A Reinforcement Learning Based Load Frequency Control for Power Systems considering Nonlinearities and Other Control Interactions.

Carlos Mauricio Bula Oyuela

Universidad Nacional de Colombia
Facultad de Ingeniería
Bogotá DC
Sede Bogotá, Colombia
2024

A Reinforcement Learning Based Load Frequency Control for Power Systems considering Nonlinearities and Other Control Interactions.

Carlos Mauricio Bula Oyuela

Tesis presentada como requisito parcial para optar por el título de:
Magíster (M, MSc) en Automatización Industrial - Investigación

Director(a):
Sergio Raúl Rivera Rodríguez
Línea de investigación:
Reinforcement Learning
Grupo de investigación:
Inteligencia Computacional Aplicada a Sistemas de Potencia

Universidad Nacional de Colombia
Facultad de Ingeniería
Bogotá DC
2024

Per aspera ad astra - A través de las dificultades a las estrellas.

Séneca

Declaración

Me permito afirmar que he realizado ésta tesis de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados el presente texto. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los he reconocido en el presente trabajo. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de tesis.

Sede Bogotá., January

Carlos Mauricio Bula Oyuela

Agradecimientos

Quiero expresar mi profundo agradecimiento a todas las personas que han contribuido de manera significativa a la realización de esta tesis. Este logro no habría sido posible sin el apoyo inquebrantable de quienes me rodean.

En primer lugar, quiero agradecer a mi director de tesis, Sergio Rivera, por su dedicación, paciencia y excelente guía para que tomara el camino al desarrollo de esta tesis. Sus valiosos consejos y apoyo han sido fundamentales para el desarrollo de este trabajo y expandir mi comprensión en el campo de la inteligencia computacional. Además del apoyo en otras decisiones que tuvieron un agradable desenlace.

Agradezco sinceramente a mis profesores y mentores que durante mi estadía en la universidad compartieron su conocimiento y experiencia durante el desarrollo de mi investigación. Su guía fue esencial para enriquecer mi comprensión de los métodos usados en esta tesis y alcanzar los objetivos propuestos.

A mis amigos más cercanos y compañeros de trabajo les agradezco por su apoyo constante y por las conversaciones estimulantes que han enriquecido mi perspectiva. Sus palabras han mentenido en alto mis ánimos y fueron cruciales en los momentos que requerían mayor exigencia de mi parte.

A mi madre, le estoy infinitamente agradecido por su amor, comprensión y paciencia. A pesar de su desconocimiento de cualquier temas que se tratan en esta tesis y en mis estudios, su apoyo incondicional y palabras de aliento me ha dado la fortaleza necesaria para llegar hasta aquí. Gracias a mi padre y mis hermanos, cuya presencia y confianza me impulsan a avanzar en mi desarrollo personal.

Por último, pero no menos importante, agradezco a todas las fuentes de financiamiento que hicieron posible llevar a cabo esta investigación. Su respaldo económico fue esencial para la recopilación de datos y la realización de experimentos.

Este trabajo no solo es mío, sino de todos aquellos que generosamente compartieron su tiempo, conocimiento y apoyo. A cada uno de ustedes, gracias por ser parte fundamental de este viaje académico.

Listado de símbolos y abreviaturas

- **UN**: United Nations
- **AC**: Alternating Current
- **ACE** - Area Control Error
- **DC**: Direct Current
- **HVDC**: High Voltage Direct Current
- **DER**: Distributed Energy Resources
- **RoCoF**: Rate of Change of Frequency
- **MDP**: Markov Decision Process
- **ISE**: Integral Square Error
- **IAE**: Integral Absolute Error
- **ITSE**: Integral Time Square Error
- **ITAE**: Integral Time Absolute Error
- **LFC**: Load Frequency Control
- **AGC**: Automatic Generation Control
- **PFC**: Primary Frequency Control
- **SFC**: Secondary Frequency Control
- **TFC**: Tertiary Frequency Control
- **EC**: Emergency Control
- **ELS**: Emergency Load Shedding
- **UFLS**: Under Frequency Load Shedding
- **UVLS**: Under Voltage Load Shedding
- **AVR**: Automatic Voltage Regulator
- **PSS**: Power System Stabilizer
- **FACTS**: Flexible AC Transmission Systems
- **RL**: Reinforcement Learning
- **PI**: Proportional-Integral

- **ADRC**: Active Disturbance Rejection Control
- **RNNs** - Recurrent Neural Networks
- **LSTM** - Long Short-Term Memory
- **PPO** - Proximal Policy Optimization
- **WECC** - Western Electricity Coordinating Council
- δ_1, δ_2 - Angles of both ends
- V_1, V_2 - Voltages of both ends
- X_{12} - Reactance of the tie-line
- $P_{tie,12}$ - Power transference through the tie-line
- $\Delta P_{tie,12}$ - Change in transferred power through the tie-line
- K_{s12} - Synchronizing coefficient between two areas
- $\Delta\omega_i$ - Angular frequency deviation of area i
- s - Laplace transform variable
- Δf_i - Frequency deviation of area i
- β_i - Frequency bias of area i
- D_i - Damping coefficient of area i
- R_i - Governor droop of area i
- H_i - Inertia constant of area i
- $S_{n,i}$ - Rated power of the generator in area i
- H_{con} - Total inertia constant in interconnected power systems
- $\Delta P_{tie,i}$ - Tie-line power variation for area i
- ΔP_{RES} - RES power fluctuation
- H_{sys} - Equivalent system inertia
- ΔP_D - Total load-generation imbalance
- θ: Parameters of the approximation function
- V: State value function
- Q: State-action value function
- π: Policy
- (r_i): Step return

Resumen

Un Control de Frecuencia basado en Aprendizaje por Refuerzo para Sistemas de Potencia, considerando No Linealidades y Otras Interacciones de Control.

Dada la evolución de las civilizaciones y la existencia limitada de recursos, hay crecientes preocupaciones sobre la sostenibilidad de ese progreso. Existe una creciente necesidad de fuentes de energía asequibles, confiables y sostenibles. Diferentes organismos internacionales reconocen el papel crucial de la energía en conseguir la sostenibilidad global, destacando la necesidad de una transición hacia alternativas energéticas modernas y ecológicas. Esta transición a la energía renovable añade complejidades en los sistemas de control, requiriendo controladores de frecuencia confiables. Este documento discute las limitaciones de las estrategias tradicionales de control proporcional-integral (PI) y estudia el uso de un algoritmo basado en aprendizaje por refuerzo, específicamente el algoritmo de Optimización de Política Próxima (PPO), en el control del sistema de potencia. Los resultados demuestran la eficacia de PPO al reducir eficazmente el error de control de área, manejar no linealidades y adaptarse a las perturbaciones en el sistema. La capacidad de adaptación de PPO a diferentes situaciones y cambios en el sistema, junto con su naturaleza agnóstica del sistema, lo convierte en una opción prometedora para mejorar el control del sistema de potencia. Finalmente, se reconocen consideraciones prácticas como requisitos computacionales, retrasos en la comunicación, ruido de medición y limitaciones de hardware, como desafíos que necesitan una exploración adicional para la implementación del mundo real de PPO en entornos de redes eléctricas. El artículo concluye instando a investigaciones adicionales para abordar estos desafíos y optimizar el algoritmo para aplicaciones en el mundo real, haciendo énfasis en sus ventajas para la transición hacia las energías renovables.

Palabras clave: control de frecuencia, aprendizaje por refuerzo, optimización de política próxima, energías renovables

Abstract

A Reinforcement Learning Based Load Frequency Control for Power Systems considering Nonlinearities and Other Control Interactions.

Considering the advances of civilizations and the limited existence of resources, concerns about the sustainability of that progress are increasing. There is a rising need for affordable, reliable, and sustainable energy sources. Different international organisms recognize the crucial role of energy in achieving global sustainability, highlighting the need for a transition to modern and green energy alternatives. This transition to renewable energy introduces complexities in control systems, requiring reliable load-frequency controllers. This document discusses the limitations of traditional PI control strategies and studies the use of a reinforcement learning-based algorithm, Proximal Policy Optimization (PPO), in power system control. Results demonstrate the efficacy of PPO in effectively reducing area control error, handling nonlinearities, and adapting to disturbances in the system. The adaptability of PPO to different situations and system changes, with its system-agnostic nature, makes it a promising candidate for improving power system control. Finally, practical considerations such as computational requirements, communication delays, measurement noise, and hardware constraints are acknowledged as challenges that need further exploration for the real-world deployment of PPO in power grid environments. The paper concludes by calling for additional research to address these challenges and optimize the algorithm for real-time applications, emphasizing its advantages in the transition to renewable energies.

Keywords: load frequency control, reinforcement learning, proximal policy optimization, renewable energies

List of Figures

List of Tables

Content

1 Introduction

1.1 Introduction

Nowadays civilizations have been characterized as the most advanced in human history. With technological progress, most of them left behind famine and massive deaths, caused by infectious diseases and plagues, and now live in peace and prosperity, which allowed world life expectancy to rise. As a consequence, while these problems and needs persist, population concerns about the future are increasing. In their 2030 Agenda for Sustainable Development, the United Nations have gathered those concerns and needs, and with them, they created a blueprint as a road map to advance as a civilization sustainably while maintaining planet ecosystems[1; 2]. The core of this blueprint is the Sustainable Development Goals. In these, the need for affordable, reliable, sustainable, and modern energy is recognized to be essential for sustainable development.

Present societies depend on fossil energy, but its finitude and gradual depletion accentuate the need to look for a replacement. Coal, natural gas, and crude oil reserves are running out at a worrying rate. According to the latest BP Statistical Review of World Energy, see Figure **1-1**, total global reserves of coal, natural gas, and crude oil would be exhausted by years 2159, 2068, and 2066, respectively [3]. Note that this depletion is not even across regions. If this dependency continues, extra and major political and economic disputes will arise as a consequence of it. Additionally, the combustion of fossil fuels is the main contributor to greenhouse gas emissions in the world. Due to its catastrophic effects, global warming has become an important motive for the search for new sustainable and greener energy sources. A call to action to combat climate change and reduce its consequences has also been made.

During the last decade, extreme events have become more frequent and their consequences are increasingly higher. Heatwaves are rising in frequency, duration, and maximum temperature, which leads to dangerous droughts in some world regions [4]. On the contrary, but following the same path, heavy downpours are also increasing in frequency and becoming heavier, which even raises problems in some cities, as they are not prepared to take that amount of water, leading to floods, possibly deaths, and poorer living conditions. Hurricanes, blizzards, and thunderstorms are other examples of extreme weather events whose frequency has risen and pose a high danger to human societies. It has already impacted directly the way of life of some societies, as more thorough strategic planning has to be done to adapt to the new living conditions and risks [5].

Electric energy promises to serve as an effective means to combat climate change and to have the necessary affordability, reliability, and sustainability characteristics. Electricity can be applied interchangeably and effectively with other types of energy, e.g. fossil energy. Recent technological advances, like solar panels, converters, and wind turbines, make it affordable and versatile enough for most populations. It also can be generated in more sustainable and clean ways. Those characteristics have made electricity a prospect of a reliable energy source for the everyday needs of customers: communication, mobility, cooking, lighting, heating, and cooling[6]. As a result, electricity has acquired a major role in the lives of people, with 90% of the global population having access to it[6]. However, the concept of electricity cannot be separated from

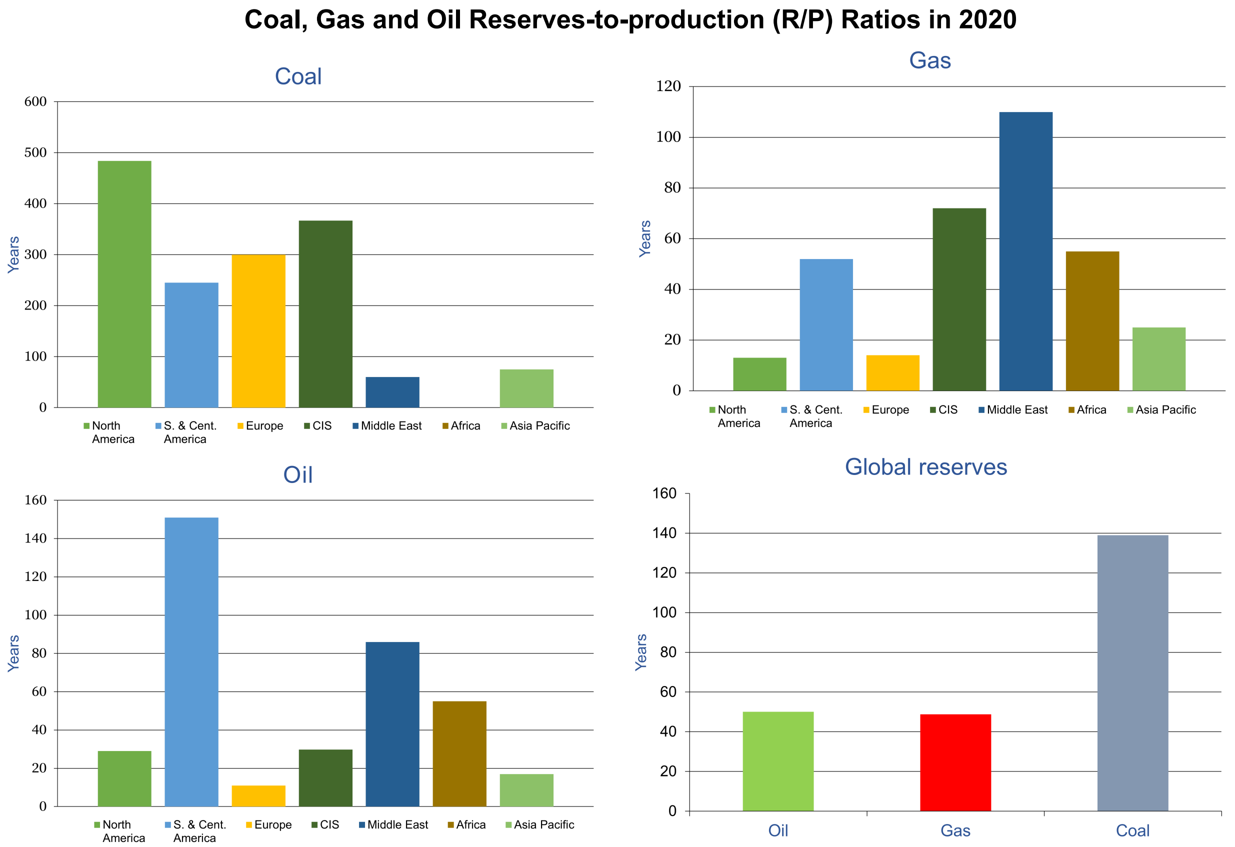


Figure 1-1: Coal, gas and oil reserve-to-production (R/P) ratios in 2020. [3].

the concept of electrical power systems, as it needs to be used.

1.2 Rationale

An electric power system refers to the interconnection of elements that are responsible for the generation, transmission, and distribution of electric energy. Most electric power systems over the world are based on alternating current (AC) and work at a unique frequency, e.g. 60 Hz in most countries of North and South America and 50 Hz in Europe. The bulk of any modern power system is composed of three fundamental machines, which exploit the well-known Faraday's Law

$$e(t) = \frac{d}{dt}\varphi(t) \tag{1-1}$$

Where e is the electromotive force (EMF) and φ is the magnetic flux.

Those are (1) Synchronous machines, where flux variation is due to a rotating flux in the rotor by a direct current flowing through its windings and its spatial rotation, (2) Transformers, where the alternating currents that flow through its windings create periodical variations in fluxes and (3) Induction Machines, where the flux of the rotor varies in space and time through its air gap due to its rotation [7].

Historically, AC systems have been preferred over DC-based without discussion[7]. Nonetheless, DC transmission and generation have been gaining interest recently. Transmission in DC has fewer losses, needs fewer conductors and less insulation, there are no skin and corona effects, and is less difficult to synchronize and stabilize a DC system [8]. The principal drawbacks of DC systems were that transmission and generation were comparatively not feasible. It was easier and cheaper for AC systems to generate electricity from mechanical energy and to obtain the high voltages needed to transport electricity long distances. With the development of semiconductors, and with them, power converters, DC systems regained attention. Renewable energy sources, such as photovoltaic generation and wind turbines, work on DC or are asynchronous. e.g. do not have a unique frequency, thus, they rely on power converters to connect themselves to the AC grid. Power converters also made High Voltage DC transmission (HVDC) systems feasible, which are preferred over AC for very long distances. Power systems will suffer a hybridization between AC and DC soon. It is expected that DC systems overtake AC systems at some point. However, the replacement cost is extremely high, which is why current AC systems still have a long life ahead of them.

This hybridization responds to the present civilization's needs and values but imposes new challenges to the operation and control of power systems. Due to the possibilities of new technological advances and the philosophy of more efficient systems, populations are heading to automation and digitization of several activities [9]. Therefore, high power quality became a new requirement. Worldwide power systems are also being operated under stressed conditions, near their stability limits, as electricity consumption has been increasing [10]. Climate change drives generation matrices to add renewable generation. Renewable resources participation in electricity generation is already more than 10% and is growing exponentially, as can be seen in Figure **1-2**. They also are allowed to generate in a distributed manner, which promotes the use of distributed energy generation. Those are usually called distributed energy resources (DER). Therefore, the intrinsic characteristics of renewable resources, such as solar and wind, as well as the quantity, location, and diversity of new technologies introduce uncertainty to the power system. These coexist with the traditional ones in an interconnected form, thus, adding even more complexity to the power system.

Evidently, the current condition of power systems is presenting new challenges. As a summary, the following factors hinder the operation and control of power systems [9]: (1) there is a large number of tightly interconnected components, (2) the relationships among various power components are so complex that mathematical theories and control methods are hard, or almost impossible, to apply; (3) components in different layers require several spatial and temporal requirements for operation. For example, power converters control works on different time scales than traditional machine controllers, hundreds of times higher; (4) simultaneous participation of different new actors, like electric vehicles or distributed generation resources, in several different places (5) random disturbances, mainly introduced by stochasticity of the load and wind and solar sources, or disturbances already known by utilities; (6) power system operation is based in multiple hierarchical layers; (7) as more and more actors participate in power systems generation, it becomes impossible for a single centralized entity to evaluate, monitor, and control all the interactions in real-time. Nonetheless, power system control is also evolving. High processing power is given by computers and processors, that are capable of handling millions of operations, and new and faster communication means, e.g. optic fiber, wireless, and satellite communications set the foundations for new control and operation paradigms, opening up new possibilities [9].

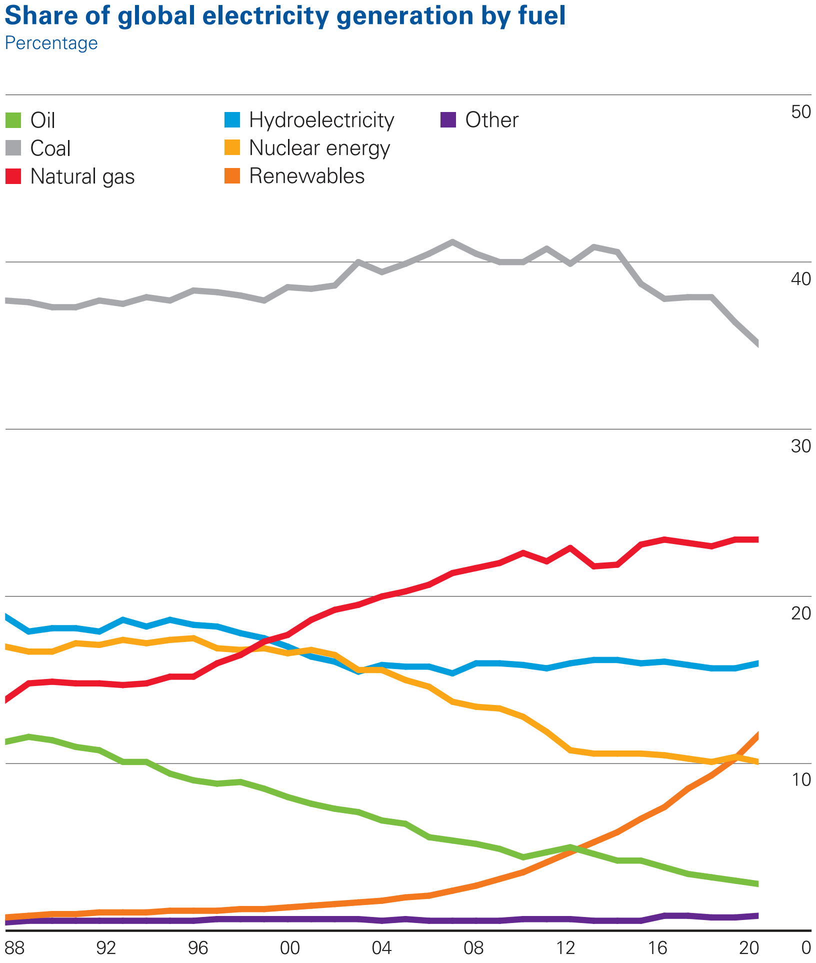


Figure **1-2**: Share of global electricity generation by fuel. [3]

1.3 Background

1.3.1 Overview of Stability and Frequency Control on Power Systems

Power Systems Stability

For AC power systems to work, they ought to have the ability to regain and maintain steady-state conditions after being physically disturbed [11]. Synchronous machines have damping capacity and large inertia, but modern transmission systems are adding more converter-based generation to increase capacity or replace older machines, which results in a system with less inertia, narrowing the operation limits of the system [12]. The loss of stability could lead the system to a blackout, which for such large and essential systems, will have a great societal and economic impact.

Stability requirements on power systems are usually stated in terms of voltage, rotor angle, and frequency stability [13]. Voltage stability refers to the ability to maintain steady voltages on all buses of the power system. Rotor angle stability refers to the ability to keep the state of synchronization between synchronous machines in the system. Its most characteristic event is the oscillatory swing of the rotor angle, where a swing with increasing amplitude indicates the loss of synchronism. Lastly, frequency stability refers to the ability of the power systems to maintain their frequency in every bus of the system.

