the above views support the commodity price and economic policy uncertain linkage, one may still ask what about comovement of commodity prices? Policy uncertainty is also found related to corporate investment [25, 26]. From an investment perspective, global economic policy uncertainty affects investment in the agricultural produces affecting supply and hence the commodity price comovement. Global economic uncertainty also increases calls and attempts for global policy coordination [27]. Such policy coordination in response to global economic policy uncertainty implies that commodity prices may share common shocks. Another channel through which global economic policy uncertain may drive commodity price comovements is the domestic agricultural policy response to global economic policy uncertainty. The intuition is that not all countries are dominant producers of all commodities, and with global economic uncertainty affecting terms of trade, commodity prices comovements will result in the equilibrium as the countries adjust to the global economic policy uncertainty. Based on the number of newspaper articles regarding policy uncertainty from national or regional leading newspapers, Baker et al. [28, 29] developed an index to measure EPU, which is more universal and applicable for comparison and can continuously track policy uncertainty.

Subsequently, numerous studies provide ample evidence on the EPU effect on commodity returns (Wang et al. [30]; Reboredo et al. [31]; Shahzad et al. [32]) and economic agents that affect commodity markets such as oil prices and stock returns [33–37]. The role of EPU as the driver of the time variation in asset correlations has also received attention. Fang et al. [38] showed that EPU harms U.S. stock and bond market correlations, while Fang et al. [39] documented a positive policy uncertainty effect on the long-run oil-stock correlation. Recently, Badshah et al. [40] documented that the EPU effect on stock-commodity correlation is stronger during weak economic conditions. As far as we know, there is no previous study that analysis the EPU effect on agricultural commodity price comovement.

This study extends the literature on the EPU effect on asset correlations comovement and examines the effect of EPU on the comovement of agricultural commodities (maize, wheat, soya bean, rice, and oats) returns. The focus on agricultural commodities is motivated by the importance of risk mitigation of food prices to food security concerns. Correlation plays an important role in asset allocation decisions, risk management, and analysis of risk transmission across assets. Specifically, the study provides new evidence by adjusting for the possible influence of EPU using partial wavelet coherence and wavelet coherence, which is lacking in the comovement of agricultural commodities in previous studies (see, for example, Živkov et al. [7]; Amrouk et al. [17]). The advantage of the use of wavelet-based methods is well documented in the literature (see [41–44]). This paper, therefore, investigates the partial correlations in the agricultural commodity prices by including (excluding) the influence of EPU.

2. Methodology

The application of wavelet transforms is widespread in time series econometrics to deal with nonstationary problems in time series. Wavelet transforms come in two forms, namely, discrete wavelet transforms (DWT) and continuous wavelet transforms (CWTs). In this paper, we employ only CWT for our analysis since DWT is useful for noise reduction and data compression, while CWT allows for good identification and isolation of periodic signals, by providing a balance between localisation of time and frequency and appears to

provide a better trade-off between detecting oscillations and peaks or discontinuities [41, 43, 45].

2.1. Continuous Wavelet Transform and Wavelet Coherence. The CWT is a powerful technique to assess the dynamics of nonstationary variables over frequency and time domain space. As required, the localized time-frequency space and zero mean must be ensured in the wavelet function decomposition of the time series. This allows for information from the local neighbourhood based on decomposition to be obtained. We specifically employ wavelet coherence under Morlet specification, which is defined as follows:

$$\Psi_{u,s}(t) = s^{-1/2}\Psi\left(\frac{t-u}{s}\right), \Psi(\cdot) \in L^2(R), \quad (1)$$

where $s^{-1/2}$ is the normalization factor, which ensures that the unit variance of the wavelet $\Psi_{u,s}(t)^2 = 1$; u is the location parameter, which provides the exact position of the wavelet; and s is the scale dilation parameter, defining how the wavelet is stretched or dilated. Thus, the Morlet wavelet can be defined as follows:

$$\varphi^M(t) = \pi^{-1/4} e^{i\omega_o t} e^{-t^2/2}, \quad (2)$$

where ω_o is the central frequency of the wavelet. We follow the extant literature to set $\omega_o = 6$ [42, 45, 46].

A continuous wavelet transform $W_x(u,s)$ is obtained via the projection of a wavelet $\Psi(\cdot)$ on the examined series $x(t)$ so that

$$W_x(u,s) = \int_{-\infty}^{\infty} x(t)\, s^{-1/2}\Psi^*\left(\frac{t-u}{s}\right)dt, \quad (3)$$

where $\Psi^*(\cdot)$ is a complex conjugate of $\Psi(\cdot)$. By projecting the specific wavelet $\Psi(\cdot)$ onto the selected time series, we easily obtain $W_x(u,s)$. The main advantage of a CWT is its ability to decompose and reconstruct the function $x(t) \in L^2(\mathrm{R})$ as follows:

$$x(t) = \frac{1}{C_\varphi}\int_0^{\infty}\left[\int_0^{\infty} W_x(u,s)\Psi_{u,s}(t)du\right]\frac{ds}{s^2}, s>0. \quad (4)$$

The power spectrum analysis can then be calculated using Equation (4), with the specification of the variance being

$$x^2 = \frac{1}{C_\varphi}\int_0^{\infty}\left[\int_0^{\infty} W_x(u,s)^2 du\right]\frac{ds}{s^2}, s>0. \quad (5)$$

The red noise background spectrum is employed to define the null hypothesis in significance tests for peaks in the wavelet power spectrum. Following Torrence et al. [47] and Torrence et al. [48], the red noise background spectrum is computed using Monte Carlo simulations. Therefore, the corresponding local wavelet power spectrum distribution for each time n and scale s can be

$$D\left(\frac{W_n^x(s)^2}{\delta_x^2} < p\right) \Rightarrow \frac{1}{2}P_f X_v^2, \quad (6)$$

where the mean spectrum at Fourier frequency f is denoted *by* P_f. The wavelet scale s corresponds to the Fourier frequency (s 1/f). The real wavelet has $v = 1$, and the complex wavelet $v = 2$. The variance of the corresponding variable is denoted by δ_x^2.

Following Rua et al. [44], we define the cross-wavelet transform of two commodities market series (X) and (Y) as follows:

$$W_n^{XY}(s) = W_n^X(s)W_n^{Y^*}(s), \quad (7)$$

where $W_n^X(s)$ and $W_n^Y(s)$ are individual wavelet spectra, u denotes the position, s denotes the scale, and $*$ indicates complex conjugation. The cross-wavelet transform shows the area in time-space with high common power. Therefore, $W_n^{Y^*}(s)$ is the complex conjugate of $W_n^Y(s)$. The cross-wavelet power $|W_n^{XY}(s)|$ measures the mutual local covariance on each scale. Therefore, the wavelet coherence of the two time series and $y = \{y_n\}$ is defined by searching the frequency bands and time intervals in which they covary. This provides a useful tool for detecting comovement in commodities markets. Grinsted et al. [38] defined it as the squared absolute value of normalizing a wavelet cross-spectrum to a single wavelet power spectrum. Similarly, we define wavelet coherence as squared wavelet coefficient as follows:

$$R^2(x,y) = \frac{\left|S\left(s^{-1}W_{xy}(u,s)\right)\right|^2}{S\left(s^{-1}\left|W_x(u,s)\right|^2\right)S\left(s^{-1}\left|W_y(u,s)\right|^2\right)}, \quad (8)$$

where S denotes the smoothing parameter, which balances resolution and significance. Also, the bias problem in the wavelet power spectrum and wavelet cross-spectrum is eliminated by the normalizing function of the wavelet coherence. The values of the wavelet coherence coefficient satisfy the inequality $0 \le R^2(x,\ y) \le 1$. Wavelet coherence close to one shows a higher similarity between the time series, while coherence near-zero depicts no relationship.

As indicated by Madaleno et al. [49] and Torrence et al. [47], the phase for wavelet depicts any lead/lag linkages between two time series and can be defined as follows:

$$\varnothing_{xy} = \tan^{-1}\left(\frac{\mathfrak{F}\{W_n^{xy}\}}{\mathfrak{R}\{W_n^{xy}\}}\right), \varnothing_{xy} \in [-\pi, \pi], \quad (9)$$

where $\mathfrak{F}$ and $\mathfrak{R}$ are the imaginary and real parts of the smoothed cross-wavelet transform, respectively. In the wavelet coherence map, directional arrows are used to distinguish different phase patterns. For instance, if $x(t)$ and $y(t)$ are in phase (antiphase), the arrow points to the right (left). Similarly, if the arrow points down (or up), this implies that $y(t)$ (or $x(t)$) is leading.

2.2. Partial Wavelet Coherence. Wavelet coherence reveals intermittent correlations and provides the significant map when a correlation is real [43, 50, 51]. In the case where there is an intervening effect from a different variable, wavelet coherence is unable to give an accurate correlation. Partial wavelet coherence (PWC) is a technique similar to the partial correlation that helps find the resulting WTC between two time series x and y after eliminating the influence

of the time series z. Similarly, in wavelet applications, PWC can help to eliminate the influence of time series $z(t)$ on the wavelet coherence between $x(t)$ and $y(t)$. Mihanović et al. [52] extended the concept of simple linear correlation and suggested that the PWC can be defined using an equation similar to the partial correlation squared as follows:

$$R_p^2(x,y,z) = \frac{\left|R(x,y) - R(x,z)\cdot R(x,y)^*\right|^2}{\left[1 - R(x,z)\right]^2\left[1 - R(y,z)\right]^2}, \tag{10}$$

where $R_p^2(x,y,z)$ ranges from 0 to 1 and has a similar interpretation as $R^2(x,y)$. Specifically, a low R_p^2 region observed where a high R^2 region is found indicates that the time series y does not have a significant influence on x. Instead, the time series z dominates the variance of x. If there is no difference between R_p^2 and R^2, both y and z have a significant influence on x. In this paper, x and y denote the agricultural commodities market returns while z denotes the EPU returns.

2.3. Data and Analysis. The data set for this analysis consists of monthly data on global economic policy uncertainty (EPU) index and monthly period average price of the five major agricultural commodities—maize, wheat, soybean, rice, and oats from January 1997 to December 2019—yielding 276 observations. The period is characterised by considerable global uncertainty, such as the global financial crisis, the sovereign debt crisis in Europe, Brexit, the US presidential election, and US-China trade tension, making it an ideal period for the study. The EPU developed by Baker et al. [28] was obtained from the website https://www.policyuncertainty.com/index.html and commodity prices from IMF Primary Commodity Prices Database. The analysis was based on the returns of monthly prices/indexes given as

$$r_t = \ln P_t - \ln P_{t-1}, \tag{11}$$

where r_t is the continuously compounded return and P_t and P_{t-1} are the respective current and previous prices/indexes.

We began the analysis by investigating the statistical distribution of returns; it is appropriate to have quick behavioural trajectories of the agricultural commodity prices and returns as well as EPU indices. Figure 1 shows the graphical representation of the trend of the time series plot of both indices and returns of the five agricultural commodities and EPU over the study period. A glance shows that all the commodity indices were trending upward and exhibited quite a similarity in dynamic patterns until the outbreak of the global financial crisis and then plunged in 2008. Recovery was observed between 2010 and 2013 and then nosedived until 2018. The recovery from 2018 has been one of a mixed.

The EPU, however, generally showed an upward trend over the study period except between 2003 and 2007. This is expected because of episodes of global financial and political uncertainty in recent years. This trajectory behaviour of all indices might induce correlations among them. Table 1 presents the pairwise correlation among the indices, and the results showed low to moderate correlation among the agricultural commodities and no correlation with EPU.

A glance at the descriptive statistics of the return presented in Table 2 reveals key features of the return rates of the series included in the study. All the return series had positive means in the period considered, and only two out of the six series had negative skewness. The negatively skewed returns indicate that higher losses are more frequent than higher gains. The returns of all series are relatively symmetrical except for rice. All the series depart from normality with high kurtosis except for oats, which means that the returns are heavy-tailed relative to a normal distribution. This is not surprising as it is a well-known stylized fact of assets in the financial literature [53].

In the first step, the pairwise wavelet coherence plots of the five selected agricultural commodities are presented as a benchmark before investigating the partial wavelet coherence. The wavelet coherence maps are used in this study to assess the strength of the interdependence both in time and frequency domains. The time component is displayed on the horizontal axis, while the vertical axis displays the frequency (the lower the frequency, the higher the scale). The region in time-frequency space where the pair-wise time series covary is indicated by the warm colour. The strength of the interdependence between paired series is indicated by the colour of the surface and depicted by the colour pallet. The warmer colour (red) represents regions with significant interrelation, while the colder colour (blue) signifies a lower dependence between the series. The crosshatch indicates regions inside the cone of influence, and the thick black contour indicates a 95% confidence level obtained from the Monte Carlo simulations, whereby cold regions beyond the significant areas represent time and frequency with no dependence in the series. An arrow in the wavelet coherence plot shows the lead/lag phase relations between the paired series. A zero-phase difference means that the two paired time series move together on a particular scale. When the arrows point to the right, the paired time series are said to be in phase (move in the same direction), while the arrows to the left indicate antiphase (move in the opposite direction). Arrows pointing to the right-down or left-up indicate that the first variable is leading, while arrows pointing to the right-up or left-down show that the second variable is leading.

Figure 2 shows the partial and bivariate wavelet coherence explaining the pairwise coherence plots between maize, wheat, soybean, rice, and oats with EPU as a covariate variable. The results of bivariate wavelet coherence are presented on the left-hand side, whereas that of partial wavelet coherence on the right-hand side. The wavelet coherence map in Figure 2(a) shows that the comovement between oats and the rest of the agricultural commodities such as maize, wheat, soybean, and rice is primarily concentrated in 0–32 month cycles. Several small islands were observed in a period of 0–8 month cycles throughout the study period and extended to 32 month cycles with intense connectedness between 2007 and 2013. Maize, however, showed strong interdependence with rice, soya bean, and wheat. It is worthy to note that maize perfectly serves as both

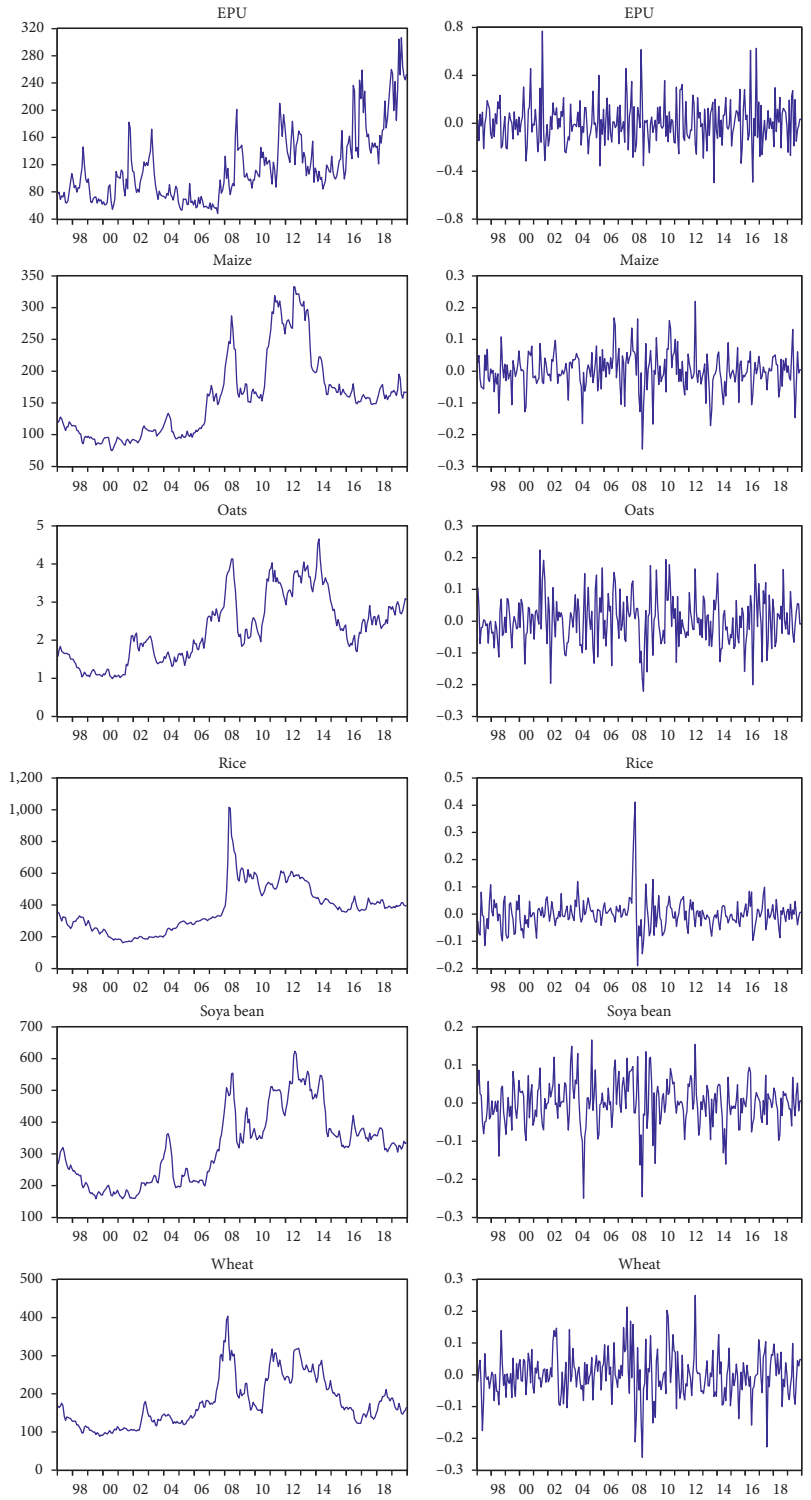


Figure 1: The plot of the raw series (left) and returns series (right).

a substitute and complements to these classes of cereals and therefore not a surprise to have its price returns highly connected. The coherence is very strong at 64–128 month cycle for all periods of study and gradually for all months' cycle after 2018. The comovements between rice, soya bean, and wheat are not different from the interrelations of maize with rice, soybean, and wheat. In general, the pattern of pairewise comovements among the agricultural commodities is in phase, but it is not clear which leads the market. It is important to note that in all cases, the correlations were profound during the period characterising the commodity price boom and the global financial crisis. This observation is consistent with the findings of prior studies that the price returns of related agricultural commodities move together

Table 1: Correlation matrix of EPU and selected agricultural commodities.

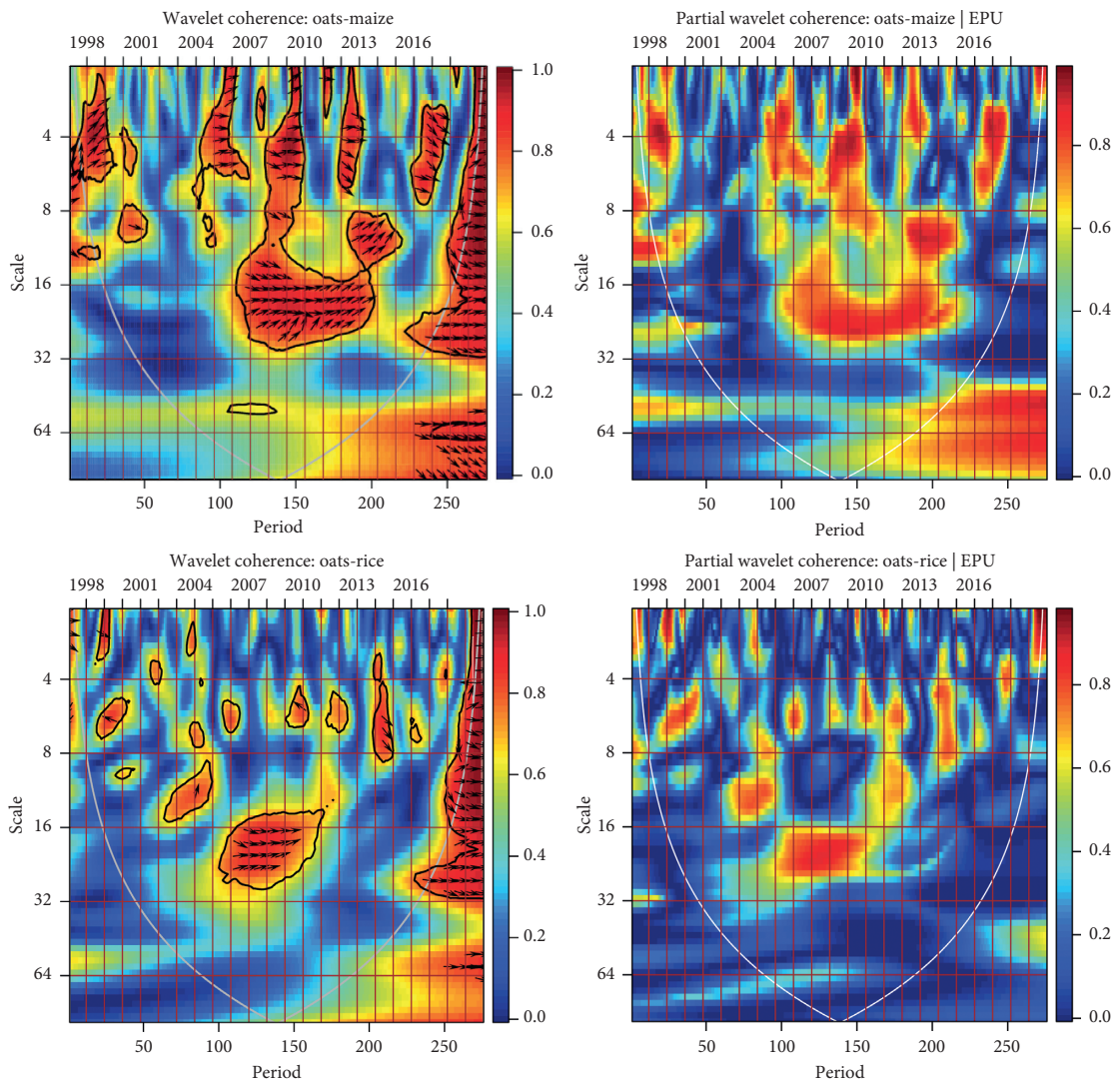
	EPU	Maize	Oats	Rice	Soya bean	Wheat
EPU	1.0000					
Maize	0.0089	1.0000				
Oats	−0.0005	0.4965***	1.0000			
Rice	−0.0086	0.1077*	0.1098*	1.0000		
Soya beans	−0.0755	0.6555***	0.3685***	0.0446	1.0000	
Wheat	−0.0076	0.5588***	0.3572***	-0.0663	0.4931***	1.0000

*, **, and *** represent 10%, 5%, and 1% significance levels, respectively.

Table 2: Descriptive statistics.

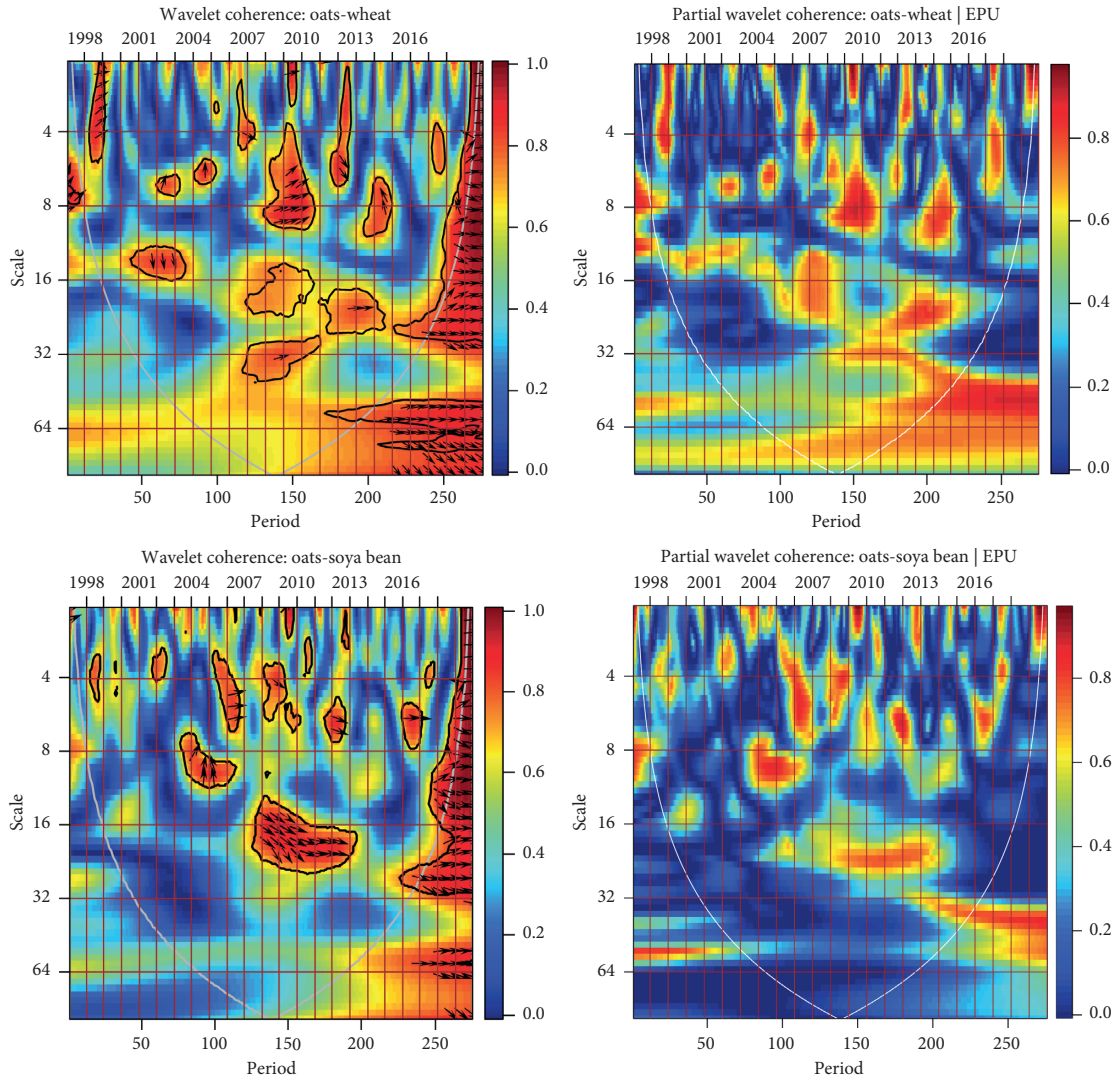
	EPU	Maize	Oat	Rice	Soya	Wheat
Mean	0.0044	0.0013	0.0024	0.0004	0.0008	0.0001
Median	-0.0061	0.0010	0.0026	-0.0027	0.0019	-0.0009
Maximum	0.7690	0.2197	0.2242	0.4116	0.1656	0.2502
Minimum	-0.4964	-0.2455	-0.2205	-0.1901	-0.2495	-0.2602
Std. dev.	0.1770	0.0583	0.0745	0.0553	0.0589	0.0695
Skewness	0.6474	-0.2158	0.1281	2.1990	-0.4524	0.0787
Kurtosis	4.9170	5.1900	3.4231	18.4220	5.0797	4.6689
Jarque–Bera	61.5388***	57.3005***	2.8136	2957.5670***	59.1504***	32.3154***
Observations	275	275	275	275	275	275

*, **, and *** indicate 10%, 5%, and 1% significance level, respectively.



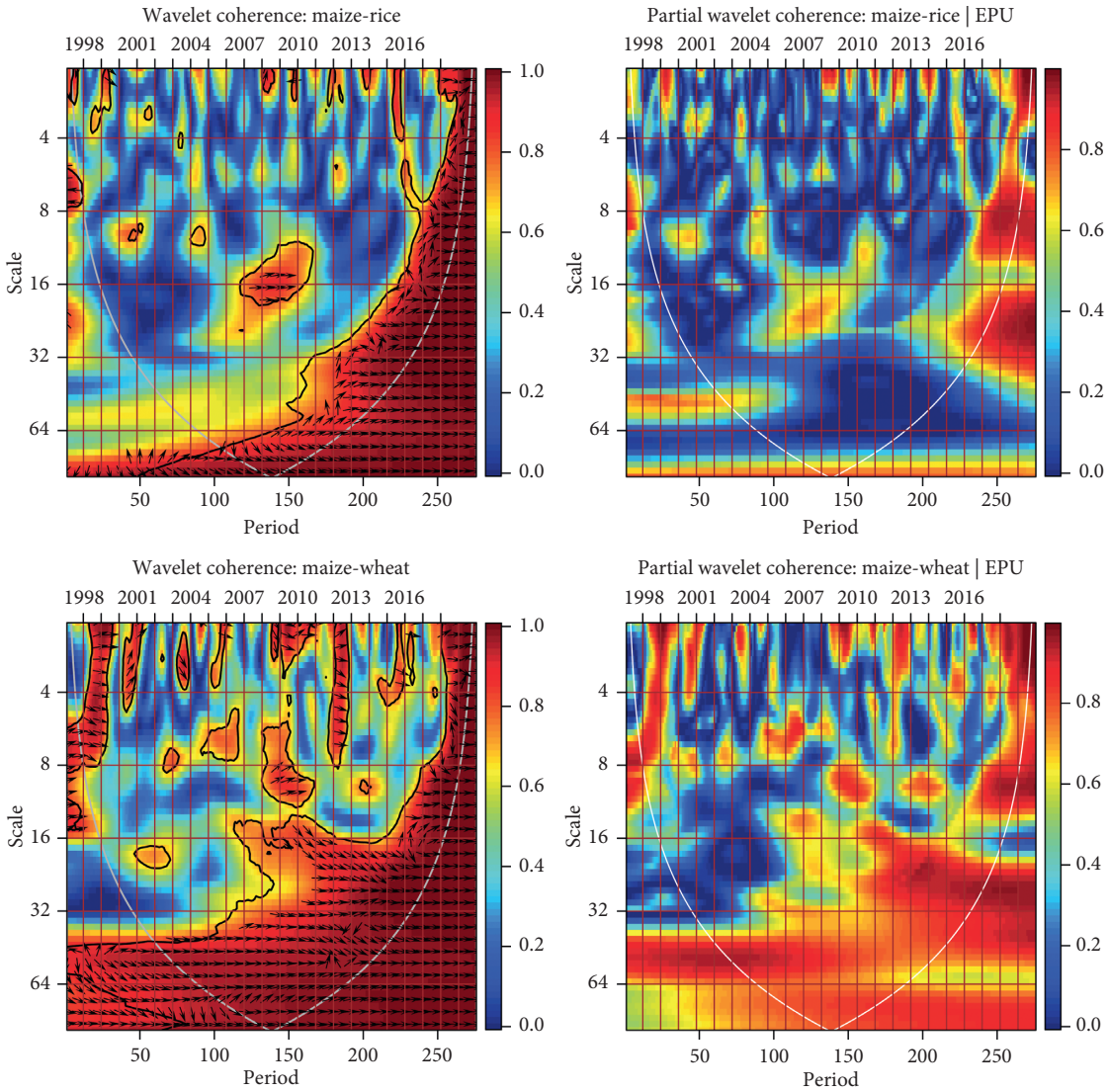
(a)

Figure 2: Continued.



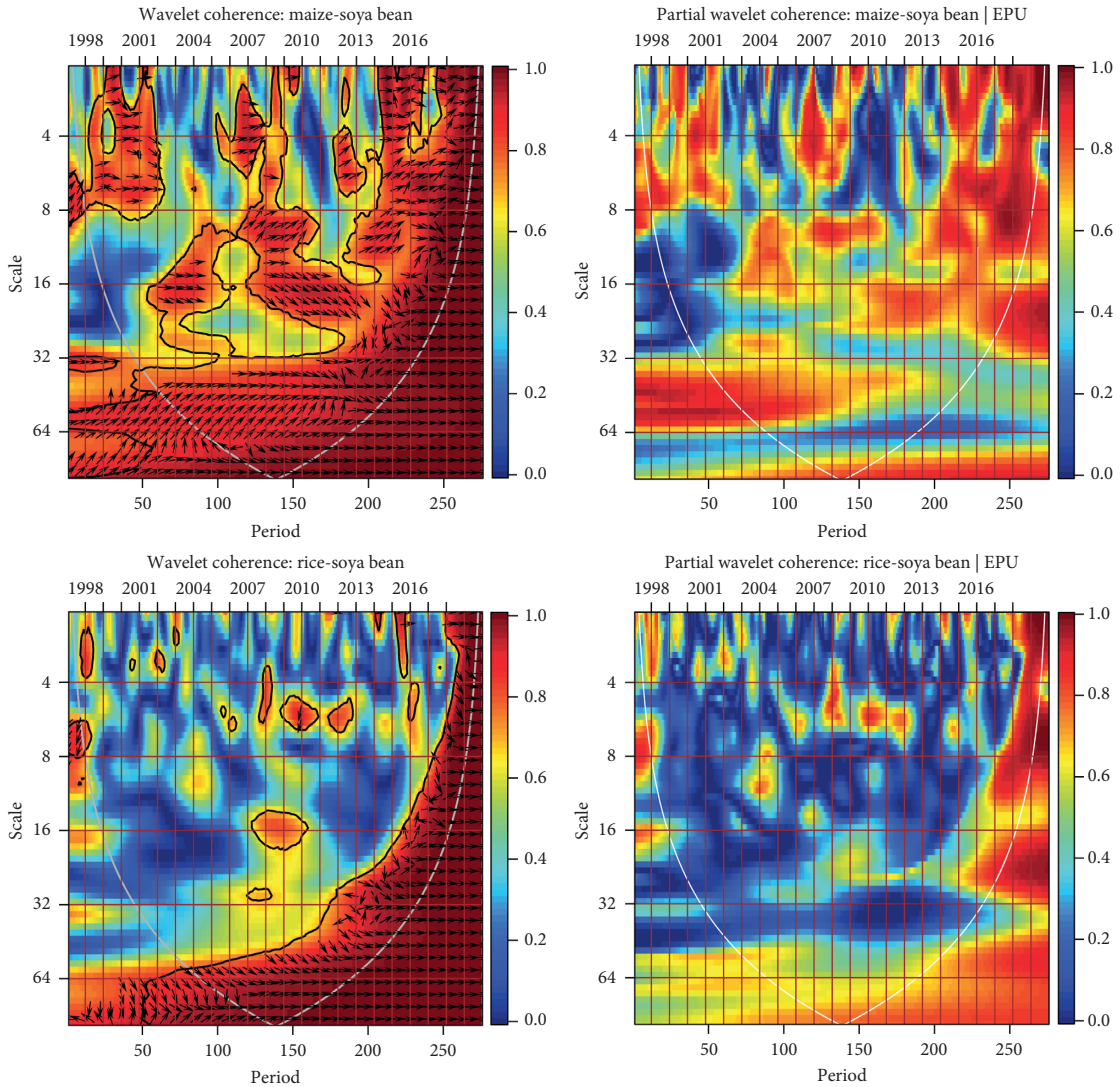
(b)

Figure 2: Continued.



(c)

Figure 2: Continued.



(d)

Figure 2: Continued.

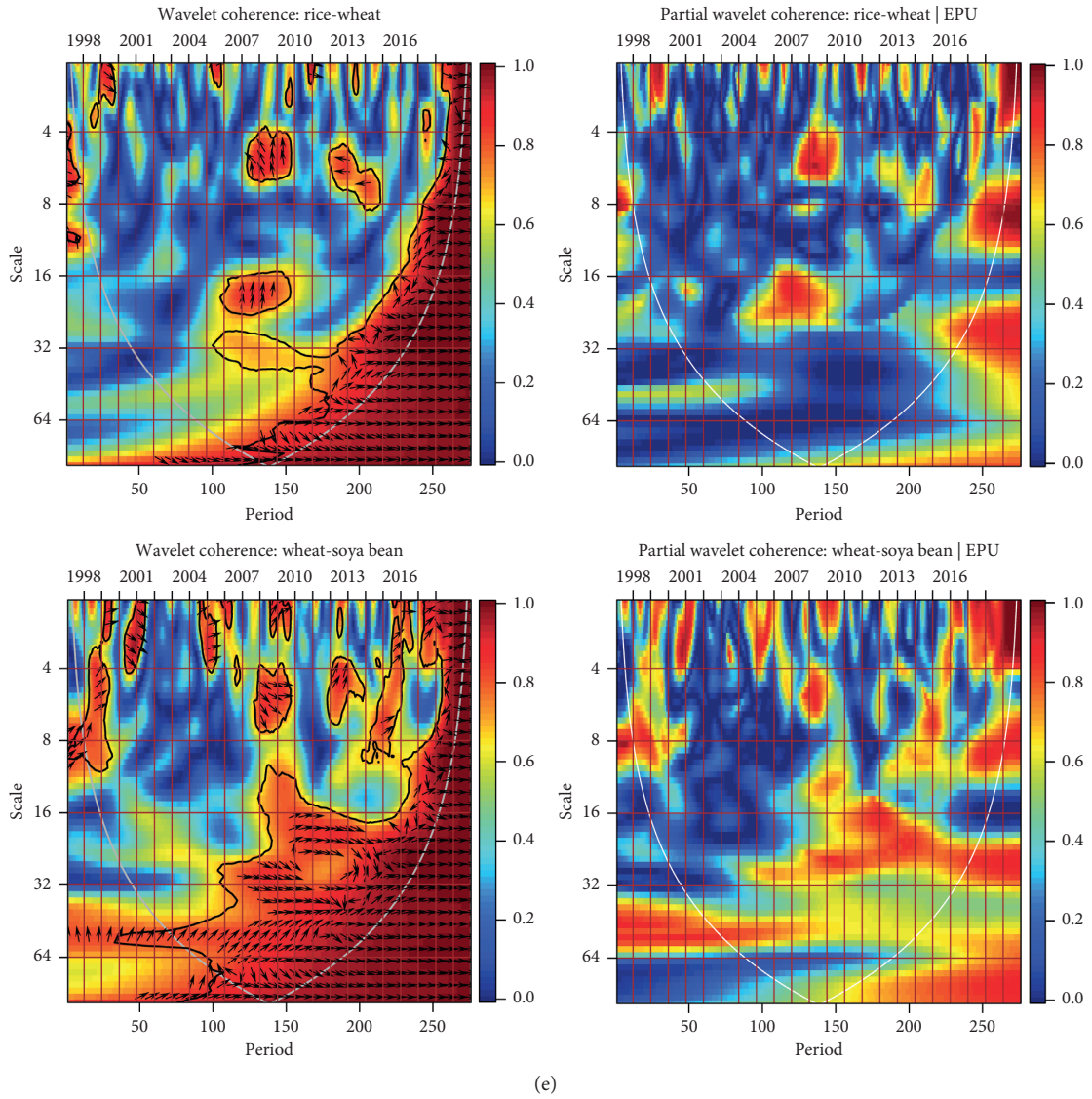


Figure 2: The left side plots bivariate wavelet coherence between commodity markets. The right side illustrates the plots of partial wavelet coherence between commodities markets when the effect of global economic policy is removed.

