Abstract
An initial study has recently been presented of surrogate-assisted evolutionary algorithms used to design vertical-axis wind turbines wherein candidate prototypes are evaluated under fan-generated wind conditions after being physically instantiated by a 3D printer. Unlike other approaches, such as computational fluid dynamics simulations, no mathematical formulations were used and no model assumptions were made. This paper extends that work by exploring alternative surrogate modelling and evolutionary techniques. The accuracy of various modelling algorithms used to estimate the fitness of evaluated individuals from the initial experiments is compared. The effect of temporally windowing surrogate model training samples is explored. A surrogate-assisted approach based on an enhanced local search is introduced; and alternative coevolution collaboration schemes are examined.

Keywords
3D printing, coevolution, fitness approximation, neural network, partnering.

1 Introduction
Renewable energy contributed over half of total net additions to global electric generating capacity from all sources in 2012, with wind power accounting for around 39% of the renewable power added (REN21, 2013, p. 13). Currently, arrays of horizontal-axis wind turbines (HAWTs) are the most commonly used form of wind farm employed to extract large amounts of wind energy. However, as the turbines extract the energy from the wind, the energy content decreases and the amount of turbulence increases downstream from each. For example, see Hasager et al. (2013) for photographs and explanation of the well-known wake effect at the Horns Rev offshore wind farm in the North Sea. Because of this, HAWTs must be spaced 3–5 turbine diameters apart in the cross-wind direction and 6–10 diameters apart in the downwind direction in order to maintain 90% of the performance of isolated HAWTs (Dabiri, 2011). The study of these wake effects is therefore a very complex and important area of research (Barthelmie et al., 2006), as is turbine placement (Mosetti et al., 1994). Thus, “modern wind farms comprised of HAWTs require significant land resources to separate each wind turbine from the adjacent turbine wakes. This aerodynamic constraint limits the amount of power that can be extracted from a given wind farm footprint. The resulting inefficiency of HAWT farms is currently compensated by using taller wind turbines to access greater wind resources at high altitudes, but this solution comes at the expense of higher engineering costs and materials.”
greater visual, acoustic, radar and environmental impact” (Dabiri, 2011, p. 1). This has forced wind energy systems away from high energy demand population centres and toward remote locations with higher distribution costs.

In contrast, vertical-axis wind turbines (VAWTs) do not need to be oriented to wind direction and the spacing constraints of HAWTs often do not apply. VAWT performance can even be increased by the exploitation of interturbine flow effects (Charwat, 1978). In addition, VAWTs can also be easier to manufacture, may scale more easily, are typically inherently lightweight with little or no noise pollution, and are more able to tolerate extreme weather conditions (Eriksson et al., 2008). This has resulted in a recent expansion of their use in urban environments (Toja-Silva et al., 2013). However, their design space is complex and relatively unexplored. Generally, two classes of design are currently under investigation and exploitation: Savonius (1930), which has blades attached directly upon the central axis structure; and Darrieus (1931), where the blades—either straight or curved—are positioned predominantly away from the central structure. Hybrids also exist. The small body of previous work considering VAWT farms/arrays has used turbines originally intended to operate alone. Our work is the only known work to consider designing VAWTs in arrays of interacting turbines.

We have presented an initial study (Preen and Bull, 2014a) of surrogate-assisted genetic algorithms (SGAs; Dunham et al., 1963) used to design VAWTs wherein candidate prototypes are evaluated under fan-generated wind conditions after being physically instantiated by a 3D printer. That is, unlike other approaches, no mathematical formulations are used and no model assumptions are made. Initially, artificial evolution was used to explore the design space of a single isolated VAWT, and subsequently a cooperative coevolutionary genetic algorithm (CGA; Husbands and Mill, 1991) was applied to explore the design space of an array of two closely positioned VAWTs. Both conventional CGA and surrogate-assisted SCGA versions were examined; increased aerodynamic performance (rotational speed) in fewer fabrications was found with surrogate assistance. For single turbine comparison, the fittest evolved designs were found to be aerodynamically more efficient than several common human designs under the experimental conditions. In this paper, we extend that work by exploring alternative surrogate modelling and evolutionary techniques. First, the accuracy of various modelling techniques used to estimate the fitness of the individuals from the initial experiments is compared. Subsequently, we compare surrogate model performance with different training samples. An alternative surrogate approach based on an enhanced local search is introduced. Finally, alternative coevolution collaboration schemes are examined, including one that considers the potential for symmetry within the task.

2 Background
2.1 Interacting Vertical-Axis Wind Turbines

Arrays of closely spaced VAWTs have long been considered for use as wind power stations. For example, Charwat (1978) observed an improvement in the performance of a pair of closely spaced S-shaped VAWT (S-rotors), whether counter or co-rotating, compared with that of a single turbine. Figure 1 illustrates the possible rotation configurations of a pair of VAWTs. Despite this, VAWTs have been restricted to niche applications because a single HAWT provides a much higher efficiency compared with a single VAWT. However, Kinzel et al. (2012) performed an experimental field study of an array of nine pairs of full-scale counter-rotating VAWTs. They found that the
wind velocity behind a turbine pair recovers to 95% of the wind velocity upwind after approximately 6 turbine diameters, compared with 4 diameters for the wake behind a single VAWT and 14 diameters for HAWTs. Thus, closely spaced VAWTs can result in an overall reduction in the average interturbine spacing as well as increasing individual performance, leading to a much greater power density. Indeed, it has recently been shown (Dabiri, 2011) that power densities an order of magnitude greater can potentially be achieved by arranging VAWTs in layouts utilising counter-rotation that enable them to extract energy from adjacent wakes and from above the wind farm.

