RoboCup Simulation Leagues

Enabling Replicable and Robust Investigation of Complex Robotic Systems

©ISTOCKPHOTO.COM/KIRSTYPARGETER

By David M. Budden, Peter Wang, Oliver Obst, and Mikhail Prokopenko

Physically realistic simulated environments are powerful platforms for enabling measurable, replicable, and statistically robust investigation of complex robotic systems. Such environments are epitomized by the RoboCup (RC) simulation leagues, which have been successfully utilized to conduct massively parallel experiments on a variety of topics, including optimization of bipedal locomotion, self-localization from noisy perception data, and planning complex multiagent strategies without direct agent-to-agent communication. Many of these systems are later transferred to physical robots, making the simulation leagues invaluable beyond the scope of simulated soccer matches.

In this article, we provide an overview of the RC simulation leagues and describe their properties as they pertain to replicable and robust robotics research. To demonstrate their utility directly, we leverage the ability to run parallelized experiments to evaluate different competition formats (e.g., round robin) for the RC two-dimensional (2-D) simulation league. Our results demonstrate that a previously proposed hybrid format minimizes fluctuations from true (statistically significant) team performance rankings within the time constraints of the RC World Finals. Our experimental analysis would be impossible with physical robots alone, and we encourage other researchers to

Digital Object Identifier 10.1109/MRA.2015.2446911
Date of publication: 18 August 2015

1070-9932/15©2015IEEE

explore the potential for enriching their experimental pipelines with simulated components, both to minimize the experimental costs and to enable others to replicate and expand upon their results in a hardware-independent manner.

Simulation in Robotics Research

Robotics researchers face many unique challenges when designing measurable, replicable, and statistically robust experiments. Robots are generally expensive and require regular maintenance, pressuring researchers to minimize the time spent evaluating algorithms and behaviors. This problem is amplified by the experimental confounds introduced by robots (e.g., motor temperature) and their environments (e.g., lighting variation), requiring more experimental iterations to yield statistically robust results. Moreover, the cost of physical robots makes platform-specific research replication particularly difficult, threatening the reliability of the peer-review process in the presumably common scenario of reviewers not having access to the robot in question.

In recent years, robot simulation has emerged as a powerful technique for replicable and robust investigation of complex robotic systems. Popular simulators include Simspark [1], Gazebo [2], Webots [3], and USARsim [4], which have been successfully applied in multi-institutional collaborations, including RC Simulation Leagues [5], [6], RobotStadium [7], the Defense Advanced Research Projects Agency Virtual Challenge [8], and the IEEE Virtual Manufacturing Competition [9]. Although these platforms are unable to perfectly model the physical sources of system stochasticity, this shortcoming is greatly outweighed by their ability to remove the requirement for physical robots and inherent support for massively parallel processing. In scenarios where imperfect system modeling is a major limitation, hybrid solutions have been developed that use simulated environments for initial experimentation (e.g., global exploration of high-dimensional parameter spaces) and physical robots for fine tuning (e.g., local optimization across principal components). The latter approach is epitomized by bipedal gait optimization and has been shown to be highly effective for the Aldebaran NAO humanoid robot [10], [11].

In this article, we provide an overview of the RC simulation leagues in the context of enabling replicable and statistically robust robotics research. We demonstrate the utility of massively parallel experimentation by evaluating different competition formats (e.g., round robin) for the RC 2-D simulation league. Tens of thousands of games were necessary to resolve nontransitivity and inherent stochasticity of team performance, which would be intractable for 10-min matches with physical robots. Our results demonstrate that a hybrid format best captures true team performance in the time constraints of the RC World Finals. This format was subsequently adopted for the competitions at RC 2014 Brazil.

RoboCup Simulation Leagues

RoboCup (the World Cup of robot soccer) was first proposed in 1997 as a standard problem for the evaluation of theories, algorithms, and architectures in the areas of artificial intelligence (AI), robotics, and computer vision [13]. This proposal followed the observation that traditional AI problems were increasingly unable to meet the requirements of appropriately/effectively evaluating theories, algorithms, and architectures and that a new challenge was necessary to initiate the development of next-generation technologies.

The overarching RC goal of developing a team of humanoid robots capable of defeating the FIFA World Cup champion team, called *the Millennium Challenge*, has been a major factor in driving research in AI and related areas for nearly two decades, with a search for the term *RC* in a major literature database yielding over 25,000 results. Since 1997, researchers and competitors have decomposed this ambitious pursuit into two complementary categories [13].

