AIMS ENERGY

VOLUME NO. 13
ISSUE NO. 2
MAY - AUGUST 2025

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

AIMS ENERGY

AIM AND SCOPE

AIMS Energy is an international Open Access journal devoted to publishing peer-reviewed, high quality, original papers in the field of Energy technology and science. We publish the following article types: original research articles, reviews, editorials, letters, and conference reports.

AIMS Energy welcomes, but not limited to, the papers from the following topics:

- Alternative energy
- Bioenergy
- Biofuel
- Energy conversion
- Energy conservation
- Energy transformation
- Future energy development
- Green energy
- Power harvesting
- Renewable energy

Editor in Chief

Department of Industrial & Systems Engineering, Chung Yuan Christian Pekka Peura University, 200, Chung Pei Road, Chungli,Taiwan 32023, ROC Levón Institute, Vaasa Energy institute, University of Vaasa, Wolffintie 34,
65200 Vaasa, Finland May Wu
System Assessment Section, Energy System Division, Argonne National Daphiny Pottmaier Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439, USA
Bin Yang
Department of Biological Systems Engineering, Washington State University, 2710 Crimson Way, Richland WA 99354, USA
Ahmad Zahedi
College of Science, Technology and Engineering, Division of Tropical Environments and Societies, James Cook University, Queensland, 4811, Australia
Zhiqiang (John) Zhai
Department of Civil, Environmental, and Architectural Engineering, University of Colorado@Boulder, UCB 428, ECCE 249, Boulder, CO 80309-0428, USA
Xiaowei Zhou
Department of Chemical & Biological Engineering, Northwestern University, Evanston, IL, USA

AIMS Energy

(Volume No. 13, Issue No. 2, May - August 2025)

Contents

The Influence of Structural Morphology on the Efficiency of Building Integrated Wind Turbines (BIWT)

Hassam Nasarullah Chaudhry 1,*, John Kaiser Calautit 2 and Ben Richard Hughes 2

1 School of the Built Environment, Heriot-Watt University, PO Box 294 345, Dubai, UAE. 2School of Civil Engineering, University of Leeds, Leeds LS2 9JT, UK.

ABSTRACT

A numerical investigation was carried out to determine the impact of structural morphology on the power generation capacity of building-integrated wind turbines. The performance of the turbines was analysed using the specifications of the Bahrain Trade Centre which was taken as the benchmark model, the results of which were compared against triangular, square and circular cross-sections of the same building. The three-dimensional Reynolds-Averaged Navier-Stokes (RANS) equations along with the momentum and continuity equations were solved for obtaining the velocity and pressure field. Simulating a reference wind speed of 6 m/s, the findings from the study quantified an estimate power generation of 6.4 kW indicating a capacity factor of 2.9 % for the benchmark model. The square and circular configurations however determined greater capacity factors of 12.2 % and 19.9 %, recording an estimated power production capability of 26.9 kW and 35.1 kW and confirming the largest extraction of the incoming wind stream. The optimum cross-sectional configuration for installing wind turbines in high-rise buildings was the circular orientation as the average wind speed at the wind turbines was accelerated by 0.3 m/s resulting in an overall augmentation of 5 %. The results from this study therefore highlighted that circular building morphology is the most viable building orientation, particularly suited to regions with a dominant prevailing wind direction.

Keywords: *Buildings; Computational Fluid Dynamics; power density; turbulence; wind turbine*

1. Introduction

The wind power industry has gone through a steady development over the years, focusing on an increasing interest in generating electricity from the wind. With the global energy concerns about the greenhouse gas emissions, considerable investment has been made into using wind as a renewable source of energy for the purpose of generating electricity. In 2007, the Bahrain World Trade Centre introduced the world's first building integrated wind turbines. This was the first instance that a commercial development integrated large-scale wind turbines within its design to harness the power of the wind. The three turbines, measuring 29 meters in rotor diameter were supported by bridges spanning between the complex's two towers. Through its positioning and the unique aerodynamic design of the towers, the prevailing on-shore Gulf breeze is intended to be funneled into the path of the turbines, helping to create power generation efficiency. The anticipated benefits determined that once the system was operational, the wind turbines delivered approximately 11–15 % of the energy needs of the building, or 1,100 to 1,300 megawatt-hours per year [1].

There are currently two major types of operational wind turbine technologies which can be used for building integration. These are classified as the Horizontal Axis Wind Turbines (HAWT) and the Vertical Axis Wind Turbines (VAWT). Horizontal axis wind turbines are the more common type of wind turbines in demand. In these turbines, the focus rotor shaft is pointed parallel to the direction of wind while the blades move perpendicular to that direction, thus providing high overall efficiency. On the other hand, the vertical axis wind turbines orientate on vertical axis where the focus rotor shaft is aligned vertically. A two-dimensional 'not-to-scale' design modeling based on the two wind turbine technologies is shown in Figure 1. The schematic provides the essential information on the individual situation of the major components in the wind turbine structure [2].

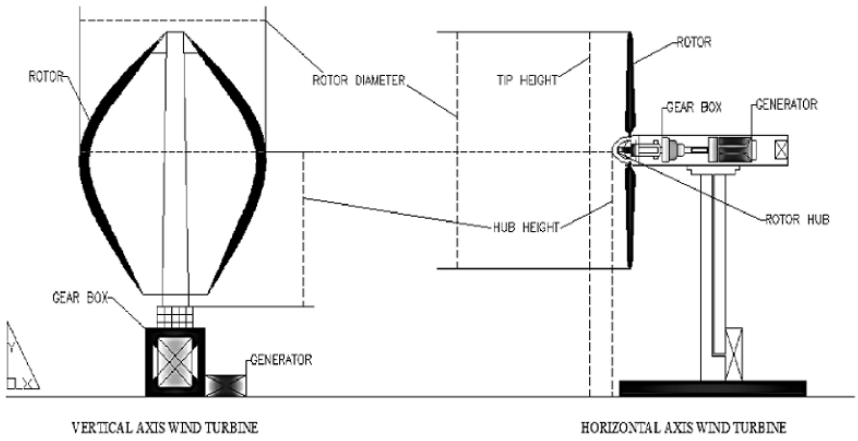


Figure 1. Structural comparison between the horizontal axis wind turbines and vertical axis wind turbines [2].

The potential of integrating wind turbines into buildings is vast since the wind speed is intensified at high altitudes resulting in an increase in electricity generation. Several wind turbine manufacturers have conducted extensive research on this technology since it aligns well with a strong appeal to shift from fossil fuels to renewables. Theoretically, the air velocity incrimination is directly proportional to increasing elevation, as it is unaffected by the urban environment and surrounding topography. Thus, installing wind turbines on top of buildings allows for taking advantage of this height efficiently. However, in many cases, the building geometry and its aerodynamic features can further assist in enhancing wind turbine performance by inducing a higher volume of air than the macro-climate [3, 4].

Numerical analyses involving Computational Fluid Dynamics (CFD) are often carried out for evaluating building design and energy performances. In the past few years, CFD has been playing an increasingly important role in the design of buildings. The information provided by CFD assists in investigating the impact of building technologies, quantifying indoor environment quality, and in integrating renewable energy systems. In this paper, the CFD modeling approach is used out in order to determine the influence of building morphology on the efficiency of building-integrated wind turbines.

2. Previous related work

Work related to implementation of renewable energy as power sources in buildings, in particular wind energy is rapidly increasing. Following is a concise assessment of previous work related to this investigation.

Chong et al. [5] conducted a study on the design of a wind turbine generator for the purpose of energy consumption in buildings. Two VAWTs in cross-wind orientation were integrated with an enclosure were installed above a cooling tower in order to harness the discharged wind for electricity generation. The enclosure in this case acts as a wind-power augmentation device to improve the performance of the wind turbine. The effect of installing the energy recovery turbine generator at the outlet of an exhaust air system was investigated via laboratory test on a scaled model of a cooling tower.

The tests were conducted with and without the VAWTs on the performance of the system. From the laboratory testing, it was observed that installing VAWTs at a correct position above the exhaust air system did not indicate any significant negative impacts on the performance of the cooling tower model. The performance of the cooling tower model was improved by the increment of intake air speed and the reduction off the motor power consumption. The rotational speed of the VAWTs were very high (> 400 rpm) while the tip speed ratios were in the range of 1.28–1.29, making it preferable for electricity generation. The exercise of using VAWTs in building was further studied by Muller et al. [6]. The study established graphical models to determine the theoretical efficiency of the Sistan type windmill. The work supports the initiative of integrating wind energy converters into buildings by employing geometry changes to the model under study in order to indicate the maximum expected convertor efficiency of 61 %.

Sharpe and Proven [7] carried out detailed work on the concept and early development of a building integrated wind turbine in Scotland. The Crossflex proposal used in this investigation is based on the Darrieus turbines approach, consisting of two or more flexible aero foil blades attached to both the top and bottom of a vertical rotating shaft. The blade design incorporated low solidity and low mass materials for its construction. Although the development of this technology is still in its early stage, the study has outlined the remaining tasks of its work. The subsequent stages will involve supplementary computer modeling using CFD analysis to model airflow over the cowling to establish the nature of augmentation and flow stability that may occur, and to optimise the design in respect of these.

Lu and Sun [8] presented an integrated method of both macro (weather data and domain topography) and

micro aspect (CFD) analysis in order to design wind turbines around numerous high-rise buildings with predominant wind in Hong Kong. Long-term wind data were compared at dense urban island and small island stations. The mean wind speeds for the urban locations were estimated at 2.93 m/s at 25 m above ground level. The study's findings determined that the wind power density at 4 m above the building roof was enhanced numerously by 1.3–5.4 times with 5–7 m/s inlet velocity.

Mithraratne [9] investigated the performance of wind turbines mounted on roof-tops, used for micro-generation in municipal houses in New Zealand. The research focused on energy consumption over a lifetime of the wind turbine with respect to the energy generation and Green House Gas (GHG) emissions at various circumstances in the life cycle. The work concluded that the feasibility level of roof-top wind turbines is generally poor as compared to large scale wind turbines for micro-generation.

Bahaj et al. [10] studied the significance of micro-wind turbines in generating energy in buildings. The work was based on various cities in the UK in accordance with accurate wind speed data, to evaluate the financial feasibility and appropriateness for domestic houses. The research concluded that areas of high wind speeds are most likely to benefit from micro-wind turbines as the payback period is notably less than the design lifetime and a significant reduction in carbon footprint is likely.

Lu and Ip [11] carried out work on the viability of implementing wind energy as a major power source for buildings in Hong Kong. The study comprised of CFD modeling based on the energy and turbulence models for analysis of various building structures. The research concluded that the height of high-rise buildings could be exploited to its optimum capability to improve the efficiency of wind turbines for power generation.

Li et al. [12] investigated the wind loads on the Pearl River Tower building in China to determine the power generation potential of wind turbines using wind tunnel testing. The Pearl River Tower, located in Guangzhou has 71 stories and rises about 310 m from the ground, which features four open holes (tunnels) equipped with four wind turbines at mechanical floors on two height levels. The objective of this study was to evaluate the wind speed amplifications in the tunnels for wind-power generation through the installation of wind turbines and to gain a better understanding of the wind effects on such a high-rise building with open holes. The findings from the analyses indicated that the bell-mouthed shape of the wind tunnels with contracted inner sections were useful in increasing the wind velocities at the location of the turbines. In addition, the work revealed that the presence of surrounding buildings influences wind speed amplifications and wind loads on the building structure.

Building upon the previous work carried out in literature, the focus of this study was to investigate the power generation potential of building-integrated wind turbines by analysing the effect of the structural morphology on its efficiency. The physical geometry of the Bahrain Trade Centre was taken as the benchmark model and the performance of wind turbines in response to the extracted wind using the building envelope was investigated using advanced numerical and theoretical analysis.

3. Computational domain

The computational domain comprised of the building geometry, which was designed according to the actual specifications of the high-rise tower and the specific wind turbines. The two 50 storey sail shaped towers are reported to measure to a height of 240 m and support three horizontal-axis wind turbines incorporating a rotor diameter of 29 m [1].

The three-dimensional Reynolds-Averaged Navier-Stokes (RANS) equations along with the momentum and continuity equations were solved using the commercial CFD code for the velocity and pressure field simulations. The model employs the control-volume technique and the Semi-Implicit Method for Pressure-Linked Equations (SIMPLE) velocity-pressure coupling algorithm with the second order upwind discretisation. The standard k-e transport model which is frequently used for incompressible flows was used to define the turbulence kinetic energy and flow dissipation rate within the model [13, 14]. The use of the standard k-e transport model on building configurations has been found in previous works of [2, 15-17]. The turbulence kinetic energy, k, and its rate of dissipation, e, are obtained from the following transport equations formulated in equation 1 and equation 2.

$$\frac{\partial}{\partial t}(\rho k)+\frac{\partial}{\partial x_i}(\rho k u_i)=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right]+G_k+G_b-\rho\epsilon-Y_M+S_k \qquad (1)$$

$$\frac{\partial}{\partial t}(\rho e)+\frac{\partial}{\partial x_i}(\rho e u_i)=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_e}\right)\frac{\partial e}{\partial x_j}\right]+C_{1e}\frac{e}{k}(G_k+C_{3e}G_b)-C_{2\epsilon}\rho\frac{e^2}{k}+S_e \qquad (2)$$

Where: Gk represents the generation of turbulence kinetic energy due to the mean velocity gradients, Gb represents the generation of turbulence kinetic energy due to buoyancy. YM represents the contribution of fluctuating dilatation in compressible turbulence to the overall dissipation rate. C1e, C2e and C3e are constants; σk and σe are the turbulent Prandtl numbers for k and e.

3.1. Mesh generation

Mesh generation is one of the most important processes in CFD simulation. The quality of the mesh plays an important role on the accuracy of results and the stability of the solution. For the investigated

computational domain, patch independent CFD tetrahedron meshing technique was applied on the geometry wherein the boundary conditions were applied on the edges and faces. The patch independent mesh algorithm for tetrahedron elements is based on the subsequent spatial subdivision algorithm which ensures refinement of the mesh where essential, but retains larger elements where feasible, therefore allowing faster computing times. The meshed model comprised of 2,013,428 nodes and 10,849,999 elements as displayed in Figure 2. The minimum face angle was 5.67° while the maximum edge length and element volume ratios were 9.9 and 44.6. Figure 2 and Figure 3 display the schematic of the benchmark geometry along with the meshed model.

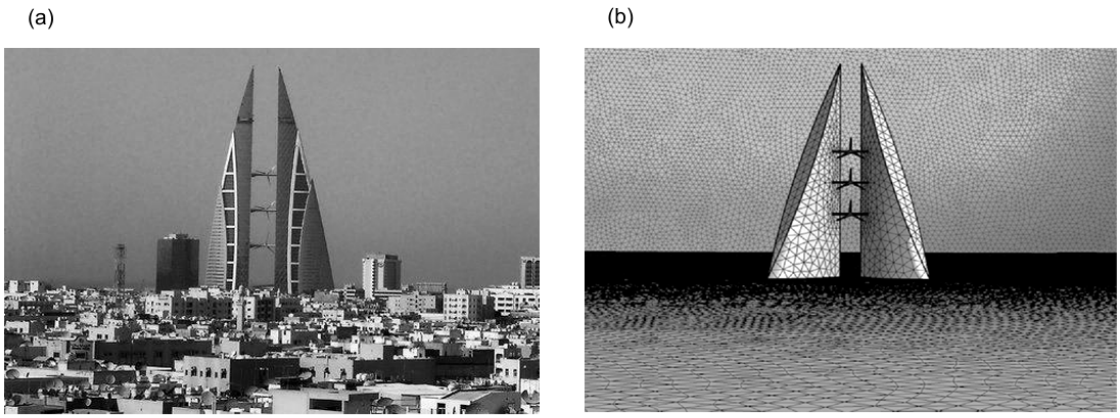


Figure 2. (a) Actual building configuration (b) meshed model.