A growing body of work has explored techniques to optimise a given wind farm layout—termed micrositing. Algorithms used include genetic algorithms (GAs; Mosetti et al., 1994), evolution strategies (ESs; Kusiak and Song, 2010), particle swarm optimisation (Aristidis et al., 2010), ant colony optimisation (Eroğlu and Seçkiner, 2012), Monte Carlo simulation (Marmidis et al., 2008), and principles from fish schooling (Whittlesey et al., 2010). See Salcedo-Sanz et al. (2011) for a review of evolutionary computation–based techniques for micrositing. Importantly, all of this work has been based on wake models of varying degrees of fidelity.

To date, heterogeneity in wind farms has been almost completely unexplored. Chamorro et al. (2014) investigated a variable-sized HAWT array composed of $3 \times 8$ model wind turbines where large and small turbines were alternately positioned. They found that size heterogeneity has positive effects on turbulent loading as a result of the larger turbines facing a more uniform turbulence distribution and the smaller turbines operating under lower turbulence levels. The interactions show the possibility that heterogeneity within wind farms has the potential to improve the overall ability to harvest energy. Our initial study (Preen and Bull, 2014a) observed that VAWT array asymmetry can be more efficient than similar symmetrical designs; for example, the individuals from the fittest SCGA array pairing after 160 fabrications were duplicated to form homogeneous arrays, and the maximum combined rotational speed was found to be lower than the heterogeneous array.
2.2 Evolutionary Design

Evolutionary algorithms (EAs) have long been used to design 3D physical objects. Notably, Hornby et al. (2011) evolved and manufactured an X-band satellite antenna for NASA’s ST5 spacecraft, representing the world’s first artificially evolved hardware in space. Significantly, the evolved antennas outperformed a design hand-produced by the antenna contractor for the mission. Most approaches, however, have used simulations to provide the fitness scores of the evolved designs before final fabrication.

Most of blade design optimisation is performed through the use of computational fluid dynamics (CFD) simulations, typically described with 3D Navier-Stokes equations (Anderson, 1995). However, 3D CFD simulations are computationally expensive, with a single calculation taking hours on a high-performance computer, making their use with an iterative search approach difficult. Moreover, assumptions need to be made, for instance, regarding turbulence or pressure distributions, which can significantly affect accuracy when modelling interacting wind turbines. Previous evolutionary studies have been undertaken with types of CFD to optimise the blade profile for both HAWT (Hampsey, 2002) and VAWT (Carrigan et al., 2012) to varying degrees of success/realism.

The evaluation of physical artifacts directly for fitness determination can be traced to the origins of evolutionary computation (Dunham et al., 1963). For example, the first ESs were used to design jet nozzles with a string of section diameters, which were then machined and tested for fitness (Rechenberg, 1971). Other well-known examples include robot controller design (Nolfi, 1992), electronic circuit design using programmable hardware (Thompson, 1998), product design via human-provided fitness values (Herdy, 1996), chemical systems (Theis et al., 2006), and unconventional computers (Harding and Miller, 2004). Boria et al. (2009) used an EA to evolve a morphing wing structure where physical designs were morphed using a set of actuators and evaluated in a closed-loop wind tunnel.

Evolution in hardware has the potential to benefit from access to a richer environment; it can exploit subtle interactions that can be utilised in unexpected ways. For example, the EA used by Thompson (1998) to work with field-programmable gate array circuits used some subtle physical properties of the system to solve problems where the properties used are still not understood. Humans can be prevented from designing systems that exploit these subtle and complex physical characteristics through their lack of knowledge; however this does not prevent exploitation through artificial evolution. There is thus a real possibility that evolution in hardware may allow the discovery of new physical effects, which can be harnessed for computation/optimisation (Miller and Downing, 2002).

Moreover, the advent of high-quality, low-cost additive rapid fabrication technology, known as 3D printing, means it is now possible to fabricate a wide range of prototype designs quickly and cheaply. 3D printers are now capable of printing a growing array of different materials; examples include chocolate (Hao et al., 2009) and meat (Lipton et al., 2010) for culinary design; sugar to help create synthetic livers (Miller et al., 2012); chemicals for custom drug design (Symes et al., 2012); cells for functional blood vessels (Jakab et al., 2008) and artificial cartilage (Xu et al., 2013); plastic for Southampton University laser sintered aircraft; thermoplastic for electronic sensors (Leigh et al., 2012); titanium for prosthetics such as the synthetic mandible developed by the University of Hasselt and transplanted into an 83-year-old woman; and liquid metal for stretchable electronics (Ladd et al., 2013). One potential benefit of the technology is the ability to perform fabrication directly in the target environment; for example, Cohen et al. (2010)
used a 3D printer to perform a minimally invasive repair of the cartilage and bone of a calf femur in situ. Lipson and Pollack (2000) were the first to exploit the emerging technology in conjunction with an EA using a simulation of the mechanics and control, ultimately printing mobile robots with embodied neural network controllers.

2.3 Surrogate-Assisted Evolutionary Algorithms

Whilst the speed and cost of rapid prototyping continues to improve, fabricating an evolved design before fitness can be assigned remains an expensive task when potentially thousands of evaluations are required; for instance, 10 minutes print time for each very simple individual (Rieffel and Sayles, 2010). However, given a sample \( D \) of evaluated individuals \( N \), a surrogate model (also known as a metamodel or response surface model) \( y = f(x) \) can be constructed, where \( x \) is the genotype and \( y \) fitness, in order to compute the fitness of an unseen data point \( x \notin D \). The use of surrogate models has been shown to reduce the convergence time in evolutionary computation and multiobjective optimisation; see Jin (2011). Refer to Viana et al. (2014) for general reviews and Forrester and Keane (2009) for aerospace design optimisation.