[19, 54]. In particular, we detected a high correlation during the financial crisis between the five agricultural commodities (corn, wheat, soybean, rice, and oats) similar to Živkov et al. [7] and Yuan [19] but more profound coherence. The difference in comovement structure emanates from the frequency of the data used; daily data are susceptible to speculation and noisy behaviour. The right side of Figure 2 illustrates the partial wavelet coherence among the commodities prices such as maize, wheat, soybean, rice, and oats with EPU as a covariate variable (i.e., excluding the influence of EPU). A substantial decrease in the strength of the coherence area among the commodities in all cases for the partial wavelet coherence indicates that EPU is a key driver of the comovement of the selected commodity prices. The finding highlights the importance of economic policy uncertainty as a driver of assets prices connectedness across financial markets, particularly commodity prices following the financialization of commodity markets [17, 55–57]. The financialization of commodity markets has affected the information transfer, improves market liquidity, and strengthens the commodity equity market comovement, making commodity prices susceptible to drivers of financial markets [4, 15, 17, 19, 20, 58, 59]. In summary, uncertainty in global economic policy as a significant driver of comovement between agricultural commodities price returns has policy implications. It suggests a greater integration of agricultural commodity markets with prices booming and crashing together during uncertain times. This calls for policymakers to devise strategies to mitigate economic policy uncertainty shocks to agricultural commodity prices. Again, the interdependency in the market poses risk to the stability of the agricultural commodities market, which requires the policymaker's intervention to protect against the spillover risk contagion effect in uncertain times. The correlation of assets is key in portfolio construction, diversification, and risk management in asset markets. The

differences in correlation structure at different time scales present unique information on both contagion and interdependence which are critical for commodity traders who desired to form a portfolio of agricultural commodities. The wavelet coherence at high frequency signifies contagion and interdependence at low frequencies. Consequently, investors must employ different trading strategies for different investment horizons. For example, oats correlate less with all other commodities in high time scales, and therefore the inclusion of oats in commodities portfolio will require a longer investment horizon to minimise the variance of the portfolio. By applying partial wavelet coherence using global economic policy uncertainty as a covariate, new information is provided to investors. A sharp decline of the comovement in any paired commodities suggests that economic policy uncertainties limit the diversification benefits of the agricultural commodities portfolio. It is therefore important that investors pay attention to global risk factors which tend to create uncertainty in economic policies. The comovement of agricultural commodities, especially upward movement, poses a threat to food insecurity to the poorest population. The evidence of sizable price spikes in times of uncertainty requires the government to implement policy interventions to minimise its impact on the poorest populations which will be enhanced through policy intervention.

3. Conclusion

This paper investigates the time-frequency effect of economic policy uncertainty on the comovement of five agricultural commodities using wavelet coherence and partial wavelet coherence. A monthly average price of maize, oat, rice, soybeans, and wheat from January 1997 to December 2019 as well as global economic policy uncertainty of the same frequency for the same period was used. The use of bivariate wavelet coherence together with partial wavelet coherence overcomes the problems of variation in time-frequency space and traditional two-variable methods of calculating comovement. The use of economic policy uncertainty as a covariate for the comovement between agricultural commodity returns provides a better understanding and information on the interaction of agricultural commodities markets.

The result from bivariate wavelet coherence shows heterogeneity in the comovement structure of the agricultural commodities market at different time-frequency scales which are profound at high frequencies. It is also evident that maize correlates strongly with all agricultural commodities studied. Similar to partial correlation, partial wavelet coherence analysis provides the resulting wavelet coherence between agricultural commodity prices after eliminating common factors of dependency of global economic policy uncertainty. By removing the effect of global economic policy uncertainty, a substantial decrease in the strength of the coherence area among the commodities in all cases was observed. Thus, according to the partial wavelet coherence analysis, global economic policy uncertainty drives the comovement of commodity prices.

The heterogeneity of correlation structure at different time scales presents unique information to commodity traders desired to form a portfolio of agricultural commodities for trading strategies to adopt. A sharp decline in the comovement of any paired commodities suggests that economic policy uncertainties limit the diversification benefits of the agricultural commodities portfolio. It is therefore important that investors pay attention to global risk factors, which tends to create uncertainty in economic policies. The comovement of agricultural commodities, especially upward movement, poses a threat to food insecurity to the poorest population. The evidence of sizable price spikes in times of uncertainty requires the government to implement policy interventions to minimise its impact on the poorest populations which will be enhanced through policy intervention.

Data Availability

The data used in this paper were collected from the IMF Primary Commodity Price System (website: https://data.imf.org/?sk=471DDDF8-D8A7-499A-81BA-5B332C01F8B9). Since this study used secondary data from the official source of IMF, the authors, therefore, decided not to reshare.

Disclosure

This study was carried out as a part of employment as lectures at the University of Cape Coast and Ghana Institute of Management and Public Administration.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] B. J. Henderson, N. D. Pearson, and L. Wang, "New evidence on the financialization of commodity markets," *The Review of Financial Studies*, vol. 28, no. 5, pp. 1287–1311, 2015.

[2] K. J. Singleton, "Investor flows and the 2008 boom/bust in oil prices," *Management Science*, vol. 60, no. 2, pp. 300–318, 2014.

[3] T. Wheeler and J. Von Braun, "Climate change impacts on global food security," *Science*, vol. 341, no. 6145, pp. 508–513, 2013.

[4] G. Tadesse, B. Algieri, M. Kalkuhl, and J. von Braun, "Drivers and triggers of international food price spikes and volatility," *Food Policy*, vol. 47, pp. 117–128, 2014.

[5] J. D. Hamilton and J. C. Wu, "Effects of index-fund investing on commodity futures prices," *International Economic Review*, vol. 56, no. 1, pp. 187–205, 2015.

[6] L. Wang, W. Duan, D. Qu, and S. Wang, "What matters for global food price volatility?" *Empirical Economics*, vol. 54, no. 4, pp. 1549–1572, 2017.

[7] D. Živkov, J. Njegić, and M. Pećanac, "Multiscale interdependence between the major agricultural commodities," *Agricultural Economics Czech*, vol. 65, pp. 82–92, 2019.

[8] M. A. Hernandez, R. Ibarra, and D. R. Trupkin, "How far do shocks move across borders? Examining volatility transmission in major agricultural futures markets," *European Review of Agricultural Economics*, vol. 41, no. 2, pp. 301–325, 2014.

[9] H. Bessembinder, "Systematic risk, hedging pressure, and risk premiums in futures markets," *Review of Financial Studies*, vol. 5, no. 4, pp. 637–667, 1992.
[10] H. Bessembinder and K. Chan, "Time-varying risk premia and forecastable returns in futures markets," *Journal of Financial Economics*, vol. 32, no. 2, pp. 169–193, 1992.
[11] G. Gorton and K. G. Rouwenhorst, "Facts and fantasies about commodity futures," *Financial Analysts Journal*, vol. 62, no. 2, pp. 47–68, 2006.
[12] K. Tang and W. Xiong, "Index investment and the financialization of commodities," *Financial Analysts Journal*, vol. 68, no. 6, pp. 54–74, 2012.
[13] Z. Adams and T. Glück, "Financialization in commodity markets: a passing trend or the new normal?" *Journal of Banking & Finance*, vol. 60, pp. 93–111, 2015.
[14] D. Matošková, "Volatility of agrarian markets aimed at the price development," *Agricultural Economics – Czech*, vol. 57, pp. 34–40, 2011.
[15] S. H. Irwin and D. R. Sanders, "Financialization and structural change in commodity futures markets," *Journal of Agricultural and Applied Economics*, vol. 44, no. 3, pp. 371–396, 2012.
[16] L. Baldi, M. Peri, and D. Vandone, "Stock markets' bubbles burst and volatility spillovers in agricultural commodity markets," *Research in International Business and Finance*, vol. 38, pp. 277–285, 2016.
[17] E. M. Amrouk, S.-C. Grosche, and T. Heckelei, "Interdependence between cash crop and staple food international prices across periods of varying financial market stress," *Applied Economics*, vol. 52, no. 4, pp. 345–360, 2020.
[18] C. Wei Su, X.-Q. Wang, R. Tao, and L. Oana-Ramona, "Do oil prices drive agricultural commodity prices? Further evidence in a global bio-energy context," *Energy*, vol. 172, pp. 691–701, 2019.
[19] X. Yuan, J. Tang, W.-K. Wong, and S. Sriboonchitta, "Modeling Co-movement among different agricultural commodity markets: a copula-GARCH approach," *Sustainability*, vol. 12, no. 1, p. 393, 2020.
[20] D. Headey and S. Fan, "Anatomy of a crisis: the causes and consequences of surging food prices," *Agricultural Economics*, vol. 39, pp. 375–391, 2008.
[21] J. M. Keynes, *A Treatise on Money*, Macmillan, New York, NY, USA, 1930.
[22] H. Working, "The theory of the price of storage," *American Economic Review*, vol. 39, pp. 1254–1262, 1949.
[23] L. Pástor and P. Veronesi, "Uncertainty about government policy and stock prices," *The Journal of Finance*, vol. 67, no. 4, pp. 1219–1264, 2012.
[24] J. A. Frankel, "The effect of monetary policy on real commodity prices," in *Asset Prices and Monetary Policy*, J. Campbell, Ed., pp. 291–333, 2008.
[25] H. Gulen and M. Ion, "Policy uncertainty and corporate investment," *The Review of Financial Studies*, vol. 29, no. 3, pp. 523–564, 2016.
[26] D. Rodrik, "Policy uncertainty and private investment," *Journal of Development Economics*, vol. 36, no. 2, pp. 229–242, 1991.
[27] J. A. Frankel, *International Coordination (No. W21878)*, National Bureau of Economic Research, Cambridge, Massachusetts, USA, 2016.
[28] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring economic policy uncertainty*," *The Quarterly Journal of Economics*, vol. 131, no. 4, pp. 1593–1636, 2016.
[29] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring economic policy uncertainty," *National Bureau of Economic Research*, vol. 131, 2015.
[30] Y. Wang, B. Zhang, X. Diao, and C. Wu, "Commodity price changes and the predictability of economic policy uncertainty," *Economics Letters*, vol. 127, pp. 39–42, 2015.
[31] J. C. Reboredo and X. Wen, "Are China's new energy stock prices driven by new energy policies?" *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 624–636, 2015.
[32] S. J. H. Shahzad, N. Raza, M. Balcilar, S. Ali, and M. Shahbaz, "Can economic policy uncertainty and investors sentiment predict commodities returns and volatility?" *Resources Policy*, vol. 53, pp. 208–218, 2017.
[33] T. P. Wisniewski and B. J. Lambe, "Does economic policy uncertainty drive CDS spreads?" *International Review of Financial Analysis*, vol. 42, pp. 447–458, 2015.
[34] M. Arouri, C. Estay, C. Rault, and D. Roubaud, "Economic policy uncertainty and stock markets: long-run evidence from the US," *Finance Research Letters*, vol. 18, pp. 136–141, 2016.
[35] Z. Liu, Y. Ye, F. Ma, and J. Liu, "Can economic policy uncertainty help to forecast the volatility: a multifractal perspective," *Physica A: Statistical Mechanics and Its Applications*, vol. 482, pp. 181–188, 2017.
[36] C. T. Albulescu, R. Demirer, I. D. Raheem, and A. K. Tiwari, "Does the U.S. economic policy uncertainty connect financial markets? Evidence from oil and commodity currencies," *Energy Economics*, vol. 83, pp. 375–388, 2019.
[37] A. M. Adam, "Susceptibility of stock market returns to international economic policy: evidence from effective transfer entropy of Africa with the implication for open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6, no. 3, p. 71, 2020.
[38] L. Fang, H. Yu, and L. Li, "The effect of economic policy uncertainty on the long-term correlation between U.S. stock and bond markets," *Economic Modelling*, vol. 66, pp. 139–145, 2017.
[39] H. Yu, L. Fang, B. Sun, and D. Du, "Risk contribution of the Chinese stock market to developed markets in the post-crisis period," *Emerging Markets Review*, vol. 34, pp. 87–97, 2018.
[40] I. Badshah, R. Demirer, and M. T. Suleman, "The effect of economic policy uncertainty on stock-commodity correlations and its implications on optimal hedging," *Energy Economics*, vol. 84, Article ID 104553, 2019.
[41] P. Owusu Junior, A. M. Adam, and G. Tweneboah, "Co-movement of real exchange rates in the west african monetary zone," *Cogent Economics & Finance*, vol. 5, no. 1, Article ID 1351807, 2017.
[42] P. Owusu junior, G. Tweneboah, and A. M. Adam, "Interdependence of major exchange rates in Ghana: a wavelet coherence analysis," *Journal of African Business*, vol. 20, no. 3, 2019.
[43] K. Wu, J. Zhu, M. Xu, and L. Yang, "Can crude oil drive the co-movement in the international stock market? Evidence from partial wavelet coherence analysis," *The North American Journal of Economics and Finance*, vol. 53, Article ID 101194, 2020.
[44] E. Asafo-Adjei, D. Agyapong, S. K. Agyei, S. Frimpong, R. Djimatey, and A. M. Adam, "Economic policy uncertainty and stock returns of Africa: a wavelet coherence analysis," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 8846507, 8 pages, 2020.
[45] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical

time series," *Nonlinear Processes in Geophysics*, vol. 11, no. 6, pp. 561–566, 2004.

[46] A. Rua and L. C. Nunes, "International comovement of stock market returns: a wavelet analysis," *Journal of Empirical Finance*, vol. 16, no. 4, pp. 632–639, 2009.

[47] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, no. 1, pp. 61–78, 1998.

[48] C. Torrence and P. J. Webster, "Interdecadal changes in the ENSO-monsoon system," *Journal of Climate*, vol. 12, no. 8, pp. 2679–2690, 1999.

[49] M. Madaleno and C. Pinho, "International stock market indices comovements: a new look," *International Journal of Finance & Economics*, vol. 17, no. 1, pp. 89–102, 2012.

[50] K. Gurley and A. Kareem, "Discussion," *Engineering Structures*, vol. 21, no. 2, pp. 149–167, 1999.

[51] K. Gurley, T. Kijewski, and A. Kareem, "First- and higher-order correlation detection using wavelet transforms," *Journal of Engineering Mechanics*, vol. 129, no. 2, pp. 188–201, 2003.

[52] H. Mihanović, M. Orlić, and Z. Pasrić, "Diurnal thermocline oscillations driven by tidal flow around an island in the Middle Adriatic," *Journal of Marine Systems*, vol. 78, pp. 157–168, 2009.

[53] E. Pereira, P. JS. Ferreira, M. F. da Silva, J. GV. Miranda, and H. BB. Pereira, "Multiscale network for 20 stock markets using DCCA," *Journal of Physics A*, vol. 529, Article ID 121542, 2019.

[54] F. D. Nicola, P. De Pace, and M. A. Hernandez, "Co-movement of major energy, agricultural, and food commodity price returns: a time-series assessment," *Energy Economics*, vol. 57, pp. 28–41, 2016.

[55] E. Nier, T. S. Sedik, and T. Mondino, "Gross private capital flows to emerging markets: can the global financial cycle be tamed?" *IMF Working Papers International Monetary Fund*, vol. 48, pp. 14–196, 2014.

[56] E. Passari and H. Rey, "Financial flows and the international monetary system," *The Economic Journal*, vol. 125, no. 584, pp. 675–698, 2015.

[57] H. Rey, "International channels of transmission of monetary policy and the mundellian trilemma," *IMF Economic Review*, vol. 64, no. 1, pp. 6–35, 2016.

[58] S. H. Irwin, "Commodity index investment and food prices: does the masters hypothesis explain recent price spikes?" *Agricultural Economics*, vol. 44, pp. 29–41, 2013.

[59] I. Goldstein and L. Yang, "Commodity financialization and information transmission," *Working Paper*, 2018.

Hindawi
Complexity
Volume 2021, Article ID 3941074, 17 pages
https://doi.org/10.1155/2021/3941074

Research Article

Predictive Trajectory-Based Mobile Data Gathering Scheme for Wireless Sensor Networks

Fan Chao,[1,2] **Zhiqin He**,[1] **Renkuan Feng**,[1] **Xiao Wang**,[1] **Xiangping Chen**,[1] **Changqi Li**,[1] **and Ying Yang**[1]

[1]*School of Electrical Engineering Guizhou University, Guiyang 550025, China*
[2]*School of Management Harbin Institute of Technology, Harbin 150000, China*

Correspondence should be addressed to Zhiqin He; 641443416@qq.com

Received 30 March 2020; Revised 15 December 2020; Accepted 24 December 2020; Published 6 January 2021

Academic Editor: Fran ois P r s

Tradition wireless sensor networks (WSNs) transmit data by single or multiple hops. However, some sensor nodes (SNs) close to a static base station forward data more frequently than others, which results in the problem of energy holes and makes networks fragile. One promising solution is to use a mobile node as a mobile sink (MS), which is especially useful in energy-constrained networks. In these applications, the tour planning of MS is a key to guarantee the network performance. In this paper, a novel strategy is proposed to reduce the latency of mobile data gathering in a WSN while the routing strategies and tour planning of MS are jointly optimized. First, the issue of network coverage is discussed before the appropriate number of clusters being calculated. A dynamic clustering scheme is then developed where a virtual cluster center is defined as the MS sojourn for data collection. Afterwards, a tour planning of MS based on prediction is proposed subject to minimizing the traveling distance to collect data. The proposed method is simulated in a MATLAB platform to show the overall performance of the developed system. Furthermore, the physical tests on a test rig are also carried out where a small WSN based on an unmanned aerial vehicle (UAV) is developed in our laboratory. The test results validate the feasibility and effectiveness of the method proposed.

1. Introduction

In recent years, the prosperous development in the Internet of Things has been validated where wireless sensor networks (WSNs) become ubiquitous. WSNs are widely used in the fields, such as intelligent transportation, agriculture, medical treatment, aerospace exploration, and other emerging applications, whereas they can contain a large number of static and mobile nodes in a self-organizing way [1–3]. Traditionally, a sink node in a WSN is fixed where the other sensor nodes (SNs) transmit data to the sink through either single-hop or multihop communication. Thus, the SNs close to the sink tend to carry more data transmission or forwarding tasks which leads to unbalanced energy consumption and results in poor data delivery in networks. The issues of energy holes and data collision may occur in such applications [4–7].

Mobile data gathering is regarded as a promising solution to tackle the problems aforementioned. In applications, mobile sinks (MSs) as data collectors gather sensing data in an efficient manner [8–15]. The issues associated with the tour planning of MS have been actively discussed in the existing literature that is crucially important to determine the performance of networks. The concern of energy consumption due to long-distance transmission is discussed in [16]. A maximum cache mechanism is proposed to enhance the transmission capability by adopting MS. In [17], an architecture of single-hop with single MS for mobile data

gathering has been investigated. The proposed scheme has the benefit of energy-saving in line with the optimal tour trajectory of MS. In the study, a heuristic optimization algorithm is employed. A joint scheme for both charging and tour planning in a MS-based WSN is proposed in [18]. Energy consumption of each sensor is leveraged which in turn mitigates energy exhaustion. To strike the balance between energy consumption and latency of data gathering, an algorithm to find the optimal number and position of cluster head nodes for data collection is proposed in [19]. The simulation results verify the effectiveness of the scheme in favor of adopting the optimal path length of MS. In order to create a robust WSN, clustering techniques are also important for energy-saving and delay reduction where SNs are partitioned into subnetworks. Taking into account the kinematic constraints of mobile nodes of similar vehicles, in [20], a mobile data acquisition algorithm based on clustering Dubins smooth curve is proposed. Aiming at the problem of sensor node data aggregation and node energy imbalance, as well as the space problem being often ignored, a heuristic search algorithm (HLSA) is proposed in [21]. A dynamic clustering algorithm to divide the SNs into clusters is employed in [22]. By using the clustering mechanism, the energy efficiency and packet reception rate are improved. Our previous work has also shown the importance of sensor clustering in WSNs [23, 24]. A WSN with respect to mobile data gathering where the work has a special focus on clustering mechanism before mobile data gathering is committed by the MS is developed in [23]. A data gathering scheme by using multiple mobile sink nodes is proposed in [24]. The average path length is dramatically shortened in this way, thereby reducing energy expenditure.

As an extension of the recent work, this study includes the main contributions as follows:

(1) We investigate the data gathering mechanism by considering energy constraints in a WSN system where the path length is minimized.

(2) The network coverage is discussed and formulated to an optimization problem. As a result, suitable node numbers are defined to strike the balance of the WSN scale and network coverage.

(3) A predicted trajectory is developed to assist the tour planning of a MS. In this manner, the overall data gathering is divided into several stages. This mechanism gives a chance to the MS to be recharged at the end of each stage to prevent energy exhaustion for a MS with insufficient power while heavy data load being required.

(4) Besides the extensive simulation on MATLAB, the trail tests are committed to verify the effectiveness of the proposed scheme where an unmanned aerial vehicle (UAV) serving as a MS is employed in a WSN.

The rest of this paper is organized as follows. Section 2 presents the systematic configuration with extensive discussion on energy constraints associated with MS traveling, activities, and the network architecture. Section 3 develops the optimization formulation for maximising network coverage while minimizing energy expenditure and path length for each cycle. Simulation setup and trail tests are both presented in Section 4 following the outcome demonstration and analysis corresponding to them. The key conclusions are drawn in Section 5.

2. Problem Formulation

The sensor network consists of a base station, a certain number of SNs, and a MS. After the sensors acquire data from the sensing field, they wait to be polled by the MS to deliver the information in each round. The WSN is shown in Figure 1.

It is assumed that there are N SNs in the network, $N = \{1,.....,n\}$, $i \neq j, i, \quad j \in N$; V represents a virtual cluster, $V = \{1, \ldots, v\}$; and trajectory between the virtual cluster head nodes is $M = \{V\ (i, j)\ i \neq j\}$. Let d_{ij} represent the Euclidean distance between the two virtual cluster head and ley x_{ij}^m indicate whether $V\ (i, j)$ is included in the tour of the MS. We have the following equation:

$$x_{ij} \begin{cases} 1, & \text{if data gathering tour contains } V\ (i, j), \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A MS gathers data periodically. The overall time in each cycle contains two segments: traveling and sojourn time [25]. Traveling time also includes two parts. One is in line with traversing time from one cluster to another. The second part is relevant to returning to the base station. Moreover, since the MS has the duty to poll the cluster head nodes, it stays on the cluster head nodes until completing data collection.

2.1. Traveling Time. In one cycle, the tour length can be calculated by

$$D = \sum_{i=1}^{V} \sum_{j=1}^{V} x_{ij} d_{ij}. \tag{2}$$

The total moving time is then calculated by

$$t_m = \frac{D}{v}, \tag{3}$$

where v is the moving speed.

2.2. Sojourn Time. In a cycle, the MS reaches sensor node i; the amount of data is C_i; τ_i is staying time at each rendezvous point; G represents the transmission rate. Therefore, there is a relationship as follows:

$$\tau_i = \frac{C_i}{G}. \tag{4}$$

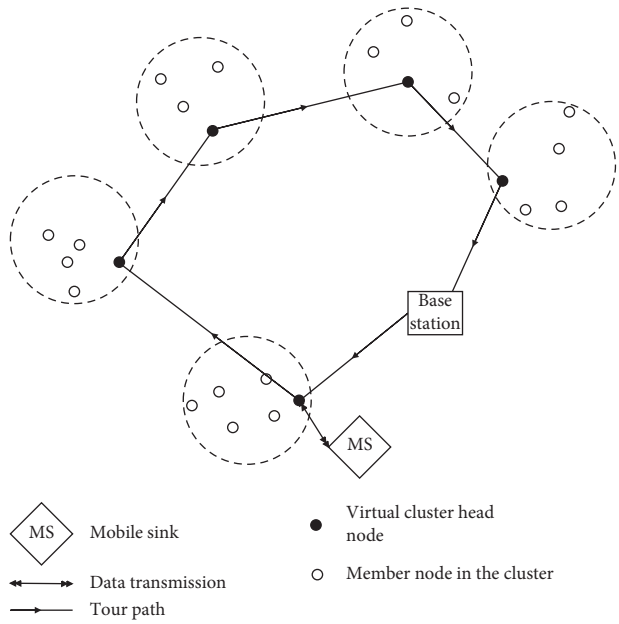


Figure 1: A wireless sensor network.

According to equations (3) and (4), the period of a cycle is obtained by

$$T = t_m + \sum_{i=1}^{v} \tau_i. \tag{5}$$

The tour planning is formulated to an optimization problem with the following equations:

$$\text{Minimize } t_m + \sum_{i=1}^{v} \tau_i, \tag{6}$$

$$\text{subject to} \sum_{i=1}^{V} x_{ij} = 1\,, \quad j \in V,\ j \neq 1, \tag{7}$$

$$\sum_{j=1}^{V} x_{ij} = 1, \quad i \in V, i \neq 1, \tag{8}$$

$$\sum_{i=1}^{V} x_{ip} - \sum_{j=1}^{V} x_{ip} = 0, \quad p \in v,\ p \neq 1, \tag{9}$$

$$\sum_{j=2}^{V} x_{1j} = 1, \tag{10}$$

$$\sum_{i=2} x_{i1} = 1, \tag{11}$$

$$T_{MN} \leq T_D, \tag{12}$$

$$L_{\text{sum}} \leq L_{\max}. \tag{13}$$

Equations (7) and (8) ensure that each cluster head node is accessed only once per round. Equation (9) guarantees tour continuity. Equation (10) guarantees that the MS starts from the base station. Equation (11) guarantees that the final destination of MS is the base station. Equation (12) is the time limit per round. Equation (13) is the limit of tour length per round, which in turn determines the energy limit.

Since the MS consumes the most energy on traversing through the sensing field, the energy cost is closely related to the tour range. This paper mainly takes energy expenditure on traveling into account. Therefore, the objective function in optimization is replaced by the target to minimize the tour length as follows:

$$\text{Minimize} \sum_{i=1}^{V} \sum_{j=1}^{V} x_{ij} d_{ij}. \tag{14}$$

3. Tour Planning

3.1. Coverage Issue with an Adjacency Matrix. The coverage issue is essential for a WSN. On the one hand, visiting fewer nodes will reduce the MS burden for data collection. On the other hand, a WSN needs enough nodes to achieve the expected monitoring quality. Thus, there is a balance to strike between the node number and the communication quality. In this paper, the network coverage with nodes deployment in random order is concerned. For simplicity purpose, the optimal density of the

deployment is determined by the communication radius, which is the basis for calculating the number of deployment nodes.

This paper uses an adjacency matrix graph to define the connectivity of deployed nodes [22–26]. We consider the network as a graph $G = (V\,E)$ where V refers to the SNs and E represents the edge between the two SNs which can communicate with each other.

In some conditions, we can convert G into a square matrix $A = (a_{ij})_{m\times m}$ with M-order $(m = |V|)$ if $a_{ij} = |e_k|, e_k = <v_i, v_j> \in E$. Therefore, the adjacency matrix G can be represented by matrix A.

Furthermore, we define $A^k = (a_{ij}^{(k)})_{m\times m}$, $(a_{ij}^{(k)}) = \sum_{k=1}^{m} a_{ih}^{k-1} a_{hj}$. If $a_{ih} \bullet a_{hj} \neq 0$, $a_{ih} \neq 0$, and $a_{hj} \neq 0$, there are k routes from the node v_i via v_h to v_j. For instance, $a_{ij}^{(2)}$ demonstrates there are two paths that can start from v_i via an intermediate node v_h to the node v_j. Therefore, if $\sum_{k=1}^{m-1} a_{ij}^{(k)} = 0$, there is no path communication between the node v_i and the node v_j.

Therefore, we use a new matrix $S = (S_{ij})_{m\times m} = \sum_{k=1}^{m-1} A^k$. If all the elements in matrix S are nonzero elements, then G is a fully connected graph. Otherwise, if there is only an element in matrix S, then G is a disconnected graph.

3.2. Cluster Forming. We uniformly deploy the nodes in the region where they are grouped into clusters. In this way, the SNs can have balanced energy consumption since they can share the workload in clusters s [27]. Moreover, it ensures that each cluster head has almost the same energy expenditure in each cluster. There are many useful and advanced algorithms to realize the deployment, for example, neighbor clustering and fuzzy clustering. In order to implement real-world tests, we want a simpler and efficient algorithm as our solution. Therefore, the K-means algorithm is adopted. We use K-means based dynamic clustering algorithm to partition the nodes into the monitoring area [24].

The K-means algorithm aims to minimize the total distance J_j between the SNs and the center of the cluster. The distance can be presented by

$$J_j = \sum_{i=1}^{N_j} \|X_i - Z_j\|^2, \quad X_i \in S_j, \tag{15}$$

where S_j is cluster j; Z_j is the center of the cluster j; and N_j is the sample number in the cluster j. Since the selection of the cluster center should make J_j extremely small [1], then $\partial J_j/\partial Z_j = 0$. Thus, equation (14) can be rewritten by

$$\frac{\partial}{\partial_{Zj}} \sum_{i=1}^{N} \|X_i - Z_j\|^2 = 0. \tag{16}$$

Therefore,

$$Z_j = \frac{1}{N_j} \sum_{i=1}^{N_j} X_i. \tag{17}$$

The algorithm is committed in the following way:

Step 1: initializing the clusters with randomly picked nodes. A cluster head is selected to be as the center. All centers are recorded by $Z_1(1), Z_2(1), \ldots\ldots, Z_k(1)$.

Step 2: allocating the rest of the SNs in cluster k around the centers according to equation (17) where Euclidean distance is applied.

Step 3: recalculating $Z_j(k+1)$, $j = 1, 2, \ldots, k$ for each cluster.

Step 4: judging if $Z_j(k+1) = Z_j(k)$, the clustering ends. Otherwise, return to Step 2 and regroup the sample iteratively.

Afterwards, the calculation completes while the results are recorded. Now, the nodes are partitioned to k clusters where k virtual cluster centers are generated at the same time. Via k-means algorithm, the overall distance from the SNs to the cluster centers reaches the least. Since the intercluster communication uses single-hop transmission, the communication costs the least energy consumption through this arrangement. The nodes are then manually placed to the locations of these virtual cluster centers where the cluster head nodes play the roles of the center. After a head node being defined, a broadcast message is sent to the members in the cluster to confirm the location of the cluster head node. Now, the location of these virtual cluster centers will be a rendezvous point for the MS. The MS starts to gather data by the members in the cluster after the MS reaches the location.

3.3. Energy Consumption Prediction. Due to the limited energy capacity of the MS, it is difficult for a single MS to complete all the data gathering without interruption in one cycle. Therefore, the whole tour planning is divided into several segments where the division method is defined by the prediction result based on the ant colony system (ACS) optimization. The algorithm follows the steps below:

Step 1: to determine the accessible order of each cluster where the access nodes are defined by $\{1, 2, \ldots, v\}$. These orders are determined by the ant colony algorithm [8–27]. The calculation is based on equations (18)–(20).

Step 2: judgement. If the length of the whole tour is within the maximum range of MS, the MS completes the data gathering at once. Otherwise, the tour is cut into multiple segments and executed in turn.

Assuming j represents the next hop point from the current i point, the possibilities of visiting point j are s that can be presented by the following equations [27]:

$$j = \begin{cases} \arg\max\left\{\left[\tau_{ij}(t)\right]\left[\eta_{ij}(t)\right]^{\beta}\right\}, & q < q_0, \\ S, \end{cases} \tag{18}$$

$$S = \begin{cases} \dfrac{\left[\tau_{ij}(t)\right]\left[\eta_{ij}(t)\right]^{\beta}}{\sum_{k\in\text{allowed}_k}\left[\tau_{ij}(t)\right]\left[\eta_{ij}(t)\right]^{\beta}}, & j \in \text{allowed}\, k \\ 0, & \text{else} \end{cases} \tag{19}$$

where $\tau_{ij}(t)$ represents pheromone variables; $\eta_{ij}(t)$ is the heuristic desirability; $\beta > 0$ represents the relative strength of heuristic factors; $q_0 \in (0, 1)$ is the initial parameters; and $q \in (0, 1)$ is a random number. Local update strategy is to make the selected edges have a better influence on the later ants, as shown in the following formula:

$$\tau_{ij}(t+1) = (1-\zeta)\tau_{ij}(t) + \zeta \cdot \tau_0, \tag{20}$$

where ζ is the pheromone decay parameter within (0, 1) and τ_0 is a constant. The global update strategy is to find the shortest path for traversing the rest of the points and return to the base station within the MS energy capability that belongs to the edges on the optimal path, as shown in the following equation:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho \cdot \Delta\tau_{ij}(t), \tag{21}$$

$$\Delta\tau_{ij} = \begin{cases} \dfrac{1}{L_k}, & (i, j) \text{ on the optimal path}, \\ 0, & \text{else}, \end{cases} \tag{22}$$

where $\Delta\tau_{ij}$ is the added pheromone on the trail from point i to j; ρ represents the evaporation coefficient, $\rho \in (0, 1)$; and L_k is the optimal path length.

Step 3: the total energy carried by the MS is E. At the beginning of the cycle, the MS arrives at the first sensor node for data collection. Afterwards, the MS measures its own remaining energy for the next stage. If the remaining energy is enough to be consumed in the next stage, the MS travels to the next node for data gathering. Otherwise, it returns directly to the base station for recharging. The energy consumption in the next stage is expected to be sufficient for visiting the remaining nodes, data collecting, and returning to the base station.

(1) Energy consumption for visiting:

It refers to the energy consumed over the traveling from node i to node j:

$$e_{ij} = e_m \bullet d_{ij}. \tag{23}$$

(2) Energy consumption for collecting data from node j:

$$e_{\text{rej}} = re_{rx} \bullet \left(t_j - t_0\right). \tag{24}$$

t_j is the time to arrive at node j, and t_0 is the starting time of the cycle.

(3) Energy consumption for returning to the base station:

$$e_{js} = e_m \bullet d_{js}. \tag{25}$$

The total energy consumption for the next stage is predicted at node i expressed by the following equation:

$$N_i = e_{ij} + e_{\text{rej}} + e_{js}. \tag{26}$$

Assuming the remaining energy when the MS reaches the node i is equal to q, the remaining energy of the MS after the completion of the data gathering is R_i. So,

$$R_i = Q - e_{\text{rei}} - e_{\text{chi}}, \tag{27}$$

If $R_i > N_i$ when MS travels to node j for data collection, MS goes on its journey to the next node. Otherwise, the MS returns to the base station for recharging.

Step 4: tour replanning.

After the MS completes the recharge, the tour planning to the remaining nodes in the next subcycle will be recalculated. The new trajectory and visiting order are calculated. Afterwards, the data gathering continues according to the updating results from Step 2.

Step 5: repeat the above process until all the cluster fields are traversed. The remaining nodes not polled are abandoned. In order to simplify the algorithm, the maximum tour length is used to represent the energy limit of the MS.

4. Results and Analysis

4.1. Simulation Environment. To verify the effectiveness of the algorithm, the simulation is carried out on a MATLAB platform. The key simulation parameters are summarized in Table 1.

4.2. Outcomes and Analysis. First, the problem of node coverage is simulated. Assuming the node number is 90 and the communication radius is 15 m, the coverage area is shown in Figure 2. After a large number of experiments being committed, the relationship between node number, communication radius, and coverage are obtained. The results are shown in Figure 3. The relationship between node density and coverage is shown in Figure 4. By synthetically analyzing these outcomes, we define the

Table 1: Parameters of the simulation.

Parameter name	Parameter values
Number of sensor nodes	70
Number of virtual cluster head nodes	10
Number of MS	1
Heuristic factor (α)	1
Expectation heuristic factor (β)	5
Information intensity (Q)	500
Pheromone volatile factor (η)	0.5
Number of ant colonies (m)	18
Required coverage (%)	0.95
Maximum endurance mileage (m)	200

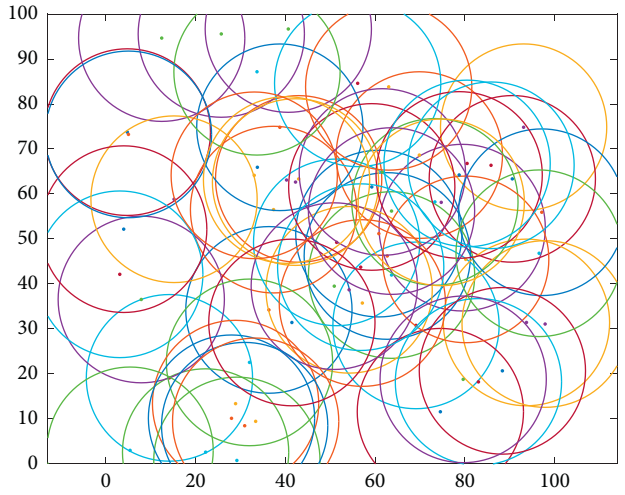


Figure 2: Coverage area in the field.

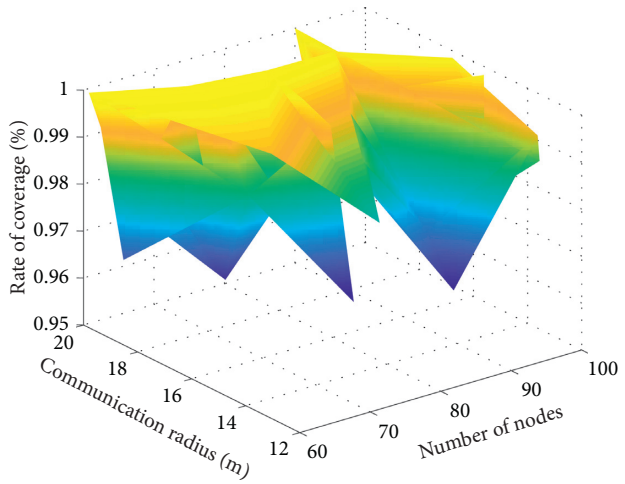


Figure 3: Relationship between the number of nodes, communication radius, and coverage.

communication radius and the node density as 15 and 0.007, respectively.

According to the coverage plan, 70 nodes are randomly placed in the sensor field as shown in Figure 5. After applying the dynamic clustering algorithm, 10 virtual cluster head nodes are generated as shown in Figure 6.

Afterwards, the ant colony optimization is used to plan the shortest path over the ten cluster head nodes, and the node path is shown in Figure 7. However, the shortest length on the tour is 348.63 according to the measurement from Figure 8. It exceeds the energy limit of the MS. The access order can be obtained as shown in Figure 9. Nevertheless, the

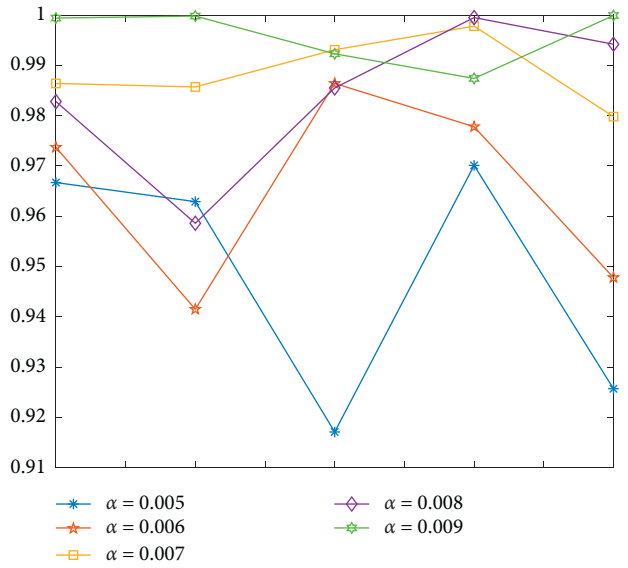


Figure 4: Relationship between node density and coverage ($r = 15$).

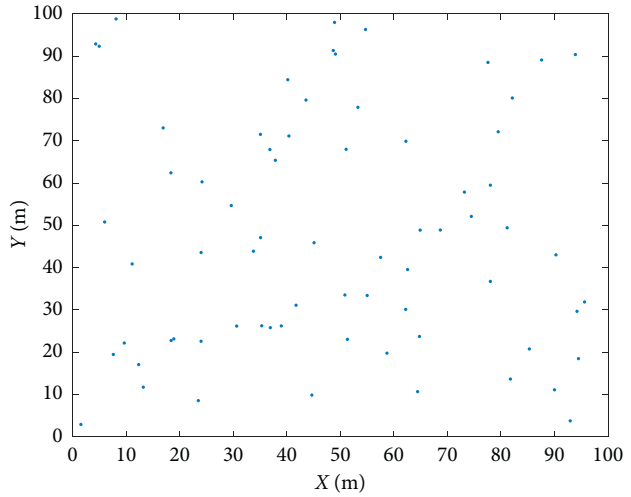


Figure 5: Node distribution.