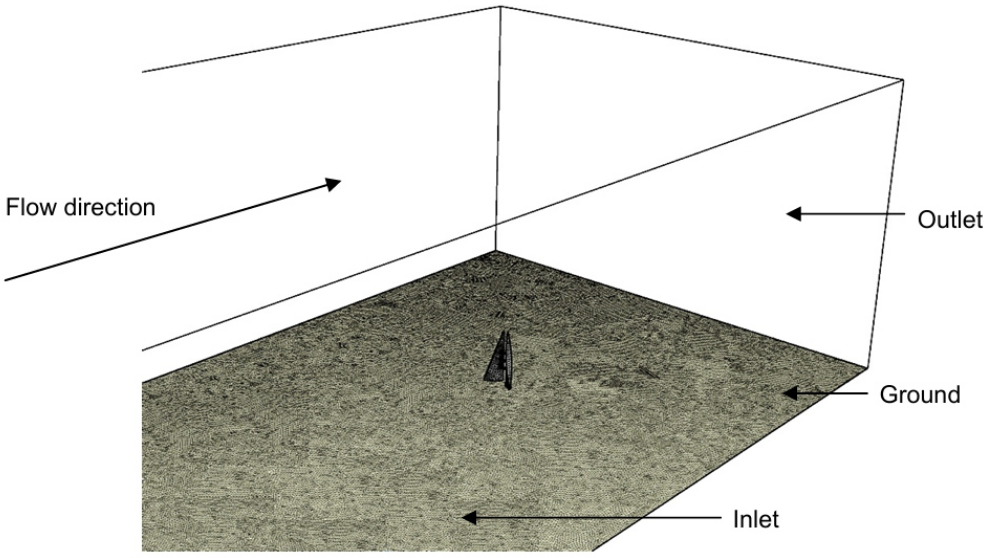


Figure 3. Representation of the flow domain.

For the current investigation, four individual shapes of the building envelope were evaluated in order to identify the optimum morphology for integrating wind turbines. Using the design of the Bahrain Trade Centre as the benchmark model, the analyses were conducted on buildings having a triangular, square and circular cross-section in order to determine the optimum configuration for mounting wind turbines. The design specifications of the investigated building models are displayed in Figure 4.

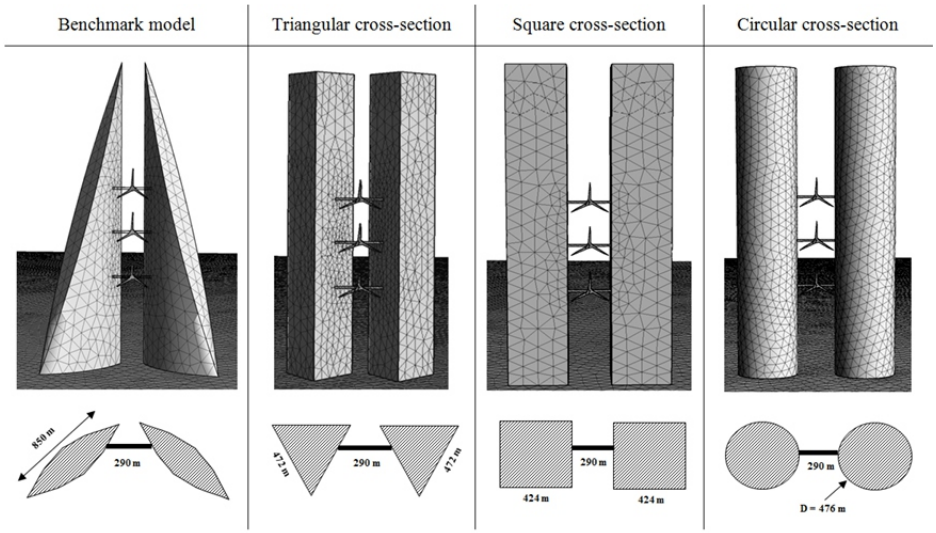


Figure 4. Design specifications of the investigated building models.

3.2. Grid independency

In order to verify the accuracy of the numerical models, a grid independency test was carried out to determine the variation in results over increasing mesh sizes. Basic concepts associated with mesh refinement deals with the refinement and evaluation of elements where the posterior error indicator is larger than the preset criterion, while mesh enrichment considers running higher order polynomials till the solution is expected to improve with a fixed mesh [14]. Grid verification was carried out using mesh refinements (h-method) in order to optimise the distribution of mesh size h over a finite element.

The area-weighted average value of the static pressure on the three wind turbines located on the building façade was taken as the error indicator, as the grid was refined from 1,519,000 to 10,849,999 elements with the average pressure value being 4.36 Pa. The grid was evaluated and refined until the posterior estimate error became insignificant between the number of elements and the posterior error indicator. The discretisation error was found to be the lowest at over ten million elements for both indicated variables and the mesh was thus selected to achieve a balance between accuracy and computational time. Figure 5 displays the variation in discretisation error at increasing number of meshed elements.

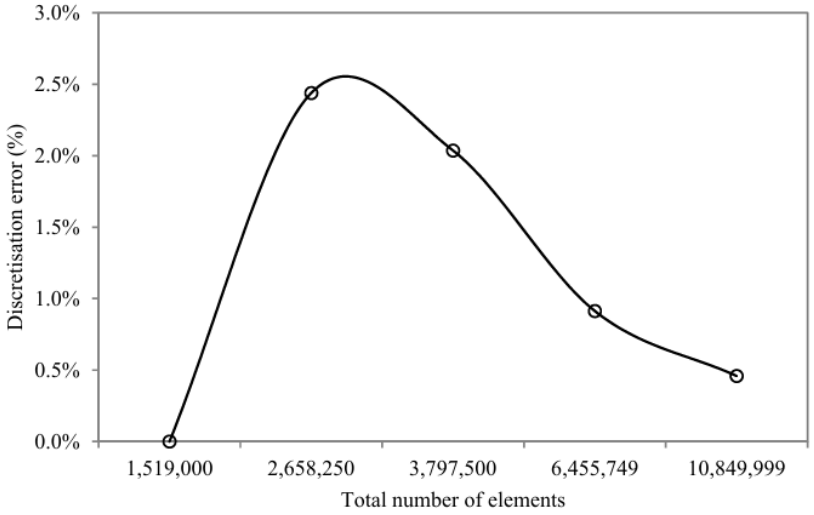


Figure 5. Variation in discretisation error at increasing number of elements.

3.3. Boundary conditions

The applied boundary conditions (Table 1) comprised of a reference velocity (uref) of 6 m/s at a height of 300 m (zref) approaching directly perpendicular to the building façade [18]. The geometry was modeled as a solid zone while the enclosure was modeled as a fluid zone for the analyses. The boundary conditions were kept identical throughout the numerical investigation for all analysed models.

Table 1. Boundary conditions.

Parameter	Type
Geometry	Solid zone
Enclosure	Fluid zone
Operating Pressure	Atmospheric
Viscous Model	k-epsilon (2 eqn)
Near-Wall Treatment	Standard Wall Functions
Velocity Formulation	Absolute
Solver Type	Pressure-Based
Time	Steady
Gravity	-9.81 m/s^2

3.4. Wind distribution and turbine specifications

In order to estimate how much energy a specific turbine will be expected to produce at a given location, the wind resource at that location must be identified. A wind turbine works by extracting kinetic energy out of the wind and converting it to mechanical and then electrical energy. The power that is available in the wind to be converted to electrical energy is defined in equation 3.

$$P_{wind} = \frac{1}{2}\rho A U^3 \qquad (3)$$

Where: Pwind is the power available in the wind, (3) is the density of air, A is the swept area of the turbine and U is the wind speed approaching the wind turbine.

Large-scale integration of wind turbines into buildings requires extensive research and development for it to operate at high efficiencies in order to balance out the high cost of installation. The specification of the wind turbines used in this study was mapped on the ones installed on the Bahrain Trade Centre. The fixed horizontal axis wind turbine comprised of a conventional nacelle design containing an enclosure with the gearbox, cooling system and the associated control system. The nominal electrical power rating of the turbine was 225 kW with the rotor diameter measuring 29 m. The rotor speed at full load was 38 rpm. The cut-in wind speed for the wind turbine operation was 4 m/s while the cut-out speed was 20 m/s [1].

Since the potential power production is proportional to the wind speed cubed, the annual mean wind power density cannot be defined by strictly using the mean annual wind speed. However, some knowledge of the wind distribution must be known to estimate the power density. The power density (Pdensity) can be used as a function of the power divided by the area and the expression is displayed in equation 4.

$$P_{density} = \frac{1}{2}\rho U^3 \qquad (4)$$

Using the computational domain, a representation of velocity boundary layer profile and turbulence intensity (I) at the windward side of the building is shown in Figure 6 wherein the wind speed is taken from the direction of the sea. The thickness of boundary layer of the atmosphere varies with wind speed, turbulence level and the type of surface. The power law is an empirical equation expressed in equation 5. For neutral stability conditions, α is approximately 1/7, or 0.143, regarded as a reasonable but conservative estimate [19].

$$u = u_{ref}\left(\frac{z}{z_{ref}}\right)^{\alpha} \qquad (5)$$

Where: u is the speed at a particular point, uref is the reference speed, z is the height at a particular point and zref is the reference height.

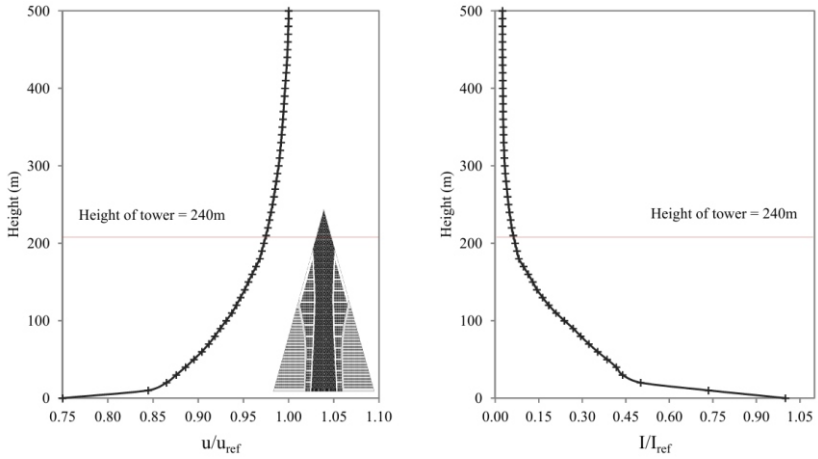


Figure 6. Representation of the boundary layer profile and turbulence intensity.

4. Results and discussion

4.1. Evaluation of airflow and pressure profiles

Using the benchmark model, Figure 7 displays the contour levels for static pressure as the air comes in contact with the building. A positive airside pressure was created as the air stream came in direct contact with the building façade. This was due to the force being directly perpendicular to the area of interaction. As a result, a negative pressure was created on the opposite end at the immediate downstream of the building where the air velocity increased due to the streamlined body of the structure. The analysis showed that a total pressure differential of 29.2 Pa was created for the entire geometry.

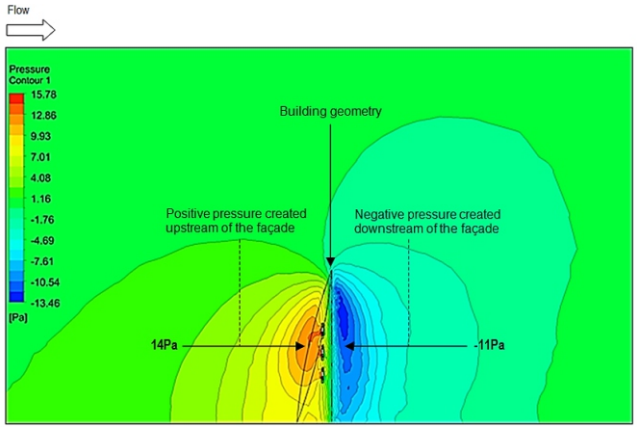


Figure 7. Contour levels of static pressure formation upstream and downstream of the building.

Figure 8 displays the contour levels of turbulence intensity, upstream and downstream of the benchmark model. At the windward side of the building, it was observed that the layers of turbulence intensified in inverse proportion to the height of the building with an average value of 0.45 %. The turbulence intensity was however found to increase (maximum value of 1.05 %) at the leeward side of the building as the

wind came in contact and sheared away towards the sides of the structure.

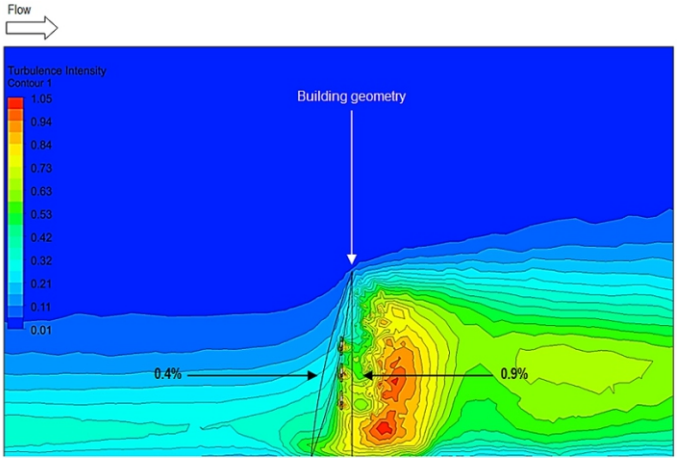
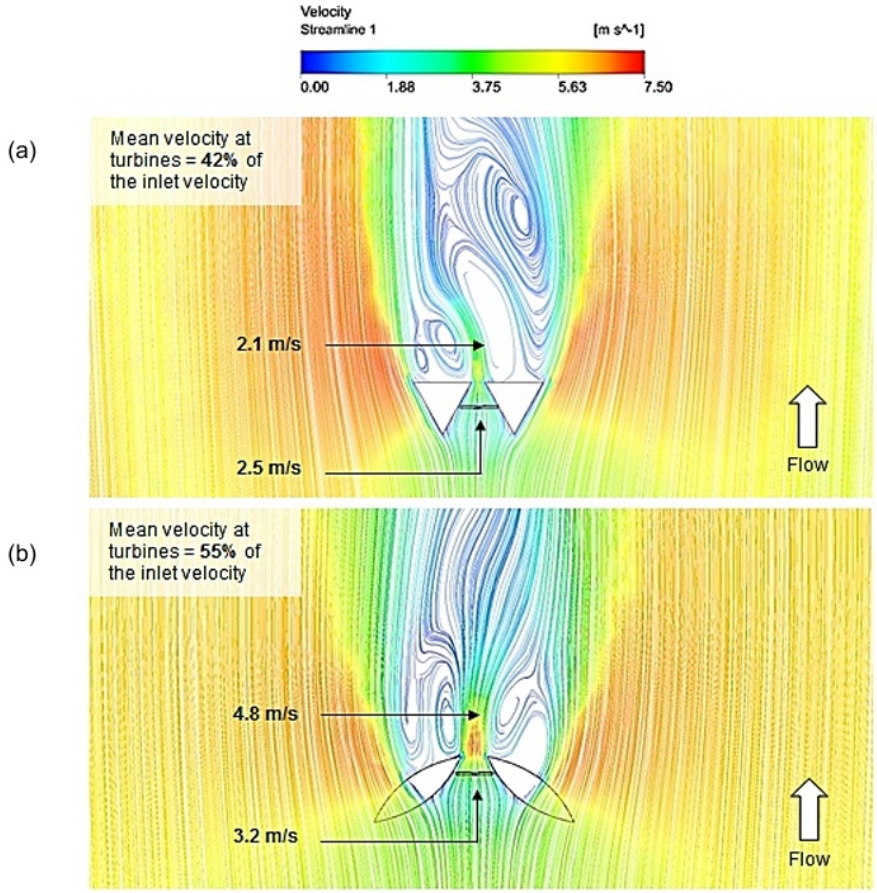


Figure 8. Contour levels of turbulence intensity upstream and downstream of the building.