Typically, a set of evaluated genotypes and their real fitness scores are used to perform the supervised training of a multilayer perceptron-based (MLP; Rosenblatt, 1962) artificial neural network (e.g., Eberhart, 1992). However, other approaches are widely used, for example, kriging (e.g., Ratle, 2001), clustering (e.g., Kim and Cho, 2001), support vector regression (e.g., Yun et al., 2009), radial-basis functions (e.g., Ong et al., 2006), and sequential parameter optimisation (e.g., Bartz-Beielstein, 2006). The surrogate model is subsequently used to compute estimated fitness values for the EA to utilise. The model must be periodically retrained with new individuals under a controlled evolutionary approach (also known as model management) to prevent convergence on local optima. Retraining can be performed by taking either an individual or generational approach. In the individual approach, \( n \) number of individuals in the population, \( P \), are chosen and evaluated with the real fitness function each generation. In the generational approach, the entirety of \( P \) is evaluated on the real fitness function each \( n \)th generation. Typically the members with the highest approximated fitness are chosen to be evaluated with the real fitness function in the individual approach, although alternative schemes have been suggested; Bull (1999) found that evaluating both the best and random individuals suggested by a neural network surrogate model resulted in a significant improvement over exclusively evaluating either the best or random individuals. Both global modelling and local modelling using trust regions (e.g., samples within a certain Euclidean distance) are popular approaches (Le et al., 2013). Resampling methods and surrogate model validation remain an important and ongoing area of research, enabling the comparison and optimisation of models (Bischl et al., 2012).

The use of approximations in a coevolutionary context has previously been shown capable of solving computationally expensive optimisation problems with varying degrees of epistasis more efficiently than conventional CGAs through the use of radial basis functions (Ong et al., 2002) and memetic algorithms (Goh et al., 2011).

2.4 Coevolution

Since the early work in evolving both competitive (Axelrod, 1987) and cooperative (Husbands and Mill, 1991) multiagent systems, the problem of how to pick evaluation partners has been noted. Many strategies have been presented, which vary from the extreme case of each individual in one population using all others in all other populations (e.g., Koza, 1991), to each individual using a subset of the others (e.g., Hillis, 1991), to
the computationally more preferable use of one individual from each population (e.g., Bull and Fogarty, 1993). The use of the current best individuals from the other species populations was examined by Potter and De Jong (1994). In their work, all populations also received a shared fitness measure. Using a generational CGA, they reported that the strategy performed well. They also suggested that for higher cross-species epistasis an additional randomly selected individual should be used as a partner and the highest obtained fitness from the two pairings assigned to the evaluating genome. Bull (1997a) examined various strategies for choosing evaluation partners from coevolving populations and also found improved performance on problems with significant cross-species epistasis when using multipartner strategies. No significant difference was observed whether the second collaborator was chosen randomly or fitness proportionately after Paredis (1994). These early results suggested that the collaboration method depends entirely upon the amount of epistatic interaction between different species. However the issue is more complex; for example, Wiegand et al. (2001) found that simply using the fittest individual as the collaborating partner worked best on a nonlinearly separable quadratic problem with cross-species epistasis. In addition, they found that when using multiple collaborative function evaluations, assigning the fitness score from the best collaboration is significantly better than using either the worst or average of the scores.

3 Methodology

Here, we investigate the use of SCGAs to design small wind farms, utilising the aggregated rotational speed of the array as fitness. Each VAWT is treated separately by evolution and approximation techniques; thus, heterogeneous designs could emerge.

A vector of 10 integers is used as a simple and compact encoding of the $x$-$y$-axis of a prototype VAWT. Each allele thus controls one-tenth of a single $z$-layer. A workspace (maximum object size) of $30 \times 30 \times 30$ mm is used so that the instantiated prototype is small enough for timely production ($\sim 30$ minutes) and with low material cost, yet large enough to be sufficient for fitness evaluation. The workspace has a resolution of $100 \times 100 \times 100$ voxels. A central platform is constructed for each individual to enable the object to be placed on to the evaluation equipment. The platform consists of a square torus, 1 voxel in width and with a centre of $14 \times 14$ empty voxels that are duplicated for each $z$ layer, thus creating a hollow tube that is 3 mm in diameter.

To translate the genome for a single $z$ layer, an equilateral cross is constructed using the 10 aforementioned genes, with four blades bent at right angles and an allele range $[1, 42]$. For northeast and southwest quadrants the baseline is a horizontal line at $y$-axis = 50, and for northwest and southeast quadrants the baseline is a vertical line at $x$-axis = 50. Starting from the central platform and translating each gene successively, the one-tenth of voxels controlled by that gene are then drawn from the allele + baseline toward the baseline; see Figure 2a. If the current allele+baseline is greater than or the same as the previous allele+baseline, the voxels are enabled from the current allele+baseline upward to the previous lower position and extended a further 2 voxels; see Figure 2c. In all other cases, 2 voxels are enabled from the current allele+baseline position toward the baseline; see Figure 2d.

An example phenotype without $z$-axis variation is shown in Figure 3. When production is desired, the 3D binary voxel array is converted to stereolithography (STL)
Figure 2: Translation of $x$-$y$-axis genome $[5, 8, 2, 4]$. (a) The voxels are enabled from the first allele (5) to the baseline (bottom). (b) Subsequently, the voxels are enabled from the second allele (8) to the previous allele (5) and extended 2 voxels. (c) The third allele (2) is less than the previous lower position (3), causing a gap, and is thus drawn from the allele (2) to the previous lower position and extended 2 voxels to provide structural support. (d) The allele (4) is less than the previous upper position (5), and 2 voxels are enabled from the allele toward the baseline.