- **Physical Robot League**: Using physical robots to play soccer games. This category now contains many different leagues for both wheeled robots (small-sized and midsized leagues) and humanoids [standard platform league (SPL) and humanoid league], with each focusing on different aspects of physical robot design, motor control and bipedal locomotion, real-time localization, and computer vision [14], [15].
- **Software Agent League:** Using software or synthetic agents to play soccer games on an official soccer server over a network. This category contains both 2-D [5], [16], [17] and three-dimensional (3-D) [6] simulation leagues.

The RC simulation leagues traditionally involve the largest number of international participating teams, reaching 40 in 2013 [18]. The ability to simulate soccer matches without physical robots removes low-level hardware and environmental issues (e.g., motor temperature and breakages), allowing teams to focus on the development of complex team behaviors and strategies for a larger number of autonomous agents. Moreover, the simulation leagues often serve as platforms for the initial development and evaluation of software modules for later integration into physical robots [10], [11]. Many of these modules have applications beyond the RC domain (e.g., localization and mapping [12]), and the hardware-independent results inherent to simulated robots promote extension and replication by other researchers.

Properties and Utility of 2-D and 3-D Leagues

The RC simulation league consists of both 2-D and 3-D competitions, which exhibit many similarities [18].

- The world model, including player and ball dynamics and kinematics, is simulated by a central soccer server [5].
- Participants develop a team of fully autonomous agents, each of which interacts with the soccer server.
  - Each agent receives information from the server regarding its current field of view.
  - Each agent determines what actions to execute and submits these requests to the server.
  - The server fulfills these requests and resolves any conflicts (e.g., two agents attempting to occupy the same spatial location).

The server proceeds in real time and imposes noise on both the agents' observations and actions [19]. It is the responsibility of each agent to submit its action requests at the appropriate times to stay synchronized with the soccer server. Furthermore, each agent is allocated an individual process/core, and no direct interprocess communication is permitted. The soccer server provides a low-bandwidth, indirect communications method between agents by support simulated verbal commands.

2-D Simulation League

The 2-D simulation league involves circular players being modeled with an (x, y) position and orientation θ. Each agent also maintains a head angle relative to its global orientation, allowing control of its field of view within human-like constraints. The action commands available to each agent include the following:

- turn body or neck by a specified angle
- dash forward or backward with a specified power
- slide tackle in a specified direction
- kick the ball in a specified direction with a specified angle, if near
- catch the ball if near (goalkeeper only)
- communicate with other players, either verbally or by pointing at a specified position.

Each team consists of 11 players and a coach, which is a nonplaying agent responsible for the allocation of players to each position given a number of randomly generated physical profiles (including characteristics such as speed and stamina). The 2-D simulation league does not model the dynamics or kinematics of any given human or robot. Instead, it encourages the development of complex player behaviors and team strategies [16]–[18]. The simulation league is also a powerful framework for evaluating the emergent downstream effects (e.g., team performance) of small perturbations to the underlying individual agents, as demonstrated in our recent study of particle filtering and self-localization [12].

A screenshot from the 2-D simulation league graphical client is presented in Figure 1(a), and the 2-D soccer server is available online at: http://sourceforge.net/projects/sserver/.

3-D Simulation League

The 3-D simulation league implements a physically realistic world model and action interface, closer to robots than human players [18]. In particular, the simulator uses the Open Dynamics Engine library for the simulation of rigid body dynamics, collision detection, and friction, based on a model of the Aldebaran NAO humanoid robot (shown in Figure 2). To remain consistent with the anatomy of the NAO, each agent simulates the following:

- 22-degrees of freedom (DOF) in a 57-cm, 4.5-kg humanoid robot (six in each leg, four in each arm, and two in the neck)
- perceptors that provide the agent with noiseless measurements of each joint position during every simulation cycle
- effectors that allow the agent to specify a direction and torque for each joint.

Although no noise is introduced to the perceptor and the effector signals (with the exception of that resulting from approximations in the physics engine), the 3-D simulation league introduces the nontrivial challenges of enabling each agent to stably walk, kick, dive, and stand up after falling. This creates an ideal framework for global optimization (and benchmarking optimization algorithms) across the high-dimensional parameter spaces characteristic of bipedal locomotion systems. Although the simulated agents do not perfectly model the stochasticity inherent to actual NAOs, this approach has proved very successful in identifying near-optimal parameter sets for subsequent local optimization on the physical robot [10], [11] (often in a lower dimensional principal component space).

