CS533 Project: Reactive Control of a Two-Body Point Absorber Using RL

Leila Ghorbanzadeh

Spring 2018

Abstract—This project represents Reinforcement learning applications for the optimal resistive control of a wave energy converter. The linear and non-linear function approximation methods are chosen to maximize energy extraction in each sea state. Damping and stiffness coefficients changes are applied to the controller to observe reward which is a function of absorbed power. The algorithm performance is evaluated through the models and simulator results for irregular waves. Two function approximators namely linear and non-linear function approximators are used to learn wave state pattern and the optimum action required to take. Results represent the satisfactory performance of non-linear function approximation. Results show that linear function approximator model convergences somehow but accuracy is not very high.

I. Introduction

Wave energy converters (WEC) are developing to get the power from ocean waves to reduce the dependency on fossil fuels. It is estimated that a potential of up to 2.1 TW of power globally can be extracted from OWCs. Different types of WEC namely attenuator, point absorber, oscillating wave surge converter, oscillating water column and overtopping/Terminator device have been developed according to different sea states and their potential performance (Falc~ao,2010). Deploying WEC is not economically viable, having effective control strategy is crucial in order to address this problem, in addition, to get high potential power without hardware costs.

Different control strategies have been applied and tested on WEC and each one has its own pros and cons in terms of mechanical and electrical performance. Model predictive control is dependent on the expected force to achieve over a specific time horizon (Brekken, 2011; Hals et al., 2011; Li and Belmont, 2014a; Richter et al., 2014). By having current excitation force and including constraints on the motions and load capacity. Resistive and reactive control are control strategies applied to WECs which mainly are dependent on time-averaged sea states, therefore stationary wave conditions should be taken into account.

The biggest disadvantage of aforementioned methods is that the optimal action depends on the internal models of the body dynamics. Therefore, the performance of these techniques could be significantly be affected by modeling errors. Moreover, as time passes, these control methods, these control strategies do not consider changes in the device dynamics because of some non- critical subsystem failures. For these reasons, the application of reinforcement learning (RL) to the reactive control of WECs has been proposed in the recent works (Anderlini, 2016). With proposed RL algorithm, the controller learns the optimal PTO damping coefficient from experiencing every ocean wave state.

II. System description

In this work, point absorber developed by Sandia National Laboratories is used as a case study. The dynamic of the WEC is represented in (Neary and Previsic, 2014) in detail which is a floating, axisymmetric, moored point absorber comprising of two bodies. The schematic of point absorber with a hydraulic PTO system as shown in Fig. 1. The mechanical energy associated resulted from ocean motion is transformed into electrical energy.

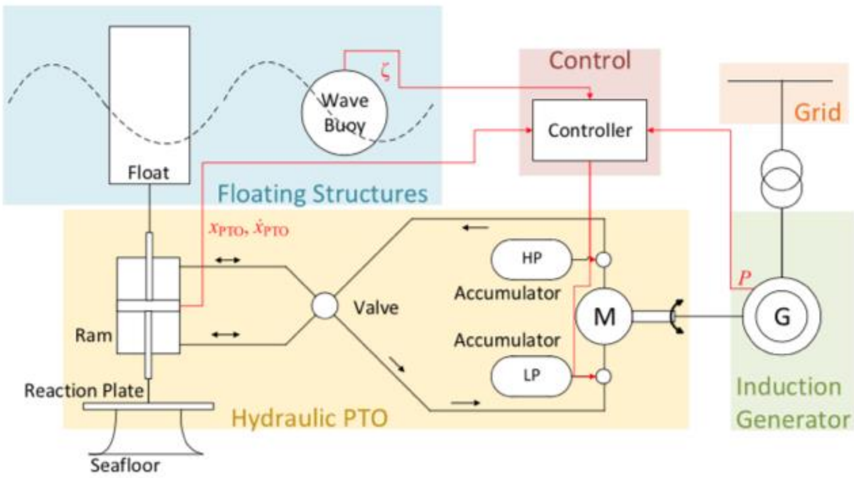


Fig. 1. Diagram of the point absorber with its hydraulic PTO.

