



UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO

DOCTORADO EN ESTADÍSTICA E INFORMÁTICA



TESIS

MODELAMIENTO DEL COMPORTAMIENTO Y RELACIONES DE PARÁMETROS FÍSICO-QUÍMICOS DE LA BAHÍA INTERIOR DEL LAGO TITICACA-PUNO

PRESENTADA POR:

EDGAR ELOY CARPIO VARGAS

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTOR EN ESTADÍSTICA E INFORMÁTICA

PUNO, PERÚ

2020



UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO

DOCTORADO EN ESTADÍSTICA E INFORMÁTICA

TESIS

MODELAMIENTO DEL COMPORTAMIENTO Y RELACIONES DE PARÁMETROS FÍSICO-QUÍMICOS DE LA BAHÍA INTERIOR DEL LAGO TITICACA-PUNO

PRESENTADA POR:

EDGAR ELOY CARPIO VARGAS

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTOR EN ESTADÍSTICA E INFORMÁTICA

APROBADA POR EL SIGUIENTE JURADO:

PRESIDENTE


.....
Dr. FÉLIX OLAGUIEL LOZA

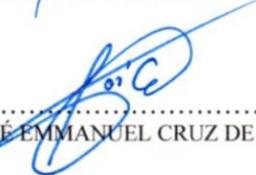
PRIMER MIEMBRO


.....
D.Sc. PERCY HUATA PANCA

SEGUNDO MIEMBRO


.....
Dr. JOSÉ PANFILO TITO LIPA

ASESOR DE TESIS


.....
Dr. JOSÉ EMMANUEL CRUZ DE LA CRUZ

Puno, 17 de enero de 2020

ÁREA: Estadística e Informática.

TEMA: Modelamiento del comportamiento de los parámetros físico químicos de la bahía interior del lago Titicaca.

LÍNEA: Modelos predictivos uni y multivariantes.



DEDICATORIA

El presente trabajo de investigación va dedicado a mis padres Teófilo y en memoria de mi madre Bernardina por haberme guiado en el sendero del bien y la ética.

A mis hermanos: Lupe, Alicia, David, Cesar, Edith y Mery por ser personas de bien y a sus queridos hijos.

A mi esposa Miriam y mis lindos hijos Briguitte Danae, Angela almendra y Ahmed Eloytzvi por su paciencia y comprensión por el tiempo que demando realizar este trabajo de investigación y dedicarlo a ellos.



AGRADECIMIENTOS

- A la escuela de posgrado de Universidad Nacional del Altiplano y a mi querida Escuela Profesional de Ingeniería Estadística e Informática.
- A mi jurados y asesor de tesis, por las orientaciones realizadas para mejorar el presente trabajo de investigación.
- Al IMARPE por las facilidades para realizar el presente trabajo de investigación.



ÍNDICE GENERAL

	Pág.
DEDICATORIA	i
AGRADECIMIENTOS	ii
ÍNDICE GENERAL	iii
ÍNDICE DE TABLAS	viii
ÍNDICE DE FIGURAS	xi
ÍNDICE DE ANEXOS	xvii
RESUMEN	xviii
ABSTRACT	xix
INTRODUCCIÓN	1

CAPÍTULO I

REVISION DE LITERATURA

1.1	Marco teórico	2
1.1.1	Caracterización del lago Titicaca y su cuenca.	2
1.1.1.1	Dimensiones y encuadre administrativo y geográfico de la cuenca.	2
1.1.2	Temperatura superficial del agua.	3
1.1.2.1	Análisis o sustento de propuesta nacional	3
1.1.3	Solidos suspendidos totales.	4
1.1.3.1	Análisis o propuesta nacional	4
1.1.4	Clorofila-a.	4
1.1.5	Demanda bioquímica de oxígeno	5
1.1.5.1	Análisis o propuesta nacional	5
1.1.6	Demanda química de oxígeno	6
1.1.6.1	Análisis o propuesta nacional	6
1.1.7	Oxígeno disuelto	7
1.1.7.1	Análisis o propuesta nacional	7
		iii



1.1.8	pH	8
1.1.8.1	Análisis o sustento de propuesta nacional	8
1.1.9	Conductividad Eléctrica	8
1.1.9.1	Análisis o propuesta nacional	9
1.1.10	Coliformes totales (microbiológico)	9
1.1.10.1	Análisis o sustento de propuesta nacional	9
1.1.11	Coliformes Termotolerantes (Microbiológico)	10
1.1.11.1	Análisis o sustento de propuesta nacional	10
1.1.12	Nitrito	10
1.1.12.1	Análisis o sustento de propuesta nacional.	10
1.1.13	Nitrato	11
1.1.13.1	Análisis o sustento de propuesta nacional	11
1.1.14	Fosfato	11
1.1.14.1	Análisis o propuesta nacional	12
1.1.15	Nitrógeno	12
1.1.15.1	Análisis o sustento de propuesta nacional	12
1.1.16	Concentración de fosforo	13
1.1.17	Nitrógeno amoniacal	13
1.1.17.1	Análisis o sustento de propuesta nacional	14
1.1.18	Transparencia	14
1.1.19	Modelos estructurales	14
1.1.19.1	Revisión metodológica de los modelos de ecuaciones estructurales.	14
1.1.19.2	Tipos de variables en modelos de ecuaciones estructurales.	15
1.1.19.3	Elementos de un modelo de ecuaciones estructurales.	16
1.1.19.4	Los diagramas estructurales. “Path-diagrams”	16
1.1.19.5	Relaciones entre variables	18
1.1.19.6	Construcción del modelo estructural	18
1.1.19.7	Bondad de ajuste de los modelos de ecuaciones estructurales	27
1.1.19.8	Ajuste del modelo estructural	27
1.1.19.9	Medidas de ajuste global	28



1.1.19.10	Medidas absolutas de ajuste	28
1.1.19.11	Medidas incrementales de ajuste	30
1.1.19.12	Medidas de ajuste de parsimonia	30
1.2.23	Análisis Factorial	30
1.2.23.1	Pasos para realizar un análisis factorial	32
1.2.24	Árboles de predicción	35
1.2.24.1	Árboles de predicción en R	36
1.2.24.2	Evitar el overfitting	36
1.2.24.3	Controlar el tamaño del árbol	36
1.2.24.4	Tree pruning	37
1.2.24.5	Ajuste del modelo	38
1.2.24.6	Podado del árbol (pruning)	38
1.2.25	Redes Neuronales	39
1.2.25.1	Neurona bilógica	39
1.2.25.2	Modelo de neurona artificial	40
1.2.25.3	Red neuronal artificial	42
1.2.25.4	Arquitectura de las RNA	42
1.2.25.5	Aprendizaje de las RNA	43
1.2.25.6	Perceptrón Multicapa	44
1.2.25.7	Problema del OR Exclusivo	44
1.2.25.8	Arquitectura	45
1.2.25.9	Entrenamiento	45
1.2.26	Inteligencia artificial	47
1.2	Antecedentes	47

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1	Identificación del problema.	55
2.2	Enunciado del problema	56
2.3	Justificación	56
2.4	Objetivos	57



2.4.1	Objetivo General	57
2.4.2	Objetivos Específicos	57
2.5	Hipótesis	57
2.5.1	Hipótesis General	57
2.5.2	Hipótesis específicas	57

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1	Lugar de estudio	59
3.2	Población	59
3.3	Muestra	59
3.4	Método de investigación	60
3.5	Descripción detallada de métodos por objetivos específicos	60

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1	Muestra y descripción de la base de datos	62
4.2	Nomenclatura y variables en estudio: Parámetros físico químicos:	62
4.3	Análisis descriptivo y tendencia del parámetro temperatura	63
4.4	Análisis descriptivo y tendencia del parámetro Coliformes Totales	66
4.5	Análisis descriptivo y tendencia del parámetro Oxígeno Disuelto	69
4.6	Análisis descriptivo y tendencia del parámetro Coliformes termotolerantes	72
4.7	Análisis descriptivo y tendencia del parámetro pH.	75
4.8	Análisis descriptivo y tendencia del parámetro Nitrito	77
4.9	Análisis descriptivo y tendencia del parámetro conectividad eléctrica	80
4.10	Análisis descriptivo y tendencia del parámetro total de sólidos en suspensión	83
4.11	Análisis descriptivo y tendencia del parámetro Clorofila-a	85
4.12	Análisis descriptivo y tendencia de la Demanda bioquímica de oxígeno	88



4.13	Análisis descriptivo y tendencia del parámetro Demanda química de oxígeno	91
4.14	Análisis descriptivo y tendencia del parámetro Nitrato	93
4.15	Análisis descriptivo y tendencia del parámetro fosfato	96
4.16	Análisis descriptivo y tendencia del parámetro Nitrógeno	99
4.17	Análisis descriptivo y tendencia del parámetro Fosforo	102
4.18	Análisis descriptivo y tendencia del parámetro Nitrógeno amoniacal	104
4.19	Análisis descriptivo y tendencia del parámetro Transparencia	107
4.20	Análisis del principal supuesto básico (normalidad)	110
4.21	Prueba de normalidad multivariante:	111
4.22	Modelamiento con análisis clúster.	115
4.23	Modelamiento mediante árboles de predicción:	116
4.24	Análisis Exploratorio de datos, formación de variables latentes	129
	CONCLUSIONES	149
	RECOMENDACIONES	151
	BIBLIOGRAFÍA	152
	ANEXOS	158



ÍNDICE DE TABLAS

	Pág.
1. Afluentes principales del lago Titicaca, Área de cuenca, caudal promedio y mínimo	3
2. Operacionalización de variables de los parámetros físico químicos del lago Titicaca.	60
3. Estadísticos descriptivos para la temperatura por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	64
4. Estadísticos descriptivos para coliformes totales por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	67
5. Estadísticos descriptivos para oxígeno disuelto por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	70
6. Estadísticos descriptivos para coliformes termo tolerantes por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	73
7. Estadísticos descriptivos para pH por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	76
8. Estadísticos descriptivos para Nitrito por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	78
9. Estadísticos descriptivos para conectividad eléctrica por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	81
10. Estadísticos descriptivos para la variable total de sólidos en suspensión por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	84
11. Estadísticos descriptivos para clorofila-a por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	86



12. Estadísticos descriptivos para la demanda bioquímica de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	89
13. Estadísticos descriptivos para la variable demanda química de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	92
14. Estadísticos descriptivos para la concentración de nitrato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	95
15. Estadísticos descriptivos para fosfato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	98
16. Estadísticos descriptivos para nitrógeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	100
17. Estadísticos descriptivos para fosforo por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	103
18. Estadísticos descriptivos para nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	105
19. Estadísticos descriptivos para la variable nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	108
20. Prueba de normalidad multivariantes Mardia-tests y Royston.	111
21. Pruebas de normalidad univariantes	112
22. correlaciones de las variables observadas.	113
23. Resultados de la Prueba de KMO y Bartlett	129
24. Varianza total explicada de las variables observadas	130
25. Formación de componentes con las variables observadas	131
26. Resultados de la Prueba de KMO y Bartlett	132
27. Varianza total explicada de las variables observadas independientes	133



28. Formación de componentes con las variables observadas independientes	134
29. Fiabilidad de la medida.	139
30. Pesos de regresión estandarizados	142
31. Ajuste del modelo	143
32. Valores significativos	145
33. Las ecuaciones estructurales quedan definidas como:	147



ÍNDICE DE FIGURAS

	Pág.
1. Convenciones “path diagrams”	17
2. “Path diagrams” de un sub-modelo de medida y un sub-modelo estructural	17
3. Neurona biológica	39
4. Modelo de neurona artificial	40
5. Modelo de Neuronas Simplificado	41
6. Arquitecturas de RNA	43
7. Valores observados para la temperatura de la bahía interior del lago Titicaca por mes según año, periodo enero 2013 – diciembre 2018.	64
8. Valores promedio de temperatura del lago por meses de la bahía interior del lago Titicaca, periodo enero 2013 – abril 2019.	65
9. Tendencia y valores predichos para el parámetro Temperatura de la bahía interior del lago Titicaca, toma puerto muelle.	65
10. valores observados para coliformes totales de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	67
11. Valores promedio de coliformes totales por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	68
12. Tendencia y valores predichos para el parámetro coliformes totales de la bahía interior del lago Titicaca, Toma Puerto muelle.	68
13. valores observados para oxígeno disuelto de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	69
14. Valores promedio de Oxígeno disuelto por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	70
15. Tendencia y valores predichos para el parámetro oxígeno disuelto de la bahía interior del lago Titicaca, Toma Puerto muelle.	71



16. valores observados para coliformes termo tolerantes de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	72
17. Valores promedio de Coliformes termo tolerantes por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	73
18. Tendencia y valores predichos para el parámetro coliformes termo tolerantes de la bahía interior del lago Titicaca, Toma Puerto muelle.	74
19. valores observados para pH de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	75
20. Valores promedio de pH por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2109.	76
21. Tendencia y valores predichos para el parámetro pH de la bahía interior del lago Titicaca, Toma Puerto muelle.	77
22. valores observados para nitrito de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	78
23. Valores promedio de Nitrito por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	79
24. Tendencia y valores predichos para el parámetro Nitrito de la bahía interior del lago Titicaca, Toma Puerto muelle.	79
25. valores observados para la conectividad eléctrica de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	80
26. Valores promedio de conectividad eléctrica por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019	81
27. Tendencia y valores predichos para el parámetro conectividad eléctrica de la bahía interior del lago Titicaca, Toma Puerto muelle.	82



28. valores observados para el total de sólidos en suspensión de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	83
29. Valores promedio de total de sólidos en suspensión por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019	84
30. Tendencia y valores predichos para el parámetro total de sólidos en suspensión de la bahía interior del lago Titicaca, Toma Puerto muelle.	85
31. valores observados para Clorofila-a de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	86
32. Valores promedio para clorofila-a por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2109.	87
33. Tendencia y valores predichos para el parámetro clorofila-a de la bahía interior del lago Titicaca, Toma Puerto muelle.	87
34. valores observados para la demanda bioquímica de oxígeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	88
35. Valores promedio para demanda bioquímica de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019	89
36. Tendencia y valores predichos para el parámetro demanda bioquímica de oxígeno de la bahía interior del lago Titicaca, Toma Puerto muelle.	90
37. valores observados para la demanda química de oxígeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	91
38. Valores promedio para demanda química de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	92



39. Tendencia y valores predichos para el parámetro demanda química de oxígeno de la bahía interior del lago Titicaca, Toma Puerto muelle. 93
40. valores observados para la Concentración de nitrato de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018. 94
41. Valores promedio para la concentración de nitrato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019. 95
42. Tendencia y valores predichos para el parámetro concentración de nitrato de la bahía interior del lago Titicaca, Toma Puerto muelle. 96
43. valores observados para la variable fosfato de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018. 97
44. Valores promedio para la concentración de fosfato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019. 98
45. Tendencia y valores predichos para el parámetro concentración de fosfato de la bahía interior del lago Titicaca, Toma Puerto muelle. 99
46. valores observados para la variable nitrógeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018. 100
47. Valores promedio para la concentración de nitrógeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019. 101
48. Tendencia y valores predichos para el parámetro concentración de nitrógeno de la bahía interior del lago Titicaca, Toma Puerto muelle. 101
49. valores observados para la variable fosforo de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018. 102



50. Valores promedio para la concentración de fosforo por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	103
51. Tendencia y valores predichos para el parámetro concentración de fosforo de la bahía interior del lago Titicaca, Toma Puerto muelle.	104
52. valores observados para nitrógeno amoniacal de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	105
53. Valores promedio para la concentración de nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	106
54. Tendencia y valores predichos para el parámetro nitrógeno amoniacal de la bahía interior del lago Titicaca, Toma Puerto muelle.	106
55. valores observados para la variable transparencia de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.	108
56. Valores promedio para la transparencia por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.	109
57. Tendencia y valores predichos para el parámetro nitrógeno amoniacal de la bahía interior del lago Titicaca, Toma Puerto muelle.	109
58. Dendograma para los parámetros físico químicos de la bahía interior del lago Titicaca: Toma puerto muelle.	115
59. árbol de predicción para establecer la relación entre la variable temperatura y los demás componentes del modelo.	117
60. árbol de predicción para establecer la relación entre la variable oxígeno disuelto y los demás componentes del modelo.	118
61. árbol de predicción para establecer la relación entre la variable pH y los demás componentes del modelo.	119
62. árbol de predicción para establecer la relación entre la variable conectividad eléctrica y los demás componentes del modelo.	121



63. árbol de predicción para establecer la relación entre la variable solidos totales en suspensión y los demás componentes del modelo.	122
64. árbol de predicción para establecer la relación entre la variable clorofila-a y los demás componentes del modelo.	123
65. árbol de predicción para establecer la relación entre la variable demanda bioquímica de oxígeno y los demás componentes del modelo.	125
66. Árbol de predicción para establecer la relación entre la variable demanda química de oxígeno y los demás componentes del modelo.	126
67. Árbol de predicción para establecer la relación entre la variable transparencia y los demás componentes del modelo.	128
68. Matriz de componentes de la matriz para formar las variables latentes que en este caso serán 3.	132
69. distribución tridimensional de los componentes del modelo factorial	135
70. Modelo de ecuaciones estructurales inicial	136



ÍNDICE DE ANEXOS

	Pág.
1. Anexo 1. Valores pronosticados, usando redes neuronales	158
2. Anexo 2. Covarianzas y correlaciones anti-imagen de las variables observadas dependientes	159
3. Anexo 3. Covarianzas y correlaciones anti-imagen de las variables observadas independientes	160

RESUMEN

El lago Titicaca es una atracción turística, es el más alto del mundo y su salud es importante para la vida de especies como flora, fauna y personas. El objetivo de la presente investigación fue: modelar el comportamiento y las relaciones entre los parámetros físico-químicos de la bahía interior del lago Titicaca. La investigación es cuantitativa de tipo descriptivo explicativo, la muestra estuvo conformada por 17 parámetros físico-químicos cada una con 76 meses tomadas del sector muelle de la bahía interior del lago. Los métodos usados fueron: redes neuronales, árboles de predicción, clúster, factorial y ecuaciones estructurales. Los resultados son: existen 7 parámetros con tendencia estacional y 10 sin patrón estacional. Las relaciones son: La temperatura con coliformes totales, conectividad eléctrica, oxígeno disuelto, transparencia, sólidos totales y nitrito. El oxígeno disuelto con nitrógeno amoniacal, clorofila-a, fósforo, demanda química y pH. El pH con fósforo, nitrógeno, transparencia, coliformes totales, temperatura y fósforo. La conectividad eléctrica con nitrato, fósforo, oxígeno disuelto, coliformes termotolerantes y demanda bioquímica. Los sólidos totales con pH, oxígeno disuelto, demanda bioquímica, nitrógeno amoniacal y clorofila-a. La clorofila-a con pH, oxígeno disuelto, temperatura, conectividad eléctrica, sólidos totales y fósforo. La demanda bioquímica con nitrógeno, nitrógeno amoniacal, oxígeno disuelto, nitrato y temperatura. La demanda química de oxígeno con coliformes termotolerantes, temperatura, conectividad eléctrica, nitrógeno amoniacal y demanda bioquímica. La transparencia con clorofila-a, pH, sólidos totales, nitrato, nitrito y demanda química. Las relaciones estructurales son: $COL_{TOTALES} = 807.521 \times L1I$; $NO3 = 0.354 \times L2I$; $pH = 0.446 \times L4D$; $SST = 5.025 \times L4D$; $Clorofila - a = 11.173 \times L4D$ y $Transp = -0.309 \times L4D$. Como conclusión se indica que las aguas del muelle Puno no son aptas para el consumo humano.

Palabras clave: árboles de predicción, componentes principales, clúster, ecuaciones estructurales, factores latentes, parámetros físico-químicos, redes neuronales.