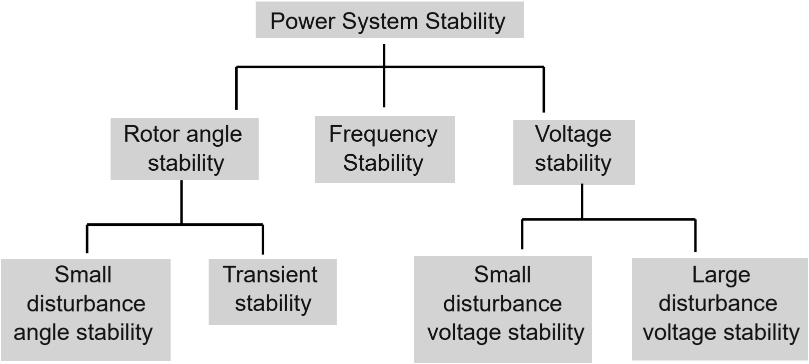

Power System Stability
Rotor angle stability
Frequency Stability
Voltage stability
Small disturbance angle stability
Transient stability
Small disturbance voltage stability
Large disturbance voltage stability

Figure **1-3**: Power systems stability classifications.
[14]

As seen in Figure Figure **1-3**, voltage and angle stability are usually divided according to the size of the disturbance, into small-signal, for small disturbances, and transient stability, for large disturbances [14]. This differentiation allows linearization for small disturbances. Power systems are inherently nonlinear systems, but they can be linearized around a steady-state operating point. Linearization is based on the following assumptions: (1) a system can be linearized around a certain operating point and (2) the system keeps an approximated linear behavior in a region not far from this point. It simplifies the analysis, as a large set of tools from linear systems theory can be applied. However, the limits to where the linearization can be used are not clear and it varies from system to system. For relatively large disturbances, the nonlinearities must be considered in the mathematical model [14].

Frequency on Power Systems

Periodic and oscillatory phenomena are characterized based on frequency by several engineering and science fields. Usually, frequency is expressed in Hertz [Hz] in the International System of Units (SI), as the number of cycles per second [15]. To maintain stability, AC power systems need to be operated at a nominal reference frequency. It has been standardized in 60 Hz, for America, and 50 Hz for Continental Europe and the United Kingdom (UK). Ideally, the frequency of the system should be constant and equal to the reference value in the whole power system. However, the frequency in power systems is never constant [16; 17; 18]. Moreover, as variations of the system state with any frequency different than the nominal one, even small ones, will decrease the power quality and increase losses on the system, they have to be eliminated or reduced, at least [7]. Variations may occur by changes in the demand or generation, due to stochastic noise, network operation, e.g. connection or disconnection of lines or buses, nonlinearities, e.g. saturation on transformers, harmonics, and occasionally faults. Sources of variations are almost impossible to predict or control. To assess that regulators have defined ranges where frequency variations are allowed. In North America, the standard frequency range accepted is ±50mHz, in continental Europe and the UK is ±50mHz and ±200mHz, whereas the maximum steady-state deviation allowed is ±200mHz, ±250mHz, and ±500mHz, for US, UK, and Europe respectively [16; 17; 18].

Notice that the definition of frequency in the SI is not suitable for transient phenomena, where there is no "repetition" of events. There is a discussion held about this in [7], where some of the most important aspects are summarized.

Therefore, in the electrical engineering field, it is pragmatic to define an *instantaneous angular frequency* in [rad/s], as the *rate of change of the phase* $v[rad/s]$, of a *sinusoidal wave* [19],

$$u(t) = sin(v(t)) \tag{1-2}$$

where w is related to v and f as,

$$w(t) = 2\pi f(t) = \frac{d}{dt}v(t) \tag{1-3}$$

and, correspondingly, the period is

$$T(t) = \frac{2\pi}{w(t)} = \frac{1}{f(t)} \tag{1-4}$$

Also, the *rate of change of frequency* (RoCoF) [Hz/s] is stated as:

$$RoCoF = \frac{\alpha(t)}{2\pi} = \frac{1}{2\pi}\frac{d}{dt}w(t) = \frac{1}{2\pi}\frac{d^2}{dt^2}v(t) \tag{1-5}$$

A nominal reference frequency is defined for standardization and regulation purposes. The reason to keep that frequency constant everywhere is due to the nature of synchronous machines.

The electromechanical behavior of the synchronous machine is given by the *swing equation*

$$M\frac{d}{dt}w(t) = \tau_m(t) - \tau_e(t) - D(w_G(t) - 1) \tag{1-6}$$

Where $\tau_m(t)$ is the input mechanical torque; $\tau_e(t)$ is the electromagnetic torque linked to the stator flux and currents; $w_G(t)$ is the per unity rate of change of the angle position $\delta_G(t)$ referred to the reference angle δ_0 of the system; D and M are the damping coefficient and inertia constants, respectively.

According to 1-6, the electromechanical torque is directly related to the real power of the synchronous machine. Note that a change in the electrical power will impose a change in the instantaneous frequency at the output of the machine. Thus, restricting the steady-state of power systems to the case where the angular frequency is also constant.

Power Systems Control

Frequency Control

To maintain a constant frequency, power systems require frequency regulation. Despite synchronous machines tending to keep synchronism and maintain the power system balance by varying their kinetic energy, they are not able to automatically recover a secure steady-state operation point by themselves after a contingency. Notice that the electro-mechanical equation (1-6) resembles that of a pendulum with damping and disturbances terms. Moreover, the oscillatory, *swing*, response of synchronous machines degrade the system's performance and stability. The system will destabilize if the power unbalances created after a large disturbance are large enough to surpass systems inertia [11]. In power systems with DER based on power converters, the total inertia of the system is reduced. As modern power systems are put forward more frequently to work under severe stress conditions, combined with less inertia due to the nature and popularity of renewable resources, frequency control gained more interest [10].

The objective of frequency regulation is to match generation and consumption while reducing losses on the system. To do that, the generators must have the capability to work under nominal or maximum available power at the moment, this saved power is called *spinning reserve*. Counting on that spinning reserve, four types of frequency regulation are defined: primary frequency control (PFC), secondary frequency control (SFC), and tertiary frequency control (TFC). They operate on different scales and in a hierarchical way. A frequency characteristic response and an overview of the control framework for frequency control, with the action time of each control type, are shown in figure **1-4**. On top of them, one more type is added, which is called emergency control (EC), which operates in extreme disturbance cases.

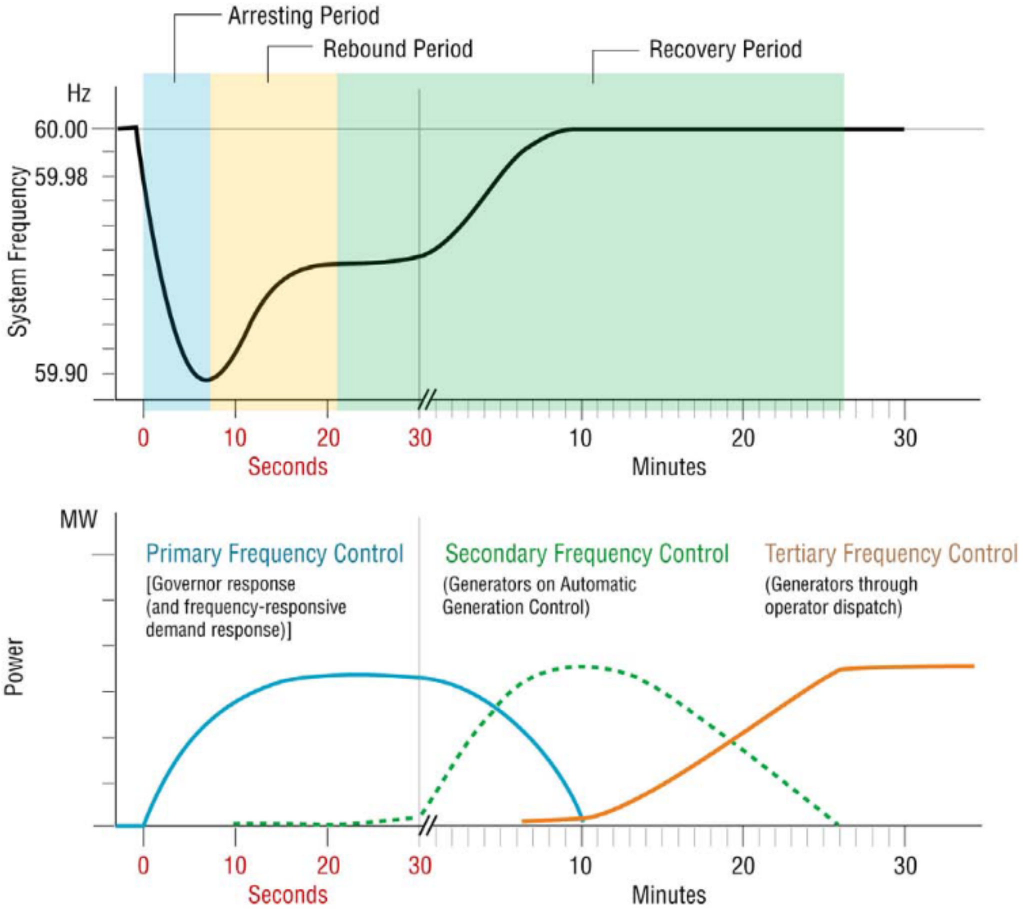


Figure **1-4**: Characteristic response of primary, secondary, and tertiary frequency controls following the sudden loss of generation.

[20]

Turbine governors (TG) compose the PFC of synchronous machines. In steady-state operation, small frequency deviations are attenuated by the PFC. The frequency deviation Δf is the input. A dead band is used to reduce the stress on the mechanical parts of the TG by chattering phenomena. The resulting signal goes through a gain filter, $1/R$, called *drop*. Its output is added to the power reference given by the SFC and the TFC.

Disregarding how complex TG models could be, PFC is deliberately based on *drop control*. Under the assumption that drop per unit gain values are similar relative to their machine sizes, it enables the unbalance of power to be distributed among all power plants according to the capacity of their relative machines [7]. This optimizes the usage of the spinning reserve. This result would not be possible for perfect reference tracking controls. Even if they are used in practice, the dead band in the TG always adds a steady-state error, thus, another control is needed to correct it.

SFC, which is often called automatic generation control (AGC), has the major function of load frequency control (LFC). LFC is responsible for eliminating this steady-state error after a disturbance. It is a regional controller, that selects a pilot bus for a frequency reference, calculates the *area control error* (ACE), and sends control signals to the turbine governor of each synchronous machine. Its working time scale is an order higher than PFC. Its objectives are to maintain the frequency and power interchanges with neighboring control areas, also known as tie-line flows, at scheduled values. LFC is recognized as one of the most important control problems in electric power systems design and operation and has been extensively researched.

TFC is done by the system operator. It is needed to restore the spinning reserve of the system after the disturbance has been cleared out. To do that, the power references of the generators are adjusted, which takes the system to a more stable steady state. The economic aspects associated with the operation of the system are also considered for optimization, e.g. costs of generation and transmission losses.

Emergency control and protection schemes have a discontinuous nature and usually act over the system structure. When the disturbance is too large, if the system maintains the same structure, it could lose stability, then, its structure needs to be modified. Switching is used for this purpose. EC includes manual control, e.g. Emergency Load Shedding (ELS) or generation shedding, or automatic control, e.g. Under Frequency Load Shedding (UFLS) or Under Voltage Load Shedding (UVLS).

Angle and Voltage Control

Power oscillations, or swings, are directly related to the voltages and angles of the terminals and rotor of the synchronous machine, respectively. To reduce the risk of losing voltage and angle stability the dynamic response of the power system needs to be smooth, thus voltage and angle regulation is needed. Analogous to the PFC, the automatic voltage regulator (AVR) is the primary voltage regulator of generators. It is responsible for maintaining the voltage at the generator's connection point and is mandatory for its connection to the transmission system. Conventional power units allow this modification through the excitation circuit of the machine.

Supplementary control is given by the power system stabilizer (PSS), which is considered the most relevant of such controllers. PSS's aim is the damping of electromechanical oscillations in the machine, due to the conflict between rotor speed and voltage dynamics, and also, between AVR and oscillatory electromechanical modes of the machine. Modern power systems also count on flexible AC transmission systems (FACTS) devices to improve voltage stability on the power system. Similar to frequency control, secondary and tertiary voltage controls are used to improve system stability and optimize the economic aspects of the transmission.

1.3.2 Reinforcement Learning for Control

Reinforcement learning (RL) is a major field of machine learning whose concern is how an agent learns to take actions from experience, called control laws or policies, to interact with an environment while maximizing long-term reward signals. An overview of the reinforcement learning framework is shown in figure **1-5**.

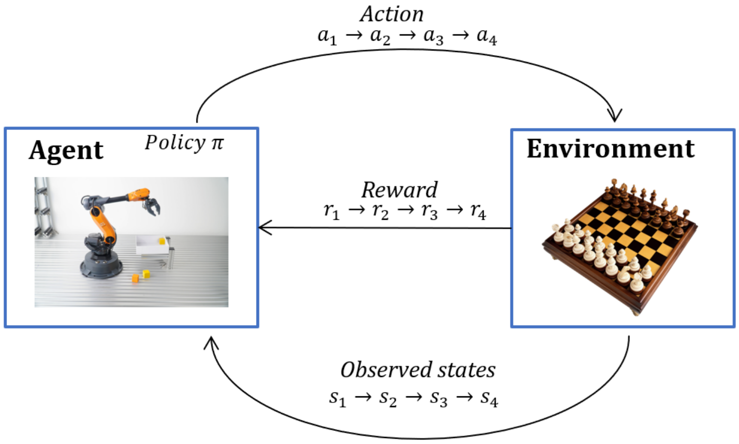

Action
$a_1 \rightarrow a_2 \rightarrow a_3 \rightarrow a_4$
Agent
Policy π
Environment
Reward
$r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_4$
Observed states
$s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4$

Figure **1-5**: Reinforcement Learning Framework.
[21]

In the RL framework is common to assume that the interaction between agent and environment is described by a Markov Decision Process (MDP). An MDP generalizes the idea of a Markov process by adding actions and rewards [21]. It is worth recalling that a Markov process is based on the assumption that the next state only depends on the current state. Thus, An MDP can be represented by a set of states S, a set of actions A, and a set of rewards R, along with a transition model $T(s, a, s')$, which dictates the probability

$$P(s', s, a) = P(s_{k+1} = s' | s_k = s, a_k = a) \tag{1-7}$$

of transitioning to the next state s' given the current state s and a chosen action a, and a reward function $R(s', s, a)$.

Common notation uses π to represent a policy, which is sometimes assumed as a mapping from states to actions $\pi : a \rightarrow s$, or as an auxiliary notation to be used in a sum of states like [21],

$$P(s', s, \pi) = \sum_{a \in A} \pi(s, a) P(s', s, a) \tag{1-8}$$

For the present document, the first representation is one to be used.

The goal of reinforcement learning is to find a policy that maximizes the expected discount cumulative reward, called optimal policy which is noted as π^*. Then, given a policy, a value function that quantifies the benefit of being in a given state is defined as,

$$V_\pi(s) = \mathbb{E}\left(\sum_k \gamma^k r_k | s_0 = s\right) \tag{1-9}$$

where $\mathbb{E}$ is the expected reward over the time steps k, subject to a *discount rate* γ. Note that future rewards are discounted, giving more value to the early rewards. Then, the best policy is given by,

$$V(s) = \max_{\pi} \mathbb{E} \left(\sum_{k} \gamma^k r_k | s_0 = s \right) \tag{1-10}$$

it can be written recursively as,

$$V(s) = \max_{\pi} \mathbb{E} \left(r_0 + \sum_{k=1}^{\infty} \gamma^k r_k | s_1 = s' \right) \tag{1-11}$$

which means that

$$V(s) = \max_{\pi} \mathbb{E}(r_0 + V(s')) \tag{1-12}$$

Thus, the optimal policy is given by,

$$\pi^* = \arg\max_{\pi} \mathbb{E}(r_0 + V(s')) \tag{1-13}$$

Sometimes, it is necessary to quantify the benefit taking into account also the chosen action. The value function $V(s)$ does not actually cares about the taken action, that is why another value function is defined as,

$$Q_{\pi}(s,a) = \mathbb{E} \left(\sum_{k} \gamma^k r_k | s_0 = s, a_0 = a \right) \tag{1-14}$$

This is called the quality or Q-value function and is the expected reward from starting at state s, and then acting with action a. It can be rewritten as,

$$\begin{aligned} Q(s,a) &= \max_{\pi} \mathbb{E} \left(R(s,s',a) + \gamma V(s') \right) && (1\text{-}15) \\ &= \sum_{s'} P(s',s,a) \left(R(s,s',a) + \gamma V(s') \right) && (1\text{-}16) \end{aligned}$$

The optimal policy is then given by,

$$\pi^* = \arg\max_{a} Q(s,a) \tag{1-17}$$

and is related to $V(s)$ as

$$V(s) = \max_{a} Q(s,a) \tag{1-18}$$

Expressions (1-18) and (1-15) are known as the *Bellmans' equation*, which are a necessary optimality condition, based on Bellman's principle of optimality.

Finally, it is worth noting that this framework is defined only for finite state and action spaces. However, it can be extended to infinity spaces with the help of neural networks.

1.4 Problem Identification

In previous sections, it was presented that current societies depend on fossil energy, but it is finite and is depleting gradually. Also, the combustion of fossil fuels is the main contributor to greenhouse gas emissions in the world. Electric energy promises to serve as an effective means to combat climate change and to have the necessary affordability, reliability, and sustainability characteristics. It has also acquired a major role in the lives of people. Electricity depends on power systems to be generated, transported, and used. Most power systems work on AC. For AC power systems to work, they ought to have the ability to regain and maintain steady-state conditions after being physically disturbed. However, renewable resources, such as solar and wind, as well as the quantity, location, and diversity of new technologies introduce uncertainty to the power system. Also, power systems are suffering a hybridization between AC and DC, adding more complexity to the analysis. On top of that, the digitization of several people's activities and lives also brought new requirements for high power quality.

As a consequence, global warming combat efforts and fossil fuel replacement are being delayed because of the problems related to the adoption of renewable and distributed generation. Due to global warming, extreme events are becoming more frequent. If the dependency on fossil fuels continues, extra and major political and economic disputes will arise as a consequence. Moreover, hybrid power systems contain less inertia, making them more vulnerable to loss of stability after a disturbance. The loss of stability could lead the system to a blackout, which for such large and essential systems, would have a great societal and economic impact.

These add complexity to the frequency control of power systems while hardening the operation requirements. Also, nonlinearities of the systems are not usually considered in research. For systems with high penetration of distributed renewable resources, meaning less inertia to withstand a disturbance, and with new power quality requirements, these nonlinearities need to be accounted for. Similarly, solar photovoltaic and wind generation are usually not considered for frequency regulation, when, due to their high penetration in power systems, they will be asked to do frequency regulation. Then, the control and operation of power systems are becoming harder. Load frequency control is one of these control problems. Most all the existing control theories have been applied to this problem, but most of them do not consider the mentioned nonlinearities and participation of renewable resources, solar and wind, and storage systems, in frequency regulation to reduce the governor's efforts of synchronous machines. Also, interactions with other controllers are not considered. Simultaneously, great processing power, given by computers and processors, set the foundations for new control and operation paradigms. Reinforcement learning is one of these paradigms. It has proved good performance addressing the LFC problem, but work on these is recognized to be still in the early stages. Therefore, the purpose of this research is to contribute to the LFC problem, addressing it from the reinforcement learning framework, leading to the following research question:

What is the performance of a LFC based on RL for a multi-area power system with high penetration of solar photovoltaic, wind generation, and energy storage systems, with the participation of these in frequency regulation, considering nonlinearities and voltage and stabilizers interaction?

1.5 General and specific objectives

General Objective

Propose a load frequency control for a multi-area power system with high penetration of solar photovoltaic, wind generation, and energy storage systems based on reinforcement learning.

Specific Objectives

1. Develop the multi-area power system environment for training and testing.
2. Design and train a controller based on reinforcement learning theory.
3. Benchmark the performance of the controller against other control strategies as optimized PI and ADRC controllers.

1.6 Thesis Organization

This document is organized as follows. Chapter 2 provides a review of Load Frequency Control (LFC) in power systems. Then it continues with an overview of neural networks as function approximators, covering recurrent neural networks, and finalizes with a review of the Proximal Policy Optimization (PPO) algorithm. Chapter 3 dives into the approach and methods employed in this study to investigate LFC in power systems. It describes the models used and the several simulation assumptions. It also describes the learning process and presents the PPO algorithm used. The last section details the Proportional Integral (PI) controller design. Then Chapter 4 presents the results of the simulation, demonstrating the generalization capabilities of the reinforcement learning algorithm. Chapter 5 includes a discussion about those results, placing emphasis on the advantages and disadvantages, and problems found. Chapter 6 closes this document with a conclusion on the work done and, finally, Chapter 7 presents a discussion on how to extend this work and other approaches or evaluations that can be performed.

2 Literature Review

This chapter provides a review of Load Frequency Control (LFC) in power systems. It traces the historical evolution of LFC, from early tie-line bias control to addressing contemporary challenges like renewable energy source integration and market deregulation. The goals of LFC, such as maintaining constant frequency and minimizing deviations, are highlighted. The chapter also emphasizes the impact of renewable energy fluctuations on LFC and their inclusion in control scheme models.

Additionally, it provides an overview of neural networks as function approximators, covering feedforward neural networks, recurrent neural networks (RNNs), and Long Short-Term Memory (LSTM) architectures. It also introduces the Adam optimizer and explores policy gradient algorithms, focusing on multi-step returns and off-policy learning. Finally, the Proximal Policy Optimization (PPO) algorithm is highlighted as a solution to noisy gradient challenges in policy gradient methods. The chapter concludes by discussing performance functions for control problems.

2.1 Load Frequency Control

The LFC problem has been one of the most researched in power systems control. Tie-line power flow and frequency control in power systems have appeared even before frequency stability was characterized in [19]. One of the first records was the work done by Cohn *et. al.* on proposing an LFC based on tie-line bias control. He also proposed control techniques based on adding coordinated inadvertent interchange and time error correction factors on area control error computations [22; 23; 24]. Since then, LFC has been continually evolving. A standardized terminology related to LFC was finally stated in [25], dynamic modeling for LFC studies began being discussed in depth in [26], and novel concepts such as optimality were added to the control philosophies. Power system restructuration also brought new challenges for LFC. Since power systems are turning into a hybrid scheme, due to distributed and renewable resources, the effects of this effect have been studied in [12], and models of the new technologies had to be added. A huge variety of these models and technologies can be found in [27]. As market deregulation also changed the rules for LFC, new studies on deregulated markets for modern power systems have been held and LFC control on deregulated markets has been extensively studied. Most of the work on deregulated power systems has been recently reviewed in [28].