A screenshot from the 3-D simulation league graphical client is presented in Figure 1(b), and the 3-D soccer server is available online at: http://sourceforge.net/projects/simspark/.

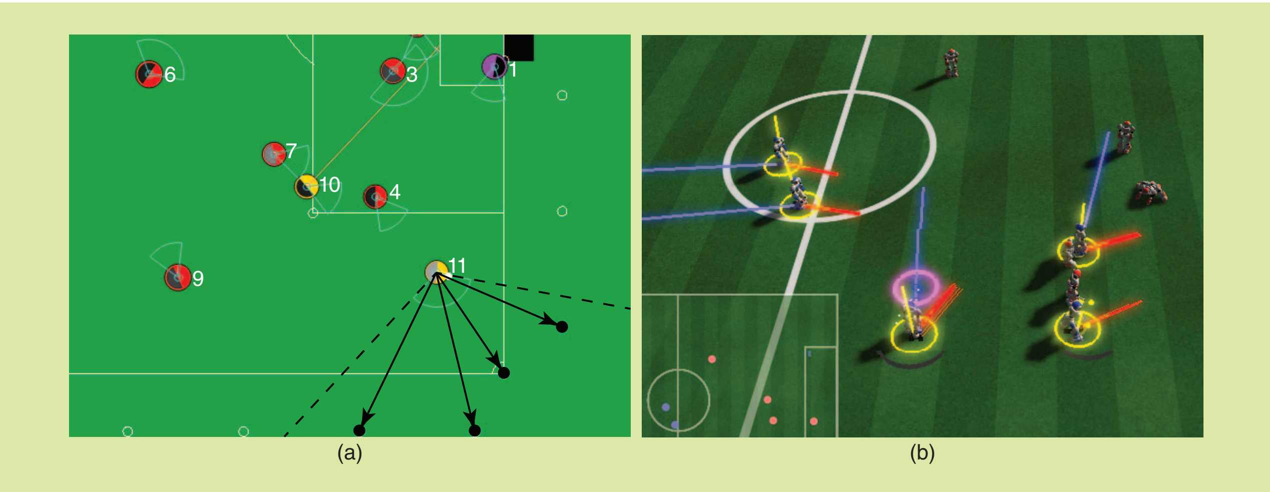


Figure 1. An example of screenshots from the RC simulation leagues: (a) 2-D [5] and (b) 3-D [1]. The 2-D simulation league screenshot demonstrates players from two teams (represented by red and yellow circles) and their respective fields of view. The black arrows around player 11 illustrate how an agent can self-localize from observations of unique landmark features [12]. The 3-D simulation league screenshot demonstrates similar tactical overlays for each player.

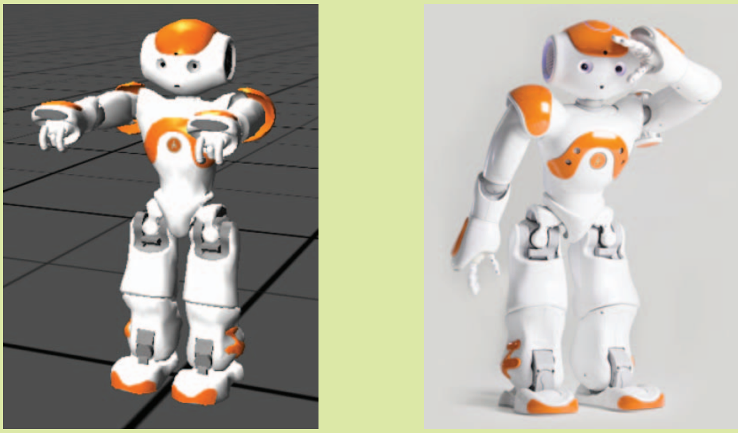
Enabling Replicable and Robust Analysis

Collectively, the simulation leagues provide ideal platforms for investigating emergent properties of complex robotic systems. Most team games and sports (both real and virtual) are characterized by rich and dynamic interactions that influence the contest outcome in a nontransitive manner. As described by Vilar et al. [21] "quantitative analysis is increasingly being used in team sports to better understand performance in these stylized, delineated, complex social systems." Early examples of such quantitative analysis include sabermetrics, which attempts to search for objective knowledge about baseball by considering the statistics of in-game activity [22]. A recent study by Fewell et al. [23] involved the analysis of basketball games as networks, with properties including degree centrality, clustering, entropy, and flow centrality (calculated from measurements of ball position throughout the game). This idea was extended by Vilar et al. [21], who considered the local dynamics of collective team behavior to quantify how teams occupy subareas of the field as a function of ball position. Recently, Cliff et al. [24] presented several information-theoretic methods of quantifying dynamic interactions in soccer games and used the RC 2-D simulation league as an experimental platform.