ABSTRACT

Lake Titicaca is a tourist attraction, it is the highest in the world and its health is important for the life of species such as flora, fauna and people. The objective of the present investigation was: to model the behavior and the relationships between the physical-chemical parameters of the inner bay of Lake Titicaca. The research is quantitative of explanatory descriptive type, the sample was made up of 17 physical-chemical parameters each with 76 months taken from the dock sector of the inner bay of the lake. The methods used were: neural networks, prediction trees, cluster, factorial and structural equations. The results are: there are 7 parameters with seasonal trend and 10 without seasonal pattern. The relationships are: The temperature with total coliforms, electrical conductivity, dissolved oxygen, transparency, total solids and nitrite. Oxygen dissolved with ammoniacal nitrogen, chlorophyll-a, phosphate, chemical demand and pH. The pH with phosphorus, nitrogen, transparency, total coliforms, temperature and phosphate. Electrical conductivity with nitrate, phosphorus, dissolved oxygen, thermo-tolerant coliforms and biochemical demand. Total solids with pH, dissolved oxygen, biochemical demand, ammoniacal nitrogen and chlorophyll-a. Chlorophyll-a with pH, dissolved oxygen, temperature, electrical conductivity, total solids and phosphate. The biochemical demand with nitrogen, ammoniacal nitrogen, dissolved oxygen, nitrate and temperature. The chemical oxygen demand with thermotolerant coliforms, temperature, electrical conductivity, ammoniacal nitrogen and biochemical demand. Transparency with chlorophyll-a, pH, total solids, nitrate, nitrite and chemical demand. The structural relationships are: $COL_TOTALES = 807.521 \times L1I$; $NO3 = 0.354 \times L2I$; $pH = 0.446 \times L4D$; $SST = 5.025 \times L4D$; $Chlorophyll-a = 11,173 \times L4D$ and $Transp = -0,309 \times L4D$. In conclusion it is indicated that the waters of the Puno dock are not suitable for human consumption.

Keywords: prediction trees, main components, cluster, structural equations, latent factors, physical-chemical parameters, neural networks

INTRODUCCIÓN

El agua es uno de los recursos naturales más preciados para la existencia de todo ser vivo, sin él no se garantiza la existencia de plantas, animales y del hombre, muchas de nuestras poblaciones se ven obligados a beber de fuentes cuya calidad deja mucho que desear y un sin fin de enfermedades a niños y adultos. El Lago Titicaca es uno de los lugares más hermosos y misteriosos y está siendo contaminado por afluentes de aguas residuales, residuos de la explotación minera entre otros. El análisis físico-químico proporciona información sobre las propiedades físico-químicas de las aguas y sobre su contenido en sustancias disueltas. Al ser estos parámetros susceptibles de variación en el tiempo, el control de la calidad del agua requiere una analítica sistemática y periódica.

Por su parte los modelos de ecuaciones estructurales establecen la relación de dependencia entre las variables, tratan de integrar una serie de ecuaciones lineales y establecer cuáles de ellas son dependientes o independientes de otras, ya que dentro del mismo modelo las variables que pueden ser independientes en una relación pueden ser dependientes en otras, por lo que se vuelve una herramienta útil para la investigación.

La investigación está organizada por capítulos, es así que en el Capítulo I se ha revisado diferentes conceptos y teorías sobre los parámetros físico químicos y la teoría de los modelos estadísticos utilizados para el análisis.

En el Capítulo II se hace una descripción del contexto en estudio, se explica el por qué se realiza el presente trabajo de investigación, los objetivos y las Hipótesis.

Luego en el Capítulo III se muestra el lugar de estudio, la población, la muestra con la que se trabajó y la metodología empleada en la investigación.

En el Capítulo IV se describe ampliamente los resultados obtenidos, priorizando las técnicas estadísticas aplicadas, luego, se presentan las conclusiones, las recomendaciones y finalmente se hace una exposición de la bibliografía utilizada.



CAPÍTULO I

REVISIÓN DE LITERATURA

1.1 Marco teórico

1.1.1 Caracterización del lago Titicaca y su cuenca.

1.1.1.1 Dimensiones y encuadre administrativo y geográfico de la cuenca.

ALT (2019) publica que en la cuenca endorreica del Altiplano se distinguen dos sistemas hidrológicos activos y separados: el lago Titicaca (3,809.5 m.s.n.m.), que se vierte en el lago Poopó (3,686 m.s.n.m.) por intermedio del río Desaguadero, que a su vez desagua en el salar de Coipasa (3,657 m.s.n.m.) durante los periodos de aguas altas, y el salar de Uyuni (3,653 m.s.n.m.), que recibe el río grande del Lipez. Las posibilidades de comunicación entre estos dos sistemas se presentan para niveles de agua superiores que fueron observados en 1986. La superficie total del lago Titicaca, hasta el comienzo del Desaguadero, es de 56,300 km² (incluidas las aguas superficiales); de esta superficie de la cuenca, las tres cuartas partes están en el Perú (subregión de Puno) y el resto en Bolivia (departamento de La Paz). El área del espejo de agua es de 8,400 km², correspondiendo el 55% a la República del Perú y el 45% a la República de Bolivia (3,790 km²). La superficie de la cuenca de drenaje del Lago es de 47,900 km², o sea, el 85% de la cuenca total, 1/5 de la cual situada en Bolivia y 4/5 en el Perú.

Las tres cuartas partes de la cuenca vertiente del Lago están drenadas por seis ríos: Ramis (31%), Ilave (15%), Coata (11%), Catari (7%), Huancané (7%) y Suchez (6%), cuyas características principales son:

Tabla 1.

Afluentes principales del lago Titicaca, Área de cuenca y caudal promedio

Rio	Área de la cuenca (km ²)	Caudal promedio (m ³ /s)	Caudal Mínimo(m ³ /s) Periodo 1956-1987
Ramis	14.700	76	25
Ilave	7.705	39	10
Coata	4.550	42	11
Catari	2.022	10	--
Huancane	3.540	20	5
Suches	2.825	11	4

Fuente: ALT-perubolivia.org

Sus longitudes están comprendidas entre 120 y 180 km, con excepción del río Ramis, que mide 283 km. Los afluentes secundarios son los ríos: Ilipa, Yanarico, Zapatilla (en Perú); Huayco, Putina, Keka, Khullu Cachi, Sehuenca y Tiahuanacu (en Bolivia).

1.1.2 Temperatura superficial del agua.

DIGESA (2011) indica que la temperatura del agua es un parámetro muy importante dada su influencia, tanto sobre el desarrollo de la vida acuática como sobre las reacciones químicas y velocidades de reacción, así como la aptitud del agua para ciertos usos útiles. La temperatura es un indicador de la calidad del agua, que influye en el comportamiento de otros indicadores de la calidad del recurso hídrico, como el pH, el déficit de oxígeno, la conductividad eléctrica y otras variables.

1.1.2.1 Análisis o sustento de propuesta nacional

A1: La temperatura aceptable para el consumo humanos para una concentración máxima aceptable de 15°C, en temperaturas altas disminuye la concentración de OD, y otras legislaciones consideran la temperatura del agua de la zona con una variación de 3°C.

1.1.3 Sólidos suspendidos totales.

DIGESA (2011) indica que los sólidos en suspensión son productos de la erosión de los suelos, detritus orgánico y plancton. Los sólidos suspendidos, tales como limo, arena y virus, son generalmente responsables de impurezas visibles. La materia suspendida consiste en partículas muy pequeñas, que no se pueden quitar por medio de deposición. Pueden ser identificadas con la descripción de características visibles del agua, incluyendo turbidez y claridad, gusto, color y olor del agua.

1.1.3.1 Análisis o propuesta nacional

Recomendaciones propuestas por SUNASS para aguas destinadas para consumo humano: A1: 25 mg/L, A2: 50 mg/L.

Se entiende por Total de sólidos en suspensión o TSS a un parámetro utilizado en la calificación de la calidad del agua y en el tratamiento de aguas residuales. Algunas veces se asocia a la turbidez del agua.

1.1.4 Clorofila-a.

IMARPE (2016) indica que la clorofila-a es el principal pigmento fotosintético de las algas. Se considera como un proxy para la biomasa de fitoplancton presente en la columna de agua. Con las partículas minerales en suspensión, contribuye a la turbidez del agua. Cuantos más aportes de nutrientes (nitrógeno y fósforo) son importantes, más el ecosistema se vuelve propicio para el desarrollo del fitoplancton. La biomasa de fitoplancton no está necesariamente correlacionada con las concentraciones de nutrientes. De hecho, en sistemas ecológicamente eficaces, el fitoplancton utiliza todos los nutrientes disponibles.

La concentración de clorofila-a se utiliza para estimar en forma indirecta la biomasa de las comunidades Fito planctónicas, debido a que es el principal pigmento fotosintético presente en las algas (Gregor y Marsálek, 2004). La clorofila-a también es un indicador del grado de contaminación de los ecosistemas acuáticos y un importante índice del estado fisiológico del fitoplancton (Pinto, Von Sperling y Moreira, 2001).

1.1.5 Demanda bioquímica de oxígeno

DIGESA (2011) indica que la demanda biológica de oxígeno o demanda bioquímica de oxígeno (DBO) es un parámetro que mide la cantidad de dióxigeno consumido al degradar la materia orgánica de una muestra líquida. Es la materia susceptible de ser consumida u oxidada por medios biológicos que contiene una muestra líquida, disuelta o en suspensión. Dado que la materia orgánica no sólo son carbohidratos, una manera más práctica de analizar el consumo de oxígeno en la degradación de la materia orgánica en general, es medir los parámetros Demanda Bioquímica de Oxígeno (DBO₅).

La DBO₅ expresan la materia orgánica en términos generales, pero no indican su composición, la cual es muy variada. Como su origen proviene de organismos, y sus productos de degradación o de metabolismo, se puede afirmar que la componen proteínas, carbohidratos y lípidos y/o sus productos de degradación: aminoácidos, monosacáridos, hidrocarburos, ácidos grasos, alcoholes, más otros componentes propios de los vegetales como pigmentos DBO: Es el parámetro de contaminación orgánica.

1.1.5.1 Análisis o propuesta nacional

A1: La DBO es la cantidad de oxígeno consumido por microorganismo para la degradación biológica en un medio aeróbico de la materia orgánica un medio aeróbico, a mayor cantidad de materia orgánica mayor contenido de microorganismos en una muestra de agua por lo que se recomienda bajas concentraciones. Algunas legislaciones recomiendan DBO₅ de 3 mgO₂/L.

A2: se recomienda concentraciones aceptables entre 5 mgO₂/L y 10 mgO₂/L para aguas dulces.

Raffo y Ruiz (2014) concluyen que las aguas superficiales son altamente susceptibles a la contaminación; siendo el vertedero tradicional a lo largo de toda la historia de la industria y las poblaciones. En el caso de los contaminantes residuos que demandan oxígeno, afectan a las corrientes de agua como a las aguas estancadas. La materia orgánica requiere oxígeno para ser degradada en un curso de agua. El alto contenido orgánico favorece

el crecimiento de bacterias y hongos. El oxígeno utilizado para la oxidación de la materia orgánica, consume el oxígeno utilizado para el desarrollo de la fauna y flora acuática. Entre los efectos al ecosistema, se encuentra el cambio en la calidad del agua, y la posible elevación del pH, provocando la desaparición de peces y plantas.

1.1.6 Demanda química de oxígeno

DIGESA (2011) en su informe indica que DQO es la cantidad de oxígeno requerido para la oxidación química de la materia orgánica e inorgánica en el agua expresada en mg/L y se emplea un oxidante (dicromato potásico), que se determina en tres horas y, en la mayoría de los casos, guarda una buena relación con la DBO por lo que es de gran utilidad al no necesitar los cinco días de la DBO. Sin embargo, la DQO no diferencia entre materia biodegradable y el resto y no suministra información sobre la velocidad de degradación en condiciones naturales, el valor de la DQO es mayor que el de la DBO₅. El valor de la DQO es siempre superior al de la DBO₅ porque muchas sustancias orgánicas pueden oxidarse químicamente, pero no biológicamente, y su contenido es de materia orgánica: es de carbohidratos, proteínas, grasas e inorgánico (hierro ferroso, nitritos, amoníaco, sulfuros y cloruros)

1.1.6.1 Análisis o propuesta nacional

El DQO se emplea para determinar el contenido de materia orgánica en aguas naturales, estos compuestos orgánicos son oxidados por el dicromato de potasio, pero no necesariamente pueden ser biodegradables (acción microbiológica) y hay algunos alcoholes que no son oxidados por el dicromato, pero sí por los microorganismos. Los valores de DQO son más altos debido a que muchas sustancias orgánicas pueden oxidarse químicamente pero no biológicamente y no determina la oxidación bioquímica en un cuerpo hídrico.

A1: algunas normas consideran valores de 10 mg/L

A2: se recomiendan concentraciones de 20 mg/L

A3: concentraciones aceptables de 30 mg/L

B1: se recomienda concentraciones de 30 mg/L

B2: concentraciones aceptables de 50 mg/L

1.1.7 Oxígeno disuelto

DIGESA (2011) explica que la oxigenación del agua se debe principalmente a la solubilización del oxígeno atmosférico y minoritariamente a su generación en la fotosíntesis, principalmente de algas. Sin embargo, el oxígeno así formado durante el día, se consume en parte durante la noche, cuando las algas consumen oxígeno para su metabolismo. Luego de la muerte de las algas la degradación de esta biomasa también consume oxígeno. La concentración del oxígeno en agua depende, de la presión parcial del oxígeno en la atmósfera y de la temperatura del agua. El conjunto de residuos orgánicos producidos por los seres humanos, ganado, etc. incluyen heces y otros materiales que se vierten a los cuerpos de agua son descompuestos por bacterias aeróbicas, es decir en procesos con consumo de oxígeno. Cuando este tipo de desechos se encuentran en exceso, la proliferación de bacterias agota el oxígeno y no es posible la existencia de peces u otros organismos vivos. El indicador para medir la contaminación por desechos o residuos orgánicos es la cantidad de oxígeno disuelto (OD) en agua o demanda bioquímica de oxígeno (DBO_5).

1.1.7.1 Análisis o propuesta nacional

El OD depende de la presión parcial del oxígeno en la atmósfera y de la temperatura, el OD debe expresarse en porcentaje de saturación, esto se debe a que el OD varía en función de la temperatura y de la altitud. A una misma temperatura de 20°C a 0 msnm el OD es de 9 mg/L para 500 msnm es de 8 mg/L y para 1000 msnm, es de 7.4 mg/L. Los valores de porcentaje de saturación de OD de 80-120% se consideran excelentes y valores menores a 60% se consideran malos.

A1: Las concentraciones establecidas por otros países recomiendan valores mayores a 75% de saturación de OD

A2: las concentraciones aceptables para el OD son de 50%

A3: se recomienda concentraciones de 30%

B1: se recomienda concentraciones no menores al 80% entre un rango de 80-120%

El oxígeno es necesario para la mayoría de los organismos acuáticos vivos. Su biodisponibilidad varía según la profundidad y la época del año. Durante los períodos de estratificación térmica (estaciones cálidas y frías), los estratos no se mezclan y la biodisponibilidad del oxígeno es heterogénea según las capas. En períodos de estratificación térmica, el epilimnio es rico en oxígeno y el hipolimnion es deficitario. Sin embargo, el déficit de oxígeno puede ser estimulado por otros factores, como la alta producción primaria (por ejemplo, durante un período de eutrofización) o por la profundidad del lago (Anderson et al., 2007).

1.1.8 pH

DIGESA (2011) concluye que el pH es el valor que determina si una sustancia es ácida, neutra o básica, calculando el número iones hidrogeno presentes. Se mide en una escala a partir de 0 a 14, en la escala 7, la sustancia es neutra. Los valores de pH por debajo de 7 indican que una sustancia es ácida y los valores de pH por encima de 7 indican que es básica. Cuando una sustancia es neutra el número de los átomos de hidrógeno y de oxhidrilos son iguales. Cuando el número de átomos de hidrógeno (H+) excede el número de átomos del oxhidrilo (OH-), la sustancia es ácida.

1.1.8.1 Análisis o sustento de propuesta nacional

A1: Es recomendable para que la desinfección con cloro sea eficiente el pH debe ser inferior a 8, se establece dentro de un rango de 6.5 – 8.5

A2: los valores recomendables por la comunidad europea 5.5-9.0

B1: se recomienda un pH dentro de un rango de pH 6-9

B2: el pH más aceptable de 6.5-8.5

1.1.9 Conductividad Eléctrica

DIGESA (2011) indica que la conductividad es una expresión numérica de la capacidad de una solución para transportar una corriente eléctrica. Esta capacidad depende de la presencia de iones y de su concentración total, de su movilidad, valencia y concentraciones relativas, así como la temperatura de la medición.

El agua pura tiene muy poca conductividad, por lo que la medida de la conductividad de un agua nos da una idea de los sólidos disueltos en la misma. De la conductividad eléctrica, que indica la presencia de sales en el agua, lo que hace aumentar su capacidad de transmitir una corriente eléctrica, propiedad que se utiliza en mediciones de campo o de laboratorio, expresadas en micro Siemens/l ($\mu\text{S/l}$).

A partir de la conductividad se puede obtener los sólidos disueltos multiplicando por un factor entre 0.55 y 0.75.

Los sólidos disueltos totales, expresados en mg/l, pueden ser obtenidos por multiplicación de la conductividad por un factor comprendido entre 0,55 y 0,75. Este factor puede ser determinado para cada cuerpo de agua, pero permanece aproximadamente constante, según las proporciones iónicas en el cuerpo de agua y si éstas permanecen estables.

1.1.9.1 Análisis o propuesta nacional

A1/A2/A3: se recomienda concentraciones de 1000 us/cm

1.1.10 Coliformes totales (microbiológico)

DIGESA (2011) en su resumen indica que pueden hallarse tanto en heces como en el medio ambiente, por ejemplo, aguas ricas en nutrientes, suelos, materias vegetales en descomposición. También hay especies que nunca o casi nunca se encuentran en las heces pero que se multiplican en el agua.

1.1.10.1 Análisis o sustento de propuesta nacional

A1: Basados en la Comunidad Económica Europea – CEE y Ecuador se propone 50 NMP/100 ml como estándar

A2: Basados en la Comunidad Económica Europea – CEE se propone 5000 NMP/100 ml como estándar

A3: Basados en la Comunidad Económica Europea – CEE se propone 50000 NMP/100 ml como estándar

Larrea et al. (2013) indican que los coliformes totales se definen como bacterias Gram negativas en forma bacilar que fermentan la lactosa a temperatura de 35 a

37 °C y producen ácido y gas (CO₂) en 24 h, aerobias o anaerobias facultativas, son oxidasa negativa, no forman esporas y presentan actividad enzimática β-alactosidasa. Entre ellas se encuentran *Escherichia coli*, *Citrobacter*, *Enterobacter* y *Klebsiella*.

1.1.11 Coliformes Termotolerantes (Microbiológico)

DIGESA (2011) indica que los termotolerantes diferentes de *Escherichia coli* pueden proceder a aguas orgánicamente enriquecidas como efluentes industriales, de materias vegetales y suelos en descomposición.

1.1.11.1 Análisis o sustento de propuesta nacional

A1: Basados en la Comunidad Económica Europea – CEE y Ecuador se propone 20 NMP/100 ml como estándar

A2: Basados en la Comunidad Económica Europea – CEE se propone 2000 NMP/100 ml como estándar

A3: Basados en la Comunidad Económica Europea – CEE se propone 20000 NMP/100 ml como estándar

1.1.12 Nitrito

(DIGESA, 2011) concluye que las heces de los animales contienen proteína no asimilada (nitrógeno orgánico) y las proteínas que queda en los cuerpos de los animales y plantas que mueren se convierten en gran medida en amoníaco por acción de las bacterias heterótrofas, en condiciones aeróbicas y anaeróbicas. El amoníaco liberado por la acción de las bacterias sobre la urea y las proteínas es utilizado por las plantas. Si se libera en exceso es oxidado por las bacterias (nitrosomas) que en condiciones aeróbicas convierten el amoníaco a nitrito.

1.1.12.1 Análisis o sustento de propuesta nacional.

Los nitritos presentes en aguas superficiales y que luego pasen por un tratamiento, sus concentraciones se verán reducidas debido a la oxidación por la cloración.