As it can be seen from Figure 1, the displacement and velocity of the PTO, power and wave elevation derived from sea state are the system inputs. By changing valves connected to the accumulators, the controller can adjust the flow of the device.

III. Reinforcement Learning

In reinforcement learning, the system learns an optimal action or policy by interacting with the environment and observing the outcome considered as a reward (Sutton and Barto, 1998). Since the state of this system is continuous, function approximation methods have been proposed and tested. Function approximation can be represented in linear and no-linear methods. Both techniques are applied to this problem and evaluated. To order to obtain function approximation, we need to set up state space, possible actions, and rewards which are as follows:

● **State space**

State space has been created from irregular sea states components like significant height ranging from 0 to 10 meters and period ranging from 5 to 15 seconds plus dynamic character of the system namely damping and stiffness coefficients. Generally, each sea states could last for around 20 minutes, but in this work, each sea state is produced for 15 minutes each.

● **Actions**

Actions for this system are increasing or decreasing damping and stiffness coefficients of the system.

● **Rewards**

Rewards are set according to the absorbed power when the range -1 MW to 1 MW. Better to mention that, the negative power is due to reactive control of the system.

IV. Approaches

As I mentioned before, for the control of point absorber, function approximation using reinforcement learning has been developed in this work. The first one is based on linear function approximation using Q-learning and the second one is neural-network based non-linear approximation called Deep Q-learning Algorithm.

- **Linear function approximation**

To get a linear function approximation, a set of features describing the model is required. During each observation, the weight of features is updated to reach the optimum correlation model between states and Q-value. ϕ and θ_0 represent features and related weights respectively of the function. Meaning that:

$$Q(s,a)=\theta_0\cdot 1+\theta_1\phi_1(s,a)+\cdots+\theta_n\phi_n(s,a)$$

In this work, state features are radiation force, excitation force, displacement and velocity dominant frequency, magnitude, and phase plus damping, stiffness and observed power (15 features in total).

- **Non-linear function approximation (Deep Q-learning Algorithm)**

Sometimes, linear functions do not converge due to some reasons like lacking enough features or maybe features cannot represent state values well. Therefore, nonlinear function approximation is proposed. There are different methods of function approximation but the one applied in this work is Q-learning algorithm with experience replay with the following pseudo-code:

initialize replay memory D

initialize action-value function Q with random weights

observe initial state s

repeat

select an action a

with probability ε select a random action

otherwise select a = argmaxa'Q(s,a')

carry out action a

observe reward r and new state s'

store experience <s, a, r, s'> in replay memory D

sample random transitions <ss, aa, rr, ss'> from replay memory D

calculate target for each minibatch transition

if ss' is terminal state then tt = rr

otherwise tt = rr + γmaxa'Q(ss', aa')

train the Q network using (tt - Q(ss, aa))^2 as loss

s = s'

until terminated

V. Results

The performance of the two employed methods is evaluated on 50 randomly chosen states. This is done by using features of each state as inputs and getting the best possible action to take using linear Q-function. As the optimum action is chosen, then the reward can be observed by taking that action. In this work, the agent is designed in Simulink environment attached beside the report.

Figure 2 and 3 represents optimal and predicted damping, stiffness coefficients and the differences between optimal power and predicted power respectively using linear function approximation.