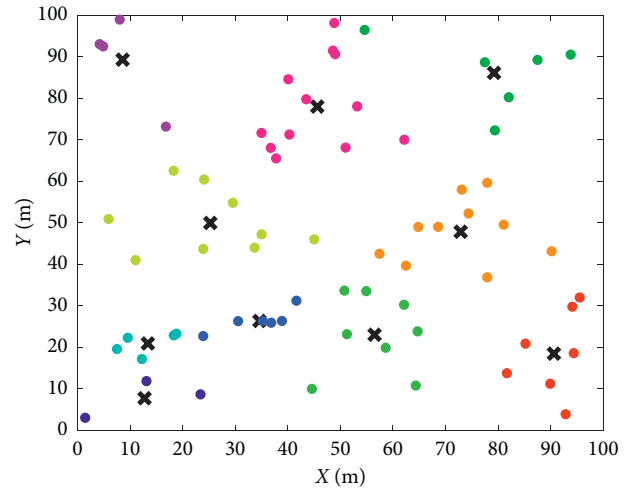


Figure 6: Node clustering and the cluster heads.

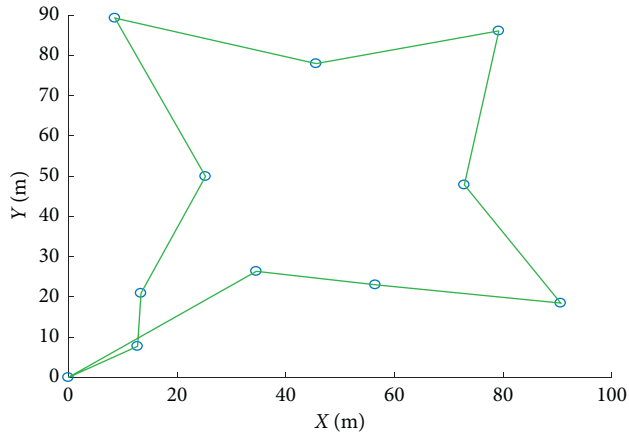


Figure 7: The planned trajectory of a MS at the beginning.

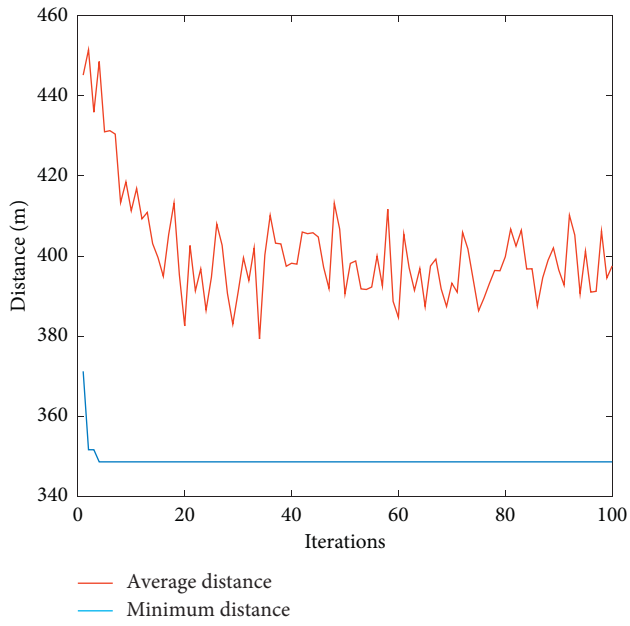


Figure 8: Predicted path distance based on prediction.

node starts the data gathering for the first subcycle from the base station (node 0).

The MS starts the tour. However, when it reaches node 3, it is found that the remaining energy is not enough to support it to complete the next stage of data collection tasks; that is, it reaches node 4 for data collection tasks and returns to the base station. Thus, it returns to the base station for energy replenishment to continue the second subcycle. After the MS is fully charged, the tour planning for traversing the remaining nodes is recalculated as shown in Figure 10. The shortest path is shown in Figure 11. At this time, traversing the remaining nodes costs more energy than the MS capacity. There will be a third subcycle for the MS. The updated access order in the second subcycle can be found in Figure 12.

Again, the MS arrives at the head node of the first virtual cluster in the second subcycle. After node 3 being poll, the MS needs more energy to reach node 4. Therefore, it returns to the base station for preparing the third subcycle.

After being fully charged, the remaining tour is reorganized. The third subcycle is then defined as shown in Figure 13. The searching map for the optimal trajectory is shown in Figure 14. The shortest length over the field cannot be completed by the MS in one cycle. Therefore, the fourth subcycle is needed. The calculation for the third subcycle with the optimal is executed and the result is shown in Figure 15.

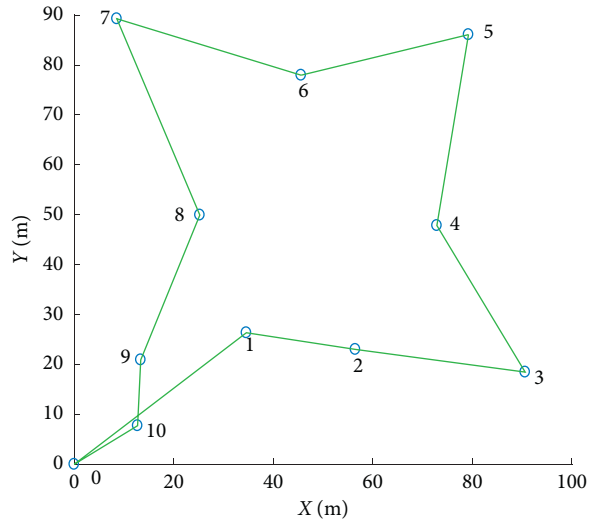


Figure 9: Planned access order for the first round prediction.

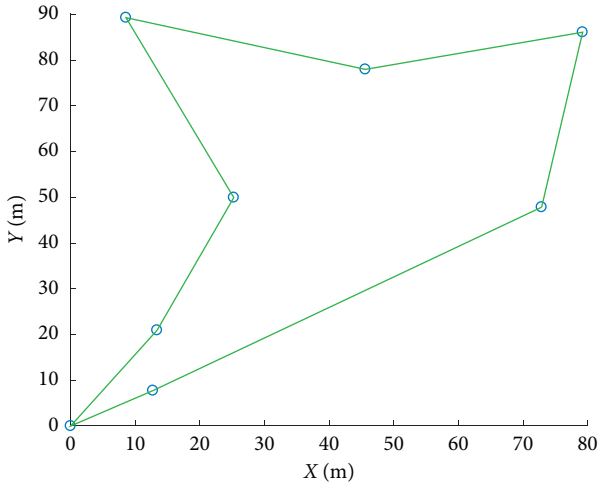


Figure 10: Planned trajectory at the beginning of the second subcycle.

The plan for the traversing, the predicted path length, and the node access order of the fourth subcycle are shown in Figures 16–18

In the fourth subcycle, it is slightly different from the previous subcycles. After polling node 2, the MS cannot reach node 3. It returns to the base station to be recharged. However, the energy cannot support it traveling to node 3. Therefore, node 3 is abandoned. The tour planning, trajectory, and the visiting order for the fourth subcycle are demonstrated in Figures 19 and 20.

The simulation results have shown the initial energy of the MS cannot support completing data collection. Therefore, the initial energy is then increased to 300. By repeating the process above, the tour planning in the final two periods can be obtained as shown in Figure 21. The path lengths of each subcycle are shown in Figure 22. With the updated initial energy, the MS can complete the data gathering by traversing all cluster head nodes over the sensing field. Therefore, the optimal energy capacity of the MS is determined by 300 m.

The simulation is carried out with the optimal tour planning with the shortest length where energy constraints and efficient data gathering are considered as a priority. Through the simulation, we also find the process of data gathering is closely related to the initial energy of the MS.

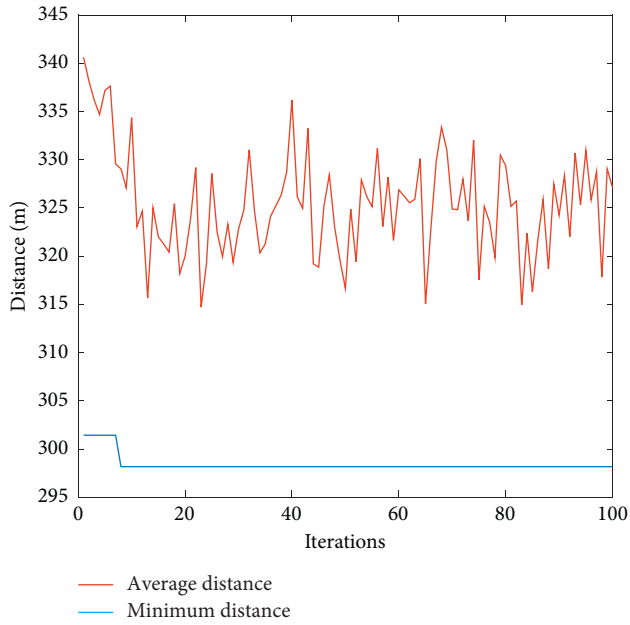


Figure 11: Predicted path length at the beginning of the second subcycle.

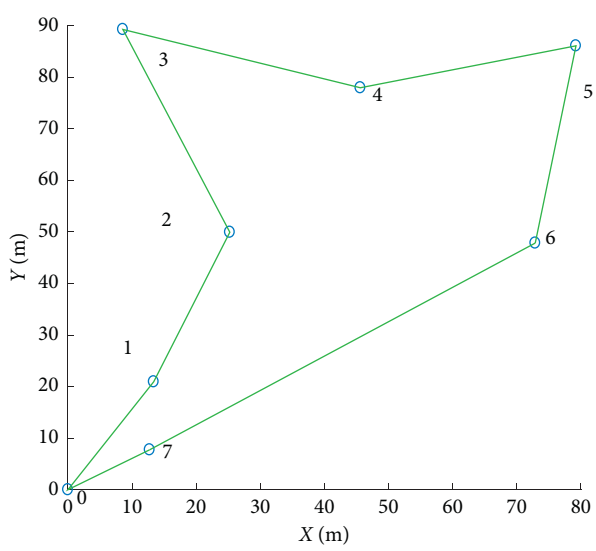


Figure 12: Planned access order in the second subcycle.

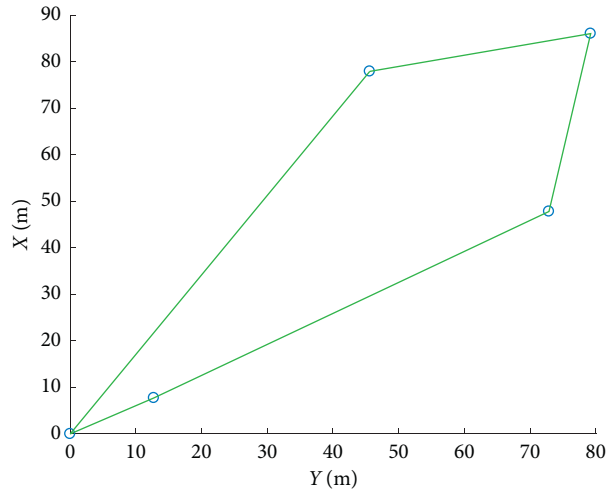


Figure 13: Planned trajectory at the beginning of the third subcycle.

Therefore, a comparative evaluation for the relationship between the tour planning and battery capacity of the MS is carried out. The key indicators are summarized in Table 2.

It is noted that the larger initial energy can lead to a smaller number of subcycles. Moreover, the shorter length means the shorter latency over data gathering.

In practice, the energy capacity of a MS is limited. Therefore, consideration is needed to strike the balance between the delay and the initial energy.

4.3. Trail Tests in a Small WSN. Trail tests are carried out in the laboratory to verify the reliability of data transition based on a small scale WSN while only the first subcycle is tested. Three SNs are placed as the terminal nodes to collect data.

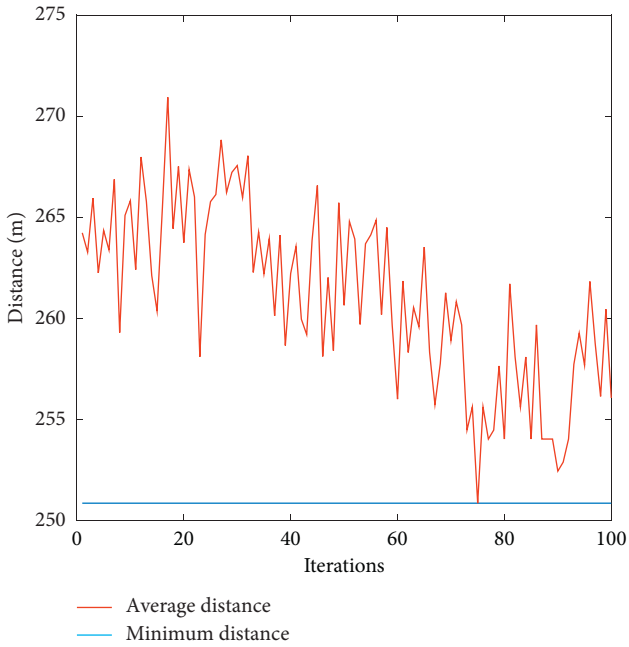


Figure 14: Predicted path length at the beginning of the third subcycle.

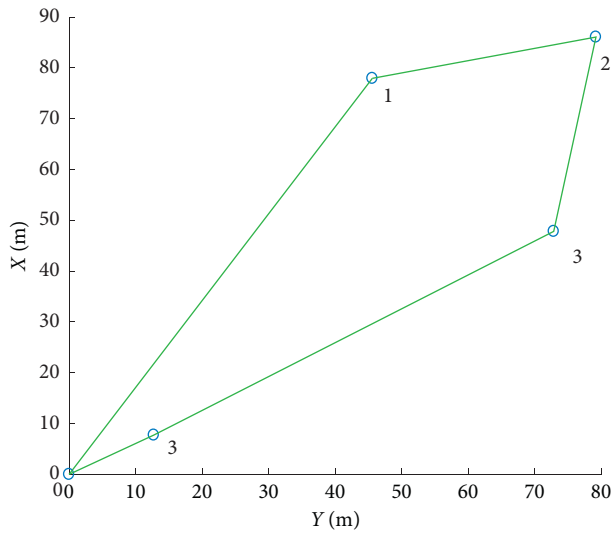


Figure 15: Planned access order at the beginning of the third subcycle.

The MS is responsible for polling the SNs and completes data gathering. The device in the experiment includes a UAV carrying the MS and three DHT11 temperature-humidity sensors as the SNs. The images of these devices are shown in Figure 23. Test results are summarized in Table 3.

Furthermore, three rounds of experiments are committed. The results are summarized in Table 4. The results show that the experimental system can effectively collect temperature and humidity data, and the data transmission is reliable.

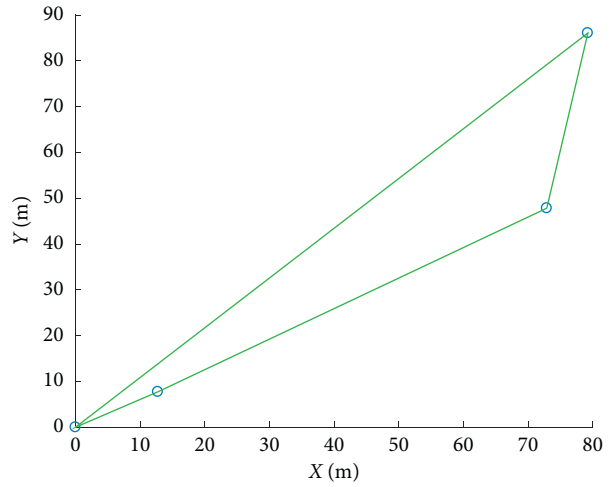


Figure 16: Planned trajectory at the beginning of the fourth subcycle.

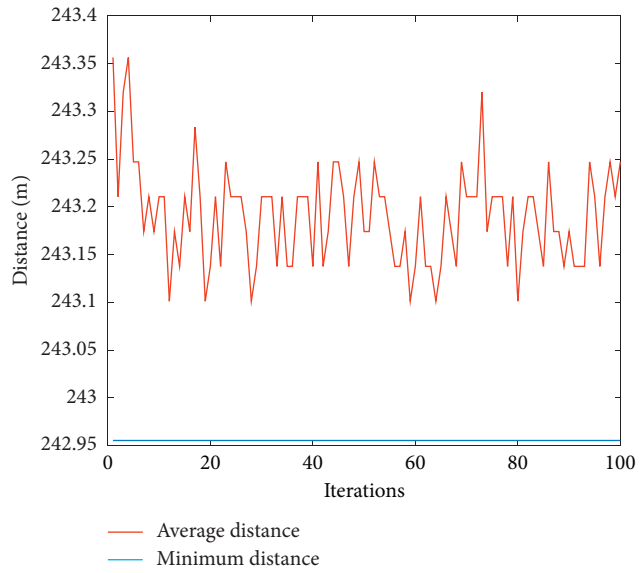


Figure 17: Predicted path length at the beginning of the fourth subcycle.

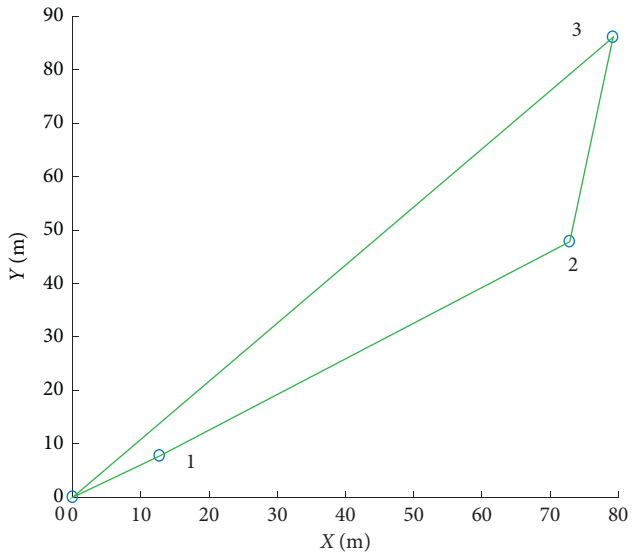


Figure 18: Planned access order at the beginning of the fourth subcycle.

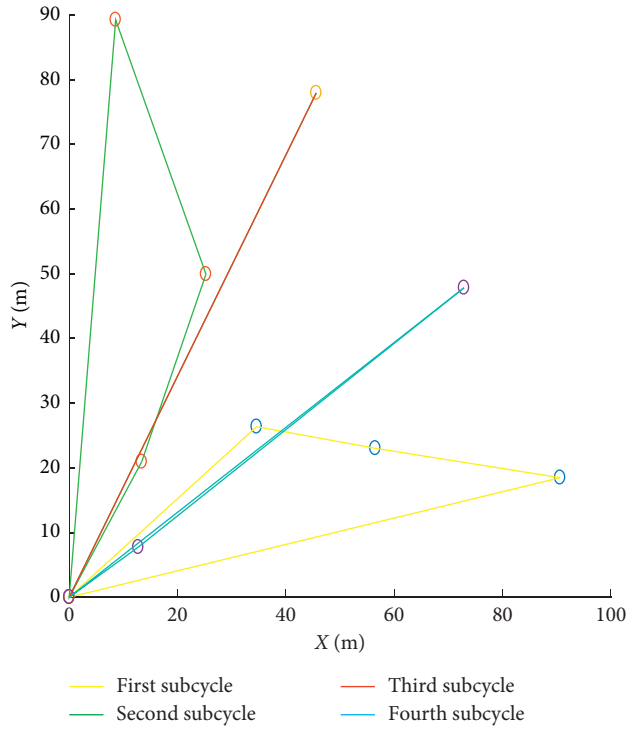
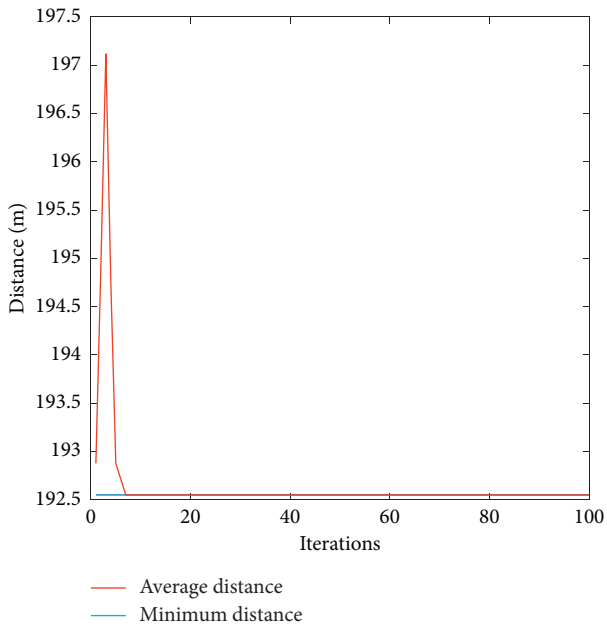
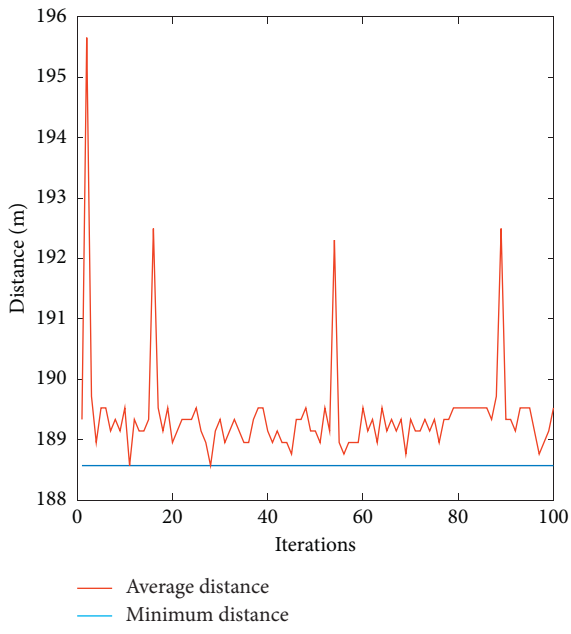


Figure 19: The whole tour planning over the four subcycles.



(a)



(b)

Figure 20: Continued.

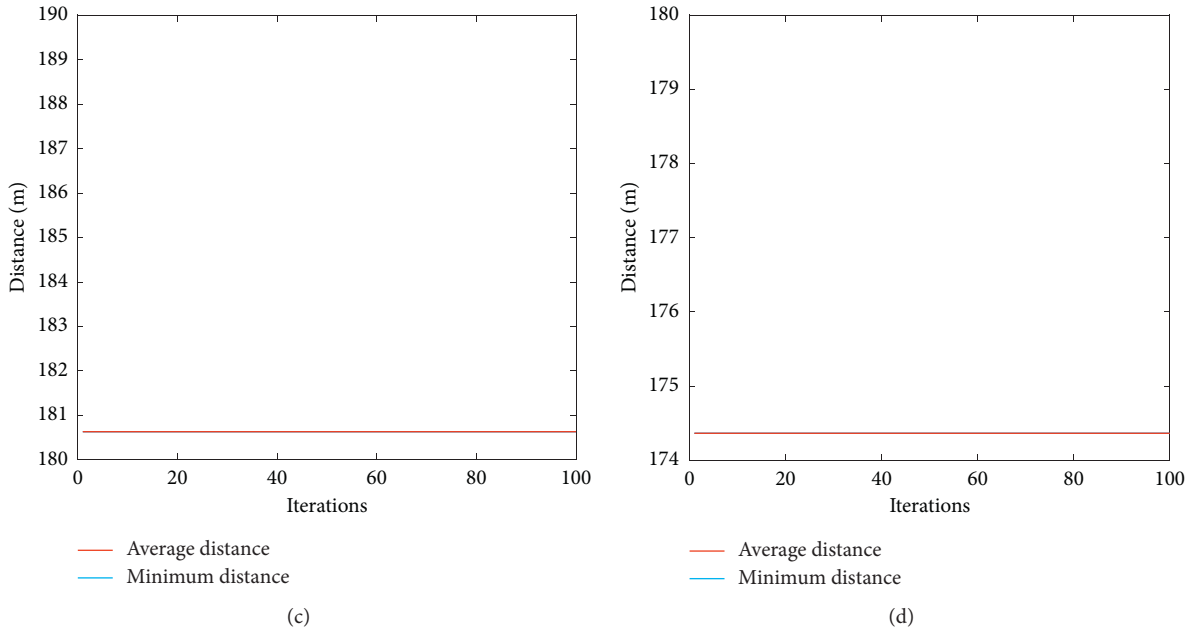


Figure 20: Tour length for each subcycle.

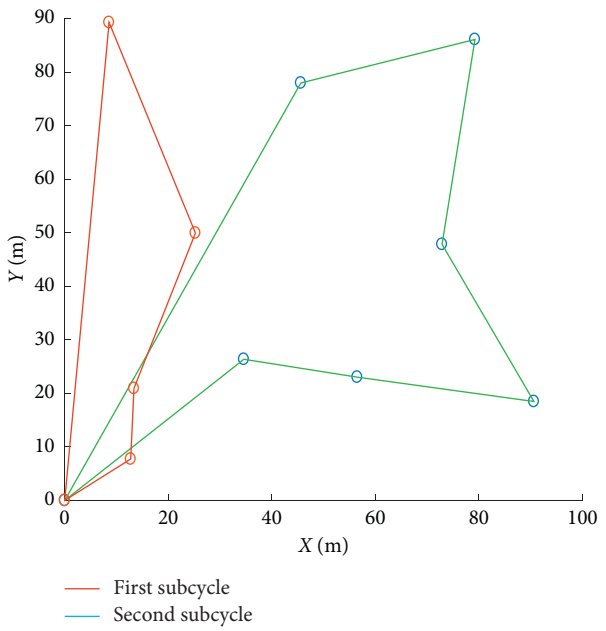


Figure 21: The first two subtours with a limitation of a maximum continuation length of 300 meters.

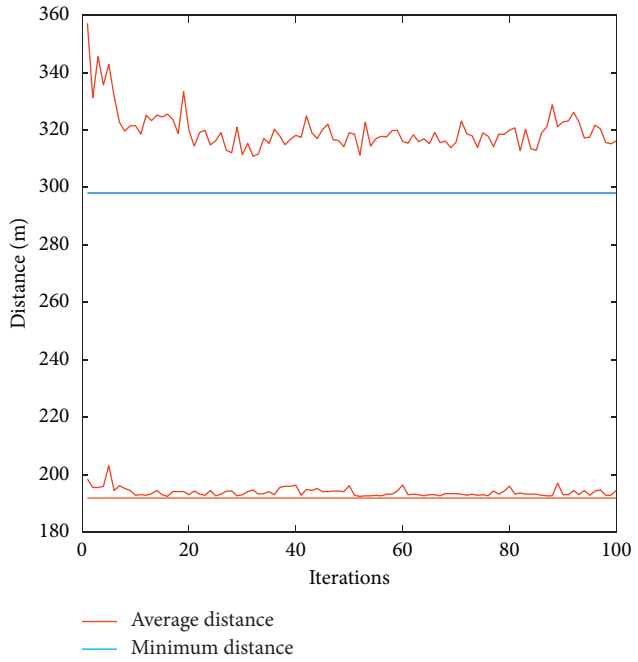


Figure 22: The path length of the first two subtours with a limitation of a maximum continuation length of 300 meters.

Table 2: Tour planning with different battery capacities.

Round	Maximum range (m)	Number of cycles	Data gathering rate (%)	Total length of the route (m)
1	200	4	90	736
2	300	2	100	490
3	400	1	100	350

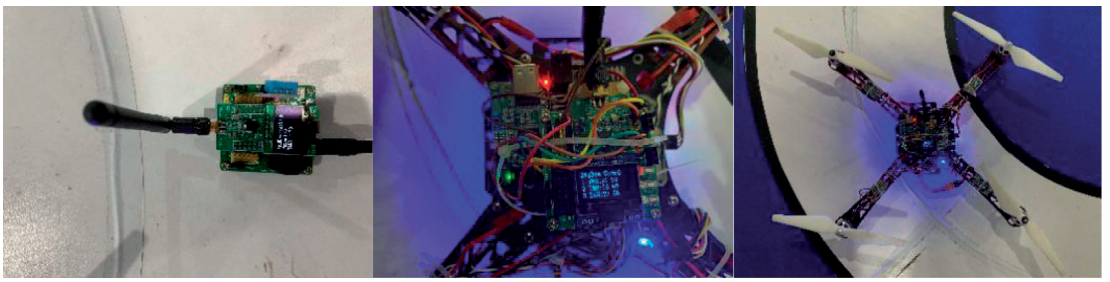


Figure 23: Device for sensing and data gathering.

Table 3: Test results of communication quality.

Distance (m)	Number of packets sent	Number of packets lost	Packet loss rate
10	1000	0	0
15	1000	0	0
20	1000	1	0.001
25	1000	23	0.022
30	1000	101	0.121
35	1000	382	0.394
40	1000	1000	1

Table 4: The experimental values of the system.

Number of experimental rounds	Node 1	Node 2	Node 3
1	21/59	19/69	22/56
2	24/50	24/55	24/51
3	22/53	21/59	24/51

5. Conclusion

This paper presents a prediction-based tour planning for data gathering in WSNs. The scheme primarily aims to mitigate the challenge when a single MS cannot complete data gathering or traverse all cluster head nodes at once. First, the issue of the network coverage is discussed in favor of a decent arrangement with the optimal number of sensor nodes across the field. Afterwards, a dynamic clustering algorithm is proposed to generate sensor clusters where the cluster head nodes are defined as the MS sojourns. The methodology of the predictive trajectory is developed for finding the optimal tour before a MS traverses the sensing field. In the study, the single tour of data gathering is divided into several subcycles to prevent energy exhaustion. Both the MATLAB simulation and physical tests are committed to demonstrating the whole procedure for data gathering under the proposed scheme. The results have proved the effectiveness and reliability of the methodology. The study outcomes provide a useful solution for tour planning in energy-constrained WSNs.

Data Availability

The data used to support the findings of this study are included within the article. If other data or programs used to support the findings of this study are needed, they can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (61861007).

References

[1] X. Li, J. Peng, M. S. Obaidat, F. Wu, M. K. Khan, and C. Chen, "A secure three-factor user authentication protocol with forward secrecy for wireless medical sensor network systems," *IEEE Systems Journal*, vol. 14, no. 1, pp. 39–50, 2020.

[2] H. Malik and W. Zatar, "Agent based routing approach to support structural health monitoring-informed, intelligent transportation system," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 1031–1043, 2020.

[3] D. Achilles, S. Maria, P. Papadopoulou, and K. Goudos, "Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: a comprehensive review," *Internet of Things*, vol. 92, Article ID 100187, 2020.

[4] A. Demertzis, K. Oikonomou, and I. Stavrakakis, "Evaluation of a proposed minimum path impotence routing policy in wireless sensor networks," *Ad Hoc Networks*, vol. 94, Article ID 101928, 2019.

[5] N. Sharmin, A. Karmaker, W. L. Lambert, M. S. Alam, and M. S. A. Shawkat, "Minimizing the energy hole problem in wireless sensor networks: a wedge merging approach," *Sensors*, vol. 20, no. 1, p. 277, 2020.

[6] R. Deng, S. He, P. Cheng, and Y. Sun, "Towards balanced energy charging and transmission collision in wireless rechargeable sensor networks," *Journal of Communications and Networks*, vol. 19, no. 4, pp. 341–350, 2017.

[7] X. Zhu, X. Tao, T. Gu, and J. Lu, "Target-aware, transmission power-adaptive, and collision-free data dissemination in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6911–6925, 2015.

[8] K.-C. Chu, D.-J. Horng, and K.-C. Chang, "Numerical optimization of the energy consumption for wireless sensor networks based on an improved ant colony algorithm," *IEEE Access*, vol. 7, pp. 105562–105571, 2019.

[9] R. Velmani and B. Kaarthick, "An efficient cluster-tree based data collection scheme for large mobile wireless sensor networks," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2377–2390, 2015.

[10] Y. Yuan, W. Liu, T. Wang, Q. Deng, A. Liu, and H. Song, "Compressive sensing-based clustering joint annular routing data gathering scheme for wireless sensor networks," *IEEE Access*, vol. 7, pp. 114639–114658, 2019.

[11] S. Guo, C. Wang, and Y. Yang, "Joint mobile data gathering and energy provisioning in wireless rechargeable sensor networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2836–2852, 2014.

[12] D. Takaishi, H. Nishiyama, N. Kato, and R. Miura, "Toward energy efficient big data gathering in densely distributed sensor networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 388–397, 2014.

[13] L. Nguyen and H. T. Nguyen, "Mobility based network lifetime in wireless sensor networks: a review," *Computer Networks*, vol. 174, Article ID 107236, 2020.

[14] V. K. Chawra and G. P. Gupta, "Load balanced node clustering scheme using improved memetic algorithm based meta-heuristic technique for wireless sensor network," *Procedia Computer Science*, vol. 167, pp. 468–476, 2020.

[15] X. Fu and Y. Yang, "Modeling and analysis of cascading node-link failures in multi-sink wireless sensor networks," *Reliability Engineering & System Safety*, vol. 197, Article ID 106815, 2020.

[16] P. Gu, C. Li, Y. Wang, S. Li, and W. Xi, "WSN energy saving strategy with a mobile data collector," *Radio Communications Technology*, vol. 45, no. 2, pp. 150–155, 2019, (in Chinese).

[17] M. Ma, Y. Yang, and M. Zhao, "Tour planning for mobile data-gathering mechanisms in wireless sensor networks," *IEEE Transactions onVehicular Technology*, vol. 62, no. 4, pp. 1472–1483, 2013.

[18] F. Liu, H. Lu, T. Wang, and Y. Liu, "An energy-balanced joint routing and charging framework in wireless rechargeable sensor networks for mobile multimedia," *IEEE Access*, vol. 7, pp. 177637–177650, 2019.

[19] R. Anwit, A. Tomar, and P. K. Jana, "Scheme for tour planning of mobile sink in wireless sensor networks," *IET Communications*, vol. 14, no. 3, pp. 430–439, 2020.

[20] Z. Meiyan and C. Wenyu, "Dubins curves based mobile data collecting algorithm for clustered wireless sensor networks," *Journal Sensor Technology*, vol. 32, no. 4, pp. 603–609, 2019.

[21] G. Yogarajan and T. Revathi, "Genetic algorithm-based mobile sink data collecting algorithm for wireless sensor networks," *Wireless Network*, vol. 24, no. 8, pp. 1095–1099, 2018.

[22] S. Ganesh and R. Amutha, "Efficient and secure routing protocol for wireless sensor networks through SNR based dynamic clustering mechanisms," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 422–429, 2013.

[23] F. Chao, Z. He, A. Pang, H. Zhou, and J. Ge, "Path optimization of mobile sink node in wireless sensor network water

monitoring system," *Computational Methods for Modeling, Simulating, and Optimizing Complex Systems*, vol. 2019, Article ID 5781620, 10 pages, 2019.

[24] A. Pang, F. Chao, H. Zhou, and J. Zhang, "The method of data collection based on multiple mobile nodes for wireless sensor network," *IEEE Access*, vol. 8, pp. 14704–14713, 2020.

[25] C. Wang, S. Guo, and Y. Yang, "An optimization framework for mobile data collection in energy-harvesting wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 2969–2986, 2016.

[26] X. Li, X. Tao, and G. Mao, "Unbalanced expander based compressive data gathering in clustered wireless sensor networks," *IEEE Access*, vol. 5, pp. 7553–7566, 2017.

[27] P. Zhou, C. Wang, and Y. Yang, "Static and mobile target -coverage in wireless rechargeable sensor networks," *IEEE Transactions on Mobile Computing*, vol. 18, no. 10, pp. 2430–2445, 2019.

Hindawi
Complexity
Volume 2020, Article ID 5792782, 17 pages
https://doi.org/10.1155/2020/5792782

Review Article

Sizing of a Standalone PV System with Battery Storage for a Dairy: A Case Study from Chile

Pablo Viveros,[1] **Francisco Wulff**,[1] **Fredy Kristjanpoller**,[1] **Christopher Nikulin**,[2] **and Tomás Grubessich**[1]

[1]*Department of Industrial Engineering, Universidad Técnica Federico Santa María, Avenida España, Valparaíso 1680, Chile*
[2]*Product Design Engineering Department, Universidad Técnica Federico Santa María, Avenida España, Valparaíso 1680, Chile*

Correspondence should be addressed to Pablo Viveros; pablo.viveros@usm.cl

Received 14 May 2020; Revised 14 October 2020; Accepted 28 November 2020; Published 12 December 2020

Academic Editor: Fran ois P r s

In this paper, a stochastic simulation model for a standalone PV system sizing is replicated and extended to supply a dairy's power demand. A detailed hourly-based simulation is conducted considering an hourly load profile and global solar radiation prediction model. The stochastic simulation model is based on a thorough statistical analysis of the solar radiation data and simulates the energy yield, the excess energy curtailed, and the state of charge of the batteries for the sizing month and the whole year, providing the designer autonomy factor values d to properly size the PV system, finding the optimum combination of installed peak power (P_m) and battery storage capacity (C_L) that meets the application load requirements, considering a preset reliability level at minimum cost. The model makes use of the NASA'S Surface Meteorology and Solar Energy database to obtain solar radiation data. Results show a substantial reduction of 44% in installed peak power and battery storage capacity when compared to conventional methodologies, considering three days of autonomy, and an 85% reduction considering four days. Considering the goodness of fit test results, the Wakeby distribution best represents the behavior of historical solar radiation data for the site in almost half of the months. This article seeks to contribute to the literature gap in the application of methodologies for the multicomponent power supply in the dairy industry through the use of renewable energy.

1. Introduction

Over the last decades, the amount of energy consumed by different types of industries has been growing enormously due to the economic growth of different countries, which in turn is triggered by the increase of world population and consumption patterns [1]. The environmental impact of human activities is contributing to rapid climate change due to CO_2 emissions, released as a result of fossil fuel combustion [2]. Solar energy is freely available and environmentally friendly, being widely adopted as an alternative for conventional electricity generation.

The sizing of standalone PV systems is an important task for the designers to find the most reliable and profitable combination of peak power and battery storage capacity that properly meets load demand, avoiding an equipment failure occurrence, considering the random behavior of solar radiation [3]. Therefore, the objective is to improve energy efficiency in terms of economic performance and reliability standards to ensure system autonomy.

In this sense, several methods have been proposed and developed in the literature to size standalone PV systems, which differ according to the approach used.

Before PV system sizing, in the design step, it is essential to identify the inherent losses present in the whole system to adequately size the system components. As for PV panels, the input energy may be reduced by shading, reflection due to the angle of incidence, dust, degradation (ageing), temperature effects, etc. Before the converted energy reaches the load demand, power losses due to the efficiency of the power conditioning units (inverters and charge controllers) need to be considered [4]. As for batteries, the charge/discharge process may lead to some power losses due to its internal resistance [5] and others in the battery storage system due to

the natural degradation of battery capacity [5]. Since the above thing directly affects PV system performance, the task consists of controlling the system operation to minimize output power losses.

It is also important to take into account the balance between energy yield by the PV array and energy consumed by the load. Any short-term mismatch between them must be counterbalanced usually by rechargeable batteries. It is for this reason that a factor of autonomy days or days of battery backup, d, has been proposed in the literature [5] to represent the number of days on which the PV system can meet the load demand in the absence of power generation and Peak Sun Hours (PSH). In other words, d represents the maximum number of days that the PV system can supply energy to loads independently.

In this sense, the research carried out by Kaplani and Kaplanis [6] presents a stochastic simulation model to determine the size of autonomous photovoltaic systems in terms of peak power (P_m) and battery storage capacity (C_L) for the critical month, thus ensuring compliance with annual reliability standards. The model, in addition to considering the days of autonomy as an exogenous variable, we consider as main inputs the analysis and simulation of daily-based solar radiation and the load profile, obtaining, as a result, the state of charge of the battery, the excess energy burnt, the energy losses, and the P_m and C_L results for sizing the system.