In addition to allowing high-level analysis of robotic systems overall, the simulation leagues provide inherent support for massively parallel processing. This property has been leveraged for the development and the analysis of algorithms with widespread applications in robotics, e.g., optimizing bipedal locomotion [10], [11], self-localization from noisy perception data [12], and planning complex multiagent strategies without direct agent-to-agent communication [16], [17]. Although simulation league agents have only noisy perception of their environment, the soccer server itself has perfect information regarding the global state, which enables replicable quantification of experimental performance (e.g., walk speed/stability and localization accuracy).

Wider Implications in Robotics Research

Robots are generally expensive to purchase, maintain, and transport, creating an intractably high entry barrier for institutes with limited access to research funding. By removing the requirement for physical robots, the RC simulation leagues allow such institutes to actively contribute to many fields of robotics research. To validate this assertion, Figure 3 presents the public expenditure on education as a percentage of gross domestic product (GDP) public expenditure on education (PEoE) at purchasing power parity per capita (GDP/cap)[20] for the home country of each participating RC 2013 team, averaged over each of the six largest RC leagues.



(a) (b)

Figure 2. The Webots-simulated NAO [3] is well-suited to global optimization across the high-dimensional parameter spaces characteristic of bipedal locomotion systems, as experiments may be automated, parallelized, and replicated exactly. This approach has proved very successful in identifying near-optimal parameter sets for subsequent local optimization on actual Aldebaran NAO robots [10], [11], highlighting the utility of simulation leagues for the investigation and the improvement of physical robotic systems. (a) The Webots-simulated NAO [3]. (b)The actual Aldebaran NAO (http://www.aldebaran.com).

Simulation League Case Study: Analysis of Competition Formats

The simulation league supports fully automated, massively parallel analysis of complex robotic systems, enabling replicable and robust investigation of algorithms and higher level emergent behaviors. In the following sections, we leverage these properties to expand upon our previous analysis of RC competition formats (e.g., round robin) to determine which best approximates the true performance rankings of competing teams [25].

The selection of an appropriate competition format is critical to both the success and the credibility of any competition. Unfortunately, this choice is not straightforward. The ideal format must minimize the randomness relative to the true performance ranking of teams while keeping the number of games to a minimum, to both satisfy time constraints and retain the interest of participants and spectators. Furthermore, maintaining competition interest introduces a number of constraints to competition formats, e.g., multiple games between the same two opponents (the obvious method of achieving a statistically significant ranking) should be avoided, making the resolution of nontransitive performance difficult.

Robocup Competition Formats

The following competition formats were adopted to determine the final rank of the top eight RC 2-D simulation league teams from 2012 to 2014.

- **2012:** The top four teams played six games each [three quarterfinals (round robin), two semifinals, and classification matches for first versus second and third versus fourth] and the bottom four teams played four games each.
- **2013:** A double-elimination system was adopted, where a team is ineligible for first place upon losing two games. A