A1: valores aceptables 1 mg/L N-nitrito

B1: se recomienda 1 mg/L Nitrito

1.1.13 Nitrato

DIGESA (2011) dice que los nitritos (NO_2) son oxidados por el grupo de nitro bacterias para formar nitrato (NO_3) Los nitratos formados pueden servir como fertilizantes para las plantas. Los nitratos producidos en exceso para las necesidades de la vida vegetal, son transportados por el agua, luego estas se filtran a través del suelo, debido a que el suelo no tiene la capacidad de retenerlos pudiendo encontrarse en concentraciones superiores en aguas subterráneas.

El uso excesivo de fertilizantes nitrogenados incluyendo el amoníaco, así como la contaminación causada por la acumulación de excretas humanos y animales puede contribuir a elevar la concentración de nitratos en el agua, estos son solubles y no adsorben a los componentes del suelo, por lo que son movilizados con facilidad por las aguas superficiales y subterráneas.

1.1.13.1 Análisis o sustento de propuesta nacional

Debido a que las practicas convencionales de tratamiento no modifican en forma apreciable los niveles de nitrato en agua y como sus concentraciones no varían en forma notable en los sistemas de distribución, las concentraciones detectadas en aguas superficiales son iguales a las aguas de consumo.

A1/A2: la EPA recomienda que la concentración de nitratos no debe excederse de 45 mg/L, dado los efectos sobre la salud de los infantes otras legislaciones establecen de 10 y 50 mg/L de N nitrato

A3: concentraciones aceptables de 50mg/L nitrato

B1: se recomienda 10 mg/L

1.1.14 Fosfato

DIGESA (2011) indica que los fosfatos son nutrientes para las plantas. Tienen aplicaciones industriales diversas y como fertilizantes. Los vertidos de fosfatos a las aguas naturales pueden causar eutrofización.

De la utilización de los fosfatos y nitratos, presentes en la materia orgánica de la basura, de los detergentes hechos a base de fosfatos ocurre el proceso eutrofización, que son arrastrados o arrojados a los ríos y lagos son un problema muy grave para

las aguas estancadas cerca de los centros urbanos o agrícolas. Durante las épocas cálidas la sobrecarga de estos productos químicos, que sirven de nutrientes, genera el crecimiento acelerado de vegetales como algas, cianobacterias, lirios acuáticos y lenteja de agua, las cuales al morir y ser descompuestas por las bacterias aeróbicas provocan el agotamiento del oxígeno disuelto en la capa superficial de agua y causan la muerte de los diferentes tipos de organismos acuáticos que consumen oxígeno, en las aguas de los lagos y ríos. Lago eutrófico es aquel de poca profundidad y poco contenido de oxígeno disuelto cual baja de 9 mg/l a 4 mg/l lo cual afecta negativamente y de inmediato a los organismos. Cuando el nivel baja a 2 mg/l todos los animales han muerto y Hay una significativa elevación de la DBO.

1.1.14.1 Análisis o propuesta nacional

A1: se recomienda valores de 0.4 mg/L P2O5

A2/A3: concentraciones aceptables 0.7 mg/L P2O5

B1/B2: (No Aplicable)

1.1.15 Nitrógeno

DIGESA (2011) indica que la cantidad de nitrógeno que se determina por el método Kjeldahl, incluye el nitrógeno de compuestos orgánicos y el nitrógeno amoniacal.

En el caso de los vegetales y animales, el nitrógeno se encuentra en forma orgánica, en un medio acuoso este rápidamente se transforma en nitrógeno amoniacal (esto requiere de condiciones anaeróbica) de esta forma encontramos mucho nitrógeno amoniacal en el agua lo que significa que estamos en presencia de materia orgánica en descomposición, en un medio que no requiere mucho oxígeno

El nitrógeno orgánico incluye productos naturales, como las proteínas y pépticos ácidos nucleicos y urea y numerosos materiales orgánicos sintéticos. La concentración típica del nitrógeno orgánico varía desde unos cientos de microgramos por litro en algunos lagos, hasta más de 20 mg/L en aguas residuales.

1.1.15.1 Análisis o sustento de propuesta nacional

A1: se recomienda concentraciones de 1 mg/L N

1.1.16 Concentración de fósforo

MVOTMA (25 de noviembre 2019) escribe que el fósforo es un componente esencial para la vida, por lo cual es uno de los nutrientes fundamentales para el crecimiento vegetal y animal, formando parte del ADN, ATP, y las moléculas de fosfolípidos de las membranas celulares. Tiene su origen natural en la erosión de las rocas y la degradación de la materia orgánica. El fósforo ocurre en aguas naturales y residuales principalmente en forma particulada inorgánica u orgánica y en una menor fracción en forma disuelta, orgánica (nucleótidos, polinucleótidos, sustancias húmicas, poli, meta y ultra fosfatos) e inorgánica (PO_4^{3-} , HPO_4^{2-} , $H_2PO_4^-$). La concentración de fósforo total mide la cantidad de fósforo disponible en forma orgánica e inorgánica, disuelta y particulada en los sistemas acuáticos.

Este indicador tiene como objetivo verificar el cumplimiento del estándar de calidad del agua establecido en la clase 3 del Decreto 253/79, el cual determina una concentración igual o menor a 0,025 mgP/L.

1.1.17 Nitrógeno amoniacal

DIGESA (2011) indica que todo nitrógeno que existe como Ion amonio o en el equilibrio se considera nitrógeno amoniacal $NH_4^+ = NH_3 + H^+$.

En una contaminación reciente la mayor parte de nitrógeno está presente en forma de nitrógeno orgánico (proteína) y amoniaco, a medida que pasa el tiempo, el nitrógeno orgánico se convierte gradualmente a nitrógeno amoniacal y más tarde si existen las condiciones aeróbicas, ocurre la oxidación del amoniaco a nitritos y nitratos.

El amoniaco y el amonio son gases que se producen de forma natural por fermentaciones microbianas de productos nitrogenados, por ejemplo, en la descomposición de proteínas o urea. El nitrógeno que proviene de la descomposición de vegetales, animales y excrementos pasa por una serie de transformaciones. En el caso de los vegetales y animales, el nitrógeno se encuentra en forma orgánica. Al llegar al agua, es rápidamente transformado en nitrógeno amoniacal, pasando después para a nitritos y finalmente a nitratos. Esas dos últimas transformaciones solamente ocurren en las aguas que contengan bastante oxígeno disuelto, pues son efectuadas por bacterias de naturaleza aerobia- llamadas nitro bacterias. De esa forma, cuando encontramos mucho nitrógeno amoniacal en el agua, estamos en

presencia de materiales orgánicos en descomposición y por lo tanto en un medio pobre en oxígeno.

1.1.17.1 Análisis o sustento de propuesta nacional

El nitrógeno amoniacal se oxida conforme avanza el tiempo, luego se forma nitritos por la acción de las bacterias nitrosomonas, los nitritos son compuesto toxico, si el nitrógeno amoniacal con concentraciones a 10 mg/L con un pH > a 8 hace que el amoniaco libre sea toxico a partir de concentraciones mayores 0.02 mg/L

A partir de concentraciones menores de 1 mg/L con un pH < 8 el amoniaco libre no presenta efectos tóxicos.

A1: La concentración para otros países de nitrógeno amoniacal de 0.5 mg/L.

1.1.18 Transparencia

DIGESA (2011) indica que la luz, que es esencial para el crecimiento de las plantas, viaja más lejos en las aguas claras que en cualquier agua turbia que contiene sólidos en suspensión o agua de con color.

Como la luz del sol proporciona la energía para la fotosíntesis, la penetración de la luz del sol determina la profundidad a la que las algas y otras plantas acuáticas pueden crecer, y la cantidad relativa de crecimiento. La transparencia decrece a medida que el color, los sólidos en suspensión o la abundancia de algas aumenta.

Acidez del agua (pH) El rango de pH requerido para mantener la vida acuática es entre 6,5 y 9. Fuera de estos umbrales, muchas funciones biológicas, como la respiración o la reproducción, se alteran.

1.1.19 Modelos estructurales

Bollen (1989) presenta la revisión metodológica del método.

1.1.19.1 Revisión metodológica de los modelos de ecuaciones estructurales.

El concepto de modelos de ecuaciones estructurales (SEM) se introdujo hace 80 años por Wrigth (1921, 1934), trabajando sobre patrones de covariación entre

varias características de cobayas, desarrolló una forma de romper las correlaciones observadas en un sistema de ecuaciones que matemáticamente describían su hipótesis respecto a las relaciones causales, estas relaciones entre variables fueron representadas en un “path diagram”. Habitualmente incluyen múltiples variables observables y múltiples variables no observables (variables latentes, también llamadas constructos). En palabras de Bollen (1989): “Podemos mirar estos modelos de diversos modos. Son ecuaciones de regresión con supuestos menos restrictivos, que permiten errores de medida tanto en las variables criterio (independientes) como en las variables dependientes. Consisten en análisis factoriales que permiten efectos directos e indirectos entre los factores. Habitualmente incluyen múltiples indicadores y variables latentes. Resumiendo, engloban y extienden los procedimientos de regresión, el análisis econométrico y el análisis factorial.”

1.1.19.2 Tipos de variables en modelos de ecuaciones estructurales.

Según sea su medición o el papel que realizan dentro del modelo estos se clasifican:

- Variables latentes, constructos, factores o variables no observadas son el objeto de interés en el análisis. Son conceptos abstractos que pueden ser observados indirectamente a través de sus efectos en los indicadores o variables observadas.
- Variable observada, o también denominada de medidas o indicadores, son aquellas variables que pueden ser medidas directamente.

Entre las variables latentes, se puede destacar tres tipos de variables, que son:

- Variable exógena. Son variables latentes independientes, es decir, afectan a otras variables y no recibe ningún efecto de ninguna de ellas. Estas variables se pueden detectar en la gráfica porque no salen ninguna de las flechas de esta variable.
- Variable endógena. Variables latentes dependientes, son aquellas que reciben el efecto de otras variables, es decir, en las gráficas son las

variables a las que llegan las flechas. Estas variables están afectadas por un término de perturbación o de error.

- Variable error. Este término tiene en cuenta todas las fuentes de variación que no están consideradas en el modelo.

1.1.19.3 Elementos de un modelo de ecuaciones estructurales.

Un modelo de ecuaciones estructurales completo consta de dos partes fundamentales: el sub-modelo de medida y el sub-modelo de relaciones estructurales. El primero contiene la manera en que cada constructo, o variable latente, está medido mediante sus indicadores observables, los errores que afectan a las mediciones y las relaciones que se espera encontrar entre los constructos cuando éstos están relacionados entre sí. Por su parte, el sub-modelo de relaciones estructurales es el que realmente se desea estimar, contiene los efectos y relaciones entre los constructos, es similar a un modelo de regresión, pero puede contener además efectos concatenados y bucles entre variables. Además, contiene los errores de predicción (que son distintos de los errores de medición).

1.1.19.4 Los diagramas estructurales. “Path-diagrams”

Los diagramas estructurales sirven para representar un modelo causal y sus relaciones, estos diagramas se denominan diagramas causales, gráfico de rutas, diagramas de senderos o diagramas estructurales.

Convenciones particulares para poder derivar las ecuaciones: Las variables observables se representan encerradas en rectángulos, Las variables no observables (latentes) se representan encerradas en óvalos o círculos, Los errores (sean de medición o de predicción) se representan sin rectángulos ni círculos (aunque algunos programas las dibujan como variables latentes), Las relaciones bidireccionales (covariación entre dos variables exógenas o dos términos de perturbación) se representan como vectores curvos con una flecha en cada extremo. El parámetro asociado indica la covarianza, La falta de flecha entre dos variables significa que dichas variables no están directamente relacionadas, aunque sí podrán estarlo indirectamente.

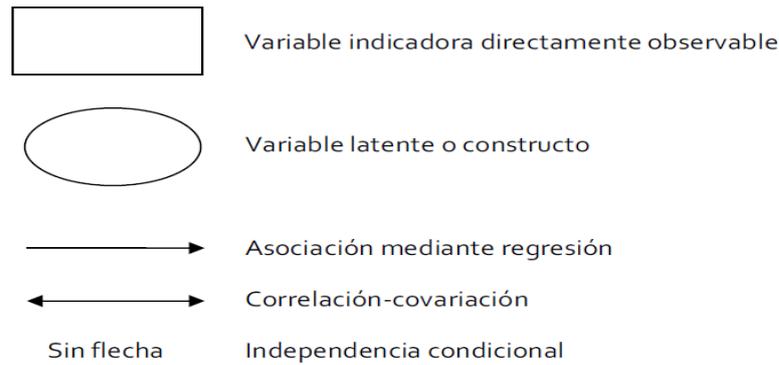


Figura 1. Convenciones “path diagrams”

Fuente: Bollen 1989.

Cualquier efecto estructural se representa como una flecha recta y unidireccional, cuyo origen es la variable predictora y cuyo final, donde se encuentra la punta de la flecha, es la variable dependiente. Las variables a las que llega alguna flecha se denominan endógenas y a las que no llega ninguna flecha se denomina exógena. Las variables endógenas están afectadas por un término de perturbación aleatorio que se incluye en el diagrama como una flecha adicional a la variable endógena.

Toda variable que reciba efecto de otras variables del modelo deberá incluir también un término error.

La Figura 2 ayudará a identificar los elementos que los conforman y, por tanto, las diferencias entre uno y otro.

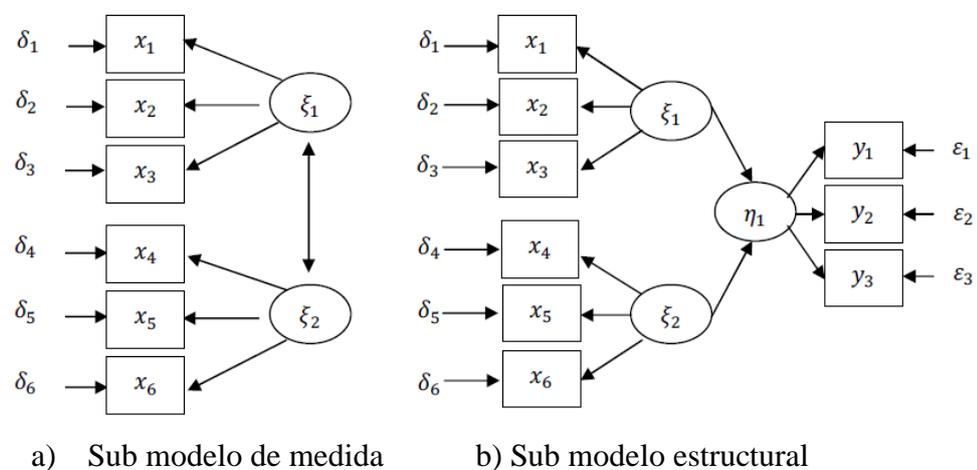


Figura 2. “Path diagrams” de un sub-modelo de medida y un sub-modelo estructural

Fuente: Bollen 1989

1.1.19.5 Relaciones entre variables

Tipos de relaciones causales que pueden establecerse entre dos variables, por ejemplo, v_1 y v_2 , para que covaríen:

- v_1 y v_2 pueden covariar si v_1 causa v_2 , asumir un modelo de regresión en el que v_2 fuese la variable respuesta y v_1 la variable predictora. Igualmente, también pueden covariar si v_1 causa v_2 , ante lo cual, estaríamos hablando de una regresión de v_1 sobre v_2 . En ambos casos, las relaciones son relaciones directas, que, a su vez, también podrían ser recíprocas.
- Así mismo, v_1 y v_2 también covariarán si ambas tienen una causa común, v_3 . Estaríamos ante lo que se ha dado en conocer como relación espuria.
- Otro tipo de relación en virtud de la cual v_1 y v_2 pueden covariar, es a través de una tercera variable interviniente v_3 ; estableciéndose entre la primer y la segunda de ellas, lo que se conoce como una relación indirecta.

Por último, otra forma de covariación entre dos variables (v_1 y v_2) es donde se emparentan con la relación indirecta y con la relación espuria.

1.1.19.6 Construcción del modelo estructural

Cupani (2012) en su artículo científico indica que el modelo de ecuaciones estructurales (Structural Equation Modeling, SEM) es una técnica que combina tanto la regresión múltiple como el análisis factorial. Permite al investigador no solo evaluar las muy complejas interrelaciones de dependencia sino también incorporar los efectos del error de medida sobre los coeficientes estructurales al mismo tiempo y propone los siguientes pasos en la construcción de un modelo estructural.

1) Fase de especificación

Se aplica conocimientos teóricos del fenómeno estudiado al planteamiento de las ecuaciones matemáticas relativas a los efectos causales de las variables latentes y a las expresiones que las relacionan con los indicadores

o variables observables. Esta distinción es importante porque cualquier relación entre variables, sin especificar por el investigador, se asume que es igual a cero. Además, en esta etapa, se formulan enunciados sobre el conjunto de parámetros, decidiendo entre los que serán libres para ser estimados o fijos, a los que se les asignará un valor dado, normalmente cero. Asimismo, se especifican los supuestos estadísticos sobre las fuentes de variación y en concreto sobre la forma de distribución conjunta, que en la mayoría de las técnicas empleadas se considera normalidad multivariante (Kaplan, 2000).

El sub-modelo estructural es aquel componente del modelo general que describe las relaciones causales entre variables latentes. Las relaciones entre las variables directamente observables se incluyen cuando éstas no actúan como indicadores de las variables latentes.

Entonces, habrá tantas ecuaciones estructurales como constructos latentes (variables endógenas) que sean explicadas por otras variables exógenas (latentes u observadas). La estructura que sigue este tipo de ecuaciones de los sub-modelos estructurales se puede expresar de la siguiente forma:

$$\eta = \beta\eta + \Gamma\xi + \zeta$$

donde:

“ η ” (eta) es un vector “ $p \times 1$ ” de variables endógenas latentes (esto es, variables predictoras o variables no explicadas por otras incluidas en el modelo).

“ ξ ” (xi) es un vector “ $q \times 1$ ” de variables exógenas latentes (esto es, variables predictoras o variables no explicadas por otras incluidas en el modelo).

“ Γ ” (gamma) es una matriz “ $p \times q$ ” de coeficientes (γ_{ij}) que relacionan las variables latentes exógenas (ξ) con las endógenas (η). Indican que una unidad de cambio en la variable exógena ξ_i resulta en un cambio en η_i γ_{ij} unidades, manteniendo todas las demás variables constantes. Este

coeficiente estructural se interpreta, al igual que β_{ij} como efectos directos de las variables endógenas.

“ β ” (beta) es una matriz “ $q \times q$ ” de **coeficientes** que relacionan las variables latentes endógenas entre sí. Cada β_{ij} indica una unidad de cambio en la variable endógena η_i , manteniendo todas las demás variables constantes. Para cada efecto hipotetizado de una variable latente endógena en otra, de las mismas características, se tendrá un coeficiente estructural β_{ij} .

“ ζ ” (zeta) es un vector “ $q \times 1$ ” de errores o términos de perturbación. Indican que las variables endógenas no son perfectamente predichas por las ecuaciones estructurales. Se supone que no existe correlación entre los errores y las variables exógenas.

Por otra parte, las variables latentes están relacionadas con variables observables a través del sub-modelo de medida, que está definido tanto por variables endógenas como por variables exógenas a través de las siguientes expresiones:

$$y = \Lambda_y \eta + \varepsilon \quad y \quad x = \Lambda_x \xi + \delta$$

donde:

η es un vector $m \times 1$ de variables latentes endógenas,

ξ es un vector $k \times 1$ de variables latentes endógenas,

Λ_x es una matriz $q \times k$ de coeficientes de variables exógenas,

Λ_y es una matriz $p \times m$ de coeficientes de variables endógenas,

δ es un vector $q \times 1$ de errores de medición para los indicadores y ,

ε es un vector $p \times 1$ de errores de medición para los indicadores endógenos.

x es el conjunto de variables observables del modelo de medida.

y es el conjunto de variables observables del modelo de estructura

2) Fase de estimación

En esta fase se pueden emplear diferentes tipos de estimación de los parámetros, con el objetivo de determinar cuál de ellos presentan un mejor ajuste: máxima verosimilitud, mínimos cuadrados ponderados y mínimos cuadrados generalizados, máxima verosimilitud, mínimos cuadrados ponderados y mínimos cuadrados generalizados. La aplicación del primero de ellos se debe a que proporciona estimaciones consistentes, eficientes y no sesgadas con tamaños de muestra suficientes, siendo capaz de facilitar la convergencia de las estimaciones con los parámetros aún en ausencia de normalidad.