The main contribution of this article lies in the lack of the literature that focuses exclusively on standalone photovoltaic system sizing methods and their implementation in the dairy industry, using direct primary energy to satisfy the simultaneous demand of multiple components for the milking process [7]. Although research such as [8, 9] addresses the sizing of PV systems in dairies, they only focus on supplying certain specific components of the milking process, unlike this research. On the other hand, this article seeks to extend the original method of [6], incorporating an hourly load profile, and extending the stochastic simulation elaborated in this article to consider an hourly basis for the determination of the capacity indicators. In order to meet these goals, there were three primary objectives of this study:

(1) To size a standalone PV system for a dairy in the south of Chile by replicating the method proposed by [6], in conjunction with reliable hourly solar radiation prediction data, performing a detailed hourly-based simulation and deepening, thus, the model proposed by these authors

(2) To test the hourly-based proposed model for a different geographical location than the original study to supplement and extend its findings

(3) To validate if the results obtained from the hourly-based method reach similar reductions in installed peak power and battery storage capacity as compared to conventional methodologies, as demonstrated by the daily-based method proposed by [6]

The methodology is based on stochastic modeling of the performance of both of the PV array and the battery storage system and provides the energy delivered, the excess energy curtailed, the load profiles, and the state of charge of the batteries for the sizing month and the whole year, as well as data on the success rate for the determination of the autonomy factor d and the sizing of the system. To generate hourly solar radiation data, the model proposed in [10] is employed.

2. Literature Review

2.1. Standalone Photovoltaic Systems. Photovoltaic (PV) power generation systems are currently one of the systems with the fastest growth in the use of direct solar energy, considering the latter a clean, environmentally friendly source of energy, and abundantly available in most locations in the world [11]. These systems are fundamentally designed to transform solar radiation useful energy into electricity, through the use of photovoltaic panels. In addition to the panels, PV systems are composed of a current controller, an inverter to supply direct and alternating current loads, and a storage battery.

Although PV systems generally require a high initial investment, they can offer a cost-effective alternative to certain isolated off-grid locations, in comparison to the installation of power lines with central distribution [11]. In such scenarios, standalone photovoltaic (SAPV) systems become a plausible alternative since they can supply off-grid power generation. Therefore, the reliability of these systems becomes a relevant aspect to address, understanding it as the system capacity to satisfy load demand during a certain evaluation period [12].

The sizing of photovoltaic systems seeks to determine the optimal combination of variables associated with the arrangement of photovoltaic panels, battery storage capacity, and tilt angle, among other parameters related to the components, to obtain the best compromise between power reliability and system cost. Given the vast diversity of approaches, Khatib et al. [11] classify the different sizing methods reviewed in the literature into intuitive methods, numerical methods, analytical methods, computational tools, computational intelligence methods, and hybrid methods. Likewise, the evaluation criteria are classified into technical parameters, considering the loss of load probability, loss of power supply probability, the state of charge (SOC), among others; economic parameters, considering the net present value (NPV) and the annualized cost; and social parameters, such as portfolio risk or even social acceptance.

Regarding the restrictions related to SAPV, the literature review developed by Fouad et al. [13] reveals that the factors with the greatest impact on the performance of photovoltaic systems analyzed are related to the battery efficiency and physical photovoltaic panels structure. Meanwhile, Khatib et al. [11] add physical space budget and energy constraints related with load profile fulfillment in the case of SAPV as essential aspects considered in the reviewed literature.

2.2. PV Systems on Daily Farms. The use of PV systems can offer more than just a cost-efficient alternative versus conventional electricity generation systems. Recently, these

types of alternatives also respond effectively to a requirement promoted both by the scientific community and by international organizations for the mitigation of anthropogenic emissions. One of the strategies recommended in research such as [14] is the proposal of primary energy saving measures through the implementation of PV systems for power generation and consumption in dairy farms.

In this sense, research such as developed by Breen et al. ([15], cited in [7]) reveals the potential of photovoltaic systems to substantially reduce carbon dioxide emissions in dairy farms, one of the main current anthropogenic emissions. In the particular case of this industry, a growth of 22% in milk production is expected for the 2018–2027 period [16], and 19% in its consumption by 2050 ([17], cited in [7]). From the energy consumption perspective, it is added the fact that electrical energy represents one of the items with the highest consumption for both conventional and organic dairy farms, representing 14% and 24% of total primary energy consumption, respectively, according to [18]. Considering the above, the incentives for photovoltaic system installation, particularly in the dairy industry, are considerable.

Regarding the review of related research, case studies such as De Blas et al. [8] and Zhang et al. [19] address the implementation and sizing of photovoltaic systems supplying certain components or subsystems within diary productive system, namely, the pumping system and the cooling system, respectively. Breen et al. [7], meanwhile, elaborate a generalizable and scalable optimization model for PV system sizing in dairy farms, using for the definition of the objective function the weighted sum method, considering both a financial criterion and the contribution ratio in terms of autoconsumed energy with respect to gross annual electricity consumption, using a trade-off parameter to search for and analyze Pareto-optimal solutions. Unlike the previously reviewed models, the model implemented in this research considers an autonomous system, which is why the need to ensure reliability and autonomy arises before the consideration of financial evaluation criteria or the contribution from renewable energy.

2.3. Estimation of Global Solar Radiation. Solar radiation measuring results are essential for the design and operation of solar energy technologies, especially for the implementation of solar collectors and the sizing of photovoltaic systems. However, the measurement of this magnitude, despite being accurate, is not always available for modeling and forecasting, mainly due to the high initial investment and maintenance costs for its measurement and recording instruments [20]. For this reason, the alternative that arises within the scientific community is to correlate radiation with other available meteorological parameters, developing a wide variety of models based on this approach.

One of the first empirical models is the one developed by Angstrom [21] for the measurement of monthly average daily global radiation, developing an empirical correlation model considering the effect of the daily average of clear-sky radiation and the sunlight duration fraction. Meanwhile, Page [9] decides to modify the correlation model, replacing the clear-sky radiation with the daily average extraterrestrial radiation on a horizontal surface. Since then, a series of empirical models have been developed that have modified and extended the Angstrom–Page model, including new parameters or different mathematical expressions to establish empirical relationships. See, for example, [22, 23] reviews that demonstrate the great variety of empirical models carried out to date, as well as the location dependence of the parameters.

Another alternative that has emerged in recent decades for estimating and forecasting global radiation is the use of artificial neural networks, where the inputs usually correspond to a great variety of meteorological and geographical parameters, while the output corresponds to the global solar radiation for different time scales [24]. The main advantages of these models for estimating global radiation are their location adaptability, and a perceptible accuracy improvement compared to traditional empirical models. However, the disadvantage of these models is its complexity and the high number of parameters in order to achieve such accuracy, not clearly establishing the effect of this inclusion, in comparison to empirical models [20].

In an attempt to classify for the first time the vast variety of models for global radiation in terms of the time span contemplated and the type of method approached, Zhang et al. [20] develop a literature review classifying them according to the output into monthly average of daily global radiation, daily global radiation, and hourly global radiation method. The authors also consider two main categories regarding the type of estimation: empirical models, which are disaggregated into sunlight duration fraction (SDF) models, modified models (MSDF), nonsunshine duration fraction (NSDF) models, and artificial neural networks (ANNs). The study also presents performance indicators such as the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) to determine the accuracy of each reviewed model based on observed and predicted values. The research concludes that the classical models of fraction of solar duration (SDF), as well as artificial neural networks (ANNs), have the best performance in terms of RMSE and MAPE. Although SDF models have slightly lower performance than ANN models, the latter generally requires a large amount of data for the training phase to improve the performance indicators presented.

Considering this last classification, the model developed by Kaplani and Kaplanis [6] falls into the category of the long-run monthly average daily global radiation. The methodology obtains a representative value of global daily radiation throughout a month, then averaging the values obtained year by year, for each month. The summarized methodology associated with this model, as well as the nomenclature used and replicated in this investigation, is presented in Section 6 and Appendix A, respectively.

The aforementioned sizing method considers a monthly fitting for daily solar radiation using probability distributions. In this regard, specialized distributions such as the Wakeby distribution have proven their effectiveness in particular for power generation modeling in planning

applications of PV power systems [25]. Also, the Johnson SB and Generalized Extreme Value distributions are usually considered for modeling meteorological measurements such as wind speed, generally providing a superior fit to one-component probability density functions [26].

3. Design of the Dairy

The case study is carried out in a dairy located in the city of Río Bueno, in the south of Chile, with coordinates south latitude 40.331 and east longitude 72.498. In the economic field, agricultural equipment prices have substantially increased and, conversely, milk prices have continued to decrease considering the producer's perspective. There are few cost-saving opportunities in these industries, being necessary to analyze more cost-effective options for power generation. In this respect, Chile is endowed with abundant solar radiation, and the amount of electrical energy that can be generated from solar panels is potentially significant.

The electrical load of the diary consists of equipment for milking, cooling, cleaning, water heating, illuminating, and ventilation, among others, such as milking machine, vacuum and water pump, cooling tank, cleaning systems, and light bulbs. The dairy performs two milkings per day, from 5:30 to 7:15 and from 15:30 to 17:15 considering ordinal hours. The milking process is described below, as well as the energy consumption required by the dairy equipment for this case study:

(i) Milking machine: its main function is the aspiration to activate the teat cups, directing the milk obtained towards the cold tank for subsequent accumulation. The machine has 14 units and is equipped with a vacuum pump, which has a 2.2 kW electric motor including an additional 10% for the operation of the pulsators. Also, a 0.5 HP motor is added to drive the milk to the cold tank. Its operation time matches the milking process, from 5:30 to 7:15 and from 15:30 to 17:15.

(ii) Cold tank: it is the device that allows the storage of milk in optimal conditions. It is equipped with a 4 kW engine and has a capacity of 4,275 liters. It has two 70 W agitators each and a ventilation system, which consists of two 200 W fans each. The operation time takes place after the milking, and all its components operate simultaneously. The period is between 7:15 and 11:00 and between 17:45 and 22: 00.

(iii) Lighting: the lighting includes the use of luminaires for the milking room, a corridor, and the machine room that contain the devices described above. This item gives a total of 472 W, but it is estimated that it reaches around 600 W by adding the warehouse and surroundings. The lighting schedule coincides with the milking in the morning, that is, between 5:30 and 7:15 hrs.

(iv) Hot water: the warm water is used for machine cleaning and to wash the cow's udders. The water heater demands 1500 W power supply, and it is estimated that it works approximately 6.4 hours per day, to maintain the water temperature at about 65°C. Schedules in which the water heater operates are between 06:45 and 08:25 and between 16:45 and 18:25 hrs (milking process), and between 13:00 and15:45 hrs (cold tank).

(v) Water pump: in order to supply the dairy with water, a 1 HP submersible pump is used. It exceeds a total height of 25 meters, with a flow of 3 m^3/hr. Average daily water consumption corresponds to 11 m^3. Therefore, the pump operates for 3 hours and 40 minutes, consuming approximately 2.74 kW per day. The hours are usually between 06: 45 and 09:15 and between 16:45 and 17:55.

(vi) Machine washing: to keep the tank in optimal conditions, it must be washed after milk collection. The closed tanks have an automatic washing system, which uses hot and cold water combined with acidic and basic cleaning products. The motor that allows this is 1 HP. The duration of the tank washing is 45 minutes, immediately after the collection by the collection truck, which occurs around 12:30 pm. In the case of the milking machine, the washing seeks to remove organic matter, fats, and minerals for the milk. The purpose is to prevent the proliferation of bacteria. The automatic washing is carried out, using a 0.5 HP automatic machine minutes after each milking. The period of milking machine washing is 30 minutes, at 08:00 hrs and 18:00 hrs.

Table 1 summarizes the consumption and operating times associated with dairy equipment.

The total load demand for the dairy was estimated to be 59.83 kWh/day, which corresponds to the peak summer demand. Summer has the highest energy consumption compared to other seasons. This is because the cooling tank has to operate for a longer period to keep milk at an appropriate temperature. Since the dairy operates at certain times of the day, especially during milking, the building of a detailed hourly load profile is required and presented in Figure 1.

4. Solar Radiation

4.1. Solar Radiation Data Analysis. In this study, based on the model proposed by [6], the daily global solar radiation data (H) were obtained from the NASA's Surface Meteorology and Solar Energy database [27] for the dairy located at south latitude 40.331 and west longitude 72.948, considering the period between 1985 and 2004. For each of the 20 years, based on the recommended average days for months [28] as cited in [29], the average day for each month, including the day both before and after the average day (i.e., three consecutive days), was taken to form the global solar radiation data for the analysis. The global solar radiation H for each month at the site concerned is shown in Figure 2.

For each month, solar radiation data were fitted by MLE (see Appendix C) considering six continuous probability

Table 1: Consumption and operation times of the diary's equipment.

Equipment	Consumption (kW)	Operation times
Cold tank	4.54	07:15–11:00; 17:45–22:00
Milking Machine	2.573	05:30–07:15; 15:30–17:15
Hot water	1.5	06:45–08:23; 16:45–18:23 (milking) 13:00–15:45 (tank)
Lighting	0.6	05:30–07:15
Water pump	0.746	06:45–9:15; 16:45–17:55
Others (Machine washing)	0.373–0.746	08:00–08:30; 18:00–18:30 (milking) 12:30–13:15 (tank)

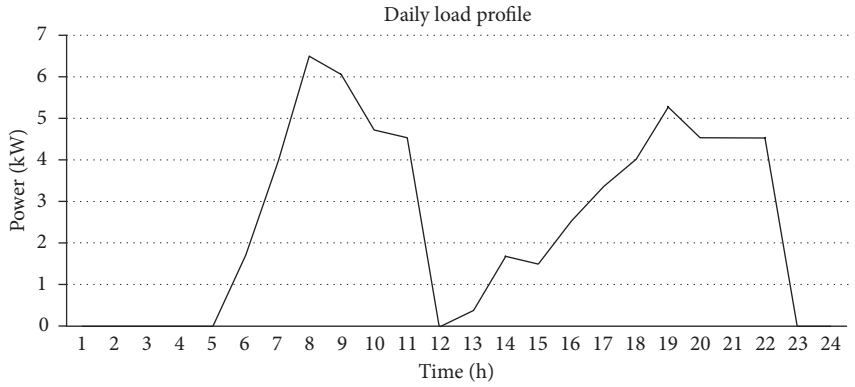


Figure 1: Load profile of the dairy.

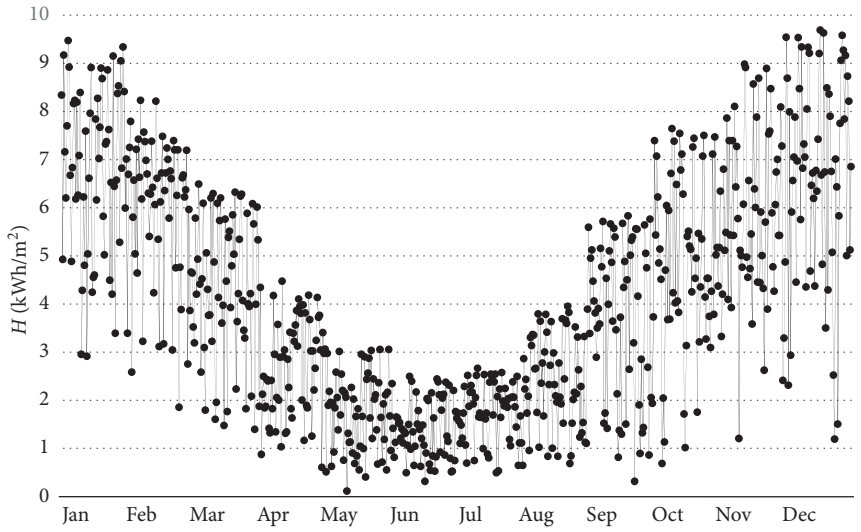


Figure 2: Daily global solar radiation H on an average day, including the day both before and after the daily average per month for the years 1985–2004 for the site concerned.

distributions, namely, Beta, Generalized Extreme Value, Johnson SB, Normal, Wakeby, and Weibull distributions (see Table 2), and were tested according to the Kolmogorov–Smirnov test (K-S test). The K-S statistic values are shown in Table 3. Then, the p value is obtained and used as selection criteria. It is important to notice that in this case, the null hypothesis H_0 establishes that the data follow the specified distribution for each distribution; therefore, the selection must consider a p value that cannot reject that hypothesis (see Appendix D). Following this observation, the distribution with the highest possible p value is chosen for each month.

The p value was calculated using a novel method named Exact-KS-FFT, expressing the p value as a double-boundary noncrossing probability for a homogeneous Poisson process, which is then efficiently computed using fast Fourier transform [30]. The method has proved his accuracy and numerical efficiency versus the algorithm proposed by Marsaglia et al. [31]. The results for the fitted distributions are presented in Table 4.

Table 2: Probability distributions to average daily global solar radiation modeling, for each month.

Distribution	Probability density function	Parameters
Beta	$f(x;a,b,\alpha_1,\alpha_2) = (1/(B(\alpha_1,\alpha_2)))\cdot((x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}/(b-a)^{\alpha_1+\alpha_2-1})$, $B(\alpha_1,\alpha_2) = \int_0^1 t^{\alpha_1-1}(1-t)^{\alpha_2-1}dt$;	a,b: distribution Limits $\alpha_1,\alpha_2>0$
Generalized Extreme Value	$f(x,k,\mu,\sigma) = \begin{cases} (1/\sigma)\exp(-(1+kz)^{-(1/k)})(1+kz)^{-1-(1/k)}), & k\neq 0 \\ (1/\sigma)\exp(-z-\exp(-z)), & k=0 \end{cases}$	$z=((x-\mu)/\sigma)$ k,μ,σ
Johnson SB	$f(x;y,\delta,\gamma,\xi) = (\sigma/(\lambda\sqrt{2\pi}z(1-z)))\exp(-(1/2)(\gamma+\delta\ln(z/(1-z)))^2)$	$z=((x-\xi)/\lambda)$ y,δ,γ,ξ
Normal	$f(x;\mu,\sigma) = (1/(\sigma\sqrt{2\pi}))e^{((-(x-\mu)^2)/(2\sigma^2))}$	μ,σ
Wakeby	$f(F;\alpha,\beta,\gamma,\delta,\xi) = \xi+(\alpha/\beta)(1-(1-F)^{\beta})-(\gamma/\delta)(1-(1-F)^{-\delta})$	$\alpha,\beta,\gamma,\delta,\xi$
Weibull	$f(x;\alpha,\beta) = (\alpha/\beta)(x/\beta)^{\alpha-1}\exp(-(x/\beta)^{\alpha})$	α,β

Table 3: The K-S statistic values of the fitted distributions on the pdf of each month for the site.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Beta	0.115	0.114	0.093	0.054	0.147	0.079	0.122	0.057	0.052	0.126	0.099	0.079
Gen. Extreme Value	0.073	0.080	0.076	0.074	0.067	0.082	0.061	0.091	0.093	0.086	0.064	0.080
Johnson SB	0.055	0.106	0.072	0.062	0.076	0.055	0.081	0.055	0.050	0.094	0.075	0.085
Normal	0.096	0.175	0.097	0.083	0.071	0.089	0.105	0.107	0.124	0.084	0.072	0.103
Wakeby	0.053	0.107	0.079	0.062	0.069	N/A	0.075	0.054	0.048	0.078	0.081	0.068
Weibull	0.082	0.166	0.111	0.100	0.126	0.087	0.136	0.086	0.172	0.144	0.077	0.166

Table 4: p values associated to the K-S statistic value for the fitted distributions.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Beta	0.377	0.387	0.643	0.991	0.135	0.820	0.308	0.984	0.994	0.273	0.565	0.820
Gen. Extreme Value	0.883	0.808	0.853	0.873	0.934	0.784	0.969	0.669	0.643	0.734	0.953	0.808
Johnson SB	0.989	0.478	0.893	0.964	0.853	0.989	0.796	0.989	0.997	0.630	0.863	0.747
Normal	0.604	0.045	0.591	0.772	0.902	0.695	0.490	0.466	0.290	0.759	0.893	0.515
Wakeby	0.993	0.466	0.820	0.964	0.919	N/A	0.863	0.991	0.998	0.831	0.796	0.926
Weibull	0.784	0.065	0.420	0.552	0.273	0.721	0.198	0.734	0.051	0.150	0.842	0.065

For the months of January, August, September, October, and December, the Wakeby distribution provides the best fit, while for the months of February, May, July, and November, the Gen. Extreme Value distribution works best. For March and June, the Johnson SB distribution is the most appropriate. Finally, for April the Beta distribution provides the best fit. Although the Normal distribution is often chosen to fit the data, in this case, it does not present substantial differences in terms of the statistic for the rest of the fitted distributions. It has to be emphasized that while this study considers only these six types of distributions, other alternatives may eventually improve the fit.

As already stated, to achieve yearly performance, PV system sizing is carried out based on the "worst month" method, i.e., the month with the lowest average solar radiation levels. Thus, June is selected for sizing in the following sections.

4.2. Model for Estimating the Hourly Global Solar Radiation. When hour-by-hour performance calculations need to be done, it may be necessary to start with daily data and then obtain hourly values from daily solar radiation estimation. The ratio of hourly global solar radiation in a day n_j, $I(h;n_j)$, to daily global solar radiation, $H(n_j)$, is calculated from equation (1) [29]. Baig et al. [10] modified the model proposed by Jain et al. [32] which tries to fit solar radiation to a Gaussian-type function, to better fit the recorded data during the start and the end periods of a day, as cited in [33]. In this model, r_h is estimated by equation (2), where h is the solar time, σ_G is the standard deviation of the Gaussian curve, and S_o is the day length of the day n_j, at a site with latitude ϕ, given by equation (3), where δ is the sun's declination. S_o correlates with σ_G, equations (4) and (5), based on the second version of the new approach to Jain's model as cited in [33]. $I(h;n_j)$ is determined from equation (1), leading to equation (6). The equations used in this section are listed as follows:

$$r_h = \frac{I\left(h;n_j\right)}{H\left(n_j\right)}, \tag{1}$$

$$r_h = \frac{1}{2\sigma_G\sqrt{2\pi}}\left\{\exp\left(-\frac{(h-12)^2}{2\sigma_G}\right)+\cos\left[180\,\frac{(h-12)}{(S_o-1)}\right]\right\}, \tag{2}$$

$$S_o = \frac{2}{15}\cos^{-1}(-\tan\phi\cdot\tan\delta), \tag{3}$$

$$S_o = 4.054\cdot\sigma_G, \tag{4}$$

$$\sigma_G = 0.246 \cdot S_o, \tag{5}$$

$$I\left(h, n_j\right) = r_h \cdot H\left(n_j\right). \tag{6}$$

5. Proposed Standalone PV System

The schematic representation of the proposed standalone PV system is shown in Figure 3. The main components are PV generator, battery system, and power conditioning units (inverters and charge controllers).

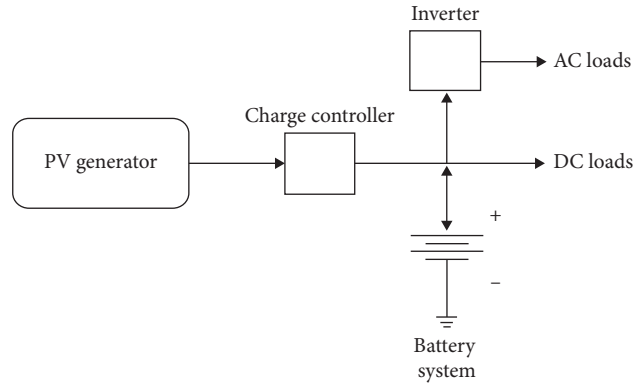


Figure 3: Standalone PV system configuration.

6. Stochastic PV Sizing Simulation Model

The stochastic PV sizing methodology corresponds to the model developed by [6], but considering some modifications specifically in steps 3 and 6 since a detailed hourly-based simulation is carried out. It consists of the following steps:

(1) An essential first step is to obtain a reliable database that contains appropriate solar radiation and temperature data for the years under consideration. The application site is defined by entering latitude and longitude values. Based on a thorough statistical analysis of the global solar radiation data H, a probability density function (pdf) is generated for each month in such a way as to find the best fitting curve to the solar radiation data for the site. Then, the parameters of the fitted distribution are extracted for all months.

(2) The global solar radiation mean (H_m) and the standard deviation (σ_H) are calculated for each month. The one that exhibits the lowest average daily solar radiation levels is selected for sizing.

(3) Depending on the application type, estimation of the energy demand is carried out based on technical datasheets and expert judgment. Unlike [6], an hourly load profile is used for a more detailed hourly-based simulation.

(4) Estimation of the correction factor F, which refers to the energy lost along the path from the PV generator to the loads during day-to-day operations, and of the correction factor F' for the operation route battery-DC/AC inverter-loads, calculated accordingly to [6]. Estimation of R_m, based on the beam and diffuse components of the monthly radiation data according to the Collares-Pereira and Rabl model as cited in [29] and stated in [6]. To estimate F, C_{Tc} is calculated from equation (7), where T_C is the cell temperature and TC is the temperature coefficient, obtained from the manufacturer's specifications. T_C is calculated using equation (8) and can be estimated from the maximum average ambient temperature T_a, the irradiance G under standard test conditions (STCs), and the nominal operating cell temperature (NOCT) as cited in [29]. Combining these two equations leads to equation (9):

$$C_{T_c} = \left(T_C - 25^\circ\mathrm{C}\right) \cdot \mathrm{TC}, \tag{7}$$

$$T_C = T_a + \left(\frac{\mathrm{NOCT} - 20^\circ\mathrm{C}}{0.8\left(\mathrm{kW/m^2}\right)}\right) \cdot G, \tag{8}$$

$$C_{T_c} = \left(\left(T_a + \left(\frac{\mathrm{NOCT} - 20^\circ\mathrm{C}}{0.8\left(\mathrm{kW/m^2}\right)}\right) \cdot G\right) - 25^\circ\mathrm{C}\right) \cdot \mathrm{TC}. \tag{9}$$

(5) The lower and upper range values for the autonomy factor d are set depending on climatic conditions and the application in question. The higher the solar radiation, the lower the autonomy factor. d is set from d_{start} to d_{end}, ranging from 1 to 10. The range values are taken from [6].

(6) The simulation algorithm starts with iterating d from d_{start} to d_{end}. After each iteration, the corrected peak power $P_{m,cor}$ and the corrected battery storage capacity $C_{L,cor}$ are calculated with the H_m and σ_H previously calculated in step 2, as stated in [6]. Each iteration of d is comprised of 100 simulations for which the energy delivered E_{PV}, the hourly load demand $Q_{L,h}$, the energy loss E_{loss}, and the battery state of charge SOC are registered, for all hours of the month selected for sizing, to estimate the success rate in percentage provided by the specific value of d, for every hour of the month.

For each day of the month, the global solar radiation $H(n_j)$ is sampled from the selected distribution with parameters extracted in step 1. Therefore, for all hours of the month r_h is calculated using equation (2), thus obtaining hourly global radiation $I(h, n_j)$ for a particular day n_j. The hourly clearness index k_T is calculated using equation (10), where I_{ext} is the hourly extraterrestrial radiation on a horizontal surface for an hour between hour angles ω_1 and ω_2 [29]. For converting the hourly solar radiation from the horizontal surface to the inclined plane, tilted at

slope β from the horizontal, R is calculated using the isotropic diffuse model as derived by [34] from the isotropic model proposed by [35], as cited in [29], given by equation (11), where (I_b/I) is the fraction of the hourly radiation on a horizontal plane which is the direct beam, R_b is the geometric factor, i.e., the ratio of beam radiation on the tilted surface to that on a horizontal surface at any time, calculated based on the hour angle ω evaluated at the midpoint of the hour, declination δ for the sampling day, site latitude ϕ, and slope β, I_d/I is the fraction of the hourly radiation on a horizontal plane which is diffuse, based on the Erbs et al. correlation [36] as cited in [10], given by equation (12), and ρ_g is the reflectance (the albedo) of the ground:

$$k_T = \frac{I\left(h, n_j\right)}{I_{ext}}, \tag{10}$$

$$R = \frac{I_b}{I}R_b + \frac{I_d}{I}\left(\frac{1+\cos(\beta)}{2}\right) + \rho_g\left(\frac{1-\cos(\beta)}{2}\right), \tag{11}$$

$$\frac{I_d}{I} = \begin{cases} 1.0 - 0.09k_T, & k_T \le 0, \\ 0.9511 - 0.1604k_T + 4.388k_T^2 - 16.638k_T^3 + 12.336k_T^4, & 0.22 < k_T \le 0.80, \\ 0.165, & k_T > 0. \end{cases} \tag{12}$$

For each hour, the energy delivered by the PV array E_{PV} is calculated using equation (13), and the remaining amount of energy after satisfying all load requirements is denoted by DE:

$$E_{PV} = P_{m,cor} \cdot I \cdot R. \tag{13}$$

To express the amount of energy remaining after consumption as a percentage, the auxiliary variable AUX is defined. In cases where the PV system is not generating any electricity ($E_{PV} = 0$), DE is estimated from equation (14), as the battery system is supplying the demand loads and AUX is estimated from equation (15):

$$DE = E_{PV} - Q_{L,h} \cdot F', \tag{14}$$

$$AUX = \frac{DE}{C_{L,cor} \cdot V}. \tag{15}$$

Alternatively, in cases where the PV system is generating power ($E_{PV} > 0$), DE is then estimated from equation (16). For the latter, three different cases are considered:

$$DE = E_{PV} - Q_{L,h} \cdot F. \tag{16}$$

(a) If DE < 0, then AUX is calculated using equation (17) since the battery bank must satisfy the load demand and the SOC of the battery is decreased:

$$AUX = \frac{DE \cdot F'}{C_{L,cor} \cdot V}. \tag{17}$$

(b) If DE = 0, then AUX is equal to zero since the energy yield by the PV system perfectly matches the load requirements.

(c) If DE > 0, then AUX is estimated using equation (15), since the PV system managed to cover all the loads and the SOC of the battery is increased.

Thus, the battery state of charge SOC is calculated from equation (18), considering the remaining energy, where SOC_a is the SOC at the moment immediately before the current period, that is, the previous hour, considering a simulation on an hourly basis. Any excess energy generated by the system is dissipated and is represented by E_{loss}, given by equation (19). For each simulation hour, the SOC is evaluated. If the value drops below the critical level 1-DOD, then a failure occurs and the simulation stops. The same process is repeated until the last hour of the month, recording the success rate for each value of d, from d_{start} to d_{end}:

$$SOC = SOC_a + AUX, \tag{18}$$

$$E_{loss} = (SOC - 1) \cdot C_{L,cor} \cdot V. \tag{19}$$

(7) Depending on the percentage of the time, it is desired to meet the load demand, and the minimum value of d that fulfills the system reliability criterion is selected.

(8) Based on the selected value of d obtained in step 7, the system sizing is carried out by substituting d into

$$P_{\mathrm{m,cor}} = Q_{\mathrm{L}} \cdot F \cdot \frac{\left(1 + \sqrt{d} \cdot 2 \cdot \left(\sigma_{\mathrm{H}}/H_{\mathrm{m}}\right)\right)}{\left(\mathrm{PSH}_{\mathrm{m}} \cdot R_{\mathrm{m}}\right)}, \tag{20}$$

$$C_{\mathrm{L,cor}} = Q_{\mathrm{L}} \cdot F' \cdot \frac{\left(1 + \sqrt{d} \cdot 2 \cdot \left(\sigma_{\mathrm{H}}/H_{\mathrm{m}}\right)\right)}{\left(V_{\mathrm{DC}} \cdot \mathrm{DOD}\right)}, \tag{21}$$

calculating $P_{\mathrm{m,cor}}$ and $C_{\mathrm{L,cor}}$, as stated in [6].

(9) The simulation is performed for the remaining months of the year, taking into account the sizing parameters obtained in step 8. For each month, the goal is to obtain the success rate and the system's performance, that is, the E_{PV}, $Q_{\mathrm{L,h}}$, E_{loss}, and SOC.

For each month, if simulation findings show that the system reliability criterion (success rate) for the selected value of d was not met, then return to step 7 and select the next incremented d.

These last two methodology steps correspond to a general vision of the system's performance during the whole year while ensuring the fulfillment of the criterion of "system reliability." The whole simulation model is presented in Appendix B. Please note that both flowchart and methodology correspond to the model proposed by [6], but with some modifications consistent with a detailed hourly-based simulation.

7. Simulation Results

Based on the "worst month" method, the month selected for sizing is June. The load demand was estimated to be 59.84 kWh/day. Assuming a maximum average ambient temperature in the daytime T_{a} of 8°C, calculated from the NASA's database for the years in question (1995 to 2004), as mentioned in the beginning of this study, the NOCT and the temperature coefficient TC are taken as (for example, from the technical specifications for the module PV-UD185MF5) 47.5°C and 0.452%/°C, respectively. Based on equation (9), C_{Tc} is calculated as 0.9215. C_{charger}, C_{inverter}, $C_{\mathrm{bat-c}}$, $C_{\mathrm{bat-d}}$, $C_{\mathrm{pv-ag}}$, and $C_{\mathrm{bat-ag}}$ are taken as 0.98 [37], 0.9 [37], 0.95 [5], 0.95 [5], 0.9936 [38], and 0.8 [39], respectively. Thus, correction factors F and F' of 1.30 and 1.46, respectively, were calculated, based on [6]. It should be kept in mind that these values are customized to the application in question and are subject to the design, equipment, and technology used. The conversion factor R was calculated considering the photovoltaic panels inclined at an angle of 25° above the horizontal. This slope was calculated based on maximizing the total estimated energy for the winter months [29]. The batteries' discharge depth DOD was set to 80%, and the DC transfer voltage V was 120 V. The autonomy factor d was set from d_{start} to d_{end}, ranging from 1 to 10, which was increased by 0.5 each iteration. The success rates obtained for the corresponding iteration value of d with the $P_{\mathrm{m,cor}}$ and $C_{\mathrm{L,cor}}$ calculated using equations (20) and (21), respectively, are shown in Figure 4. From this figure, the minimum value of d is selected to achieve at least 95% success for noncritical loads and at least 99% for critical loads. The success rates obtained for the $P_{\mathrm{m,cor}}$ and $C_{\mathrm{L,cor}}$, calculated by replacing d with its corresponding iteration value, are, respectively, depicted in Figures 5 and 6. It is worth highlighting that the PV system sizing, apart from being affected by the autonomy factor d, strongly depends on the weather conditions, since the $P_{\mathrm{m,cor}}$ and $C_{\mathrm{L,cor}}$ are also affected by the correction parameter $(\sigma_{\mathrm{H}}/H_{\mathrm{m}})$, that is, the global solar radiation data at the site. Regarding the first claim, it is logical

that the more autonomy days a PV system has, the larger the size of the PV generator and battery bank will be.

For June, an autonomy factor d of 3 assures at least an average of 97% success for noncritical loads. The $P_{\mathrm{m,cor}}$ calculated is 97.074 kWp and $C_{\mathrm{L,cor}}$ is calculated to be 2.308 kAh. Contrasting the results with the conventional methodology [5], d_{cr} is calculated from linear equation from [6]. For noncritical loads, $d_{\mathrm{n-cr}}$ is estimated to be 4.532. Thus, P_{m} and C_{L} of 173.443 kWp and 4.124 kAh, respectively, are calculated, based on equations (1) and (2) from [6]. The proposed stochastic methodology, based on the model proposed by [6], far exceeds the conventional methodology, achieving a 44% reduction in installed peak power and battery storage capacity. Repeating the same analysis, for $d = 4$, the proposed methodology assures at least an average of 99% success for critical loads for June. In this case, the $P_{\mathrm{m,cor}}$ and $C_{\mathrm{L,cor}}$ obtained are 106.171 kWp and 2.525 kAh, respectively, as summarized in Table 5. From the conventional point of view, d is calculated using d_{cr} linear equation from [6], which results in $d_{\mathrm{cr}} = 18.11$ and gives $P_{\mathrm{m}} = 693.083$ kWp, based on equation (1) and $C_{\mathrm{L}} = 16.481$ kAh, based on equation (2) from [6], as summarized in Table 6. Comparing the results, an 85% reduction in installed peak power and battery storage capacity is achieved with the proposed methodology, getting better results than with the conventional methodology.

The hourly performance of the proposed system throughout the sizing month June and $d = 3$, in terms of E_{PV}, $Q_{\mathrm{L,h}}$, E_{loss}, and SOC, is shown in Figure 7. It can be seen from Figure 7 that the proposed standalone PV system successfully meets the load demand without the SOC falling below 60%. The battery bank was able to cover the energy demand for those hours at which no energy was generated by the PV modules. Besides, the amount of energy dissipated was considerable.

Conversely, Figure 8 shows a failed case where the low levels of solar radiation and, thus, the E_{PV} value caused the batteries to be unable to meet the demand in loads with a SOC falling below 20%.

7.1. Designed Standalone PV System Hourly Performance for the Other Months. The algorithm is run for the remaining months using the corresponding values of d, $P_{\mathrm{m,cor}}$, and $C_{\mathrm{L,cor}}$, as previously calculated, in such a way to achieve at least 95% and 99% of system autonomy. As the sizing is based on the month with the lowest solar radiation, it is expected to perform better. The success rates for each month

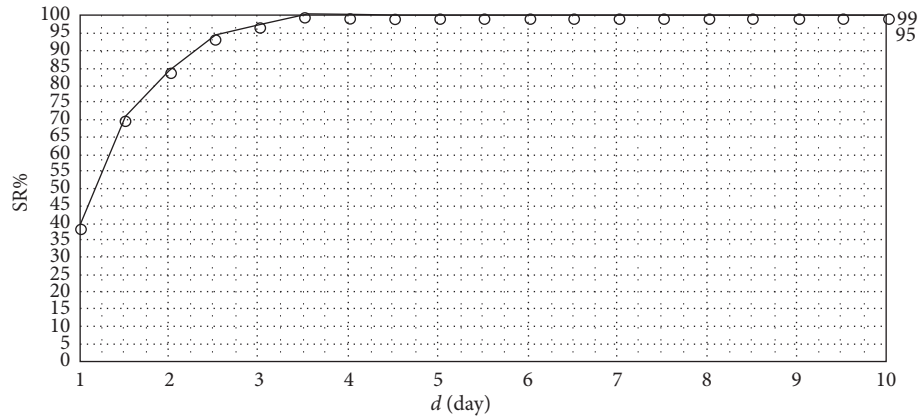


Figure 4: Success rates SR% vs different values for the autonomy factor d.

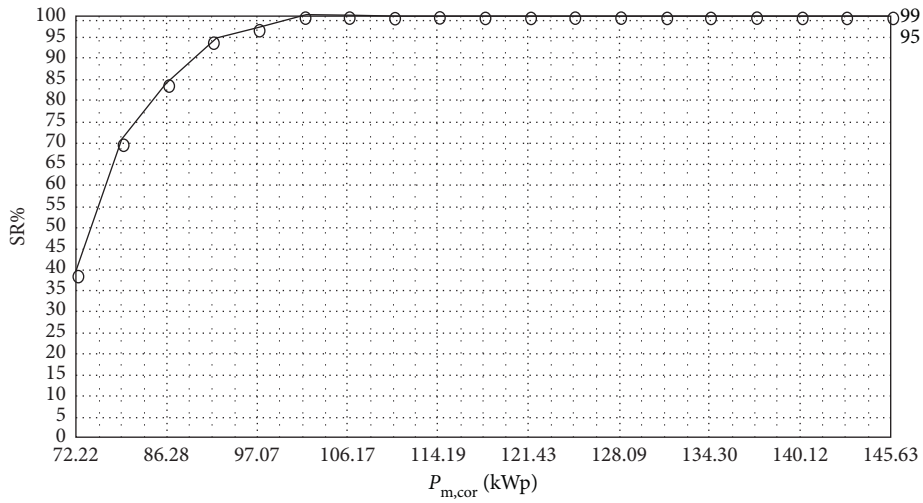


Figure 5: Success rates SR% vs corrected peak power $P_{\mathrm{m,cor}}$.