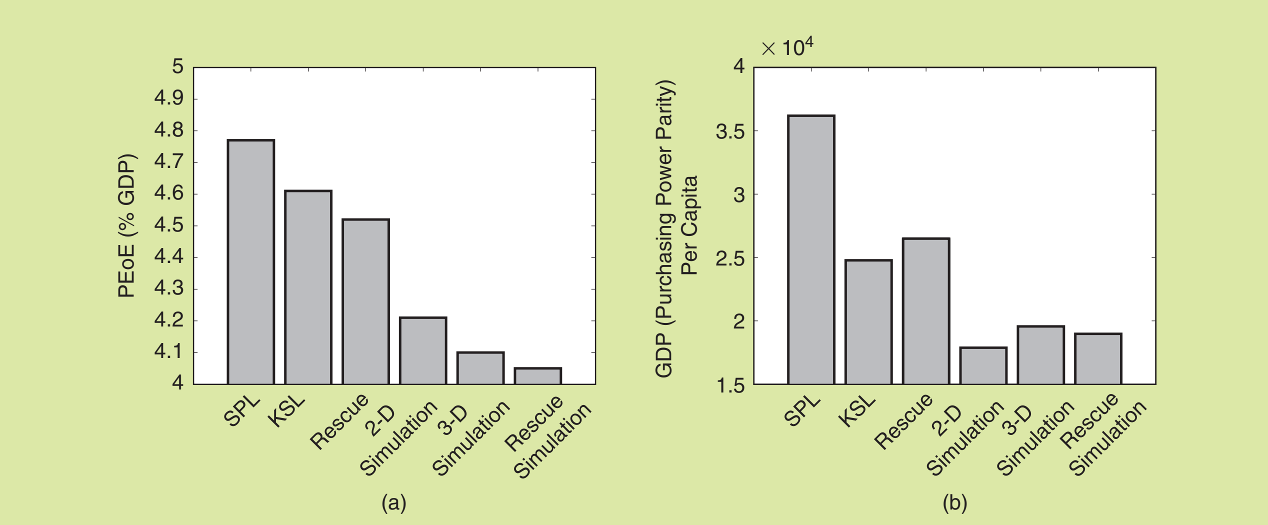


Figure 3. (a) The PEoE and (b) GDP/cap [20] for the home country of each participating RC 2013 team, averaged over each of the six largest RC leagues. Each of the three major simulation leagues (2-D, 3-D, and simulated rescue) exhibits significantly lower values than those requiring the purchase or development of physical robots [SPL, kid-sized league (KSL), and physical rescue].

total of 14 games were played in double-elimination format (i.e., $2n - 2$, $n = 8$) followed by two classification games.

- **2014:** Our proposed format was adopted [25]. In particular, round-robin games were conducted for the top eight teams (28 games) followed by four classification games for first versus second, third versus fourth, and so on.

Previously, it has been unclear whether these changes in the competition format improve the fairness and the reproducibility of the final team rankings. In general, lack of reproducibility is due to the nontransitivity of team performance (a well-known phenomena that occurs frequently in actual human team sports).

Methods of Ranking Team Performance

Before evaluating different competition formats, it is necessary to establish a fair (i.e., statistically significant) ranking of the top eight RC 2-D simulation league teams from previous years. This was accomplished by conducting an eight-team round robin for previous years, where all 28 pairs of teams play approximately 1,000 games against one another. In addition, two different schemes were considered for point calculation [25].

- **Continuous Scheme:** Teams are ranked by the sum of average points obtained against each opponent across all 1,000 games.
- **Discrete Scheme:** The average score between each pair of teams (across all 1,000 games) is rounded to the nearest integer (e.g., 1.9:1.2 is rounded to 2:1). Next, points are allocated for each pairing based on these rounded results: 3 for a win, 1 for a draw, and 0 for a loss. Teams are then ranked by the sum of these points received against each opponent.

The final rankings generated for the 2012 and 2013 RoboCup 2-D simulation league teams under these two schemes are presented in [25]. Although both the schemes have been shown to generate statistically robust results, we have chosen to adopt the continuous scheme for this article to avoid the boundary effects inherent to discretization. In order to formally capture the overall difference between two rankings $\mathbf{r^a}$ and $\mathbf{r^b}$, the L_1 distance was utilized:

$$d_1(\mathbf{r^a}, \mathbf{r^b}) = \| \mathbf{r^a} - \mathbf{r^b} \|_1 = \sum_{i=1}^{n} | r_i^a - r_i^b |, \quad (1)$$

where i is the index of the ith team in each ranking, $1 \le i \le 8$.

Results

Following the iterated round-robin and continuous ranking scheme described in the "Methods of Ranking Team Performance" section, statistically significant rankings were generated for the top eight RC 2-D simulation league teams for 2012–2014. The full set of experiments is described in [25], which verified that the proposed format (later adopted for RC 2014) consistently outperformed the other candidates in terms of approximating the true rankings for RoboCup 2012 and 2013. We expand upon this analysis to incorporate the results for RC 2014.

The L_1 distance [see (1)] was used to capture the discrepancy between RC final results, $\mathbf{r^a}$, and the statistically significant rankings generated from the 28,000-game round robin, $\mathbf{r^c}$:

$$\begin{aligned} d_1(\mathbf{r^a}, \mathbf{r^c})_{2012} &= 12 \\ d_1(\mathbf{r^a}, \mathbf{r^c})_{2013} &= 12, \end{aligned} \quad (2)$$

where the competition formats for RC 2012 and 2013 are described in the "RC Competition Formats" section. We can also quantify that the corresponding discrepancy, $\mathbf{r^p}$, had our proposed competition format [25] been used for those competitions:

$$\begin{aligned} d_1(\mathbf{r^p}, \mathbf{r^c})_{2012} &= 4 \\ d_1(\mathbf{r^p}, \mathbf{r^c})_{2013} &= 6. \end{aligned} \quad (3)$$