Sin embargo, algunos estudios sugieren que en condiciones de no normalidad las estimaciones aun siendo consistentes no resultan necesariamente eficientes, resultando recomendable con muestras pequeñas, la estimación por Máxima verosimilitud con la aplicación de procedimientos bootstrap, puesto que permite obtener empíricamente (mediante técnicas de remuestreo) estimaciones de los errores estándar de los parámetros del modelo, independientemente de su distribución. Además, la aplicación de métodos bootstrap permite determinar intervalos de confianza para las estimaciones.

Cuando los datos no siguen una distribución normal, el método de estimación más recomendado es un método de distribución libre asintótica (ADF), siendo el método de mínimos cuadrados ponderados uno de los más utilizados.

De esta forma, en el caso de que la distribución de las variables observadas no esté muy alejada de una distribución normal, se recomienda emplear Máxima verosimilitud.

Estimación por Máxima Verosimilitud (ML)

Es el método de estimación más utilizado en el ajuste de modelos de ecuaciones estructurales. La estimación por ML requiere que las variables observadas mantengan una distribución normal, sin embargo, la violación de la condición de normalidad multivariante no afecta a la

capacidad del método para estimar de forma no sesgada los parámetros del modelo. Además, cuando el tamaño de muestra aumenta la distribución de los estimadores se aproximan a la distribución normal.

$$\log L = \frac{1}{2}(N - 1) \left\{ \log |\Sigma(\theta)| + \text{tr} \left[S \sum (\theta)^{-1} \right] \right\} + c$$

dónde:

\log es un logaritmo natural,

L es la función de verosimilitud,

N es el tamaño de la muestra,

θ es el vector de parámetros,

S es la matriz de covarianzas empírica,

$\Sigma(\theta)$ es la matriz de covarianzas del modelo y $|\Sigma(\theta)|$ su determinante,

tr es la traza de la matriz y, c es una constante.

Maximizar $\log L$ es equivalente a minimizar la siguiente función:

$$F_{ML} = \log |\Sigma(\theta)| - \log |S| + \text{tr} \left[S \sum (\theta)^{-1} \right] - p$$

donde:

F_{ML} es el valor de la función de ajuste al finalizar la estimación y, p es el número de variables observadas.

a) Estimación por Mínimos Cuadrados Ponderados (WLS)

Cuando la condición de normalidad de los datos no se cumple, una de las posibilidades para la estimación de los parámetros son métodos alternativos como el criterio de mínimos cuadrados ponderados (Weighted Least Squares), para cuya aplicación no es necesaria dicha condición; método de estimación que en el programa AMOS se encuentra bajo la nomenclatura de método de estimación de distribución libre asintótica (ADF).

Una de las ventajas que presenta este método es que permite introducir en los análisis variables ordinales, variables dicotómicas y variables continuas que no se ajusten a criterios de normalidad.

$$F_{WLS} = [S - \sigma(\theta)]'W^{-1}[S - \sigma(\theta)]$$

dónde:

S es el vector de elementos no redundantes en la matriz de covarianzas empírica, $\sigma(\theta)$ es el vector de elementos no redundantes en la matriz de covarianzas del modelo, θ es un vector de $(t \times 1)$ parámetros, W^{-1} es una matriz $(k \times k)$ definida positiva con $k = p(p+1)/2$ y $p =$ número de variables observadas.

A pesar de presentar ciertas ventajas, la estimación por mínimos cuadrados ponderados también presenta importantes restricciones y desventajas.

Si bien, su punto fuerte reside en poder realizar la estimación de los parámetros sin la necesidad de que las variables observadas sigan una distribución normal, el valor del estadístico ji-cuadrado que proporciona únicamente será preciso cuando la muestra sea lo suficientemente grande.

La limitación de este método se puede comprobar con el hecho de que la matriz de pesos aumentará rápidamente al incrementarse el número de variables indicadoras. De este modo, la matriz de covarianzas asintótica, de orden $(k \times k)$, donde $k = p(p + 1)/2$ y p es el número de variables observadas, la matriz de pesos correspondiente a un modelo con 10 variables debería ser del orden (55×55) , con 1540 elementos no redundantes. Es por ello, por lo que la estimación por mínimos cuadrados ponderados, en comparación con el método de máxima verosimilitud, requiere muestras mucho más grandes para obtener estimaciones consistentes y eficientes.

3) Estimación por Mínimos Cuadrados Generalizados (GLS)

Se trata de un caso especial dentro de la estimación por mínimos cuadrados ponderados. Bajo condiciones de normalidad multivariante, la función de ajuste del método de estimación por mínimos cuadrados ponderados puede reescribirse de la siguiente forma:

$$F_{WLS} = \frac{1}{2} tr \left\{ \left[S - \sum (\theta) \right] V^{-1} \right\}^2$$

donde:

tr es la traza de la matriz,

S es la matriz de covarianzas empírica,

$\Sigma(\theta)$ es la matriz de covarianzas del modelo,

θ es un vector de parámetros (t x 1) y,

V^{-1} es la matriz de pesos p x p. Se trata de una matriz de pesos de bajas dimensiones.

El método de estimación por mínimos cuadrados generalizados sigue una función de ajuste similar, reemplazando de la expresión anterior, V^{-1} por inversa de la matriz de covarianzas empírica, de modo que su función de ajuste se puede expresar como:

$$F_{GLS} = \frac{1}{2} tr \left\{ \left[S - \sum (\theta) \right] S^{-1} \right\}^2$$

En general, el método GLS se utiliza con frecuencia como un método de estimación asintóticamente equivalente a F_{ML} , puesto que se basa en los mismos criterios que la estimación por máxima verosimilitud y se puede emplear bajo las mismas condiciones. Sin embargo, la estimación por máxima verosimilitud resulta más apropiada cuando el tamaño de muestra es pequeño. Además, resulta inapropiado cuando el modelo aumenta en tamaño y complejidad.

4) Fase de identificación del modelo.

En esta fase se identifica el modelo, en donde debemos asegurar que pueden ser estimados los parámetros del modelo. El modelo está identificado si todos los parámetros lo están, es decir, si existe una solución única para cada uno de los parámetros estimados. Existe una serie de reglas generales aplicables para identificar un modelo, una de ellas es la regla de los grados de libertad (Número de variables observadas \times [número de variables + 1J] /2). Se espera que los grados de libertad del modelo deban ser mayores o iguales a cero. Esto corresponde a lo que se denomina como modelo identificado o modelo sobre identificado. Un modelo identificado tiene exactamente cero grados de libertad ($gl = 0$). Aunque esto ofrece un ajuste perfecto del modelo, la solución no tiene interés puesto que no se puede generalizar. Un modelo sobre identificado es el objetivo de todos los modelos de ecuaciones estructurales. Tiene más información en la matriz de datos que el número de parámetros a estimar, lo que significa que tiene un número positivo de grados de libertad ($gl > 0$). Esto asegura que el modelo sea tan generalizable como sea posible. Finalmente, un modelo infraestimado tiene grados de libertad negativo ($gl < 0$), lo que significa que se intentan estimar más parámetros de los que permite la información disponible (Kaplan, 2000).

La aplicación de los modelos estructurales tiene por finalidad estimar los parámetros desconocidos del modelo especificado, para después contrastarlo estadísticamente. Un modelo estará identificado si los parámetros del modelo pueden estimarse a partir de los elementos de la matriz de covarianzas de las variables observables.

La regla conocida como “regla del conteo”, se emplea para identificar los modelos de ecuaciones estructurales. Se denotará al número total de variables con $s = p + q$, siendo p las variables endógenas y q las exógenas. Por consiguiente, el número de elementos no redundantes en Σ es igual a $\frac{1}{2} s(s+1)$.

Además, se denota al número total de parámetros a ser estimados en el modelo como t , por lo que, para realizar la identificación del modelo se debe tener la siguiente condición necesaria $t \leq \frac{1}{2} s(s+1)$.

- Si se tiene la igualdad, se dice que el modelo está identificado.
- Si t es estrictamente menor que $\frac{1}{2} s(s+1)$, se dice que el modelo está sobre identificado.
- Si t es mayor que $\frac{1}{2} s(s+1)$, entonces el modelo no está identificado.

Cuando el modelo de ecuaciones estructurales es recursivo, está también identificado.

5) Fase de evaluación e interpretación de modelos ecuaciones estructurales

Una vez que el modelo esté identificado restan dos pasos fundamentales para cualquier análisis: la evaluación e interpretación del mismo a fin de contrastar si los datos se han ajustado al modelo propuesto.

La utilidad del modelo vendrá dada por su capacidad para explicar la realidad observada. Esta capacidad debe evaluarse tanto para el conjunto del modelo, como para cada una de las relaciones expresadas en él.

En sentido general, podemos decir que existen tres alternativas para evaluar el modelo: la evaluación del ajuste del modelo global, la evaluación del ajuste del modelo de medida y la evaluación del ajuste del modelo estructural.

En definitiva, hay que destacar que para cada estimación aparecerán asociados un gran número de índices de bondad de ajuste que serán necesarios interpretar para concluir si el modelo es consistente con los datos empíricos. Será su correcta interpretación, tanto de forma individual, como en sentido global, lo que determine si podemos aceptar el modelo planteado.

1.1.19.7 Bondad de ajuste de los modelos de ecuaciones estructurales

Se obtiene el ajuste del sub-modelo de medida, seguido del ajuste del sub-modelo estructural y finalmente, se desarrollarán las medidas de ajuste global más importantes en la evaluación e interpretaciones de los modelos de ecuaciones estructurales.

El modelo de medida se construye a través de la aplicación de análisis factorial, relacionando indicadores con variables latentes; relación en la que subyace la existencia de una relación lineal entre el concepto no directamente observable y su(s) manifestación(es) observable(s).

El análisis factorial tiene mucho en común con el análisis de componentes principales. Ambos métodos buscan crear nuevas variables o factores (no directamente observables) que expliquen los datos, para lo que resulta necesario que las variables originales no estén incorreladas, porque de estarlo no habría nada que explicar de las variables.

Dado que los factores son variables latentes no directamente observables, se puede fijar de forma arbitraria su media en 0 y su varianza en 1, en otras palabras, las consideramos variables estandarizadas que, como ya se ha dicho, están incorreladas entre sí, de modo que los pesos factoriales resultan ser las correlaciones entre las variables directamente observables y los factores.

1.1.19.8 Ajuste del modelo estructural

El análisis "Path" o de senderos es el que modeliza las relaciones entre las variables latentes, o lo que es lo mismo, el modelo de estructura; en donde se incorporan el error de medida y la relación entre constructos o variables latentes.

Para evaluar el ajuste del modelo estructural, corresponde en primer lugar, analizar la significación alcanzada por los coeficientes estimados. De este modo, cualquier parámetro estimado debe ser estadísticamente diferente de cero, o lo que es lo mismo, si consideramos un nivel de significación igual a 0.05.

Consecuentemente, un parámetro no significativo indicaría que la relación propuesta no tiene ningún efecto sustancial, por lo que debería ser eliminado y, a continuación, reformular el modelo.

Hay que tener en cuenta que, la magnitud de los coeficientes no está únicamente determinada por su significación de los parámetros, ya que otros factores como el tamaño muestral y la varianza de las variables dependientes e independientes (cuanto mayor es la magnitud de la relación y el tamaño muestral y cuanto menor es la varianza de las variables dependientes e independientes, mayor es la probabilidad de obtener una relación estadísticamente significativa).

1.1.19.9 Medidas de ajuste global

Diremos que existe un ajuste perfecto cuando haya una correspondencia perfecta entre la matriz reproducida por el modelo y la matriz de observaciones. Entre los índices identificados como de ajuste global, podemos diferenciar entre tres tipos de medidas: medidas absolutas de ajuste, medidas incrementales de ajuste y medidas de ajuste de parsimonia.

1.1.19.10 Medidas absolutas de ajuste

Determinan el grado en que el modelo globalmente predice la matriz de datos inicial. En este grupo podemos encuadrar: al estadístico ji-cuadrado y el estadístico ji cuadrado no centrado (NCP). Junto a ellas, aparecen otras medidas como la raíz cuadrada del error cuadrático medio o “RMSEA” (“Root Mean Square Error of Approximation”), en virtud de la cual la discrepancia entre la matriz reproducida por el modelo y la matriz de observaciones está medida en términos de la población y no en términos de la muestra; el conocido como índice de bondad de ajuste o “GFI”, que aporta información sobre la variabilidad explicada por el modelo y cuyos valores oscilan entre 0 (no hay ajuste) y 1 (ajuste perfecto); y, por último, dentro de las medidas absolutas de ajuste resta mencionar al índice de bondad de ajuste relativo o “RGFI”, que ayuda a evaluar la bondad de ajuste del modelo considerando el tamaño muestral y el número de indicadores.

Es decir, en general, la mayor parte de estas medidas están comprendidas en el rango entre cero (mal ajuste) y uno (ajuste perfecto), sin embargo, autores como Hu y Bentler (1995) consideran que la distribución de estas medidas es desconocida, a excepción obviamente, del test χ^2 .

Análogamente, aparecen dentro de este grupo de medidas: RMR (“Root Square Residual” o Residuo cuadrático medio) y SRMR (“Standarized Root Mean Square Residual”).

De todos ellos, la única medida de bondad de ajuste con un test de significación asociado es el test χ^2 , mientras que el resto de medidas e índices son descriptivos.

La bondad de ajuste será mayor cuanto menor sea la diferencia entre ambas matrices, de modo que si la diferencia es muy pequeña se podrá entender que el modelo reproduce el comportamiento de los datos observados de un modo razonablemente bueno.

Al tratarse de parámetros poblacionales, estas matrices son desconocidas, por lo que será necesario examinar la matriz de covarianzas empírica S y la matriz de covarianzas estimada del modelo $\Sigma(\hat{\theta})$, donde $\hat{\theta}$ es un vector ($t \times 1$) de parámetros estimados. Si se verifica la hipótesis nula, el valor mínimo de la función de ajuste $N-1$ converge a una χ^2 .

$$\chi^2(df) = (N - 1)F \left[S, \sum (\hat{\theta}) \right]$$

con $df = s - t$ grados de libertad, donde:

s es el número de elementos no redundantes en S ,

t es el número total de parámetros a estimar,

N es el tamaño de muestra,

S es la matriz de covarianzas empírica, y

$\Sigma(\hat{\theta})$ es la matriz de covarianzas estimada

AMOS, proporciona diferentes valores para el test χ^2 para los métodos de estimación por máxima verosimilitud, mínimos cuadrados generalizados y estimación por distribución libre asintótica; es decir, el valor del test dependerá del método de estimación que se utilice. En este sentido y, de modo general, se puede decir que valores altos de dicho test en relación al número de grados de libertad indican que la matriz de covarianzas de la población Σ y la matriz de covarianzas del modelo $\Sigma(\theta)$ son significativamente distintas una de otra.

Por su parte, los residuos (los elementos resultantes de $S - \Sigma(\hat{\theta})$), deben ser nulos o cercanos a cero en un buen ajuste, por lo que se buscará obtener un test χ^2 no significativo. De este modo, si el p-valor asociado a dicho test es mayor que .05, se aceptará la hipótesis nula y el modelo será compatible con la matriz Σ de covarianzas de la población, en cuyo caso el modelo se ajustaría bien a los datos. En caso contrario, se puede concluir que el modelo propuesto no es consistente con los datos observados.

Sin embargo, ha de tenerse en cuenta la existencia de tres factores que hacen que el test χ^2 pueda perder eficacia:

- Violación de la condición de normalidad de las variables observadas.
- Complejidad del modelo. Uno de los puntos débiles del estadístico X^2 es que su valor decrece cuando incrementan los parámetros del modelo. Es decir, para modelos muy parametrizados y de gran complejidad tiende a producir valores más pequeños de este estadístico en comparación con modelos más sencillos dada la reducción de los grados de libertad.

1.1.19.11 Medidas incrementales de ajuste

Nos referimos a ellas como medidas incrementales de ajuste o también como medidas descriptivas basadas en la comparación de modelos.

1.1.19.12 Medidas de ajuste de parsimonia

La parsimonia de un modelo es el grado en que alcanza ajuste para cada coeficiente o parámetros estimados, además de funcionar como criterio de selección entre modelos alternativos.

1.2.23 Análisis Factorial

El análisis factorial es un método de interdependencia (no hace distinción entre variables dependientes e independientes) y tiene como objetivo simplificar las múltiples y complejas relaciones que puedan existir en un conjunto de variables métricas. Para ello trata de encontrar dimensiones comunes o factores (grupos homogéneos de variables) que ligan a las aparentemente no relacionadas variables.

Una vez que se dispone de esta matriz concierne examinarla para comprobar si sus características son adecuadas para realizar un Análisis Factorial. Uno de los requisitos que deben cumplirse para que el Análisis Factorial tenga sentido es que las variables estén altamente correlacionadas. Este examen puede realizarse con diferentes métodos:

Hay diversas formas de afrontar esas dificultades que dan lugar a otros tantos procedimientos de extracción de factores;

1.2.23.1 Pasos para realizar un análisis factorial

Luque (2000), indica que los pasos para realizar un análisis factorial es:

a) Fase de preparación: formulación del problema.

En esta fase debe seleccionarse las variables a analizar, así como los elementos de la población a observar. Aunque pueden realizarse análisis factoriales con variables discretas y/o ordinales lo habitual es que las variables sean cuantitativas continuas.

b) Calcular la matriz de correlaciones. Matriz R.

i) El modelo factorial.

Tiene como objetivo simplificar las múltiples y complejas relaciones que puedan existir entre un conjunto de variables observadas X_1, X_2, \dots, X_p , para ello trata de encontrar dimensiones comunes o factores ($k < p$) no directamente observables F_1, F_2, \dots, F_k que expliquen suficientemente a las variables observadas perdiendo el mínimo de información. Además, los factores han de extraerse de forma que sean independientes entre sí, es decir, que sean ortogonales. Además, se supone que son variables tipificadas (con media cero y varianza uno).

ii) Correlaciones

La existencia de una estructura subyacente presupone la existencia previa de correlaciones entre las variables observables. Podemos optar por calcular las correlaciones entre variables o atributos (análisis R), o por obtener la asociación entre individuos u objetos (análisis Q).

Evidentemente, antes de realizar un análisis factorial nos plantearemos si las p variables originales están correlacionadas entre sí o no lo están.

- Una aproximación para comprobar que esta condición se cumple, es inspeccionando visualmente la matriz de correlaciones. Para determinar si existe correlación, el coeficiente debe ser superior a 0.5 aunque otros autores toman como referencia 0.3.
- Determinante de la matriz de correlaciones.
- Contraste (test) de esfericidad de Bartlett.
- Coeficiente de correlación parcial.
- La medida (índice) de adecuación muestral de Kaiser-Meyer-Olkin (KMO)
- Medida de adecuación de la muestra (MSA: Measurement of Sampling Adequacy).
- Correlación Múltiple al cuadrado

c) **Extracción y selección de los factores**

Existen muchos métodos para obtener los factores comunes. Daremos una breve referencia de algunos de ellos: Método de las Componentes Principales, Método de los Ejes Principales o componentes principales iteradas, Método de los mínimos cuadrados no ponderados, Método de mínimos cuadrados generalizados, Método de la Máxima Verosimilitud, Método alfa (alpha), Método de Thurstone para obtener los factores, Método del factor principal para obtener factores, Método MINRES, ULS y GLS para obtener factores, Criterio, regla o test de Kaiser, Scree-test de Cattell, Velicer, Bartlett, Análisis Paralelo fue sugerido por Horn, Determinación "a priori" del número de factores, Criterio del porcentaje de la varianza, Gráfico de Sedimentación, Criterio de división a la mitad.

d) Rotación de factores

Con frecuencia es difícil interpretar los factores iniciales, por lo tanto, la extracción inicial se rota con la finalidad de lograr una solución que facilite la interpretación. Hay dos sistemas básicos de rotación de factores: los métodos de rotación ortogonales (mantienen la independencia entre los factores rotados: varimax, quartimax y equamax) y los métodos de rotación no ortogonales (proporcionan nuevos factores rotados que guardan relación entre sí). En el presente estudio aplicaremos los métodos de rotación ortogonales, específicamente el Método de Rotación Varimax. Éste es, actualmente, uno de los métodos más utilizados.