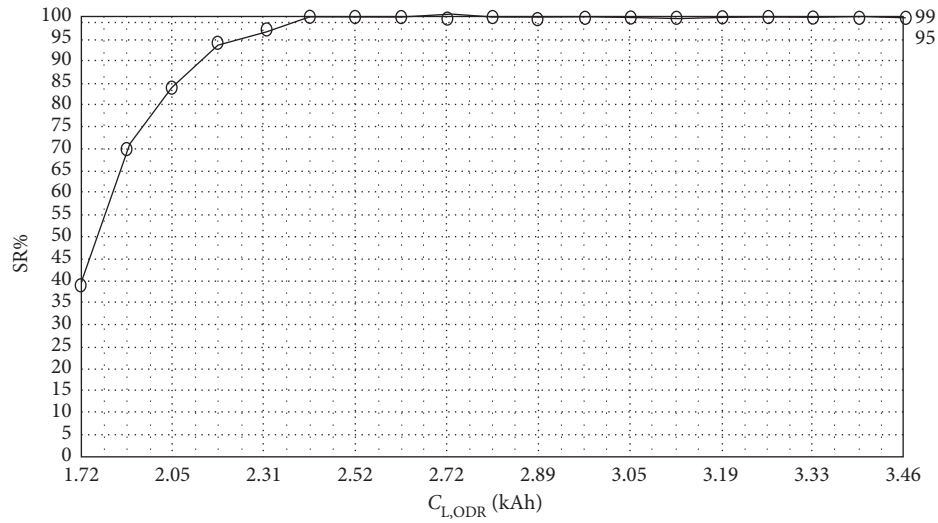


Figure 6: Success rates SR% vs corrected battery storage capacity $C_{\mathrm{L,cor}}$.

Table 5: Success rates SR%, corrected peak power $P_{m,cor}$, and corrected battery storage capacity $C_{L,cor}$ values for different values of the autonomy factor d for the proposed stochastic methodology.

d	1	1, 5	2	2, 5	3	3, 5	4	4, 5	5
SR%	39	70	84	94	97	100	100	100	100
$P_{m,cor}$ (kWp)	72.22	79.85	86.28	91.95	97.07	101.79	106.17	110.29	114.2
$C_{L,cor}$ (kAh)	1.72	1.90	2.05	2.19	2.31	2.42	2.52	2.62	2.72

Table 6: Success rates SR%, peak power P_m, and battery storage capacity C_L values for different values of the autonomy factor d for the conventional methodology.

d	$d_{n-cr} = 4.53$	$d_{n-cr} = 18.11$
SR%	95	99
P_m (kWp)	173.44	693.08
C_L (kWA)	4.12	16.48

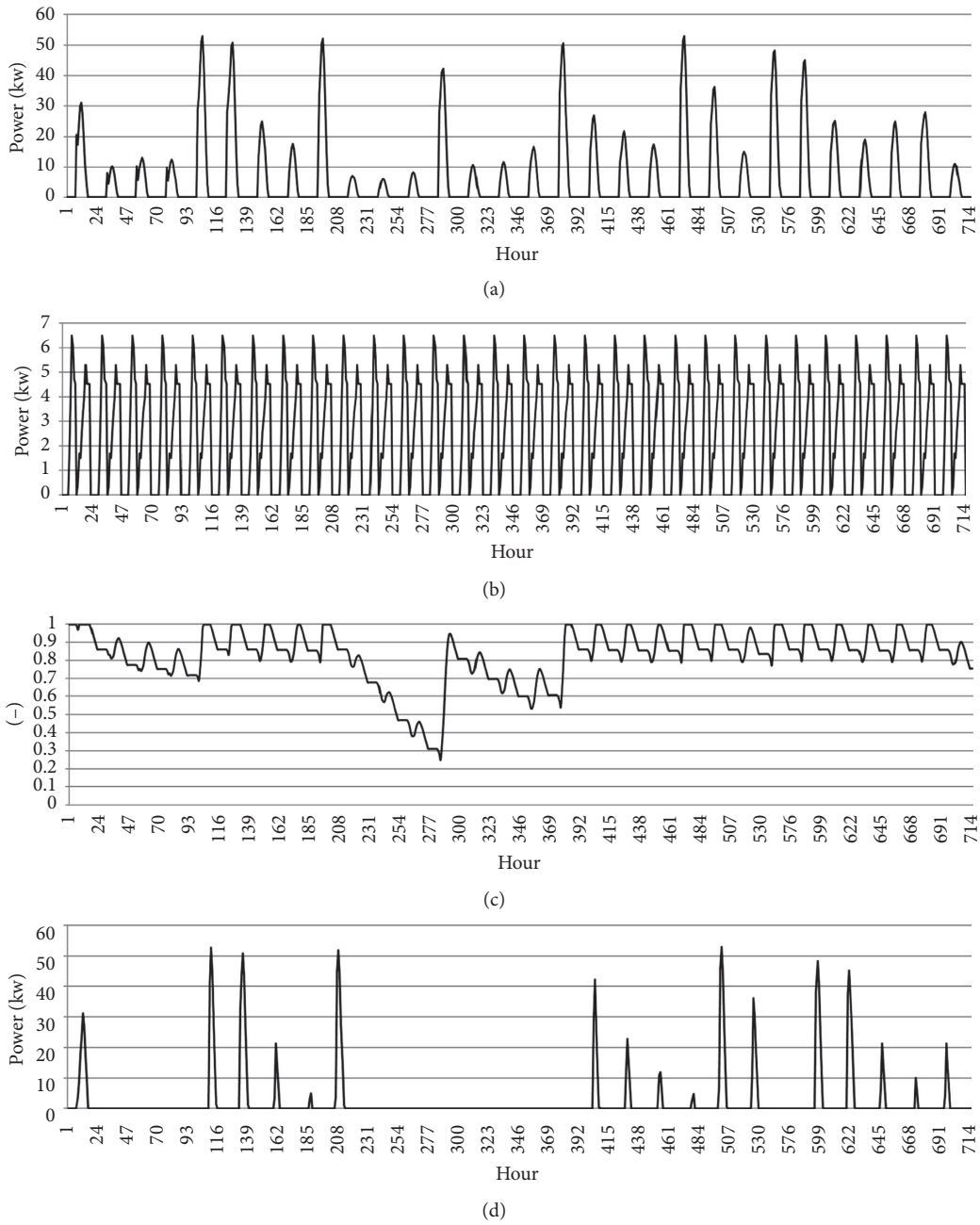


Figure 7: Designed system hourly performance for all hours of June, with $d = 3$, in terms of the energy delivered E_{PV}, load demand $Q_{L,h}$, excess energy curtailed E_{loss}, and state of charge of the batteries SOC. (a) E_{PV}, (b) Q_L, (c) SOC, and (d) E_{LOSS}.

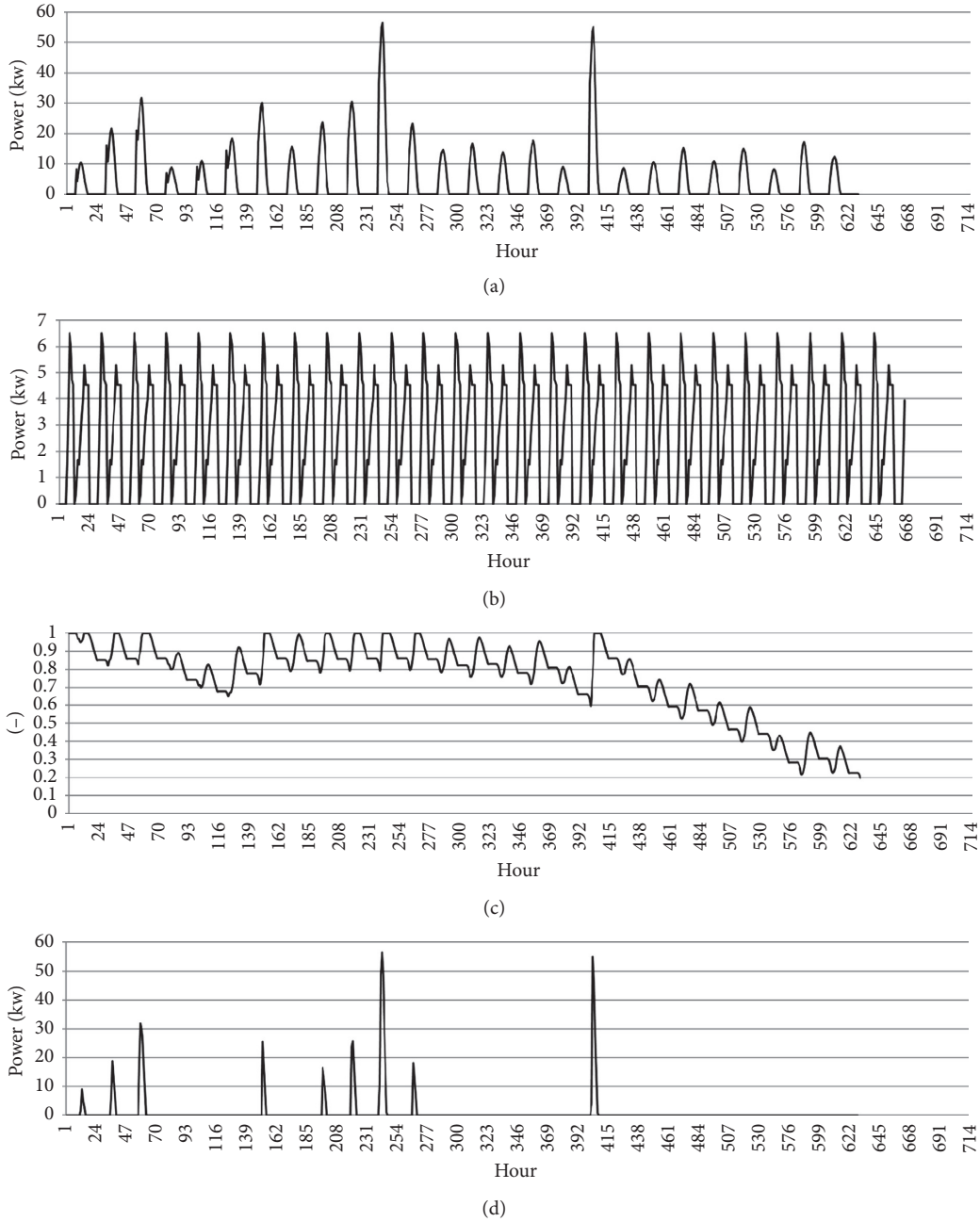


FIGURE 8: Designed system hourly performance for a failed case for June, with $d = 3$, in terms of the delivered energy E_{PV}, load demand $Q_{L,h}$, excess energy curtailed E_{loss}, and state of charge of the batteries SOC. (a) E_{PV}, (b) Q_L, (c) SOC, and (d) E_{LOSS}.

are shown in Table 7, including both critical and noncritical operation. Although the values are not identical, these success rates assure a minimum of 95% when operating with noncritical loads.

8. Discussion

The results obtained by adopting the proposed methodology reveal substantive differences with respect to the forecast made using the conventional approach. However, it is necessary to discuss the assumptions and considerations made throughout the research to elucidate the causes of this difference.

Regarding the estimation of solar radiation, the same period was considered both for the adjustment using probability distributions and for determining the Kolmogorov–Smirnov statistic to measure the goodness of the fit. On the other hand, the number of parameters of the different distributions was not considered when performing the statistical test. This can result in overfitting based on the period analyzed, affecting its adaptability to estimate or forecast more recent data.

Table 7: Success rates SR% for all months of the year in order to achieve at least 95% and 99% of system autonomy.

	d	$P_{m,cor}$ (kWp)	$C_{L,cor}$ (kAh)	SR%											
				Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Noncritical loads	3	97.07	2.31	100	100	100	100	96	97	99	100	100	100	100	99
Critical loads	4	106.17	2.52	100	100	100	100	99	100	100	100	100	100	100	100

The method considered for estimating radiation is classified into long-term monthly average of daily solar radiation. According to the research carried out by [20], this type of estimate can result in lower estimation errors for the contemplated period, although the superiority and usability of the model are questionable. It is mainly because of the estimation error that a relatively stable and long enough period was considered in the radiation data to avoid the problems that the method itself entails. The aforementioned one may partially explain the difference between the results obtained versus conventional methodology. Another additional aspect to consider is the uncertainty that the use of meteorological data from sources such as NASA or METEONORM can provide, according to [40], precisely because the radiation data have not been routinely observed in most of the world's meteorological stations due to the high cost of instruments and technical requirements.

In the design phase of the dairy, it is necessary to mention that the proposed scenario combines the month with the lowest average daily global radiation according to the data provided, with the summer month with peak consumption. Therefore, the capacity required to meet the energy demand is being overestimated. The scenario, therefore, delivers a solution that is overestimating the storage capacity and thus becoming suboptimal, at least for the critical month, to meet the reliability standards at the lowest cost.

As in the research proposed in [6], this case considers a daily estimation method for solar radiation, but the main difference lies in the update of the different state variables involved to measure the compliance of reliability standards. In this sense, while the base method has a daily scale, the method proposed in this research updates the state variables and decomposes the global solar radiation considering an hourly scale. Furthermore, the extended method is not validated by a real implementation of the case study, not being able to determine which sizing method is more accurate.

9. Conclusions

The results obtained from the stochastic simulation model show that the proposed standalone PV system can successfully meet the dairy's load demand at specified reliability, achieving a reduction of 44% in installed peak power and battery storage capacity for noncritical operation, and an 85% reduction for critical operation, as compared to conventional methodology, which is consistent with the results obtained in base research, thus validating the hourly-based method for a different geographical location.

However, given the characteristics of the sizing method, its result has to be contrasted with sophisticated and highly accurate models for the estimation and designing phase in order to overcome the limitations previously analyzed. In this sense, another aspect that can be expanded is the consideration of stochastic components for solar radiation on an hourly scale. The conditions for the design of this case study may also consider the possibility of simulating the energy supplied by the PV system with the estimated load profile for each month of the year, thus avoiding capacity overestimation.

One of the characteristics of the sizing of autonomous photovoltaic systems is the compliance with certain reliability levels; in this sense, it is possible to consider for future research the incorporation of a multiobjective function that considers not only the storage capacity, but also the implementation costs and other relevant performance indicators, as well as the incorporation of budget or size constraints for the installation of PV modules.

Given the current world energy scenario, it is now desirable to think of different sustainable alternatives for electricity generation in such a way as to minimize the environmental impact. Solar energy is widely available, and its use is increasing enormously, which leads to new challenges in improving the efficiency of PV systems and balancing supply and demand. If some loads can be shifted from night to day, better results can be obtained. Thus, the concept of load management becomes relevant to sizing standalone PV systems. In this sense, dynamic load management functionality can be introduced. In more recent studies, the effect of load profile uncertainty on the off-grid PV systems optimum design is analyzed, highlighting the importance when defining load profiles for off-grid PV-battery systems.

Finally, it is therefore recommended to perform an hourly, or even a more detailed simulation (e.g., simulation at minutely resolution), to have better control of the load profile and the power output fluctuations, clearly identifying the time when peak demand occurs, and utilize this valuable information for getting optimal results.

Appendix

A. Nomenclature Used, Based on Kaplanis and Kaplani [6]

Nomenclature

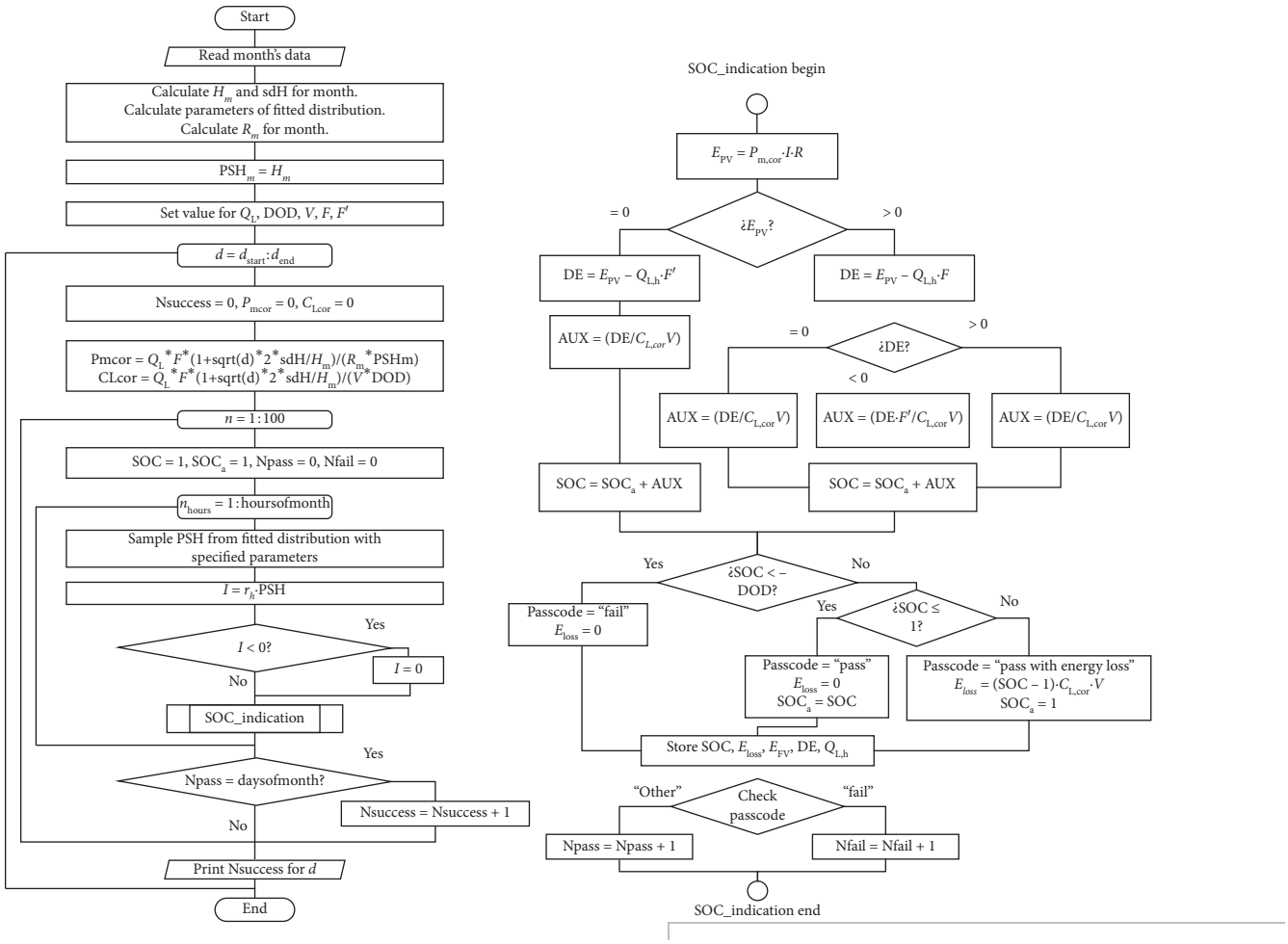
AUX: auxiliary variable
a: lower bound of beta distribution
b: upper bound of beta distribution

C_{bat-ag}: for the ageing of the battery system due to cycling
C_{bat-c}: the efficiency of the batteries during charging process
C_{bat-d}: efficiency of the batteries during the discharging process
$C_{charger}$: efficiency of the charger
$C_{inverter}$: efficiency of the inverter
C_L: battery storage capacity (kAh)
$C_{L,cor}$: C_L corrected (kAh)
C_{pv-ag}: for the effect of the ageing of the PV
C_{Tc}: correction term for the temperature effect on PV efficiency
DE: the remaining amount of energy after satisfying all load requirements
DOD: depth of discharge of battery (dimensionless)
d: number of days of autonomy of a PV system
d_{cr}: number of days of autonomy of a PV system operating with critical loads
d_{n-cr}: number of days of autonomy of a PV system operating with noncritical loads
E_{loss}: energy dissipated due to excess energy and fully charged battery (kWh)
E_{PV}: energy delivered by the PV system (kWh)
F: loss of energy between the route of the PV generator and load demand (dimensionless)
F': loss of energy between the route of the battery bank and loads (dimensionless)
G: irradiance under standard test conditions (STC) (1 kW/m^2)
H: daily global solar radiation on the horizontal (kWh/m^2)
H_m: mean value of H (kWh/m^2)
I: hourly global solar radiation on the horizontal in a day n_j (kWh/m^2)
I_b: hourly direct beam radiation
I_d: hourly diffuse radiation
I_{ext}: hourly extraterrestrial radiation on a horizontal surface for an hour period
k: shape parameter of Gen. Extreme Value distribution
k_T: hourly clearness index
NOCT: nominal operating cell temperature (°C)
P_m: peak power (kWp)
$P_{m,cor}$: P_m corrected (kWp)
PSH: Peak Sun Hours (h/day)
PSH_m: mean value of PSH (h/day)
PSH_{min}: minimum value of PSH (h/day)
Q_L: total loads in a day (Wh/day)
$Q_{L,h}$: load demand at any time (kWh)
r_h: ratio of hourly global solar radiation to daily global solar radiation (dimensionless)
R_m: factor for converting the global solar radiation from the horizontal to the inclined plane of the PV panels, mean value for a specific month (dimensionless)
R: factor for converting the hourly solar radiation from the horizontal surface to the inclined plane, for a specific hour (dimensionless)
R_b: ratio of beam radiation on the tilted surface to that on a horizontal surface at any time
S_o: length of the day (h)
SOC: state of charge of battery % (dimensionless)
SOC_a: the remaining SOC carried onto the next hour (dimensionless)
TC: temperature coefficient
T_a: maximum average ambient temperature in the daytime for the sizing month (°C)
T_C: cell temperature (°C)
V: DC transfer voltage. Also, nominal battery operating voltage (V)

Greek letters

α: scale parameter of Wakeby distribution; shape parameter of Weibull distribution
α_1, α_2: shape parameters of beta distribution
β: shape parameter of Wakeby distribution; scale parameter of Weibull distribution; slope
γ: shape parameter of Johnson SB distribution; scale parameter of Wakeby distribution
δ: declination; shape parameter of Johnson SB distribution; shape parameter of Wakeby distribution
λ: scale parameter of Johnson SB distribution
ξ: location parameter of Johnson SB and Wakeby distributions
μ: location parameter of Gen. Extreme Value and Normal distributions
ρ_g: albedo
σ: scale parameter of Normal and Gen. Extreme Value distributions
σ_H: standard deviation of H
σ_G: standard deviation of the Gaussian curve
ϕ: site latitude
ω: hour angle

B. Flowchart of the Stochastic Proposed Methodology, Based on Kaplani and Kaplanis [6]



C. Maximum Log-Likelihood Estimation

To adjust the distributions for the subsequent comparison of the goodness of fit, the parametric estimation of the distributions is performed using the maximum likelihood estimation (MLE). Considering the set $\{X_i, i = 1, \ldots, n\}$ of n independent and identically distributed random variables, taken from a continuous probability distribution characterized by the distribution parameter θ, and representing its probability density function as $f(x|\theta)$, the likelihood function L is given by the following joint probability function P:

$$L(x_1, x_2, \ldots, x_n|\theta) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|\theta), \tag{C.1}$$

where the realization of said random variables, that is, the observations, is represented by the set $\{x_i\}_n$. Developing the previous expression and considering the condition of random variables (iid), the likelihood function is given by

$$L(x_1, x_2, \ldots, x_n|\theta) = f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta) f(x_2|\theta), \ldots, f(x_n|\theta). \tag{C.2}$$

Therefore, the maximum likelihood estimator $\widehat{\theta}_{\text{ML}}$ is the estimator of the parameter θ that maximizes the value of this joint probability function. That is,

$$L\left(x_1, x_2, \ldots, x_n|\widehat{\theta}_{\text{ML}}\right) = \max_{\theta} L(x_1, x_2, \ldots, x_n|\theta). \tag{C.3}$$

The above sentence applies without loss of generality for the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$.

D. Kolmogorov–Smirnov Test

The Kolmogorov–Smirnov (K-S) test is commonly used to decide if a sample from a population comes from a determined continuous probability distribution, from which its specification parameters are known. The test is defined by the hypotheses:

H_0: the sample follows a specified distribution

H_a: the sample does not follow a specified distribution

And the K-S statistic D_n is defined as

$$D_n = \sup_x \left|F_n(x) - F(x)\right|, \tag{D.1}$$

where $F_n(x)$ is the empirical distribution and $F(x)$ is the theoretical cumulative distribution function calculated over ordered observations $\{X_i, i = 1, \ldots, n\}$ of the sample. This value is compared with the K-S tables of two-tailed critical values $D_{\mathrm{cr},\alpha}$ for a given significance percentage α, which usually takes a value of 5%. Therefore, the null hypothesis is rejected when the condition $D_n > D_{\mathrm{cr},\alpha}$ is satisfied. To quantify this decision, avoiding the dependence of the significance value set a priori by the researcher to reject the null hypothesis, the p value is calculated to choose between the fitted distributions.

Data Availability

The information used in this study is referenced, respectively, and the data from the case study are obtained from a dairy in southern Chile.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. W. Sun, "Changes in energy consumption and energy intensity: a complete decomposition model," *Energy Economics*, vol. 20, no. 1, pp. 85–100, 1998.

[2] G. P. Peters, G. Marland, C. Le Quéré, T. Boden, J. G. Canadell, and M. R. Raupach, "Rapid growth in CO_2 emissions after the 2008-2009 global financial crisis," *Nature Climate Change*, vol. 2, no. 1, pp. 2–4, 2012.

[3] D. Watts, M. F. Valdés, D. Jara, and A. Watson, "Potential residential PV development in Chile: the effect of net metering and net billing schemes for grid-connected PV systems," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1037–1051, 2015.

[4] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Conversion and Management*, vol. 51, no. 10, pp. 1929–1937, 2010.

[5] M. Roger and V. Jerry, *Photovoltaic Systems Engineering*, CRC Press, Boca Raton, FL, USA, 2000.

[6] E. Kaplani and S. Kaplanis, "A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations," *Applied Energy*, vol. 97, pp. 970–981, 2012.

[7] M. Breen, J. Upton, and M. D. Murphy, "Photovoltaic systems on dairy farms: financial and renewable multi-objective optimization (FARMOO) analysis," *Applied Energy*, vol. 278, p. 115534, 2020.

[8] M. De Blas, J. Appelbaum, J. L. Torres, A. García, E. Prieto, and R. Illanes, "A refrigeration facility for milk cooling powered by photovoltaic solar energy," *Progress in Photovoltaics: Research and Applications*, vol. 11, no. 7, pp. 467–479, 2003.

[9] J. Page, "The estimation of monthly mean values of daily total shortwave radiation on vertical and inclined surfaces from sunshine records for latitude 40°N-40°S," in *Proceedings of the UN Conference on New Sources of Energy*, pp. 378–390, Rome, Italy, August 1961.

[10] A. Baig, P. Akhter, and A. Mufti, "A novel approach to estimate the clear day global radiation," *Renewable Energy*, vol. 1, no. 1, pp. 119–123, 1991.

[11] T. Khatib, I. A. Ibrahim, and A. Mohamed, "A review on sizing methodologies of photovoltaic array and storage battery in a standalone photovoltaic system," *Energy Conversion and Management*, vol. 120, pp. 430–448, 2016.

[12] S. Alsadi and T. Khatib, "Photovoltaic power systems optimization research status: a review of criteria, constrains, models, techniques, and software tools," *Applied Sciences*, vol. 8, no. 10, p. 1761, 2018.

[13] M. M. Fouad, L. A. Shihata, and E. I. Morgan, "An integrated review of factors influencing the performance of photovoltaic panels," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1499–1511, 2017.

[14] C. J. Lanigan, T. Donncilan, K. Hanrahan et al., *An Analysis of Abatement Potential of Greenhouse Gas Emissions in Irish Agriculture 2021–2030*, Teagasc, Carlow, Ireland, 2018.

[15] M. Breen, J. Upton, and M. D. Murphy, "Photovoltaic systems on dairy farms: an economic analysis," in *Proceedings of the Annual International Meeting*, Detroit, MI, USA, July 2018.

[16] Food and Agriculture Organization, *Outlook 2018–2027 Dairy and Dairy Products 2018*, Food and Agriculture Organization, Rome, Italy, 2018, http://www.agrioutlook.org/commodities/Agricultural-Outlook-2018-Dairy.pdf.

[17] J. Bruinsma and N. Alexandratos, *World Agriculture Towards 2030/2050: The 2012 Revision,* Food and Agriculture Organization, Rome, Italy, 2012, http://www.fao.org/docrep/016/ap106e/ap106e.pdf.

[18] P. Shine, J. Upton, P. Sefeedpari, and M. D. Murphy, "Energy consumption on dairy farms: a review of monitoring, prediction modelling, and analyses," *Energies*, vol. 13, no. 5, p. 1288, 2020.

[19] C. Zhang, P. Campana, J. Yang, J. Zhang, and J. Yan, "Can solar energy be an alternative choice of milk production in dairy farms?—A case study of integrated PVWP system with alfalfa and milk production in dairy farms in China," *Energy Procedia*, vol. 105, pp. 3953–3959, 2017.

[20] J. Zhang, L. Zhao, S. Deng, W. Xu, and Y. Zhang, "A critical review of the models used to estimate solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 314–329, 2017.

[21] A. Angstrom, "Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation," *Quarterly Journal of the Royal Meteorological Society*, vol. 50, no. 210, pp. 121–126, 1924.

[22] K. Bakirci, "Models of solar radiation with hours of bright sunshine: a review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 9, pp. 2580–2588, 2009.

[23] F. Besharat, A. A. Dehghan, and A. R. Faghih, "Empirical models for estimating global solar radiation: a review and case study," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 798–821, 2013.

[24] J. Qin, Z. Chen, K. Yang, S. Liang, and W. Tang, "Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products," *Applied Energy*, vol. 88, no. 7, pp. 2480–2489, 2011.

[25] A. S. Al-Sumaiti, M. H. Ahmed, S. Rivera, M. S. El Moursi, M. M. A. Salama, and T. Alsumaiti, "Stochastic PV model for power system planning applications," *IET Renewable Power Generation*, vol. 13, no. 16, pp. 3168–3179, 2019.

[26] C. Jung and D. Schindler, "Global comparison of the goodness-of-fit of wind speed distributions," *Energy Conversion and Management*, vol. 133, pp. 216–234, 2017.

[27] Nasa, Surface Meteorology and Solar Energy, Sol Energy 2011, http://org/commodities/Agricultural-Outlook-2018-Dairy.pdf.
[28] S. A. Klein, "Calculation of monthly average insolation on tilted surfaces," *Solar Energy*, vol. 19, no. 4, pp. 325–329, 1977.
[29] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, Wiley, Hoboken, NJ, USA, 4th edition, 2013.
[30] D. S. Dimitrova, V. K. Kaishev, and S. Tan, "Computing the Kolmogorov-smirnov distribution when the underlying CDF is purely discrete, mixed, or continuous," *Journal of Statistical Software*, vol. 95, no. 10, pp. 1–42, 2020.
[31] G. Marsaglia, W. W. Tsang, and J. Wang, "Evaluating Kolmogorov's distribution," *Journal of Statistical Software*, vol. 8, no. 18, pp. 1–4, 2003.
[32] P. C. Jain, S. Jain, and C. F. Ratto, "A new model for obtaining horizontal instantaneous global and diffuse radiation from the daily values," *Solar Energy*, vol. 41, no. 5, pp. 397–404, 1988.
[33] S. N. Kaplanis, "New methodologies to estimate the hourly global solar radiation; comparisons with existing models," *Renewable Energy*, vol. 31, no. 6, pp. 781–790, 2006.
[34] B. Y. H. Liu, R. C. Jordan, S. Dongaonkar, and M. A. Alam, "The long-term average performance of flat-plate solar-energy collectors," *Solar Energy*, vol. 7, no. 2, pp. 53–74, 1963.
[35] H. C. Hottel and B. B. Woertz, "The performance of flat-plate solar–heat collectors," *Transactions of the American Society of Mechanical Engineers*, vol. 64, pp. 91–104, 1942.
[36] D. G. Erbs, S. A. Klein, and J. A. Duffie, "Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation," *Solar Energy*, vol. 28, no. 4, pp. 293–302, 1982.
[37] A. Luque and S. Hegedus, *Handbook of Photovoltaic Science and Engineering*, Wiley, Hoboken, NJ, USA, 2nd edition, 2011.
[38] N. Lukač, S. Seme, D. Žlaus, G. Štumberger, and B. Žalik, "Buildings roofs photovoltaic potential assessment based on LiDAR (light detection and ranging) data," *Energy*, vol. 66, pp. 598–609, 2014.
[39] A. Green, "Life cycle costing for batteries in telecom applications," in *Proceedings of the INTELEC—Twentieth International Telecommunications Energy Conference (Cat. No.98CH36263)*, pp. 1–7, San Francisco, CA, USA, October 1998.
[40] L. Wang, O. Kisi, M. Zounemat-Kermani, G. A. Salazar, Z. Zhu, and W. Gong, "Solar radiation prediction using different techniques: model evaluation and comparison," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 384–397, 2016.

Hindawi
Complexity
Volume 2020, Article ID 1065810, 17 pages
https://doi.org/10.1155/2020/1065810

Research Article

Bioinspired Mitigation Scheme for Cascading Failures in Farmland Wireless Sensor Networks

Jun Wang,[1] **Zhuangzhuang Du**,[2] **Xunyang Wang**,[3,4] **and Zhengkun He**[5]

[1]*School of Electrical Engineering, Henan University of Science and Technology, Luoyang, Henan 471000, China*
[2]*School of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, Henan 471003, China*
[3]*Department of Applied Mathematics, Lanzhou University of Technology, Lanzhou, Gansu 730050, China*
[4]*Postdoctoral Research Station in Gansu Electric Power Research Institute, Wanxin South Road, Anning District, Lanzhou, Gansu 730000, China*
[5]*School of Computer Science and Engineering, Central South University, Changsha, Hunan 410000, China*

Correspondence should be addressed to Xunyang Wang; 12198114@163.com

Received 12 June 2020; Accepted 16 October 2020; Published 5 November 2020

Academic Editor: Fran ois P r s

Existing mitigation strategies on wireless sensor networks (WSNs) against cascading failures cannot appropriately adapt the particular characteristics of farmland WSNs. Spider web provides a new reference for improving network invulnerability. In this study, a bionic network scheme is built based on symmetric analysis of a series of spider-web vibration transmission trials, which include networking methodology, communication rules, and load capacity model. The basic idea of this scheme is to apply the cascading-failure coping mechanism inspired by spider web into the construction and operation of farmland WSNs. We found that the link backup contributed by a topological structure and communication rules had positive effects on suppressing the spread of cascading failures. The study showed that the damages of cascading failures can be efficiently lowered by regulating the adjustment coefficient of the load capacity model. The difference between the inner-layer node failures and outer-layer node failures for network invulnerability was verified under deliberate attack circumstances. Based on these results, the proposed network scheme can be utilized to enhance the invulnerability performance of farmland WSNs.

1. Introduction

Wireless sensor networks (WSNs) are a distributed network system consisting of a large number of sensor nodes to wirelessly cooperate in perceiving and processing various information [1, 2]. Due to the coupling relationship between node traffics, failure nodes will lead to malfunctions of neighboring nodes resulting in a cascading effect [3]. Cascading faults inevitably lead to topological segmentation, reduced communication connectivity, limited network coverage, and increased likelihood of network paralysis [4]. Therefore, cascading failures are an essential factor affecting network invulnerability and should be widely investigated both by theoretical analyses and experimental characterizations.

Some aspects of cascading failures in WSNs have been discussed in recent works, such as the cascade control, defense strategy, and analytical calculation of load distribution [5, 6]. Identification and reinforcement of the critical nodes that guarantee the network functionality have been proved to have an inhibiting effect on the network disruptions produced by cascading failures [7, 8]. However, it is still challenging to assure the accuracy of quantifying the node importance and the stability of the network topology after adding redundant nodes [9]. Predetected attack mode or dynamic topology repair has been discovered to improve the response efficiency for cascading failures [10, 11]. Nevertheless, the targeted repairs based on the real-time detection of attacks are more feasible [12]. Furthermore, it is revealed that the initial load assignment of nodes and the load distribution strategy of fault nodes have a positive impact on avoiding the spread of cascading failures [13]. But for different application scenarios, these established load

capacity models demonstrate notable performance differences [14].

As an innovative data acquisition method in precision agriculture, farmland WSNs can continuously monitor indispensable environmental factors for crop growth and generate remarkable labor-saving benefits to farmers [15]. Compared with other WSN applications, influenced by the agriculture production mode, farmland WSNs present unique characteristics in terms of network topology, information transmission, and load capacity [16]. First, due to the large-scale monitoring area, multiple transmission constraints, and limited cost budget, it is essential to introduce multilevel relay nodes to achieve reliable data transfer with low-density deployment, and the network topology displays the feature of hierarchical clustering [17, 18]. The combined factors of node mobility, network heterogeneity, and chain-directed transmission dramatically boost the dynamic uncertainty of network behavior, which make farmland WSNs exhibit substantial topological variability [19]. Besides, the pattern of multinode collaborative monitoring causes overlap and interaction between interlayer nodes, intercluster nodes, and communication links, resulting in network transmission with both interlayer coupling association and intersecting coupling association [20]. Lastly, the entire network shows multilevel differences in load capacity, as a consequence of differences in detecting tasks and monitoring frequencies for nodes [21]. In summary, farmland WSNs are a complex network system with the characteristics of coupling and vulnerability, and a few node or link faults are prone to provoke cascading failure disasters. The current mitigation techniques of cascading failures have the shortcoming of overidealized assumptions and sole coping means [22]. It is an urgent requirement to develop a specialized countermeasure scheme for farmland WSNs by systematic consideration of cascading failure mechanism and mitigation measures.

Over hundreds of millions of years, the spider web has evolved an elegant, ultralight, and destruction-resistant structure [23]. The previous studies tested the hypothesis that the orb spider web has a stunning similarity in topological structures and component functions with farmland WSNs and can be used to promote the network capability against cascading failures [24]. The existing research focuses on the invulnerability performance of artificial spider-web topology, but the enlightenment of vibration transmission characteristics of destructed spider web on the alleviation of cascading failures has not been discussed [25]. By investigating the transmission mechanism of spider-web vibration, we can develop a bioinspired network scheme for improving the invulnerability of farmland WSNs.

The objective of this study is to build a bionic network scheme of cascading-failure mitigation for farmland WSNs inspired by the vibration transmission reactions of artificial spider web under various destruction cases. The major contributions of this study are summarized as follows:

(1) To properly characterize the response process of spider web to cascading damages, an experimental method based on artificial spider-web vibration testbed is proposed.

(2) A network scheme comprised of networking methodology, communication rules, and load capacity model is presented in the face of random and deliberate attacks on nodes.

(3) Through extensive trials, the soundness and effectiveness of the network scheme against cascading failures are verified. The impact of scheme parameters on invulnerability is also been explored.

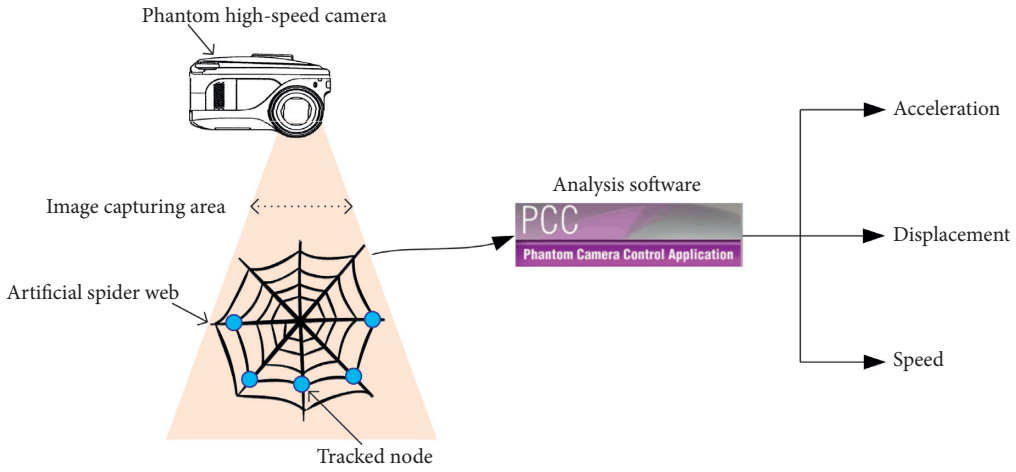
The remainder of this paper is arranged as follows: the damage experiments of artificial spider web and the suppression mechanism of cascading propagation are given in Section 2; the details of the proposed network scheme for farmland WSNs are displayed in Section 3; the evaluations on the performances of network scheme and the impacts of key parameters are exhibited in Section 4; finally, conclusions are shown in Section 5.