Our proposed format was subsequently adopted for the RC 2014 finals. The divergence from statistically robust team rankings was equivalently small:

$$d_1(\mathbf{r}^a, \mathbf{r}^c)_{2014} \Longleftrightarrow d_1(\mathbf{r}^p, \mathbf{r}^c)_{2014} = 4. \tag{4}$$

To statistically validate that the proposed competition format is significantly more appropriate than those adopted at RC 2012 and RC 2013, 10,000 tournaments were generated for each format by randomly sampling the game results from the 28,000-game round robin. For each tournament, the L_1 distance $d_1(\mathbf{r}^a, \mathbf{r}^b)$ [see (1)], was calculated to capture the discrepancy between the tournament and the true team rankings. These results are presented in Figure 4 for the top eight teams from RC 2012–2014. It is evident that the proposed format yields more statistically robust rankings (i.e., smaller L_1 distance) than the formats adopted in previous years.

In addition to comparing the accuracy of team rankings under different competition formats, it is interesting to compare the team performance against a consistent benchmark. Before 2013, it was commonplace for the simulation league teams to optimize their performance against the default Agent2D code [26], which is reflected in the high Spearman's rank correlation coefficient $(\rho_{2012} = 0.98)$ between true rankings and goal difference against Agent2D (across 1,000 games per team). Since 2013, these correlations have decreased substantially $(\rho_{2013} = 0.55$ and $\rho_{2014} = 0.57)$with teams opting to optimize behavior against the binaries published by top-performing teams postcompetition to gain a competitive advantage with opponent-specific strategy. The average goal difference against Agent2D decreased for the top four RC teams between 2013 and 2014 accordingly. Importantly, this level of behavioral complexity (in addition to our analysis of competition formats) would be impossible without the support for massively parallel processing inherent to simulation leagues.

Summary and Discussion

Continual increases in data volume and computational power have led to increased complexity in the experimental methodologies across most fields of research. Therefore, it is unsurprising that many fields (particularly in the life sciences [27]) have recently placed increased focus on enabling measurable, replicable, and statistically robust results. Although robotics researchers face many unique challenges due to the expense and stochasticity inherent to physical robots, we propose that physically realistic simulated environments (epitomized by the RC simulation leagues) have an important and widespread role to play in the future of robotics.

The simulation leagues often serve as platforms for the initial development and the evaluation of software modules for later integration into physical robots [10], [11], and many of these modules have applications beyond the RC domain (e.g., localization and mapping [12]). They also enable the investigation of high-level emergent properties of complex robotic systems, as demonstrated in a recent study by Cliff et al. [24] that presents novel information-theoretic methods for quantifying dynamic interactions in a multiagent context.