Figure 9 displays the formation of air velocity profile for all analysed models in order to illustrate the effect of the building morphology on the wind speed received by the wind turbines.



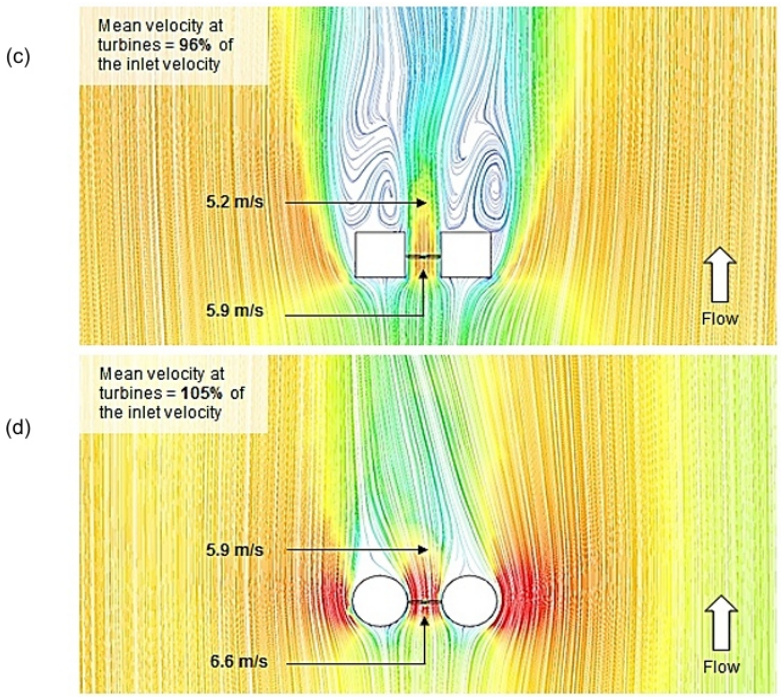


Figure 9. Air velocity streamlines representing the formation of resulting flow using: (a) triangular cross-section (b) benchmark model (c) square cross-section (d) circular cross-section.

At a reference speed of 6 m/s, the only successful augmentation of the speed was achieved using the circular cross-section with the mean velocity at wind turbines determined at 6.6 m/s. The investigation showed that the mean air velocity at the wind turbines using the configuration of the existing benchmark model was 55 % of the inlet velocity. This was superior to an equilateral triangular cross-section which showed the least recovery in wind speeds with the turbines capturing only 42 % of the inlet velocity.

A graphical representation of the variation in velocity and pressure profiles with respect to the height of the building is shown in Figure 10. A periodic increase in air velocity was observed for the square and circular cross-sections with the wind speed augmenting with increasing amplitude. The three regions of distinct peaks represent the elevation levels of the three wind turbines used in this study. The average wind speed across the profile was 6.6 m/s and 6.0 m/s for the circular and square cross-sections, thereby indicating the optimum extraction of the prevailing inlet wind. The average wind speeds for the triangular and benchmark configurations were 2.9 m/s and 3.7 m/s thus confirming the inability of the structural design to capture the incoming wind when blown perpendicular to the building. With regards to the static pressure formations, negative pressure gradients were recorded for the circular and square cross-sections while the other two designs displayed positive pressure profiles, thus confirming the reason for the reduction in wind speed.

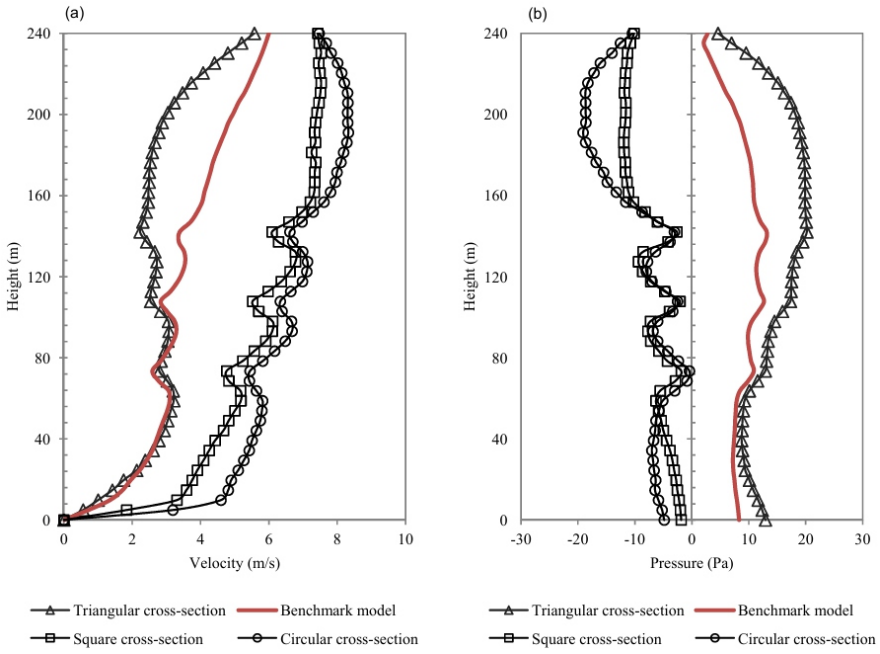


Figure 10. Formation of velocity and pressure layers with periodic rise in altitude.

4.2. Power density and wind turbine capacity factor

Following the numerical investigation, an estimation of the wind power characteristics was calculated theoretically. Using the rotor diameter of 29 m for the wind turbine used in the Bahrain Trade Centre (the benchmark model), a comparison between power in the wind and the tip speed ratio is illustrated in Figure 11. Wind turbines are designed to operate at their optimum tip speed ratio in order to extract as much power possible from the air stream. For a grid connected wind turbine with three rotor blades, the optimum tip speed ratio is recommended between 6 and 8 [20]. While a linear increase in wind power was obtained with increasing wind speeds, the recommended tip speed ratio values were found to be at 4 m/s and 5 m/s respectively.

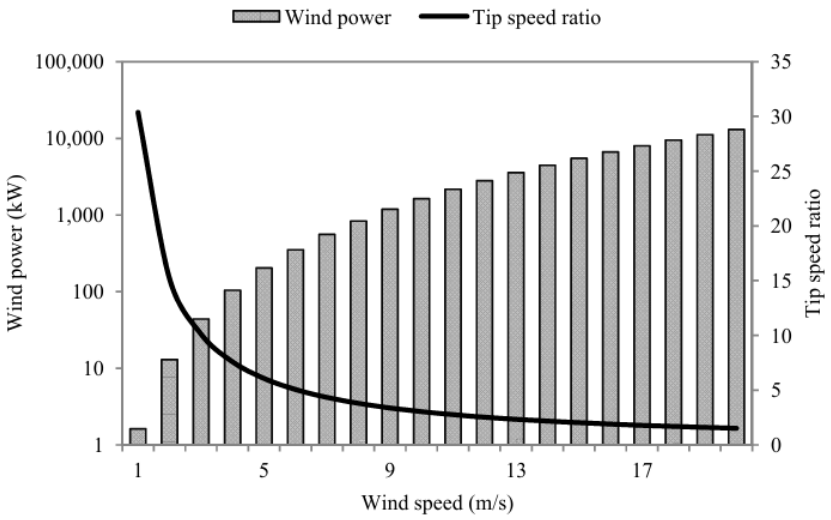


Figure 11. Relationship between wind power and tip speed ratio at increasing wind speeds.

Using the wind speeds obtained from the CFD models, the conservative figures of estimated power generation capability of wind turbines for all analysed cases is displayed in Table 2. Three wind turbines were used on the buildings which were categorised as Turbine 1 (Low-altitude), Turbine 2 (Mid-altitude) and Turbine 3 (High-altitude). For the triangular cross-section, the results showed that none of the wind turbines were able to meet the cut-in wind speed requirement, thereby indicating no production of power. Using the benchmark model, the study quantified an estimate power generation of 6.4 kW indicating a capacity factor of 2.9 %. The square and circular cross-sections however determined greater capacity factors of 12.2 % and 19.9 %, recording an estimated power production capability of 26.9 kW and 35.1 kW and confirming the largest extraction of the incoming wind stream. The results quantified that the wind turbines located between the circular building morphology had the highest power density of 203 W/m2 using the high-altitude turbine. Figure 12 displays the breakdown of results in a pie-chart representation for the estimated power generation capacity of wind turbines integrated in buildings having different structural morphologies.

Table 2. Estimated power density and capacity factor.

Cross-section	**Turbine**	**Velocity** (m/s)	**Estimated power** (W)	**Power density** (W/m^2)	**Capacity Factor** (%)
Triangular cross-section	Wind turbine 1	2.8	-	-	-
	Wind turbine 2	2.5	-	-	-
	Wind turbine 3	2.3	-	-	-
Benchmark model	Wind turbine 1	2.6	-	-	-
	Wind turbine 2	3.3	-	-	-
	Wind turbine 3	4.1	6,430	43.1	2.9%
Square cross-section	Wind turbine 1	4.8	14,060	69.4	6.2%
	Wind turbine 2	5.9	27,470	127.2	12.2%
	Wind turbine 3	6.6	39,380	179.7	17.5%
Circular cross-section	Wind turbine 1	5.4	21,200	99.2	9.4%
	Wind turbine 2	6.6	39,400	127.2	17.5%
	Wind turbine 3	6.9	44,700	203.1	19.9%

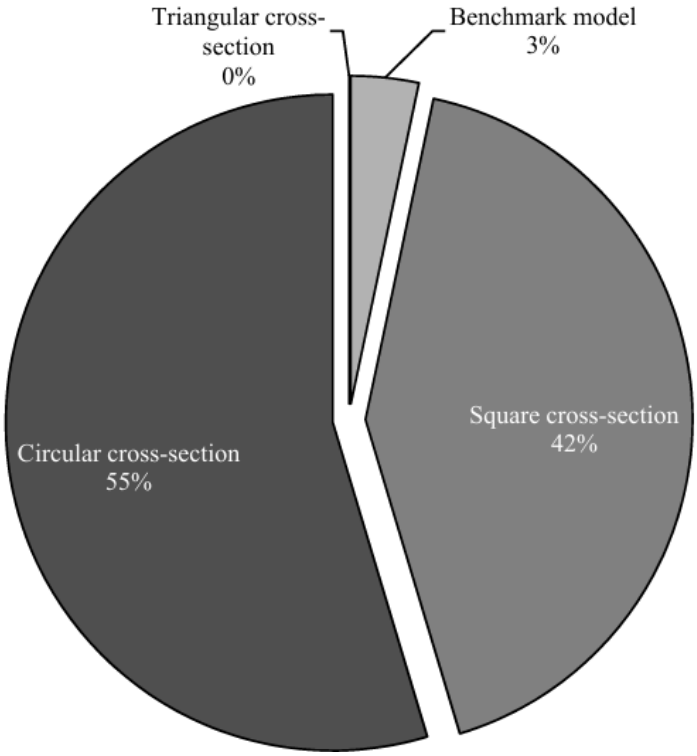


Figure 12. Power generation comparison between the analysed structural morphologies.

The extracted wind speeds by the three wind turbines for all four cases is shown in Figure 13. Despite its streamlined façade, the average speed at the wind turbines for the benchmark model was 2.7 m/s lower than the incoming speed of 6 m/s. For the square cross-section, the average speed was only 0.2 m/s lower than the incoming speed this indicating a high percentage of wind extraction. The square cross-sectional configuration indicated a 41 % superior extraction of incoming wind than the benchmark model. However, the optimum cross-sectional configuration for installing wind turbines in high-rise buildings was the circular orientation as the average wind speed at the wind turbines was 0.3 m/s greater than the incoming wind resulting in an average augmentation of 5 %. The maximum wind speed of 6.9 m/s was recorded at wind turbine 3 which was located at an altitude of approximately 160 m from the ground level. The circular configuration depicted a 49 % superior extraction of incoming wind than the benchmark model. In general, the findings of this study highlighted that circular cross-section is the most suitable orientation for integrating wind turbines, particularly suited to regions with a dominant prevailing wind direction.

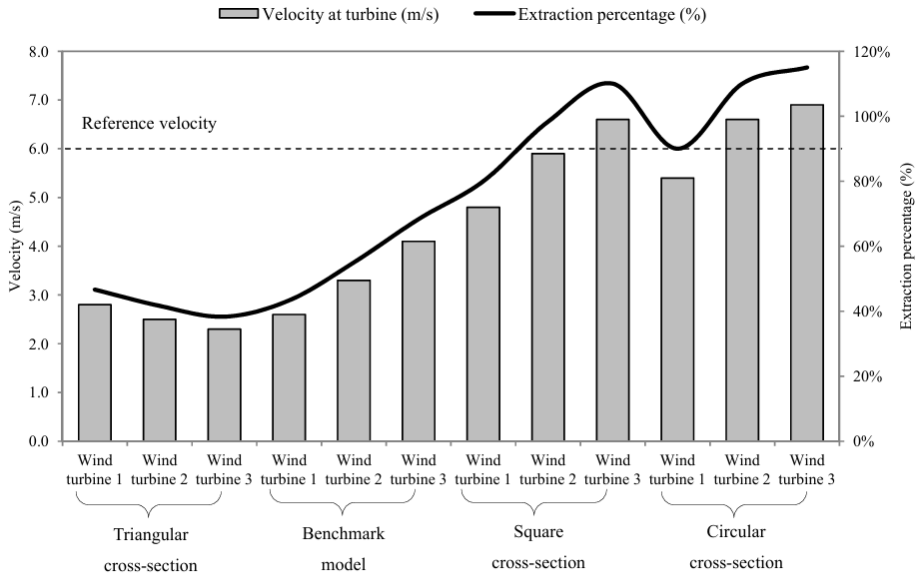


Figure 13. Wind speeds recorded at individual wind turbines for the four analysed cases.

5. Conclusion

In this paper, the feasibility of implementing building-integrated wind turbines was determined by investigating the effect of structural morphology on the extraction of prevailing inlet wind. The power generation capacity of wind turbines was determined using the specifications of the Bahrain Trade Centre which was taken as the benchmark model and the findings were compared against triangular, square and circular cross-sections of the same building. The three-dimensional Reynolds - averaged Navier-Stokes (RANS) equations along with the momentum and continuity equations were solved using the commercial CFD code for velocity and pressure field simulations. Using a reference wind speed of 6 m/s, the study quantified an estimate power generation of 6.4 kW indicating a capacity factor of 2.9 % for the benchmark model. The square and circular cross-sections however determined greater capacity factors of 12.2 % and 19.9 %, recording an estimated power production capability of 26.9 kW and 35.1 kW and confirming the highest extraction of the incoming wind stream. The findings from this study highlighted that circular cross-section is the most viable building orientation, particularly suited to regions with a dominant prevailing wind direction as a mean wind speed augmentation of 5 % was achieved at the turbines.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. Killa S, Smith RF. (2008) Harnessing Energy in Tall Buildings: Bahrain World Trade Center and Beyond. Council of Tall Buildings and Urban Habitat (CTBUH) 8th World Congress; Dubai, United Arab Emirates.