2. Artificial Spider-Web Experiments and Mechanism Analysis

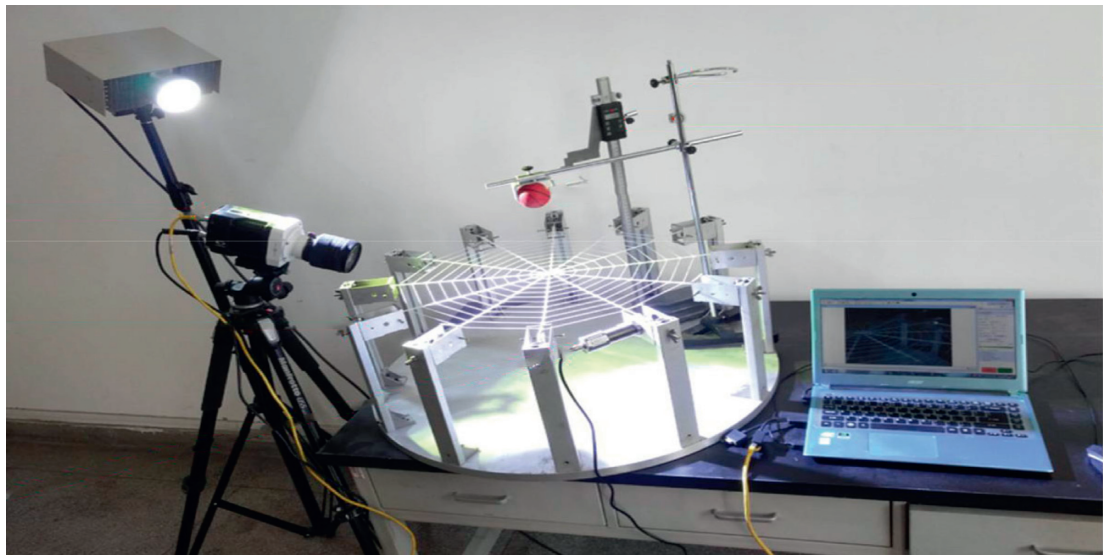
We use the developed artificial spider-web vibration testbed (Figure 1) to clarify the cascading spread characteristics of spider web after damage to partial components [26]. In the experiments, a rubber ball with a mass of 20 g and a diameter of 50 mm is utilized as the excitation source. Each radius line from the central node and the horizontal connection line of each layer that encircled the central node are defined as a radial line and a spiral line, respectively. The intersection of a radial line and a spiral line is designated as a node. Moreover, the radial line between adjacent spiral lines is described as a radial path, and the spiral line between adjacent radial lines is identified as a spiral path. Moreover, radial paths, spiral paths, and nodes constitute the artificial spider-web topology.

2.1. Node Damage. Figure 2 indicates the variations in peak-to-peak values of longitudinal vibrations of 51 nodes on radial lines 6–10 before and after cutting off four nodes at intersections of radial lines 6–7 and spiral lines 1–2. It can be noticed that the vibrations at these nodes decrease from the inner layers to the outer layers without changing the attenuation characteristics of nodes on a radial line, except the vibrations of each node increase insignificantly. The peak-to-peak values of node vibrations in areas A1 and B1 are 1.20 and 1.10 times of that in the complete web, and the vibrations of nodes in areas $C1$, $D1$, and $D2$ are consistent with that of the complete web. These results show that the node failures have a slight effect on the vibration transmission of spider web, and the affected area is within the adjacent layers, manifesting that the cascading propagation range is limited. Furthermore, in trapezoid area A1 and triangle area B1, the amplitude increments of vibrations for six nodes are the same, showing that the triangle and trapezoid structures can play the role of dissipation and inhibition of cascading spread.

2.2. Radial Line Damage. Figure 3 displays the variations in peak-to-peak values of longitudinal vibrations of nodes on radial lines 6–10, 7–11, and 8–12 as cutting out radial line 5,



(a)

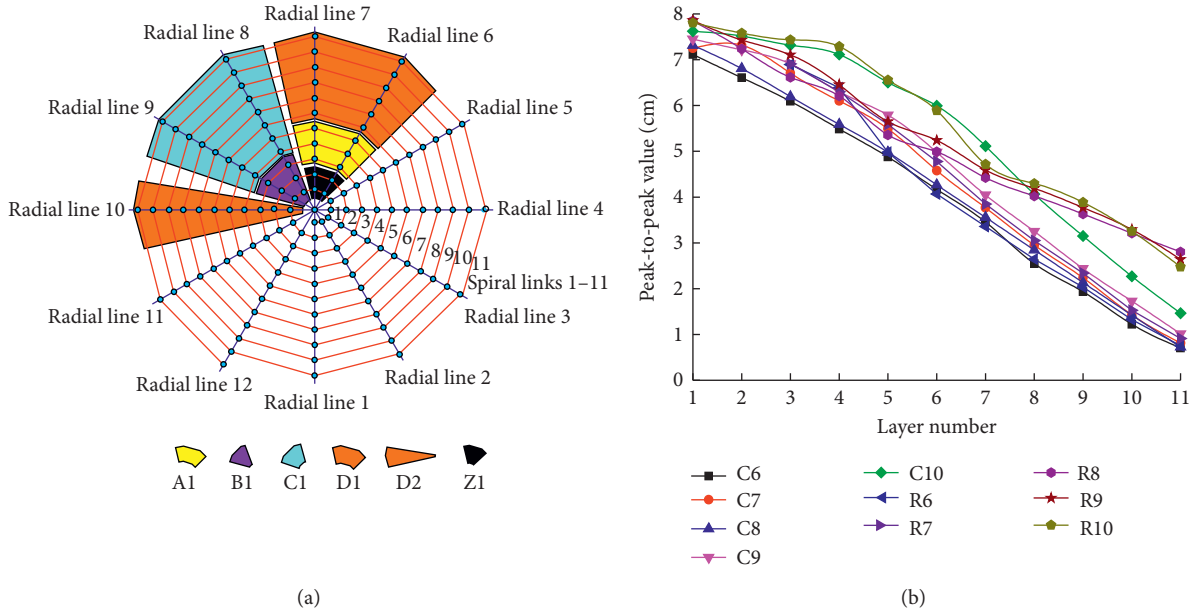


(b)

Figure 1: Artificial spider-web vibration testbed.

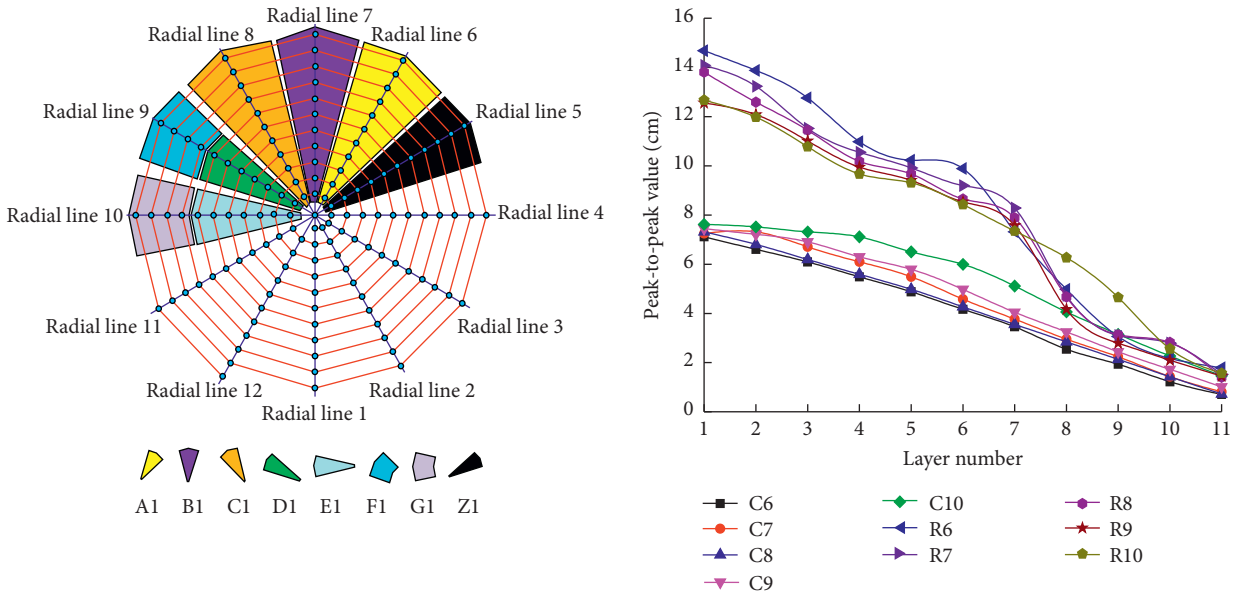
radial lines 5–6, and radial lines 5–7, respectively. As can be seen from Figures 3(a) and 3(b), with the elimination of radial line 5, the node vibrations on radial lines 6–10 dramatically enhance. In areas $A1$, $B1$, $C1$, D1, and $E1$, the peak-to-peak values rise to 2.0, 1.9, 1.8, 1.6, and 1.4 times of that in the complete web on average, respectively. For areas $F1$ and $G1$, the values grow to 1.3 and 1.2 times of that in the complete web on average, respectively. It can be concluded that the closer to the failure positions, the more intense the vibrations of nodes on the radial lines, and vice versa. In Figures 3(c) and 3(d), with radial lines 5–6 being destroyed, the increments of peak-to-peak values of nodes on radial lines 7–11 are observable. Compared with the corresponding positions of the complete web, the average peak-to-peak values increase by 1.2, 1.0, 0.8, 0.6, and 0.6 times in areas $A1$, $B1$, $C1$, $D1$, and E1, raise by 1.4, 1.2, 0.4, and 0.3 times in areas $B2$, $C2$, $D2$, and $E2$, and boost by 1.1 and 0.8 times in areas $D3$ and $E3$. It can be found that, with the increase in damaged radial lines, the coupling relationship between layers and adjacent radial lines causes the cascading spread of the radial-line failure effect. As shown in Figures 3(e) and 3(f), when the number of damaged radial lines jumps to 3, the average peak-to-peak values of nodes in areas $A1$, $B1$, $C1$, $D1$, and $E1$ enhance by 1.31, 1.07, 0.80, 0.61, and 0.57 times and that in areas $A2$, $B2$, $C2$, $D2$, and $E2$ grow by 2.30, 1.83, 0.60, 0.33, and 0.31 times, respectively. It can be recognized that the nodes of layers 1–5 with dense deployment mutually undertake the cascading propagation generated by radial line damages, which results in the vibrations lower than that of nodes at layers 6–11 with sparse deployment.

The results show that, with the growth of the number of damaged radial lines, the node vibrations on the remaining radial lines raise and obey the principle of nearby sharing, which is identical to the phenomenon of cascading-failure spread in WSNs. Besides, the difference between the vibration tracking areas increase with the increment of the number of failed radial lines. Furthermore, the increasing deviations of node vibrations in these triangle and trapezoid



(a) (b)

Figure 2: Differences in peak-to-peak values of longitudinal vibrations for nodes on radial lines 6–10 before and after node damages. (a) Analysis area division. *Z*1 denoted the node failure area, and *A*1, *B*1, *C*1, *D*1, and *D*2 expressed a vibration tracking area, respectively. (b) Peak-to-peak values of longitudinal vibrations of the tracked nodes. *C*6-10 depicted the vibrations of 55 nodes on radial lines 6–10 during the complete web, and *R*6-10 represented the vibrations of 51 nodes on radial lines 6–10 after the node failures, respectively.



(a) (b)

Figure 3: Continued.

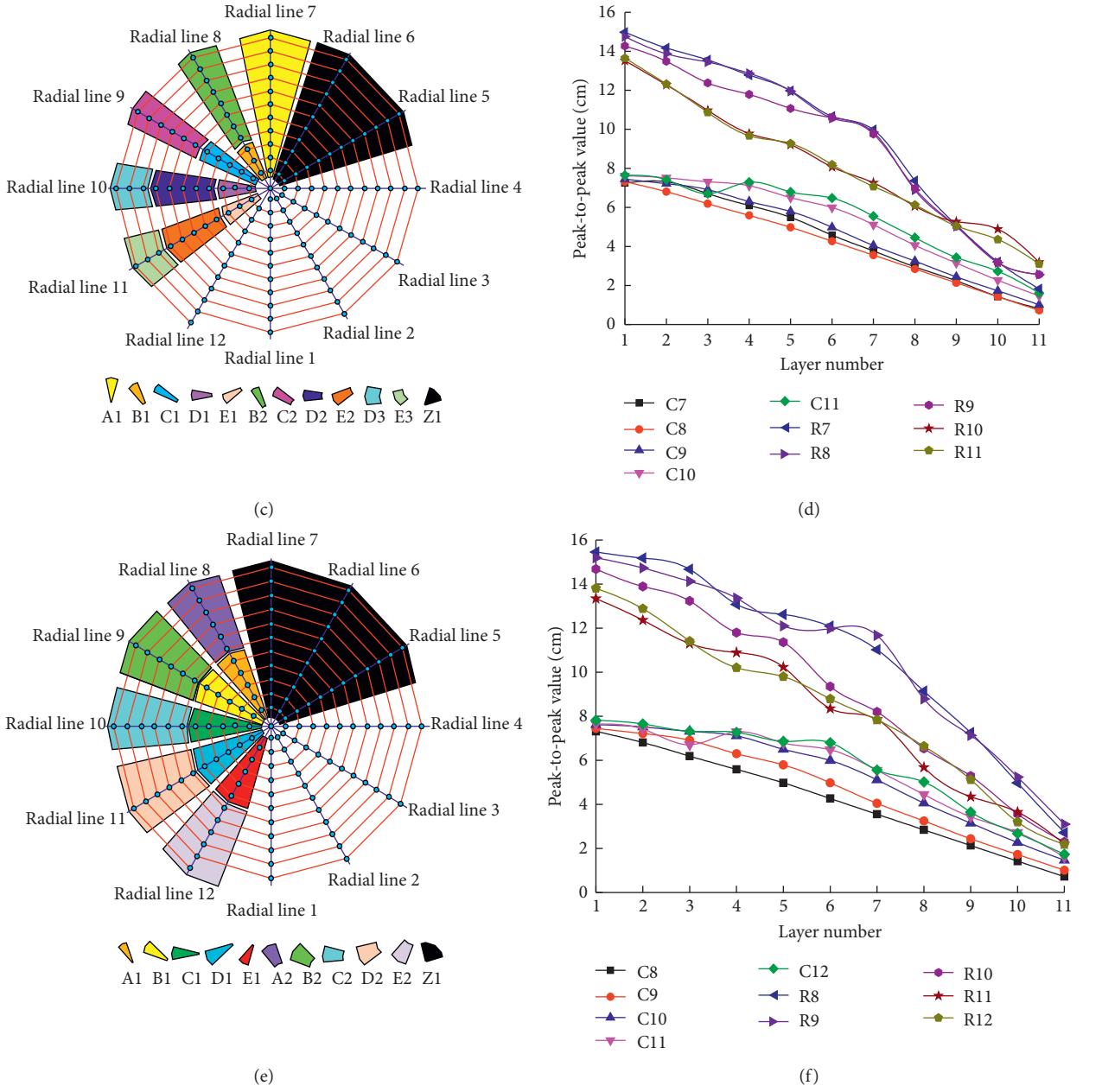


Figure 3: Variations in peak-to-peak values of longitudinal vibrations for nodes on radial lines 6–10, radial lines 7–11, and radial lines 8–12 before and after radial line damages: (a) division of tracking areas at the failure of radial line 5; (b) peak-to-peak values of longitudinal vibrations of the tracked nodes on radial lines 6–10; (c) distribution of tracking areas at the failure of radial lines 5–6; (d) peak-to-peak values of longitudinal vibrations of the tracked nodes on radial lines 7–11; (e) allocation of tracking areas under the damage condition of radial lines 5–7; (f) peak-to-peak values of longitudinal vibrations of the tracked nodes on radial lines 8–12.

areas are comparatively small, which illustrates that the triangle and trapezoid structures can accomplish the hierarchical reduction of vibrations in the inner layers.

2.3. Spiral Line Damage. Figure 4 manifests the differences in peak-to-peak values of longitudinal vibrations for the nodes on radial lines 8-9 under the condition that the spiral lines 1–5 are removed layer by layer. It can be observed that as the spiral lines 1–5 are destroyed one by one, the node vibrations of radial lines 8–9 strengthen gradually. The average peak-to-peak values of nodes on radial line 8 reach 1.10, 1.30, 1.46, 1.56, and 1.70 times of the complete web, respectively. The average peak-to-peak values of nodes on radial line 9 become 1.16, 1.30, 1.39, 1.51, and 1.62 times of the complete web, respectively. It is obvious that the number of broken spiral lines is positively correlated with the vibration intensity of the tracked nodes. Meanwhile, the influence of the damaged spiral lines on the cascading propagation lies between node failures and radial-line

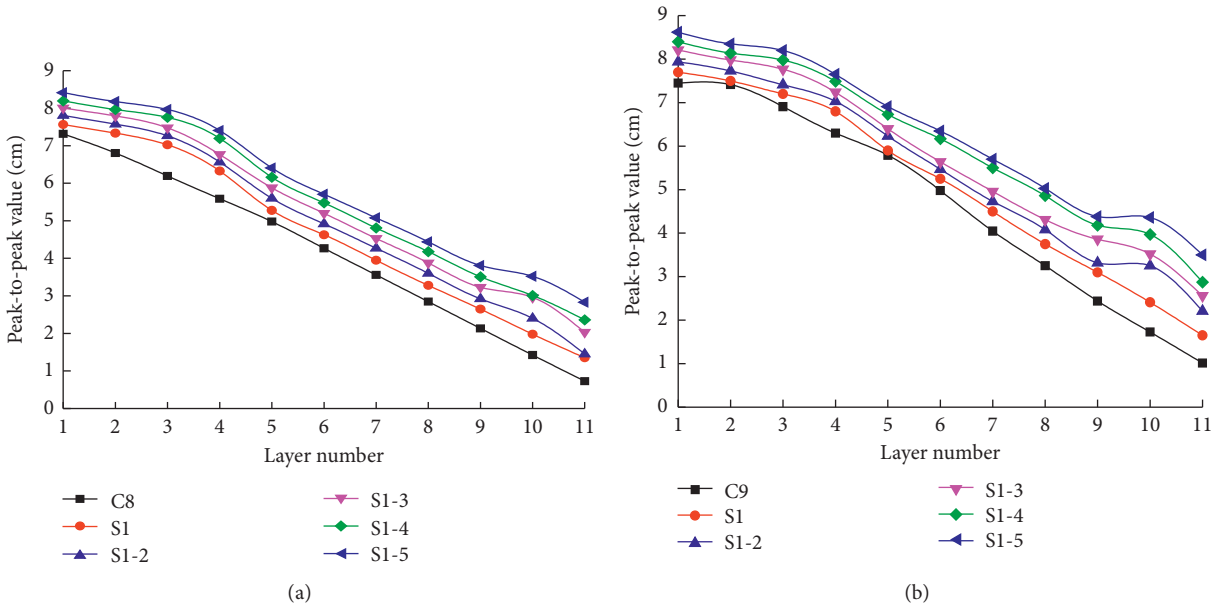


FIGURE 4: Changes in peak-to-peak values of longitudinal vibrations for nodes on radial lines 8–9 under situation of successive failures of spiral lines 1–5: (a) peak-to-peak values of longitudinal vibrations of nodes on radial line 8; (b) peak-to-peak values of longitudinal vibrations of nodes on radial line 9.

failures, indicating that the spiral lines only serve as an aid to the vibration transmission.

By analyzing the influences of different component failures on the vibration distributions, we can summarize the mitigation mechanism of cascade failures in the artificial spider web as follows:

(1) The hierarchical redundancy arrangement of nodes and the internal-external differentiation of node densities can share the burden of inner-layer nodes to cope with the cascading spread.

(2) In the case of node and path damages, the peripheral nodes and paths will carry the vast majority of vibration variations, and the topological structures of triangles and trapezoids can reduce the cascading diffusion effect by using the hierarchical weakening process.

(3) Except for the central node and the outermost nodes, all the other nodes simultaneously connect with the four ambient nodes through radial paths and spiral paths, which has distinct feature of link backup. In addition, radial paths shoulder more transmission responsibilities than spiral paths.

Therefore, on the basis of the inspirations of artificial spider-web topology against cascading failures (topological structures, node deployment mode, and link backup mechanism), combining with the characteristics of farmland WSNs, we should be able to develop a bionic network scheme to alleviate the cascading effect for potential network attacks. Furthermore, the artificial spider-web topology has been proven to be more invulnerable than other traditional topologies in harsh network environments, with the possibility for improving the capability of farmland WSNs against cascading failures [27].

3. Network Scheme for Mitigating the Cascading Failures

In this section, through characterizing the mitigation process of cascading spread in artificial spider web, we describe how to establish a comprehensive network scheme for cascade control guidance of farmland WSNs by assembling networking methodology, communication rules, and load capacity model.

3.1. Networking Methodology. To discuss the network formation procedure, the assumptions are presented as follows:

(1) Each node in the network has a unique physical coordinate, the physical coordinates of the sink node are set to (0, 0), and the physical coordinates of other nodes are determined by the communication with the sink node.

(2) The sink node has the capacity to communicate throughout the network, and other nodes barely interact with nodes in the adjacent areas.

(3) The nodes are in gradient distribution according to the distance from the sink node, and the deployment density gradually increases from the inside out to ensure network coverage. In addition, no isolated nodes exist in the network, and every communication link is symmetrical.

(4) The network adopts carrier monitoring multiple access with collision avoidance (CSMA/CA) protocol to prevent channel conflicts during data transmission [28].

In this premise, we propose the networking methodology of farmland WSNs to perform the hierarchical clustering

routing protocol and topological frame structures composed of triangle and trapezoid units, which are motivated by spider web. The detailed networking steps are listed as follows:

(1) The layer number of the sink node is initialed as 0. The sink node broadcasts the networking packets to the whole network for discovering child nodes. All nodes that have received packets are regarded as child nodes of the sink node and automatically estimate individual physical coordinates by measuring the distances and orientations from the sink node [29]. Afterwards, the physical coordinates of each child node are added into the corresponding routing table (Table 1).

(2) Each child node acquires its layer number by Euclidean distance from the sink node and stores the layer number in the routing table. The layer number coding of child node positively correlates with the distance from the sink node, and the interlayer spacing division can be regulated to balance the number of nodes in each layer. Followed by the assignments of layer numbers, the sink node is indicated as N_0. Besides, the nodes of the nth layer uniformly are coded in the clockwise direction according to orientations and expressed as N_{n-1}, N_{n-2}, N_{n-3}, etc.

(3) Once the layer numbers and physical coordinates of the whole-network nodes are fixed, the networking frames of sink node are sent out layer by layer starting from the child nodes in the first layer. Apparently, the nodes not only transmit the networking frames to the neighboring child nodes in different layers but also receive the frames from other child nodes in the same layer. By comparing its layer number with the layer number of source nodes, each child node receiving the network frames can be classified into three categories as follows:

   (a) The two-layer numbers were equal, it means that the sender and the receiver are in the same layer, and the distance between the two child nodes is recorded. If the networking process of the first layer is accomplished, the two senders closest to each receiver will be registered as nodes of transverse links in the routing table of receiver.

   (b) The layer number of the receiver is less than that of the sender; it indicates that the receiver belongs to the upper layer, and then. the distance between the two nodes is noted. At the end of the networking process of the first layer, the node with the smallest distance from the sender among all the receiving nodes is designated in the routing table as the radial-link node of the sender.

   (c) The number of layers of the receivers is higher than that of the senders, and no operation will be done.

(4) Through iterating step (3), all the layers can finish the networking processes in turn, and a topological framework modeled from spider web is developed (Figure 5).

Table 1: Routing table for child node.

Index labels	Categories	Lengths
1	Layer number	8
2	Physical coordinates	32
3	Node of radial link	8
4	Node 1 of transverse link	8
5	Node 2 of transverse link	8

As shown in Figure 5, the farmland WSNs topology is formed of numerous topological units of triangles and trapezoids. The basic triangle structure is constituted of the child node in the first layer and its transverse-link node, the sink node, and the corresponding links. The fundamental trapezoid structure is included of any two adjacent-layer nodes with a radial link, their transverse-link nodes, and related links. It can be perceived that the triangle and trapezoid structures in the same area have interlayer coupling and intersecting coupling relations, and the radial-link coupling relationship exists between adjacent triangle or trapezoid structures.

3.2. Communication Rules. The link backup and differential distribution of load are the core methods of artificial spider-web topology to overcome cascading failures. Thus, the communication rules of the farmland WSNs based on the invulnerability characteristics of spider web are defined as follows:

(1) In the network, a nonsink node connects with three neighboring nodes by transmission logic, including a radial-link node in the upper layer and two transverse-link nodes in the same layer. The transfer probability of radial link and transverse links is defined as α and β ($1 \geq \alpha > \beta > 0$ and $\alpha + 2\beta = 1$), respectively, and that of a failure link is set as 0.

(2) If the node traffic exceeds its load capacity, the node is assessed to be a failure node, and the associated outer node transmits data to the sink node through its transverse-link nodes.

(3) In the occurrence of a node failure, all the links connect with it failed, and data receiving and delivering cannot be carried out. The function of the failed node is assumed by its transverse-link nodes. Assuming that the communication reliability between a certain node and the three linked nodes is λ ($0 \leq \lambda \leq 1$), the transfer probabilities and transmission success rates of the node under different circumstances are assigned in Table 2.

(4) At the time that the ratio of the failure nodes to the whole-network nodes surpasses the upper limit of node failure rate γ, the sink node will rebroadcast the networking request.

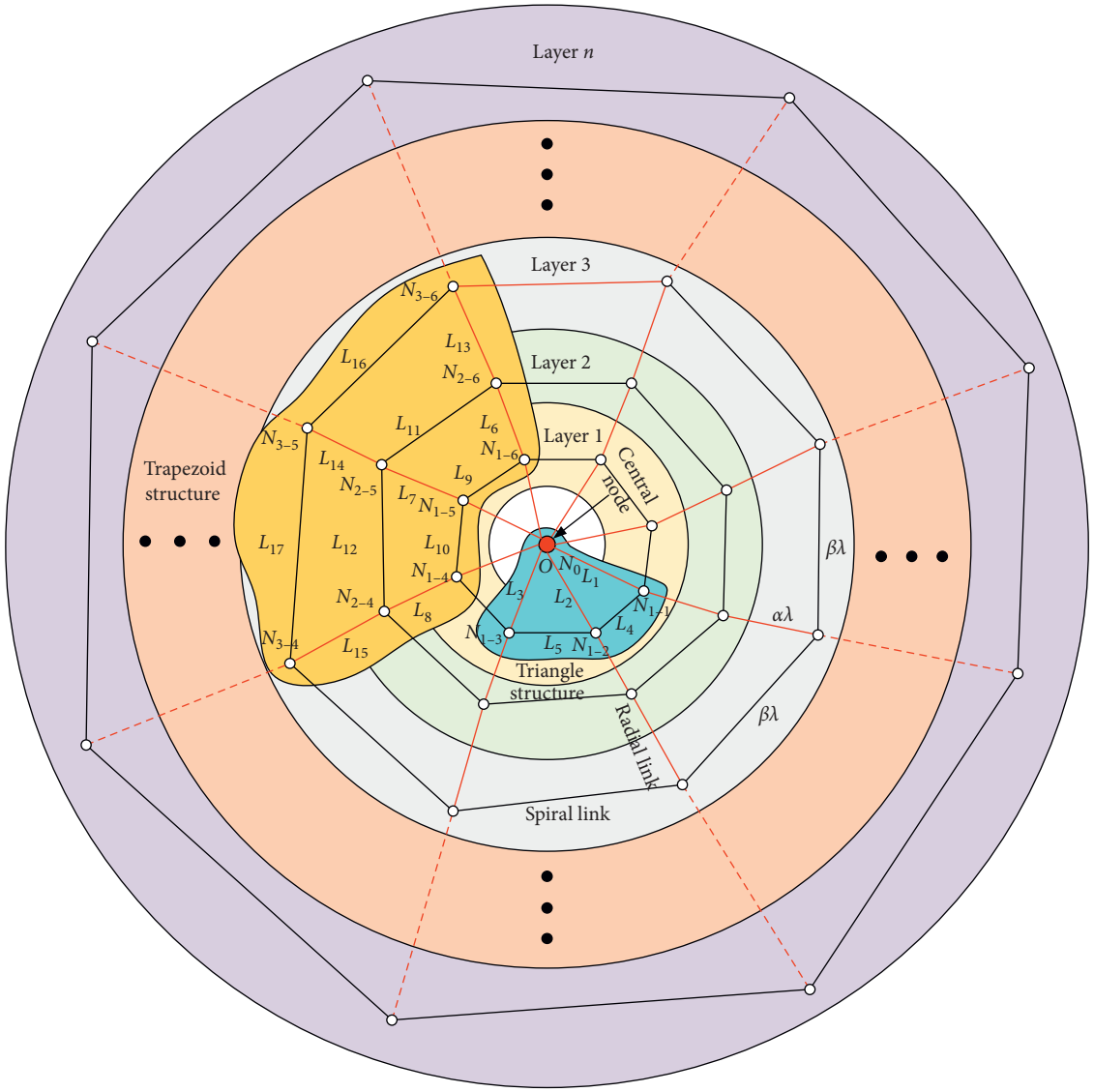


Figure 5: Bionic networking for farmland WSNs based on artificial spider-web topology.

Table 2: Node communication settings in various conditions.

Categories	Descriptions	Transfer probability	Transmission success rate
Normal condition	Trouble-free operation of nodes and links	$\alpha + 2\beta = 1$	$\alpha\lambda + 2\beta\lambda$
Node failures	Failure of radial-link node	$2\beta = 1$	$2\beta\lambda$
	Failures of both transverse-link nodes	$\alpha = 1$	$\alpha\lambda$
	Failures of radial-link node and a transverse-link node	$\beta = 1$	$\beta\lambda$
	Failures of radial-link node and transverse-link nodes	0	0
Link failures	Failure of radial link	$2\beta = 1$	$2\beta\lambda$
	Failures of both transverse links	$\alpha = 1$	$\alpha\lambda$
	Failures of radial-link and a transverse link	$\beta = 1$	$\beta\lambda$
	Failures of radial link and transverse links	0	0

Furthermore, we select a triangle structure and a trapezoid structure to analyze the cascading-failure mitigation means of the particular topological structures stimulated by artificial spider-web topology.

Figure 6 shows the triangle structure comprised of nodes N_0, N_{1-1}-N_{1-3}, radial links L_1-L_3, and transverse links L_4-L_5. In traditional point-to-point mode, the data transmission between nodes is entirely dependent on the link between them. In contrast, the radial and transverse links of nodes in the triangle structure can both participate in data transfer, and the link backup improves the invulnerability of the network. For instance, the transmission success rate between

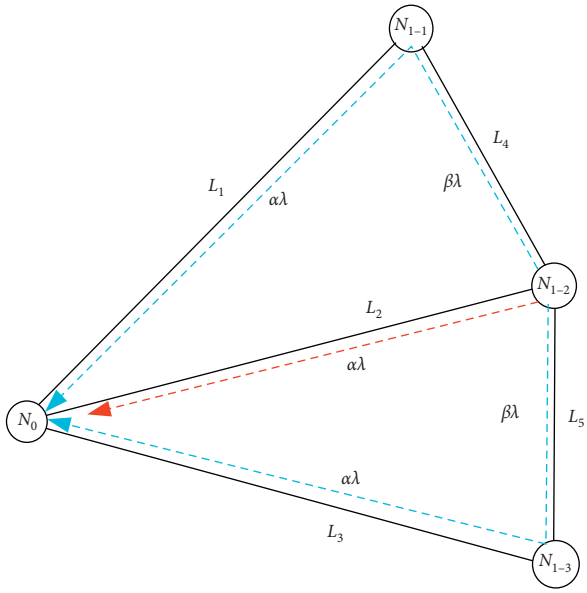


Figure 6: Triangle structure.

$N_{1\text{-}2}$ and N_0 is $\alpha\lambda + 2\beta\alpha\lambda$ under normal case, while in the situation of radial link L_2 damaged, N_{1-2} can transfer data to N_0 through the relay nodes N_{1-1} and N_{1-3}, and the transmission success rate becomes $2\beta\alpha\lambda$.

Figure 7 displays the trapezoid structure composed of nodes N_{1-4}-N_{1-6}, N_{2-4}-N_{2-6}, and N_{3-4}-N_{3-6} and links L_6-L_{17}. The transmission success rates of $N_{3-5} \longrightarrow N_{2-5}$, $N_{2-5} \longrightarrow N_{1-5}$, and $N_{3-5} \longrightarrow N_{1-5}$ are computed in state of the normal condition, L_{14} failed, N_{2-5} damaged, L_7 disabled, and N_{1-5} corrupted, respectively.

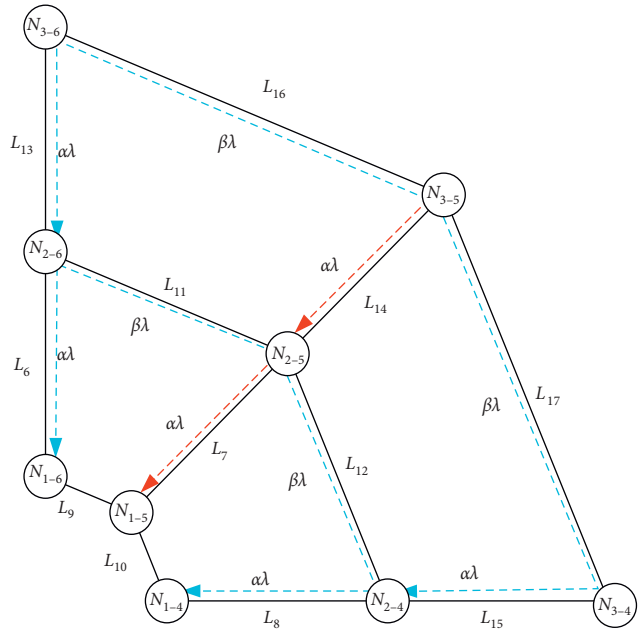


Figure 7: Trapezoid structure.

It can be seen that the failure of nodes or links at different positions on the radial link will not cause the communication interruption depending on the trapezoidal structure. The analysis indicates that the trapezoid structure can lessen the cascading-failure effect of failed nodes or links to guarantee data transmission by utilizing link backup.

Normal condition:

$$\begin{cases} P(N_{3-5} \longrightarrow N_{2-5}) = \alpha \cdot \lambda + 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{2-5} \longrightarrow N_{1-5}) = \alpha \cdot \lambda + 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{3-5} \longrightarrow N_{1-5}) = \alpha^2 \cdot \lambda^2 + \alpha \cdot \lambda \cdot (\alpha \cdot \lambda + 2\beta \cdot \alpha \cdot \lambda^2) + 2\alpha \cdot \beta \cdot \lambda^2(\alpha \cdot \lambda + \beta \cdot \alpha \cdot \lambda^2). \end{cases} \tag{1}$$

L_{14} failed:

$$\begin{cases} P(N_{3-5} \longrightarrow N_{2-5}) = 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{2-5} \longrightarrow N_{1-5}) = 2\alpha \cdot \lambda + 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{3-5} \longrightarrow N_{1-5}) = 4\beta \cdot \alpha \cdot \lambda^2 \cdot (\alpha \cdot \lambda + \alpha \cdot \beta \cdot \lambda^2). \end{cases} \tag{2}$$

$N_{2\text{-}5}$ damaged:

$$\begin{cases} P(N_{3-5} \longrightarrow N_{2-4} \text{ and } N_{2-6}) = 2\alpha \cdot \beta \cdot \lambda^2, \\ P(N_{2-4} \text{ and } N_{2-6} \longrightarrow N_{1-5}) = 2\alpha \cdot \lambda, \\ P(N_{3-5} \longrightarrow N_{1-5}) = 2\beta \cdot \alpha^2 \cdot \lambda^3. \end{cases} \tag{3}$$

L_7 disabled:

$$\begin{cases} P(N_{3-5} \longrightarrow N_{2-5}) = \alpha \cdot \lambda + 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{2-5} \longrightarrow N_{1-5}) = 2\beta \cdot \alpha \cdot \lambda^2, \\ P(N_{3-5} \longrightarrow N_{1-5}) = 4\beta \cdot \alpha^2 \cdot \lambda^3. \end{cases} \tag{4}$$

N_{1-5} corrupted:

$$\begin{cases} P(N_{3-5} \longrightarrow N_{2-5}) = \alpha \cdot \lambda + 2\alpha \cdot \beta \cdot \lambda^2, \\ P(N_{2-5} \longrightarrow N_{1-4} \text{ and } N_{1-6}) = 2\alpha \cdot \beta \cdot \lambda^2, \\ P(N_{3-5} \longrightarrow N_{1-4} \text{ and } N_{1-6}) = 4\alpha^2 \cdot \beta \cdot \lambda^3. \end{cases} \tag{5}$$

3.3. Load Capacity Model. The hierarchical clustering of farmland WSNs makes the nodes closer to the sink node undertake more communication tasks, and preventing the premature failures of these nodes is the premise of network invulnerability. Due to the gradient distribution of nodes, the inner layers have fewer nodes than the outer layers, and

the inner-layer nodes with higher load capacities can offset the negative impact of more extra traffic. Therefore, based on the classical load capacity model and the topological properties of constructed farmland WSNs, we address the load capacity model to optimize the distribution of traffic and promote the network capabilities against cascade failures in this section.

The load capacity of a single node is described as follows:

$$C_n = \left(1 + \frac{\varepsilon}{n}\right) L_n, \quad n = 1, 2, \ldots, H, \tag{6}$$

where C_n is the capacity of a single node in the nth layer, L_n is the initial load of a single node of the nth layer, ε is the adjustment coefficient used to regulate the node capacity value, n represents the layer number, and H denotes the total number of layers in the network.

The initial load L_n of a single node in the nth layer is computed by

$$L_n = \frac{S_{n+1}}{T_n}, \tag{7}$$

where L_n is the initial load of a single node at the nth layer, S_{n+1} is the total initial load of all nodes at the $n+1$th layer, and T_n signifies the total number of nodes in the nth layer.

Hence, the load capacity model can be defined as

$$C_n = \left(1 + \frac{\varepsilon}{n}\right) \cdot \frac{S_{n+1}}{T_n}, \quad n = 1, 2, \ldots, H. \tag{8}$$

4. Analysis on the Invulnerability of Farmland WSNs

4.1. Simulation Settings. In the simulations, the number of layers of farmland WSNs is set to 5, the interval distance between layers is 15 m, and the sink node is placed at the center of the simulation area. A total of 200 nodes are distributed in the simulation area. We configure the initial traffic of each node to be 1 unit and the communication reliability λ to be 0.8. The transfer probability of radial link α and the transfer probability of transverse link β are initially set to 0.6 and 0.2, respectively. Each node in the initial network is in a normal state before attacks. In our model, we assume that cascading failures are triggered by the faulty nodes caused by overload of load capacity. All the simulation results are the average of 50 simulations.