Figure 3: Example phenotype; genome = $[2, 2, 3, 4, 5, 8, 13, 20, 34, 40]$.

format. Once encoded in STL, it then undergoes postprocessing with the application of 50 Laplacian smoothing steps using MeshLab; see smoothed example phenotype in Figure 4. Finally, the object is converted to printer-readable G-code and is subsequently fabricated on a BFB 3000 printer (0.25 mm resolution) using a polylactic acid (PLA) bioplastic. Figure 5 shows the smoothed object after fabrication.

To enable prototypes with $z$-axis variability, the genome also includes five additional integers in the range $[-42, 42]$, each controlling one-sixth of the $z$-axis. After drawing the initial $z$-layer as previously described, each $z$ gene transforms the $x$-$y$ genome for the next successive $z$-layer by uniformly adding the allele value, after which it is drawn as described. For example, with an $x$-$y$-axis genome of $[2, 2, 3, 4, 5, 8, 13, 20, 34, 40]$

---

1MeshLab is an open source, portable, and extensible system for the processing and editing of unstructured 3D triangular meshes. http://meshlab.sourceforge.net/
and z-axis genome of [2, −5, 10, 3, −2], the next z-layer is translated using the x-y-axis genome of [4, 4, 5, 6, 7, 10, 15, 22, 36, 42] and the following z-layer is translated with [1, 1, 1, 2, 5, 10, 17, 31, 37], and so on.

Since the interaction effects of rotor direction in the context of differing turbine morphology, relative phase angle, separation distance, positional layout, wind velocity, and so on, are not well understood, a Boolean gene is added to designate the rotation direction of a VAWT. Thus, a total of 16 genes define an individual. In a previous CGA experiment (Preen and Bull, 2014a) with two Savonius turbines positioned 0.1 diameters adjacently and in perpendicular position to the air flow, the average combined rotational speed of the final 20 counter-rotating pairs ($M = 1,760$, $SD = 206$, $N = 20$) was significantly less than the average combined rotational speed of the final 20 co-rotating pairs ($M = 2,048$, $SD = 95$, $N = 20$) using a two-tailed Mann-Whitney test ($U = 17$, $p \leq 7.39 \times 10^{-7}$), showing that co-rotation was found to result in faster combined rotational speed. It can be noted that this result is similar to Sun et al. (2012) where two counter-rotating S-rotors performed worse than co-rotating versions when placed at small separation distances.

The CGA used herein proceeds with two species populations, a maximum mutation step size of ±10, a per allele mutation rate of 25%, and a crossover rate of 0%. A tournament size of 3 takes place for both selection and replacement. To help increase performance and save fabrication time, each species population initially consists of the first 10 (>0 rpm) individuals from a single VAWT (z-varying) experiment of Preen and Bull (2014a), both normally rotated and counter-rotated. These individuals still retain a good degree of randomness while possessing some useful aerodynamic properties. Each
species thus maintains $P = 20$ individuals. The individuals in each species population are initially evaluated in collaboration with a single randomly selected individual from the other species population. Thereafter, the CGA proceeds by alternating between species after each offspring is formed and evaluated with the elite member from the other species; see Algorithm 1.

For the surrogate-assisted architecture described in this article, the basic CGA remains unchanged except that fitness evaluations are obtained from a forward pass of the genome through a neural network when the real fitness value is unknown. Initially the entire population is fabricated and evaluated on the real fitness function and added to an evaluated set. The model is trained using backpropagation for 1,000 epochs, where an epoch consists of randomly selecting, without replacement, an individual from the evaluated set and updating the model weights. Each generation thereafter, the individual with the highest approximated fitness as suggested by the model and a randomly chosen unevaluated individual are fabricated and evaluated on the real fitness function and added to the evaluated set. The model weights must be reinitialised each time before training because of the temporal nature of pairing with the elite member. The genetic algorithm runs for one generation (using the model-approximated fitnesses where real fitness is unknown) before the individual with the highest approximated fitness and a randomly selected unevaluated individual are evaluated with the elite member from the other species; see Algorithm 2. The model parameters are $\beta = 0.3, \theta = 0, \text{elasticity} = 1, \text{calmingrate} = 1, \text{momentum} = 0, \text{elasticityrate} = 0$.

The fitness of each individual is the maximum combined array rotational speed achieved over the period of 1 minute during the application of constant wind generated by a propeller fan after fabrication by a 3D printer. The rotational speed is the significant measure of aerodynamic efficiency, since the design space is constrained (including rotor radius and turbine height). However, in future work, the power generated will be preferred, which will take into account any slight weight variations that may affect performance. The rotational speed is here measured in number of revolutions per minute (rpm) using a digital photo laser tachometer (PCE-DT62; PCE Instruments UK Ltd) by placing a $10 \times 2$ mm strip of reflecting tape on the outer tip of one of

<table>
<thead>
<tr>
<th>Algorithm 1: Coevolutionary genetic algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Generate and fabricate individuals for all species;</td>
</tr>
<tr>
<td>2. for each species population do</td>
</tr>
<tr>
<td>3. Select random representative from each species;</td>
</tr>
<tr>
<td>4. for each individual in population do</td>
</tr>
<tr>
<td>5. Evaluate;</td>
</tr>
<tr>
<td>6. end</td>
</tr>
<tr>
<td>7. end</td>
</tr>
<tr>
<td>8. while fabrication budget not exhausted do</td>
</tr>
<tr>
<td>9. for each species population do</td>
</tr>
<tr>
<td>10. Create an offspring using evolutionary operators;</td>
</tr>
<tr>
<td>11. Select representatives for each species;</td>
</tr>
<tr>
<td>12. Fabricate and evaluate the offspring;</td>
</tr>
<tr>
<td>13. Add offspring to species population;</td>
</tr>
<tr>
<td>14. end</td>
</tr>
<tr>
<td>15. end</td>
</tr>
</tbody>
</table>
R. J. Preen and L. Bull

the individual’s blades. The experimental setup can be seen in Figure 6, showing the 30 W, 3500 rpm, 304.8 mm propeller fan, which generates 4.4 m/s wind speed, and two turbines mounted on rigid metal pins 1 mm in diameter and positioned 33 mm adjacently and 30 mm from the propeller fan. That is, there is a 3 mm spacing between the blades at their closest point.