2. Chaudhry HN, Hughes BR. (2011) Computational analysis of dynamic architecture. Journal of Power and Energy, P I Mech Eng Part A 225, 85-95.

3. Hughes BR, Chaudhry HN. (2011) Power generation potential of dynamic architecture, World Acad Sci, Eng Technol 5: 1-24.

4. Hughes BR, Chaudhry HN, Ghani SA. (2011) A review of sustainable cooling technologies in buildings, Renew Sust Energ Rev 15: 3112-3120.

5. Chong WT, Yip SY, Fazlizan A, et al. (2013) Design of an exhaust air energy recovery wind turbine generator for energy conservation in commercial buildings. Renew Energ 67: 252-256.

6. Muller G, Jentsch MF, Stoddart E. (2008) Vertical axis resistance type wind turbines for use in buildings. Renew Energ 34: 1407-1412.

7. Sharpe T, Proven G. (2010) Crossflex: Concept and early development of a true building integrated wind turbine. Energ Buildings 42: 2365-2375.

8. Lu L, Sun K. (2014) Wind power evaluation and utilization over a reference high-rise building in urban area. Energ Buildings 68: 339-350.

9. Mithraratne N. (2009) Roof-top wind turbines for microgeneration in urban houses in New Zealand. Energ Buildings 41: 1013-1018.

10. Bahaj AS, Myers L, James PAB. (2006) Urban energy generation: Influence of micro-wind turbine output on electricity consumption in buildings. Energ Buildings 39: 154-165.

11. Lu L, Ip, KY. (2007) Investigation on the feasibility and enhancement methods of wind power utilization in high-rise buildings of Hong Kong. Renew Sust Energ Rev 13: 450-461.

12. Li QS, Chen FB, Li YG, et al. (2013) Implementing wind turbines in at all building for power generation: A study of wind loads and wind speed amplifications. J Wind Eng Ind Aerod 116: 70-82.

13. Launder BE, Spalding DB. (1972) Lectures in mathematical models of turbulence. London, England: Academic Press.

14. Chung TJ. (2002) Computational Fluid Dynamics, Cambridge University Press; illustrated edition, ISBN-0521594162.

15. Calautit JK, Hughes BR, Ghani SA. (2013) Numerical investigation of the integration of heat transfer devices into wind towers. Chem Eng T 34: 43-48.

16. Calautit JK, Hughes BR, Ghani SA. (2013) A Numerical Investigation into the Feasibility of Integrating Green Building Technologies into Row Houses in the Middle East. Architectural Sci Rev 56:

: 279-296.

17. Hughes BR, Calautit JK, Ghani SA. (2012) The Development of Commercial Wind Towers for Natural Ventilation: A Review. Appl Energ 92: 606-627.

18. Wind & weather statistics Bahrain Airport 2014, Available from http://www.windfinder.com/windstatistics/Bahrain.

19. Kubik ML, Coker PJ, Hunt C. (2011) Using meteorological wind data to estimate turbine generation output: a sensitivity analysis. World Renewable Energy Congress; Linkoping. Sweden.

20. Cetin NS, Yurdusev MA, Ata R, et al. (2005) Assessment of optimum tip speed ratio of wind turbines. Math Comput Appl 10: 147-154.

Phenological cycle and physicochemical characteristics of avocado cultivars in subtropical conditions

Geovanna Cristina Zaro 1,*, Paulo Henrique Caramori 2, Cíntia Sorane Good Kitzberger 2, Fernanda Aparecida Sales 2, Sergio Luiz Colucci de Carvalho 2, and Cássio Egidio Cavenaghi Prete 1

1 Universidade Estadual de Londrina, Centro de Ciências Agrárias, Londrina-PR, Brasil 2 Instituto Agronômico do Paraná, Área de Ecofisiologia, Londrina-PR, Brasil

A B S T R A C T

Avocado has a great potential as a commercial crop in southern Brazil, for its high productivity, rusticity, and multiple uses. Its high oil content can be explored for biodiesel production with advantages over other crops. This study comprised six avocado cultivars-Geada, Fortuna, Fuerte, Margarida, Primavera and Quintal-belonging to the Agronomic Institute of Paraná (IAPAR) collection, from Londrina, Brazil (23° 23' S, 50° 11' W). Analysis of fruit growth (length and diameter) allowed the classification of the cultivars into groups showing early, midseason, and late maturation, which were harvested in March/April, May/June, and July/August, respectively. The fruits were analyzed to assess their pulp, peel, and seed proportions, and their levels of oil (pulp) and starch (seed). Results showed these six cultivars are good alternatives for oil extraction. Fuerte stands out as the most adequate for biodiesel production from pulp and seed due its higher yield of oil and starch. The fruit cycle diversity of these cultivars might allow them for prolonged fruit production, both for fresh fruit marketing and biodiesel supply, as well as possibly using fruit pulp for oil extraction and seed starch for alcohol production.

Keywords: *biodiesel; growth curves; oil content; starch content; acidity; peroxide*

1. Introduction

Avocado is considered a tropical fruit, although its genetic diversity allowed it to adapt to subtropical climate conditions. In Brazil, it is present in nearly every state; however, the commercial plantations that are concentrated in the South, Southeast, and Northeast were responsible for producing 157,000 t in 2013, according to the Brazilian Institute of Geography and Statistics [1]. Several avocado cultivars are found throughout the Brazilian territory presenting different shapes, colors, sizes, and chemical composition.

In Brazil, avocado is mainly consumed naturally (raw fruit), or used in confectionery, cosmetics, and oil productions [2]. Another potential use is the extraction of its pulp oil for the production of biodiesel. Given the diversity of cultivars, wide climate adaptation, and high oil content with adequate physical and chemical quality [3], avocado is a promising crop for biodiesel production. In addition, it has the advantage of enabling the extraction of two important raw materials, the oil from the pulp and the ethanol from seed [4]. The fruit oil content is third after palm and olive [5]. Previous studies showed a variation from 5 to 30% in pulp oil content, and from 8 to 26% in seed starch content between different avocado

cultivars. This pointed out the need for a thorough evaluation of cultivars, in order to select those with high levels of pulp oil and seed starch for further uses of avocado's raw material [2,6].

Avocado has an extremely low fruit set, typically less than 0.3% under natural pollination to about 5% with manual pollination [7]. In the first month after flowering a large drop in potential fruit production occurs due to low fertility and abnormal flowers; later, a natural fertilized seed dropping also occurs due to competition for vegetative flux [8]. Therefore, evaluations of the development cycle of the fruits should only start after this natural seed dropping. Avocado also presents protogynous dichogamous flowers [9] i.e., female parts of flowers mature before the male parts separating cultivars into two groups (A and B). However, in order to ensure pollination it is necessary to mix cultivars from both groups in the same area.

Fruits (raw material) need to be produced for a prolonged period in order to supply the industry and/or the market, so that the processing structure and the consumer may have full access to the product. In this sense, species with diverse production cycles such as avocado have a great potential for industrial use.

Knowledge on the periodicity of climate conditions, on soil and environmental conditions and on the effects of these factors on plants biological cycle, including reproductive organs development and vegetative growth, is important for the study of avocado phenology [10]. Information on phenology then enables planning of planting and fruit harvesting and marketing, allowing taking advantage of market opportunities and maximizing industrial activity. According to [11], the different thermal conditions and cycles of avocado cultivars enable fruit harvesting throughout most of the year, providing favorable conditions for supplying the biodiesel production industries in Paraná state, Brazil.

Nevertheless, for the raw avocado materials to be used in the industry it is necessary to fulfill certain efficiency and quality requirements. In this work, we examine the production cycle, harvest season, and physicochemical properties of six avocado cultivars from Paraná, in order to determine their potential for biodiesel production.

2. Materials and Methods

The field study was based on data collected in the orchards of the experimental farms belonging to the Agronomic Institute of Parana State (IAPAR), in Paranavaí (23° 05′ S, 480 m of altitude) and Londrina (23° 23′ S, 566 m of altitude), Paraná, Brazil. The climate of this region is Cfa, according to Koeppen classification. The average annual temperature is 21 °C to 22 °C with precipitation from 1300 to 1600

, according to IAPAR historical records. The rainy season begins in September and ends in March-April [12]. Severe frosts are infrequent and do not constitute a risk factor for avocado [13]. Between 1986 and 1996, data of 24 avocado cultivars were collected from these two locations [11], from flowering to harvest, which was determined as the time when fruits started to fall naturally. These data allowed classifying the cultivars into three groups according to their phenological cycle: early, midseason, and late harvest. Based on the results of that study a more detailed research was carried out in Londrina between June 2011 and December 2012, comprising data from six cultivars belonging to the three harvest groups: Geada and Fuerte (early), Fortuna and Quintal (midseason), Primavera and Margarida (late). Geada and Fuerte fruits are pear-shaped whereas the others are rounded. The climatic conditions during the experimental period in Londrina are in Figure 1. The assessments in this study were as follows.

2.1. Fruit growth curves

Three representative plants were selected from each cultivar. Within each plant, four branches belonging to the middle third and distributed across the four quadrants of the tree canopy were selected. Every week, fruits from each tree were counted and their length and larger diameter were measured using a caliper; the average value per quadrant and per plant was obtained for each evaluation. As avocado fruits grow in both length and diameter, and this growth can occur unevenly, an index (Gi) corresponding to the sum of these two parameters was defined. Measurements were taken after the natural fruit dropping, starting November 4, 2011, and continuing until Gi values remained constant for each cultivar, when harvest time was assumed to start. The values of Gi for each cultivar were adjusted to sigmoid functions.

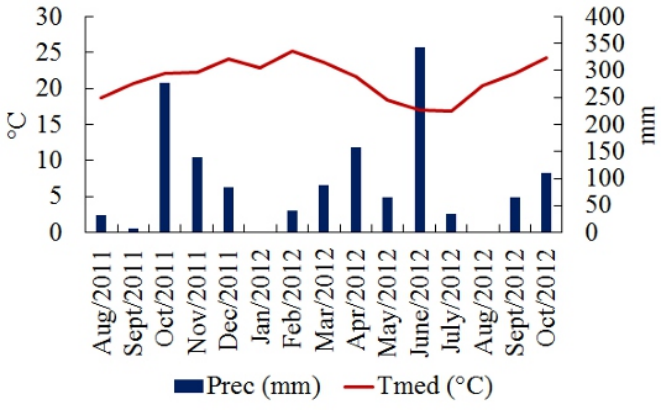


Figure 1. Monthly means of temperature and rainfall from August 2011 to October 2012 in Londrina, PR, Brazil.

2.2. Weight percentage of pulp, peel, and seed

After growth stabilization, four fruits of each cultivar were collected randomly, one from each quadrant, but avoiding the marked branches. The fruits were weighed and stored in a cool place until they were mature (indicated by their soft consistency). After maturation, fruits were weighed once again to determine the weight of the pulp, peel, and seed relative to the total weight of each fruit.

2.3. Seed starch content

A portion of seeds taken from four fruits aleatory chosen from each cultivar was dried in a forced circulation stove at 50–60 °C until it reached a constant weight. This seed portion was then ground and kept at –20 °C until further analyses. To hydrolyze seed starch, 50 mg of ground seed were transferred to a test tube with a screw cap; after adding 25 mL of 0.7 M HCl, this mix was stirred and placed in a 96 °C water bath for 90 min; following a cooling period, this solution was neutralized with 3M NaOH in a 500 mL volumetric flask to which distilled water was supplemented; the content of reducing sugars (RS) (RS fraction of the hydrolyzed starch) was determined after filtering this solution using Whatman NO.1 filter paper [14]. To perform the alcoholic extraction of glucose (RS fraction of the alcoholic extract), 200 mg of ground seed were added to 10 mL of 80% ethyl alcohol contained in a centrifuge tube. After thorough agitation, the mixture was placed in a shaking water bath at 80 °C for 30 min and then centrifuged at 804 × g; the supernatant was then collected in a beaker. This operation was repeated three times, resulting in 30 mL of alcoholic extract that was subsequently evaporated at 50 °C until there were only about 5 mL of extract left in the beaker. This residue was transferred into a 100 mL volumetric flask, the volume completed with distilled water, and the content of RS further determined. The acidic hydrolysis of sucrose (RS fraction of the hydrolyzed sucrose) was performed by adding 2.5 mL of concentrated HCl (purity of 37%) to 25 mL of the alcoholic extract diluted (to 100 mL volumetric flask) obtained during glucose extraction. This mixture was kept at room temperature for 16 h and neutralized with 3 M NaOH in a 100 mL volumetric flask to which distilled water was supplemented, before proceeding with the RS fraction determination. In all the fractions mentioned above, i.e., starch, glucose, and sucrose, the RS was determined employing the method of Somogyi and Nelson [15] as follows: 1 mL of Somogyi's reagent was added to 1 mL of each fraction, and this mixture was boiled for 10 min; after its immediate cooling on ice, 1 mL of Nelson's reagent was added and the mixture was stirred; 7 mL of distilled water was subsequently added and the mixture was again stirred. The RS percent content of each fraction was then determined by spectrophotometry at 535 nm. The standard curve was obtained with a glucose solution at 0.1% of benzoic acid (1 to 10 mg 100 g–1). All determinations were performed in duplicate. Seed starch, sucrose, and glucose contents were expressed as a percentage, according to the following equations:

Starch content = 0.9 (% RS of the hydrolyzed starch – % RS of the sucrose hydrolyzate)
Glucose content = % RS of the alcoholic extract
Sucrose content = 0.9 (% RS of the sucrose hydrolyzate – % RS of the alcoholic extract)

2.4. Fruit pulp oil content

Oil content was determined using the Soxhlet method [16]. Sample preparation consisted of collecting a part of the pulp of four fruits aleatory chosen for each cultivar. The pulp was homogenized and macerated to obtain a paste and dried in a forced circulation oven at a temperature ranging from 50 to 60 °C, until constant weight. The dried pulp was then grounded in a blender and frozen at –20 °C. One gram of this pulp was transferred to filter paper cartridges and dried again at 60 °C for 12 h. The cartridge was subsequently transferred to a Soxhlet extractor where it was kept under reflux for 16 h, using petroleum ether as the solvent extractor. After extraction, the cartridges were dried at 60 °C and weighed. Oil content was expressed in 100 g–1 of dry weight of pulp. Three replicates were performed for each cultivar.

2.5. Acidity index (Ai)

The Ai of avocado oil for each cultivar was determined using the following solution: 0.5 g of avocado oil extracted as described above were dissolved in 10 mL of ethyl alcohol, previously neutralized by titration with 0.01 N NaOH. The oil titration was carried out using phenolphthalein as an indicator and until a pink color persisted for 30 seconds [17]. The oil acidity index was determined according to the equation:

Ai (%) = (mL NaOH 0.01N × 0.282)/avocado oil weight

2.6. Peroxide index (Pi)

The Pi was determined according to the methodology described in [17] and using 1 g of the avocado oil extracted as described above. This gram of oil was dissolved in 30 mL of acetic acid: chloroform solution (3:2 volume ratio), to which 0.5 mL of saturated potassium iodide were added; after stirring this solution for 1 min, 30 mL of distilled water and 0.5 mL of a 1% starch solution were added. The starch solution was prepared by dissolving starch in distilled water, heating this solution up to 80 °C, followed by cooling titration with 0.01 N sodium thiosulfate. The peroxide index value was calculated by the following equation:

Pi (mEq kg^{-1}) * = (mL sodium thiosulfate × 1000)/avocado oil weight
* milliequivalents of active oxygen per kg of oil

2.7. Refractive index (Ri)

The analyzes were performed employing a refractometer Astral bench (mod. EEQ9001). About 3 to 4 drops of avocado oil were transferred to the prism and performed refractometer readings of the refractive index at 20 °C.