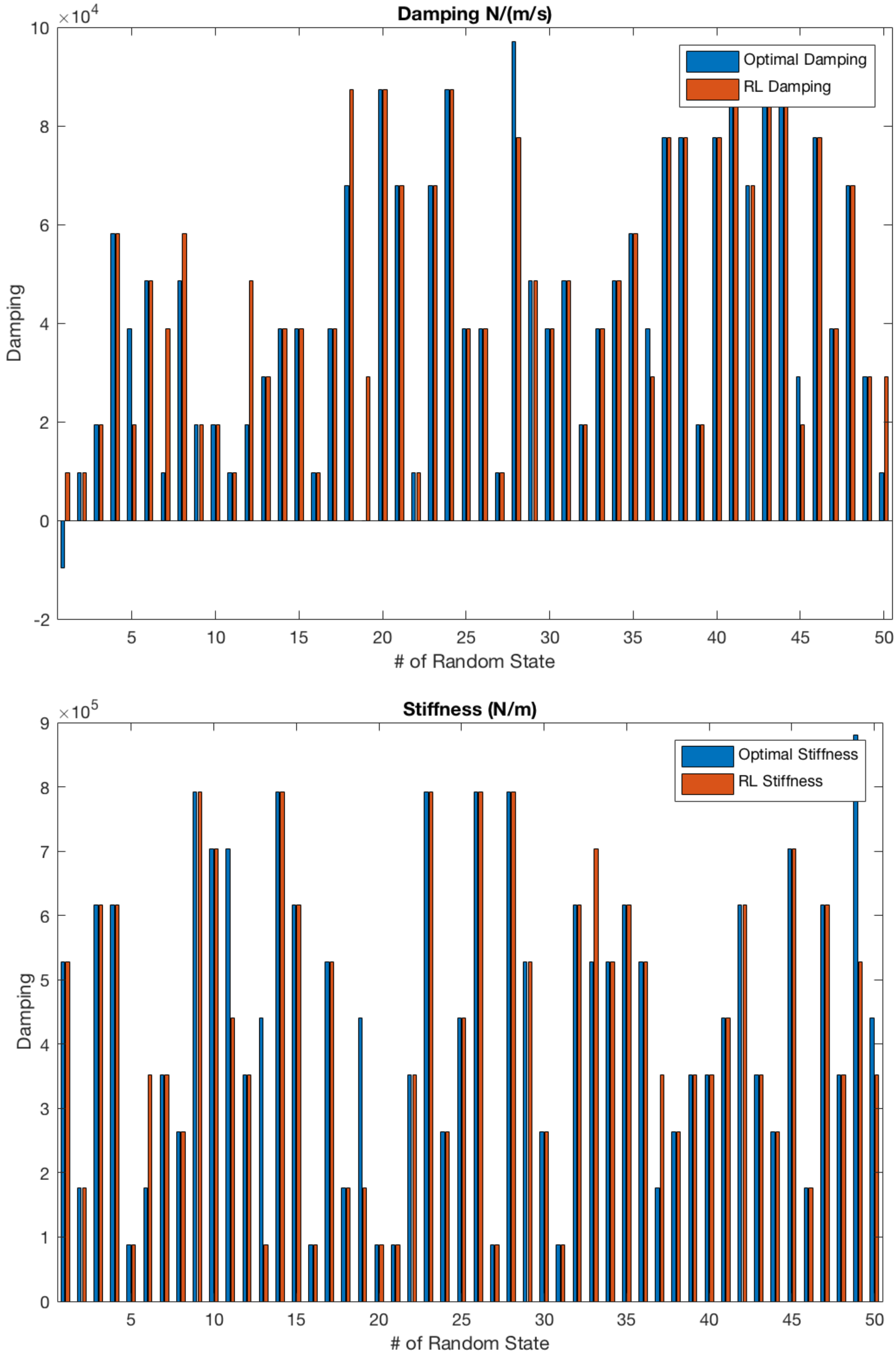


Fig. 2: Optimal vs estimated optimal daming and stiffness Coefficients by Linear function approximation