In instances where the generated power falls short of the load demanded, the speed and frequency of generator units decrease, and conversely, if the generated power exceeds the load, the speed and frequency increase. Typically, normal frequency variations range around 5% between light load and full load conditions [29; 10].

In the context of frequency regulation, there are two primary strategies: flat frequency regulation and parallel frequency regulation. The former occurs when only the generator closest to the altered load adjusts its generation set point. On the other hand, parallel frequency regulation involves multiple generators adjusting their set points simultaneously [29].

To maintain stability, the change in a specific area is managed by the generators within that region, ensuring that tie-line loading remains constant. All generators within such an area form a coherent group, synchronizing their speed and power angles to collectively speed up or slow down. This defined area is termed a control area, typically aligning with the boundaries of an individual electricity board company [30; 10]. The objectives of load frequency control are [10]:

- Hold a constant frequency and reduce deviations from its set-point.
- Maintain the tie-lines power flow to their scheduled value.

The power transference through the tie-line is defined by the expression [13]:

$$P_{tie,12} = \frac{|V_1||V_2|}{X_{12}} \sin(\delta_1 - \delta_2) \tag{2-1}$$

where V_1 and V_2, and δ_1 and δ_2 represent the voltages and their respective angles of both ends and X_{12} the reactance of the tie-line. Small changes in frequency can be linearly approximated by the following expression [30; 10].

$$\Delta P_{tie,12} = \frac{|V_1||V_2|}{X_{12}} cos(\delta_1 - \delta_1)(\Delta\delta_1 - \Delta\delta_1) \tag{2-2}$$

Grouping terms

$$K_{s12} = \frac{|V_1||V_2|}{X_{12}} cos(\delta_1 - \delta_1) \tag{2-3}$$

where K_{s12} is referred as the synchronizing coefficient between the two areas [10]. Then, a change in transferred power can be expressed as

$$\Delta P_{tie,12} = K_{s12}(\Delta\delta_1 - \Delta\delta_2) \tag{2-4}$$

Considering that

$$\Delta\delta = \int_0^T \Delta\omega dt \tag{2-5}$$

Replacing the angle variation $\Delta\delta$ in the power variation equation,

$$\Delta P_{tie,12} = K_{s12} \left(\int_0^T \Delta\omega_1 dt - \int_0^T \Delta\omega_2 dt \right) \tag{2-6}$$

Applying the Laplace transform and considering any pair of areas i and j,

$$\Delta P_{tie,ij} = \frac{K_{s12}}{s}\left(\Delta\omega_i(s) - \Delta\omega_j(s)\right) \tag{2-7}$$

In general, the tie line power variation for any area is defined by [10],

$$\Delta P_{tie,i} = \sum_{j=1, j\neq i}^{n} \Delta P_{tie,ij} \tag{2-8}$$

Area Control Error (ACE) is considered an essential indicator for the imbalance in generation-load power for the transmission system operator [10]. Defined as,

$$ACE_i = \beta_i 2\pi \Delta\omega_i + P_{tie,i}$$

or

$$ACE_i = \beta_i \Delta f_i + \Delta P_{tie,i} \tag{2-9}$$

where β_i represents the frequency bias and it is given by,

$$\beta_i = D_i + \frac{1}{R_i} \tag{2-10}$$

D_i, and R_i represent frequency bias, damping coefficient, and governor droop, respectively. The D_i determines the LFC system's ability to rapidly reduce frequency deviations in the power system [10]. The higher the D_i the quicker the LFC to respond to frequency variations and stabilize the power system. The damping coefficient D_i of conventional power system can be calculated as [30]:

$$D_i = \frac{4H_i\omega_n\xi}{\omega_m} \tag{2-11}$$

where H_i, ω_n, and ξ represent the inertia constant, natural frequency, and damping factor, respectively. Likewise, the inertia constant, H_i, is important in determining the system's capability to respond to variations in load demand and maintain frequency stability. For a single generator, H_i is determined by

$$H_i = \frac{1}{2}\frac{J_i\omega_m^2}{S_{n,i}} \tag{2-12}$$

where J_i, ω_m, and $S_{n,i}$ represent the moment of inertia, rotor speed, and rated power of the generator, respectively. the moment of inertia, J_i, rotor speed, ω_m, and rated power, $S_{n,i}$, of the generator

In interconnected power systems with multiple generators, the total inertia constant, H_{con}, is computed by considering the contributions of individual generators based on their rated powers

$$H_{con} = \sum_{i=1}^{n} H_i \frac{S_{n,i}}{S_n} \tag{2-13}$$

where n and S_n are the total number of connected generators and the rated power of the power system.

2.1.1 Renewable Energy Sources Impact on Frequency Response

Renewable power plants are an additional source of variation to an already variable system. Analyzing the overall effect of fluctuations caused by Renewable Energy Sources (RES) is crucial, as instantaneous fluctuations in load and RES power output can either amplify, be unrelated, or cancel each other out. The slow dynamics of RES power fluctuations and total average power variation negatively impact power imbalance and frequency deviation. Therefore, these fluctuations should be considered in Load Frequency Control (LFC) schemes, requiring their inclusion in the conventional LFC structure [10].

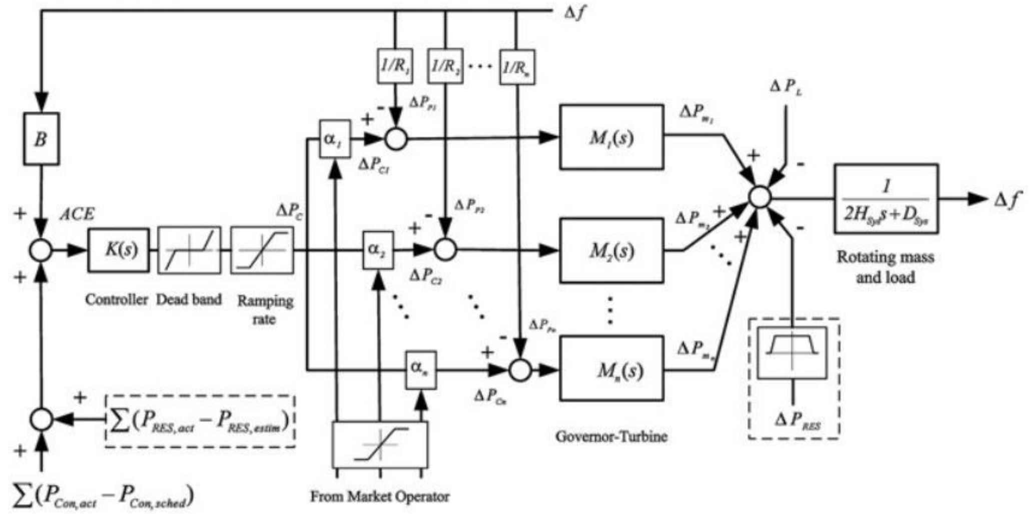


Figure **2-1**: LFC model with considering RES power fluctuation.
[10]

A generalized LFC model in the presence of RES is shown in Fig. **2-1**. Here, to cover the variety of generation types in the control area, different values for turbine governor parameters and the generator regulation parameters are considered [10]. Figure **2-1** shows the block diagram of a typical control area with n generator units. Here, Δf is frequency deviation, ΔPm is mechanical power, ΔP_C is secondary control action, ΔP_L is load disturbance, ΔH_{Sys} is equivalent inertia constant, ΔD_{Sys} is equivalent damping coefficient, B is frequency bias, R_i is drooping characteristic, ΔP_P is primary control, α_i is participation factors, ΔP_{RES} is RES power fluctuation, $K(s)$ is LFC controller, and $M_i(s)$ is governor turbine model [10]. Right after a load disturbance in the control area, the frequency experiences a transient change, and a feedback mechanism generates suitable signals for participating generator units based on their participation factors (α_i) to align generation with the load [10]. Participation factors have the characteristic that their sum in a given area equals one, $\sum \alpha_i = 1$. They are also time-dependent signals that are computed by an independent organization based on several market and power system factors, e.g. availability, bid prices, congestion problems, costs, etc. In a steady state, generation matches the load, driving tie-line power and frequency deviations to zero. The control signal is distributed proportionally among the various generators within the area [10; 29]. In Fig. **2-1**, the frequency performance is represented by a lumped load-generation model using equivalent frequency, inertia, and damping factors [10]. Wind units, with significant kinetic energy, are more important than other Renewable Energy Sources (RESs). The equivalent system inertia is defined by total inertia constants (H_C and H_W) for conventional and wind turbine generators, respectively [10].

$$H_{\text{sys}} = H_C + H_W = \sum_{i=1}^{N_1} H_{C_i} + \sum_{i=1}^{N_2} H_{W_i} \tag{9.2}$$

Wind turbines provide substantial inertia for frequency control support, with typical inertia constants of 2–6 s. For context, typical values for grid power generators are in the range of 2–9 s [10]. The total effect of power fluctuation should consider (ΔP_{RES}). The resulting ACE signal must reflect the total RESs power generation changes [10]

$$ACE = \beta \nabla f + \sum (P_{Con,act} - P_{Con,sched}) + \sum (P_{RES,act} - P_{RES,estim}) \tag{2-14}$$

It can also be shown that frequency deviation in steady state for power systems with RES can be obtained as [10]:

$$\Delta f_{ss} = \frac{R_{sys}(\Delta P_C - \Delta P_{RES} - \Delta P_L)}{D_{sys} R_{sys} + 1} \tag{2-15}$$

where

$$\frac{1}{R_{sys}} = \frac{1}{\sum_i R_i} \tag{2-16}$$

$$D_{sys} = \sum_i D_i \tag{2-17}$$

$$\tag{2-18}$$

Where R_{sys} and D_{sys} are the equivalent system drooping characteristic and damping coefficient respectively [10]. The magnitude of total load-generation imbalance ΔP_D immediately after the occurrence of disturbance can be expressed as [10]:

$$\Delta P_D = 2 H_{sys} \frac{d\Delta f}{dt} \tag{2-19}$$

It shows that the frequency gradient is proportional to the imbalance, with the factor of proportionality being the inertia of the system. The inertia constant is loosely defined by the combined mass of all synchronous rotating generators and motors connected to the system. A higher inertia constant (H) results in a slower decrease in frequency following a load decrease, and conversely, a lower inertia results in faster changes in the frequency.

2.1.2 LFC control types

Conventional Control Approaches

In the early stages, LFC was implemented using PID controllers. It is still the most common in literature and the preferred control strategy used where the bulk power system is composed of conventional generators due to its simplicity [10]. PI-based controllers are mostly known as conventional, or classical, controllers. A basic formulation of a PI-based LFC can be encountered in [7]. Zigler-Nichols Tunning and other methods for a PID controller can be found in [31]. Internal model control (IMC), another conventional control strategy, has been applied for LFC in [32]. The use of IMC for tuning a PID can be found in [33]. The performance of classical controllers for conventional power systems has been assessed in [34], concluding that its performance decreases if non-linearities of the system are taken into account. However, PID controllers are still used as benchmarks for evaluating the performance of the other types of controllers.

The efforts to improve PID performance can be summarized in looking for new topologies, better tunning methods, and adding new features, e.g. adaptability and specific functions. Patel *et. al.* studied a double loop PD+PI controller in [35] for a three-area power system, with thermal and hydro units. Its purpose was to improve the performance of typical PID controllers against non-linearities, even in the presence of an HVDC link. Bakken and Grande in [36] proposed a PI controller with an additional ramp following controller (RFC) to overcome the problem of power deviation during HVDC ramps when modifying the tie-line scheduled power.

Metaheuristic Optimization Strategies

With the concept of optimality introduced in the control of power systems, optimization strategies gained popularity in Load Frequency Control. However, the nonlinearities and complexity of modern power systems involve non-convex and hard-to-solve optimization problems, that deterministic optimization algorithms cannot solve or cannot find the global optima. To overcome this, several metaheuristic optimization algorithms were developed. Some examples are genetic algorithms (GA), Particle Swarm Optimization (PSO), Grey wolf optimization (GWO), Ant Colony Optimization (ACO), Genetic Algorithms (GA), Cuckoo search algorithm (CSA), etc. Kumari has proposed a control using Particle Swarm Optimization (PSO) and Gradient Descent (GD) algorithms for tunning a PI for LFC, the governors drop constant, and the area frequency bias parameters in a multi-area thermal-wind-hydro power plant[37]. The purpose was to compare the LFC performance of PSO and GD on these systems, obtaining better system performance with PSO. Kumar and Antwar have proposed a Fractional Order PID (FOPID) in a parallel control structure [38]. This allowed them to decouple set-point tracking and load disturbance rejection. PSO was used to optimize the integral time absolute error (ITAE) of the frequency deviation for parameter tuning. The performance of this proposal was compared against IMC-based control proposed in [39] showing a performance improvement, however, not exceeding also the IMC-based one in [32] in a test carried out in network topology similar to the standard IEEE 39 bus system (New England 10 machine test system). The difference between the last two resides in that the first one used the direct synthesis approach, which is based on a desired closed-loop transfer function, for design, while the second used the principle of the worst-case plant. Both of them are also robust approaches. Lastly, Otchere *et al.* proposed an adaptative PI control that uses a genetic algorithm (GA) also for tunning and compared it against adaptative PI-PSO controller in a power system with high penetration of renewable energy, such as photovoltaic and wind energy, obtaining a better performance with GA [40].

Robust Control Approaches

Modern power systems make the task of modeling and parameter identification hard and complex. While they are the largest systems that ever existed, the operating conditions increase that level of complexity. The operating point is constantly changing due to inherited demand behavior and rapidly varying sources that renewable resources depend on. Power systems are also constantly subjected to disturbances that can make the system structure change. Moreover, the system parameters vary with time conditions. Adding all those complex dynamics makes any model-based controller, that does not consider them, not feasible. In [41], Bevrani *et al.* suggested a decentralized LFC as a multi-objective optimization problem, considering the regulation against random disturbances, based on a mixed H_2/H_∞ control. The author considers that disturbances are better addressed by H_2 or linear quadratic Gaussian (LQG) control, while H_∞ is considered more useful for holding closed-loop stability and formulation of physical control constraints. The suggested control showed better performance than considering separated strategies. Also, Niimi *et al.* proposed a control strategy with H_∞ to add frequency support to a solar photovoltaic power plant, improving the dynamic frequency response of a small system [42]. Shayeghi *et al.* suggested a novel mixed decentralized H_2/H_∞ control technique trained by a new decentralized radial basis function neural network (RBFNN) based controller, which can account for large modeling uncertainties [43]. This has been compared against other H_2/H_∞ mixed techniques and it has been shown to keep a robust performance, minimizing the effects of area load disturbances in the presence of plant parameter changes and system nonlinearities. Sliding mode controls are other types of robust controllers. A second-order sliding mode control (SMC) has been suggested by Kumar in [44], where the characteristic problem of SMC, chattering, is addressed by using a second-order control law. The LFC problem is solved by stating the system equation in sliding mode and treating the problem as a linear optimal state regulator problem to solve it. Improvement in performance against PI controller and low chattering has been shown. More about robust approaches and control strategies for power systems can be found in Bevrani's book [10].

Advanced Control Approaches

Fuzzy logic solves the control problem based on a linearized mathematical model and experience and knowledge of this system. It has been successfully applied to the LFC problem. You *et al.* have proposed an adaptative PI control based on fuzzy logic. The latter is used as a parameter variation logic for online tunning the PI controller based on a lookup table created with Ziegler-Nichols and critical proportions methods [45]. The method was tested on a diesel generator linearized model suggesting good performance. Singh *et al.* suggested a new fuzzy logic-based approach in PI-controlled LFC two-area linear systems [46]. Here, a fuzzy logic control was used alongside PI to improve the system dynamics, showing less overshoot and less setting time. Non-linearities and HVDC links are addressed in [47] for a two-area hydrothermal system using a fuzzy logic control. A comparison against I, PI, and PID controllers is shown, however, it is not clear if the FLC had better performance. There is more literature in the mix of both, PID and fuzzy logic, which shows better performance, for example in [48].

Artificial neural networks (ANN) are also used for control and are equivalent to fuzzy logic. The knowledge of the system is acquired during training from optimization. Demiroren *et al.* suggested a multilayer ANN controller that uses the area power deviations as inputs to calculate the control signals for minimizing the error signal [49]. It was proposed for a three-area interconnected power system that has two areas including the steam turbine and the other hydro turbine. Chaudhary *et al.* proposed a multilayer ANN controller for a hybrid power system entailing non-reheat, re-heat, and hydropower generating units. It compared it with a fuzzy logic controller, obtaining pretty similar results [50].

A combination of FLC has been used with artificial neural networks, in what is called, an adaptative neuro-fuzzy inference system (ANFIS) controller. This controller was developed based on an adaptative controller framework. Shree and Kamaraj have proposed an ANFIS controller for LFC for a three-area hydrothermal power system using a multilayer perceptron structure for the ANN. The purpose was to improve the system's dynamic performance for large disturbances [51]. It has been compared against other optimization-based and ANN controllers, showing a better performance.

A linear quadratic regulator (LQR) is a state feedback controller. It takes advantage of the linearized system model and state measurements to design an optimal control law based on state feedback and gains. When state measurements are not available or noisy, an observer can be used to observe the states of the system. Prakash *et al.* suggested an application of a linear quadratic regulator-based proportional-integral (LQR-PI) controller for load frequency control for a two-area thermal power system, including solar photovoltaic systems and wind turbines, [52]. In the control, no observers were used. A sensitivity analysis has been done, showing robust performance to generation power variations and tie-line power schedule changes. On the other hand, an LQR based on a state observer is proposed in [53], for addressing the control of a single area power system. Panwar *et al.* has proposed the use of the Jaya optimization algorithm to optimize the definition of the state cost weight matrix (Q) for an LQR control for a two-area hydropower system [54].

Model predictive control (MPC) estimates the future performance of the system based on the present outputs, measured disturbances, and unmeasured disturbances and control signals over a finite horizon, and uses optimization to find the set of states and control inputs that gives the best performance. Ali *et al.* have proposed a decentralized MPC control for a four-area smart grid, including renewables and electric vehicles, and compared it against a centralized version [55]. No frequency regulation was provided from EVs or DER. The results have shown that the decentralized version is better than its centralized counterpart, which is not usual, as it is more common for centralized controls to have better performance. However, this behavior can also be found in this comparative study of several MPC [56]. This different MPC has been tested on a three-area power system considering GRC. Kumtepeli *et al.* also proposed an MPC control for a four-area power system considering nonlinearities but aroused concern about the robustness of the optimization-based controls [57].

Active disturbance rejection (ADR) methods for control are based on techniques to estimate disturbances and then compensate for them in control input. Liu *et al.* proposed a two-layer active disturbance rejection control (ADRC) method with the compensation of estimated equivalent input disturbances (EID) for a two-area power system, which uses ADRC for its first layer and IM principles for the second layer [58]. This scheme has also been studied for wind-diesel microgrids and power systems with photovoltaic generation [59; 60]. Liu *et al.* later proposed a robust version of ADRC by including a generalized state observer (GSO), which is tested for the two-area power system to different power disturbances and compared against PI-based LFC [61]. A comparison between the ADR control strategies is found in [62].

Reinforcement learning (RL), as an artificial intelligence technique, has gained interest recently for LFC. Rekhasree in [63] proposed multi-agent reinforcement learning for a two-area hydro-thermal power system. The purpose was to compare the proposed strategy with a GA control. Early results show that GA control was performing better, however, in a larger system, both had similar behavior. Eftekharnejad has proposed a control multi-agent strategy based on RL for a two-area power system with FACTS [64]. It consists of a layer of local agents and a global agent that coordinates the behavior of the local agents. It was shown to handle different cases with various system parameters, and nonlinearities. More thorough research has been done by Yan and Xu, where they compare multi-agent, single-agent, deep Q networks (DQN) reinforcement learning techniques against PID classical controllers. First, the models were tunned and trained in a linear three-area power system [65]. Then, they considered the impact of nonlinearities and studied the stability and rotation speed of the generators in a fully-modeled New England 39-bus system, with solar photovoltaic generation. The RL techniques showed better performance and stability than the PID control technique, with MA having the best performance.

2.1.3 Wind and Solar Photovoltaic Participation on Frequency Control

A power reserve is needed to give solar and wind generation the possibility to participate in frequency regulation. Unlike conventional generators where the energy source can be controlled, the sun and wind cannot. Then, de-loading techniques, that allow having a power reserve for frequency support and regulation participation of solar and wind generation systems, have been developed. These are mainly based on maximum power point (MPP) tracking with the additional function for selecting a power reference point, to select the desired power under the MPP. Some of these techniques have been explained in [66; 67]. With this power reserve, PV systems and wind turbines can have an inertia response and participate in frequency regulation. One of the most common mechanisms for that is aggregators, also called virtual power plants (VPP). Those aggregate all the generation mainly from converter-based sources, and energy storage systems also, and give them the ability to participate as a larger virtual generator, some of them, emulating the electromechanical synchronous generator behavior. However, they also can be treated separately, and the amount of energy available for frequency support in PV systems or wind turbines varies whether they have energy storage systems or not.

To simulate, models of the technology are needed. Most of the ones used in the literature are simplified linear models only considering the governors and frequency response of the converters. However, more complex and complete models can be encountered in [27; 13].

2.1.4 EMC-UN previous works

Its research seedbed *Inteligencia Computacional Aplicada a Sistemas de Potencia* has been working on modeling, simulation, control, and optimization of power systems. Regarding this thesis topic, a decentralized ADRC has been proposed for a three-area power system [68]. Solar photovoltaic systems and wind turbine participation in frequency regulation were considered. However, there is still room to add uncertainties of the system parameters and interaction with the other control types.