4 Experiments

4.1 Modelling Techniques

To explore whether there is any significant advantage in replacing the neural network used previously as the surrogate model, several algorithms using the Weka 3.6.10 machine learning collection\(^2\) were run over the data generated by Preen and Bull (2014a). It is important to note that a surrogate model with a larger fitness prediction error may be more accurate in predicting the rank order of individuals than a model with lower error. Since the rank order is the guiding influence on the evolutionary process, alternative quality measures for approximate models have therefore been suggested. However, Bischl et al. (2012) found no difference between using the root mean squared error

---

(RMSE), mean absolute error (MAE), or Spearman’s rho, which measures the correlation between the actual and predicted ranking. Furthermore, in practice, optimisation based on the fitness accuracy has often been found to perform the best (e.g., Hüskens et al., 2005). Thus, the MAEs of the fitness predictions are here used as the performance metric.

Table 1 shows the average MAEs for various algorithms predicting the fitness of evolved individuals. Each number represents the average MAE over 100 runs using 10-fold cross-validation (CV). The results are marked in bold font where they are statistically different from the standard MLP at the 95% confidence level using a two-sample t-test assuming unequal variances. GA is the z-axis varying GA-only single VAWT experiment ($N = 100$); SGA is the surrogate-assisted GA z-axis varying VAWT experiment ($N = 100$); CGA-1 is the first species from the CGA experiment ($N = 80$); CGA-2 is the second species from the CGA experiment ($N = 80$); SCGA-1 is the first species from the surrogate-assisted CGA experiment ($N = 80$); and SCGA-2 is the second species from the surrogate-assisted CGA experiment ($N = 80$). Default Weka parameters were used unless otherwise specified.

The results show that there is little statistical difference between the algorithms in predicting the fitness of the evolved individuals. This is perhaps due to the very small and noisy training samples available for this task. We therefore continue with neural network surrogate models, which are well suited to problems with a high-dimensional input space (design space) and limited number of samples (Jin, 2005). In addition, neural network modelling frequently outperforms more complex approaches such as coevolutionary active learning (Ly and Lipson, 2014).