To assess the performance of mitigation of cascading failures for the built farmland WSNs, the number of simulation rounds and network efficiency ratio are adopted as evaluation metrics:

(1) A simulation round is defined as the period for all normal nodes to complete data transmissions to the sink node. The number of simulation rounds while the termination condition of node failure rate is satisfied can be expressed as R.

(2) Network efficiency ratio is applied to measure the damage degree of cascading failures and can be achieved by

$$M = \frac{E_{\text{normal}}}{E_{\text{initial}}} = \frac{\left(1/N_{\text{normal}}\left(N_{\text{normal}} - 1\right)\right)\sum_{i \neq j} 1/d'_{ij}}{\left(1/N_{\text{initial}}\left(N_{\text{initial}} - 1\right)\right)\sum_{i \neq j} 1/d_{ij}}, \tag{9}$$

where E_{initial} and E_{normal} represent the network efficiency before and after network damage, respectively; N_{normal} is the number of normal nodes, and N_{initial} is the total number of nodes in the initial network; and d_{ij} and d'_{ij} denote the number of hops along the shortest path from node i to node j in the initial network and the faulty network, respectively.

4.2. Invulnerability for Random Attacks. Figure 8 depicts the number of simulation rounds R and network efficiency ratio M with varying adjustment coefficient ε under different levels of node failure rate γ. From Figure 8(a), It can be observed that, with the growth of γ from 0.1 to 0.3, R raises less with ε rising from 0 to 40 (average number of nodes per layer) but more with γ increasing from 0.4 to 0.9. In particular, when γ increases from 0.7 to 0.9, R rises 23.2, 26.8, and 37.5 rounds, respectively. Moreover, when ε is in the range of 25 to 40, the trend of R growing with ε is more obvious. Therefore, to weak the impact of cascading failures, the designer should choose ε as large as possible. Meanwhile, as ε stays the same and γ grows, the increments in R increase gradually. The performance of inhibiting the spread of cascading failures may benefit from the spiderweb-like topology and communication rules. As is shown in Figure 8(b), when γ is within the range [0.1, 0.6], with the growth of ε from 0 to 40, M increases by 11.94%, 16.35%, 12.31%, 23.93%, 20.19%, and 27.74%, respectively. M is comparatively small, while γ varies from 0.7 to 0.9, but the rise is up to 53.5%. The results suggest that the farmland WSNs with a higher ε can lower the damages caused by cascading failures.

Figure 9 demonstrates the relationship between R and M, in the case that $\gamma = 0.8$. It can be noticed that M continually decreases with the increase of R, indicating that M is inversely proportional to R. For different ε, when R raises from 0 to 50, M rapidly declines from 0.8 to about 0.1, and the smaller the ε is, the more notable M reduces. Taking $R = 50$ as an example, with the growth of ε from 0 to 40, M decreases by 97.29%, 89.48%, 88.37%, 85.49%, 85.54%, 85.67%, 85.87%, 84.19%, and 84.00%, respectively. Thus, ε has the function of regulating M. However, as R continues to go up from 50, M drops slowly. It can be concluded that, after a certain degree of damage generated by cascading failures, the network scheme for farmland WSNs can improve R with trifling effect on M.

4.3. Invulnerability for Deliberate Attacks. Figures 10(a)–10(j) show R and M with varying ε after the failures of the same number of nodes from the 1st layer to the 5th layer caused by deliberate attacks when $\gamma = 0.8$, respectively. As shown in Figures 10(a), 10(c), 10(e), 10(g), and 10(i), it can be easily discovered that when ε is fixed, R declines with the increase in the number of failure nodes for each layer. For instance, when $\varepsilon = 10$, the failure nodes in the 3r d layer grow from 1 to 11, and R is 69.8, 67, 61, 60.4, 57.4,

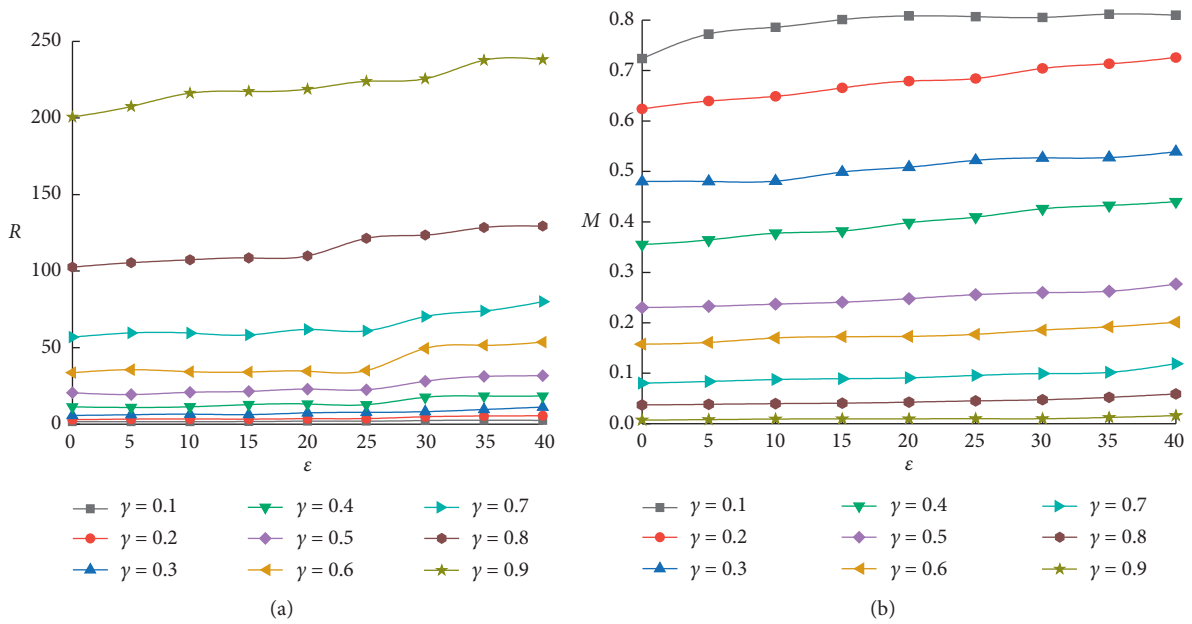


Figure 8: Evaluation metrics with varying ε for random attacks: (a) number of simulation rounds with varying ε; (b) network efficiency ratio with varying ε.

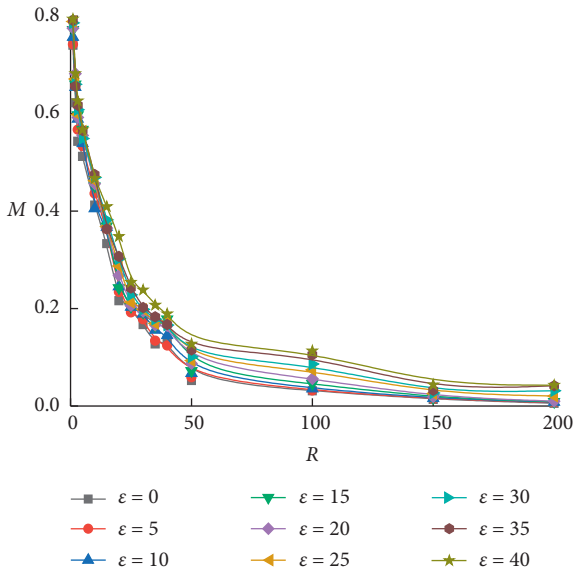


Figure 9: Relevance between R and M for random failure ($\gamma = 0.8$).

and 56.6, respectively. With the growth of ε, R has an increasing trend under the number of failure nodes is the same. For the 1st layer, in the case that the number of failure nodes is 5, R with ε rising from 0 to 40 is 66.6, 66.8, 69.4, 68.2, 70, 70.2, 80.4, 86.2, and 98.8, respectively. The results infer that the farmland WSNs adopting the network scheme can mitigate the cascading failures in the event of node failures caused by targeted attacks. Furthermore, it can be also observed that when ε and the number of failure nodes are maintained the same, R reduces more drastically in the inner layers than that in the outer layers. The result shows that ε has a greater influence on the inner layers by ensuring the load capacity of inner-layer nodes in the spiderweb-like hierarchical topology, and the effect of outer-layer nodes on network invulnerability is less. Besides, ε is within the range (30, 40), and R is improved significantly for the different numbers of failure nodes in each layer. Taking the 2nd layer as an instance, when ε is in this range, the average increment of R for each number of failure nodes reaches 16.26%, 24.01%, and 32.04%, respectively.

From Figures 10(b), 10(d), 10(f), 10(h) and 10(j), we can note that, under the identical ε, with the rise of the number of failure nodes, the fluctuation of M is insignificant, which is the result of the link backup mechanism in the farmland WSNs to assure network connectivity at small hop costs. Moreover, it can be seen that M boosts with the increase in ε for each layer under the circumstance that the number of failure nodes stays the same. For example, when the number of failure nodes in the 5th layer is 7 and ε raises from 0 to 40, M is 0.035, 0.037, 0.046, 0.063, 0.072, 0.086, 0.113, 0.116, and 0.135, respectively. The performance of cascading-failure reduction may originate from the balance of load and traffic of nodes with the continuous growth of ε. Meanwhile, with the same ε and the same number of failure nodes, M in the inner layers is smaller than that in the outer layers. Due to the more limited influence of edge-layer nodes on the shortest paths between nodes, the results are contrary to the variation characteristics of R under the same condition and consistent with the inversely proportional relationship between R and M in the random-failure simulations.

To distinguish the difference in cascading invulnerability between different degrees of deliberate attacks with varying ε, when $\gamma = 0.8$, we investigate the influence of ε after the failure of 2, 4, 6, 8, and 10 nodes and 10, 8, 6, 4, and 2 nodes from the 1st layer to the 5th layer on the evaluation metrics,

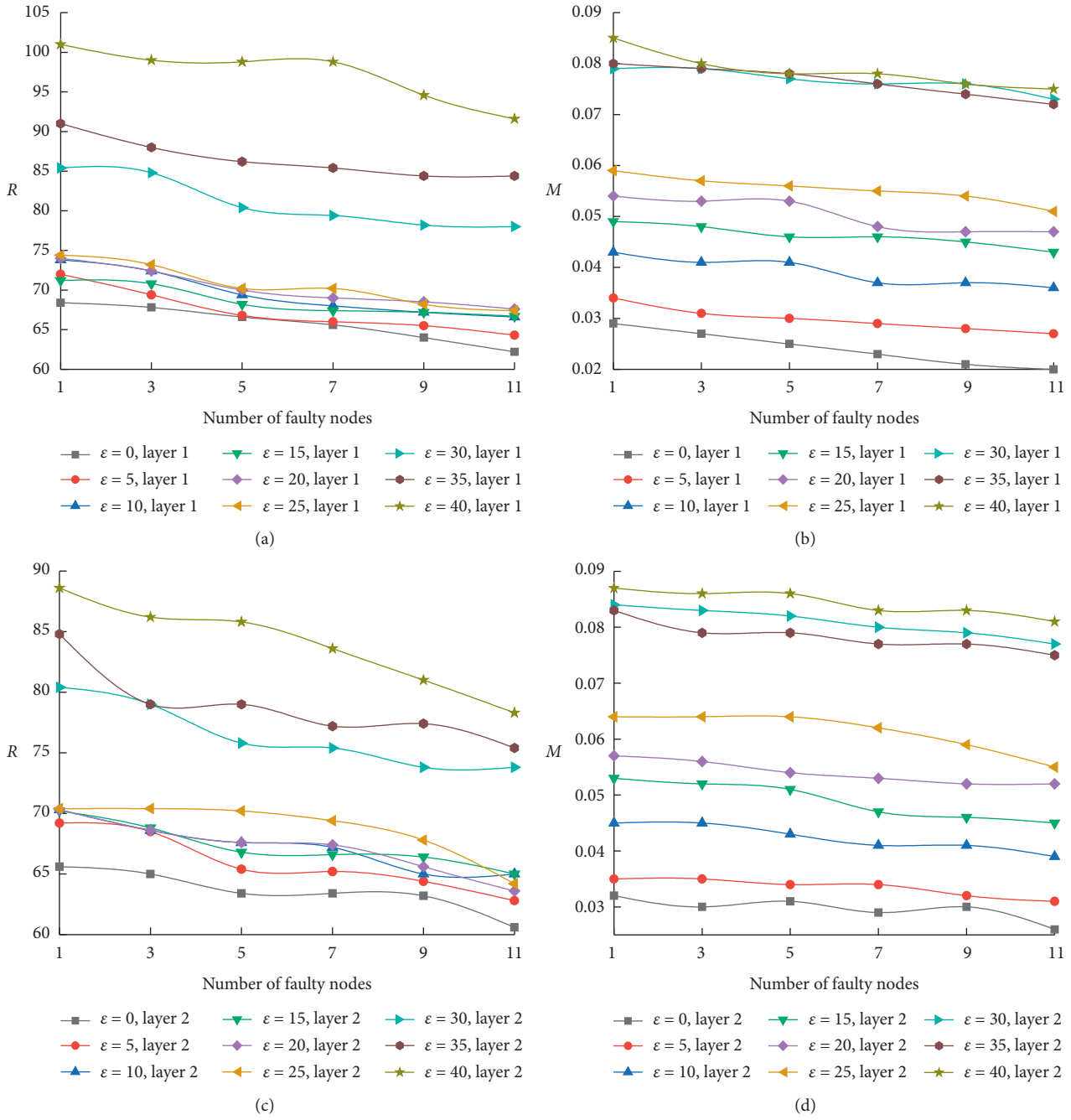


Figure 10: Continued.

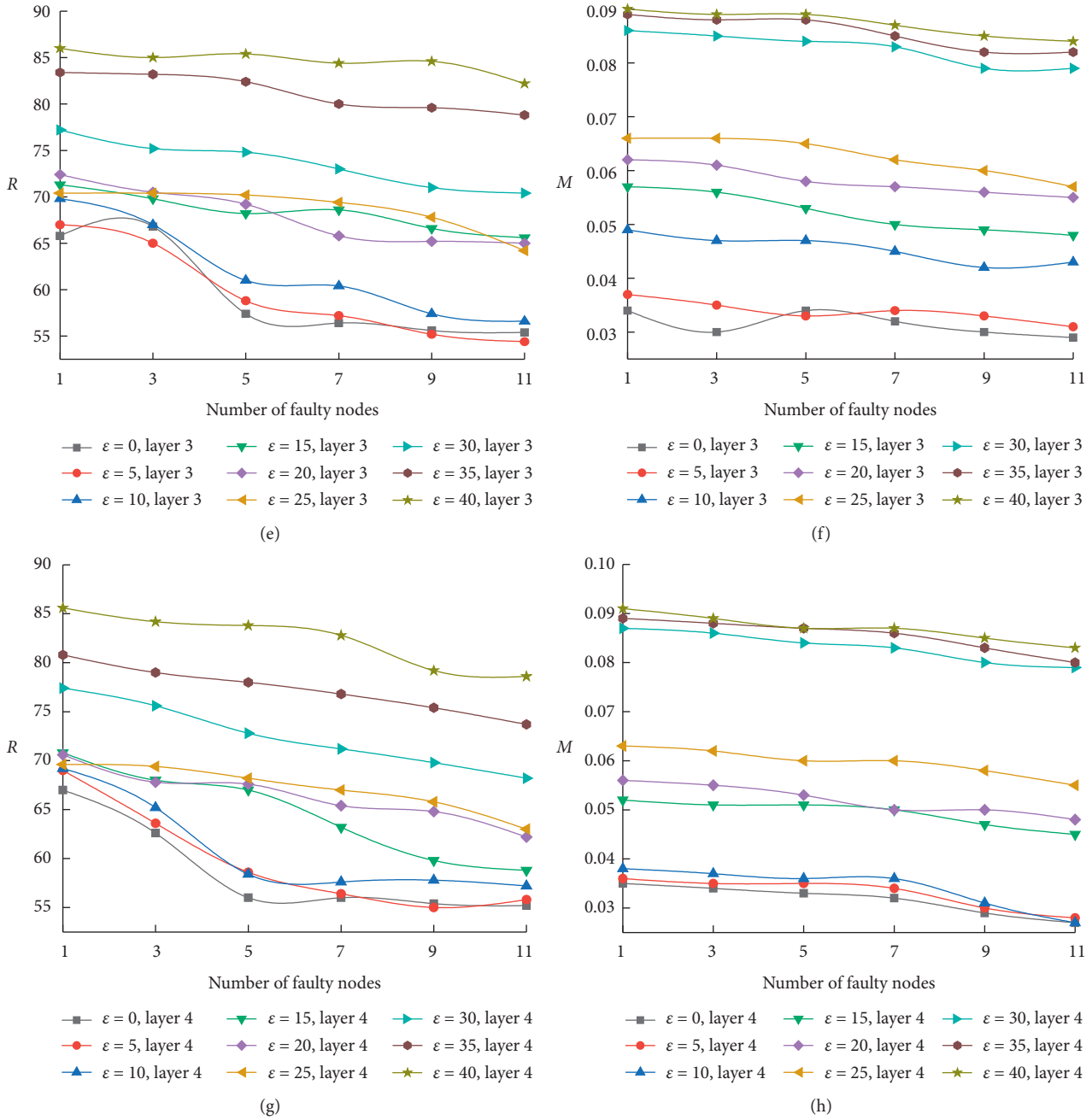


Figure 10: Continued.

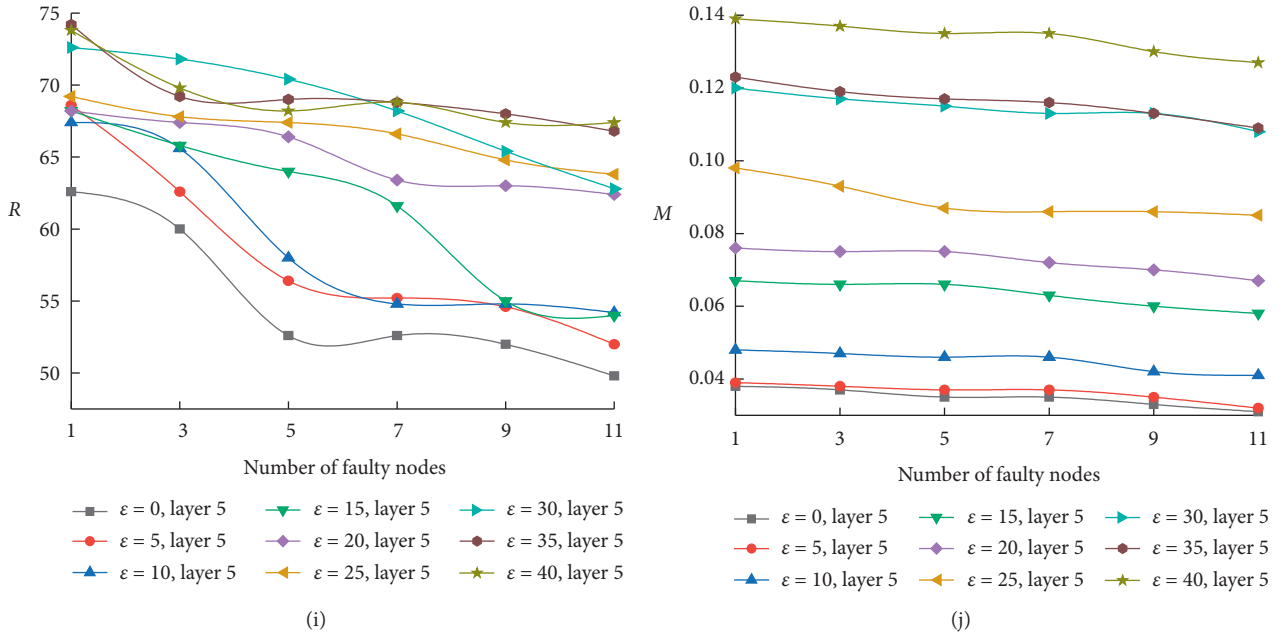


Figure 10: Evaluation metrics with varying ε for the failures of the same number of nodes in each layer ($\gamma = 0.8$): (a) R with varying ε in the 1st layer; (b) M with varying ε in the 1st layer; (c) R with varying ε in the 2nd layer; (d) M with varying ε in the 2nd layer; (e) R with varying ε in the 3rd layer; (f) M with varying ε in the 3rd layer; (g) R with varying ε in the 4th layer; (h) M with varying ε in the 4th layer; (i) R with varying ε in the 5th layer; (j) M with varying ε in the 5th layer.

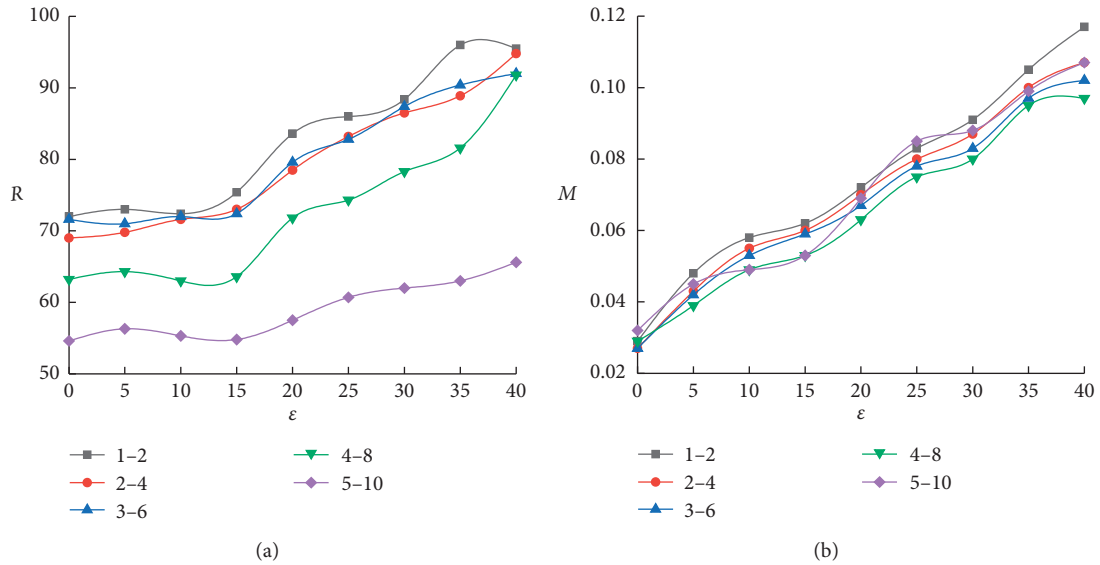


Figure 11: Continued.

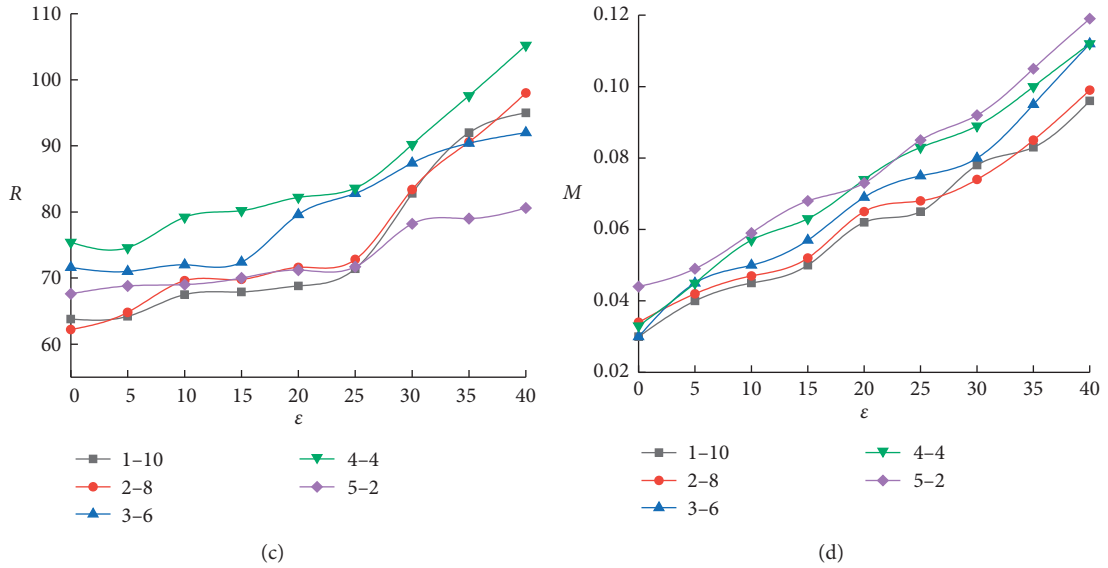


FIGURE 11: Comparison of ε on the evaluation metrics for different degrees of deliberate attacks ($\gamma = 0.8$): (a) influence of ε on R after the incremental failures; (b) influence of ε on M after the incremental failures; (c) influence of ε on R after the decreasing failures; (d) influence of ε on M after the decreasing failures.

respectively. As shown in Figures 11(a) and 11(c), R demonstrates an upward trend with the increase in ε for the different number of failure nodes in each layer. When the number of failure nodes rises from the inner layers to the outer layers, compared with that of $\varepsilon = 0$, R under the case of $\varepsilon = 40$ increases by 32.63%, 37.39%, 28.49%, 45.25%, and 20.14%, respectively. While the number of failure nodes reduces from the inner layers to the outer layers, compared with that of $\varepsilon = 0$, R under the circumstance of $\varepsilon = 40$ grows by 57.66%, 57.55%, 28.49%, 39.52%, and 19.23%, respectively. It can be discovered that, with the increase in the number of failure nodes in the 1st and 2nd layers, R raises markedly with the growth of ε. The results manifest that the more the number of failure nodes in the inner layers, the more the R is affected by ε, which is the result of rising the load capacity of the normal nodes in the inner layers. Nevertheless, for the outermost layer, R is lightly affected by ε, and the number of failure nodes slightly impacts the effect of ε on R. Our data imply that the greater ε in the face of severe deliberate attacks will promote the network invulnerability. From Figures 11(b) and 11(d), when the number of failure nodes increases from the inside out, in the case that $\varepsilon = 40$, M is 3.03, 2.96, 2.77, 2.34, and 2.34 times higher than that of $\varepsilon = 0$, respectively. In the condition that the number of failure nodes lessens from inside to outside, compared with the case of $\varepsilon = 0$, M of each layer when $\varepsilon = 40$ enhances by 2.20, 1.91, 2.77, 2.39, and 1.70 times, respectively. Therefore, for the two modes of deliberate attacks, M presents a remarkable positive correlation with the increase in ε and negatively correlates with the number of failure nodes. Moreover, the detection method of deliberate attack should be added in the future to avoid the catastrophic cascading damages induced by a large number of failure nodes.

5. Conclusions

Aiming to relieve cascading failures in farmland WSNs, we introduce a bionic research method in this study. On the basis of the destructive vibration testing, the mitigation mechanism of cascade failures of artificial spider web including topological structures, node deployment mode, and link backup is analyzed theoretically. The network scheme consisting of networking methodology, communication rules, and load capacity model is proposed based on the combination of the characteristics of farmland WSNs and the inspirations of artificial spider web. We verified the performance of the developed network scheme by using the number of simulation rounds and network efficiency ratio as evaluation metrics. The results showed that the network scheme achieved the ideal invulnerability for both random and deliberate attacks. The adjustment coefficient in the load capacity model was confirmed for the regulatory effect on network invulnerability. Furthermore, the use of the topology and communication rules inspired by spider web should be advised for lowering the damages of cascading failures. The nodes in the inner layers had a greater impact on the propagation of cascading failures. Our results suggest that the bionic network scheme is an effective approach of cascading-failure mitigation for farmland WSNs.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

J.W. conceptualized the study, did formal analysis, gathered resources, performed project administration, and obtained funding acquisition. J.W. and Z.D. formulated methodology and investigated and visualized the study. Z.H. analyzed using software. Z.D. and X.W. validated and supervised the study and reviewed and edited the manuscript. Z.D. and Z.H. curated the data. J.W., Z.D., and X.W. wrote the original draft preparation. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 61771184), Program for Science and Technology Innovation Talents in Universities of Henan Province (Grant no. 20HASTIT029), Program for Innovative Research Team (in Science and Technology) in University of Henan Province (Grant no. 19IRTSTHN021), and Science and Technology Major Project of Henan Province (Grant no. 181100110100).

References

[1] S. Sharma, R. K. Bansal, and S. Bansal, "Issues and challenges in wireless sensor networks," in *Proceedings of the 2013 International Conference on Machine Intelligence and Research Advancement*, pp. 58–62, IEEE, Katra, India, December 2013.

[2] T. Ojha, S. Misra, and N. S. Raghuwanshi, "Wireless sensor networks for agriculture: the state-of-the-art in practice and future challenges," *Computers and Electronics in Agriculture*, vol. 118, pp. 66–84, 2015.

[3] X. Fu, Y. Yang, and O. Postolache, "Invulnerability of clustering wireless sensor networks against cascading failures," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1431–1442, 2019.

[4] A. More and V. Raisinghani, "A node failure and battery-aware coverage protocol for wireless sensor networks," *Computers & Electrical Engineering*, vol. 64, pp. 200–219, 2017.

[5] D. Chen, X. Ren, Q. Zhang, Y. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.

[6] S. Misra, R. Singh, and S. V. R. Mohan, "Information warfare-worthy jamming attack detection mechanism for wireless sensor networks using a fuzzy inference system," *Sensors*, vol. 10, no. 4, pp. 3444–3479, 2010.

[7] R. Ghanbari, M. Jalili, and X. Yu, "Correlation of cascade failures and centrality measures in complex networks," *Future Generation Computer Systems*, vol. 83, pp. 390–400, 2018.

[8] Z. Zhao, "Research on invulnerability of wireless sensor networks based on complex network topology structure," *International Journal of Online Engineering (iJOE)*, vol. 13, no. 3, pp. 100–112, 2017.

[9] R.-R. Yin, B. Liu, H.-R. Liu, and Y.-Q. Li, "Research on invulnerability of the random scale-free network against cascading failure," *Physica A: Statistical Mechanics and Its Applications*, vol. 444, pp. 458–465, 2016.

[10] M. Yuvaraja and M. Sabrigiriraj, "Fault detection and recovery scheme for routing and lifetime enhancement in WSN," *Wireless Networks*, vol. 23, no. 1, pp. 267–277, 2017.

[11] A. Majdandzic, B. Podobnik, S. V. Buldyrev, D. Y. Kenett, S. Havlin, and H. Eugene Stanley, "Spontaneous recovery in dynamical networks," *Nature Physics*, vol. 10, no. 1, pp. 34–38, 2014.

[12] O. Osanaiye, A. Alfa, and G. Hancke, "A statistical approach to detect jamming attacks in wireless sensor networks," *Sensors*, vol. 18, no. 6, Article ID 1691, 2018.

[13] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, no. 6, Article ID 065102, 2002.

[14] Y. Wang, B. Chen, X. Chen, and X. Gao, "Cascading failure model for command and control networks based on an m-order adjacency matrix," *Mobile Information Systems*, vol. 2018, Article ID 6404136, 11 pages, 2018.

[15] H. M. Jawad, R. Nordin, S. K. Gharghan, A. M. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: a review," *Sensors*, vol. 17, Article ID 1781, 2017.

[16] Y. S. Miao, L. Yuan, H. R. Wu, and Q. X. Li, "Optimization of energy heterogeneous cluster-head selection in farmland WSN," *Applied Mechanics and Materials*, vol. 441, pp. 1010–1015, 2013.

[17] F. Huang, C. Zhao, F. Li, and H. Wu, "An improved farmland WSN topology based on YG and clustering algorithm," *Information Technology Journal*, vol. 12, no. 21, pp. 6463–6468, 2013.

[18] F. Huang, C. J. Zhao, F. F. Li, and H. R. Wu, "A WSN routing and sleeping algorithm in agriculture intermittent data collection," *Applied Mechanics and Materials*, vol. 513-517, pp. 606–611, 2014.

[19] H. Wu, H. Zhu, and Y. Miao, "An energy efficient cluster-head rotation and relay node selection scheme for farmland heterogeneous wireless sensor networks," *Wireless Personal Communications*, vol. 101, no. 3, pp. 1639–1655, 2018.

[20] Z. Erasmus, X. Liu, and S. M van Ruth, "Study on the relations between hyperspectral images of bananas a path planning method for mobile sink in farmland wireless sensor network," in *Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference*, pp. 1157–1160, IEEE, Chengdu, China, December 2017.

[21] B.-L. Miao, X.-G. Zhao, S.-Y. Wu, and H. Zhu, "Research on power consumption weighted state machine of farmland WSN node," in *Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, pp. 1141–1144, IEEE, Hangzhou, China, October 2012.

[22] B. Dou, X. Wang, and S. Zhang, "Robustness of networks against cascading failures," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 11, pp. 2310–2317, 2010.

[23] Z. Qin, B. G. Compton, J. A. Lewis, and M. J. Buehler, "Structural optimization of 3D-printed synthetic spider webs for high strength," *Nature Communications*, vol. 6, Article ID 7038, 2015.

[24] A. Canovas, J. Lloret, E. Macias, and A. Suarez, "Web spider defense technique in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 348606, 10 pages, 2014.

[25] R. Zaera, A. Soler, and J. Teus, "Uncovering changes in spider orb-web topology owing to aerodynamic effects," *Journal of the Royal Society Interface*, vol. 11, Article ID 20140484, 2014.

[26] J. Wang, Z. Du, Z. He, and J. Wang, "Exploring vibration transmission rule of an artificial spider web for potential application in invulnerability of wireless sensor network," *Applied Bionics and Biomechanics*, vol. 2019, Article ID 5125034, 18 pages, 2019.

[27] J. Wang, Z. Du, and Z. He, "Quantitative invulnerability analysis of artificial spider-web topology model based on end-to-end delay," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 4617239, 11 pages, 2020.

[28] D. De Guglielmo, B. Al Nahas, S. Duquennoy, T. Voigt, and G. Anastasi, "Analysis and experimental evaluation of IEEE 802.15.4e TSCH CSMA-CA algorithm," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1573–1588, 2017.

[29] Y. Zhang and Y. I. Wu, "Multiple sources localization by the WSN using the direction-of-arrivals classified by the genetic algorithm," *IEEE Access*, vol. 7, pp. 173626–173635, 2019.

Hindawi
Complexity
Volume 2020, Article ID 6535927, 10 pages
https://doi.org/10.1155/2020/6535927

Research Article

Triangulation Applied to the Intra-European Union Tomato Market

Jaime De Pablo Valenciano,[1] **José Antonio Torres Arriaza**,[2] **Juan Uribe-Toril**,[1] **and José Luis Ruiz-Real**[1]

[1]*Faculty of Economics and Business, University of Almeria, Ctra. De Sacramento, s/n, Almería 04120, Spain*
[2]*Faculty of Computer Science, University of Almeria, Ctra. De Sacramento, s/n, Almería 04120, Spain*

Correspondence should be addressed to Jaime De Pablo Valenciano; jdepablo@ual.es

Received 3 March 2020; Accepted 1 May 2020; Published 15 May 2020

Guest Editor: François Pérès

An understanding of the intracommunity trade is essential for the agents involved in the fresh tomato market (farmers, entrepreneurs, public administrations, and consumers). The purpose of this paper is to analyze the interdependent relationships between exporting and importing countries within the European Union for a specific product such as fresh tomatoes and thus understand which have been the key countries in three specific years (2002–2007–2017). The methodology used to study the interrelationships of trade flows in the countries of the European Union (EU) is that of triangulation by means of the Leontief input-output model. Artificial intelligence techniques are used to process and triangulate the data based on pathfinding techniques using a cost function. The triangulation results have created a hierarchy of countries (suppliers and customers). This type of methodology has not been applied to the field of foreign trade. The results show that Netherlands and Spain are key countries in intracommunity trade as they have a strong impact both with regard to their exports and their imports and are fundamental when analyzing the growth of specific sectors and how they are able to stimulate the economies of other countries.

1. Introduction

According to Leontief [1], input-product matrices are a fundamental tool for studying the interrelationships of economic structures and for finding the best way to activate the economy. In particular, there has been a long tradition of using them for establishing hierarchies in economic structures, although their use has proven to be very complex, since it is impossible to arrive at a strictly hierarchical structure in view of the persistence of both direct (binary) and indirect induction cycles (circular interrelations). Therefore, it is necessary to reorganize the branches that form the whole economic structure in order to preserve the hierarchical classification to the extent possible.

Input-out triangulation studies are based on the idea of inter-industrial transactions can be recorded in a matrix between the origin (in our case, exporting countries) and destination (importing countries).

The years analyzed in this study are 2002, 2007, and 2017. The reason for choosing these years is to consider whether the economic and financial crisis that began in 2007 and that affected the economies of the European Union had an impact on the hierarchy of tomato producing countries and their customers with regard to intracommunity trade. It should be borne in mind that there is a close relationship between exporting countries, importing countries and several other countries that act as forwarders within the intra-EU market itself.

As can be seen in Table 1, in 2017, the main intracommunity destinations and importing countries were Germany, the United Kingdom, France, and the Netherlands. These four countries account for 57.18% of the total. In 2002, the participation was 73.95%. This reduction is due to the fact that Poland, Spain, Italy, and Lithuania have increased their participation as import destinations.

Table 1: Evolution in the participation percentage of importing countries in the intra-EU tomato market (2002–2007–2017).

Participation of importing countries (%)				
	2002	2007	2017	Var
Austria	1.82	2.01	1.86	0.04
Belgium	2.96	2.76	3.72	0.76
Bulgaria	0.01	0.05	0.8	0.87
Cyprus	0.00	0.01	0.02	0.02
Czech Republic	3.01	3.41	3.50	0.49
Denmark	1.09	1.31	1.36	0.27
Estonia	0.35	0.40	0.52	0.17
Finland	0.91	0.86	1.14	0.23
France	11.20	10.44	9.02	−2.18
Germany	34.87	29.87	28.72	−6.15
Greece	0.15	0.28	0.26	0.11
Hungary	0.33	0.61	0.49	0.16
Ireland	1.04	1.08	1.11	0.07
Italy	2.94	4.23	4.51	1.58
Latvia	0.53	0.57	1.06	0.53
Lithuania	0.34	0.61	4.06	3.72
Luxembourg	0.20	0.22	0.20	0.00
Malta	0.00	0.02	0.05	0.05
Netherlands	11.89	9.13	5.81	−6.07
Poland	2.15	2.58	5.30	3.16
Portugal	2.32	1.15	1.29	−1.03
Romania	0.02	0.17	1.27	1.26
Slovakia	0.40	0.66	1.10	0.70
Slovenia	0.31	0.29	0.51	0.20
Spain	2.10	5.24	5.08	2.99
Sweden	3.08	3.58	3.51	0.43
United Kingdom	15.99	18.46	13.62	−2.37

Source: European Commission. Documents from the forecast working group. https://ec.europa.eu/agriculture/fruit-and-vegetables/product-reports/tomatoes_en.

Table 2: Evolution in the participation percentage of exporting countries in the intra-EU tomato market (2002–2007–2017).

Participation of exporting countries (%)				
	2002	2007	2017	Var
Austria	0.26	0.55	0.32	0.06
Belgium	8.31	8.31	7.82	−0.50
Czech Republic	0.00	0.27	0.29	0.29
France	5.52	5.00	8.81	3.29
Germany	1.35	1.58	0.71	−0.65
Italy	5.37	3.88	3.96	−1.41
Netherlands	29.95	33.60	36.08	6.13
Poland	0.00	2.46	1.35	1.35
Portugal	1.71	2.33	4.01	2.30
Spain	46.08	41.17	34.25	−11.83
Others	1.44	0.85	2.41	0.97

Source: European Commission. Documents from the forecast working group. https://ec.europa.eu/agriculture/fruit-and-vegetables/product-reports/tomatoes_en.

With regard to the countries of origin (Table 2), in 2017, the Netherlands and Spain represent 70.33% of the total market, while in 2002, this figure rose to 76.03%. The lead exporting country in 2002 was Spain although it has subsequently been replaced by the Netherlands.