Las variables obtenidas son latentes, no observables, y su descripción e interpretación debe contribuir a una mejor comprensión del tema objeto del estudio.

- Método Varimax
- Método Quartimax
- Método Equamax
- Rotación oblicua.
- Método Promax

e) Cálculo de puntuaciones factoriales

Una vez determinados los factores rotados el siguiente paso es calcular la matriz de puntuaciones factoriales F . Las posibilidades de analizar las puntuaciones factoriales de los sujetos son muy variadas según lo que se pretenda:

- conocer qué sujetos son los más raros o extremos, es decir, la representación gráfica de las puntuaciones factoriales para cada par de ejes factoriales puede ayudar a detectar casos atípicos;
- conocer dónde se ubican ciertos grupos o subcolectivos de la muestra (los jóvenes frente a los mayores, los de clase alta frente a los de baja, los más católicos frente a los no católicos, los de una provincia frente a los de otras provincias, etc);

- conocer en qué factor sobresalen unos sujetos y en qué factor no, etc.
- explicar, analizando las informaciones anteriores, por qué han aparecido dichos factores en el análisis realizado

El Análisis Factorial es en otras ocasiones un paso previo a otros análisis, como, por ejemplo, Regresión Múltiple o Análisis Cluster, en los que se sustituye el conjunto de variables originales por los factores obtenidos. Por ello, es necesario conocer los valores que toman los factores en cada observación.

Métodos de cálculo de las puntuaciones

- Método de regresión
- Método de Barlett
- Método de Anderson-Rubin

1.2.24 Árboles de predicción

Existen muchos árboles de predicción, entre ellos tenemos:

- Árboles bagging,
- Random forest,
- boosting
- C5.0

Para la presente investigación se usará Árboles Bagging y Random forest

Los métodos predictivos como la regresión lineal o polinómica generan modelos globales en los que una única ecuación se aplica a todo el espacio muestral. El estudio de árboles implica múltiples predictores, que interaccionan entre ellos de forma compleja y no lineal, es muy difícil encontrar un modelo global que sea capaz de reflejar la relación entre las variables. Existen métodos de ajuste no lineal (step functions, splines,...) que combinan múltiples funciones y que realizan ajustes locales, sin embargo, suelen ser difíciles de interpretar. Los métodos estadísticos basados en árboles engloban a un conjunto de técnicas supervisadas no paramétricas

que consiguen segmentar el espacio de los predictores en regiones simples, dentro de las cuales es más sencillo manejar las interacciones.

1.2.24.1 Árboles de predicción en R

Existen diferentes paquetes en R que permiten la creación de modelos basados en árboles:

- `tree` y `rpart`: contiene funciones para la creación y representación de árboles de regresión y clasificación.
- `rpart.plot`: permite crear representaciones detalladas de modelos creados con `rpart`.
- `randomForest`: implementa algoritmos para crear modelos random forest.
- `gbm`: implementa algoritmos de boosting.
- `C50`: contiene los algoritmos C5.0 para árboles de clasificación.

1.2.24.2 Evitar el overfitting

El proceso de construcción de árboles tiende a reducir rápidamente el error de entrenamiento, es decir, el modelo se ajusta muy bien a las observaciones empleadas como entrenamiento. Como consecuencia, se genera un overfitting que reduce su capacidad predictiva al aplicarlo a nuevos datos. La razón de este comportamiento radica en la facilidad con la que los árboles se ramifican adquiriendo estructuras complejas. De hecho, si no se limitan las divisiones, todo árbol termina ajustándose perfectamente a las observaciones de entrenamiento creando un nodo terminal por observación. Existen dos estrategias para prevenir el problema de overfitting de los árboles: limitar el tamaño del árbol y el proceso de podado (pruning).

1.2.24.3 Controlar el tamaño del árbol

El tamaño final que adquiere un árbol puede controlarse mediante reglas de parada que detengan la división de los nodos dependiendo de si se cumplen o no determinadas condiciones. El nombre de estas condiciones puede variar

dependiendo del software o librería empleada, pero suelen estar presentes en todos ellos.

- Observaciones mínimas para división:
- Observaciones mínimas de nodo terminal:
- Profundidad máxima del árbol:
- Número máximo de nodos terminales:
- Reducción mínima de error.

Todos estos parámetros son lo que se conoce como hiperparámetros porque no se aprenden durante el del modelo. Su valor tiene que ser especificado por el usuario en base a su conocimiento del problema y mediante el uso de validación cruzada.

1.2.24.4 Tree pruning

La estrategia de controlar el tamaño del árbol mediante reglas de parada tiene un inconveniente, el árbol crece seleccionando la mejor división en cada momento hasta alcanzar una condición de parada. Al evaluar las divisiones sin tener en cuenta las que vendrán después, nunca se elige la opción que resulta en el mejor árbol final, a no ser que también sea la que genera en ese momento la mejor división. A este tipo de estrategias se les conoce como greedy.

Una alternativa no greedy que consigue evitar el overfitting consiste en generar árboles grandes, sin condiciones de parada más allá de las necesarias por las limitaciones computacionales, para después podarlos (pruning), manteniendo únicamente la estructura robusta que consigue un test error bajo. La selección del sub-árbol óptimo puede hacerse mediante cross-validation, sin embargo, dado que los árboles se crecen lo máximo posible (tienen muchos nodos terminales) no suele ser viable estimar el test error de todas las posibles subestructuras que se pueden generar. En su lugar, se recurre al cost complexity pruning o weakest link pruning.

1.2.24.5 Ajuste del modelo

La función `tree()` del paquete `tree` permite ajustar árboles de predicción. La elección entre árbol de regresión o árbol de clasificación se hace automáticamente dependiendo de si la variable respuesta es de tipo numérico o factor. Es importante tener en cuenta que solo estos dos tipos de vectores están permitidos, si se pasa uno de tipo `character` se devuelve un error.

A continuación, se ajusta un árbol de regresión empleando como variable respuesta `medv` y como predictores todas las variables disponibles. La función `tree()` crece el árbol hasta que encuentra una condición de stop. Por defecto, estas condiciones son:

- `mincut = 5`: número mínimo de observaciones que debe de tener al menos uno de los nodos hijos para que se produzca la división.
- `minsize = 10`: número mínimo de observaciones que debe de tener un nodo para que pueda dividirse.
- `depth = 31`: profundidad máxima que puede alcanzar el árbol.

Esto implica que, no todas las variables pasadas como predictores en el argumento `formula`, necesariamente tienen que acabar incluidas en el árbol.

Como en cualquier estudio de regresión, no solo es importante ajustar el modelo, sino también cuantificar su capacidad para predecir nuevas observaciones. Para poder hacer la posterior evaluación, se dividen los datos en dos grupos, uno se emplea como `training data set` y el otro como `test data set`.

1.2.24.6 Podado del árbol (pruning)

Con la finalidad de reducir la varianza del modelo y así disminuir el test error, se somete al árbol a un proceso de pruning. El proceso de pruning intenta encontrar el árbol más sencillo (menor tamaño) que consigue explicar las observaciones. La función `cv.tree()` emplea `cross validation` para identificar el valor óptimo de penalización α . Por defecto, esta función emplea la deviance para guiar el proceso de pruning.

1.2.25 Redes Neuronales

Las Redes Neuronales son un campo muy importante dentro de la Inteligencia Artificial. Inspirándose en el comportamiento conocido del cerebro humano (principalmente el referido a las neuronas y sus conexiones), trata de crear modelos artificiales que solucionen problemas difíciles de resolver mediante técnicas algorítmicas convencionales

1.2.25.1 Neurona biológica

Fue (Ramón y Cajal, 1911) quién descubrió la estructura celular (neurona) del sistema nervioso. Defendió la teoría de que las neuronas se interconectaban entre sí de forma paralela, y no formando un circuito cerrado como el sistema sanguíneo.

Una neurona consta de un cuerpo celular (soma) de entre 10 y 80 μm , del que surge un denso árbol de ramificaciones (dendritas) y una fibra tubular (axón) de entre 100 μm y un metro.

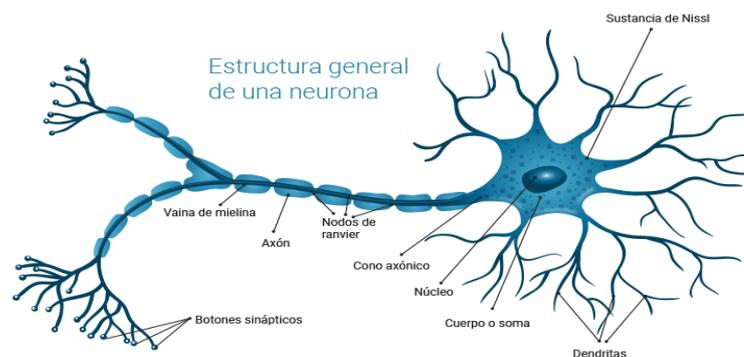


Figura 3. Neurona biológica

Fuente: Ramon y Cajal

De alguna forma, una neurona es un procesador de información muy simple:

- Canal de entrada: Dendritas
- Procesador: soma
- Canal de salida: axón

Una neurona cerebral puede recibir unas 10.000 entradas y enviar a su vez su salida a varios cientos de neuronas.

La conexión entre neuronas se llama sinapsis. No es una conexión física, sino que hay unos 2 mm de separación. Son conexiones unidireccionales, en la que la transmisión de la información se hace de forma eléctrica en el interior de la neurona y de forma química entre neuronas; gracias a unas sustancias específicas llamadas neurotransmisores.

1.2.25.2 Modelo de neurona artificial

El modelo de Rumelhart y McClelland (1986) define un elemento de proceso (EP), o neurona artificial, como un dispositivo que, a partir de un conjunto de entradas, x_i ($i=1...n$) o vector \mathbf{x} , genera una única salida.

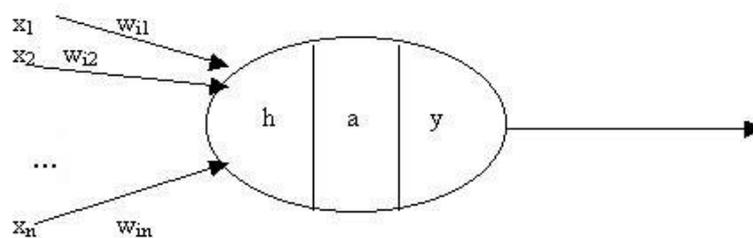


Figura 4. Modelo de neurona artificial

Fuente: Rumelhart y McClelland

Esta neurona artificial consta de los siguientes elementos:

- Conjunto de entradas o vector de entradas \mathbf{x} , de n componentes
- Conjunto de pesos sinápticos w_{ij} . Representan la interacción entre la neurona presináptica j y la postsináptica i .
- Regla de propagación $d(w_{ij}, x_j(t))$: proporciona el potencial postsináptico, $h_i(t)$.
- Función de activación $a_j(t) = f(a_j(t-1), h_i(t))$: proporciona el estado de activación de la neurona en función del estado anterior y del valor postsináptico.

- Función de salida $F_i(t)$: proporciona la salida $y_j(t)$, en función del estado de activación.

Las señales de entrada y salida pueden ser señales binarias (0,1 – neuronas de McCulloch y Pitts), bipolares (-1,1), números enteros o continuos, variables borrosas, etc.

La regla de propagación suele ser una suma ponderada del producto escalar del vector de entrada y el vector de pesos.

$$h_i(t) = \sum w_{ij}x_j$$

También se usa a menudo la distancia euclídea entre ambos vectores:

$$h_i(t) = \sum (x_j.w_i)^2$$

Existen otro tipo de reglas menos conocidas como la distancia de Voronoi, de Mahalanobis, etc.

La función de activación no suele tener en cuenta el estado anterior de la neurona, sino sólo el potencial $h_i(t)$. Suele ser una función determinista y, casi siempre, continua y monótona creciente. Las más comunes son la función signo (+1 si $h_i(t) > 0$ -1 en caso contrario la función de salida suele ser la identidad. En algunos casos es un valor umbral (la neurona no se activa hasta que su estado supera un determinado valor).

Con todo esto, el modelo de neurona queda bastante simplificado:

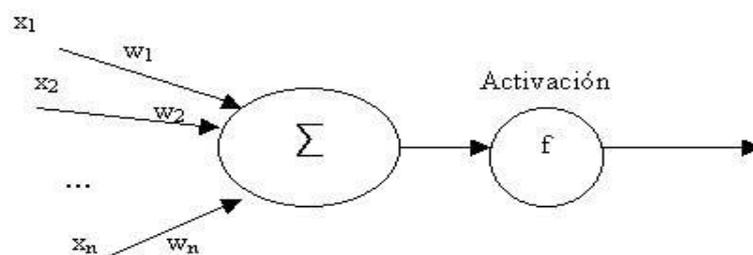


Figura 5. Modelo de Neuronas Simplificado

Fuente: Rumelhart y McClelland

1.2.25.3 Red neuronal artificial

Una red neuronal artificial (RNA) se puede definir (Hecht – Nielssen) como un grafo dirigido con las siguientes restricciones:

- 1) Los nodos se llaman elementos de proceso (EP).
- 2) Los enlaces se llaman conexiones y funcionan como caminos unidireccionales instantáneos
- 3) Cada EP puede tener cualquier número de conexiones.
- 4) Todas las conexiones que salgan de un EP deben tener la misma señal.
- 5) Los EP pueden tener memoria local.
- 6) Cada EP posee una función de transferencia que, en función de las entradas y la memoria local produce una señal de salida y / o altera la memoria local.
- 7) Las entradas a la RNA llegan del mundo exterior, mientras que sus salidas son conexiones que abandonan la RNA.

1.2.25.4 Arquitectura de las RNA

La arquitectura de una RNA es la estructura o patrón de conexiones de la red. Es conveniente recordar que las conexiones sinápticas son direccionales, es decir, la información sólo se transmite en un sentido.

En general, las neuronas suelen agruparse en unidades estructurales llamadas capas. Dentro de una capa, las neuronas suelen ser del mismo tipo. Se pueden distinguir tres tipos de capas:

- De entrada: reciben datos o señales procedentes del entorno.
- De salida: proporcionan la respuesta de la red a los estímulos de la entrada.
- Ocultas: no reciben ni suministran información al entorno.

Generalmente las conexiones se realizan entre neuronas de distintas capas, pero puede haber conexiones intracapa o laterales y conexiones de realimentación que siguen un sentido contrario al de entrada-salida.

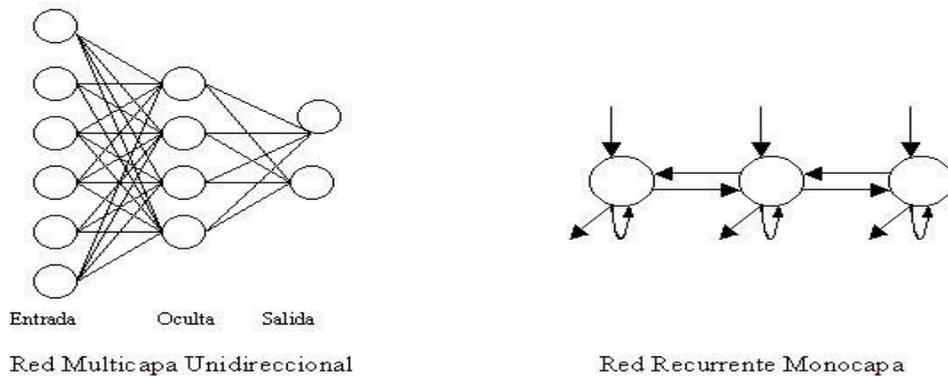


Figura 6. Arquitecturas de RNA

Fuente: Rumelhart y McClelland

1.2.25.5 Aprendizaje de las RNA

Es el proceso por el que una RNA actualiza los pesos (y, en algunos casos, la arquitectura) con el propósito de que la red pueda llevar a cabo de forma efectiva una tarea determinada.

Hay tres conceptos fundamentales en el aprendizaje:

- Paradigma de aprendizaje: información de la que dispone la red.
- Regla de aprendizaje: principios que gobiernan el aprendizaje.
- Algoritmo de aprendizaje: procedimiento numérico de ajuste de los pesos.

Existen dos paradigmas fundamentales de aprendizaje:

- Supervisado: la red trata de minimizar un error entre la salida que calcula y la salida deseada (conocida), de modo que la salida calculada termine siendo la deseada.
- No supervisado o autoorganizado: la red conoce un conjunto de patrones sin conocer la respuesta deseada. Debe extraer rasgos o agrupar patrones similares.

En cuanto a los algoritmos de aprendizaje, tenemos cuatro tipos:

- Minimización del error: reducción del gradiente, retropropagación, etc. La modificación de pesos está orientada a que el error cometido sea mínimo.
- Boltzmann: para redes estocásticas, donde se contemplan parámetros aleatorios.
- Hebb: cuando el disparo de una célula activa otra, el peso de la conexión entre ambas tiende a reforzarse (Ley de Hebb).
- Competitivo: sólo aprenden las neuronas que se acercan más a la salida deseada.

Los algoritmos, y en general el proceso de aprendizaje, son complejos y suelen llevar bastante tiempo computacionalmente hablando. Su ventaja es que una vez ha aprendido, la red puede congelar sus pesos y funcionar en modo recuerdo o ejecución.

1.2.25.6 Perceptrón Multicapa

Veamos el modelo más importante (por su antigüedad y utilidad) de red neuronal. Este modelo, el perceptrón multicapa, se basa en el perceptrón simple y resuelve de forma eficiente problemas de clasificación y reconocimiento de patrones.

1.2.25.7 Problema del OR Exclusivo

Una de las razones por la que las redes neuronales desaparecieron de la mayoría de los proyectos de investigación durante la década de los 70 y parte de los 80 fue que las redes conocidas hasta entonces, cuyo exponente máximo era el perceptrón (hoy perceptrón simple); no eran capaces de solucionar problemas no linealmente separables. Para problemas linealmente separables, existían soluciones algorítmicas tradicionales eficientes, luego no tenía sentido utilizar la tecnología neuronal. Para que nos hagamos una idea, un problema linealmente separable es aquél que puede dividirse en dos áreas claramente diferenciadas mediante una línea.

Vale cualquier línea para separarlos, tan compleja como se desee. Por ejemplo, si el problema que queremos resolver es saber dadas dos entradas, cuál es el resultado de realizar la operación lógica OR entre ellas, podemos dividir el problema linealmente de forma sencilla.

1.2.25.8 Arquitectura

La arquitectura de este applet consta de 3 capas, cada una totalmente conectada con la siguiente:

- Capa de Entrada de dos neuronas, que tomará los dos valores para hacer la operación. Aunque estos valores son binarios, la red está programada de forma que podría aceptar valores continuos y dar la respuesta adecuada.
- Capa Oculta, con un número configurable de neuronas mediante la opción nº de neuronas ocultas. El hecho de que exista la capa oculta es la que da al perceptrón multicapa la capacidad de resolver problemas no separables linealmente. La cantidad de neuronas que se utilicen hace que varíe el tiempo que se tarda en minimizar el error. 1 neurona hace el trabajo muy lento, 2 o 3 son el número más adecuado (de 6000 a 40000 iteraciones para un error de 0.1). 4 o más son muy eficientes (menos de 15000 its.), pero aumentan la carga de trabajo en cada iteración al aumentar el número de pesos a calcular.
- Capa de salida. Una sola neurona que nos da el valor del XOR entre las dos entradas. Se ha mantenido un valor continuo de salida para que pueda observarse el efecto de tolerar más o menos error.

El botón cargar red nos permite crear una red ya entrenada con un error muy bajo, para comprobar la máxima eficiencia del perceptrón sin necesidad de entrenar la red.