4.2 Windowing

In interactive evolutionary computation, a user’s evaluation is relative within each generation, and therefore fitness values from generations long past may differ from recent generations even if the evaluated candidate is identical. Consequently, Wang and Takagi (2005) found that a neural network surrogate model using only recent training data was more effective than using all past data. In addition, many other surrogate-assisted approaches have used fixed-length training sets (e.g., Bull, 1997b), although...
Table 1: Average mean absolute errors for algorithms predicting the fitness of evolved individuals from the experiments of Preen and Bull (2014a). Results marked in bold font are statistically different from the standard MLP at the 95% confidence level using a two-sample *t*-test assuming unequal variances.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>GA</th>
<th>SGA</th>
<th>CGA-1</th>
<th>CGA-2</th>
<th>SCGA-1</th>
<th>SCGA-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard MLP</td>
<td>197.89</td>
<td>237.86</td>
<td>430.30</td>
<td>432.64</td>
<td>462.57</td>
<td>394.04</td>
</tr>
<tr>
<td>Linear regression</td>
<td>158.91</td>
<td>170.78</td>
<td>275.13</td>
<td>390.79</td>
<td>361.97</td>
<td>379.08</td>
</tr>
<tr>
<td>Isotonic regression</td>
<td>135.57</td>
<td>145.97</td>
<td>327.22</td>
<td>462.68</td>
<td>367.10</td>
<td>523.73</td>
</tr>
<tr>
<td>Pace regression</td>
<td>160.75</td>
<td>172.57</td>
<td>257.57</td>
<td>397.99</td>
<td>380.85</td>
<td>384.50</td>
</tr>
<tr>
<td>Least med squares</td>
<td>208.80</td>
<td>237.78</td>
<td>306.71</td>
<td>432.26</td>
<td>334.00</td>
<td>447.14</td>
</tr>
<tr>
<td>SMO regression</td>
<td>152.70</td>
<td>191.33</td>
<td>280.94</td>
<td>370.18</td>
<td>334.65</td>
<td>356.75</td>
</tr>
<tr>
<td>MLP <em>N</em> = 1,000 <em>H</em> = 10</td>
<td>207.21</td>
<td>241.89</td>
<td>451.91</td>
<td>427.81</td>
<td>503.93</td>
<td>418.42</td>
</tr>
<tr>
<td>MLP <em>N</em> = 1,000 <em>H</em> = 100</td>
<td>265.10</td>
<td>202.33</td>
<td>385.61</td>
<td>414.38</td>
<td>416.14</td>
<td>408.88</td>
</tr>
<tr>
<td>MLP <em>N</em> = 10,000 <em>H</em> = 10</td>
<td>256.52</td>
<td>305.41</td>
<td>562.50</td>
<td>481.27</td>
<td>654.00</td>
<td>487.14</td>
</tr>
<tr>
<td>MLP <em>N</em> = 500 <em>H</em> = 1</td>
<td>191.63</td>
<td>216.88</td>
<td>397.40</td>
<td>401.16</td>
<td>405.01</td>
<td>401.27</td>
</tr>
<tr>
<td>MLP <em>N</em> = 500 <em>H</em> = 1</td>
<td>142.13</td>
<td>174.45</td>
<td>384.69</td>
<td>338.37</td>
<td>410.49</td>
<td>333.47</td>
</tr>
<tr>
<td>MLP <em>N</em> = 500 <em>H</em> = 1</td>
<td>142.32</td>
<td>177.98</td>
<td>395.50</td>
<td>340.86</td>
<td>418.94</td>
<td>343.24</td>
</tr>
<tr>
<td>MLP <em>N</em> = 10,000 <em>H</em> = 1</td>
<td>159.20</td>
<td>193.07</td>
<td>429.34</td>
<td>364.05</td>
<td>446.37</td>
<td>366.36</td>
</tr>
<tr>
<td>MLP <em>N</em> = 500 <em>H</em> = 2</td>
<td>154.28</td>
<td>181.66</td>
<td>392.26</td>
<td>389.26</td>
<td>390.21</td>
<td>337.27</td>
</tr>
<tr>
<td>MLP <em>N</em> = 1,000 <em>H</em> = 2</td>
<td>156.30</td>
<td>181.51</td>
<td>408.23</td>
<td>407.67</td>
<td>408.15</td>
<td>358.50</td>
</tr>
<tr>
<td>Gaussian processes</td>
<td>168.00</td>
<td>193.03</td>
<td>281.32</td>
<td>369.74</td>
<td>323.65</td>
<td>361.45</td>
</tr>
<tr>
<td>RBF <em>B</em> = 2</td>
<td>172.28</td>
<td>180.44</td>
<td>432.09</td>
<td>533.81</td>
<td>482.91</td>
<td>596.46</td>
</tr>
<tr>
<td>RBF <em>B</em> = 10</td>
<td>213.94</td>
<td>158.23</td>
<td>1253.08</td>
<td>512.91</td>
<td>404.72</td>
<td>3981.20</td>
</tr>
<tr>
<td>REP tree</td>
<td>196.11</td>
<td>164.33</td>
<td>346.77</td>
<td>343.68</td>
<td>394.93</td>
<td>377.64</td>
</tr>
<tr>
<td>Decision stump</td>
<td>169.99</td>
<td>193.78</td>
<td>325.33</td>
<td>488.70</td>
<td>390.42</td>
<td>519.27</td>
</tr>
<tr>
<td>M5P</td>
<td>165.70</td>
<td>168.34</td>
<td>311.37</td>
<td>337.33</td>
<td>355.37</td>
<td>350.04</td>
</tr>
</tbody>
</table>

they did so to save computational time rather than to intentionally aid convergence. Because of the temporal nature of partnering with the elite species members it may be advantageous to exclude older training samples for SCGAs. To explore whether there is any benefit from using different windowing approaches to training the surrogate model, 13 different models from Weka were run across the four datasets consisting of the 80 evaluated individuals in each species from the coevolution experiments: two species for CGA and two for SCGA. The final two individuals from each set were used for testing and were excluded from any training.

The models used are as follows. Weka default parameters were used unless otherwise stated: (1) Gaussian processes (support vector, RBF kernel); (2) isotonic regression; (3) linear regression (M5 attribute selection); (4) multilayer perceptron; (5) SMOreg (RBF kernel, RegSMO improved); (6) RBF network (*B* = 2); (7) RBF network (*B* = 100); (8) multilayer perceptron (*N* = 1,000, *H* = 10, *M* = 0); (9) multilayer perceptron (*N* = 500, *H* = 1); (10) multilayer perceptron (*N* = 1,000, *H* = 1); (11) multilayer perceptron (*N* = 500, *H* = 5); (12) multilayer perceptron (*N* = 1,000, *H* = 5); (13) multilayer perceptron (*N* = 1,000, *H* = 100).

One set of experiments used the full 78 training samples for each dataset (*T* = 78); one set used the most recent 40 training samples (*T* = 40); one set used the most recent 20 training samples (*T* = 20); and one set used the most recent 10 training samples (*T* = 10). The average of the 13 algorithms’ MAEs over the four datasets with *T* = 20
Figure 7: Array rotational speed-based evolution. Fittest array pairs. The SCGAs are used for comparison only after 120 evaluations (three generations) of the CGA, since sufficient training data are required for the surrogate models.

(M = 197.11, SD = 30.54, N = 13) is significantly less than T = 78 (M = 336.39, SD = 18.34, N = 13) using a two-sample t-test assuming unequal variances, t(14) = 4.76, p = .0003. In addition, the average MAE of T = 20 is also significantly less than T = 40 (M = 269.94, SD = 16.43, N = 13), t(16) = 3.31, p = .0045. However, the average MAE of T = 10 (M = 190.14, SD = 15.03, N = 13) was not significantly less than T = 20, t(17) = 0.74, p = .47.

These results suggest that using the most recent 20 training samples (i.e., \( P \)) produces a more accurate model than using the full evaluation set. The SCGA was therefore rerun as before, using only the most recent 20 evaluated individuals for training (SCGA-20T); the results are shown in Figure 7. The average rotational speed of the original SCGA (M = 2,112, SD = 307, N = 40) was significantly greater than CGA (M = 1,905, SD = 223, N = 40) using a two-tailed Mann-Whitney test, \( U = 331, p \leq 6.4 \times 10^{-6} \).