The purpose of this paper is to analyze the interdependent relations of the transactions between exporting and importing countries of different member states of the EU, in a specific product market such as that of tomatoes and to understand which were the key countries in three specific years (2002–2007–2017). This vegetable is considered the most important fresh products within the EU market, both in terms of production and commercialization.

2. Materials and Methods

In the analysis of relationships between data, the scientific literature is extensive and varied. DEMATEL-based approaches to the analysis of financial risks in banks are current, as demonstrated in the works of [2–4]. The use of Influence Network Relation Map (INMR) is a working tool that has been proven in the analysis of complex processes of sustainable urbanism, as shown by the works of [5, 6]. Against this, the analysis of economic variables by input-output table triangulation has been for many years the preferred tool of economic researchers. Our work is an improvement on this last method, without diverting attention, in future studies, in the comparison with other procedures of analysis of relationships between variables.

The input-output methodology has been applied in various fields, including industry [7–13], energy [14], water [15, 16], price systems [17], structural changes [18, 19], and human capital [20], among others. It has also been used to analyze key sectors, such as the work by Haji [21] related to key sectors in Kuwait's productive structure, the article by Cassetti [22] which identifies sectors by transactions in different countries, and the work by Díaz et al. [23] who identify key sectors by means of multiplier, diffusion, topological hierarchy, technological level, diffusion of innovation, and polluting capacity.

To learn about the interdependent relationships within a matrix, it must be based on a structural analysis that presents the following situations (Figure 1):

(i) Total dependency or interdependency is defined as each sector depending on the rest of the sectors, acting as a supplier and receiver of inputs.

(ii) Interdependence occurs when some sectors depend on others, even if some can act independently.

(iii) Hierarchy corresponds to a nonreal situation but rather is obtained by rearranging of certain criteria. In the case of a perfectly triangular matrix, the sectors that are above and below the row corresponding to a given sector have very different relationships [24]. Those located below are suppliers of the sector, causing an increase in the final demand of their product which then generates indirect demands that are precipitated by the main diagonal of the matrix without affecting the sectors located above the sector in question. Those located above the row are considered to be customers. Any increase in demand corresponding to the output of any of them generates indirect demand for the output of the sector in question.

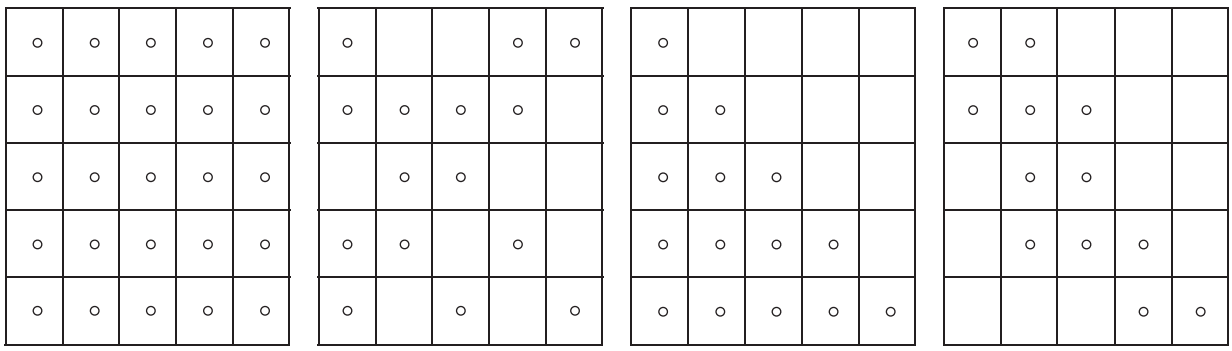


Figure 1: Basic concepts of structural analysis: (a) interdependence; (b) independence; (c) hierarchy; (d) circularity.

(iv) Circularity is where perfect triangulation is not achieved and the asymmetry is broken due to another demand and supply sector or multiregional interdependence.

This study will focus on creating a hierarchy through the triangulation methodology. According to Korte and Oberhofer [25], this methodology performs, among others, the following tasks:

(i) Sheds light on the functioning of the economy from a structural point of view

(ii) Facilitates comparison between the economies of different countries and periods

(iii) Optimally influences cycle and growth

(iv) Makes matrix inversion easier

(v) Useful for forecasting and economic planning

This method of triangulation by means of input-output tables was pioneered by Korte and Oberhofer [25, 26] although it has since been subject to a number of variations [27–33]. Grötschel et al. [34] utilizes new polyhedral results for the triangulation problem in a linear programming cutting plane framework. Professor Göram Östblom [35] introduced the alternative approach of maximizing the sum of negative differences between the elements below and the symmetrical elements above the principal diagonal. A series of suboptimal solutions, converging to the optimal solutions, is proposed as an approximate solution for tables of large dimension.

In order to study the interrelationships of trade flows in the tomato market in EU countries, three matrices ($M11 \times 11$) have been developed, namely, through the trade flows corresponding to 2002, 2007, and 2017 (Tables 3–5).

An outline of the trade flows can be seen in Table 6. The columns correspond to imports (tons), and the rows correspond to exports (tons). The total sum of the transactions of the exporting and importing countries coincide. The data have been obtained from Eurostat [36] and from the reports of the European Working Group in "tomatoes" [37].

Once the table of commercial transactions is completed, the technical coefficients table (A) is retrieved. The components of this table represent the direct commercial effort carried out by country i per unit of consumption of country j. The value of each of the coefficients would be $a_{ij} = M_{ij}/c_j$.

From the matrix of coefficients, we develop the technological matrix which is the difference between the identity matrix and that of the technical coefficients. In a classic input-output analysis [1], (A) would be a technical matrix used to analyze the direct relationships between total demand (internal and external) and trade flows. The Leontief inverse matrix is defined as $B = (I–A)^{-1}$ (Tables 7–9).

To determine the hierarchy of specific countries according to the impact that they could have on a specific economic sector in other countries, Chenery–Watanabe coefficients have been traditionally used [38], but these have significant limitations [39]:

(a) They use the direct coefficients of the transaction matrix without taking into account the indirect effects

(b) The coefficients obtained represent measures without regard to their deviations

(c) These indices are not weighted

Given these limitations, another form of triangulation is proposed by means of artificial intelligence based on pathfinding techniques using cost functions.

The use of cost functions for the minimization of functions is a technique widely used in artificial intelligence, statistics, and other technical disciplines when it comes to solving an optimization problem iteratively where the gradient calculation of the cost function is used to determine the direction of modification of the problem resolution parameters.

There is extensive literature on this subject, whether it be generally about these techniques, their application, or as a way to optimize functions in economics [40–42]. The search for optimal paths has continued its development in

Table 3: Matrix of trade transactions 2002.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	**0**	1429	20160	21558	21507	308731	1	66317	609	106501	195761
ES	202873	0	156883	21742	21915	204400	42676	270520		183431	617299
FR	3182	2286	**0**	0	7629	27913	36	1361	382	5089	67692
BE	11519	1047	29495	**0**	1932	73583	56	28904	1007	4168	12625
IT	776	46	7375	1365	**0**	66118	0	36319	401	8886	124268
DE	13326	1889	5082	1106	4475	**0**	2675	193	1902	2607	7875
PT	0	34217	22	0	8	7	**0**	0	0	1022	26
PL	47	0	0	54	0	315	0	**0**	137	0	160347
EL	0	0	176	0	3	80	0	1000	**0**	93	1738
UK	325	114	9	81	0	428	0	564	0	**0**	6432
REU	856	0	129	12103	31	1319	0	15311	14	1415	65289

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 4: Matrix of trade transactions 2007.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	**0**	7049	28488	22742	30278	376111	149	14499	3154	174265	173305
ES	138203	0	128124	21177	24931	158701	29180	44742	7	159568	109599
FR	6311	3716	**0**	22508	16189	49390	449	5308	571	8555	16172
BE	21472	2701	44924	**0**	2102	60145	25	2005	3333	6123	10527
IT	3759	210	8622	739	**0**	50398	7	2004	1674	14302	31319
DE	14239	36	3855	0	6977	**0**	2801	7756	3678	16187	30254
PT	0	205201	1341	0	160	7	**0**	0	0	1342	66
PL	1210	546	2862	306	996	1768	0	**0**	2050	26832	37849
EL	33	0	16	0	0	437	0	288	**0**	1613	1747
UK	1595	474	747	0	0	338	0	32	0	**0**	5047
REU	1038	94	359	1380	292	5044	0	796	312	1356	**0**

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 5: Matrix of trade transactions 2017.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	**0**	14195	33206	39547	46763	415675	104	33665	154	176984	273235
ES	118008	**0**	150798	16385	37233	169804	32255	57324	16	143183	130053
FR	19488	5558	**0**	29545	20100	68153	626	16739	0	12319	30139
BE	24372	2289	57325	**0**	5395	47936	0	1306	689	7081	17081
IT	1759	695	6397	1401	**0**	35896	0	4141	1175	9759	43478
DE	4867	900	4670	5105	7804	**0**	1141	17888	1608	18296	34469
PT	216	58902	1883	118	411	0	**0**	599	0	2736	107
PL	524	4219	915	110	1057	2441	83	**0**	501	5806	13160
EL	0	28	0	0	0	183	0	48	**0**	0	17803
UK	2387	1451	913	72	0	131	1	22	0	**0**	9785
REU	2725	179	158	535	247	4405	1	741	363	1117	**0**

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 6: Trade flows.

	Country 1	Country 2	Country *n*	XI	ID=C + XE	DT = XE + XI + C
Country 1	M11	M12	M1*n*	X1	ID1	DT1
Country 2	M21	M22	M2*n*	X2	ID2	DT2
Country *n*	M*n*1	M*n*2	M*nn*	X*n*	ID*n*	DT*n*
II	II1	II2	II*n*			
IE	IE2	IE2	IE*n*			
P	P1	P2	P*n*			
C	C1	C2	C*n*			

ID = intermediate demand; II_i = imports from other EU countries (intra-EU) to country "i;" IE_i = imports from countries outside the EU to country "i;" XI_i = exports to other EU countries from country "i;" XE_i = exports to countries outside the EU from country "i;" P_i = production of country "i;" C_i = apparent consumption of country "i;" $C_i = P_i - XI_i - XE_i + II_i + IE_i$.

Table 7: Inverse matrix 2002.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	1.08153	0.00227	0.03774	0.21442	0.02440	0.50196	0.01480	0.29490	0.00184	0.28758	0.09892
ES	1.88948	1.01532	0.26784	0.59621	0.06643	1.20549	0.47438	1.55313	0.00427	0.96260	0.46674
FR	0.03673	0.00174	1.00204	0.01074	0.00814	0.05834	0.00272	0.01941	0.00052	0.02343	0.02917
BE	0.11798	0.00125	0.04072	1.02594	0.00519	0.16167	0.00555	0.13963	0.00141	0.04321	0.02312
IT	0.02140	0.00029	0.01096	0.02241	1.00100	0.10651	0.00305	0.14302	0.00065	0.02852	0.05540
DE	0.12321	0.00222	0.01123	0.03469	0.00706	1.05952	0.02984	0.03814	0.00208	0.04035	0.01524
PT	0.04391	0.02359	0.00625	0.01386	0.00155	0.02802	1.01102	0.03609	0.00010	0.02485	0.01086
PL	0.00169	0.00001	0.00031	0.00679	0.00006	0.00203	0.00006	1.00438	0.00015	0.00073	0.05820
EL	0.00004	0.00000	0.00022	0.00010	0.00001	0.00016	0.00000	0.00370	1.00000	0.00024	0.00085
UK	0.00323	0.00009	0.00018	0.00161	0.00008	0.00224	0.00010	0.00325	0.00001	1.00092	0.00278
REU	0.02055	0.00015	0.00476	0.11018	0.00077	0.02282	0.00075	0.07406	0.00019	0.01027	1.03003

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 8: Inverse matrix 2007.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	1.72267	0.02251	0.08597	0.45347	0.05109	0.94586	0.10963	0.47000	0.01403	0.67010	0.29166
ES	7.29160	1.22545	0.54313	2.22485	0.24469	4.29244	1.49188	2.72100	0.06397	3.25458	1.37238
FR	0.60625	0.01410	1.04419	0.40714	0.03262	0.42493	0.06588	0.25817	0.00731	0.26656	0.12427
BE	1.02650	0.01671	0.10211	1.28364	0.03348	0.65199	0.07722	0.32236	0.01299	0.41885	0.18708
IT	0.23077	0.00414	0.02214	0.07208	1.00747	0.19824	0.02213	0.09937	0.00438	0.12330	0.06792
DE	0.70427	0.02137	0.04195	0.19067	0.02710	1.39118	0.14106	0.30508	0.01053	0.31694	0.14900
PT	0.91546	0.15369	0.06966	0.27965	0.03087	0.53903	1.18719	0.34165	0.00803	0.41127	0.17238
PL	0.06745	0.00134	0.00688	0.02248	0.00291	0.04051	0.00505	1.01981	0.00308	0.08024	0.04271
EL	0.00244	0.00004	0.00016	0.00069	0.00008	0.00197	0.00022	0.00462	1.00003	0.00440	0.00199
UK	0.07173	0.00128	0.00447	0.01924	0.00215	0.04005	0.00493	0.02039	0.00059	1.02811	0.01634
REU	0.06632	0.00107	0.00456	0.03270	0.00231	0.04480	0.00514	0.03029	0.00105	0.02971	1.01200

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 9: Inverse matrix 2017.

	NL	ES	FR	BE	IT	DE	PT	PL	EL	UK	REU
NL	1.9387	0.0293	0.1260	0.7894	0.1263	1.0925	0.0198	0.5585	0.0053	0.7270	0.4568
ES	7.0214	1.1219	0.6308	3.0707	0.5102	4.2069	0.3331	2.4324	0.0206	2.9344	1.7747
FR	1.5104	0.0289	1.1166	0.8947	0.1256	0.9588	0.0231	0.5575	0.0049	0.6016	0.3883
BE	1.4360	0.0241	0.1549	1.6050	0.1022	0.8774	0.0162	0.4370	0.0053	0.5573	0.3555
IT	0.1521	0.0031	0.0180	0.0794	1.0107	0.1333	0.0023	0.0761	0.0026	0.0785	0.0707
DE	0.3599	0.0072	0.0317	0.1960	0.0334	1.2079	0.0133	0.2169	0.0039	0.1731	0.1136
PT	0.3559	0.0544	0.0336	0.1568	0.0263	0.2129	1.0163	0.1263	0.0010	0.1534	0.0899
PL	0.0598	0.0045	0.0056	0.0266	0.0054	0.0380	0.0022	1.0193	0.0010	0.0348	0.0247
EL	0.0024	0.0001	0.0002	0.0010	0.0002	0.0016	0.0000	0.0011	1.0000	0.0009	0.0139
UK	0.1398	0.0033	0.0103	0.0582	0.0092	0.0795	0.0018	0.0411	0.0004	1.0529	0.0405
REU	0.1569	0.0026	0.0107	0.0692	0.0106	0.0944	0.0017	0.0506	0.0011	0.0614	1.0373

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

works using genetic algorithm techniques, distributed programming, and colony-based algorithm models [43–46].

In the present work, the number of elements we have to deal with (the set of countries on which the interdependence analysis is performed) is small enough to rule out a minimum distributed path search model or one based on genetic algorithms. Instead, we will follow the idea of generating a triangulation path through permutations of rows/columns in the matrix directed by a cost calculation function.

This function is constructed in such a manner as to create a theoretical global minimum to a perfectly triangulated matrix, with all zeros above the diagonal. In reality, however, this state may not be attainable. The value of the cost function will give us, in this case, the level of triangulation achieved.

In order to develop this technique, a definition of the series of concepts in our problem is needed.

U is defined as the space of possible solutions to a problem f.

$y \in \mathrm{U}$ is the possible solution that satisfies the conditions of f.

$J(y)$ is defined as $\mathrm{U} \longrightarrow \mathrm{R}$ as a cost function of f if it meets the following conditions:

Table 10: Triangulated matrix 2002.

	FR	DE	UK	BE	EL	PL	PT	ES	IT	NL	REU
FR	1.00204	0.05834	0.01941	0.02343	0.00052	0.01074	0.00272	0.00174	0.00814	0.03673	0.02917
DE	0.01123	1.05952	0.03814	0.04035	0.00208	0.03469	0.02984	0.00222	0.00706	0.12321	0.01524
UK	0.00031	0.00203	1.00438	0.00073	0.00015	0.00679	0.00006	0.00001	0.00006	0.00169	0.05820
BE	0.00018	0.00224	0.00325	1.00092	0.00001	0.00161	0.00010	0.00009	0.00008	0.00323	0.00278
EL	0.00022	0.00016	0.00370	0.00024	1.00000	0.00010	0.00000	0.00000	0.00001	0.00004	0.00085
PL	0.04072	0.16167	0.13963	0.04321	0.00141	1.02594	0.00555	0.00125	0.00519	0.11798	0.02312
PT	0.00625	0.02802	0.03609	0.02485	0.00010	0.01386	1.01102	0.02359	0.00155	0.04391	0.01086
ES	0.26784	1.20549	1.55313	0.96260	0.00427	0.59621	0.47438	1.01532	0.06643	1.88948	0.46674
IT	0.01096	0.10651	0.14302	0.02852	0.00065	0.02241	0.00305	0.00029	1.00100	0.02140	0.05540
NL	0.03774	0.50196	0.29490	0.28758	0.00184	0.21442	0.01480	0.00227	0.02440	1.08153	0.09892
REU	0.00476	0.02282	0.07406	0.01027	0.00019	0.11018	0.00075	0.00015	0.00077	0.02055	1.03003

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

(1) $\forall a, b \in U \longrightarrow J(a) < J(b) \leftrightarrow$ a is a better solution than b for problem f

(2) J, at least a minimum, corresponds to the optimal solution at the local level

The process of minimizing J can be straightforward if it is defined as a derivable and continuous analytical function and there is an analytical procedure to determine the minimums of said function. Conversely, it can be an iterative procedure, when J does not have an analytical form or there is no direct procedure to achieve its minimum values.

For the problem at hand, U is defined as a set of possible solutions to the problem of triangulation of an input-output matrix (f).

We define the cost function $J(y)$ on a matrix $y \epsilon U$ as the number of "zeros" above the main normalized diagonal according to the following expression:

$$J(y) = 1 - \frac{n_{\text{zeros}}}{(1/2)\text{lines} * \text{columns}}. \quad (1)$$

The cost function will define the level of triangulation of the matrix in each iteration. In order to select the best permutation, the cost function will be the discriminating element. This function has a maximum value of 1 for a perfectly triangulated matrix.

This cost function meets the conditions defined above although it requires the definition of the concept of zero in an input-output matrix.

In this work, the concept of zero is fundamental for the operation of the algorithm. There are general two possibilities:

(i) An absolute threshold based on the nature of the interdependence matrix data

(ii) A relative threshold related to the minimally significant interrelation factor for a study of these characteristics

For the present work, we have opted for the second option and define zero as that value of the matrix that is below the ten percentiles in an analysis of the distribution of values. This definition has been taken into account, compared to other possible ones (for example, a minimum threshold) as it allows the process of selecting interdependence values not relevant to the study to be automated.

The matrix triangulation process is as follows:

Step 1. Determine the threshold value of zero to triangulate on the matrix.

Step 2. Repeat this step a predetermined number of times or until a stop condition is met.

Step 2.1. Swap two rows of the array randomly.
Step 2.2. Calculate the value of J for the new matrix.
Step 2.3. If the value of J is less than that of the original matrix, replace the original with the permuted one.

The choice of rows and columns to be exchanged is methodical in order to test all possible permutations.

The stop criteria for Step 2 are, either by time, or a situation where the cost function does not decrease despite having tried all possible permutations. When this occurs, the gradient of function J is close to zero, and a minimum has been found for the cost function.

The implementation of this process has been carried out using Matlab, as computing support to develop numerous tests, with different values for execution times and stop conditions and obtaining the results set out below.

3. Results

The results are displayed in Tables 10–12. The countries that lead the table are those that depend on the imports from other countries, while those at the base are clearly the exporting countries, and their activity is essential for the operation of the system.

For the year 2002 (Figure 2), France, Germany, and the United Kingdom are the main clients leading the table, while Spain, Italy, the Netherlands, and the rest of EU countries are located at the bottom, the majority of these being exporters.

For 2007 (Figure 3), the top of the table is headed by Poland, the United Kingdom, and the Netherlands and at the bottom by Spain, France, and Belgium.

For 2017 (Figure 4), the countries leading the table are Germany, Belgium, and France with the lower quadrant being occupied by the Netherlands, the United Kingdom, and Spain.

When comparing the years 2002 and 2007, a significant change is observed in the intra-EU tomato trade and how the

Table 11: Triangulated matrix 2007.

	PL	UK	NL	REU	IT	DE	PT	EL	ES	FR	BE
PL	1.72267	0.67010	0.01403	0.29166	0.05109	0.94586	0.10963	0.47000	0.02251	0.08597	0.45347
UK	0.07173	1.02811	0.00059	0.01634	0.00215	0.04005	0.00493	0.02039	0.00128	0.00447	0.01924
NL	0.00244	0.00440	1.00003	0.00199	0.00008	0.00197	0.00022	0.00462	0.00004	0.00016	0.00069
REU	0.06632	0.02971	0.00105	1.01200	0.00231	0.04480	0.00514	0.03029	0.00107	0.00456	0.03270
IT	0.23077	0.12330	0.00438	0.06792	1.00747	0.19824	0.02213	0.09937	0.00414	0.02214	0.07208
DE	0.70427	0.31694	0.01053	0.14900	0.02710	1.39118	0.14106	0.30508	0.02137	0.04195	0.19067
PT	0.91546	0.41127	0.00803	0.17238	0.03087	0.53903	1.18719	0.34165	0.15369	0.06966	0.27965
EL	0.06745	0.08024	0.00308	0.04271	0.00291	0.04051	0.00505	1.01981	0.00134	0.00688	0.02248
ES	7.29160	3.25458	0.06397	1.37238	0.24469	4.29244	1.49188	2.72100	1.22545	0.54313	2.22485
FR	0.60625	0.26656	0.00731	0.12427	0.03262	0.42493	0.06588	0.25817	0.01410	1.04419	0.40714
BE	1.02650	0.41885	0.01299	0.18708	0.03348	0.65199	0.07722	0.32236	0.01671	0.10211	1.28364

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

Table 12: Triangulated matrix 2017.

	DE	BE	FR	EL	IT	REU	PT	PL	NL	UK	ES
DE	1.93870	0.45680	0.00530	0.72696	0.02929	0.12598	0.78945	0.12631	1.09254	0.01983	0.55853
BE	0.15686	1.03731	0.00106	0.06138	0.00256	0.01074	0.06920	0.01063	0.09441	0.00172	0.05061
FR	0.00236	0.01394	1.00002	0.00094	0.00006	0.00017	0.00105	0.00016	0.00165	0.00003	0.00108
EL	0.13982	0.04048	0.00039	1.05286	0.00333	0.01031	0.05817	0.00921	0.07945	0.00176	0.04109
IT	7.02141	1.77472	0.02064	2.93438	1.12189	0.63081	3.07072	0.51019	4.20686	0.33312	2.43237
REU	1.51040	0.38826	0.00491	0.60161	0.02891	1.11663	0.89468	0.12558	0.95885	0.02306	0.55750
PT	1.43599	0.35550	0.00532	0.55727	0.02412	0.15490	1.60500	0.10222	0.87738	0.01624	0.43697
PL	0.15211	0.07073	0.00256	0.07852	0.00315	0.01800	0.07939	1.01073	0.13332	0.00226	0.07609
NL	0.35994	0.11357	0.00385	0.17311	0.00722	0.03173	0.19605	0.03338	1.20790	0.01328	0.21695
UK	0.35587	0.08991	0.00105	0.15338	0.05442	0.03362	0.15678	0.02628	0.21294	1.01627	0.12627
ES	0.05984	0.02468	0.00102	0.03484	0.00448	0.00558	0.02663	0.00539	0.03799	0.00222	1.01926

NL, Netherlands; ES, Spain; FR, France; BE, Belgium; IT, Italy; DE, Germany; PT, Portugal; PL, Poland; EL, Greece; UK, United Kingdom; REU, rest EU.

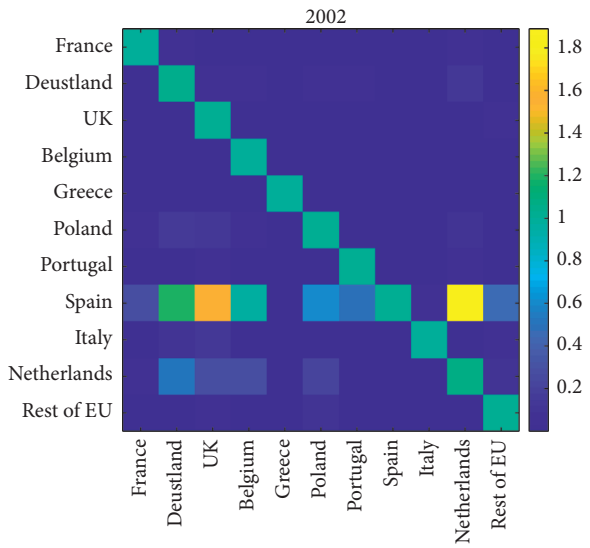


Figure 2: Triangulation of inverse matrix 2002. Source: own compilation.

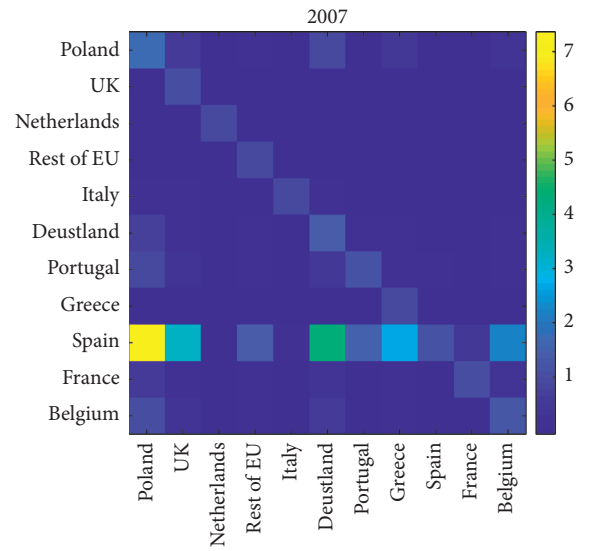


Figure 3: Triangulation of inverse matrix 2007. Source: own compilation.

crisis of 2017–13 has caused a virtual return to the initial situation of 2002 in terms of the main customers and suppliers. In this sense, Germany, France, and Belgium stand out in the upper part as importing countries, particularly during the winter season. At the base of the matrix, the Netherlands and Spain are shown to be principal intracommunity exporters.

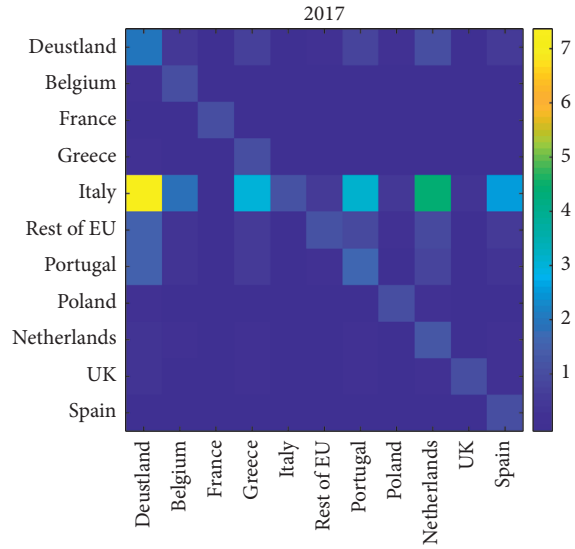


Figure 4: Triangulation of inverse matrix 2017. Source: own compilation.

4. Discussion

Traditional triangularizing methods that considery–Watanabe coefficients give acceptable results, but they present a series of limitations as discussed in the methodology section.

The implementation of pathfinding techniques is in line with other works such as that of Kondo [7] who proposes a new method to triangulate input-output tables based on mixed integer programs to compare the production structures of multiple economies.

The identification of key sectors/countries is essential to understand the operation of the system, as reflected in the works of Haji [16], Cassetti [17], Díaz et al. [18], and García Muñiz et al. [47]. In principle, the leading country in tomato exports within the EU is Spain; however, despite this fact, it is not considered to be a key player. The same cannot be said for the Netherlands and Belgium which are considered to be key players and operate efficiently with regard to their own trade transactions as well as in the reexportation of products from other countries. The question that arises and that should be identified is which process is the most profitable, production or commercialization. In the present case, the answer is clearly commercialization.

In the case of Spain, it is losing a great opportunity to act as an exporting country to Morocco. Despite being the lead exporter of tomatoes within the EU and a net exporting country, Spain is also becoming an import market. The reason for this is shift in purchasing patterns from traditional channels to that of large distributors: discount stores, supermarkets, and small independent stores are strongly committed to fresh food. Likewise, as reflected in the latest MERCASA study [48], these establishments, through their platforms, buy fresh products outside Spain. On the contrary, Table 13 shows how the Mercadona supermarket, regional chains such as IFA and Euromandi, and the Lidl discount store are becoming major market drivers as they continue to increase their market share.

Table 13: Market share of supermarket chains in Spain (2018–15).

	Market share (%)	Variation (2018–2015)
Auchan group	3.5	0.9
Carrefour	8.4	−0.6
Dia group	7.5	−1.7
Eroski group	5.3	−1.0
Lidl	4.8	2.5
Mercadona	24.9	0.6
Regional supermarkets*	11.8	1.1

*Regional supermarkets do not include Eroski or Arbol supermarkets. Source: Kantar Worldpanel, Data to end of 2018.

5. Conclusions

The application of pathfinding techniques using a cost function is a consistent tool for the triangulation process of input-output tables. Consequently, the choice of a suitable cost function is essential to obtain a solution to triangulation.

The algorithm used for triangulation makes it possible to carry out the triangulation process by avoiding local minimums in the cost function by exploring time paths with a cost function value worse than the best found until a certain moment. This feature substantially improves the result. The size of this speculative scan slows the process by a factor dependent on the number of rows (columns) in the matrix.

Developing a table which displays the trade flows of intraindustrial tomato trade between the main countries of the European Union is a starting point for obtaining later the tables of technical coefficients and Leontief´s inverse matrix. Likewise, comparing several years gives us an in-depth look at how these countries perform.

A structural analysis is essential to understand the interrelationships of a sector or an economy. The present case study has focused on hierarchy through triangulation. The sectors shown below the diagonal in the matrix are suppliers

of the tomato sector, resulting in an increase in the final demand while those above the diagonal are considered to be customers in said sector.

The leading supplier countries are the Netherlands and Spain, with the former also acting as an export forwarder. As for customers, Germany stands out above the rest.

The tendency for the loss of power of wholesale markets at the expense of large distributors has been particularly noticeable in Spain which had traditionally been a net tomato exporter since it is the largest producer of fresh tomatoes and, but which in recent years, has seen the level of imports gradually increase.

Data Availability

The data used to support this study are free and publicly available from European Commission.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] W. Leontief, *The Structure of the American Economy, 1919–1939*, Oxford University Press, Oxford, UK, 1941.

[2] H. Dinçer and S. Yüksel, "Financial sector-based analysis of the G20 economies using the integrated decision-making approach with DEMATEL and TOPSIS," in *Emerging Trends in Banking and Finance*, pp. 210–223, Springer, Berlin, Germany, 2018.

[3] S. Altuntas and T. Dereli, "A novel approach based on DEMATEL method and patent citation analysis for prioritizing a portfolio of investment projects," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1003–1012, 2015.

[4] W.-S. Lee, A. Y. Huang, Y.-Y. Chang, and C.-M. Cheng, "Analysis of decision making factors for equity investment by DEMATEL and Analytic Network Process," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8375–8383, 2011.

[5] C.-H. Cheng, J. Liou, and C.-Y. Chiu, "A consistent fuzzy preference relations based ANP model for R&D project selection," *Sustainability*, vol. 9, no. 8, p. 1352, 2017.

[6] Q.-G. Shao, J. Liou, S.-S. Weng, and Y.-C. Chuang, "Improving the green building evaluation system in China based on the DANP method," *Sustainability*, vol. 10, no. 4, p. 1173, 2018.

[7] J. W. Foster and R. C. Higgins, "Modelling the quality of a manufacturing process using input-output analysis," *International Journal of Production Research*, vol. 10, no. 2, pp. 141–146, 1972.

[8] T. J. Mules, "Input-output analysis in Australia: an agricultural perspective," *Review of Marketing and Agricultural Economics*, vol. 511 pages, 1983.

[9] H. Verbeek, *Innovative Clusters*, Erasmus University, Rotterdam, Netherlands, 1999.

[10] M. Sonis, G. J. D. Hewings, and D. Guo, *Industrial Clusters in the Input-Output Economic System*, University of Illinois, Champaign, IL, USA, 2007.

[11] P. Midmore, M. Munday, and A. Roberts, "Assessing industry linkages using regional input-output tables," *Regional Studies*, vol. 40, no. 3, pp. 329–343, 2006.

[12] Y. Kondo, "Triangulation of input-output tables based on mixed integer programs for inter-temporal and inter-regional comparison of production structures," *Journal of Economic Structures*, vol. 3, no. 1, p. 2, 2014.

[13] R. Harthoorn, *Production Chains*, Central Bureau of Statistics, Netherlands, 1986.

[14] E.W. Henry, "A capacity growth input-output model with forward recursive solution," *Energy Economics*, vol. 17, no. 1, pp. 3–13, 1995.

[15] E. Velázquez, "An input-output model of water consumption: analysing intersectoral water relationships in Andalusia," *Ecological Economics*, vol. 56, no. 2, pp. 226–240, 2006.

[16] E. Dietzenbacher and E. Velázquez, "Analysing andalusian virtual water trade in an input-output framework," *Regional Studies*, vol. 41, no. 2, pp. 185–196, 2007.

[17] O. J. Asger, "Aggregation in input-output models: prices and quantities," *Economic Systems Research*, vol. 5, no. 3, pp. 253–264, 1993.

[18] F. Aroche-Reyes and A. S. García Muñiz, "Modelling economic structures from a qualitative input-output perspective: Greece in 2005 and 2010," *Metroeconomica*, vol. 69, no. 1, pp. 251–269, 2018.

[19] D. Lind, *Value Creation and Structural Change during the Third Industrial Revolution*, Lund University, Lund, Sweden, 2014.

[20] H. Zhang and X. Chen, ""An extended input–output model on education and the shortfall of human capital in China," *Economic Systems Research*, vol. 20, no. 2, pp. 205–221, 2008.

[21] J. A. Haji, "Key sectors and the Structure of Production in Kuwait - an input-output approach," *Applied Economics*, vol. 19, no. 9, pp. 1187–1200, 1987.

[22] M. Cassetti, "A new method for the identification of patterns in input-output matrices," *Economic Systems Research*, vol. 7, no. 4, pp. 363–382, 1995.

[23] B. Díaz, L. Moniche, and A. Morillas, "A fuzzy clustering approach to the key sectors of the Spanish economy," *Economic Systems Research*, vol. 18, no. 3, pp. 299–318, 2006.

[24] W. Leontief, *Análisis Económico Input-Output*, Planeta-Agostin, Barcelona, Spain, 1973.

[25] B. Korte and W. Oberhofer, "Triangularizing input-output matrices and the structure of production," *European Economic Review*, vol. 1, no. 4, pp. 482–511, 1970.

[26] B. Korte and W. Oberhofer, "Zwei Algorithmen zur Lösung eines komplexen Reihenfolgeproblems," *Unternehmensforschung Operations Research–Recherche Opérationnelle*, vol. 12, no. 1, pp. 217–231, 1968.

[27] J. Lamel, J. Richter, S. Teufel, and W. Bauer, "Triangulation," *Economic Bulletin for Europe*, vol. 23, pp. 59–75, 1971.

[28] Z. Drabek, "A comparison of technology in centrally-planned and market-type economies," *European Economic Review*, vol. 25, no. 3, pp. 293–318, 1984.

[29] Y. Fukuy, "A more powerful method for triangularizing input-output matrices and the similary of production structures," *Econometrica*, vol. 54, no. 6, pp. 1425–1433, 1986.

[30] E. C. Howe, "A more powerful method for triangularizing input-output matrices: a Comment," *Econometrica*, vol. 59, no. 2, pp. 521–523, 1991.

[31] O. Haltia, "A triangularization algorithm without ringshift permutation," *Economic Systems Research*, vol. 4, no. 3, pp. 223–234, 1992.

[32] F. A. Reyes, "Evolution of the productive structure of Mexico - exercise for triangularizing the input-output matrix," *Trimestre Económico*, vol. 247, pp. 397–416, 1995.

[33] T. F. M. Chang, L. C. Piccinini, L. Iseppi, and M. A. Lepellere, "The black box of economic interdependence in the process of structural change. EU and EA on the stage," *Italian Journal of Pure and Applied*, vol. 31, pp. 285–306, 2013.

[34] M. Grötschel, M. Jünger, and G. Reinelt, "Optimal triangulation of large real world input-output matrices," *Statistische Hefte*, vol. 25, no. 1, pp. 261–295, 1983.

[35] G. Östblom, "Use of the convergence condition for triangularizing input-output matrices and the similarity of production structures among Nordic countries 1970, 1980 and 1985," *Struct Chang Econ Dyn*, vol. 8, no. 1, pp. 481–500, 1997.

[36] C. Europea, "Statistics EU trade flows inside & outside Europe," 2017, http://exporthelp.europa.eu/thdapp/display.htm?page=st2fst_Statistics.html&docType=main&languageId=en.

[37] C. Europea, "Document de travail. Groupe de travails prévisions tomates," 2018, http://ec.europa.eu/agriculture/fruit-and-vegetables/product-reports/tomatoes/index_en.htm.

[38] H. B. Chenery and T. Watanabe, "International comparisons of the structure of production," *Econometrica*, vol. 26, no. 4, pp. 487–521, 1958.

[39] E. Jahangard and V. Keshtvarz, "Turkey economies: a network theory approach," *Iranian Economic Review*, vol. 16, no. 32, pp. 45–58, 2012.

[40] J. Noceral and S. Wright, *Numerical Optimization*, Springer Science & Business Media, Berlin, Germany, 2006.

[41] A. Antoniou and L. Wu-Sheng, *Practical Optimization: Algorithms and Engineering Applications*, Springer Science & Business Media, Berlin, Germany, 2007.

[42] F. Audrino and G. Barone-Adesi, "Functional gradient descent for financial time series with an application to the measurement of market risk," *Journal of Banking & Finance*, vol. 29, no. 4, pp. 959–977, 2005.

[43] K. Kulankara, S. Satyanarayana, and S. N. Melkote, "Iterative fixture layout and clamping force optimization using the genetic algorithm," *Journal of Manufacturing Science and Engineering*, vol. 124, no. 1, pp. 119–125, 2002.

[44] A. Richardson and E. Olson, "Iterative path optimization for practical robot planning," in *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3881–3886, IEEE, San Francisco, CA, USA, September 2011.

[45] J. Y. Hwang, J. S. Kim, S. S. Lim, and K. H. Park, "A fast path planning by path graph optimization," *IEEE Transactions on Systems, Man, and Cybernetics-Part a: Systems and Humans*, vol. 33, no. 1, pp. 121–129, 2003.

[46] J. Isern-González, D. Hernández-Sosa, E. Fernández-Perdomo, J. Cabrera-Gámez, A. C. Domínguez-Brito, and V. Prieto-Marañón, "Path planning for underwater gliders using iterative optimization," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, pp. 1538–1543, IEEE, Shanghai, China, May 2011.

[47] A. S. García, A. Morillas, and C. Ramos, "Key sectors: a new proposal from network theory," *Regional Studies*, vol. 42, no. 7, pp. 1013–1030, 2008.

[48] MERCASA, *Informe del Consumo en España 2017*, Ministerio de Agricultura, Madrid, Spain, 2018.