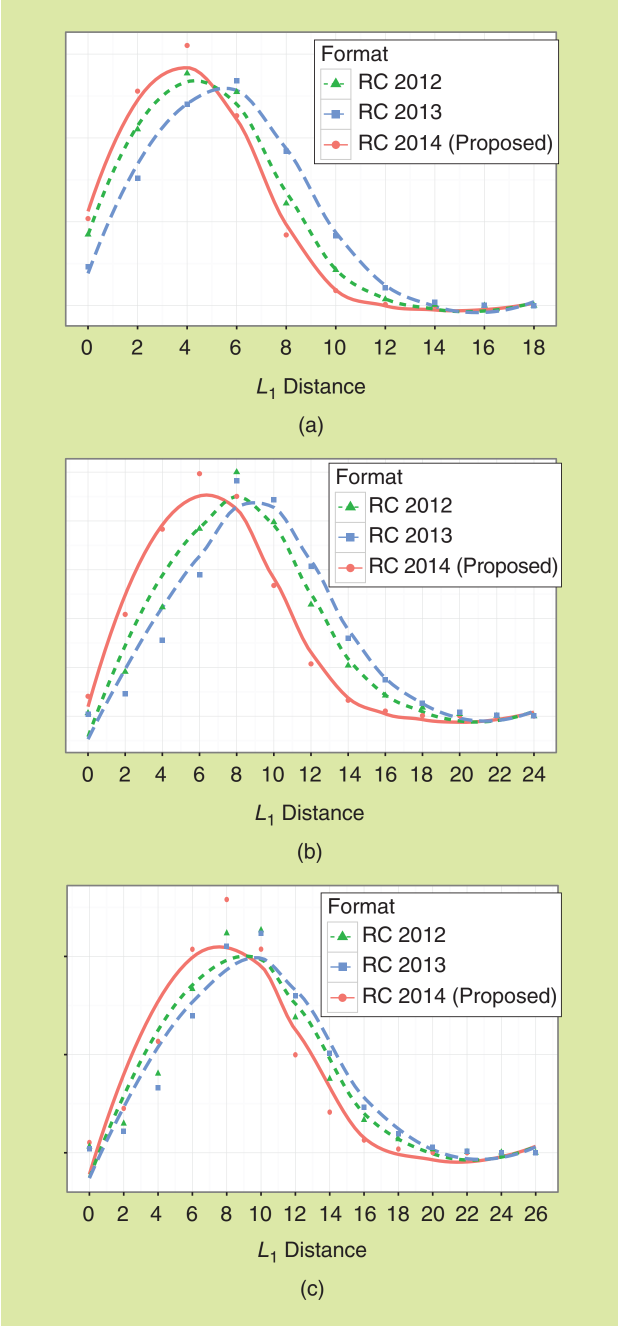


Figure 4. The discrepancy between tournament and true team rankings, captured as an L_1 distance (1), for 10,000 randomly generated tournaments structured according to the three considered formats: (a) 2012, (b) 2013, and (c) 2014. It is evident that the proposed format (red) yields more statistically robust rankings (i.e., a smaller L_1 distance) than the formats adopted in RC 2012 (green) and 2013 (blue), considering the top eight teams from each RC.

In this article, we have provided an overview of the RoboCup simulation leagues (both 2-D and 3-D) and described their properties as they pertain to replicable and robust robotics research. To demonstrate their utility directly, we leverage the ability to run massively parallelized experiments to evaluate different competition formats (e.g., round robin) for the RC 2-D simulation league. Our results demonstrate that a hybrid format [25] minimizes fluctuations from true (statistically significant) team performance rankings within the time constraints of the RC world finals.

Our experimental analysis and many others in [10], [11], [12], and [24] would be impossible with physical robots alone, and have widespread applications beyond the scope of simulated soccer matches. We encourage other researchers to explore the potential for enriching their experimental pipelines with simulated components to minimize the experimental costs and enable others to replicate and expand upon experimental results in a hardware-independent manner.

References

[1] J. Boedecker and M. Asada, "SimSpark—Concepts and applications in the RoboCup 3D soccer simulation league," in *Proc. Workshop SIMPAR Int. Conf. Simulation, Modeling Programming Autonomous Robots*, 2008, pp. 174–181.

[2] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2004, vol. 3, pp. 2149–2154.

[3] O. Michel, "Webots: Symbiosis between virtual and real mobile robots," in *Virtual Worlds*. Berlin, Germany: Springer-Verlag, 1998, pp. 254–263.

[4] S. Carpin, M. Lewis, J. Wang, S. Balakirsky, and C. Scrapper, "USARSim: A robot simulator for research and education," in *Proc. IEEE Int. Conf. Robotics Automation*, 2007, pp. 1400–1405.

[5] M. Chen, E. Foroughi, F. Heintz, Z. Huang, S. Kapetanakis, K. Kostiadis, J. Kummeneje, I. Noda, O. Obst, P. Riley, T. Steffens, Y. Wang, and X. Yin, "RoboCup soccer server," Tech. Rep., Aug. 2002.

[6] RoboCup Technical Committee. (2013). RoboCup Soccer Simulation League 3D competition rules and setup. [Online]. Available: http://homepages.herts.ac.uk/sv08aav/RCSoccerSim3DRules2013.1.pdf

[7] O. Michel, Y. Bourquin, and J.-C. Baillie, "Robotstadium: Online humanoid robot soccer simulation competition," in *RoboCup 2008: Robot Soccer World Cup XII*. Berlin, Germany: Springer, 2009, pp. 580–590.

[8] C. Aguero, N. Koenig, I. Chen, H. Boyer, S. Peters, J. Hsu, B. Gerkey, S. Paepcke, J. L. Rivero, J. Manzo, E. Krotkov, and G. Pratt, "Inside the virtual robotics challenge: Simulating real-time robotic disaster response," *IEEE Trans. Automat. Sci. Eng.*, vol. 12, no. 2, pp. 494–506, 2015.