2.1.5 Conclusion on Literature Review

In conclusion, it may be considered that extensive work has been done in the area of LFC. To the author's best knowledge, almost all the existing control theories have been applied to the problem of LFC: linear, robust, and nonlinear controllers, centralized and decentralized controllers, and optimization techniques for control or for improving other proposed controllers. However, despite that some nonlinearities were included, such as generation rate control, governor dead-bands, or boiler dynamics, few articles actually consider them for control design and testing; most of the models were linear representations of the power systems. Others only consider the governor's dynamics in their models. Also, voltage and stabilizing controls are not usually considered, when their effects on LFC are demonstrated. Similarly, solar photovoltaic and wind generation are usually not considered for frequency regulation, when tendencies show that these technologies' participation in power systems is increasing, and in the future, they will be asked to do frequency regulation as well.

Following these future tendencies, reinforcement learning has also proved good performance in addressing the LFC problem, but work on these is recognized to be still in the early stages and it still needs further testing and improvements.

2.2 Function Approximation and Neural Networks

Function approximation is one of the key mechanisms used in reinforcement learning to deal with high-dimensional action and observation spaces. It is based on the intuition that there exists a parameterized function f with parameters θ that can be used to approximate the values of the target function. In this context, they can be used to approximate the main functions used in the reinforcement learning framework, e.g. state value function V, state-action value function Q, or policy π. Among the large set of techniques for function approximation, machine learning is the most popular nowadays. Machine learning is a discipline that comprises a set of algorithms and techniques developed with the purpose of allowing computers to learn by themselves and perform tasks autonomously. It includes deep learning techniques, that employ neural networks to achieve this goal. The advantages over methods are their flexibility, less programming overhead, and the ability to capture the non-linearity of some relations.

This section introduces the neural networks and states the intuition behind them and their learning process. It advances to recurrent neural networks and finishes with its Long Short-Term Memory (LSTM) variation.

2.2.1 Neural Networks

Neural networks are biologically inspired models of computation. Generally, a neural network consists of a set of artificial neurons, commonly referred to as nodes or units, and a set of directed edges between them, which intuitively represent the synapses in a biological neural network. Associated with each neuron j is an activation function $l_j(\cdot)$, sometimes called a link function.

Associated with each edge from node j' to j is a weight $w_{jj'}$. Following the convention in several foundational papers, neurons are indexed with j and j', and $w_{jj'}$ denotes the "to-from" weight corresponding to the directed edge to node j from node J. The value v_j of each neuron j is calculated by applying its activation function to a weighted sum of the values of its input nodes:

$$v_j = l_j\left(\sum_{j'} w_{jj} \cdot v_{j'}\right)$$

For convenience, we term the weighted sum inside the parentheses of the incoming activation and notate it as a_j. We represent this computation in diagrams by depicting neurons as circles and edges as arrows connecting them. When appropriate, we indicate the exact activation function with a symbol, e.g., σ for sigmoid.

Common choices for the activation function include:

- (Sigmoid)

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{2-20}$$

- Tanh

$$\tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{2-21}$$

- Rectified linear unit (ReLU)

$$l_j(z) = \max(0, z) \tag{2-22}$$

Among these, ReLU has been demonstrated to improve the performance of many deep neural networks on various tasks.

The activation function at the output nodes depends on the task. For multiclass classification with K alternative classes, a softmax nonlinearity is applied in an output layer of K nodes [69]. The softmax function calculates

$$\hat{y}_k = \frac{e^{a_k}}{\sum_{k_0=1}^{K} e^{a_{k'}}}, \quad \text{for } k = 1 \text{ to } K \tag{2-23}$$

The denominator is a normalizing term ensuring that the outputs of all nodes sum to one. For multilabel classification, the activation function is simply a point-wise sigmoid, and for regression, we typically have a linear output.

2.2.2 Feedforward Neural Network

A feedforward neural network is a type of artificial neural network where the information flows in one direction—from the input layer to the output layer—without forming cycles or loops. The absence of cycles allows the network to be organized into layers, with each layer comprising nodes (or neurons). In a feedforward network, the input values are set in the lowest layer, and the computation proceeds layer by layer until an output is generated at the topmost layer. This type of neural network is often used for supervised learning tasks such as classification and regression.

Neural networks are trained by adjusting their weights to minimize a specified loss function $\mathcal{L}(\hat{y}, y)$, which penalizes the difference between the predicted output $\hat{y}$ and the target output y. The most successful algorithm for training neural networks is backpropagation [70]. in 1985. The backpropagation algorithm uses the chain rule to calculate the derivative of the loss function with respect to each parameter (weight) in the network. The weights are then adjusted using gradient descent.

The training process involves the following steps:

1. **Forward Pass:** An input example is propagated forward through the network to produce outputs at each layer, ultimately resulting in the predicted output. During this process, each value v_j computed at each node is saved.
2. **Loss Calculation:** The loss function value is computed based on the predicted output and the target output, $\mathcal{L}(\hat{y}, y)$.
3. **Backward Pass (Backpropagation):** The derivatives of the loss function with respect to each parameter are calculated. This is done by propagating the error backward through the network and updating the derivatives for each layer. For each output node, the gradient is given by

$$\delta_k = \frac{\delta \mathcal{L}(\hat{y}_k, y_k)}{\delta \hat{y}_k} \cdot l'_j(a_k) \tag{2-24}$$

Given these values, δ_k, each immediately prior node derivative can be computed as

$$\delta_j = l'_j(a_j) \sum_k \delta_k \cdot w_{kj} \tag{2-25}$$

Which represents the derivative $\partial\mathcal{L}/\partial a_j$ of the total loss function with respect to that node's incoming activation.

Following the chain rule, the gradient of the loss function with respect to the parameters is calculated

$$\frac{\partial\mathcal{L}}{\partial w_{jj'}} = \frac{\partial\mathcal{L}}{\partial a_j}\frac{\partial a_j}{\partial w_{jj'}} = \delta_j v_{j'} \tag{2-26}$$

4. **Weight Update:** The weights are adjusted using the calculated derivatives and a learning rate for the gradient descent algorithm. The learning rate determines the size of the steps taken during the weight updates.

   $$\omega = \omega - \eta\nabla_\omega F_i \tag{2-27}$$

   where $\nabla_\omega F_i$ is the gradient of the objective function calculated on one example.

5. **Stochastic Gradient Descent (SGD):** Nowadays, neural networks are usually trained with stochastic gradient descent using mini-batches. The SGD update equation involves multiplying the learning rate by the gradient of the objective function with respect to the parameters, calculated on a mini-batch of examples.

Various variants of SGD, such as AdaGrad, AdaDelta, RMSprop, and momentum methods, are employed to accelerate learning and adaptively tune the learning rate for each feature. Despite the non-convex nature of the loss surface, heuristic pre-training and optimization techniques, along with these SGD variants, have led to the successful training of neural networks on many supervised learning tasks.

2.2.3 Adam Optimizer

Another popular optimization algorithm is the Adam optimizer [71]. It combines ideas from RMSprop and momentum methods. Adam adapts the learning rates of each parameter individually by considering both the first-order momentum and the second-order acceleration of the gradients. This helps in achieving faster convergence and handling sparse gradients. The adaptive nature of Adam makes it well-suited for a wide range of neural network architectures and tasks.

The updated equations for the Adam optimizer are as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{2-28}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{2-29}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{2-30}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{2-31}$$

$$\theta_{t+1} = \theta_t - \alpha\frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} \tag{2-32}$$

where:

θ_t is the parameter at time t,
g_t is the gradient of the objective function at time t,
m_t, v_t are the first and second moments of the gradients,
$\hat{m}_t, \hat{v}_t$ are bias-corrected estimates of m_t, v_t,
α is the learning rate,
β_1, β_2 are the exponential decay rates for the moment estimates,
ϵ is a small constant to prevent division by zero.

2.2.4 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) extend feedforward neural networks by incorporating edges that span adjacent time steps, introducing a temporal aspect to the model. Unlike feedforward networks, RNNs can have recurrent edges that form cycles, allowing information to persist across different time steps. At each time step, nodes with recurrent edges receive input from the current data point $x^{(t)}$ and hidden node $h^{(t)}$ values from the network's previous state, see Figure **2-2**. The output at each time step is calculated based on the hidden node values.

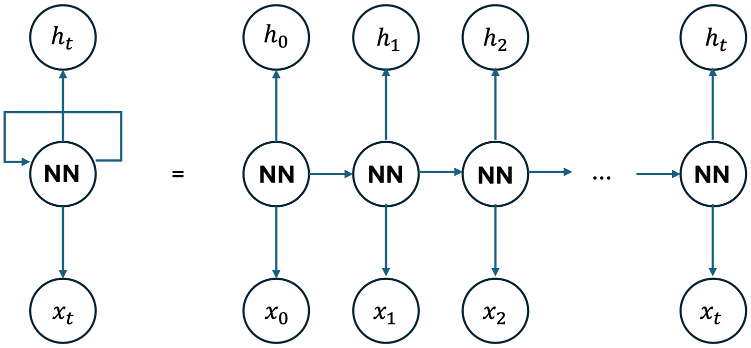


Figure **2-2**: Unfold of a Recurrent Neural Network Over Time.

The computations in a simple recurrent neural network are governed by two equations, involving matrices of conventional weights (between input and hidden layers) and recurrent weights (between the hidden layer and itself at adjacent time steps), as well as bias parameters.

$$h^{(t)} = \sigma(W^{hx}x^{(t)} + W^{hh}h^{(t-1)} + b_h) \tag{2-33}$$

$$\hat{y}^{(t)} = Softmax(W^{yh}h^{(t)} + b_y) \tag{2-34}$$

To understand the network dynamics, it can be unfolded across time steps, visualizing it as a deep network with one layer per time step and shared weights across time. This unfolded network allows for training across multiple time steps using backpropagation through time (BPTT) [72]. BPTT is a fundamental algorithm applied to train recurrent networks, enabling them to learn and capture temporal dependencies.

Learning with recurrent networks has long been considered difficult, especially due to challenges in learning long-range dependencies. Long-range dependencies pose a challenge, leading to issues such as vanishing and exploding gradients when backpropagation is done over many time steps. The vanishing or exploding gradient phenomenon depends on whether the weight of the recurrent edge is greater or less than 1 and on the activation function in the hidden node. The Long Short-Term Memory (LSTM) architecture is introduced as a solution to the vanishing gradient problem. It utilizes carefully designed nodes with recurrent edges with fixed unit weight, preventing the vanishing gradient problem.

2.2.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a recurrent neural network architecture designed by Hochreiter and Schmidhuber in 1997 to mitigate the challenge of vanishing gradients in standard recurrent neural networks (RNNs) [73]. The LSTM model aims to address this issue by replacing ordinary nodes with *memory cells* in the hidden layer, each equipped with a self-connected recurrent edge of fixed weight one. So, gradients can pass across numerous time steps without vanishing or exploding.

The term "long short-term memory" comes from the intuition of the model's unique ability to maintain both long-term memory, encoded in the form of slowly changing weights, and short-term memory, represented by transitory activations passing through the nodes. LSTM introduces an intermediate storage mechanism through the memory cell, which is a composite unit built from simpler nodes in a specific connectivity pattern, incorporating multiplicative nodes [].

Components of an LSTM Cell

1. **Input Node (g_c):** This unit receives activation from the input layer at the current time step ($x(t)$) and from the hidden layer at the previous time step ($h(t-1)$). The summed weighted input typically undergoes a tanh activation function.

2. **Input Gate (i_c):** A distinctive feature of LSTMs is the presence of gates. The input gate is a sigmoidal unit that takes activation from the current data point ($x(t)$) and the hidden layer at the previous time step. The value of the input gate (i_c) multiplies the value of the input node, acting as a gate to control information flow.

3. **Internal State (s_c):** At the heart of each memory cell lies a node (s_c) with linear activation, referred to as the "internal state." This internal state has a self-connected recurrent edge with a fixed unit weight, often called the constant error carousel. This structure allows error to flow across time steps without vanishing or exploding.

4. **Forget Gate (f_c):** Forget gates, introduced by Gers et al. in 2000, provide a mechanism for the network to learn to erase or reset the contents of the internal state. This feature is particularly beneficial in scenarios involving continuously running networks.

5. **Output Gate (o_c):** The final value produced by a memory cell is the result of multiplying the internal state (s_c) by the output gate (o_c). Typically, the internal state undergoes a tanh activation function before being passed through the output gate.

Forget gates have become a standard component in modern implementations, offering improved learning capabilities. Peephole connections allow direct information transfer from the internal state to the input and output gates, proving beneficial in tasks that demand precise timing.

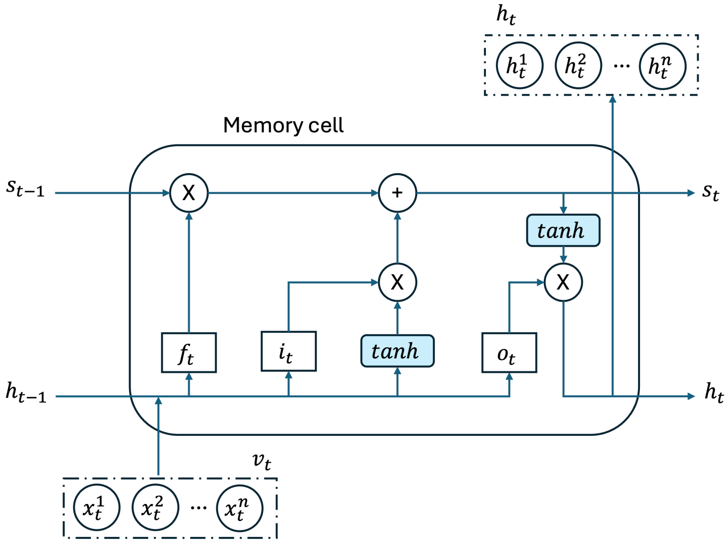


Figure **2-3**: Long Short Term Memory Cell Diagram.

$$g^{(t)} = \tanh\left(W^{gx} \cdot x(t) + W^{gh} \cdot h^{(t-1)} + b_g\right) \tag{2-35}$$

$$i^{(t)} = \sigma\left(W^{ix} \cdot x^{(t)} + W^{ih} \cdot h^{(t-1)} + b_i\right) \tag{2-36}$$

$$f^{(t)} = \sigma\left(W^{fx} \cdot x^{(t)} + W^{fh} \cdot h^{(t-1)} + b_f\right) \tag{2-37}$$

$$o^{(t)} = \sigma\left(W^{ox} \cdot x^{(t)} + W^{oh} \cdot h^{(t-1)} + b_o\right) \tag{2-38}$$

$$s^{(t)} = g^{(t)} \cdot i^{(t)} + s^{(t-1)} \cdot f^{(t)} \tag{2-39}$$

$$v^c = \tanh(s^{(t)}) \cdot o^{(t)} \tag{2-40}$$

To summarize, the LSTM model operates through a set of equations, including those for forgetting gates, providing a comprehensive algorithm for modern LSTMs. The architecture's success lies in its ability to capture and balance long-term and short-term dependencies in sequential data.

2.3 Proximal Policy Optimization Algorithms Theory

This section introduces the basics of policy gradient algorithms and the technique used in the Proximal Policy Optimization (PPO) algorithm to improve the learning process. It is worth mentioning this chapter is based on the DeepMimic paper review of proximal policy algorithms [74]. However, only the important information is considered, while adding more information is required to understand the algorithm used in this study.

2.3.1 Policy Gradient Algorithms

Policy gradient methods learn the parameters of a policy based on the gradient of some scalar performance metric $J(\theta)$ with respect to the policy parameters. Their goal is to maximize the performance, so their updates approximate gradient ascent in J [75].

$$\theta_{t+1} = \theta_t + \alpha \widehat{\nabla J(\theta_t)} \tag{2-41}$$

where $\widehat{\nabla J(\theta_t)} \in \mathbf{R}^{d'}$ is a stochastic estimate whose expectation approximates the gradient of the performance measure with respect to its argument θ. Methods that follow this learning schema are called policy gradient methods. The method that learns approximations to value and policy functions is often called the actor-critic method, where the actor refers to the policy and criticizes the value function, usually a state-value function. Policy gradient methods work by computing an estimator of the policy gradient and plugging it into a stochastic gradient ascent algorithm [75]. The most commonly used gradient estimator is the form

$$\hat{g} = \hat{\mathbb{E}}_t \left[\nabla_\theta \log \pi_\theta(a_t|s_t) \hat{A}_t \right] \tag{2-42}$$

where π_0 is a stochastic policy and $\hat{A}_t$ is an estimator of the advantage function at time-step t. This function comes as a result of considering the weighted sum of rewards over a set of trajectories under a policy π. Note that no derivative with respect to the reward functions is needed, and neither is a system model.

Several software implementations build an objective function whose gradient is the policy gradient estimator [76]; this function has the form:

$$J(\theta_t) = \mathbb{E} \left[f(\theta_t) \hat{A}_t \right] \tag{2-43}$$

where f is the objective function estimator that approximates the policy update and $\hat{A}_t$ is the advantage function, which measures how much better or worse an action is compared to the average action at a given state. It reflects the impact of a given action on the total reward [76].

2.3.2 Multi-Step Return

To obtain an unbiased sample of the expected return, the Monte-Carlo return can be utilized [75]:

$$R_t = \sum_{l=0}^{T-t} \gamma^l r_{t+l}$$

Nonetheless, the accuracy of this estimator suffers from significant variability because r_t of the influence of a stochastic environment and policy. In an effort to address this problem, an alternative approach called the n-step return has been suggested. This method resigns on the zero-bias characteristic and reduces the variability by limiting the Monte-Carlo return to n steps, and employing the value function $V(s)$ to approximate the remaining steps:

$$R_t^{(n)} = \sum_{l=0}^{n-1} \gamma^l r_{t+l} + \gamma^n V_{(s_{t+n})} \tag{2-44}$$

For $n = 1$, this yields the low-variance high-bias estimator $R_t^{(n)} = r_t + \gamma V_{(s_t)}$, commonly employed in Q-learning. Setting $n = \infty$ retrieves the Monte-Carlo return if all rewards after time T are 0. Thus, the parameter n trades off bias versus variance in the estimator [75]. This bias-variance trade-off can also be addressed by using an exponentially weighted average of n-step returns with a decay parameter λ:

$$R_t(\lambda) = (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}R_t^n \tag{2-45}$$

Assuming all rewards after T are 0, $R_t^n = R_t^{T-t}$ for all $n \geq T-t$. This expression becomes:

$$R_t(\lambda) = (1-\lambda)\sum_{n=1}^{T-t-1}\lambda^{n-1}R_t^n + \lambda^{T-t-1}R_t^{T-t-1} \tag{2-46}$$

Here, $\lambda = 1$ corresponds to the 1-step return, and $\lambda = 0$ yields the Monte-Carlo return. Varying λ in the interval $(0,1)$ provides an estimator with high variance and low bias or vice versa. Empirically, the λ-return has demonstrated better performance than using the n-step return. When applied to updating the value function with the temporal difference method, this approach results in the TD(λ) method. Similarly, employing it to estimate advantages leads to the generalized advantage estimator GAE(λ):

$$\hat{A}_t = R_t(\lambda) - \hat{V}(s_t) \tag{2-47}$$

The algorithm uses the TD error. The temporal difference method defines the TD error as

$$\delta_t = (r_{t+1} + \gamma\hat{V}(s_{t+1},\theta_t)) - \hat{V}(s_t,\theta_t) \tag{2-48}$$

where the expected return is calculated using the estimates of the state-value function, $r_{t+1} + \gamma\hat{V}(s_{t+1},\theta_t)$. Applying it to the λ-return using the state-value function, the advantage estimator becomes

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \ldots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \tag{2-49}$$

which then can be used to calculate the expected return for the TD(λ) method,

$$R_t(\lambda) = \hat{A}_t + \hat{V}(s_t) \tag{2-50}$$

2.3.3 Off-Policy Learning

For policy gradient, the update of the parameters is done after the gradient of the expected return is calculated with samples using the current policy. After an update, new samples with the new policy need to be collected. This usage of data is inefficient. An alternative to improve data efficiency is importance sampling, which provides an unbiased estimator of the policy gradient using only off-policy samples from an older policy [74; 75]. This allows to reuse of samples from previous episodes run with older policies, avoiding the early discard of useful samples.

$$\nabla_\theta J(\theta) = \mathbb{E}_{s_t \sim d_{\theta_{old}(s_t)}, a_t \sim \pi_{old}(s_t|a_t)} \left[w_t(\theta) \nabla_\theta \log \pi_\theta(a_t|s_t) \hat{A}_t \right] \tag{2-51}$$

where θ_{old} are the parameters of the old policy, $w_t(\theta)$ is

$$w_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{2-52}$$

and is known as the probability or likelihood ratio [76].

In the importance-sampling policy gradient, a surrogate objective function can be used instead for the optimization

$$L^{IS} = \mathbb{E}_{s_t \sim d_{\theta_{old}(s_t)}, a_t \sim \pi_{old}(s_t|a_t)} \left[w_t(\theta) \hat{A}_t \right] \tag{2-53}$$

Finally, L^{IS} allows the use of a batch of data to update θ multiple times.

2.3.4 Proximal Policy Optimization

As a general description, the Proximal policy optimization (PPO) algorithm is a model-free, on-policy, online, policy gradient learning method. The practice has shown that policy gradient algorithms can still be extremely unstable due to the variance of estimators. Noisy gradients can lead to instability in learning. To mitigate methods to prevent big deviations of policy between updates and improve instability in learning have been proposed. Trust region policy optimization (TRPO) adds a hard constraint to the optimization problem of optimizing function $J(\theta)$ [76]. This hard constrain limits the Kullback–Leibler divergence between updates as

$$\max_\theta J(\theta) \tag{2-54}$$

$$s.t. \mathbb{E}_{s_t \sim d_\theta(s_t)} \left[KL(\pi_{\theta_{old}}(\cdot|s_t) | \pi_\theta(\cdot|s_t) \right] \leq \delta_{KL} \tag{2-55}$$

where δ_{KL} is another hyperparameter of the optimization problem. This problem can be approximately solved in an efficient manner using the conjugate gradient algorithm, after making a linear approximation to the $J(\theta)$ and a quadratic approximation to the constrain [76].