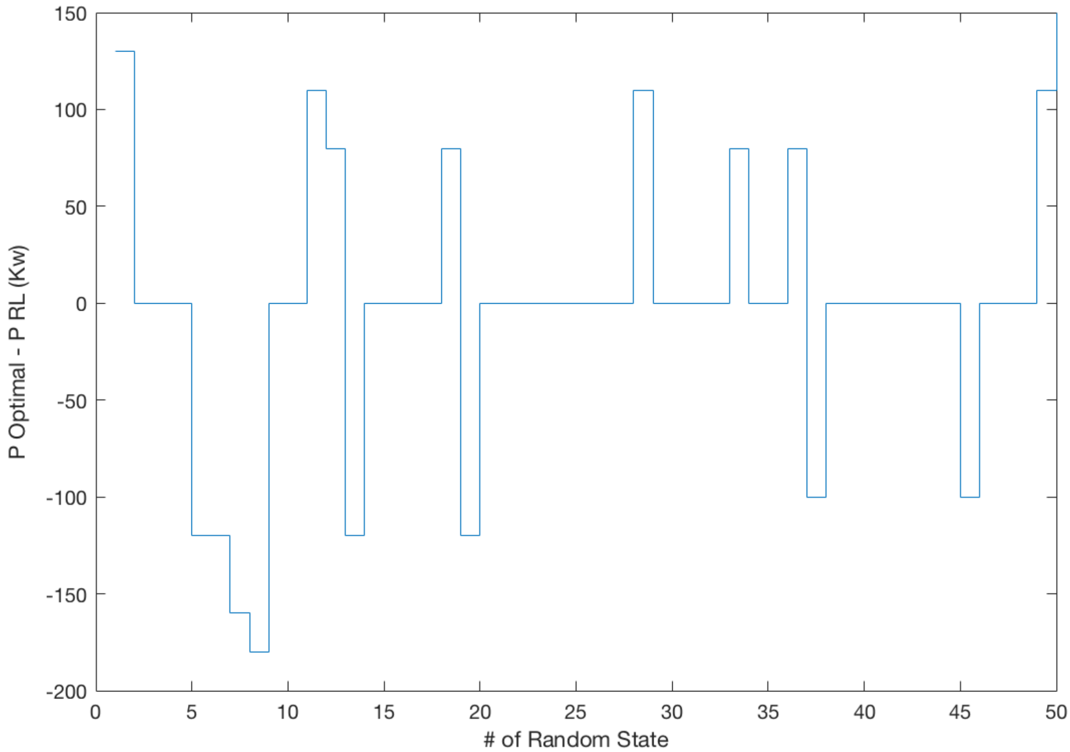
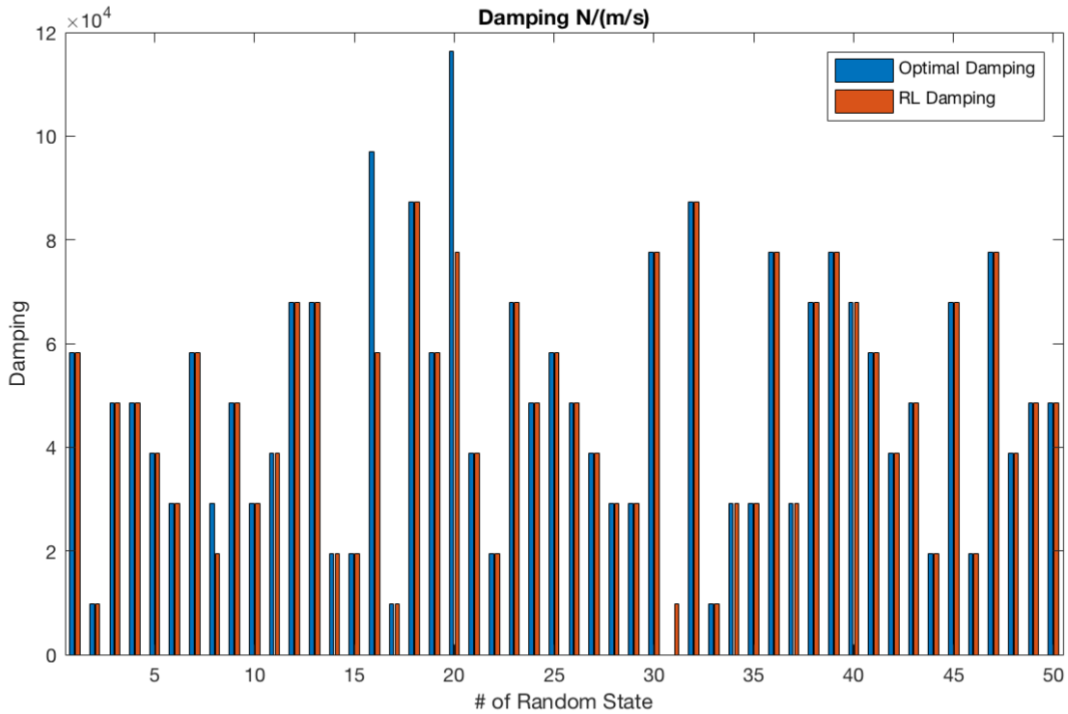