2.8. Statistical analyses

To compare oil and starch contents between cultivars, data were subjected to an analysis of variance (ANOVA) followed by Tukey's tests at the 5% probability level in order to evaluate the significance of means' differences. Principal component analysis (PCA), applying the Waard clustering method, was used to determine the parameters associated with characteristics favorable to the production of biodiesel. Analyses were performed using the statistical software XLSTAT version 2008.4.02 [18].

3. Results and Discussion

The study of the characteristics of the phenological cycle of these avocado cultivars can give indications of the periods in which there will be supply of raw material for in natura consumption or production of biodiesel, ensuring the continuity of production and commercialization throughout the year. Fruit diameter plus length (Gi) growth curves adjusted to sigmoid functions are shown in Figure 2. Fuerte has the smallest Gi and Primavera and Quintal cultivars have the highest. Table 1 shows each cultivar's adjusted equation with its respective coefficient of determination (R2) and number of days needed to complete a cycle (DCC). Considering that flowering started in August but mainly occurred in the beginning of September, the average duration of a complete cycle, from flowering to harvest, could be obtained by adding 60 days (measurements started to be taken in the beginning of November) to the values of DCC shown. The ideal commercial maturation point for avocado is not easy to be determined. Visual aspects and size of the fruit are mostly used in the field [19,20]. There is an uncertainty in using the start of ripening as a criterion, since it is not always accompanied by visual external changes of the fruit [21]. Periodical fruit size measurements in the orchard could be a good criterion to determine the exact harvesting time for each cultivar.

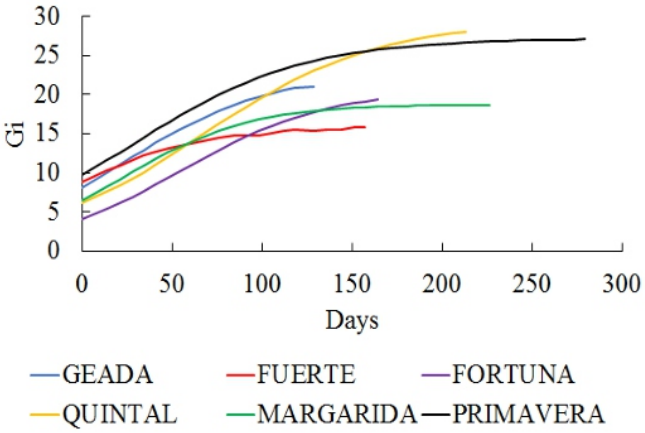


Figure 2. Sigmoid growth curves of the Growth index (Gi) for the six avocado cultivars. Measurements started on November 4th 2011, after the natural fruit dropping. Londrina, PR, Brazil.

Table 1. Sigmoid growth curves adjusted to each avocado cultivar with the respective coefficient of determination (R^2) and days to complete the fruit cycle (DCC).

Cultivar	Sigmoid Function	R^2	DCC
Geada	22.9982/(1 + exp (– (x – 24.4293)/40.9597))	0.9979	130
Fuerte	15.7753/(1 + exp (– (x + 8.3585)/35.6181))	0.9981	150
Fortuna	20.6347/(1 + exp (– (x + 55.6300)/40.0704))	0.9986	170
Quintal	20.6347/(1 + exp (– (x + 55.6300)/40.0704))	0.9986	215
Margarida	18.7321/(1 + exp (– (x + 22.4931)/34.7987))	0.9962	230
Primavera	27.1541/(1 + exp (– (x + 28.0527)/47.0111))	0.9968	280

Avocado fruits grew according to the genetic characteristic of each cultivar (Figure 1): Geada was the first to reach the maximum Gi in the beginning of March (120 days), followed by Fuerte and Fortuna in the middle of April (160 days); Quintal completed fruit growth in late May (200 days), therefore being classified as midseason. The fruits of Primavera and Margarida cultivars were the last to reach their maximum growth-late June for Margarida (240 days) and early August for Primavera (280 days).

These results are very important to the raw fruit production and market distribution (in natura consumption) as well as for the production of biodiesel from avocado. Thus, by combining the different cultivars in orchards, one can have avocado fruits available from mid-March to late September-October, considering that after reaching their maximum growth the fruits remain on plant for an extended period. This is interesting for fresh fruit consumption, as this long supply period reduces the risk of getting a low market price associated with productions concentrated in short harvesting intervals. In addition, this longer harvest period is also important for the biodiesel production industry as it implies a shorter idle

time of the processing facilities [2,11].

Yield of pulp, seeds and bark of avocado fruits as well as oil content of pulp and starch in seeds provide important information indicating potential cultivars for biodiesel production. Table 2 shows the percentages of pulp, seed, and peel of the fruits from the six cultivars. The pulp content ranged from 63.20 to 80.10%, with a coefficient of variation (CV) of 8%. The influence of the local thermal regime on the cultivars cannot be neglected, as different cultivars adapt differently to the environmental [11,22].

Table 2. Percentage of the components pulp, seed, and peel in fruits of six cultivars of avocado. Londrina-PR, Brazil.

Cultivar	Pulp (%)	Seed (%)	Peel (%)
Fuerte	63.20	24.30	12.49
Quintal	74.87	16.89	8.23
Fortuna	76.60	16.14	7.14
Geada	80.06	11.39	8.53
Primavera	72.30	15.73	11.95
Margarida	77.31	13.25	9.42
Mean	74.06	16.28	9.63
CV (%)	8%	27%	22%

The highest oil percentage was found in Fuerte, followed by Primavera (Table 3). Oliveira et al. (2013) obtained the highest percentage of oil in Fuerte and Haas in a comparison among eleven cultivars of avocado in the state of São Paulo, Brazil [20]. Found highest oil concentration in the pulp of the cultivars Hass and Fuerte and the lowest values in Fortuna and Margarida. Regarding the seed's starch content, this ranged from 45.37 to 62.22% but no significant differences were found between cultivars. The starch content values obtained in the present study for Fortuna, Fuerte, and Quintal are very close to those reported by [6]: 51.60% for Fortuna, 56.60% for Fuerte, and 37.20% for Quintal. These authors evaluated the oil (pulp) and starch (seed) contents in 24 cultivars and found that cultivars with high oil content showed a starch content above the average; our results also seem to agree with this, as the cultivar with the highest oil percentage-Fuerte-also presented the highest starch percentage (72.14 and 62.22%, respectively).

Table 3. Percentage of oil (pulp) and starch (seed) of avocado cultivars. Londrina-PR, Brazil.

Cultivars	Oil (pulp)	Starch (seed)
Fuerte	72.14 ± 2.8^{a}	62.22 ± 7.4^{a}
Quintal	55.25 ± 1.6^{c}	48.15 ± 3.6^{a}
Fortuna	55.64 ± 1.5^{c}	48.52 ± 6.7^{a}
Geada	55.84 ± 2.0^{c}	45.37 ± 4.5^{a}
Primavera	63.00 ± 1.0^{b}	45.58 ± 9.0^{a}
Margarida	$57.10 \pm 2.4^{b,c}$	59.25 ± 1.4^{a}

Means followed by same letter vertically do not differ to each other by the Tukey test at 5% probability.
Values are on a dry basis.

In previews studies, the authors pointed out that the ethanol extracted from avocado seeds can be a secondary raw material for biodiesel production [4]. The avocado seed has potential for the production of ethyl alcohol due to its high total starch content that, when subjected to a hydrolysis process, raises the Brix content of the broth providing conditions for fermentation to occur.

Acidity index, peroxide index and refractive index are evaluations of avocado oil that may reveal the quality and stability to oxidation processes of this raw material (crude oil). The average values of the acidity, peroxide, and refraction indices obtained for the avocado oil extracted from each of the six cultivars are shown in Table 4.

Table 4. Acidity index (Ai), peroxide index (Pi), and refractive index (Ri) of the oil of six cultivars of avocado.

Cultivar	Ai (%)	Pi (mEq kg^{-1})	Ri
Fuerte	0.39 ± 0.01	1.79 ± 35.9	1.469 ± 0.0
Quintal	1.16 ± 0.01	1.01 ± 48.8	1.478 ± 0.0
Fortuna	0.87 ± 0.02	0.68 ± 15.5	1.468 ± 0.0
Geada	0.78 ± 0.00	0.76 ± 47.0	1.473 ± 0.0
Primavera	0.57 ± 0.01	0.56 ± 68.2	1.468 ± 0.0
Margarida	0.63 ± 0.01	0.70 ± 13.7	1.466 ± 0.0

3.1. Acidity index (Ai)

Quintal presented the highest Ai among cultivars (1.16%) while Fuerte presented the lowest (0.39%); the other cultivars had intermediate values ranging from 0.57 to 0.87%. Acidity values indicate the amount of free fatty acids present in the oil. According to [23], the oil conservation status is related to the nature and quality of the raw material, to the quality and degree of oil purity, and to oil processing. The Ai of vegetable oils should be the lowest possible. High levels of Ai indicate alterations that prevent its use as a fuel or for human consumption [24]. Storage conditions are also very important as heat and light accelerate the decomposition of glycerides (rancidity), and this is almost always accompanied by the formation of free fatty acids [25]. These free fatty acids form fatty acid salts (soap) in basic media, which is responsible for the formation of emulsions during the purification step of biodiesel production, leading to a reduction of the yield. High acidity may catalyze intermolecular reactions, which affect the thermal stability of the fuel in the combustion chamber, and have a corrosive action on the metallic components of the engine.

3.2. Peroxide index (Pi)

The peroxide content of Fuerte was the highest among cultivars. Average values ranged from 1.79 mEq kg–1 in Fuerte to 0.56 mEq kg–1 in Primavera, and can be considered low when compared to other species [26,27]. The peroxide content in vegetal oils estimates the degree of degradability of raw material. The values found in this work are much lower then the maximum accepted values of 10 mEq kg–1 [27], and indicate therefore a low possibility of oxidative deterioration in avocado.

3.3. Refractive index (Ri)

The Ri obtained for the oil of each of the six cultivars ranged from 1.466 to 1.478. The refractive index of oils and fats is widely used as identification and quality criteria, as this index increases with increasing iodine content and can thus be used for controlling unsaturated oils hydrogenation processes [28].

3.4. Principal components analysis (PCA)

Bilateral relations between the variables account for physiological events, but when the intention is to verify the possibility of discriminating a group of samples, the best choice is the analysis by multivariate techniques. Among the several multivariate techniques, Principal Component Analysis (PCA) was used to reduce the dimensionality of a data group, explaining the variance. In this process, the original variable data is linearly transformed into a small number of uncorrelated components. From these components, the samples are projected in the plane formed by the components and their proximity to the projection of the variables allows to infer the similarities between them [29]. The PCA performed allowed to analyze the simultaneous influence of pulp, seed, and peel percentages, as well as oil and starch contents, in the Ai, Pi, and Ri of each cultivar (Figure 3), in order to discriminate avocado cultivars and verify which components, or set of components, could be used for selecting cultivars for biodiesel production. The first two components explained 81% of the total variance according to the following equations:

$$F1 = 0.87\ Ai + 0.81\ Ri - 0.66\ Oil - 0.63\ Starch + 0.88\ Peel - 0.80\ Pulp$$

$$F2 = 0.85\ Pi + 0.61\ Oil + 0.71\ Seed.$$

Peel, seed, and pulp (Table 2), and oil and starch contents (Table 3) refer to percentages of these fractions.

The principal component F1 was positively correlated to the acidity index, refractive index, and percentage of peel, and negatively correlated with oil content and starch and pulp percentage; F2 was positively correlated with the peroxide index value, oil content, and percentage of seed.

As can be observed from Figure 2, the cultivars Quintal, Geada, and Primavera are characterized by high

refractive and acidity indices, and high percentage of peel (F1 +) whereas Fuerte, Margarida, and Fortuna differed from the others by their high oil and starch contents, and high percentage of pulp (F1 –). Regarding cultivars' differentiation according to F2 +, Fuerte and Quintal cultivars showed the highest levels of oil, peroxide index, and percentage of seed.

The cultivar Fuerte presented high oil and starch content, and high percentages of seed and peel (Tables 1, 2 and 3). Although its percentage of pulp was the lowest among all cultivars, it showed the highest oil extraction and peroxide values (value still within the allowable range). This profile is interesting for obtaining biodiesel through seed and pulp processing, because the highest oil content was obtained from the lowest percentage of pulp and the highest starch content was obtained from the highest seed percentage.

The cultivars Margarida and Fortuna showed very similar profiles, with intermediate levels of oil and starch, pulp, seed, and peroxide index.

The cultivars Primavera and Geada also performed similarly to each other. They were characterized by higher refractive and acidity indices and percentage of peel, and lower values of percentage of oil, peroxide index, and percentage of seed.

The cultivar Quintal separated from the others for its larger values of refractive, acidity and peroxide indices, percentage of peel, seed, and lower oil content (Tables 1, 2 and 3). This cultivar has the least favorable profile for biodiesel production due its higher acidity and peroxide indices. This fact would result in a greater risk of oil degradation, which would imply additional steps of oil pretreatment.

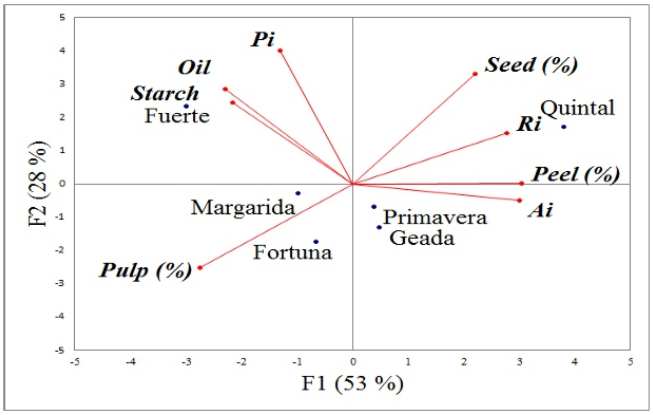


Figure 3. Biplot of the principal components 1 and 2 for oil yield, parameters analyzed in the oil, proportions of the fractions (pulp, peel and seed) and starch content in the seed of the cultivars. Pi = peroxide index, Ri = refractive index, Ai = acidity index.

In general, all cultivars showed potential to be used as a source of biodiesel given their physiochemical properties and the diversity of maturation cycles that would enable a long production period, but we can highlight the cultivar Fuerte that stood out for presenting the composition and physicochemical characteristics more suitable for potential use in the production of biodiesel. A study considering each cultivar's oil stability is thus necessary in order to combine this high potential of extraction with a higher oil quality, avoiding the use of additional steps in raw material treatment, which would increase biodiesel production costs.

4. Conclusions

The diversity of maturation cycles observed in this study shows the advantage of combining avocado cultivars with different harvest times in production orchards, enabling the market and biodiesel production supply at different times and for a longer period.