Proximal Policy Optimization is a variant of TRPO, which replaces the hard constraint by optimizing a surrogate loss. The most common surrogate loss for PPO is the clipped surrogate loss $L^{CLIP}(\theta)$ defined as

$$L^{CLIP}(\theta) = \mathbb{E}_{s_t,a_t} = \Big[\min(w_t(\theta)\hat{A}_t, clip(w_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\Big] \tag{2-56}$$

When $\theta = \theta_{old}$, $w_t = 1$, but when θ is different, the likelihood ratio differs from 1. Note, the first term inside the min is L^{CPI}. The second term clips the likelihood ratio, removing the incentive of moving outside of the interval $[1-\epsilon, 1+\epsilon]$.This term serves a similar purpose to the KL divergence constrain. Finally, the minimum of clipped and unclipped objectives is taken to obtain a lower bound (i.e., a pessimistic bound) on the unclipped objective. With this METHOD, the changes in likelihood ratio, when it would make the objective improve, are limited and when it makes the objective worse remain unlimited [76].

2.4 Performance Metrics for Control Problems

Several objective functions have been proposed to measure the performance of controllers, which can be used in the optimization problem of LFC in interconnected power systems to minimize the frequency and tie-lines power deviations. Integral square error (ISE), integral absolute error (IAE), integral time square error (ITSE), and integral time absolute error (ITAE) are the four most significant objective metrics for control problems:

$$J_{ISE} = \int_0^T \left(\Delta f_i^2 + \Delta P_{tie,i}^2\right) dt \tag{2-57}$$

$$J_{IAE} = \int_0^T \left(|\Delta f_i| + |\Delta P_{tie,i}|\right) dt \tag{2-58}$$

$$J_{ITSE} = \int_0^T \left(\Delta f_i^2 + \Delta P_{tie,i}^2\right) tdt \tag{2-59}$$

$$J_{ITAE} = \int_0^T \left(|\Delta f_i| + |\Delta P_{tie,i}|\right) tdt \tag{2-60}$$

3 Methodology

This chapter dives into the approach and methods employed in this study to investigate load frequency control (LFC) in power systems. The main focus is to simulate the Kundur case, a widely used power system model in dynamic stability studies. Proposed by P. Kundur [13], this model is composed of two mirrored areas connected by a single transmission line, acting as a weak link. The network topology consists of 10 buses and a virtual bus, featuring generators, transformers, loads, shunt capacitors, and tie-lines.

In the context of LFC, the study involves replacing one of the generators with a wind turbine to explore the control strategies in response to transient events. Two critical events, line disconnection, and re-connection, and a sudden load change, are considered, necessitating high-resolution time-domain simulations in the order of milliseconds.

To model the system, several standard models are employed. These models follow the modeling guidelines stated by the Western Electricity Coordinating Council (WECC) in order to meet strict specifications for their use in standards and regulations [77; 78; 79]. These models are widely used in power system simulation software like PSS/E, and Power World, or can be easily modeled using MATLAB/Simulink, and in Python using the ANDES library.

The chapter also provides an overview of the policy and value networks implemented for reinforcement learning, leveraging artificial neural networks (ANNs) with Long Short-Term Memory (LSTM) networks as feature extractors. The load frequency control with the reinforcement learning algorithm is detailed, including its key parameters, such as discount factor, likelihood ratio clipping threshold, and step size.

Considerations for the simulation, such as the smallest simulation step, the absence of noise in measurements, and the use of LSTM networks for recurrent processing, are discussed. The return function, defining the reward in reinforcement learning, is introduced, emphasizing its role in capturing the power system's stability.

The section on Policy and Value Networks addresses the choice of ANNs and LSTM networks for function approximation.

Finally, it concludes with an overview of the Proximal Policy Optimization algorithm, providing a comprehensive understanding of the learning process and parameter tuning. The last section details the Proportional Integral (PI) controller design.

3.1 System Model

Kundur case is a power system model commonly used in benchmarks for power system dynamic stability studies and research. It was proposed by P. Kundur. It is divided into two mirrored areas joined by a single transmission line, the tie-line, acting as a weak link. Its network topology is composed of 10 buses and a virtual bus in the middle of the weak link that joins the two areas. It contains 4 generators and 4 transformers, 2 loads, and 2 shunt capacitors at each end of the weak link. All are connected by 4 lines and the tie-line. That tie-line is divided from the middle in two sections. Two voltage levels are present, $20kV$ and $230kV$. For this study, generator no. 4, connected to bus 4, is replaced by a wind turbine, see Figure **3-1**.

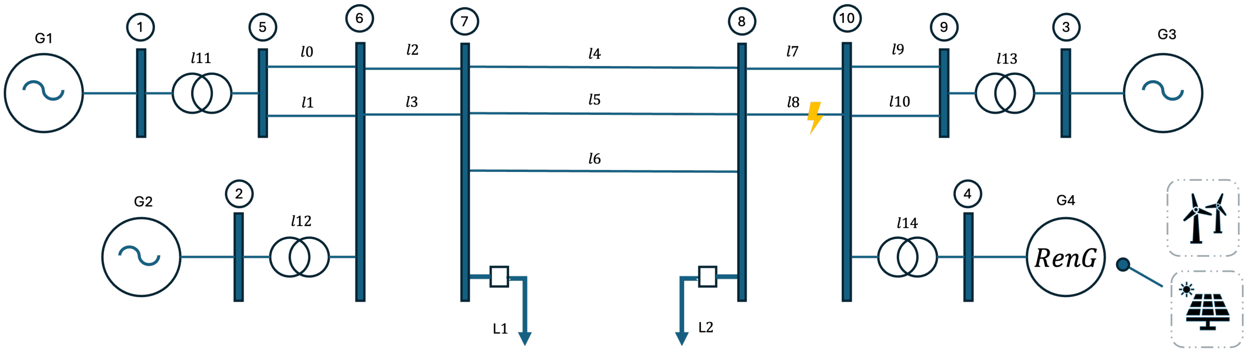


Figure **3-1**: Kundur System Diagram.
[13].

This study is about load frequency controllers. This requires analysis of transient events in the system to capture the transient response of controllers to these events. The two events considered are line disconnection and re-connection and a big sudden load change. This requires time-domain simulations in the order of milliseconds [79]. Suited models to accurately simulate the behavior of electrical machines in this temporal resolution are needed.

All dynamics of elements in the power system are modeled by differential-algebraic equations (DAE), which depend on having consistent initial values, making it harder to solve than ordinary differential equations (ODE). In this case, initial values are taken from the result of a power flow simulation previous to the simulation. Then the trapezoid rule is used to solve the differential equations.

All models used in this study to model electrical machines are standard models used in multiple power systems simulation software, like PSS/E or Power World, or are easily modeled using MATLAB/Simulink. These models follow the modeling guidelines stated by the WECC in order to meet strict specifications for their use in standards and regulations [79]. Their modeling was able thanks to the ANDES library in Python. This library contains the standard parts to recreate the different models.

The replication was considered crucial when choosing the models. Considering this, all models are available in the ANDES library. Moreover, any of the module's parameters and complete models can be found in the Annex A. So, they can be replicated not only in Python but in any other simulation program available for power system simulations using the WECC guidelines [77; 78; 79].

3.1.1 Modeling of Traditional Power System Elements

Synchronous Generator

The synchronous generator used in this study only comprises a machine model, an exciter model, and a governor Model. They are all interconnected as shown in Figure **3-2** and connect to the network through current commands [79]. The modules used to model the synchronous generators are listed below:

- Machine model GENROU, for machines with a solid round rotor with quadratic saturation
- Turbine governor TGOV1 without deadband
- Exciter model EXDC2A

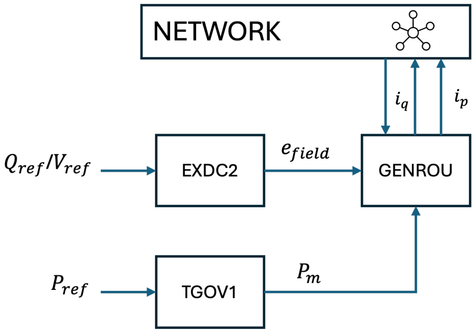


Figure **3-2**: Synchronous Generator Model: GENROU, TGOV1 AND EXDC2A standar models. [79]

Load model The load model uses a constant P-Q load model during initialization, employing results from the power flow. This model is then internally transformed into a constant Z model for transient simulations.

Line/Transformer Loads and transformers use the equivalent π model, as shown below in Figure **3-3**. This model is modified to be able to shift the phase to also model phase shifts in transformers.

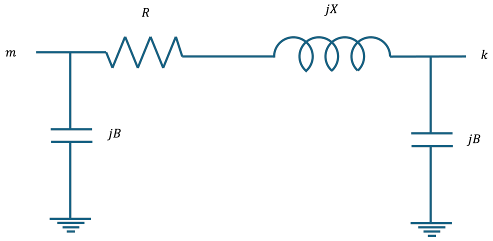


Figure **3-3**: π model employed for lines and transformers.

3.1.2 Modeling of Renewable Energies

Strictly speaking, renewable generation does not have a machine model, an exciter model, or a governor model. However, this block distribution can still be useful, by replacing the machine model with a converter model, the exciter model for electrical control, and the governor model for mechanical or plant control for renewable. The models presented in this section follow the WECC guidelines for renewable energy system dynamics [79].

For this study, only two systems will be considered: a wind turbine generator and a solar PV and battery storage hybrid plant. Choosing these two systems encompasses the three main systems proposed earlier in this document; wind turbines, energy storage systems, and PV systems. As per the experience of the WECC Modeling Validation Subcommittee, for transient studies the dynamics observed in the DC side of the converters are faster than power systems dynamics, this side is purposely ignored [78]. Then, solar panels or battery dynamics, elements placed on the DC side of the converter, are not incorporated into the modeling. Additionally, they use similar components to form their models. Consequently, and aiming to balance the modeling efforts versus the significance of the results, their mechanics are lumped into a single hybrid system. Its impact is seen in the turbine type and the minimum power of the converter model, from zero to a negative value.

Wind Turbine Generator

This model includes the three core components, a converter or generator, an electrical control, and a plant control. Additionally, it includes models of elements of the wind turbine like the drive train, a turbine torque controller, and a pitch controller [77]. And then, a model of the turbine aerodynamics. The models used to model the wind turbine generator are listed below:

- Renewable energy generator REGCA1
- Renewable energy electrical controller REECA1
- Pitch controller WTPTA1
- Wind turbine torque controller WTTQA1
- Renewable energy plant control REPCA1
- Wind turbine generator drive train WTDTA1
- Wind turbine aerodynamics WTARA1

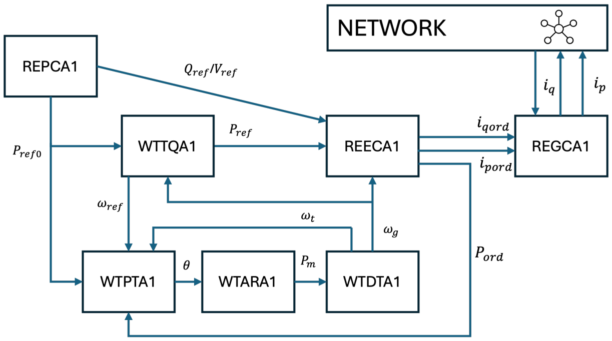


Figure **3-4**: Wind Turbine Model Integration: REGCA1, REPCA1, REECA1, WTPTA1, WTTQA1, WTDTA1, and WTARA1 standard models.
[77]

Solar PV and Battery Storage Hybrid Plant

The solar PV and Battery Storage Hybrid Plant is modeled using just the three components, due to the reasons mentioned in the introduction of this section. The models used to model the hybrid plant are listed below:

- Renewable energy generator REGCA1
- Renewable energy electrical controller REECA1
- Renewable energy plant control REPCA1

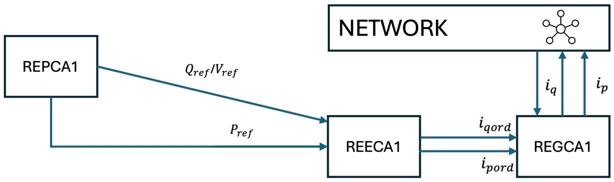


Figure **3-5**: Solar PV and Battery Storage Hybrid Plant Model Integration: REGCA1, REPCA1 and REECA1 standard models.

3.2 Simulation and Training-Testing Considerations

3.2.1 Tests Descriptions

Line Disconnection Test

To simulate a disturbance in the system, a disconnection event was initiated in the tie-line. This disconnection occurred 2 seconds after the simulation commenced and lasted for 500 milliseconds. During this period, the Load Frequency Control (LFC) needed to adjust the generated power of all synchronous generators to mitigate the Area Control Error (ACE) and stabilize the frequency.

Load Change Test

This scenario is set up to test the generalization of the policy. It consists of changing the load PQ_0, decreasing it by 25% at 2 seconds after the simulation started, and increasing it again by 10% at 30 seconds.

3.2.2 Power System as an Environment for Reinforcement Learning

Before integrating this model into the reinforcement learning framework, certain definitions must be established, and considerations must be addressed.

In this study case, the power system functions as the environment. The observation space comprehends all relevant measurements for the algorithm to determine the power system's state, including variables such as generator speed (ω) and ACEs. The agent interacts with the environment by modifying an auxiliary input (P_{aux}), which is added to the power reference in the governors of the synchronous generators.

There are several considerations to the simulation. The smallest simulation step possible is 33 milliseconds. After each step, the values of all system variables, including ACEs, and the phase angle (δ) and speed (ω) of each generator, are known, for a total of 11 measurements. However, actions are not executed after every simulation step but rather at intervals of 100 milliseconds. The algorithm receives all preceding measurements in between actions each 100-millisecond. The observations from each 33-millisecond step within each 100-millisecond interval are aggregated and sent to the agent. To effectively use this information, the policy employs a recurrent neural network known as a Long Short-Term Memory Network (LSTM) with only forward processing. This network is also shared with the value function to assess state values.

It is important to note that no noise is considered in measurements, and it is assumed that all information about the power system state is available; there are no missing measurements. Furthermore, no communication delays are taken into account, and the speed at which actions are executed is limited to 100 milliseconds, conceiving impractical for the algorithm to counteract frequency deviations that occur more rapidly.

3.2.3 Reward Function

The return function is a negative sum of all $ACEs$:

$$r_t = -\sum_i |ACE_{ti}| - 0.5 \sum_i |Ptie_{t_i} - Ptie_{0_i}| \tag{3-1}$$

Other return functions were considered. For example, the quadratic cost function is used in Linear Quadratic Regulators. This was selected because it was the most simple one, avoiding observed behaviors like obtaining policies similar to the ones in other algorithms.

3.2.4 Termination Logic and Training Description

For learning, the algorithm only runs for 20 seconds. During those 20 seconds, there is a disturbance on line 8 that causes a disconnection for 500 milliseconds. After the 20 seconds, the episode is considered truncated, rather than terminated, so that the updates of the policy and value functions are completed correctly. The simulation is stopped and the done flag is set to true in two cases, when the stability criteria for the power system are not met or, in a more extreme case, the system is busted. In that case, the return is -200. The resulting policy will be used in the load change scenario.

3.3 Policy and Value Networks

Reinforcement learning faces a problem with continuous state and action variables. Managing large spaces becomes impractical in terms of memory and computational resources. It's nearly impossible to go through every state during training in such vast spaces. To address this, one needs to leverage the similarities between observed states and new ones to make informed decisions. This principle extends to actions as well, where even a slight difference in states should prompt a corresponding adjustment in actions. The main challenge is making sure the system generalizes well.

In the context of power systems, states, and actions inherently have a continuous nature. In policy gradient algorithms, adapting to continuous spaces involves approximating the policy and value functions. Among the various methods for function approximation, artificial neural networks (ANNs) stand out for their proven success across a diverse set of applications. ANNs exhibit key features such as adaptability and flexibility, allowing them to effectively learn complex patterns in data, including non-linear patterns. This adaptability is particularly valuable in reinforcement learning applications, where good generalization is essential. In this specific study, the policy parameters, denoted as θ, and the value function parameters, denoted as ψ, are parameters representing neural networks. Both the policy and value functions share a common base, specifically, an LSTM (Long Short-Term Memory) neural network for feature extraction, which is later used by two different feed-forward neural networks, one for each function, policy and value see Figure **3-6**.

This feature extractor was selected due to its flexibility in encoding information from a series of any length in time. Allowing the use of any set of measurements before acting. The feature extractor is updated along with both, value function and policy. Thus, it encodes information from both, value and actions.

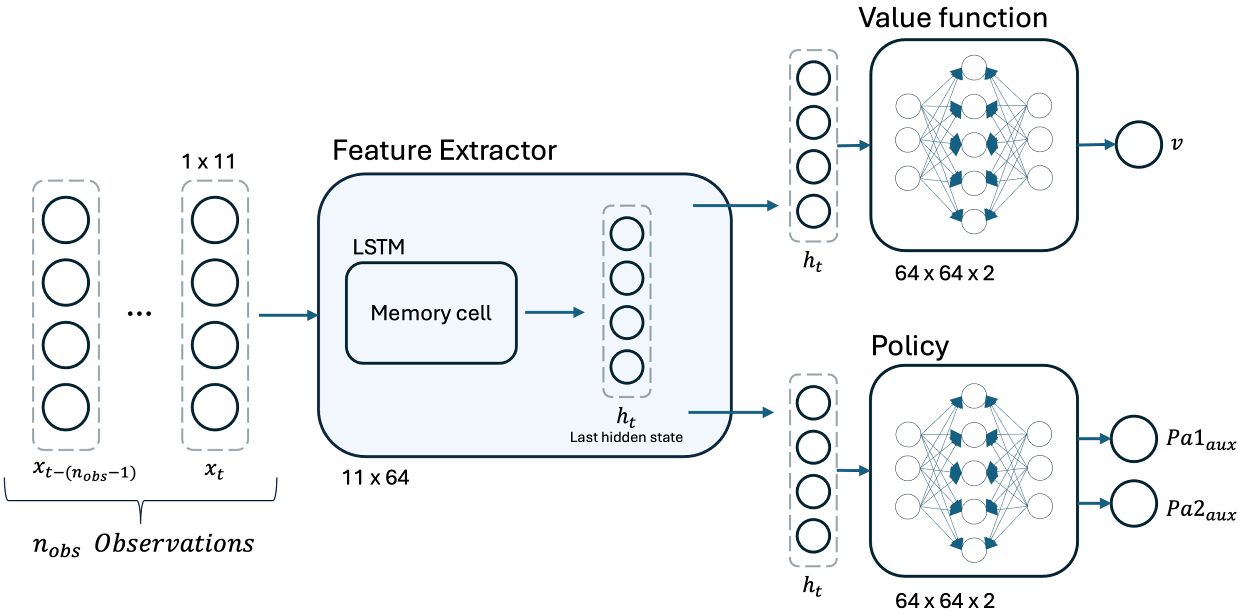


Figure **3-6**: Policy and Value Functions with Common Feature Extractor.

3.4 Load Frequency Control with Proximal Policy Optimization

In this section, the algorithm is described. The learned policy resulting from this algorithm is then tested in the same state and actions framework used in reinforcement learning. Each step is a simulation of 100 milliseconds. After each step is finished, observations are used to compute the action to modify the auxiliary inputs of the generators.

3.4.1 Algorithm

Algorithm 1 summarizes the common learning procedure used to train all policies. A discount factor $\gamma = 0.99$ is used for all motions. Discounter reward parameter $\lambda = 0.95$ is used for both TD(λ) and GAE(λ). Updates on the policy are performed after a batch of $m = 2048$ samples has been collected. Mini batches of size $n = 64$ are then sampled from the data for each gradient step. The likelihood ratio clipping threshold is set to $\epsilon = 0.2$. An initial stepsize of $\alpha = 1e - 4$ is used for the value function and policy updates. This could be modified if more accuracy in the learning process. Once gradients have been computed, the network parameters are updated using Adam optimizer, with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e - 8$. The same hyper-parameter settings are used for all characters and skills, with the exception of the step size. This algorithm, and other implementations, can be found in Stable-Baselines3 library [80]. This same library was used in this study to train the control policy with their implementation of the PPO Algorithm 1. Policies for this problem require about 600 thousand samples to train, requiring about 10 hours on a 4-core machine. All simulation and network updates are performed on the CPU and no GPU acceleration is used.

Algorithm 1 Proximal Policy Optimization

```
$\theta \leftarrow$ random weights
$\psi \leftarrow$ random weights
while not done do
    $s_0 \leftarrow$ simulate power system for j steps to get initial state
    Initialize character to state $s_0$
    for step $= 1, \dots, m$ do
        $s \leftarrow$ start state
        $a \sim \pi_\theta(a|s)$
        Apply $a$ and simulate forward one step
        $s' \leftarrow$ end state
        $r \leftarrow$ reward
        record $(s, a, r, s')$ into memory (roll-out buffer) $B$
    end for
    $\theta_{\text{old}} \leftarrow \theta$
    for each update step do
        Sample minibatch of $n$ samples $\{(s_i, a_i, r_i, s'_i)\}$ from $B$

        Update value function:
        for each $(s_i, a_i, r_i, s'_i)$ do
            $y_i \leftarrow$ compute target values using TD($\lambda$)
        end for
        $\psi \leftarrow \psi + \alpha_v \frac{1}{n} \sum_i \nabla_\psi V_\psi(s_i)(y_i - V(s_i))$

        Update policy:
        for each $(s_i, a_i, r_i, s'_i)$ do
            $A_i \leftarrow$ compute advantage using $V_\psi$ and GAE($\lambda$)
            $w_i(\theta) \leftarrow \frac{\pi_\theta(a_i|s_i)}{\pi_{\theta_{\text{old}}}(a_i|s_i)}$
        end for
        $\theta \leftarrow \theta + \alpha\pi \frac{1}{n} \sum_i \nabla_\theta \min(w_i(\theta)A_i, \text{clip}(w_i(\theta), 1-\epsilon, 1+\epsilon)A_i)$

    end for
end while
```

3.5 PI design

The design of the Proportional Integral (PI) controllers is done according to Figure **2-1**. Each area has its own controller, that follows the common structure of a PI controller. The output of each PI controller is determined by the proportional and integral components. For a generic PI controller, the control signal $u(t)$ is given by

$$u(t) = K_p \cdot e(t) + K_i \cdot \int_{t_0}^{t} e(\tau)\, d\tau$$

Here, K_p is the proportional gain, K_i is the integral gain, and $e(t)$ is the error signal, which in this case is the ACE.