Fig. 3: Power differences between optimal and predicted optimal power by Linear function approximation

Figure 4,5 represents optimal vs predicted damping, stiffness coefficients and the difference between optimal power and predicted power using Deep_Q learning method respectively.



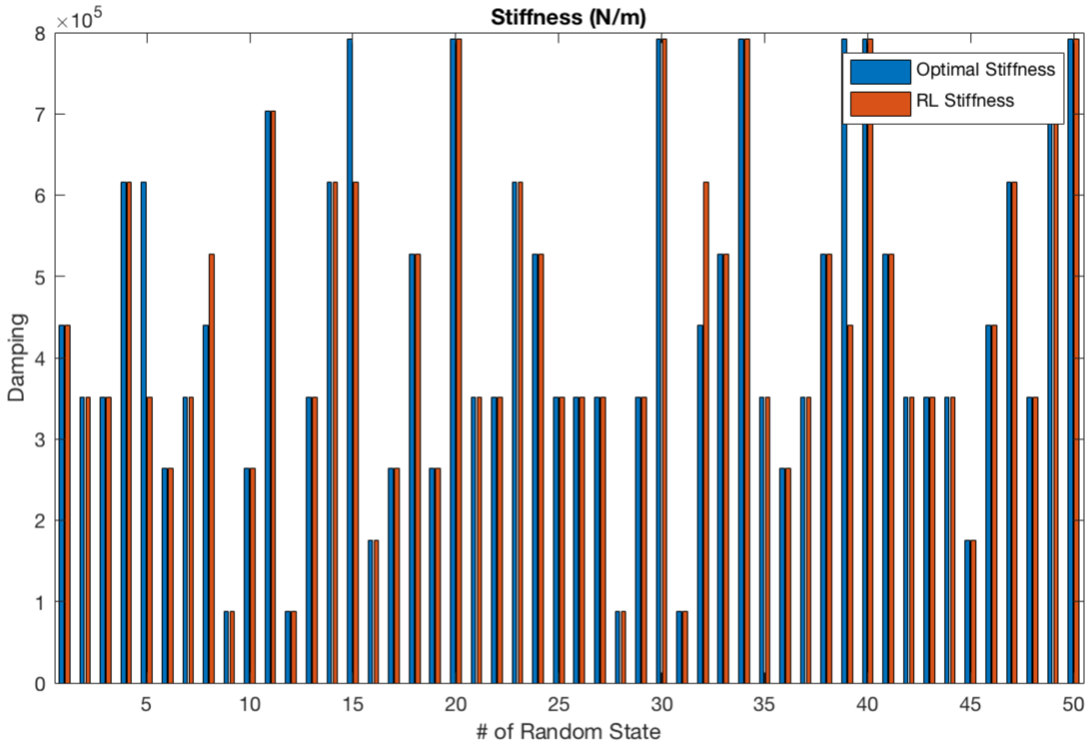


Fig. 5: Optimal vs estimated optimal daming and stiffness Coefficients by non-Linear function approximation