Although all avocado cultivars are potentially suitable for biodiesel production, Fuerte stands out as the most adequate for biodiesel production from pulp and seed due to its higher yield of oil and starch, when compared to the remaining five.

Acknowledgments

The authors are grateful to Fundação Araucária for the financial support to the first author, Instituto Agronômico do Paraná (IAPAR) for the experimental facilities and Universidade Estadual de Londrina.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. IBGE: Instituto Brasileiro de Geografia e Estatística, 2014. Available from: http://www.ibge.gov.br.

2. Oliveira MC, Ramos PR, Lima JD, et al. (2013) Fenologia e características físico-químicas de frutos de abacateiros visando à extração de óleo. Ciência Rural 43: 411–418.

3. Knothe G (2013) Avocado and olive oil methyl esters. Biomass Bioenerg 58: 143–148.

4. Menezes ML, Lopes LFS, Passaretti FJ (2010) Síntese de biodiesel empregando óleo de abacate. Analytica 44: 68–78.

5. Werman MJ, Neeman I (1987) Avocado oil production and chemical characteristics. J Am Chem Soc

64: 229–232.

6. Tango JS, Carvalho CRL, Soares NB (2004) Caracterização física e química de frutos de abacate visando a seu potencial para extração de óleo. Rev Bras Frutic 26: 17–23.

7. Evans LJ, Goodwin RM, Mcbrydie HM (2010) Factors affecting 'Hass' avocado (Persea americana) fruit set in New Zealand. New Zealand Plant Protect 63: 214–218.

8. Donadio LC (1995) Abacate para exportação: aspectos técnicos da produção, 2nd ed., Brasília: DENACOOP.

9. Salazar GS, Garner LC, Lovatt CJ (2013) Reproductive biology, In: Schaffer B, Wolstenholme BN, Whiley AW, The avocado: botany, production and uses, 2nd ed., Boston: CABI.

10. Wielgolaski FE (1974) Phenology in agriculture, In: Lieth H, Phenology seasonality modeling, London: Chapman & Hall, 369–381.

11. Zaro GC, Ricce WS, Caramori PH, et al. (2014) Zoneamento agroclimático para a cultura do abacateiro no Estado do Paraná. Rev Bras Frutic 36: 363–372.

12. Pereira LMP, Caramori PH, Ricce WS (2008) Determinação do início e término da estação chuvosa no estado do Paraná. Geografar 3: 1–12.

13. Caramori PH, Caviglione JH, Wrege MS, et al. (2001) Zoneamento de riscos climáticos para a cultura do café (Coffea arabica L.) no estado do Paraná. Rev Bras Agrometeorol 9: 486–494.

14. Rickard JE, Behn KR (1987) Evaluation of acid and enzyme hydrolytic methods of the determination of cassava starch. J Sci Food Agr 41: 373–379.

15. Southgate DAT (1976) Determination of food carbohydrates. London: Applied Science.

16. AOAC (2012) Association of Oficcial Anlytical Chemists. Official methods of analysis of AOAC international. Rockville.

17. Quast LB, Aquino AD (2004) Oxidação dos lipídeos em café arábica (Coffea arabica L.) e café robusta (Coffea canephora P.). CEPPA 22: 325–336.

18. ADDINSOFT (2010) XLStat: software for statistical analysis. Versão 2008.4.02 (2008). Paris. 1 CD-ROM.

19. Ozdemir F, Topuz A (2004) Changes in dry matter, oil content and fatty acids composition of avocado during harvesting time and post-harvesting ripening period. Food Chem 86: 79–83.

20. Silva FOR, Ramos JD, Oliveira MC, et al. (2014) Fenologia reprodutiva e caracterização fisicoquímica de abacateiros em Carmo da Cachoeira, Minas Gerais. Ceres 61: 105–111.

21. Martinez NL, Moreno RV (1995) Parameters for determining the maturity of avocados. IAA.

22. Soares NB, Pedro JMJ, Teixeira LAJ, et al. (2002) Tolerância a baixas temperaturas de cultivares de abacate (Persea americana Mill.). Rev Bras Frutic 24: 721–723.

23. Ribeiro EP, Seravalli EAG (2004) Química de alimentos. São Paulo: Edgard Blücler.

24. Siqueira WC, Fernandes HC, Silva AC, et al. (2014) Caracterização de misturas de óleo diesel com óleo de pinhão manso para acionamento de um trator agrícola. Reveng 22: 552–561.

25. Bhojvaid PP (2006) Raising large-scale jatropha plantation in India–present status and future prospects. In: Bhojvaid PP, Biofuels–towards a greener and secure energy future, New Delhi: TERI, 119–136.

26. Kobori CN, Jorge N (2005) Caracterização dos óleos de algumas sementes de frutas como aproveitamento de resíduos industriais. Ciênc Agrotec 29: 1008–1014.

27. Conceição LRV, Costa CEF, Rocha FGN (2011) Obtaining and characterization of biodiesel from jupati (Raphia taedigera Mart.) oil. Fuel 90: 2945–2949.

28. Cecchi HM (2003) Fundamentos teóricos e práticos em análise de alimentos, 2 nd ed. Campinas: UNICAMP.

29. Hair JF, Black WC, Babin BJ, et al. (2009) Multivariate data analysis (7th Edition), Editorial Prentice Hall.

Estimating solar irradiance using genetic programming technique and meteorological records

Rami Al-Hajj 1,* and Ali Assi 2

1 Department of Math and Statistics for Engineering, American University of the Middle East, Egaila, Kuwait 2 Department of Electrical and Electronics Engineering, International University of Beirut, Lebanon

A B S T R A C T

Solar irradiance is one of the most important parameters that need to be estimated and modeled before engaging in any solar energy project. This article describes a non-linear regression model based on genetic programming technique for estimating solar irradiance in a specific region in the United Arab Emirates. The genetic programming is an evolutionary computing technique that enables automatic search for complex solutions. The best nonlinear modeling function that can estimate the global solar radiation on horizontal will be developed taking into account measured meteorological data. A reference approach to model the solar radiation is first presented. An enhanced approach is then presented which consists of multi nonlinear functions of regression in a parallel structure where each function is designed to estimate the global solar irradiance in a specific seasonal period of the year. Statistical analysis measures have been used to evaluate the performance of the proposed approaches. The obtained results are comparable with the outcomes of models developed by other researchers in the field.

Keywords: *global solar radiation; estimation; genetic programming; evolutionary computation; meteorological data; statistical analysis; RMSE; MBE*

1. Introduction

With the increased concern and interest in energy preservation and environmental protection, the world today is moving into a new era; transition from almost total dependence of the fossil fuel to an increased use of alternative sources of energy. Solar radiation is one of the promising and potential renewable energy sources especially in regions like UAE.

An accurate and detailed long-term knowledge of the available global solar irradiance on horizontal surfaces is of a major importance for the design and development of solar energy systems in a given region. Information about solar radiation can be obtained by installing expensive measuring sensors (pyranometers) at as many locations as possible in this region thus, requiring daily maintenance and data acquisition; consequently, increasing the cost of collecting solar radiation data. In most of the cases, the potential sites for solar energy implementation are not covered by measuring stations, especially in the deserted regions. Many countries do not have sufficient network of weather stations for collecting solar data. For such regions, empirical models have to be developed using meteorological data from available measurement stations. These models are then used to estimate solar irradiance values at other locations in the region where solar energy systems are planned [2].

UAE is among countries having potential for solar energy where the solar irradiance has significant strength, the average annual solar hours is approximately 3568 h (i.e. 9.7 h/day), which corresponds to an average annual global solar irradiance of approximately 2285 kWh/m2 (i.e. 6.3 kWh/m2 per day) [2].

Numerous researchers have developed statistical and empirical regression models to predict the monthly average daily global solar irradiance in their regions using various weather parameters [3–20]. The mean daily sunshine duration and air delta temperature were the most available and commonly used parameters. The most popular model developed by researchers was the linear model by Angström-Prescott. This model establishes a linear relationship between global solar irradiance and sunshine duration taking into account extra-terrestrial solar irradiance and the theoretical maximum daily sunshine hours. Many studies with empirical regression and machine learning models were presented in the literature for many regions around the world. Recently, different models predicting global solar irradiance using various meteorological and climatological variables have been published [16–37].

Assi et al. [25,26] used four meteorological 12-years data between 1995 and 2007 to train and validate a Feed Forward ANN-based estimation system of solar radiation in Al-Ain city in UAE. The authors examined several MLP architectures and tested more than twelve alternatives based on various derivatives of back-propagation training algorithms.

Antonanzas et al. [27] presented a new methodology to build parametric models for the estimation of global solar irradiation. The models were adjusted to specific on-site characteristics based on an evaluation of the variables importance. Authors have adjusted general parametric models such as the Bristow and Campbell BC models [27] with the on-site particularities. The presented methodology was appropriate for the investigated case study. The daily range of maximum and minimum temperatures, the logical variable of rainfall, and the daily mean wind speed were among the parameters that showed higher correlation with solar irradiance and that were included in the newly developed models.

Ahmed and Adam [28] applied a feed forward back- propagation neural network on weather data measured at Qena-Egypt during the year 2007. The proposed approach used location coordinates and sunshine hours to estimate monthly average daily global solar radiation. The authors presented a comparative study between the described MLP-based approach and other empirical models. Based on their experimental results, authors showed the advantages of the MLP-based estimation technique for solar radiation estimation over the existing empirical regression models. Khatib et al. [29] developed a feed forward multi-layer perceptron with four inputs: longitude,latitude, day of the month, and sunshine radiation to predict the clearance index. The clearance index helped in calculating the solar irradiation. The models used long term solar radiation data for 28 sites in Malaysia measured between years 1984

and 2004.

Ramedani et al. [30] investigated two models based on Support Vector Regression technique (SVR) which is a type of Support Vector Machines (SVM) for predicting Global solar Radiation GSR in Tehran province. The authors examined two kernel functions for SVR: a radial basis function and a polynomial function. The authors designed and validated their approach on a measured daily data consisting on Temperature and sunshine parameters and belonging to seven-year period. The proposed approach, mainly the one based on radial basis function (SVR-rbf) showed better performance when compared to an ANN-based and a Neuro-Fuzzy based systems. Olatomiwa et al. [31] proposed a hybrid approach for predicting solar radiation based on SVMs coupled with the meta-heuristic Firefly algorithm FFA. The FFA has been applied to detect the optimal parameters of the SVM algorithm. The performance of the proposed approach showed superiority comparing to others based on ANN and GP when they have been tested on temperature and sunshine hour's data records collected from three different regions in Nigeria. Mohamadani et al. [32] presented a comparative evaluation among three soft computing methodologies for estimating global solar radiation in a specific region in Iran based on temperature measures. The developed models are an Adaptive Neuro-Fuzzy inference system ANFIS, a radial basis function SVR (SVR-rbf), and a polynomial basis function SVR (SVR-poly). The statistical analysis showed a superiority for the SVR-rbf over the two remaining examined models when validated on daily temperature measures. Kizi [33] proposed a Fuzzy-Genetic FG approach to model and predict solar radiation. The heuristic genetic algorithm has been used to find the optimum parameters of the Fuzzy inference method. The author used latitude, longitude, and altitude as inputs to the FG model to estimate one month ahead solar radiation in some regions in Turkey.

Recently, many researchers tried to instigate the most significant meteorological variables and parameters for estimating and predicting GSR [34,35]. Mohamadi et al. [34] examined the influence of meteorological parameters on horizontal GSR. They examined nine climatological parameters collected from three different cities in Iran. The authors applied an adaptive neuro-fuzzy inference technique in their selection procedure, and they determined the most influential parameters' combinations for each city and have concluded that it is not possible to introduce an optimal combination of inputs for all cities. They justified their conclusion by the fact that GSR, which is special for each region, depends on climate conditions and geographical location that are special for each region.

Demirhan and Atilgan [35] presented a robust coplot optimization approach coupled with a GP technique for solar radiation estimation. The robust coplot analysis technique has been applied on the measures of solar radiation and other related parameters to identify the optimal set of covariates in a data that consists of solar radiation, meteorological, and terrestrial variables. The main goal was to handle the

multicollinearity problem that may exist among variables, and to eliminate the effect of outliers on the space of solar radiation modeling. The optimal set of covariates have been then used in a GP technique to construct monthly and yearly solar radiation estimation models. Pan I et al. [36] presented a GP-based approach for predicting solar radiation using six geographical and sunshine duration data from India. The authors introduced what they called Multi-Gene Genetic Programming (MGGP) models where each individual solution, named a gene, is composed by a weighted combination of sub-individuals named Single-Gene Genetic Programming (SGGP) models. The authors indicated that the MGGP based approach has outperformed the other ones based on simple individual SGGP models as well as other classical regression models.

The current article investigates the prediction of global solar irradiance on horizontal using evolutional computational technique, namely the genetic programming (GP). Recently, the GP techniques showed good performance and flexibility in modelling non-linear regression problems [38,39]. Practically, The GP demonstrates its advantages in dynamically building complex formulas (solutions), and its flexibility to choose a set of functions and operators that match the problem to be solved. Such flexibility is possible due to the fact that the structure of the binary trees that represent solution candidates can be dynamically changed during the evolutionary process. These characteristics give the GP the ability to skip out of the local minima problem commonly found in the neural networks models especially in their feed-forward structures with back-propagation training algorithms. In the solar radiation estimation literature, the Genetic Algorithm GA has been used to select the optimal parameters of machine learning based models [33], whereas the GP algorithm has been used as core models of estimation [35,36].

In this article, the design and validation of a new GP based approach to estimate global solar irradiance using meteorological data will be described. The main idea is to find the best model for the relation between a set of meteorological parameters and the solar irradiance on a specific geographical area. Two approaches have been validated: A reference approach that consists of one global model that estimates the solar irradiance with respect to four climatological parameters, and a second approach that consists of a set of several models in a parallel structure. Each model consists of a nonlinear function that is dedicated to estimate the global solar irradiance in a specific seasonal or bi-seasonal period of the year. The experimental results indicated the advantages of using such type of multi-model structure when dealing with a set of data with large variability during the year. The remaining of this article is organized as follows: In section 2, the genetic programming is explained as an optimization heuristic technique, the GPLAB toolbox of MATLAB® that has been used in the adopted approach is then introduced. In section 3, the GP based reference approach that consists of one estimation function is described. Then, an enhanced approach also based on GP is given. Moreover, the dataset used for the design and the validation of the proposed approaches is introduced and described in this section. Discussion of the

results is presented in section 4. Finally, section 5 includes conclusions and future perspectives.

2. Genetic Programming and GPLAB Toolbox

2.1. Genetic Algorithm

The GP is an extension of the conventional genetic algorithm [40,41]. Genetic Algorithm (GA) is a metaheuristic method usually used to find an optimal solution in optimization problems based on a natural selection process. The GA starts by an initial population of random individuals. Each individual is represented by what is called chromosome that is an array of genes. Each gene represents a parameter to be optimized. Every individual (chromosome) represents a possible solution of the optimization problem and has its own fitness measure. The fitness is a measure for each individual that indicates how the solution related to this individual is suitable to solve the problem.

The GA uses the so called genetic operators: crossover, mutation, and cloning to evolve from a population to another until it reaches a population that consists of an optimal solution based on a chosen fitness objective function [40,41]. It starts by an initial random population of individuals.