1.2.25.9 Entrenamiento

El entrenamiento se realiza con retropropagación batch (se pasan los cuatro juegos de ensayo antes de modificar los pesos). En el gráfico podemos

observar el descenso del error cuadrático medio según aumentan las iteraciones

El entrenamiento se realiza con retropropagación batch (se pasan los cuatro juegos de ensayo antes de modificar los pesos). En el gráfico podemos observar el descenso del error cuadrático medio según aumentan las iteraciones. Podemos configurar:

- **Momento:** controla en qué grado el estado en el que se encuentra la red neuronal influye en el estado que tendrá en el siguiente paso del aprendizaje.
- **Factor de Aprendizaje:** controla el efecto del error en la modificación de los pesos de las capas de neuronas.
- **Nº de Iteraciones:** número máximo de iteraciones que usaremos para entrenar la red. Si no se alcanza el error máximo tolerado en esas iteraciones, la red trabajará con el error al que haya llegado.
- **Error Máximo:** error cuadrático medio (MSE, Medium Square Error) máximo que tolerará la red. Mientras no se haya alcanzado, continuará el entrenamiento hasta que se alcance, o hasta que agotemos las iteraciones estipuladas, lo que ocurra primero. Se observa claramente que el error máximo influye en la respuesta de la red. Con un error de 0.1, la red responde a una entrada 0,0 con un 0.15, mientras que si el error es de 0.01 nos da un 0.08. Mediante el botón Cargar Red podemos usar unos pesos entrenados para un error de 0.01 para una red de 3 neuronas ocultas.

Existen otras redes multicapa que se harán uso en el presente trabajo de investigación, por lo que solo se mencionan:

- reconocimiento de caracteres escritos.
- Comprensión de imágenes
- Compresión de Imágenes a Color

- Reconocimiento de Disparos
- Reconocimiento de Dígitos Hablados
- Memoria Auto asociativa. La aplicación más conocida de las redes de Hopfield es su uso como Memoria Autoasociativa
- Redes de Hopfield

Las Redes de Hopfield pertenecen a un grupo especial de redes, llamadas redes recursivas o recurrentes, que se empezaron a desarrollar a principios de los 80. Son muy apropiadas para simulación de memorias asociativas.

1.2.26 Inteligencia artificial

Inteligencia artificial Según Rich & Knight (1994) “El estudio de cómo lograr que las computadoras realicen tareas que, por el momento, los humanos hacen mejor.” La Inteligencia Artificial como ciencia y tecnología ha ido acumulando conocimientos sobre como emular diversas capacidades del ser humano para exhibir comportamiento vigente y se han desarrollado sistemas cada vez más perfeccionados que producen parcialmente dichas capacidades. La Inteligencia Artificial trata de abarcar y estudiar muchas de las capacidades del hombre para poder ofrecerlas a la máquina y también al propio hombre en el entendimiento de los principios de su inteligencia. Ejemplos: agentes, lógica, redes neuronales, sistemas expertos, vida artificial, tecnología del habla y síntesis, visión artificial, robótica entre otros.

1.2 Antecedentes

Aguilar (2019), usa el análisis de componentes principales para datos sets tabular donde observa una mejora respecto a la disminución de dimensiones y aumento de performance de aprendizaje. Ya que, de los 15 data sets, 53.33% obtuvieron una mejora en los datos de test usando PCA en comparación con el 33.33% usando el método clásico. Asimismo, concluye que ambos modelos obtuvieron resultados iguales en un total de 13.33%.

Crombet, Abalos y Zamora (2018), en su estudio de parámetros fisicoquímicos concluyen que las aguas residuales de la comunidad universitaria Antonio Maceo se caracterizan por una gran variabilidad en los valores de sus parámetros fisicoquímicos y forman dos

componentes principales: DQO, DBO5, turbidez, HS-, NH4+, AyG, ST y STF y logran explicar el 87,31 % de la variabilidad total contenida en las aguas de la residencia.

Ruiz (2012), usa el análisis de componentes principales, para transformar un conjunto de indicadores simples de calidad ambiental a un nuevo conjunto de indicadores compuestos incorrelacionados o componentes. Se obtuvieron tres componentes principales, la primera agrupó cinco indicadores referidos a calidad de vida en el hogar, la segunda agrupó a dos indicadores referidos a la temperatura ambiental y la tercera agrupó a tres indicadores de calidad ambiental fuera de la vivienda.

Vias, Perles y Andreo (2003), Concluyen que la cartografía resultante del análisis clúster muestra una distribución espacial de la vulnerabilidad del acuífero similar a la ofrecida por otros métodos, e incluso se observa en algunas zonas del acuífero una mayor capacidad de discriminación espacial. El grado de vulnerabilidad, de las distintas unidades espaciales homogéneas obtenidas a partir del clúster, se ha establecido mediante un análisis de la distancia al punto ideal.

Gasco y Quesada (2007), llegaron a la conclusión que todas las clasificaciones efectuadas poseen como principales elementos de agrupación: el nivel de conductividad, la contaminación biológica y la proporción de iones sodio y cloro con respecto a la concentración total de solutos. además, acota que los parámetros básicos analizados son: pH, conductividad, conductividad por catión, cloruros, sulfatos, alcalinidad, iones calcio, nitratos, oxígeno disponible, fosfatos y temperatura. del estudio se desprende que la conductividad es muy alta.

Calvo, Martínez y Juanatey (2013), indican que los modelos de ecuaciones estructurales (Structural Equation Modeling) son una herramienta estadística que establece modelos de medida y modelos estructurales. Los resultados revelan que el modelo propuesto que incorpora variables latentes como mediadoras, posee mayor número de relaciones significativas que el modelo que no los incorpora. Los resultados obtenidos muestran, sin embargo, algunas limitaciones empíricas debido a que los dos modelos se comparan para un contexto determinado.

Hernández (2016), este artículo presenta una descripción de los principales elementos y componentes de los modelos de ecuaciones estructurales, así como de los valores

estadísticos de referencia para evaluar estos modelos. Además, se muestran los principales modelos desarrollados para el análisis de fatiga humana.

Madrigal, Madrigal y Avila (2019), esta investigación muestra la relevancia del uso de Modelos de Ecuaciones Estructurales (SEM) en estudios enfocados al comportamiento de compra de los jóvenes universitarios de la generación. Millennials. A través de la revisión de artículos científicos, se destaca la importancia y vigencia del análisis multivariable por medio de SEM. Los hallazgos permiten establecer la pertinencia del uso de SEM en la investigación de marketing, ya que se facilita la explicación de teorías del comportamiento humano particularmente la toma de decisiones en el consumo.

Pereira, De Araujo y Moya (2019), indican que el modelo de ecuación estructural (SEM) basado en mínimos cuadrados parciales (PLS) a través de Smartpls R software. Se probaron cinco hipótesis, en las cuales las partes interesadas involucradas tienen en cuenta el SEM el modelo no fue estadísticamente significativo en relación con el desempeño de la logística de la ciudad, lo que indica la necesidad para prestar atención a esto en términos de la alineación estratégica de una nueva definición de los agentes cubiertos en la distribución de bienes en el área urbana. Además, el 56.4% de la variación en el desempeño logístico de la ciudad variable fue explicada por el modelo.

Okamoto, Kojima, Yamagishi, Kato y Tamada (2018), indican que el desarrollo de una estructura multicapa de palabras adjetivas que explica las relaciones semánticas entre las respuestas perceptivas y afectivas humanas a los estímulos es fundamental en el diseño de los aspectos afectivos de los productos.

Pulido, Roldan, Lopez y Gutierrez (2002), indica que la demanda de agua es la magnitud de referencia en la gestión óptima de los sistemas de distribución. En este trabajo se propone la estimación de la demanda en las próximas 24 horas en un sistema de distribución de agua para riego, y se utilizan, junto con los métodos tradicionales de predicción de regresión múltiple y de modelos univariantes de series temporales (ARIMA), las Redes Neuronales Computacionales (RNCs).

Samboni, Carbajal y Escobar (2007), proponen que, para facilitar la interpretación de los datos físico, químicos y biológicos, para determinar la calidad y contaminación del agua las variables que se tienen en cuenta en este proceso son: pH, oxígeno disuelto, Demanda Bioquímica de Oxígeno (DBO), una forma de nitrógeno, fosfatos y sólidos totales (ST).

García, Rodríguez, López y Tenorio (2010), aplicó un modelo de Red Neuronal Artificial para predecir el transporte de contaminantes (cobre y cadmio) en medios saturados, homogéneos e isotrópicos para diferentes clases texturales. Los modelos fueron entrenados y evaluados a partir de la ecuación propuesta por Ogata y Banks, que considera los términos advectivo y difusivo.

Hallasi (2018), realiza investigación en las islas flotantes de los UROS en los meses de mayo a julio para determinar los índices de bacterias coliformes totales, fecales y termotolerantes y concluye que el promedio general de coliformes totales es 4424 NMP/100 ml de mayo a junio. Estos valores evidencian una importante carga bacteriana. Coliformes fecales 27,33 NMP/100 mL que supera los límites A-1 (0 NMP/100 mL). Temperatura 14,74 °C, relativamente templado. pH 8,38. DBO5 8,54 mg/L, importante carga orgánica. Nitratos 0,03 mg/l cumplen con la normativa de los ECAs, porque se encuentran debajo de los LMP que es de 50 mg/l para la calidad 1 A-1. Fosfato 0,04 mg/l las aguas cumplen con la Normativa de los ECAs, por lo tanto, son aptas para consumo humano ya que el límite evidenciado en el (MINAM, 2017) es de 0.1 mg/l para 1 A-1. Sólidos totales disueltos 11147,80 mg/l Al confrontar estos valores con los ECAs (MINAM, 2017) en la Categoría 1 A-1 se aprecia que nuestros valores superan ampliamente lo establecido en la norma, ya que aguas con fines de consumo humano no deberían pasar los SDT de 1000 mg/l, por lo tanto, estas aguas no son adecuadas en este parámetro para el consumo humano. Oxígeno disuelto 6,32 mg/l Al confrontar estos valores con los ECAs (MINAM, 2017) en la Categoría 1 A-1 se aprecia que nuestros valores se encuentran dentro de lo establecido en la norma, ya que aguas con fines de consumo humano deberían pasar el OD de ≥ 6 mg/l, por lo tanto, estas aguas son adecuadas en este parámetro.

Diaz, Alvarado y Calacho (2012), indican que el tratamiento de aguas residuales domésticas en asentamientos humanos dispersos es un problema que compete al desarrollo local sostenible, para el saneamiento de estas aguas han surgido una serie de tecnologías, una de ellas son las denominadas alternativas, cuyas características son adecuadas para esos casos.

Larios, Moeño y Gonzales (2015), indican que la población de América Latina se encuentra concentrada en ciudades en más de un 80%. Sin embargo, la provisión de agua es insuficiente. Más aun, el 70% de las aguas residuales no tienen tratamiento, lo cual

dificulta alcanzar el ciclo del agua, particularmente por el reuso del agua debido a su contaminación. En Perú, solamente se ha ejecutado el 30% de la inversión pública en tratamiento de agua, de acuerdo al Plan Nacional de Saneamiento Urbano y Rural 2006-2015.

Romero, Colin, Sánchez y Ortiz (2009), en su investigación, concluyen que el tratamiento de las aguas residuales es una cuestión prioritaria a nivel mundial, ya que es importante disponer de agua de calidad y en cantidad suficiente, lo que permitirá una mejora del ambiente, la salud y la calidad de vida. En México, debido a la insuficiente infraestructura, los altos costos, la falta de mantenimiento y de personal capacitado, sólo 36 % de las aguas residuales generadas reciben tratamiento, lo cual crea la necesidad de desarrollar tecnologías para su depuración. Los humedales artificiales son una alternativa de tratamiento debido a su alta eficiencia de remoción de contaminantes y a su bajo costo de instalación y mantenimiento.

Beltrán, Palomino, Moreno, Peralta y Montesinos (2011), en la investigación llevada entre diciembre del 2010 y abril 2011 tomas realizadas próximas a la laguna de espinar concluyen los valores de transparencia del agua fueron bajos, la conductividad eléctrica registró valores elevados, los valores de alcalinidad fueron altos (75 – 150 mg/L) y muy altos (>150 mg/L), la dureza total registrada indica aguas duras (121 – 180 mg/L) y muy duras (>180 mg/L), las altas cantidades de coliformes fecales (*E. coli*) en aguas cercanas a la isla Espinar serian el resultado de las descargas de aguas residuales de la ciudad de Puno, sin un tratamiento adecuado.

Romero (2018), concluye que la evaluación de los parámetros físicos y químicos mostró en el punto 1(ZAM) menor concentración de OD con 4.80 mg/L; en el punto 3(ZAP) mayor concentración de nitritos con 0.095 mg/L, mayor DQO con 18.6 mg/L y mayor DBO con 11.5 mg/L; en el punto 4(ZAP) se presentó la mayor profundidad con 25.90 m, mayor transparencia con 5.60 m y el mayor pH con 8.81, en el punto 6(ZSV) se presentó la mayor concentración de nitratos con 0.224 mg/L, en el punto 7(ZSV) se presentó la menor temperatura con 11.3°C, la mayor conductividad con 3092.0 μ S/cm, la mayor salinidad con 1864.0 mg/L, y la mayor concentración de nitratos con 0.142 mg/L, mientras que en el punto 8(ZTR) se presentó el mayor número de coliformes totales con 7.20×10^3 NMP.

Jiménez (2016), concluye que las aguas residuales de la bahía de Puno, presentan valores de 2.21 mgNT/l, y Fósforo Total 1.36 mgP-PO₄/L, estos valores se confirman cuando evaluados los parámetros (pH, conductividad eléctrica, Sólidos totales disueltos, DBO₅ y DQO) nos indican que sobre pasan los límites permisibles, por tanto, la bahía se encuentra en proceso de eutrofización. En cuanto al estudio de iones de Fosforo Total, en aguas residuales en acuarios de cada sector diluidas (50% control + 50%) Elodea ser más eficaz remueve a 29 días y sin diluir hasta los 37 días eficiente en remover fósforo Total 3902.8 mg/Kg.

Callata (2015), realiza estudios en la Bahía Interior de Puno a 3810 m.s.n.m. La investigación se llevó acabo en 07 estaciones de muestreo. La temperatura del agua tuvo valores entre 13 y 15.60 °C, la transparencia llegó hasta 0.81 m, el pH del agua en la superficie tuvo valores entre 8.86 a 9.63 unidades de pH, la conductividad eléctrica 1901 µS/cm, la turbiedad 43 NTU, el oxígeno disuelto presentó valores de 2.10 a 3.28 mg/l, los sólidos disueltos totales 953 mg/l, la DBO₅ 205 mg/l, DQO hasta 514 mg/l, para nitratos 2.085 mg/l y fosfatos 0.333 mg/l. Para los parámetros bacteriológicos de coliformes fecales 2900 NMP/100 ml, coliformes totales 11000 NMP/100 ml y coliformes termotolerantes 1900 NMP/100 ml. De los resultados obtenidos para los parámetros indicadores de contaminación (OD, pH, Sólidos disueltos totales, DBO₅, y coliformes fecales y totales) al ser comparados con los estándares nacionales de calidad ambiental para agua; excedieron grandemente lo recomendado, y también se ha identificado mediante los mapas temáticos que la zona más crítica en la bahía interior de Puno, corresponde zonas cercanas a la isla Espinar y Muelle de Puno.

Carpio (2016) llega a la conclusión de que las aguas de los tributarios presentaron una temperatura entre 11,4 y 13,1 °C, la conductividad eléctrica entre 1738,4 y 1742 µS/cm, el pH entre 7,12 y 8,14 unidades, la turbidez entre leve y elevada, el oxígeno disuelto entre 5,6 y 5,7 mg/l, el color del agua fue aparente en un 80% de las muestras, el olor fue de leve a fétido en mayor porcentaje. En ambas épocas, la biomasa de las macrófitas evaluadas (*E. potamogeton*, *M. elatinoides* y *P. strictus*) no están catalogadas como hiperacumuladoras de metales pesados. Los metales Cd, Cr, Cu, Ni, Pb y Hg presentes en las macrófitas *M. elatinoides* y *P. strictus* no superan los niveles permisibles, mientras que Mn, Al, As, Zn y Hg superan los niveles de 0.1 mg/kg.

Moreno, Argota, Alfaro, Aparicio, Atencio, Goyzueta (2018) concluyen que, en ecosistemas acuáticos, los sedimentos son reservorios de metales pesados cuando sus aguas se encuentran expuestas. El propósito del estudio fue cuantificar metales en sedimentos superficiales de la bahía interior de Puno, Lago Titicaca-Perú. De enero a noviembre del 2016 fueron analizadas las concentraciones de Cu, Zn, Pb, Cd, As y Hg en los sedimentos superficiales de seis estaciones de muestreo ambiental seleccionadas mediante un muestreo no probabilístico por conveniencia. Los metales se analizaron por digestión ácida donde su cuantificación fue mediante espectrometría de plasma inductivamente acoplado con vista axial (ICP-AES). Al comparar los resultados con la norma ambiental seleccionada (Interim Sediment Quality Guidelin, Canadá).

Arohuanca (2105) indica que la concentración de nitrógeno y fósforo, en los efluentes del lago Titicaca de las ciudades de Puno, Juliaca, Ayaviri, Yunguyo, fueron evaluadas en dos estaciones una en el mes de octubre y las concentraciones de nitrógeno total en el mes de octubre 48 ; 39.1 ; 18.4 ; 0.446 mg/L respectivamente, en el mes de febrero presentaron las siguientes concentraciones de 54.6 ; 42.9 ; 34.4 mg/L; no se detectó la concentración (hay que utilizar otro reactivo de más baja concentración de nitrógeno total), la planta de tratamiento de la ciudad de Puno (EMSA –PUNO) no cumple con este parámetro, la concentración de fósforo total en el mes de octubre son 12.23 ; 13.2 ; 9.8 ; 4.1 mg /L, y en el mes de febrero presentaron las siguientes concentraciones de 27.6 ; 12.6 ; 9.8 ; 2.7 mg P/L, ninguna planta de tratamiento cumple con este parámetro, según LMPs dados por SUNASS. Por lo que se afirma la hipótesis de la investigación que los efluentes de las plantas de tratamiento se encuentran alteradas.

Escobar (2019), concluye que los parámetros físico-químicos del agua no superan los límites de tolerancia de los ECA. Así, el valor del pH del agua (8.79) está dentro de rango 6,0-9,0 UpH; el OD en agua ($6.81 > 5$ mg/L). A pesar de las diferencias significativas de variabilidad entre jaulas y blanco $p < 0.05$; se optimizó el mejor modelo estadístico de parámetros de dependencia y el nivel de muestreo a 1.20 m de profundidad. La concentración de metales pesados Hg, Cd, Pb, Zn, Cu en agua y sedimentos están dentro de los límites permisibles comparado con la norma canadiense, excepto As que presenta un efecto probable ($41.1 > 17$ mg/Kg), sin superar el valor ECA (50 mg/kg). Las tendencias de crecimiento de la actividad productiva de truchas en el lago, comprometerán la calidad de agua y por ende la sustentabilidad del Titicaca.

Jiménez (2016) llega a las siguientes conclusiones, las aguas residuales de la bahía interior de Puno presentan en promedio 2.21 mg NT/L, y en Fósforo Total 1.36 mg PT/L, los intervinientes evaluados, presentan valores altos, por tanto, la bahía interior de Puno esta eutrofizado. Por otra parte; la eficiencia de macrófitos en acuarios en aguas diluidas, *Myriophyllum* es más eficiente en absorber Nitrógeno Total (NT) hasta 35 días como óptimo. Por otro lado, se evaluó, *Elodea canadensis*, presenta más eficaz para remover iones de fósforo Total a los 29 días con dilución, y sin diluir hasta los 37 días. Este método es una alternativa para remediar aguas de la bahía interior de Puno.

Cuevas (2018), concluye que los parámetros fisicoquímicos como bioindicadores en la calidad de aguas del CITT Chucuito - Puno registraron: en la temperatura del agua se registró un promedio de 13° C, una máxima de 14.5°C en la Z3 a 8 m de profundidad y una mínima de 11.6°C en la Z3 a 8 m de profundidad. Localizamos valores altos de pH, en promedio 8.6 un mínimo de 8.0 y un máximo de 9.5, asumiendo los valores de la transparencia del agua un máximo de 4 m en la Z2 a 4 m de profundidad y un mínimo de 1 m en la Z1 a 2 m de profundidad. En definitiva, el O.D. registro un valor promedio de 9.0 mg/l se ha tenido un mínimo de 7.0 mg/l en la Z2 a 4 m de profundidad y un máximo de 10.5 mg/l en la Z3 a 8 m de profundidad.