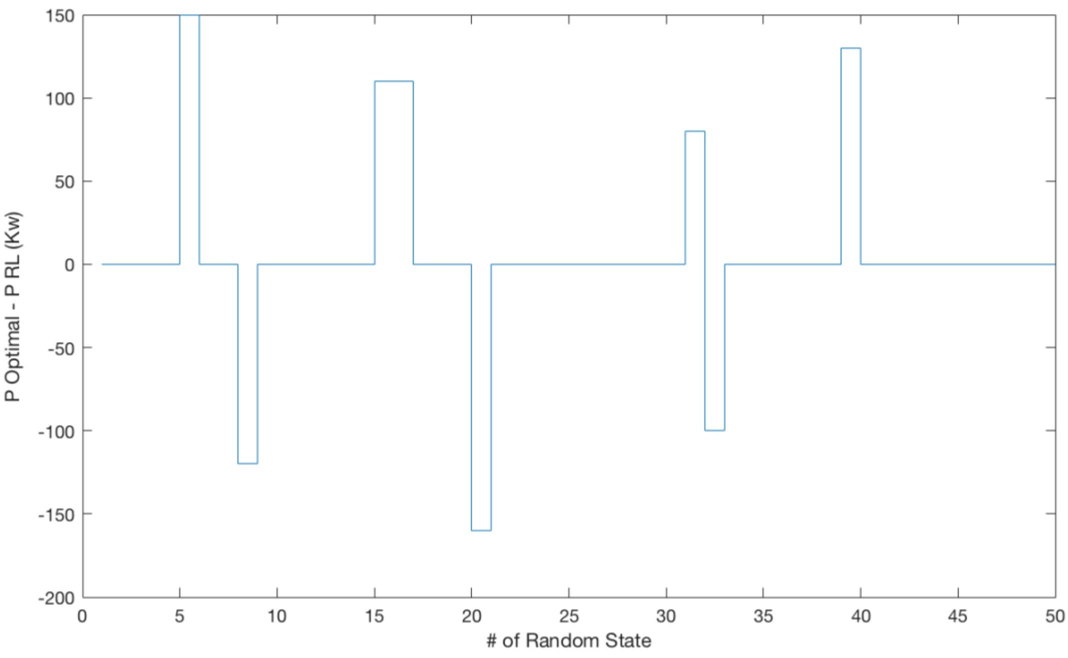


Fig. 5: Power differences between optimal and predicted optimal power by non-Linear function approximation

As it can be seen from the plots the accuracy of the linear approximators is not very good but still giving correct answers for around %66 of the evaluated cases.

On the other hand, the accuracy of Deep Q-Learning is noticeable reaching %84 success for the evaluated states.

VI. Conclusion

Two approaches are applied and evaluated in this work to control the point absorber and results represent that linear function may not converge to the optimal policy very well or convergence rate may not be very high but Deep Q-Learning can provide the model with better function approximation. From the results, it can be seen that, even for states with wrong policy, still performance is good and power drop is about 100 KW. There is also negative power in the results which represent reactive control characteristic of the system.

VII. References

Falc~ao, A.F.D.O., 2007. Modelling and control of oscillating-body wave energy converters with hydraulic power take-off and gas accumulator. Ocean Eng. 34 (14–15), 2021–2032.

Brekken, T.K.A., 2011. On Model Predictive Control for a point absorber Wave Energy Converter. In: Proceedings of the IEEE Trondheim PowerTech, 1–8.

Anderlini, E., Forehand, D.I.M., Stansell, P., Xiao, Q., Abusara, M., 2016. Control of a point absorber using reinforcement learning. Trans. Sustain. Energy 7 (4), 1681–1690.

Neary, V.S., Previsic, M., Jepsen, R.a., Lawson, M.J., Yu, Y.-H., Copping, A.E., Fontaine, A.a., Hallett, K.C., 2014. Methodology for Design and Economic Analysis of Marine Energy Conversion (MEC) Technologies. Tech. Rep. March, Sandia National Laboratories.

Previsic, M., Shoele, K., Epler, J., 2014. Validation of theoretical performance results using wave tank testing of heaving point absorber wave energy conversion device working against a subsea reaction plate. 2nd Marine Energy Technology Symposium, 1–8.

Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning. Hardcover Edition, MIT Press. Tedeschi, E., Carraro, M., Molinas, M., Mattavelli, P., 2011.