Then, in each generation, the GA performs the following steps: -
Select from the present population the individuals that have the best computed fitness. - -
Use the best individuals to generate the next population by the crossover of those individuals.
This is similar to the biological reproduction based on natural selection. A new individual has part of its genes coming from the first parent and the other part from the second parent.
Based on a computed probability, a mutation operator may be applied to one or many chromosomes by changing the value of one of its genes. Similarly, and based on another computed probability, a chromosome with good fitness may be cloned and promoted to evolve to the next generation.

This procedure is repeated until an optimum individual is found. Figure 1 illustrates the operations of crossover and mutation of two individuals to produce new individuals for the next generation.

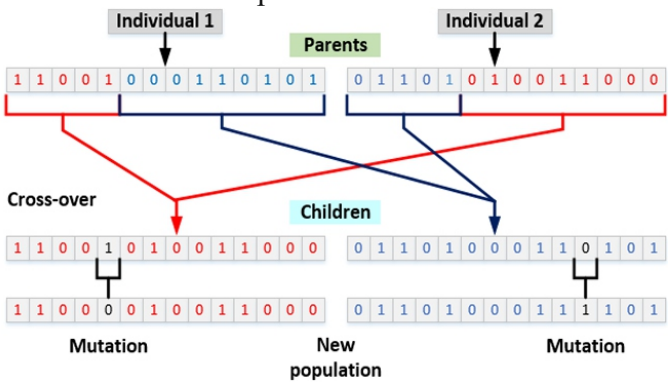


Figure 1. The crossover and mutation operators of the GA. The crossover produces two new children individuals from two parent individuals, and the mutation changes randomly the values of one or more than one gene.

2.2. Genetic Programming

The GP aims to find the best computer program (function) that is composed of both data and operators and that solves a specific problem [41,42,43]. A chromosome in GP is represented by a binary tree data structure where internal nodes represent algebraic and/or logical operators whereas the external ones represent numbers and parameters related to the problem to be solved. Figure 2 shows examples of binary trees that represent mathematical expressions/functions. A function can be considered as a computer program that consists of a set of data (terminals) and actions (operators).

The process of evolution in GP starts by an initial random population of chromosomes that represent possible solutions (functions) and tries to generate new populations subsequently [42].

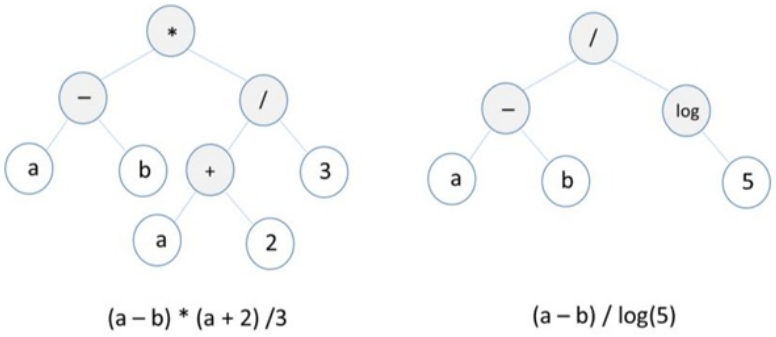


Figure 2. Examples of binary trees that represent algebraic functions.

The evolution is controlled by a fitness function that is equivalent to the objective function adopted in local heuristic search techniques. The fitness function is special for each optimization problem and allows the evaluation of fitness of each chromosome (solution) in a population. Figure 3 illustrates the effect of the crossover operator on two selected individuals.

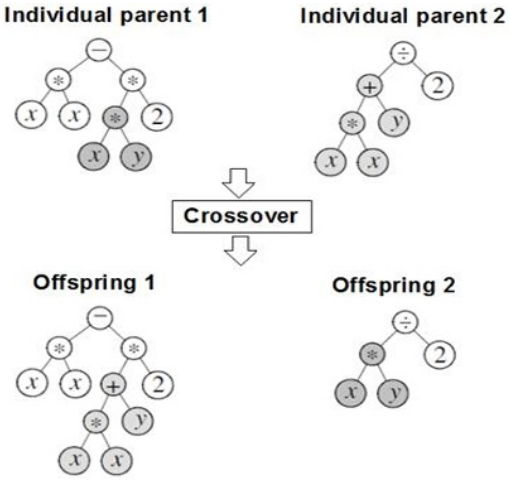


Figure 3. The crossover of two individuals to produce two new ones for the next generation.

2.3. GPLAB toolbox

GPLAB is a Genetic Programming toolbox for MATLAB® [44]. GPLAB provides most of the features and operators commonly used in GP. Its modular structure allows considering it as an extendable tool that is suitable for prototyping new techniques of heuristic local search in GP. GPLAB enables a set of facilities and features to handle and control the structure and the size of both: the chromosomes that represent individual solutions and the populations that represent sets of those individuals. In addition, GPLAB allows the dynamic control of the variable size of populations during run time. This feature is indeed important in case of limited computational resources [45,46,47].

Moreover, GPLAB implements a technique for automatically adjusting the probabilities of adopted genetic operators during runtime. This feature allows the use of the GPLAB toolbox as a test workbench for new genetic operators

3. Genetic Programming Based Systems

In this work, a GP based approach to estimate solar irradiance is designed, implemented, and validated. In the first phase of this work, a reference system is proposed, which consists of a single function that can model the relation between solar irradiance and a set of climatological factors in a specific geographical area. The reference system showed promising performance. In the second phase, the performance of the reference system is analyzed and an enhanced one that consists of multiple independent models is suggested. Each model is dedicated to estimate the solar radiation amount in a specific seasonal period of the year. All components of the two proposed approaches were designed, implemented, and validated using the GPLAB toolbox.

In the design and validation phases, a meteorological dataset provided by the National Center of meteorology and Seismology (NCMS) in Abu-Dhabi—UAE is used. The dataset consists of daily data records for the period between 2004 and 2007. Each daily record includes the measures of: air temperature, wind speed, relative humidity, and sunshine duration. The dataset has been divided into two subsets; a design subset having records for the years between 2004 and 2006 inclusive and a test subset that includes records for the year 2007. Table 1 shows some samples of the dataset. In this table, the first four columns represent the four meteorological records.

Table 1. Samples of meteorological records (dataset).

Max Temperature	Mean wind Speed (knot)	Sun Hours	Mean Relative Humidity %	Total Radiation (kWh/sq.m)
25.7	5.6	8.9	68	462.8
24.5	7.1	9.6	67	459.3
23.0	5.4	10.2	54	515.5
25.1	6.4	10.1	47	505.0
23.6	8.3	1.6	63	230.6

3.1. The Reference System

The problem of developing an appropriate function that models the relation previously discussed, looks like a search problem for an optimal state that represents the expected function. In our case, the targeted function includes operand and operators. The operands consist of the climatological parameters and other constants that may appear in the resulting function. Besides, the set of operators may include arithmetic operators, exponential operators, and any other algebraic or non-algebraic ones like: square root, natural log, exponentials, etc.

The functions investigated in this work can be represented by binary trees data structures, where the internal nodes represent the operators whereas the terminals represent the operands. Figure 4 illustrates an example of a binary tree that represents a solution candidate. In Figure 4, X1, X2, X3, and X4 represent temperature, wind speed, sun hours, and humidity respectively.

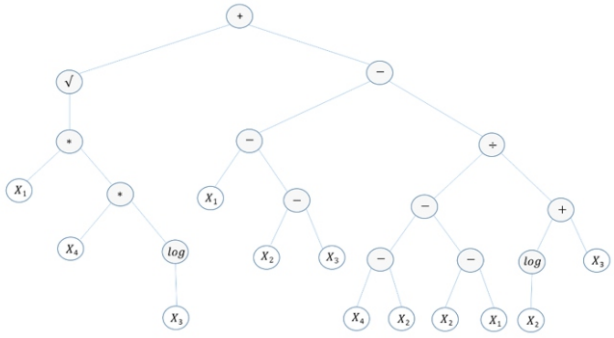


Figure 4. An example of a binary tree that represents a function for modeling (estimating) solar radiation.

Equation (1) shows the algebraic function that can be obtained using the binary tree of Figure 4. In this equation, sr, tmp, sh, ws, and hum stand for solar radiation, temperature, sun hours, wind speed, and humidity respectively. The internal nodes in the tree include the four arithmetic operators: addition (+), difference (-), and division (/), and multiplication (*), plus it includes the decimal log, as well as the square root ( ). Figure 4 shows that the depth of the tree is equal to six. Actually, the maximum depth

allowed for trees in each population is one of the adjustable parameters in a GP evolution process.

$$sr = \sqrt{tmp * hum * \log(sh)} + (tmp - (ws - sh)) - \left(\frac{(hum - ws) - (ws - tmp)}{\log(ws) + sh}\right) \quad (1)$$

In this work, fixed values for the following parameters have been adopted:

- The population size is set to be equal to 500 individuals. In our experiments the performance of optimization has been slightly affected by variations in the size of population.
- The fitness function that evaluates the efficiency of candidate solutions (chromosomes). A fitness function related to the root mean square error (RMSE) has been used.
- The depth of each of the binary trees that represents chromosomes (solution candidate) is set to be dynamic. The maximum value of that depth is chosen to be equal to six. GPLAB provides a technique that permits to start by an initial depth of trees that can be dynamically increased until a selected maximum value.

On the other hand, alternatives have been investigated:

- The probability of applying each of the genetic operators (crossover and mutation).
- The sampling method to select individuals from the current population to participate in generating new individuals for the next generation.

The first three columns in Table 2 show combinations of parameters adopted in designing the proposed approach. The implemented fitness function computes, for each individual, the RMSE between the set of exact output values available in the design dataset and the output values returned by that individual. Equation (2) describes the RMSE computation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(H_{pi} - H_i)^2}{N}} \quad (2)$$

Where H_{pi} represents the estimated value of global solar irradiance, H_i is the measured value that is available in the design dataset, and N is the total number of records in that dataset.

Table 2. Combinations of parameters with the fitness of best individuals in the last population for each combination.

Operators probability	Sampling method	Set of functions	Best fitness in last population
Crossover prob. : 0.85 Mutation prob. : 0.15	Roulette	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.402
Crossover prob. : 0.85 Mutation prob. : 0.15	Tournament	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.491
Dynamic	Roulette	{arithmetic , $\sqrt{}$, $\log_{10}$}	4.503
Dynamic	Tournament	{arithmetic , $\sqrt{}$, $\log_{10}$}	**4.174**

As for the probabilities of applying each of the genetic operators, the GPLAB allows either to fix the values of those probabilities or to dynamically compute them at each iteration during the run time. The computation in this case is based on the history of each operator in producing individuals with the best fitness and on statistics about the newly produced individuals [44]. The results presented in the right most column of Table 1 indicate that the best fitness value (in this case the lowest) is the one related to

dynamic probabilities of operators and tournament sampling method. Figure 5 shows the binary tree associated to the best individual in the last population of the best combination. Equation (3) represents the function stored in that binary tree.

$$sr = tmp + sh + sh * \log(sh) - \log|tmp - hum| + sh - \left(\frac{hum - ws}{tmp + ws}\right) \quad (3)$$

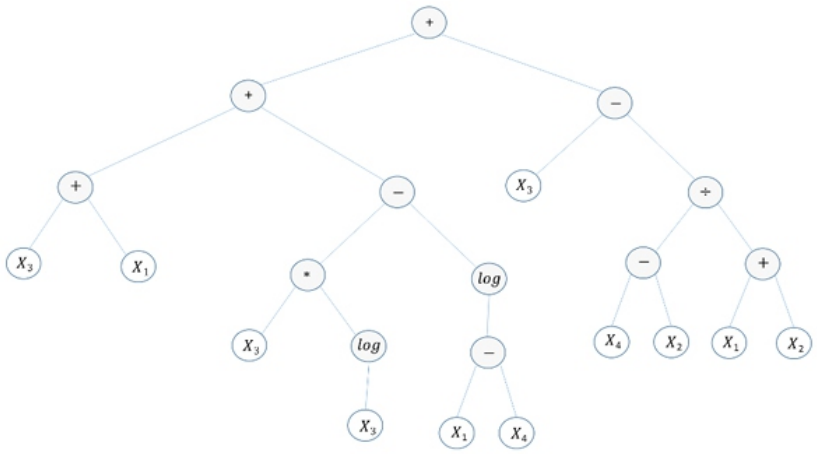


Figure 5. Binary tree related to the function of the fittest individual, where X_1 = temperature, X_2 = wind speed, X_3 = sun hours, and X_4 = humidity.

3.2. The Multi-Model System

The results obtained using the reference system show remarkable difference between the measured values and the estimated ones. Figure 7 compares the measured and the estimated monthly average daily global solar irradiance values.

The records of the design dataset have been investigated, and the values of each meteorological factor have been analyzed. Analysis showed that the values of some factors, especially the humidity, have wide variations, i.e. a large deviation around the average value over a year. Such variations make the search for an optimal model quite difficult.

One of the suggestions to improve the whole performance is by estimating the global average solar irradiance over relatively short period of time in a year by using a multi-model approach. The main idea is to find the function with best fitness for estimating the global solar irradiance for each seasonal period of the year. Applying this strategy lead to build proficient functions. A function has been built for each two consecutive months of the year. Thus, the multi-mode system consists of six nonlinear functions. Figure 6 illustrates the structure of the proposed approach.

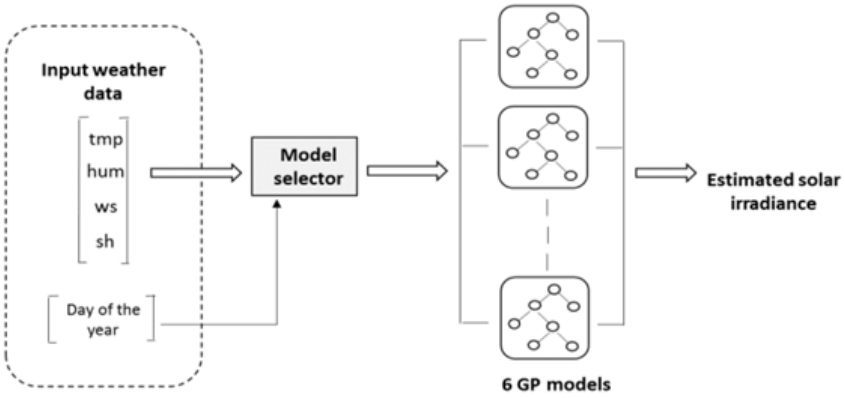


Figure 6. Structure of the multi-model approach.

The evolutional computation process is launched with the same combination of parameters described in the 1st and 4th rows of Table 2. The estimation performance is significantly improved. Table 3 shows the obtained enhancement in terms of fitness of the best individual when the multi-model strategy is applied. The best performance is obtained with dynamic probabilities of genetic operators, tournament sampling, and set of functions that contains arithmetic, algebraic and logarithmic operators.

Table 3. The fitness of the best individual of the last population for the multi-model approach.

Operators probability	Sampling method	Set of functions	Best fitness in last population
Crossover prob. : 0.85 Mutation prob. : 0.15	Roulette	{arithmetic , $\sqrt{\ }$, $\log_{10}$}	2.137
Dynamic	Tournament	{arithmetic , $\sqrt{\ }$, $\log_{10}$}	**1.218**

Error statistical analysis showed good improvement as will be described in the next section.

4. Results

The estimation performance of the suggested approach was assessed through a statistical analysis of error. The analysis was conducted by computing the RMSE and Mean Bias Error MBE that measure the variation of estimated values against the measured available ones. Low RMSE and MBE values are desired and indicate an accurate estimation. The RMSE computation is described earlier in equation (2), whereas the MBE computation is described in equation (3).

$$MBE = \frac{\sum_{i=1}^{N}(H_{pi} - H_i)}{N} \quad (3)$$

Table 4 compares the values of RMSE of the best two reference models and the new model that consists of parallel multi-functions.