Furthermore, the fittest array combination designed by SCGA (2429 rpm) was greater than CGA (2209 rpm) after 160 fabrications. However, the average rotational speed of SCGA-20T (M = 1,971, SD = 386, N = 40) is not significantly different than SCGA (M = 2,112, SD = 307, N = 40) using a two-tailed Mann-Whitney test, \( U = 629, p \leq .1 \), and the fittest array combination was nearly identical after 160 fabrications; SCGA-20T 2444 rpm versus SCGA 2429 rpm. The fittest CGA, SCGA, and SCGA-20T evolved arrays after 160 fabrications can be seen in Figure 8.

4.3 Model Enhanced Local Search

Typically \( n \) individuals are evaluated with the real fitness function each generation, where \( n < P \). This results in a population consisting of a mix of individuals evaluated with the real fitness function and those whose fitnesses are approximated. However, Ulmer et al. (2004) introduced a preselection approach to the surrogate assistance of ESs. The approach consists of creating \( \lambda_{Pre} \) offspring, where \( \lambda_{Pre} > \lambda \), and evaluating each with the surrogate model; whereupon the \( \lambda \) individuals with the highest approximated fitness are then selected to form the offspring population and are evaluated with the real fitness function. The main difference is that with preselection all offspring are...
generated from parent individuals evaluated directly with the real fitness function. This enables an enhanced local search through the evaluation of a large number of offspring while preventing the evolutionary search from drifting too far from the evaluated design space, which can occur when repeatedly creating offspring from approximated individuals. They found that the bigger $\lambda_{Pr}$ is, the better the algorithm performed. Here we explore the use of a SCGA with an enhanced local search, SCGA-ELS. The CGA runs as usual except that each time a parent is chosen, $m$ number of offspring are created and evaluated with the surrogate model, and the single offspring with the highest approximated fitness is then fabricated and evaluated with the real fitness function before being added to the population; see Algorithm 3. Here $m = 1,000$. The results are shown in Figure 9.

The average rotational speed of SCGA-ELS ($M = 2,264, SD = 322, N = 40$) is significantly greater than SCGA ($M = 2,112, SD = 307, N = 40$) using a two-tailed Mann-Whitney test, $U = 469, p \leq .0015$. Furthermore, the fittest SCGA-ELS array pairing (2692 rpm; see Figure 10) was greater than SCGA (2429 rpm) after 160 fabrications. However, this comparison may not be fair to the original SCGA, which fabricates a randomly selected individual as well as the best each epoch, whereas SCGA-ELS only fabricates the best. Therefore, a two-tailed Mann-Whitney test was performed for the rotational speed of SCGA-ELS over the first 20 model-suggested offspring ($M = 2,303, SD = 134, N = 20$) with only the suggested best offspring from SCGA ($M = 2,087, SD = 386, N = 20$), and the result was found to be statistically significant, $U = 114, p \leq .02$. The same comparison was made for SCGA-ELS over the first 20 model-suggested offspring with only the suggested best offspring from SCGA-20T ($M = 1,907, SD = 443, N = 20$),
Figure 9: Array rotational speed-based evolution. Fittest array pairs. The SCGAs are used for comparison only after 120 evaluations (three generations) of the CGA, since sufficient training data are required for the surrogate models.

and the result was also statistically significant, $U = 73.5$, $p \leq .00063$. SCGA-ELS therefore seems to offer a clear benefit in these experiments.
4.4 Collaboration Strategies

While it is relatively costly to fabricate a new VAWT prototype (~30 minutes), the rotational speed measurement with the laser tachometer is relatively cheap (≤ 1 minute). Therefore, while it may not be the case for more elaborate wind tunnel testing conditions, here reevaluating some fabricated individuals with other partners may provide a relatively cheap way to increase performance. To see whether a two-partner strategy can increase the performance in designing two interacting VAWTs, the CGA was rerun with the same initial population as before, but collaboration was performed with an individual that was fitness proportionately selected (i.e., roulette wheel) in addition to the usual elite member from the other species population, and the larger of the two fitness scores assigned to the individual being evaluated (CGA-2). Since initially no fitness scores are known, each individual in the first generation of a species population was evaluated in conjunction with a single randomly selected individual from the other population (that is, as before) and also with a second randomly selected individual, uniquely chosen for each evaluation, and the larger of the two shared fitness scores assigned.

So far, it had been assumed that there is a significant degree of asymmetry within the array solution and hence no cross-species gene sharing had been permitted. However, it is possible that there may be sufficient symmetry that can be exploited to increase performance. Therefore, the CGA was also rerun with the same initial population as before, but an offspring was also evaluated in the alternative physical position and in collaboration with the elite member from its own species population (CGA-CROSS). If the fitness is greater than the population best fitness, it is also added to that species population, replacing the worst individual.

Figure 11 shows the rotational speed of the fittest array pairings evolved by the original CGA, CGA-2, and CGA-CROSS over 120 fabrications. As can be seen, the fittest CGA-2 array pairing after 120 fabrications (2343 rpm) is greater than CGA (2158 rpm). Furthermore, the average combined rotational speed of the final 40 offspring formed by CGA-2 ($M = 2,053$, $SD = 295$, $N = 40$) is significantly greater than the original CGA ($M = 1,882$, $SD = 316$, $N = 40$) using a two-tailed Mann-Whitney test, $U = 481$, $p \leq .0022$. The extra evaluation during initialisation was used 38 out of 40 times, which provided an initial performance boost. However, during steady state evolution the extra evaluation was only used 4 out of 80 times. That is, the roulette wheel–selected partner resulted in a combined rotational speed greater than when paired with the best individual only 5% of the time. Moreover, the average difference of those fitness scores was only 12 rpm, which probably would have made little difference to the GA. While not used for fitness determination during the experiment, a randomly selected partner was also evaluated with each offspring and found to have a near identical effect to the roulette wheel collaborator, being greater than the best partner 4 out of 80 times with an average difference of 46 rpm. Thus it appears that performing extra evaluations...
during initialisation provides a significant boost to performance over only partnering with a single random collaborator, but thereafter there is little performance benefit from performing extra evaluations.