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1 Identificación del problema.

Uno de los problemas que enfrenta la ciudad de Puno, es la contaminación de la bahía interior del lago Titicaca. Este tema es enfocado por muchos gobiernos nacionales, regionales y locales quienes incluyen dentro de sus campañas políticas diferentes alternativas de solución, pero, hasta la fecha solo son ofrecimientos. Por otro lado, la contaminación se ha venido incrementando de forma alarmante en los últimos años debido a las descargas de nutrientes, fósforos y nitratos y otros por causas humanas, siendo la bahía interior del lago Titicaca la zona más contaminada donde se perciben aguas de muy baja calidad y alta peligrosidad, con temperaturas muy altas (hasta 20°C), baja transparencia (menor de 0.5 m), concentración alta de nutrientes (nitrógeno y fósforo disueltos), sobresaturación de oxígeno disuelto en aguas superficiales, demanda biológica de oxígeno muy elevada y niveles de oxígeno excesivos (3mg/l en la superficie y 0 en la profundidad a 1m), presencia de grandes volúmenes de lemna SP y turbidez en el agua causada por materiales sólidos causando alteraciones en los mecanismos fotosintetizadores (Tapia, 2017), por otra parte, el Organismo de Evaluación y Fiscalización Ambiental (OEFA) afirma que solo el 32% de las aguas residuales en las redes de alcantarillado recibe tratamiento. Existen diferente tipo de modelos para establecer relaciones entre factores entre ellos redes neuronales, árboles y ecuaciones estructurales. por esta razón, la pregunta de investigación queda formulada de la siguiente manera:

2.2 Enunciado del problema

El problema de investigación queda de la siguiente manera:

¿Cómo es el comportamiento y en qué medida se dan las relaciones entre los parámetros físico químicos de la bahía interior del lago Titicaca?

Interrogantes específicas

- 1) ¿Cuál es el comportamiento de las parámetros físico-químicos de la bahía interior del lago Titicaca?
- 2) ¿Cuáles son las relaciones entre las variables observadas entre los parámetros físicos químicos de la bahía interior del lago Titicaca?
- 3) ¿Cuáles son variables latentes (factores) de los parámetros físicos químicos de la bahía interior del lago Titicaca?
- 4) ¿Cómo son las relaciones entre las variables observadas y latentes utilizando ecuaciones estructurales entre los parámetros físicos químicos de la bahía interior del lago Titicaca?

2.3 Justificación

La contaminación de los espacios naturales se ha incrementado notablemente en la actualidad debido a los avances tecnológicos e industriales que el hombre pone en práctica, todo esto en búsqueda del desarrollo de la sociedad. Existe hoy en día diversos tipos de contaminación que el hombre utiliza de manera descontrolada sin tomar conciencia de los daños ocasionados al medio ambiente.

La FAO define las aguas residuales como: “las aguas residuales son cualquier tipo de agua cuya calidad se vio afectada negativamente por influencia antropogénica. Las aguas residuales incluyen las aguas usadas domésticas y urbanas, y los residuos líquidos industriales o mineros eliminados, o las aguas que se mezclaron con las anteriores (aguas pluviales o naturales). Su importancia es tal, que requiere sistemas de canalización, tratamiento y desalojo, su tratamiento nulo o indebido genera graves problemas de contaminación. Las llamadas aguas negras son las aguas residuales que están contaminadas con heces u orina.

Modelar los parámetros permitirá establecer relaciones entre variables y así poder buscar soluciones para la descontaminación.

2.4 Objetivos

2.4.1 Objetivo General

Modelar el comportamiento y las relaciones entre los parámetros físico-químicos de la bahía interior del lago Titicaca.

2.4.2 Objetivos Específicos

- 1) Analizar el comportamiento de los parámetros físico-químicos de la bahía interior del lago Titicaca.
- 2) Identificar las relaciones entre las variables observadas entre los parámetros físico-químicos de la bahía interior del lago Titicaca.
- 3) Identificar las variables latentes (factores) de los parámetros físicos químicos de la bahía interior del lago Titicaca.
- 4) Identificar las relaciones de variables observadas y latentes utilizando ecuaciones estructurales entre los parámetros físico-químicos de la bahía interior del lago Titicaca.

2.5 Hipótesis

2.5.1 Hipótesis General

Las ecuaciones estructurales y los árboles de predicción modelan adecuadamente el comportamiento y muestran relaciones significativas entre los parámetros físico-químicos de la bahía interior del lago Titicaca.

2.5.2 Hipótesis específicas

- El comportamiento de los parámetros físico-químicos de la bahía interior del lago Titicaca son estacionales.
- Las relaciones entre las variables observadas entre los parámetros físico-químicos de la bahía interior del lago Titicaca siguen un modelo no paramétrico.



- 2 son las variables latentes independientes y 2 las variables latentes dependientes que representan a los parámetros físico-químicos de la bahía interior del lago Titicaca.
- El modelo estructural representa adecuadamente el comportamiento y las relaciones entre los parámetros físicos químicos observados y latentes de la bahía interior del lago Titicaca.



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 Lugar de estudio

La región Puno se encuentra localizada en la sierra del sudeste del Perú, en la meseta del Collao entre $13^{\circ}00'66''00''$ y $17^{\circ}17'30''$ de latitud sur y los $71^{\circ}06'57''$ y $68^{\circ}48'46''$ de longitud oeste del meridiano de Greenwich. Limita por el Sur, con la región Tacna; por el Este, con la República de Bolivia y por el Oeste, con las regiones de Cusco, Arequipa y Moquegua y por el Norte, con Madre de Dios. La región Puno posee altitudes variadas, se encuentra en el altiplano entre los 3,812 y 5,500 msnm, entre la ceja de selva y la selva alta, entre los 4,200 y 500 msnm. Cabe mencionar que la capital, Puno, está ubicada a orillas del Lago Titicaca, así el estudio de investigación se realizó en la capital de la provincia de Puno.

3.2 Población

Está conformada por todos los parámetros fisicoquímicos que puedan ser obtenidos de la bahía interior del lago Titicaca en el tiempo.

3.3 Muestra

La muestra para el siguiente trabajo de investigación está conformada por datos entre enero del 2013 hasta abril del año 2019, siendo 76 meses para cada variable en estudio. Los datos fueron tomados del laboratorio Continental del IMARPE (Instituto del mar del Perú).

3.4 Método de investigación

El método de investigación es descriptivo porque en su primera parte describe el comportamiento de las variables utilizando la estadística descriptiva y explicativa porque busca relaciones entre variables observadas y transformadas.

3.5 Descripción detallada de métodos por objetivos específicos

Los métodos utilizados para cumplir con los objetivos específicos fueron:

Variables en estudio

Tabla 2

Operacionalización de variables de los parámetros físico químicos del lago Titicaca.

NOMBRE DE LA VARIABLE	INDICADOR	INDICE
Transparencia	Nivel de transparencia del agua	M
Temperatura	Nivel de temperatura del agua	° C
pH	Nivel de pH del agua	0-14
Conductividad eléctrica	Nivel de conductividad eléctrica del agua	(μ s/cm)
Demanda química de oxígeno	Nivel de demanda química de oxígeno	mg/L
Coliformes totales	Volumen de coliformes totales	NMP/100L
Coliformes termo tolerantes	Volumen de coliformes termo tolerantes	NMP/100L
Clorofila-a	Nivel de clorofila-a	mg/m ³
Demanda bioquímica de Oxígeno	Nivel de DBO del agua	mg/L
Nitrato (NO ₃)	Volumen de Nitrato del agua	mg/L
Nitrito (NO ₂)	Volumen de Nitrito del agua	mg/L
Fosfato (PO ₄)	Volumen de Fosfatos del agua	mg/L
Fosforo (P-T)	Volumen de Fosforo del agua	mg/L
Nitrógeno Amoniacal	Volumen de nitrógeno amoniacal	mg/L
Nitrógeno (N-T)	Volumen de Sulfatos del agua	mg/L
Solidos Totales en suspensión	volumen de solidos totales en suspensión	mg/L
Oxígeno disuelto	Nivel de oxígeno disuelto	mh/mL

Materiales para el análisis estadístico.

Para Analizar el comportamiento de los parámetros físico-químicos, se ha recurrido a la estadística descriptiva, gráficos uni y bivariantes y para evaluar la tendencia y realizar predicciones, se ha recurrido a la técnica de machine learning denominada redes neuronales artificiales.

Para obtener las relaciones entre las variables, se ha recurrido a la técnica machine learnig denominada árboles de predicción, la cual es una técnica no paramétrica que encuentra modelos de predicción.

Para identificar las variables latentes (factores) de los parámetros físico-químicos se ha recurrido a las técnicas del análisis multivariado las cuales son el análisis clúster y el análisis factorial exploratorio.

Para identificar las relaciones de las variables observadas y latentes, se ha recurrido a la técnica del análisis de ecuaciones estructurales utilizando para ello el software AMOS.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Luego de la obtención de datos de los laboratorios del del IMARPE-Puno se procedió a la limpieza de datos y a la aplicación de técnicas estadísticas para explicar la modelización de los parámetros físico químicos, los resultados son los siguientes:

4.1 Muestra y descripción de la base de datos

El estudio se centra en la evaluación de los parámetros físico químicos obtenidos de los laboratorios Continental del IMARPE Puno-Perú. cito en la Av. Circunvalación Sur 1911 de la provincia de Puno. Los datos corresponden a la toma del puerto muelle de la bahía interior del lago Titicaca. La muestra de la investigación está constituida por 76 datos mensuales, entre los años 2013 y abril del 2018.

4.2 Nomenclatura y variables en estudio: Parámetros físico químicos:

- Oxígeno disuelto (mh/ml):
- pH
- Clorofila-a (mg/m^3): pigmento fotosintético de las algas
- SST (mg/l): Total de sólidos en suspensión en miligramos por litro
- DBO5 (mg/l): Demanda bioquímica de oxígeno en miligramos por litro
- PO4 (mg/l): Fosfato en miligramos por litro
- Temperatura: °C.

- DQO (mg/l): Demanda química de oxígeno en miligramos por litro
- Transp (m): Transparencia
- N-NH₃ (mg/l): Nitrógeno amoniacal en miligramos por litro
- P-T: (mg/l): fosforo en miligramos por litro
- N- T (mg/l): Nitrógeno en miligramos por litro
- NO₂ (mg/l): Nitrito, concentración de nitrito en miligramos por litro
- NO₃ (mg/l): Nitrato, concentración de nitrato en miligramos por litro
- Conectividad Eléctrica (Us/cm):
- Col Term (NMP/100ml): coliformes termotolerantes, número más probable en 100 mililitros.
- Col Totales (NMP/100ml): Coliformes totales, número más probable en 100 mililitros

4.3 Análisis descriptivo y tendencia del parámetro temperatura

La temperatura del agua es un parámetro importante, dada su influencia sobre el desarrollo de la vida acuática como sobre las reacciones químicas y velocidades de reacción. La temperatura es un indicador de la calidad del agua, que influye en el comportamiento de indicadores de calidad del recurso hídrico, como el pH, el déficit de oxígeno, la conductividad eléctrica y otras variables fisicoquímicas (DIGESA, 2011). Para observar su comportamiento, se muestra las tendencias de la variable temperatura por mes y años.

La figura 7 muestra que, en el año 2013, en el mes de junio y julio se presentaron los niveles de temperatura más bajas y en los meses de octubre, noviembre y diciembre los niveles de temperatura más altas.

En forma general se puede apreciar que la temperatura en el lago tiene una tendencia poco cambiante en el periodo estudiado, los meses donde la temperatura del lago baja son mayo, junio, julio, agosto y setiembre (líneas punteadas), presentando el mes de

septiembre una temperatura intermedia. La temperatura aumenta en los meses de octubre, noviembre, diciembre, enero, febrero, marzo y abril.

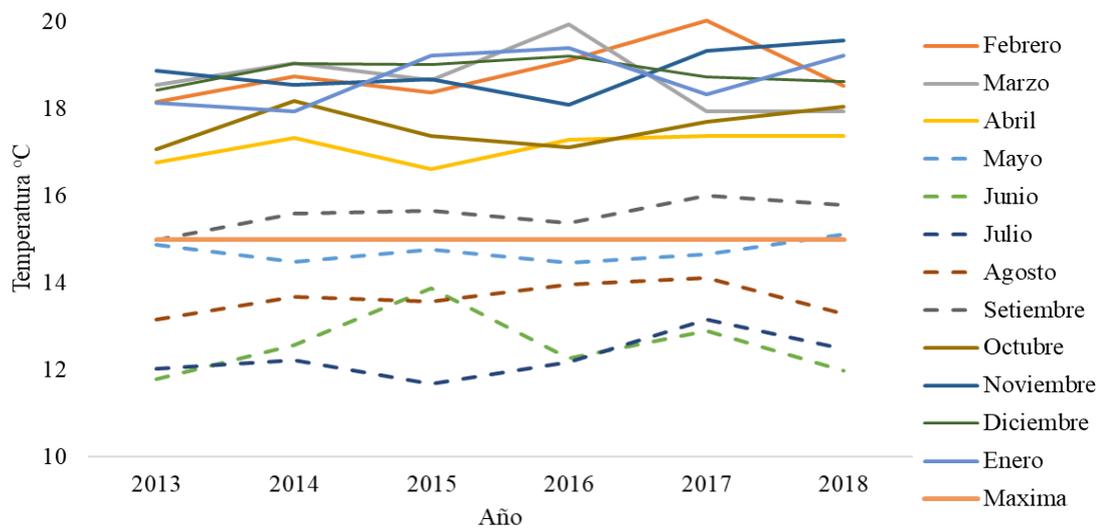


Figura 7. Valores observados para la temperatura de la bahía interior del lago Titicaca por mes según año, periodo enero 2013 – diciembre 2018.

En la figura 7, la línea continua perpendicular sin oscilaciones corresponde a 15 °C que según la DIGESA sería la temperatura máxima aceptable.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada temperatura, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca, toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 3

Estadísticos descriptivos para la temperatura por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

Estadísticos	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Promedio	18.77	18.80	18.56	17.16	14.73	12.57	12.30	13.63	15.57	17.58	18.85	18.83
Desv Est	±0.61	±0.62	±0.75	±0.32	±0.24	±0.75	±0.50	±0.37	±0.35	±0.47	±0.53	±0.29
Mín	17.94	18.16	17.87	16.62	14.47	11.80	11.68	13.15	14.99	17.08	18.10	18.43
Max	19.39	20.03	19.93	17.37	15.11	13.87	13.16	14.11	16.01	18.17	19.57	19.20

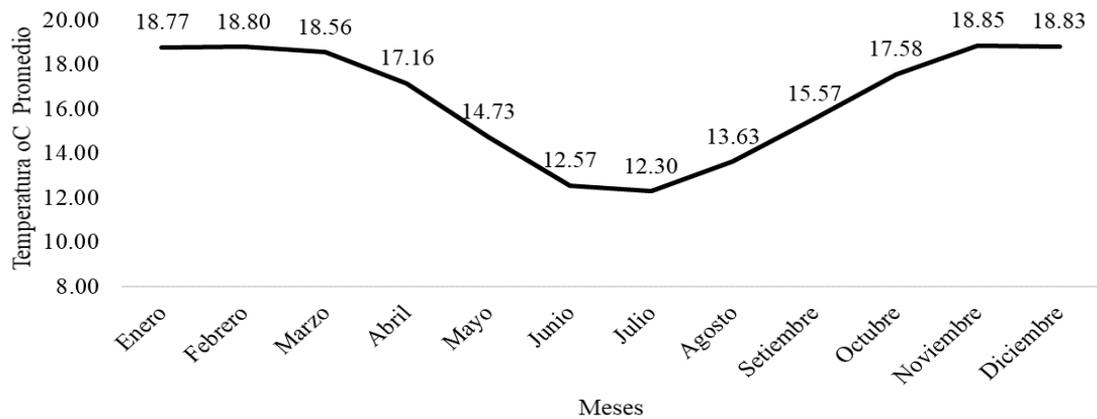


Figura 8. Valores promedio de temperatura del lago por meses de la bahía interior del lago Titicaca, periodo enero 2013 – abril 2019.

La tabla 3 y figura 8 muestra que la temperatura promedio en el periodo de estudio, tiene una tendencia decreciente desde el mes de marzo, descendiendo paulatinamente en los meses de abril, mayo y junio, llegando al mes de julio con valor mínimo y pasa a un patrón ascendente desde el mes de agosto hasta diciembre en el periodo de un año.

Análisis de la tendencia. La figura siguiente muestra la tendencia del parámetro temperatura, tomada en el puerto muelle de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente a los datos.

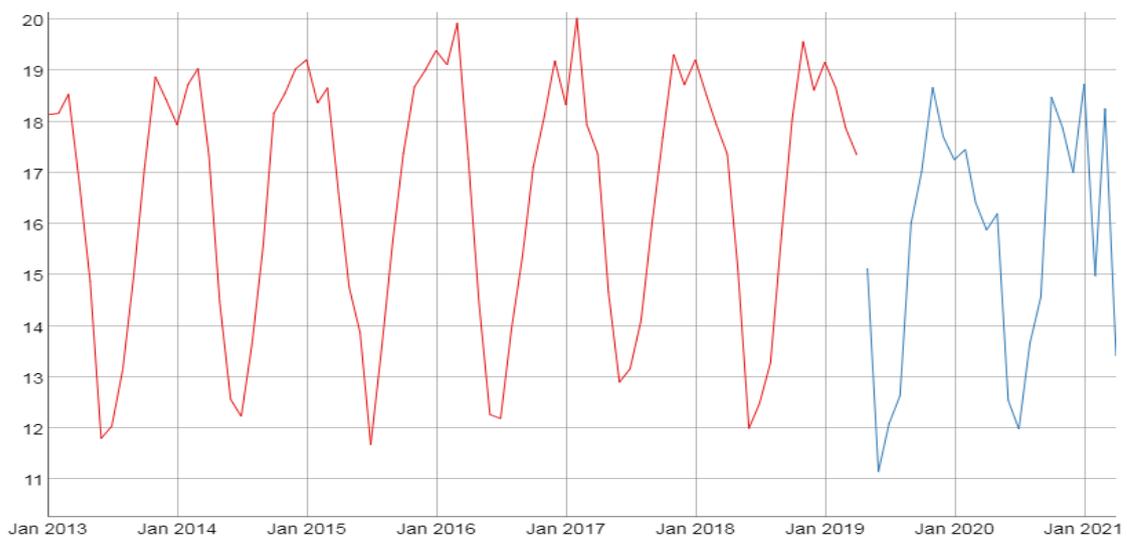


Figura 9. Tendencia y valores predichos para el parámetro Temperatura de la bahía interior del lago Titicaca, toma puerto muelle.

La figura 9 muestra la tendencia del parámetro temperatura en forma mensual desde enero del 2013 hasta marzo del 2019 y las proyecciones hasta el 2021 (desde la parte entrecortada). La tendencia tiene un efecto estacional durante el año, y cuya proyección muestra que el patrón continuará a lo largo de los dos años siguiente (ver proyecciones anexo 1). Para obtener el modelo de red neuronal se usó 11 retardos y 5 nodos, con un RMSE de 0.0859.

Beltran et al. (2011) reporta valores inferiores (diciembre: 16.38, enero:15, febrero:14.67, marzo: 16.5 y abril: 15.83, promedio: 15.68) a los encontrados en la presente investigación, es decir para el año 2013 la temperatura del lago se ha incrementado a un promedio mayor a de 18 °C frente a la etapa de comparación. En los meses de mayo, junio, julio y agosto la temperatura del lago se torna peligrosa para los bañistas ya que alcanzan temperaturas menores a 15°C. (DIGESA). Los resultados que se presentan en la isla los uros (14.74 °C) en los meses de mayo a julio son similares a los del presente estudio, concluyendo una temperatura relativamente templada (Hallasi, 2018)

4.4 Análisis descriptivo y tendencia del parámetro Coliformes Totales

Los coliformes totales se definen como bacterias Gram negativas en forma bacilar que fermentan la lactosa a temperatura de 35 a 37 °C y producen ácido y gas (CO₂) en 24 h, aerobias o anaerobias facultativas, son oxidasa negativa, no forman esporas y presentan actividad enzimática β- alactosidasa. Entre ellas se encuentran *Escherichia coli*, *Citrobacter*, *Enterobacter* y *Klebsiella* (Larrea et al., 2013).