The PI controller parameters (K_p and K_i) are obtained through optimization using the Nelder-Mead method for a max of 100 steps. The negative of the total sum of rewards serves is the objective function during this optimization process, making sure that the controllers are tuned to minimize the same objective function as the RL algorithm.

The control signals generated by the PI controllers are distributed among the generators within each area based on their α_i coefficients. In this study, they are set to $1/n$, where (n) is the number of generators in an area that can participate in the frequency regulation. The auxiliary signal for the (i)-th generator $(P_{aux_i}(t))$ is determined by multiplying the control signal $(u_i(t))$ with the associated coefficient (α_i):

$$P_{aux_i}(t) = u_i(t) \cdot \alpha_i \tag{3-2}$$

This distribution allows each generator to contribute to the overall control effort based on its characteristics.

4 Results

To implement load frequency control (LFC) within power systems, this study employs a reinforcement learning (RL) methodology. The following section presents a detailed analysis of the results obtained, exploring the effectiveness of RL in tackling the complexities associated with LFC.

This results section is organized to present a detailed analysis of the system dynamics after a load disturbance with PI and RL controllers, allowing for a detailed understanding of the RL algorithm's performance in load frequency control under a disturbance scenario. Results are presented separately, to allow a comprehensive view of the system transient response. Along with a comparison analysis of the conventional PI controller versus the reinforcement learning-based controller.

It begins by examining the mean reward over simulation steps, followed by the performance of the algorithm, and finally computing criteria metrics for comparison. The evaluation criteria include ISE and IAE. These metrics serve as benchmarks for comparing results with existing methodologies.

4.1 PPO Training and PI Controller Optimization

In the following Figure **4-1**, the mean reward across the training process is shown. The initial value of the reward is low, as the controller still does not know a good policy. As it starts exploring and learning the correct values, it improves. There are no peaks or high variations in the mean reward, which indicates that the training was stable. As expected, it plateaus at some values where there is a need for the change of learning rate to continue improving. From these changes, there are small improvements. On the other hand, the yellow line is the maximum value reached by the PI controller, which is surpassed by the mean reward of the PPO algorithm. Additionally, it is important to mention that the total real training time was 10 hours, with only a 4-core CPU.

As per the PI controller optimization, it took half an hour, far less than the computation resources used for the PPO controller. The values obtained were $K_{p_1} = 0.13937$ and $K_{i_1} = -0.00075$ for area 1, and $K_{p_2} = 0.01519$, $K_{i_2} = -0.00019$ for area 2.

4.1.1 A Note on Hyperparameters Optimization

A thorough optimization was not performed in this work, but it does not mean there was not a tuning of hyperparameters. The selection of the structure size varied from values from 16 to 128 nodes per layer. For the low limit, the model did not have the capabilities to cover the whole range of the generators' power input, while keeping the numerical accuracy to perform the task correctly. While the bigger the network the

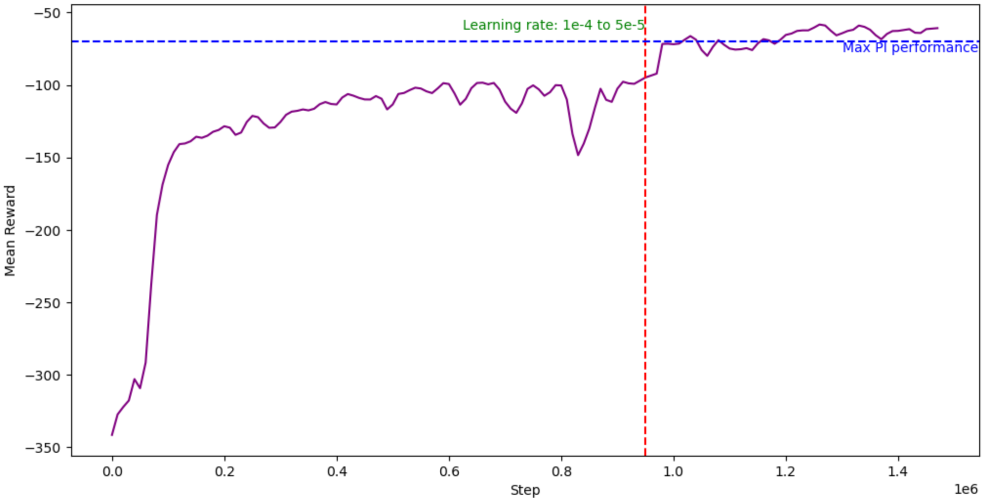


Figure **4-1**: Reward versus Learning Steps

higher the accuracy, the model tends to unlearn the objective task.

To get an acceptable behavior, the reward function is crucial. So, this was also a parameter to consider. First, the negative of a quadratic loss function was used as a reward function. However, this enforces behaviors in the model that can be similar to known controllers. It was also way harder to avoid unwanted behaviors, like learning a suboptimal task. To avoid this, the reward function was simplified considering just l_1 distance to the desired frequency values. A study on the reward functions and their outcomes in comparison to known controllers is worth exploring.

Finally, Another tuning was done in the learning rate, as it needed to be adjusted to reach accurate values to keep the working point of the system. It was reduced once the training reward stopped increasing.

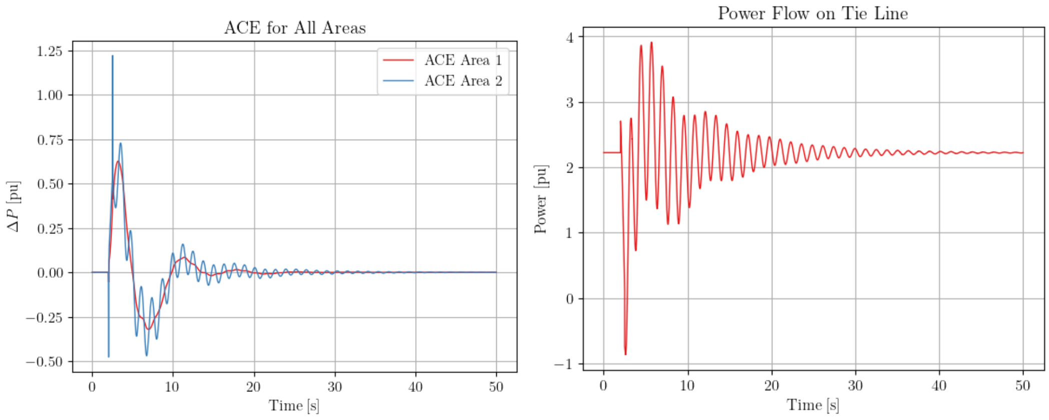
4.2 Line Disconnection Test

4.2.1 System with Wind Turbine Generator

For the system with a wind generator, Figure **4-2**a shows the ACEs, before and after the disturbance in line 8. The sudden load drop increased the frequency on the nodes, which triggered a response in the governors of the generators. The drop controller then tries to adjust the generation to meet the new load, while keeping the frequency changes in a secure level and stabilizing the generators. However, the tie-line power transference also suffers some changes, this can be seen in Figure **4-2**b.

After the disturbance is cleared, it takes more than 20 seconds for the primary controllers to stabilize the frequency. The ACE for Area 1 has fewer changes and ripples. This is the area with the majority of the inertia of the system. Also, the ACE of Area 2 now has bigger peaks and dips and seems to be coupled with the frequency changes and tie-line power transference from the changes in Area 1. This is explained by the presence of a mechanical coupling. This is the signal used in the PI controller as input and among others for the RL controller.

Now, considering the LFC PI controller designed using the optimization strategy, Figure **4-3**a shows the

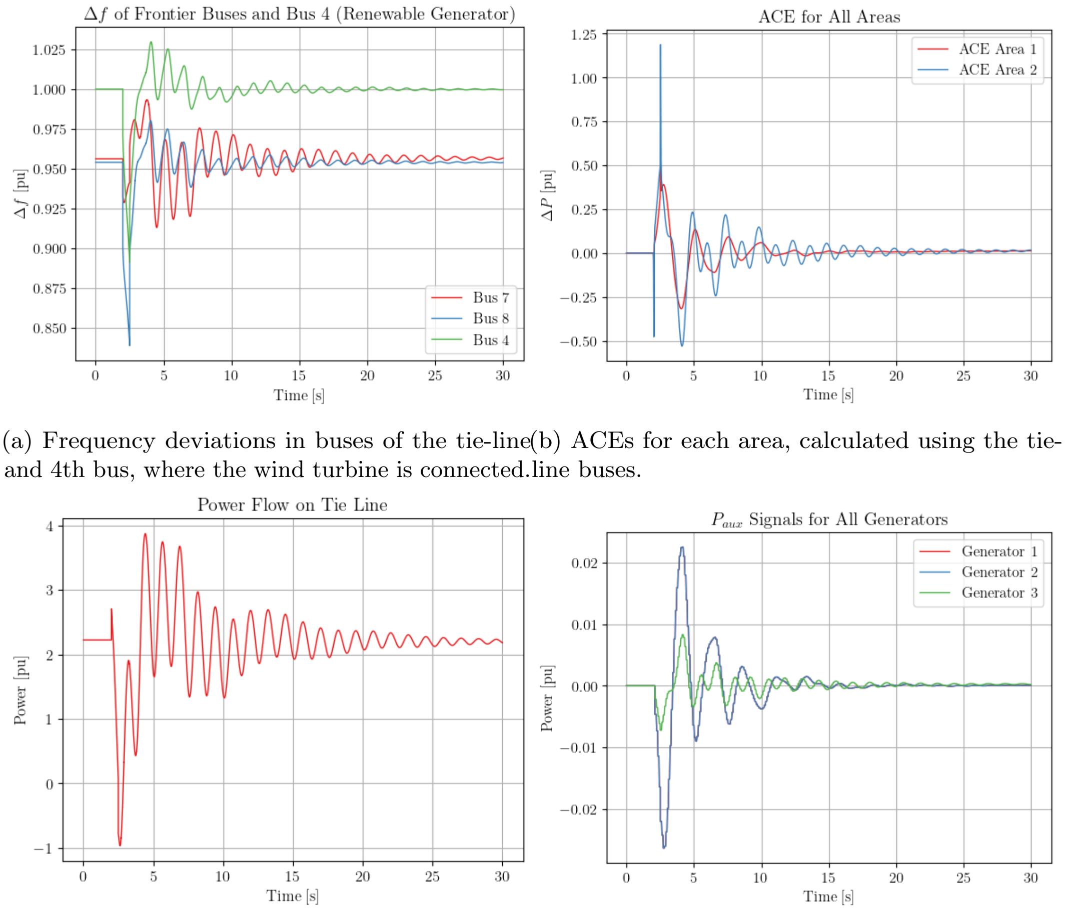


(a) ACEs without load frequency control (b) Tie-power without load frequency control

Figure **4-2**: Results of 8th line disconnection without Load Frequency Control for system with Wind Turbine Plant replacing 4th Gen.

frequency deviations. In there, the stabilization appears faster. The ACEs shown in Figure **4-3**b allow us to see peaks earlier than the system without a controller, due to the complexity added by the acting time resolution in the system. The auxiliary control signals are shown in Figure **4-3**d. The control actions are trying to compensate for the disturbance effects. Note that the tie-power transferred between areas in Figure **4-3**c suffered a subtle reduction compared to the system without a controller. After the stabilization, it goes back to the previous disturbance level.

Then, for the PPO controller, Figure **4-4**a shows the frequency deviations. Frequency reaches smaller peaks and dips than with the PI controller. The ACEs show a lower frequency deviation, especially in the peaks. However, at the start of the episode, due to the nature of the network, it was a hard task for the network to learn to output a zero auxiliary reference so it adds a non-zero signal to area 2. A closer look at the auxiliary signals in Figure **4-4**d, confirms this. Observe that ACE is being reduced to zero, but there is some chattering. The tie-line power in Figure **4-4**c is also maintained near the initial value, but keeping some chattering, that later disappears.



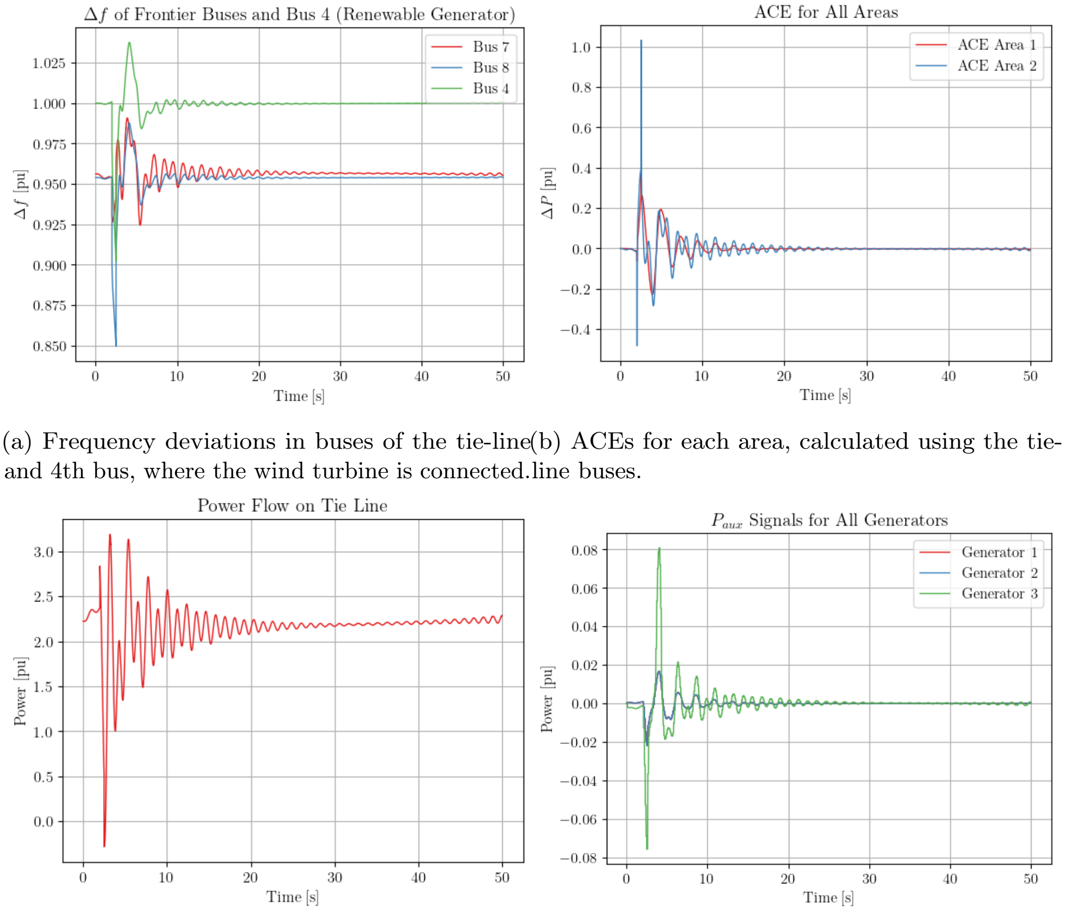
(a) Frequency deviations in buses of the tie-line and 4th bus, where the wind turbine is connected.

(b) ACEs for each area, calculated using the tie-line buses.

(c) Tie-line power transference.

(d) Auxiliary inputs for the generator set-points.

Figure **4-3**: Line disconnection results: frequency deviations, ACEs, tie-line power transference, and auxiliary inputs for generators power set-points with the PI controller and a Wind Turbine Plant replacing 4th Gen.



(a) Frequency deviations in buses of the tie-line and 4th bus, where the wind turbine is connected.

(b) ACEs for each area, calculated using the tie-line buses.

(c) Tie-line power transference.

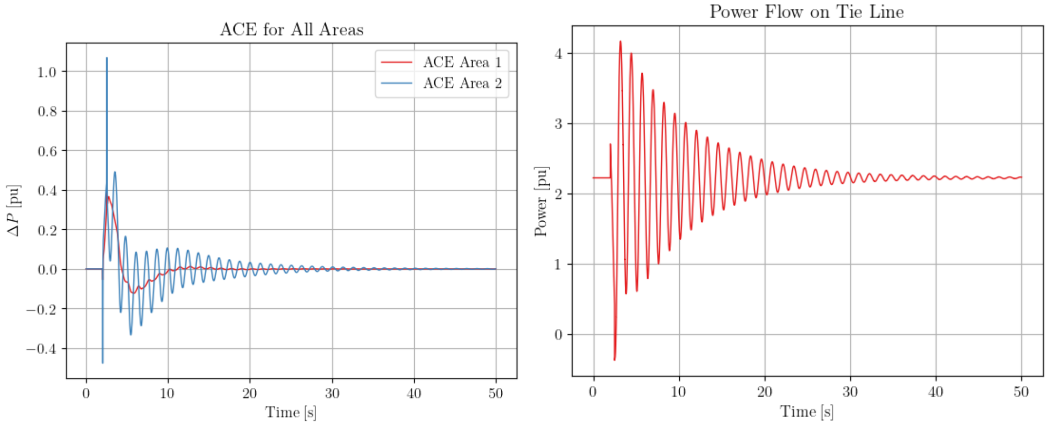
(d) Auxiliary inputs for the generator set-points.

Figure **4-4**: Line disconnection results: frequency deviations, ACEs, tie-line power transference, and auxiliary inputs for generators power set-points with the PPO controller and a Wind Turbine Plant replacing 4th Gen.

4.2.2 System with Solar-Battery Hybrid Plant

For the system with a solar-battery hybrid plant generator, Figure **4-5**a shows the ACEs, before and after the disturbance in line 8. The response is similar to the system with the wind turbine. However, the faster response from the drop controller, which acts as modifying the generation, to meet the transient frequency changes reduces the frequency changes in Area 1. For Area 2 the lower inertia causes an increase in frequency peaks and dips, and those changes do not follow the Area 1 changes. This is explained by the lower mechanical coupling.

Again, after the disturbance is cleared, it takes around 20 seconds for the controllers to stabilize the frequency. The ACEs are shown in Figure **4-5**a. The tie-line power transference is shown in Figure **4-5**b, where is possible to see that the decay of the transient signal is lower than the system with the wind turbine.



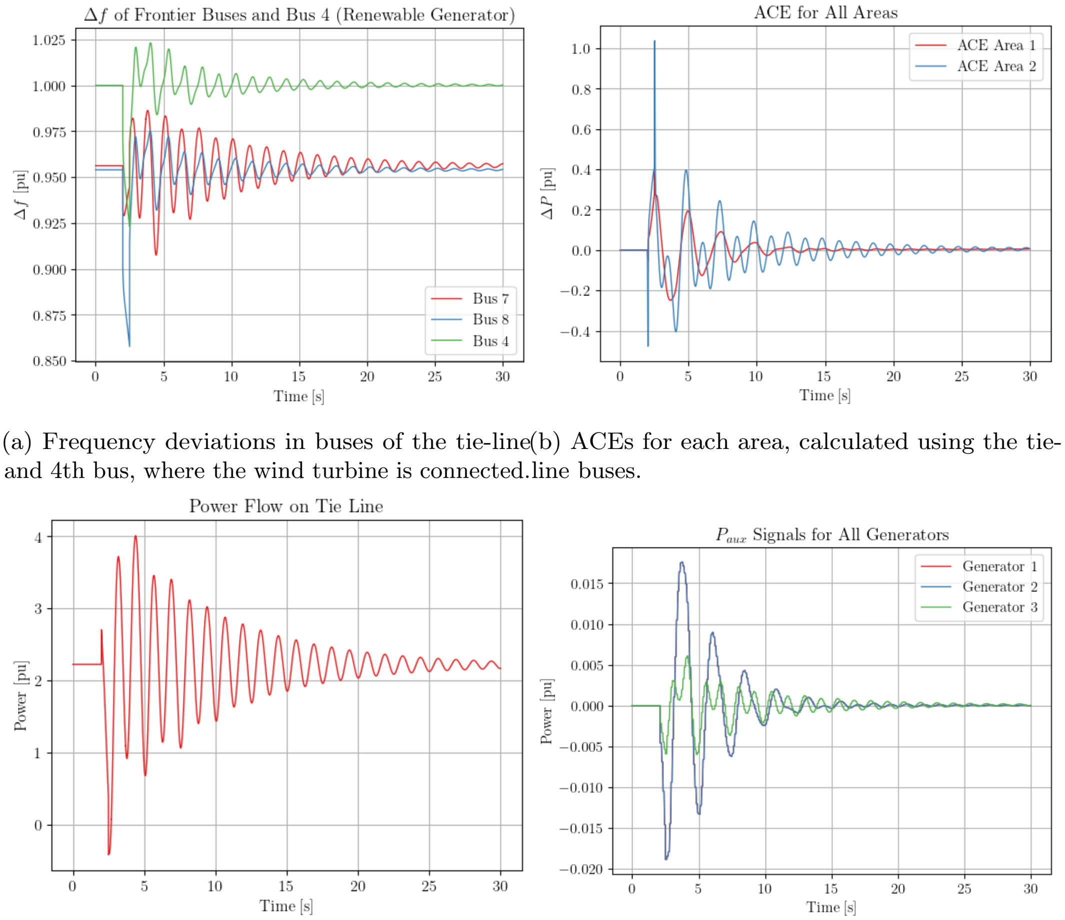
(a) ACEs without load frequency control

(b) Tie-power without load frequency control

Figure **4-5**: Results of 8th line disconnection response without Load Frequency Control for system with Solar-Battery Hybrid Plant replacing 4th Gen.

For the PI controller, the stabilization effect on frequencies, in Figure **4-6**a, is not noticeable and there are even more peaks. The ACEs shown the Figure **4-6**b, follow the frequency changes in their respective areas, and also the power transference through the tie-line. The auxiliary control signals shown in Figure **4-6**d show that the control actions are helping to compensate for the disturbance effects and try to reduce the settling time and that the tie-power transferred between areas in Figure **4-6**c is also being affected, but their changes are small. After the stabilization, it goes back to the previous disturbance level, as expected.