The fittest CGA-CROSS array pairing after 120 fabrications (2373 rpm) is similar to CGA-2 (2343 rpm). In 7 out of the 80 extra evaluations performed in the CGA-CROSS experiment, evaluating the offspring in the alternative position resulted in a faster combined rotational speed than the fittest individual in that species. Furthermore, both individuals in the fittest array pairing were offspring produced in the alternative species, showing that there is a high degree of symmetry in the task. However, despite the extra evaluations having resulted in a VAWT pair with a higher rotational speed than the fittest pair from the original CGA (2158 rpm), the average combined rotational speed of the final 40 CGA-CROSS offspring evaluated in the original species position ($M = 1,988, SD = 261, N = 40$) is not statistically different from CGA using a two-tailed Mann-Whitney test, $U = 655, p \leq .163$. The fittest CGA-2 and CGA-CROSS interacting VAWTs after 120 fabrications are shown in Figure 12.
5 Conclusions

This article has explored a range of surrogate modelling and evolutionary techniques used to design interacting VAWTs wherein candidate prototypes are fabricated by a 3D printer and evaluated under fan-generated wind conditions. The experiments here have shown that it is possible to use SCGAs to iteratively increase the performance of closely positioned Savonius turbines, exploiting the interturbine flow effects, which is extremely difficult to achieve under simulation. The SCGA represents a scalable approach to the design of wind turbine arrays, since the number of inputs to the surrogate-models remains constant regardless of the number of turbines undergoing evolution.

The accuracy of various modelling algorithms used to estimate the fitness of evaluated individuals from the initial experiments was compared, finding that there is little difference between the algorithms for the current task. The effect of temporally windowing the surrogate model training samples was shown to be a promising approach, but it resulted in no performance benefit in practice. Subsequently, a SCGA based on an enhanced local search was introduced and found to produce more efficient designs within the same number of fabrications when compared with the original SCGA approaches. Finally, alternative coevolution collaboration schemes were examined, finding an initial performance increase resulting from more accurate fitness assignments but with no significant improvement thereafter. This highlights that while much can be learned by developing optimisation algorithms using simulated problems and mining data, significant differences in performance are often found when applying the results to the real application. Thus, whilst general learning algorithms provide a starting point for exploration, there is no escape from further development directly in the problem application.

The use of 3D printing to physically instantiate candidate designs completely avoids the use of 3D computer simulations, with their associated processing costs and modelling assumptions. In this case, 3D CFD analysis was avoided, but the approach is equally applicable to other real-world optimisation problems, for example, those requiring computational structural dynamics or computational electromagnetics simulations. We anticipate that in the future such design mining approaches will yield unusual yet highly efficient designs that exploit characteristics of the environment or materials that are difficult to capture formally or in simulation. This has the potential to place knowledge discovery at the core of engineering design, particularly within an iterative framework such as in agile approaches.

Although the print resolution used here to build the prototypes was set at the printer default, the resolution can be adjusted to provide coarser designs at a faster rate for preliminary studies (e.g., for early evolutionary candidates), or slower, higher-resolution prints for more subtle optimisation. Further, only PLA plastic was used here to fabricate designs, but other materials such as flexible rubbers or multimaterial designs can be constructed and explored by an EA. Thus 3D printing offers a range of ways to customise the evolutionary instantiation to the design task. Multiple 3D printers can also easily be used to perform parallel fabrication to speed up the process.

Future work will include the use of the power generated as the fitness computation under various wind tunnel conditions; the coevolution of larger arrays, including the turbine positioning; the exploration of more advanced assisted learning systems to reduce the number of fabrications required; examination of the effect of seeding the population with a given design; investigation of alternative representations that provide
more flexible designs including variable number of blades, for example, supershapes (Preen and Bull, 2014b); and the production of 1:1 scale designs.

The issue of scalability remains an important future area of research. When the scale of designs is increased, the changes in dimensionality will greatly affect performance; however, it remains to be seen how performance will change in the presence of other significant factors such as turbine wake interactions in the case of arrays. One potential solution is to simply use larger 3D printing and wind tunnel capabilities whereby larger designs could be produced by the same method. On the opposite end of the spectrum, micro wind turbines that are 2 mm in diameter or smaller can be used to generate power, such as for wireless sensors (Howey et al., 2011), and in this case more precise 3D printers would be required. Moreover, wind turbines can find useful applications on any scale. For instance, a feasibility study (Park et al., 2012) for powering wireless sensors on cable-stayed bridges examined turbines with a rotor diameter of 138 mm in wind conditions with an average of 4.4 m/s (similar to the artificial wind conditions used in this article).

If the recent speed and material advances in rapid-prototyping continues, along with the current advancement of evolutionary design, it will soon be feasible to perform a wide array of automated complex engineering optimisation in situ, whether on the microscale (e.g., drug design), or the macroscale (e.g., wind turbine design). That is, instead of using mass manufactured designs, EAs will be used to identify bespoke solutions that are manufactured to compensate and exploit the specific characteristics of the environment in which they are deployed, for example, local wind conditions, nearby obstacles, and local acoustic and visual requirements for wind turbines.

Acknowledgments

This work was supported by the UK Leverhulme Trust under Grant RPG-2013-344.

References