[9] S. Balakirsky, R. Madhavan, and C. Scrapper, "NIST/IEEE virtual manufacturing automation competition: From earliest beginnings to future directions," in *Proc. 8th Workshop Performance Metrics Intelligent Systems*, 2008, pp. 214–219.

[10] P. MacAlpine, S. Barrett, D. Urieli, V. Vu, and P. Stone, "Design and optimization of an omnidirectional humanoid walk: A winning approach at the robocup 2011 3D simulation competition," in *Proc. 26th Conf. Artificial Intelligence*, 2012, pp. 1047–1053.

[11] C. Niehaus, T. Röfer, and T. Laue, "Gait optimization on a humanoid robot using particle swarm optimization," in *Proc. IEEE 2nd Workshop Humanoid Soccer Robots*, 2007, pp. 1–7.

[12] D. Budden and M. Prokopenko, "Improved particle filtering for pseudo uniform belief distributions in robot localisation," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer-Verlag, 2013.

[13] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "RoboCup: The robot world cup initiative," in *Proc. 1st Int. Conf. Autonomous Agents*, 1997, pp. 340–347.

[14] D. Budden and A. Mendes, "Unsupervised recognition of salient colour for real-time image processing," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[15] J. Fountain, J. Walker, D. Budden, A. Mendes, and S. Chalup, "Motivated reinforcement learning for improved head actuation of humanoid robots," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[16] M. Prokopenko, O. Obst, P. Wang, D. Budden, and O. Cliff, "Gliders2013: Tactical analysis with information dynamics," in *Proc. RoboCup Symp. Competitions: Team Description Papers*, Eindhoven, The Netherlands, June 2013.

[17] M. Prokopenko, P. Wang, and O. Obst, "Gliders2014: Dynamic tactics with voronoi diagrams," in *Proc. RoboCup Symp. Competitions: Team Description Papers*, Joao Pessoa, Brazil, July 2014, pp. 1–6.

[18] A. Bai, X. Chen, P. MacAlpine, D. Urieli, S. Barrett, and P. Stone, "Wrighteagle and UT Austin villa: RoboCup 2011 simulation league champions," in *RoboCup 2011: Robot Soccer World Cup XV*. Berlin, Germany: Springer, 2012, pp. 1–12.

[19] I. Noda and P. Stone, "The RoboCup soccer server and CMUnited clients: Implemented infrastructure for MAS research," *Auton. Agents Multi-Agent Syst.*, vol. 7, nos. 1–2, pp. 101–120, 2003.

[20] World Bank. (2012). World development indicators 2012: GDP per capita, PPP (current international $). [Online]. Available: http://data.worldbank.org/data-catalog/world-development-indicators and http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

[21] L. Vilar, D. Araújo, K. Davids, and Y. Bar-Yam, "Science of winning soccer: Emergent pattern-forming dynamics in association football," *J. Syst. Sci. Complexity*, vol. 26, no. 1, pp. 73–84, 2013.

[22] D. Grabiner. (2004). The sabermetrics manifesto. [Online]. Available: http://seanlahman.com/baseball-archive/

[23] J. Fewell, D. Armbruster, J. Ingraham, A. Petersen, and J. Waters, "Basketball teams as strategic networks," *PloS one*, vol. 7, no. 11, p. e47445, 2012.

[24] O. Cliff, J. Lizier, R. Wang, P. Wang, O. Obst, and M. Prokopenko, "Towards quantifying interaction networks in a football match," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[25] D. Budden, P. Wang, O. Obst, and M. Prokopenko, "Simulation leagues: Analysis of competition formats," in *RoboCup 2014: Robot Soccer World Cup XVIII*. Berlin, Germany: Springer, 2014.

[26] H. Akiyama and T. Nakashima, "HELIOS Base: An open source package for the RoboCup soccer 2D simulation," in *RoboCup 2013: Robot World Cup XVII*. Berlin, Germany: Springer, 2014, pp. 528–535.

[27] D. G. Hurley, D. M. Budden, and E. J. Crampin, "Virtual reference environments: A simple way to make research reproducible," *Briefings in Bioinformatics*, 2014.

David M. Budden, The University of Melbourne, Parkville, Victoria, Australia. E-mail: david.budden@unimelb.edu.au.

Peter Wang, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia. E-mail: peter.wang@csiro.au.

Oliver Obst, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia. E-mail: Oliver.Obst@csiro.au.

Mikhail Prokopenko, The University of Sydney, New South Wales, Australia. E-mail: mikhail.prokopenko@sydney.edu.au.