Additionally, Figure **4-7**a shows the frequency deviations with the PPO controller. Frequency reaches lower peaks and dips. The ACEs show a lower frequency deviation, especially in the peaks. However, at the start of the episode, due to the nature of the network, it was a hard task for the network to learn to output a zero auxiliary reference, inducing some offset in the frequency. A closer look at the auxiliary signals in Figure **4-7**d, confirms this. Also, note that ACE is being reduced to zero, but there is also some chattering left. The tie-line power in Figure **4-7**c is also maintained near the initial value but keeps some chattering. See that the controller tries to keep the ACEs near zero as fast as possible, by using a higher control action.



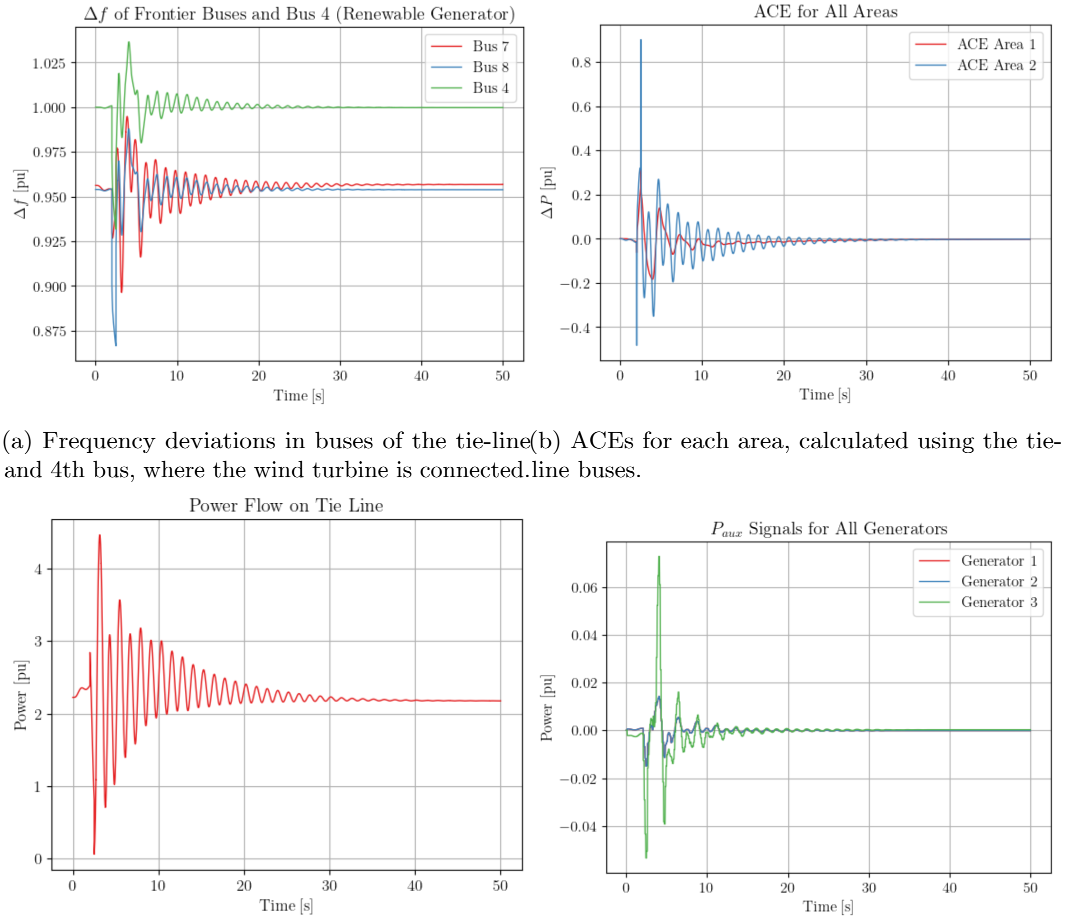
(a) Frequency deviations in buses of the tie-line and 4th bus, where the wind turbine is connected.

(b) ACEs for each area, calculated using the tie-line buses.

(c) Tie-line power transference.

(d) Auxiliary inputs for the generator set-points.

Figure **4-6**: Line disconnection results: frequency deviations, ACEs, tie-line power transference, and auxiliary inputs for generators power set-points with the PI controller and a Solar-Battery Hybrid Plant replacing 4th Gen.



(a) Frequency deviations in buses of the tie-line and 4th bus, where the wind turbine is connected.

(b) ACEs for each area, calculated using the tie-line buses.

(c) Tie-line power transference.

(d) Auxiliary inputs for the generator set-points.

Figure **4-7**: Line disconnection results: frequency deviations, ACEs, tie-line power transference, and auxiliary inputs for generators power set-points with the PPO controller and a Solar-Battery Hybrid Plant replacing 4th Gen.

4.2.3 Controllers Performance Comparison

To better observe the difference between the two controllers, the ACEs for both controllers for the system with the wind turbine generator are shown in Figure **4-8**. The PPO controller presents a slightly better performance than the optimized PI controller. However, the PPO controller also poses a bigger burden on generator 2, in area 2, as seen in Figure **4-4**d. The algorithm learned that is better and more straightforward to modify the power output of generator 2, but this is going to present some problems once it is tested on generalization, as observed in the next section.

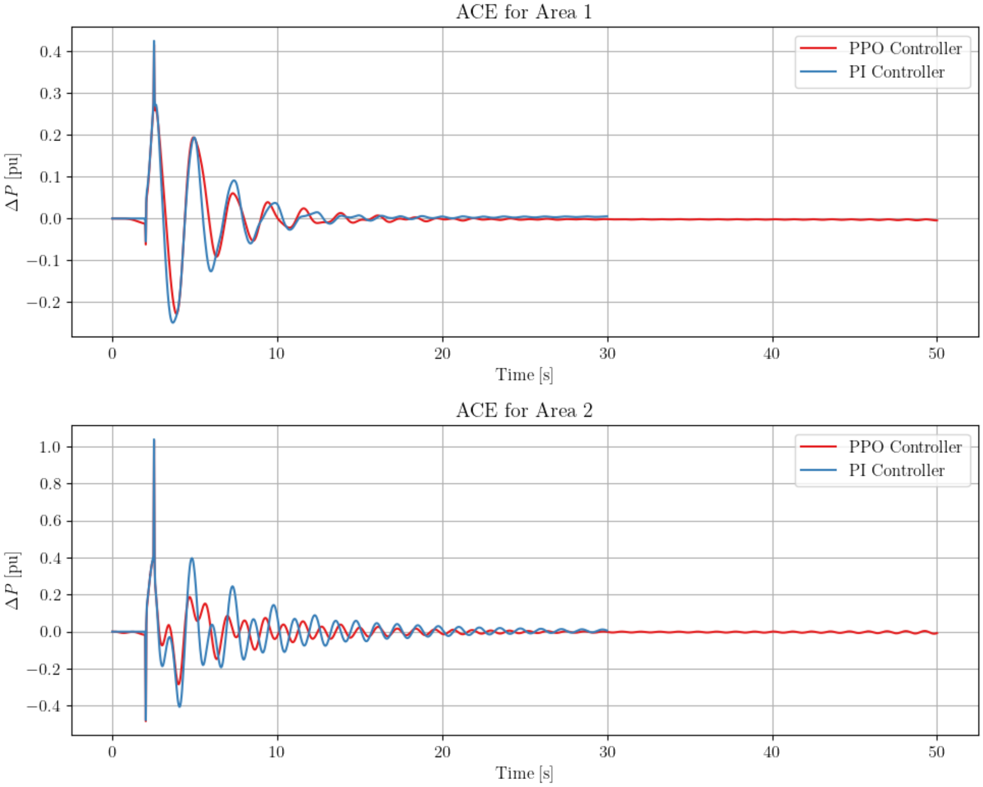


Figure **4-8**: ACEs comparison between PI and PPO controllers in the power system with Wind Turbine Plant replacing 4th Gen.

The following tables show the ISE and IAE metrics in the six scenarios: The system with a wind turbine without the controller, and with PI and PPO controllers, and for the system with a solar-battery hybrid plant without the controller, and with PI and PPO controller. Notice that the system can have better performance with PPO than with PI, or without a controller. This demonstrates an improvement in performance to meet the two objectives, reducing the frequency deviations while maintaining the tie-line power transference inside the limits. However, the difference is subtle. Next, we are going to test the generalization of the policy to the meet goals of the LFC.

As per the ADRC controller, that is approached by a colleague in his thesis [68]. This controller design is based on advanced control theory and leverages more information from the system than the PI controller. Then, its stabilization time for this specific problem falls into the order of 5-10 seconds frame, which is far better than the actual PPO controller. However, dynamics not modeled in the system, like delays, noises, communication network problems, and discrete signals and acting times, would hinder the performance of this controller. To provide a fair comparison, the ADRC controller needs to be adapted to consider the discrete acting times and re-tested using the same environment used in this study, which is out of the scope of this document, and is proposed as future work.

Wind Turbine Generator			
Variable	**Controller**	**IAE**	**ISE**
P_{tie}	PI Controller	322.12	298.58
	PPO Controller	**311.79**	**246.55**
	No Control	395.66	377.68
ACE_1	PI Controller	29.86	4.53
	PPO Controller	**28.93**	**2.41**
	No Control	33.52	6.67
ACE_2	PI Controller	53.36	10.62
	PPO Controller	**47.69**	**7.27**
	No Control	64.23	13.53

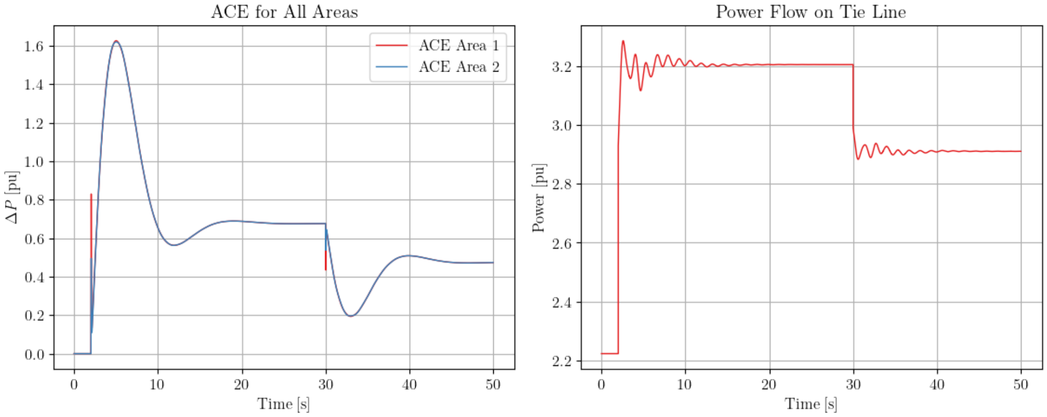
Table **4-1**: Control Performance Metrics for Power System with Wind Turbine Generator

Solar-Battery Hybrid Plant			
Variable	**Controller**	**IAE**	**ISE**
P_{tie}	PI Controller	313.57	360.81
	PPO Controller	**311.79**	**246.55**
	No Control	342.93	340.22
ACE_1	PI Controller	37.18	6.92
	PPO Controller	**28.93**	**2.41**
	No Control	78.7	27.83
ACE_2	PI Controller	53.57	12.78
	PPO Controller	**47.69**	**7.27**
	No Control	91.38	31.55

Table **4-2**: Control Performance Metrics for Power System with Solar-Battery Hybrid Plant

4.3 Load Change Test

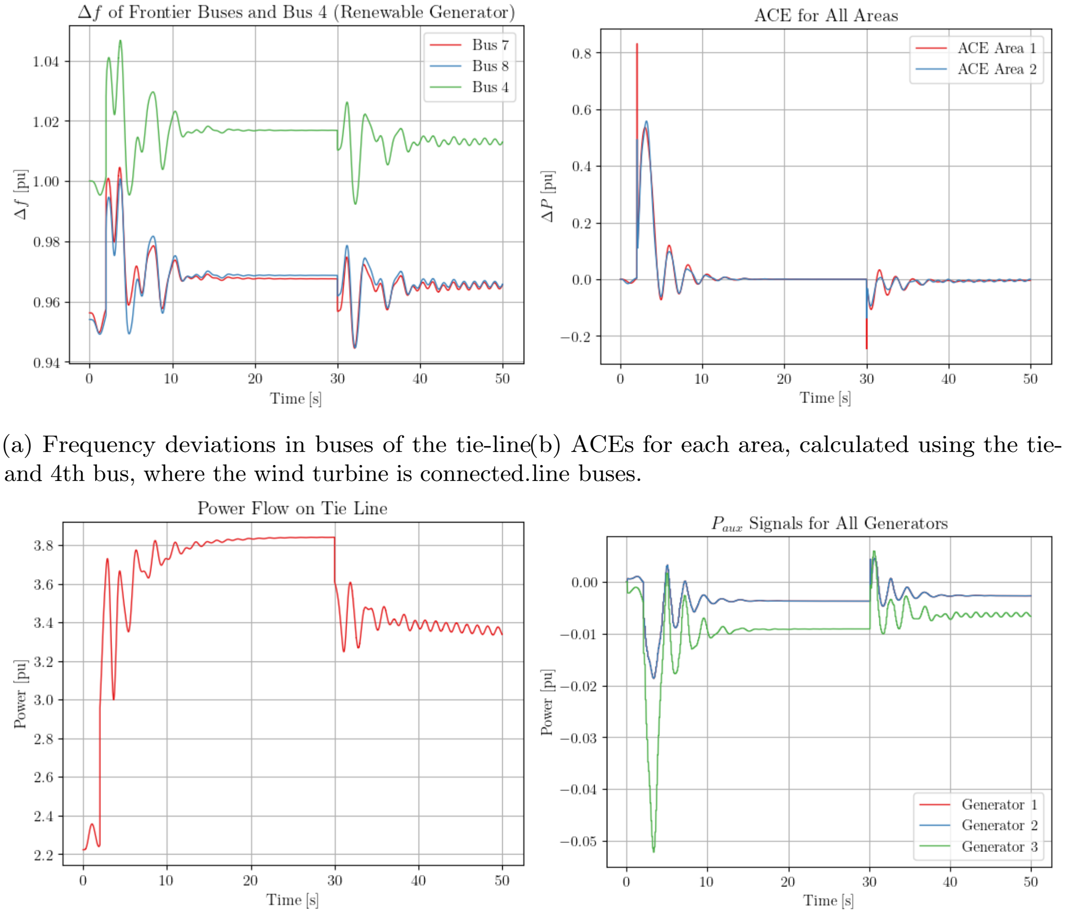
The load change test enables to observation of the behavior of the policy trained using disturbance data, ensuring it learns to help the primary frequency controllers to keep the frequency changes between desired and secure levels. In Figure **4-9**, the ACEs and the tie-line power transference are shown. The increase in power transference also coincides with a positive change in the ACEs, but the differences do not match, indicating an important impact due to frequency deviations.



(a) ACEs without load frequency control (b) Tie-power without load frequency control

Figure **4-9**: Results of load change response without load frequency control for system with Wind Turbine Plant replacing 4th Gen.

The same PPO controller used in the previous section is used here. After the stabilization, in Figure **4-10**a, frequency settling values changed from levels previous to the load changes. Also, the ACEs are around zero, as shown in the Figure **4-10**b. While the algorithm was able to reduce the frequency changes, the tie-line power, in Figure **4-10**c, presents an offset. This indicates that the algorithm may require another mechanism to ensure the offset is reduced while keeping the ACEs down to zero. As the change occurs on the load that is in Area 1, the auxiliary inputs of the generators in Area 1 should be the ones addressing the load change, but the power output of the generator 2, in Area 2, suffered the most modifications, as shown in the Figure **4-10**d. This behavior is suspected as a result of the policy the algorithm learned during training, which tends towards modifying the output of that generator to cope with disturbances on the 8th line.



(a) Frequency deviations in buses of the tie-line and 4th bus, where the wind turbine is connected.

(b) ACEs for each area, calculated using the tie-line buses.

(c) Tie-line power transference.

(d) Auxiliary inputs for the generator set-points.

Figure **4-10**: Load change results: frequency deviations, ACEs, tie-line power transference, and auxiliary inputs for generators power set-points with the PPO controller and a Wind Turbine replacing 4th Gen.

5 Discussion

The introduction of renewable energies to the systems has increased the complexity of the control problem. Less inertia increases the need for reliable load frequency controllers that can be able to handle the collective deviations of frequency, and other system parameters, from the steady and stable states. Wind turbines and solar plants, despite being able to participate with their drop controllers, are not able to participate in a coordinated strategy to reduce those deviations. Among control strategies, PI control strategies have been spread across industry and control of power systems also, because of their simplicity and easy implementations. It is also very intuitive. However, it falls short under different situations where there is a need for better coordination.

Considering the results, the application of Proximal Policy Optimization in the load frequency control of power systems demonstrated good results. The experiments show that PPO is able to provide a control signal to effectively reduce the area control error with a good performance, and leveraging all the available information from the system, in situations the algorithm has seen. PPO is able to handle nonlinearities and disturbances in the system for effective control applications. Nevertheless, generalizing to other scenarios, where the enforcement-specific conditions is required is hard for this algorithm. It needs to see more scenarios during training to improve the policy. Moreover, its behavior is complicated to asses under unseen scenarios.

One of the advantages of PPO is the possibility to adapt the algorithm to continue learning and improving in different situations and system changes as it is system agnostic, this is called online learning. It can also use real information from the current system performance as a baseline to continue improving. This could help mitigate encountering unseen situations, and the robustness of the controller. Some solutions have been proposed, like the use of the neural network to tune the Proportional Integral controller parameters [81], or a Lyapunov function previous to the neural networks to ensure the convergence of the output signals and the stability of the system [82].

However, these solutions, aside from employing sophisticated solutions, lack the core essence of reinforcement learning and limit the way the algorithm can interact with the environment. This highlights the need for solutions more related to those employed under the reinforcement learning framework in other applications, like game simulations.

6 Conclusions

In Chapter 3, standard models were employed by following the WECC guidelines to develop a version of the test bench Kundur power system in Python. By employing the tools provided by the ANDES library, the power system was modeled. The use of these standard models enables replication and results validation. later, this model is adapted to the reinforcement learning framework with the logic stated in the second section of this chapter. Then, the learning algorithm using Proximal Policy Optimization (PPO) was developed. Alongside a quick strategy to develop a Proportional Integral (PI) conventional controller. After successfully using the algorithm and developing the controllers, the results are shown in Chapter 4. A discussion in Chapter 5 follows the results, summarizing and discerning about the observations.

As a result, while the findings in this document highlight the positive aspects of PPO in power system control, some considerations should be taken into account for real-world applications. Such as computational requirements, communication delays, measurement noise, and hardware constraints may impact the feasibility of deploying PPO in practical power grid environments. Additionally, reinforcement learning-based controllers required mechanisms to ensure the correct suppression of any deviations from the nominal values of the system and to enforce generalization to unseen scenarios. Nevertheless, these mechanisms should rely on the same framework elements and techniques, like modifying the rewards modifying the approximation networks, or the learning algorithm itself. More research is needed to address these challenges optimize the algorithm for real-time applications and adapt the different techniques used in other applications to power system control.

7 Future Work

Addressing the specific limitations of the learning algorithms to ensure hard enforcement of the desired behavior and requirements for reinforcement learning algorithms seems to be a complex task. Their inherent stochastic characteristics are opposed to the deterministic behaviors a controller is expected to have for real-world applications. Acknowledging this, there is still an open field for investigation on the mechanisms available, and the different strategies, that can be developed as solutions that can allow these algorithms to be employed in real-world scenarios with confidence that they meet specific control requirements. This document can be a starting point for PPO algorithms.

Additionally, all the advantages mentioned about these algorithms could allow to reduce the overhead of control designs, by not having to worry about all the control problems that conventional controllers face, e.g. delays or noise. However, more studies on how good this algorithm can do in addressing the problems are needed. Despite the confidence, reinforcement learning algorithms employing neural network models suffer from the same problems as other solutions using neural networks. They behave as black boxes, with the necessity of large quantities of data and computational resources to be trained on. Knowing their limitations becomes crucial to having reasonable expectations of their performance once deployed in the real world. This is still under investigation and is pending for future work.