Table 4. The RMSE and MBE of the best functions.

Operators probability	Best fitness in last population	RMSE	MBE
Operators prob.: [0.85, 0.15] Sampling: roulette Model: reference model	4.402	1.862	1.322
Operators prob. : dynamic Sampling: tournament Model: reference model	4.174	1.026	0.68
Operators prob. : dynamic Sampling: tournament Model: Parallel	**1.218**	**0.210**	**0.052**

Figure 7 and Figure 8 show the measured and estimated values of monthly average daily global solar irradiance for the best reference model and the multi- model. The later shows better performance in estimating monthly average daily global solar irradiance.

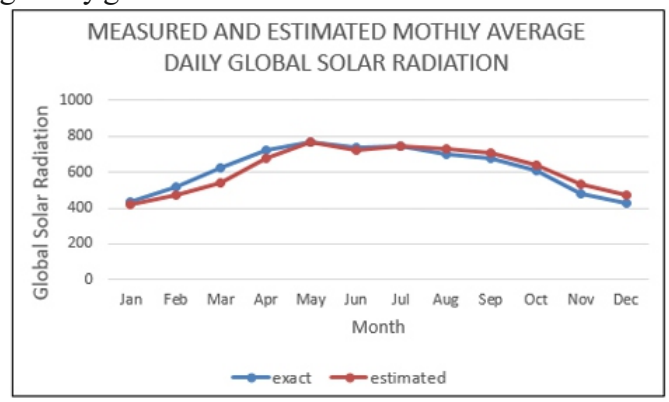


Figure 7. Monthly average daily global solar irradiance estimation of the best reference system.

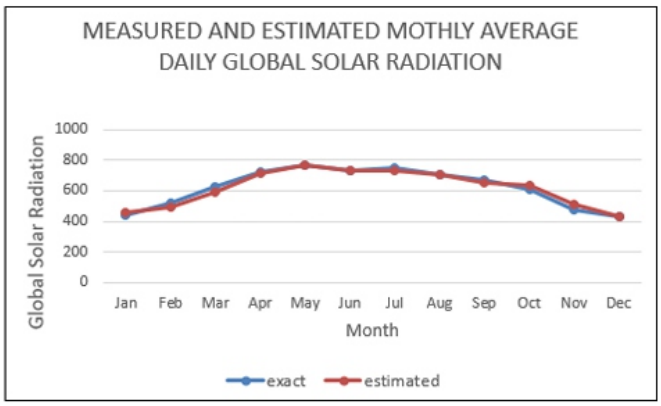


Figure 8. Monthly average daily global solar irradiance estimation of the multi-model system.

Figure 9 shows the binary tree that represents one of the best functions in the multi-model approach. Equation (4) represents the function of the binary tree shown in this Figure.

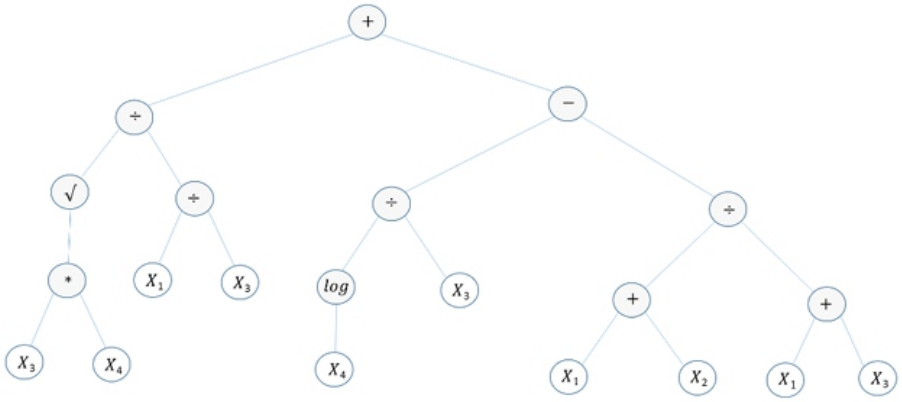


Figure 9. Binary tree that represents one of the functions in the multi-model system.

$$sr = \frac{\sqrt{sh * hum}}{\frac{tmp}{sh}} + \frac{log\ (hum)}{sh} - \frac{tmp + ws}{tmp + sh} \tag{4}$$

The suggested genetic programming based approaches show comparable performance with respect to other empirical regression and neural models. Table 5 compares the RMSE of the results obtained by the suggested approach to those obtained by other models conducted by other groups.

Table 5. The RMSE of the multi-functions model compared to other models.

Model	RMSE
Quadratic Regression model	0.214
Logarithmic Regression model	0.259
ANN 3-10-4-1	0.374
ANN 3-20-8-1	0.391
GP model with parallel functions	**0.210**

5. Conclusions and Perspectives

This article described new approaches for estimating global solar irradiance using meteorological records. The suggested methods are based on a GP heuristic technique. The first method (reference system) consists of estimating the nonlinear function that can model the relation between solar irradiance and four meteorological parameters. The performance of the reference system is promising. An enhanced model that consists of multi-function was proposed and it showed better performance with respect to the first method. The performance of the proposed approaches was evaluated using statistical

analysis measures.

The GP showed its advantages in dynamically building complex formulas that represent solution candidates for the problem to be solved. As an evolutionary process, the GP shows its ability to resolve the local minima problem and to converge toward a global minima. The problem of local minima is commonly found in the neural networks models especially in their feed-forward structures commonly used in the literature coupled with the classical back-propagation training algorithm. Moreover, the GP technique provides analytical expressions as solutions, like the expression given in Equation (3), which is not available in most of the machine learning techniques, for instance the neural and the neuro-fuzzy models. The later property is important for researchers to understand the contribution of each variable input in the calculation of the dependent variable output.

In our experiments, we controlled the well-known bloat (inflation) problem of GP by using a set of techniques provided by the GPLAB environment [44]. Some of those techniques consist of automatic resizing of the population in runtime to save computational resources. Those techniques are adequate in cases of complex problems when the complexity of the expressions' models increases dramatically during the evolutionary process.

A similar approach to our enhanced one has been proposed in [35] with two main differences: First, in that approach the input data has been pre-analyzed by using an optimization technique to handle the multicollinearity problem that may exist among the variables, which is not available in our approach that has been applied on data records of four meteorological variables. Second, the approach in [35] consists of a set of twelve models, each model is dedicated to estimate the solar radiation in a specific month of the year whereas our approach consists of six models. Each of our parallel models is devoted to estimate the solar radiation for a semi-seasonal period of the year. In general, the increase of number of learning-based models requires more training data records to adjust those models during the design phase which may not be always possible. One of our future suggestions is to automate the splitting of data into seasonal subsets based on an automatic learning technique in order to optimize the estimation performance.

In this work, the proposed GP approach is not compared to other learning based estimation techniques of the same type. Such comparison needs comparing the convergence time of each approach as well as the performance of estimation using the same data sets and same leaning parameters. This could be one of our future perspectives. On the other hand, the obtained results in this work are comparable to those obtained by mathematical regressions and neural models that were conducted by other research groups. Finally, the obtained results showed the advantage of using the parallel modular structure over the global

one.

Three main future perspectives can be drawn for this work. The first one consists of finding a way to automatically splitting the data into seasonal or semi seasonal periods in order to optimize the performance of estimation. The second perspective consists of comparing or GP based approach with ther machine learning based techniques by using the same data sets. The third perspective consists of validating the proposed approach using new meteorological datasets with larger number of weather parameters in each record.

Acknowledgments

The authors would like to thank the National Center of Meteorology and Seismology (NCMS), Abu Dhabi for providing the weather data.

Conflict of Interest

All authors declare on conflicts of interest in this paper.

References

1. Kassem AS, Aboukarima AM, EL Ashmawy NM (2009) Development of Neural Network Model to Estimate Hourly Total and Diffuse Solar Radiation on Horizontal Surface at Alexandria City (Egypt). J Appl Sci Res 5: 2006-2015.

2. Assi A, Jama M (2010) Estimating Global Solar Radiation on Horizontal from Sunshine Hours in Abu Dhabi –UAE. Advances in Energy Planning, Environmental Education and Renewable Energy Sources, 4th WSEAS international Conference on Renewable Energy Sources, 101-108.

3. Podestá G, Núñez L, Villanueva C, et al. (2004) Estimating daily solar radiation in the Argentine Pampas. Agr Forest Meteorol 123: 41-53.

4. Almorox J, Benito M, Hontoria C (2008) Estimation of global solar radiation in Venezuela. Interciencia 33: 280-283.

5. Falayi E, Adepitan J, Rabiu A (2008) Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria. Int J Phys Sci 3: 210-216.

6. Fortin J, Anctil F, Parent L, et al. (2008) Comparison of empirical daily surface incoming solar radiation models. Agr Forest Meteorol 148: 1332-1340.

7. Togrul I, Togrul H (2002) Global solar radiation over Turkey: Comparison of predicted and measured data. Renew Energy 25: 55-67.

8. Walthall C, Dulaney W, Anderson M, et al. (2004) A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. Remote Sens Environ 92: 465-474.

9. Bakirci K (2009) Correlation for estimation of daily global solar radiation with hours of bright sunshine in Turkey. Energy 34: 485-501.

10. Tadros M (2000) Uses of sunshine duration to estimate the global solar radiation over eight meteorological stations in Egypt. Renew Energy 21: 231-246.

11. Al-Lawati A, Dorvlo A, Jervase J (2003) Monthly average daily solar radiation and clearness index contour maps over Oman. Energ Convers Manage 44: 691-670.

12. Zhou J, Yezheng Wu, Gang Y (2005) General formula for estimation of monthly average daily global solar radiation in China. Energ Convers Manage 46: 257-268.

13. Ball R, Purcell C, Carey S (2004) Evaluation of Solar Radiation Prediction Models in North America. Agron J 96: 391-397.

14. Menges H, Ertekin C, Sonmete M (2006) Evaluation of solar radiation models for Konya, Turkey. Energ Convers Manage 47: 3149-3173

15. Şahin A (2007) A new formulation for solar irradiation and sunshine duration estimation. Int J Energy Res 31: 109-118.

16. Ulgen K, Hepbasli A (2002) Comparison of solar radiation correlations for Izmir, Turkey. Int J Energy Res 26: 413-430.

17. Angström A (1924) Solar and terrestrial radiation. Quart J Roy Met Soc 50: 121-125.

18. Boccol M, Willington E, Arias M (2010) Comparison of Regression and Neural Networks Models to estimate Solar Radiation. Chilean J Agr Res 70: 428-435.

19. Mohandes M, Rehman S, Halawani T (1998) Estimation of Global Solar Radiation Using Artificial Neural Networks. Renew Energy 14: 179-184.

20. Mohandes M, Balghonaim A, Kassas M, et al. (2000) Use of Radial Basis Functions for Estimating Monthly Mean Daily Solar Radiation. Solar Energy 68: 161-168.

21. Rehman S, Mohandes M (2008) Artificial neural network estimation of global solar radiation using air temperature and relative humidity. Energy Policy 36: 571-576.

22. Tasadduq I, Rehman S, Bubshait K (2002) Application of neural networks for the prediction of hourly mean surface temperature in Saudi Arabia. Renew Energy 25: 545-554.

23. Krishnaiah T, Srinivasa RS, Madhumurthy K, et al. (2007) A Neural Network Approach for Modelling Global Solar Radiation. Appl Sci Res 3: 1105-1111.

24. Elminir H, Areed F, Elsayed T (2005) Estimation of solar radiation components incident on Helwan site using neural networks. Solar Energy 79: 270-279.

25. Assi A, Al-Shamisi M, Jama M (2010) Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks, Proceedings of the 4th International Conference

on Renewable Energy Sources , Tunisia, 109-113.

26. Assi A, Al-Shamisi M (2010) Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks. Proceedings of the 25th European Photovoltaic Solar Energy Conference, Spain, 508-512.

27. Antonanzas-Torres F, Sanz-Garcia A, Martınez-de-Pison F, et al. (2013) Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain. Renew Energy 60: 604-614.

28. Ahmed A, Adam M (2013) Estimate of Global Solar Radiation by using Artificial Neural Nework in Qena, Upper Egypt. J Clean Energy Technol 1(2): 148-150.

29. Khatib T, Mohamed A, Mahmoud M, et al. (2012) Estimating Global Solar Energy Using Multilayer Perception Artificial Neural Network. Int J Energy 6(1): 82-87.

30. Ramedani Z, Omid M, Keyhani A, et al. (2014) Potential of radial basis function based support vector regression for global solar radiation prediction. Renew Sust Energy Rev 39:1005-1011.

31. Olatomiwa L, Mekhilefa S, Shamshirband S, et al. (2015) A support vector machine–firefly algorithm-based model for global solar radiation prediction. Solar Energy 115: 632-644.

32. Mohammadi K, Shamshirband S, Danesh AS, et al. (2016) Temperature-based estimation of global solar radiation using soft computing methodologies. Theor Appl Climatol 125: 101-112

33. Kisi O (2014) Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. Energy 64: 429-436

34. Mohammadi K, Shamshirband S, Kamsin A, et al. (2016) Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. Renew Sust Energy Rev 63: 423-434.

35. Demirhan H, Atilgan Y (2015) New horizontal global solar radiation estimation models forTurkey based on robust coplot supported genetic programming technique. Energy Convers Manage 106: 1013-1023.

36. Pan I, Pandey DS, Das S (2013) Global solar irradiation prediction using a multi-gene genetic programming approach. J Renew Sust Energy 5: 063129.

37. Baser F, Demirhan H (2017) A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. Energy 123: 229-240.

38. Schwaerzel R, Bylander T (2006) Predicting currency exchange rates by genetic programming with trigonometric functions and high-order statistics. Genetic and Evolutionary Computation Conference Gecco'06 1: 955-956.

39. Agapitos A, Dyson M, Kovalchuk J, et al. (2008) On the genetic programming of time-series predictors for supply chain management, proceedings of the 10th annual conference on genetic and evolutionary computation, Atlanta 1: 1163-1170.

40. Srinivas M, Patnail L (1994) Genetic algorithms, A survey. IEEE Computer 27(6): 17-26.

41. Poli R, Langdon W, McPhee N, et al. (2007) Genetic programming, An introductory tutorial and a survey of techniques and applications. University of Essex, UK, Tech Rep CES-475. Available from: http://cswww.sx.ac.uk/staff/rpoli/technical-reports/tr-ces-475.pdf

42. Poli R, Langdon WB, McPhee NF (2008) A field guide to genetic programming. Available from: http://www.gp-field-guide.org.uk/

43. Riolo R, Vladislavleva E, Moore J (2011) Genetic programming theory and practice IX. Springer Science & Business Media, ISBN 1461417708.

44. Silva S, Almeida J (2003) A Genetic programming toolbox for MATLAB, Version 3. ECOS Evolutionary and complex Systems Group, University of Coimbra-portugal.

45. Silva S, Costa E (2005) Resource-Limited genetic programming: The dynamic approach. Proceedings of GECCO-2005, 1673-1680.

46. Silva S, Costa E (2004) Dynamic limits for bloat control - Variations on Size and Depth. Proceedings of GECCO'04, 666-677.

47. Silva S, Almeida J (2003) Dynamic maximum tree depth - A Simple technique for avoiding bloat in tree-based GP. Proceedings of GECCO-2003, 1776-1787.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Note