A System Parameters

Table **A-1**: Bus Parameters

Bus	Vn	vmax	vmin	v0	a0	area
1	20	1.1	0.9	1	0.570254917	1
2	20	1.1	0.9	0.99761	0.368746183	1
3	20	1.1	0.9	0.96263	0.185317315	2
4	20	1.1	0.9	0.81691	0.462358663	2
5	230	1.1	0.9	0.97928	0.480202909	1
6	230	1.1	0.9	0.95796	0.283886529	1
7	230	1.1	0.9	0.9362	0.126901145	1
8	230	1.1	0.9	0.87904	-0.080592324	2
9	230	1.1	0.9	0.89054	0.093617716	2
10	230	1.1	0.9	0.82958	0.336600709	2

Table **A-2**: Transmission Line Parameters

name	From	To	Sn (MVA)	fn (Hz)	Vn1 (kV)	Vn2 (kV)	r	x	b
Line_0	5	6	100	60	230	230	0.005	0.05	0.075
Line_1	5	6	100	60	230	230	0.00501	0.05001	0.075
Line_2	6	7	100	60	230	230	0.002	0.02	0.03
Line_3	6	7	100	60	230	230	0.00201	0.02001	0.03
Line_4	7	8	100	60	230	230	0.02201	0.22001	0.33
Line_5	7	8	100	60	230	230	0.02202	0.22002	0.33
Line_6	7	8	100	60	230	230	0.022	0.22	0.33
Line_7	8	10	100	60	230	230	0.002	0.02	0.03
Line_8	8	10	100	60	230	230	0.00201	0.02001	0.03
Line_9	9	10	100	60	230	230	0.005	0.05	0.075
Line_10	9	10	100	60	230	230	0.00501	0.05001	0.075
Line_11	1	5	100	60	20	230	0.001	0.012	0
Line_12	2	6	100	60	20	230	0.001	0.012	0
Line_13	3	9	100	60	20	230	0.001	0.012	0
Line_14	4	10	100	60	20	230	0.001	0.012	0

Table **A-3**: Load Parameters

name	PQ_0	PQ_1
bus	7	8
Vn (kV)	230	230
p0 (MW)	11.59	15.75
q0 (MVAr)	-0.735	-0.899
vmax	1.1	1.1
vmin	0.9	0.9

Table **A-4**: Synchronous Generator Parameters

name	G1 - Slack	G2	G2	G3
Sn (MVA)	900	900	900	900
Vn (kV)	20	20	20	20
bus	1	2	3	4
p0 (MW)	7.45861	7	7	7
q0 (MVAr)	1.43612	3	5.5	-1
pmax (MW)	9	9	9	20
pmin (MW)	0	0	0	0
qmax (MVAr)	6	3	5.5	-1
qmin (MVAr)	0	3	5.5	-1
v0	1	1	1	1
vmax	1.4	1.4	1.4	1.4
vmin	0.6	0.6	0.6	0.6
ra	0	0	0	0
xs	0.25	0.25	0.25	0.25

Table **A-5**: Generator Model Parameters

name	GENROU_1	GENROU_2	GENROU_3
bus	1	2	3
gen	1	2	3
Sn (MVA)	900	900	900
Vn (kV)	20	20	20
fn (Hz)	60	60	60
D	0	0	0
M	13	13	12.35
ra	0	0	0
xl	0.06	0.06	0.06
xd1	0.3	0.3	0.3
kp	0	0	0
kw	0	0	0
S10	0	0	0
S12	1	1	1
xd	1.8	1.8	1.8
xq	1.7	1.7	1.7
xd2	0.25	0.25	0.25
xq1	0.55	0.55	0.55
xq2	0.25	0.25	0.25
Td10 (s)	8	8	8
Td20 (s)	0.03	0.03	0.03
Tq10 (s)	0.4	0.4	0.4
Tq20 (s)	0.05	0.05	0.05

Table **A-6**: Turbine Governor Model Parameters

name	TGOV1_1	TGOV1_2	TGOV1_3
syn	1	2	3
Tn (s)	900	900	900
wref0	1	1	1
R	0.05	0.05	0.05
VMAX	33	33	33
VMIN	0.4	0.4	0.4
T1 (s)	0.49	0.49	0.49
T2 (s)	2.1	2.1	2.1
T3 (s)	7	7	7
Dt	0	0	0

Table **A-7**: Excitation System Model Parameters

name	EXDC2_1	EXDC2_2	EXDC2_3
syn	1	2	3
TR (s)	0.02	0.02	0.02
TA (s)	0.02	0.02	0.02
TC	1	1	1
TB	1	1	1
TE	0.83	0.83	0.83
TF1	1.246	1.246	1.246
KF1	0.0754	0.0754	0.0754
KA	20	20	20
KE	1	1	1
VRMAX	5.2	5.2	5.2
VRMIN	-4.16	-4.16	-4.16
E1	0	0	0
SE1	0	0	0
E2	1	1	1
SE2	1	1	1

Table **A-8**: REECA1 Parameters

Parameter	Value
PFFLAG	0
VFLAG	0
QFLAG	0
PFLAG	0
PQFLAG	0
Vdip	0.8
Vup	1.1
Trv	0.02
dbd1	-0.02
dbd2	0.02
Kqv	20
Iqh1	999
Iql1	-999
Vref0	1
Iqfrz	0
Thld	-2
Thld2	1
Tp	0.02
QMax	999
QMin	-999
VMAX	999
VMIN	-999
Kqp	1
Kqi	0.1
Kvp	1
Kvi	0.1
Vref1	1
Tiq	0.02
dPmax	999
dPmin	-999
PMAX	999
PMIN	-999
Imax	10
Tpord	0.02
Vq1	0.2
Iq1	2
Vq2	0.4
Iq2	4
Vq3	0.8
Iq3	8
Vq4	1
Iq4	10
Vp1	0.2
Ip1	2
Vp2	0.4
Ip2	4
Vp3	0.8
Ip3	8
Vp4	1
Ip4	10

Table **A-9**: REGCA1 Parameters: Part 1

Parameter	Value
name	REGCA1_1
bus	4
gen	4
Tg	0.1
Rrpwr	999
Brkpt	0.8
Zerox	0.5
Lvpl1	1
Volim	1.2
Lvpnt1	1
Lvpnt0	0.4
Iolim	0
Tfltr	0.1
Khv	0.7
Iqrmax	999
Iqrmin	-999
Accel	0
Iqmax	999
Iqmin	-999
ra	0
xs	0.25

Table **A-10**: REPCA1 Parameters: Part 2

Parameter	Value
name	REPCA1_1
ree	1
line	Line_14
busr	
busf	BusFreq_4
VCFlag	1
RefFlag	1
Fflag	0
Tfltr	0.02
Kp	1
Ki	0.1
Tft	1
Tfv	1
Vfrz	0.8
Rc	
Xc	
Kc	1
emax	999
emin	-999
dbd1	-0.02
dbd2	0.02
Qmax	999
Qmin	-999
Kpg	1
Kig	0.1
Tp	0.02
fdbd1	-0.01
fdbd2	0.01
femax	0.05
femin	-0.05
Pmax	999
Pmin	-999
Tg	0.02
Ddn	10
Dup	10

Table **A-11**: WTDTA1 Parameters

Parameter	Value
name	WTDTA1_1
ree	1
Sn	1000
fn	60
Ht	3
Hg	3
Dshaft	1
Kshaft	1

Table **A-12**: WTARA1 Parameters

Parameter	Value
name	WTARA1_1
rego	1
Ka	1
theta0	0

Table **A-13**: WTGPTA1 Parameters

Parameter	Value
name	WTGPTA1_1
rea	1
Kiw	0.1
Kpw	0
Kic	0.1
Kpc	0
Kcc	0
Tp	0.3
thmax	30
thmin	0
dthmax	5
dthmin	-5

Table **A-14**: WTGTRQA1 Parameters

Parameter	Value
name	WTGTRQA1_1
rep	1
Kip	0.1
Kpp	0
Tp	0.05
Twref	30
Temax	1.2
Temin	0
Tflag	0
p1	0.2
sp1	0.58
p2	0.4
sp2	0.72
p3	0.6
sp3	0.86
p4	0.8
sp4	1

Bibliography

[1] Grammarly, ‘‘Grammarly handbook,’’ 2024. Accessed on July 09, 2024.

[2] G. A. United Nations, ‘‘Transforming our world: the 2030 Agenda for Sustainable Development,’’ 2015.

[3] BP, ‘‘Statistical Review of World Energy,’’ *BP Energy Outlook 2021*, vol. 70, pp. 8--20, 2021.

[4] Melillo et al., ‘‘Climate Change Impacts in the United States: The Third National Climate Assessment,’’ tech. rep., 2014.

[5] S. Curtis, A. Fair, J. Wistow, D. V. Val, and K. Oven, ‘‘Impact of extreme weather events and climate change for health and social care systems,’’ *Environmental Health*, vol. 16, p. 128, Dec 2017.

[6] IEA, ‘‘World Energy Outlook 2021,’’ 2021.

[7] F. Milano, *Frequency Variations in Power Systems.* John Wiley and Sons, Ltd, 2020.

[8] H. Wang and M. Redfern, ‘‘The advantages and disadvantages of using hvdc to interconnect ac networks,’’ in *45th International Universities Power Engineering Conference UPEC2010*, pp. 1--5, 2010.

[9] M. Shahidehpour and Y. Wang, *Communication and Control in Electric Power Systems: Applications of Parallel and Distributed Processing.* Wiley-IEEE Press, 2003.

[10] H. Bevrani, *Robust Power System Frequency Control Second.* 2 ed., 2014.

[11] F. Report, ‘‘First Report of Power System Stability,’’ *Transactions of the American Institute of Electrical Engineers*, vol. 56, no. 2, pp. 261--282, 1937.

[12] Y. Liu, S. You, J. Tan, Y. Zhang, and Y. Liu, ‘‘Frequency response assessment and enhancement of the u.s. power grids toward extra-high photovoltaic generation penetrations—an industry perspective,’’ *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3438--3449, 2018.

[13] P. Kundur, *Power System Stability and Control.* 2009.

[14] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control.* John Wiley and Sons, Ltd, 2 ed., 2008.

[15] E. Benham, ‘‘Si measurement system chart,’’ 2017-08-24 2017.

[16] National Fire Protection Association, *NFPA 70: National Electrical Code.* Quincy, MA, USA: NFPA, 2023 ed., 2023.

[17] British Standards Institution, *Requirements for Electrical Installations: IET Wiring Regulations - 18th Edition.* London, UK: IET, 2018.

[18] U. Governmen, ‘‘The electricity safety, quality and continuity regulations 2002 part vii regulation 27,’’ 2002.

[19] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, E. T. McGunigal, J. A. Mullen, W. L. Smith, R. L. Sydnor, R. F. Vessot, and G. M. Winkler, ‘‘Characterization of Frequency Stability,’’ *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, no. 2, pp. 105--120, 1971.

[20] Eto et al., ‘‘Use of Frequency Response Metrics to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation,’’ no. December 2010, pp. LBNL--4142E, 2010.

[21] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control.* Cambridge University Press, 2019.

[22] N. Cohn, ‘‘Some aspects of tie-line bias control on interconnected power systems [includes discussion],’’ *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 75, no. 3, pp. 1415--1436, 1956.

[23] N. Cohn, ‘‘Methods of controlling generation on interconnected power systems,’’ *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 80, no. 3, pp. 270--279, 1961.

[24] N. Cohn, ‘‘Decomposition of time deviation and inadvertent interchange on interconnected systems, part i: Identification, separation and measurement of components,’’ *IEEE Power Engineering Review*, vol. PER-2, no. 5, pp. 37--37, 1982.

[25] IEEE, ‘‘Ieee standard definitions of terms for automatic generation control on electric power systems,’’ *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 6, pp. 1356--1364, 1970.

[26] P. Anderson, A. Fouad, I. of Electrical, E. Engineers, and I. P. E. Society, *Power System Control and Stability*. IEEE Press power engineering series Power system control and stability, Iowa State University Press, 1977.

[27] F. Milano, *Power System Modelling and Scripting*. Power Systems, Springer Berlin Heidelberg, 2010.

[28] A. Pappachen and A. Peer Fathima, ‘‘Critical research areas on load frequency control issues in a deregulated power system: A state-of-the-art-of-review,’’ 2017.

[29] M. V. RAO, ‘‘Load frequency control,’’ 2024. Dept. of EEE, JNTUA College of Engineering, Kalikiri, Chittoor District, A P, India.

[30] R. Asghar, F. Riganti Fulginei, H. Wadood, and S. Saeed, ‘‘A review of load frequency control schemes deployed for wind-integrated power systems,’’ *Sustainability*, vol. 15, no. 10, 2023.

[31] J. C. Basilio and S. R. Matos, ‘‘Design of PI and PID controllers with transient performance specification,’’ *IEEE Transactions on Education*, vol. 45, no. 4, pp. 364--370, 2002.

[32] S. Saxena and Y. V. Hote, ‘‘Stabilization of perturbed system via IMC: An application to load frequency control,’’ *Control Engineering Practice*, vol. 64, no. January, pp. 61--73, 2017.

[33] W. Tan, ‘‘Unified tuning of PID load frequency controller for power systems via IMC,’’ *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 341--350, 2010.

[34] P. Bhanu, C. Bhushan, K. Sujatha, M. Venmathi, and A. Nalini, ‘‘Load frequency controller with PI controller considering non-linearities and boiler dynamics,’’ *IET Conference Publications*, vol. 2011, no. 583 CP, pp. 290--294, 2011.

[35] N. C. Patel, D. Manoj Kumar, S. Binod Kumar, and D. Pranati, ‘‘Solution of LFC Problem using PD+PI Double Loop Controller Tuned by SCA,’’ pp. 337--342, 2018.

[36] B. H. Bakken and O. S. Grande, ‘‘Automatic Generation Control in a Deregulated Power System,’’ vol. 62, no. 4, pp. 294--298, 1998.

[37] N. Kumari and A. N. Jha, ‘‘Particle swarm optimization and gradient descent methods for optimization of PI Controller for AGC of multi-area thermal-wind-hydro power plants,’’ *Proceedings - UKSim 15th International Conference on Computer Modelling and Simulation, UKSim 2013*, pp. 536--541, 2013.

[38] S. Kumar and M. N. Anwar, ‘‘Fractional order PID Controller design for Load Frequency Control in Parallel Control Structure,’’ *2019 54th International Universities Power Engineering Conference, UPEC 2019 - Proceedings*, pp. 17--22, 2019.

[39] M. N. Anwar and S. Pan, ‘‘A new PID load frequency controller design method in frequency domain through direct synthesis approach,’’ *International Journal of Electrical Power and Energy Systems*, vol. 67, pp. 560--569, 2015.

[40] I. K. Otchere, K. A. Kyeremeh, and E. A. Frimpong, ‘‘Adaptive pi-ga based technique for automatic generation control with renewable energy integration,’’ in *2020 IEEE PES/IAS PowerAfrica*, pp. 1--4, 2020.

[41] H. Bevrani, Y. Mitani, and K. Tsuji, ‘‘Robust decentralized LFC design in a deregulated environment,’’ *Proceedings of the 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004)*, vol. 1, pp. 326--331, 2004.

[42] K. Niimi, K. Yukita, T. Matsumura, and Y. Goto, ‘‘Verification of Load Frequency Control Using H-infinity Control,’’ *7th International IEEE Conference on Renewable Energy Research and Applications, ICRERA 2018*, vol. 5, pp. 1174--1178, 2018.

[43] H. Shayeghi, H. A. Shayanfar, and O. P. Malik, ‘‘Robust decentralized neural networks based LFC in a deregulated power system,’’ *Electric Power Systems Research*, vol. 77, no. 3-4, pp. 241--251, 2007.

[44] K. V. Kumar, T. A. Kumar, and V. Ganesh, ‘‘Chattering free sliding mode controller for load frequency control of multi area power system in deregulated environment,’’ in *2016 IEEE 7th Power India International Conference (PIICON)*, pp. 1--6, 2016.

[45] X. Z. Yu and F. Y. Hai, ‘‘A new fuzzy PI control algorithm for marine electric governor system,’’ *ISCID 2009 - 2009 International Symposium on Computational Intelligence and Design*, vol. 1, pp. 276--279, 2009.

[46] V. K. Singh and R. Dahiya, ‘‘Automatic generation control system using pi and fis controller.,’’ in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1--4, 2018.

[47] M. Chandrashekar and R. Jayapal, ‘‘Design and comparison of i, pi, pid and fuzzy logic controller on agc deregulated power system with hvdc link,’’ in *2013 International conference on Circuits, Controls and Communications (CCUBE)*, pp. 1--6, 2013.

[48] N. El Yakine Kouba, M. Menaa, M. Hasni, K. Tehrani, and M. Boudour, ‘‘A novel optimized fuzzy-pid controller in two-area power system with hvdc link connection,’’ in *2016 International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 204--209, 2016.

[49] A. Demiroren, H. Zeynelgil, and N. Sengor, ‘‘The application of ann technique to load-frequency control for three-area power system,’’ in *2001 IEEE Porto Power Tech Proceedings (Cat. No.01EX502)*, vol. 2, pp. 5 pp. vol.2--, 2001.

[50] R. Chaudhary and A. P. Singh, ‘‘Intelligent load frequency control approach for multi area interconnected hybrid power system,’’ in *2017 International Conference on Technological Advancements in Power and Energy (TAP Energy)*, pp. 1--4, 2017.

[51] S. Baghya Shree and N. Kamaraj, ‘‘Hybrid Neuro Fuzzy approach for automatic generation control in restructured power system,’’ *International Journal of Electrical Power and Energy Systems*, vol. 74, pp. 274--285, 2016.

[52] A. Prakash and S. K. Parida, ‘‘LQR based PI controller for load frequency control with distributed generations,’’ *2020 21st National Power Systems Conference, NPSC 2020*, pp. 2--6, 2020.

[53] S. K. Pandey, P. Gupta, and S. S. Dwivedi, ‘‘Full order observer based load frequency control of Single Area Power System,’’ *Proceedings - 2020 12th International Conference on Computational Intelligence and Communication Networks, CICN 2020*, no. 3, pp. 239--242, 2020.

[54] A. Panwar, V. Agarwal, and G. Sharma, ‘‘Studies on Frequency Regulation of Hydro System via New JAYA Optimized LQR Design,’’ *icABCD 2021 - 4th International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems, Proceedings*, pp. 2--5, 2021.

[55] A. Ali, B. Khan, C. A. Mehmood, Z. Ullah, S. M. Ali, and R. Ullah, ‘‘Decentralized mpc based frequency control for smart grid,’’ in *2017 International Conference on Energy Conservation and Efficiency (ICECE)*, pp. 1--6, 2017.

[56] I. E. Uyioghosa, ‘‘A Comparative Analysis of Different MPC Controllers for Load Frequency Control for Interconnected Power System,’’ no. 5, pp. 1--6, 2020.

[57] V. Kumtepeli, Y. Wang, and A. Tripathi, ‘‘Multi-Area Model Predictive Load Frequency Control : A Decentralized Approach,’’ no. October, pp. 25--27, 2016.

[58] F. Liu, Y. Li, S. Member, Y. Cao, and S. Member, ‘‘A Two-Layer Active Disturbance Rejection Controller Design for Load Frequency Control of Interconnected Power System,’’ vol. 31, no. 4, pp. 3320--3321, 2016.

[59] M. Yang, C. Wang, Z. Liu, and S. He, ‘‘An EID Load Frequency Control Method for Two-Area Interconnected Power System with Photovoltaic Generation,’’ pp. 5662--5666, 2021.

[60] C. Wang and J. Li, ‘‘Frequency Control of Isolated Wind-Diesel Microgrid Power System by Double Equivalent-Input-Disturbance Controllers,’’ 2019.

[61] F. Liu, Z. Xu, L. Liu, F. Yang, and D. Sidorov, ‘‘A Robust Active Disturbance Rejection Controller Design for LFC in Two-area Power System,’’ *Chinese Control Conference, CCC*, vol. 2018-July, pp. 8858--8863, 2018.

[62] Y. Du, W. Cao, J. She, and M. Fang, ‘‘A comparison study of three active disturbance rejection methods,’’ in *2020 39th Chinese Control Conference (CCC)*, pp. 135--139, 2020.

[63] R. L. Rekhasree and J. A. Jaleel, ‘‘A comparison of agc of power systems using reinforcement learning and genetic algorithm with a case study,’’ in *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, pp. 34--38, 2014.

[64] S. Eftekharnejad and A. Feliachi, ‘‘Stability enhancement through reinforcement learning: Load frequency control case study,’’ in *2007 iREP Symposium - Bulk Power System Dynamics and Control - VII. Revitalizing Operational Reliability*, pp. 1--8, 2007.

[65] Z. Yan and Y. Xu, ‘‘A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system,’’ *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4599--4608, 2020.

[66] V. A. K. Pappu, B. Chowdhury, and R. Bhatt, ‘‘Implementing frequency regulation capability in a solar photovoltaic power plant,’’ in *North American Power Symposium 2010*, pp. 1--6, 2010.

[67] Y.-z. Sun, Z.-s. Zhang, G.-j. Li, and J. Lin, ‘‘Review on frequency control of power systems with wind power penetration,’’ in *2010 International Conference on Power System Technology*, pp. 1--8, 2010.

[68] S. A. Dorado-rojas, ‘‘Decentralized Load Frequency Control for a Power System with High Penetration of Wind and Solar Photovoltaic Generation,’’ 2020.

[69] Z. C. Lipton, J. Berkowitz, and C. Elkan, ‘‘A critical review of recurrent neural networks for sequence learning,’’ *arXiv preprint arXiv:1506.00019*, 2015.

[70] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ‘‘Learning representations by back-propagating errors,’’ *Nature*, vol. 323, no. 6088, pp. 533--536, 1986.

[71] D. P. Kingma and J. Ba, ‘‘Adam: A method for stochastic optimization,’’ *arXiv preprint arXiv:1412.6980*, 2014.

[72] P. Werbos, ‘‘Backpropagation through time: what it does and how to do it,’’ *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550--1560, 1990.

[73] S. Hochreiter and J. Schmidhuber, ‘‘Long short-term memory,’’ *Neural computation*, vol. 9, no. 8, pp. 1735--1780, 1997.

[74] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, ‘‘Deepmimic: Example-guided deep reinforcement learning of physics-based character skills,’’ *ACM Trans. Graph.*, vol. 37, pp. 143:1--143:14, July 2018.

[75] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[76] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, ‘‘Proximal policy optimization algorithms,’’ 2017.

[77] WECC Modeling and Validation Work Group, *WECC Wind Power Plant Dynamic Modeling Guide.* Western Electricity Coordinating Council, November 2010. `https://transmission.bpa.gov/Business/Operations/GridModeling/WECCWindPlantDynamicModelingGuide.pdf`.

[78] WECC Modeling and Validation Work Group, *WECC PV Power Plant Dynamic Modeling Guide.* Western Electricity Coordinating Council, April 2014. `https://www.wecc.org/Reliability/WECC%20Solar%20Plant%20Dynamic%20Modeling%20Guidelines.pdf`.

[79] WECC, *WECC Generator Unit Model Validation Guideline*. Western Electricity Coordinating Council, April 23 2020. https://www.wecc.org/Reliability/WECC%20Generator%20Unit%20Model%20Validation%20Guideline.pdf.

[80] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, ‘‘Stable-baselines3: Reliable reinforcement learning implementations,’’ *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1--8, 2021.

[81] O. Dogru, K. Velswamy, F. Ibrahim, Y. Wu, A. S. Sundaramoorthy, B. Huang, S. Xu, M. Nixon, and N. Bell, ‘‘Reinforcement learning approach to autonomous pid tuning,’’ *Computers Chemical Engineering*, vol. 161, p. 107760, 2022.

[82] W. Cui, Y. Jiang, and B. Zhang, ‘‘Reinforcement learning for optimal primary frequency control: A lyapunov approach,’’ 2021.

[83] C. of United States of America, ‘‘Energy Policy Act of 2005,’’ 2005.

[84] A. Demiroren and E. Yesil, ‘‘Automatic generation control with fuzzy logic controllers in the power system including smes units,’’ *International Journal of Electrical Power and Energy Systems*, vol. 26, no. 4, pp. 291--305, 2004. 2002 Conference on Probabilistic Methods Applied to Power Systems.

[85] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, ‘‘Proximal policy optimization algorithms,’’ *arXiv preprint arXiv:1707.06347*, 2017.

[86] Power World Corporation, *Transient Models in PowerWorld Simulator*. Power World Corporation, year. Accessed: January 30, 2024.