Los coliformes totales pueden hallarse tanto en heces como en el medio ambiente, por ejemplo, aguas ricas en nutrientes, suelos, materias vegetales en descomposición. También hay especies que nunca o casi nunca se encuentran en las heces pero que se multiplican en el agua (DIGESA, 2011).

La figura siguiente muestra la tendencia de la variable coliformes totales con respecto al tiempo, para ello se clasifican los datos por meses y años.

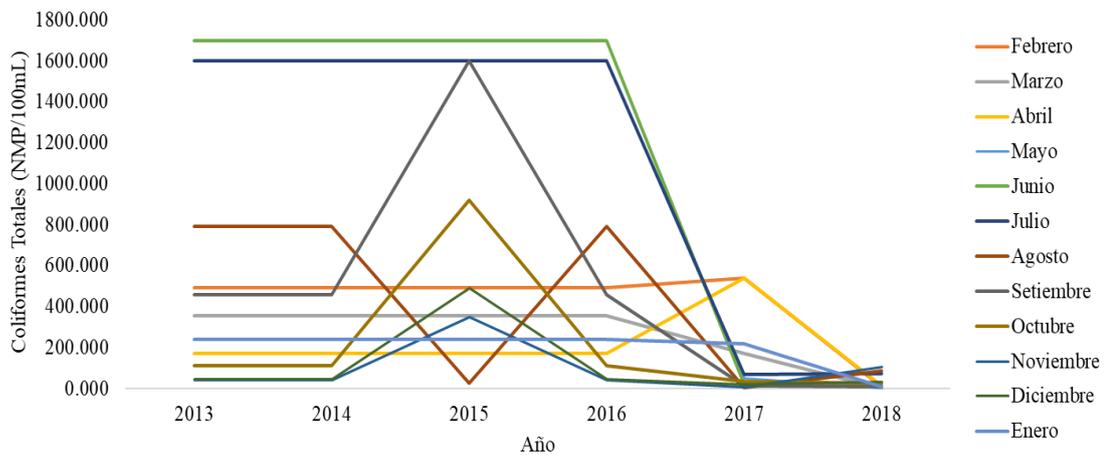


Figura 10. valores observados para coliformes totales de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013, en los meses de diciembre y enero se presentaron los niveles más bajos de la concentración de coliformes totales y en los meses de junio y julio los niveles de concentración de coliformes totales más alta, esto se repite en el año 2014. En el año 2105, existe una variación en los niveles de concentración de coliformes totales, tal es así que el nivel mínimo se presentó en el mes de agosto. Se puede apreciar que en los años 2017 y 2018 los niveles de concentración de coliformes totales han disminuido en todos los meses.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada coliformes totales, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos. Para este efecto, se ha consolidado todo el periodo.

Tabla 4

Estadísticos descriptivos para coliformes totales por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

Estadíst	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Prom	170.51	489.69	242.40	226.71	1075.17	1139.33	1090.25	414.75	499.63	218.00	97.30	113.00
DE	112.08	264.55	148.73	170.08	813.20	868.59	789.70	411.95	582.61	346.20	127.90	184.96
Max	240.00	920.00	355.00	540.00	1600.00	1700.00	1600.00	790.00	1600.00	920.00	350.00	490.00
Min	6.80	7.80	1.80	12.00	2.00	13.00	70.00	4.50	6.80	25.00	7.80	20.00

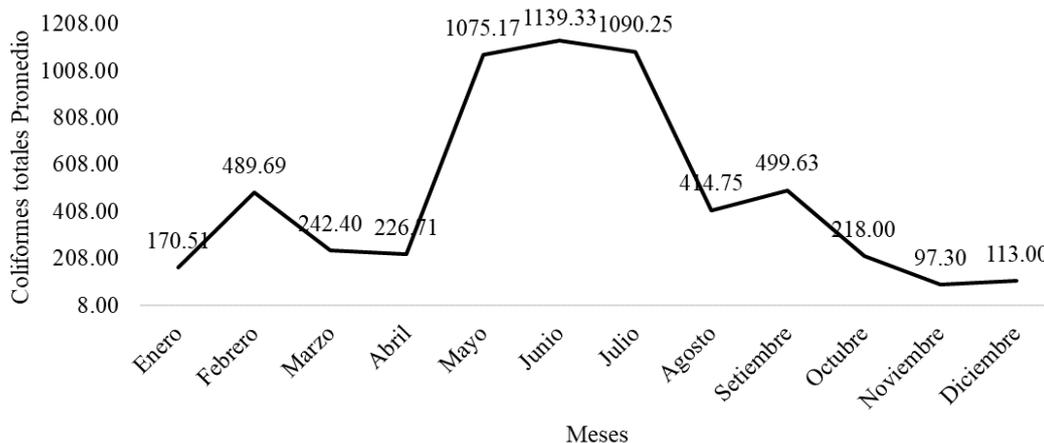


Figura 11. Valores promedio de coliformes totales por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 4 y figura 11 muestra que la concentración de coliformes totales promedio en el periodo de estudio, tiene una concentración mayor en los meses de mayo, junio y julio, descendiendo en los meses de enero, febrero, marzo, abril agosto, setiembre, octubre, noviembre y diciembre.

Análisis de la tendencia. La figura siguiente muestran la tendencia de la concentración de coliformes totales de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales.

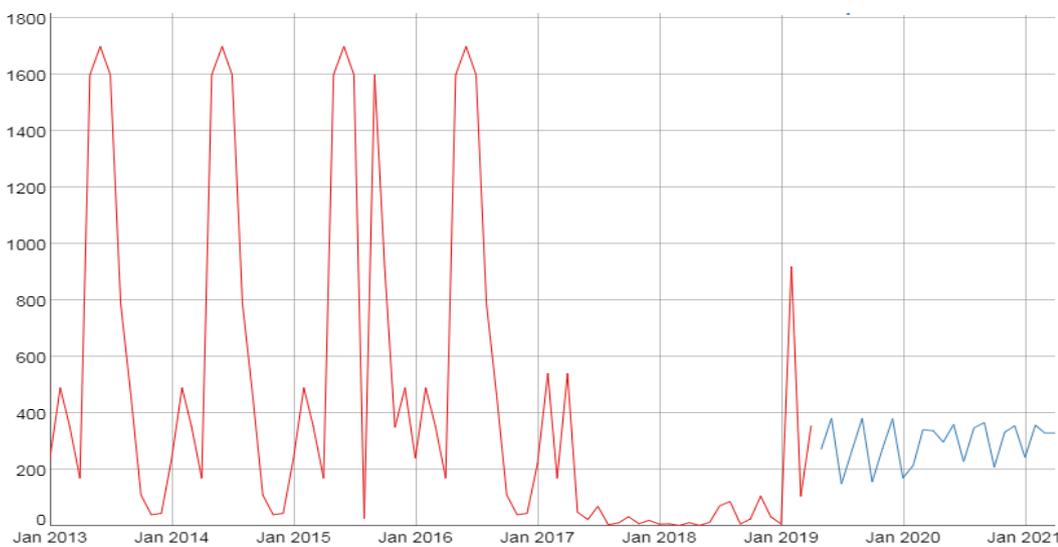


Figura 12. Tendencia y valores predichos para el parámetro coliformes totales de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro temperatura en forma mensual desde enero del 2013 hasta marzo del 2019 y las proyecciones hasta el año 2021. La tendencia indica que los valores han disminuido a mediados del año 2017 excepto un valor en el año 2019. Las proyecciones indican que la oscilación se mantendrá en niveles entre 150 y 400. El modelo de red neuronal se usó 1 y 12 retardos y 8 nodos, con un RMSE de 190.469.

Hallasi (2018) en la isla los Uros obtienen 4424 NMP/100 ml de coliformes totales indicando una importante carga bacteriana, que difiere de los resultados de la toma del puerto muelle de Puno. si basamos los resultados promedio con los de DIGESA en el cual A1: Basados en la Comunidad Económica Europea – CEE y Ecuador se propone 50 NMP/100ml como estándar y A2: Basados en la Comunidad Económica Europea – CEE se propone 5000 NMP/100 ml como estándar, los valores estarían por encima de A1, pero por debajo de A2.

4.5 Análisis descriptivo y tendencia del parámetro Oxígeno Disuelto

El oxígeno disuelto se puede explicar como la medida de la concentración de oxígeno en el agua, usando como referencia el 100% de saturación de oxígeno en el aire. La cantidad de oxígeno disuelto en agua tiene una gran incidencia en el desarrollo de la vida y de muchos procesos que se dan en el medio acuático. Los organismos vivos necesitan oxígeno para mantener su metabolismo, y su captación se realiza a través de la respiración. Por este motivo, el oxígeno ha sido siempre una medida imprescindible en los estudios de la calidad del agua.

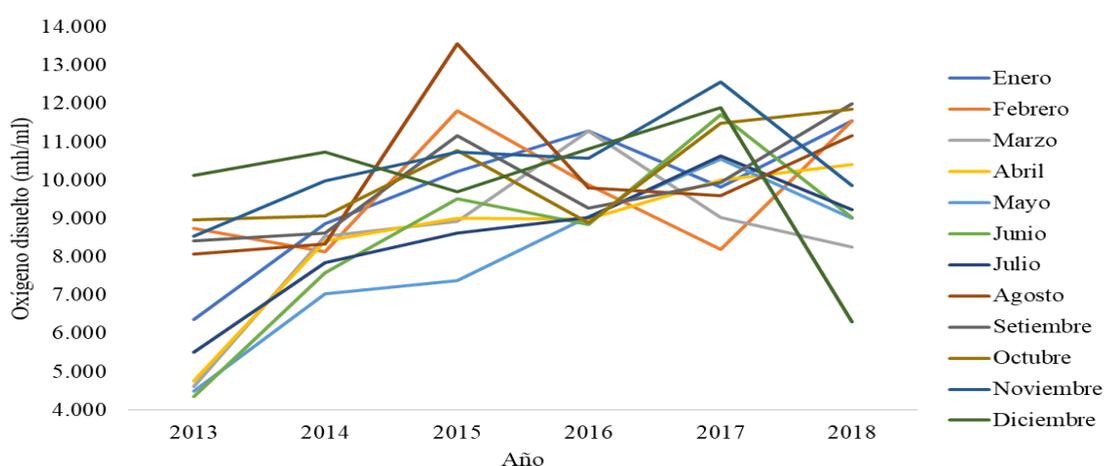


Figura 13. valores observados para oxígeno disuelto de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013, en los meses de abril, mayo, junio y julio se presentaron los niveles más bajos de concentración de oxígeno disuelto y en el mes de diciembre los niveles de concentración de oxígeno disuelto más alta. Los niveles de oxígeno disuelto aumentan desde el año 2014 y estos valores se mantienen hasta el 2018. la figura muestra un valor bajo en el 2018 en comparación a los demás valores y esto ocurre en el mes de diciembre del año 2018.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada oxígeno disuelto, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 5

Estadísticos descriptivos para oxígeno disuelto por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

Estadísticos	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Promedio	9.47	9.66	8.76	8.66	7.92	8.50	8.48	10.09	9.91	10.17	10.38	9.92
desv est	1.70	1.40	1.99	1.72	2.11	2.44	1.72	2.04	1.43	1.36	1.32	1.92
Mínimo	11.56	11.82	11.29	10.40	10.56	11.71	10.64	13.57	11.99	11.86	12.57	11.89
Max	6.36	8.14	4.60	4.75	4.49	4.34	5.50	8.06	8.42	8.89	8.54	6.30

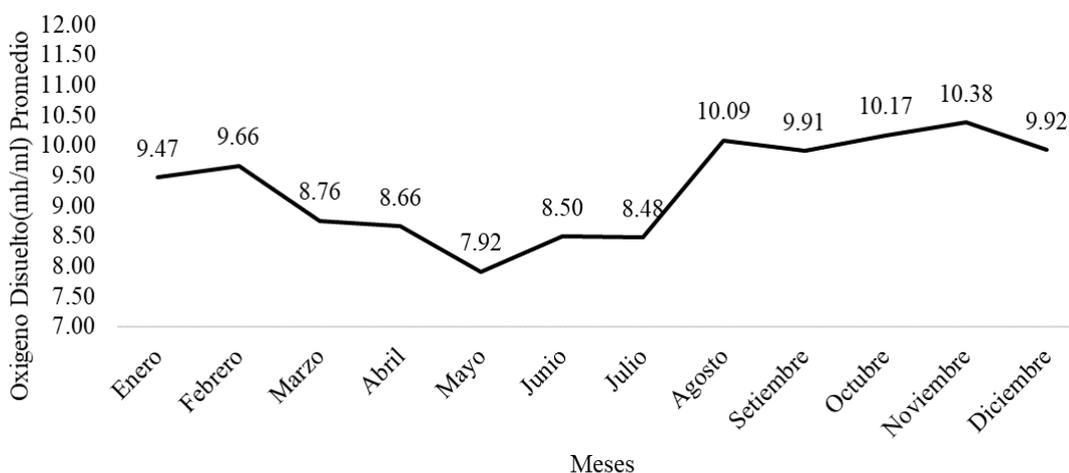


Figura 14. Valores promedio de Oxígeno disuelto por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 5 y la figura 14 muestra que la concentración de oxígeno disuelto promedio en el periodo de estudio (2013-2019) tiene una concentración mayor en los meses de agosto a diciembre respectivamente, muestra valores menores entre los meses de enero a julio, siendo sus puntos más bajos en los meses de abril, mayo, junio y julio.

Análisis de la tendencia. La figura siguiente muestran la tendencia de la concentración de oxígeno disuelto de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

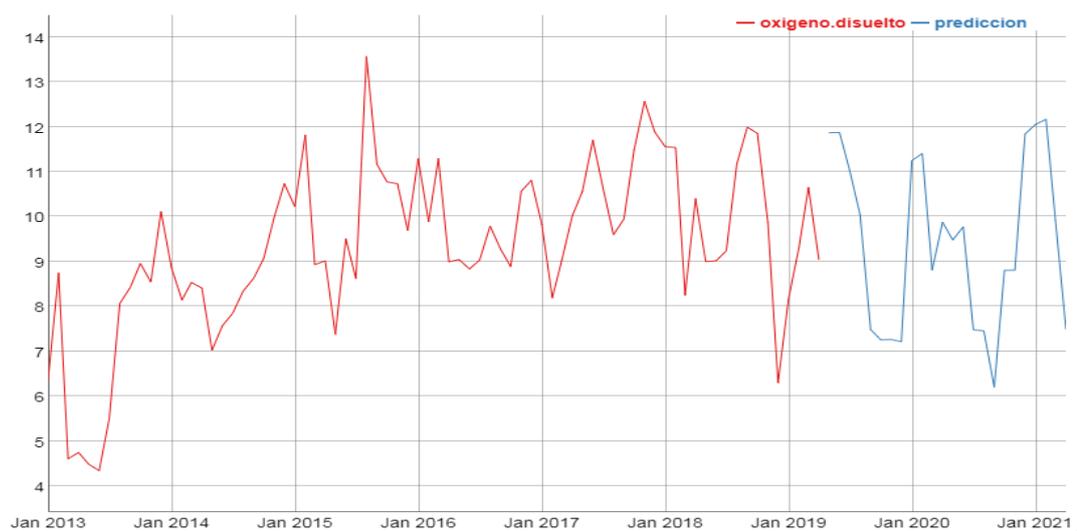


Figura 15. Tendencia y valores predichos para el parámetro oxígeno disuelto de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro oxígeno disuelto de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica que a partir del año 2014 este indicador ha aumentado y las predicciones muestran que la tendencia mantendrá los niveles actuales. Para obtener el modelo de red neuronal se usó 1,3,6,8 y 9 retardos y 5 nodos, con un RMSE de 0.6694.

Hallasi (2018) obtiene concentración de oxígeno disuelto promedio 6,32 mg/l y concluye que estos valores se encuentran dentro de lo establecido en la norma, ya que aguas con fines de consumo humano deberían pasar el OD de ≥ 6 mg/l, por lo tanto, estas aguas son adecuadas en este parámetro, resultados que concuerdan con la presente investigación en la cual los valores sobrepasan los 6,32 mg/l.

4.6 Análisis descriptivo y tendencia del parámetro Coliformes termotolerantes

Los coliformes termotolerantes (CF), denominados así porque soportan temperaturas hasta de 45 °C, comprenden un número muy reducido de microorganismos, los cuales son indicadores de calidad por su origen. En su mayoría están representados por *E. coli*, pero se pueden encontrar de forma menos frecuente las especies *Citrobacter freundii* y *Klebsiella pneumoniae*. Estas últimas forman parte de los coliformes termotolerantes, pero su origen normalmente es ambiental (fuentes de agua, vegetación y suelos) y solo ocasionalmente forman parte del microbiota normal. Por esto, algunos autores plantean que el término de coliformes fecales, comúnmente utilizado, debe ser sustituido por coliformes termotolerantes (Larrea et al, 2013).

Es necesario observar la tendencia de la variable en el tiempo, para ello se clasifican los datos por meses y años, los resultados se muestran a continuación.

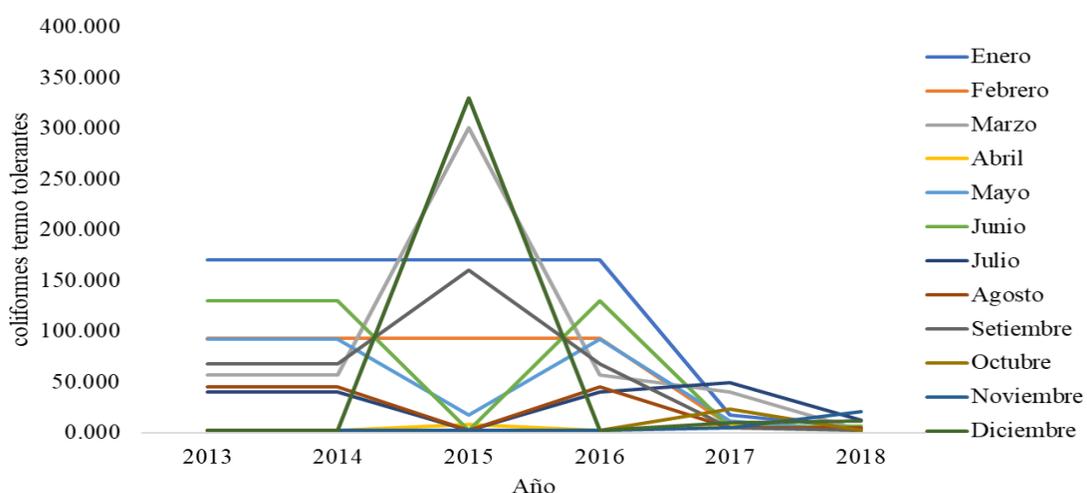


Figura 16. valores observados para coliformes termo tolerantes de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

Los valores mínimos se presentaron en el mes de abril y diciembre del año 2013. En el año 2015 se muestra en el mes de setiembre valores anómalos en algunos meses. Los coliformes termo tolerantes han disminuido a partir del año 2017 y se mantiene con valores bajos en comparación a los demás años.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada coliformes termo tolerantes, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 6

Estadísticos descriptivos para coliformes termo tolerantes por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

Estadísticos	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Prom	100.86	77.61	77.26	4.31	50.97	67.07	30.55	24.33	61.72	5.37	5.45	59.28
De	86.34	54.92	100.29	3.10	45.21	68.95	18.77	22.66	57.62	8.64	7.69	132.69
Max	170.00	160.00	300.00	8.00	92.00	130.00	49.00	45.00	160.00	23.00	21.00	330.00
Min	4.50	4.50	1.80	1.80	1.80	1.80	1.80	2.00	1.80	1.80	1.80	1.80

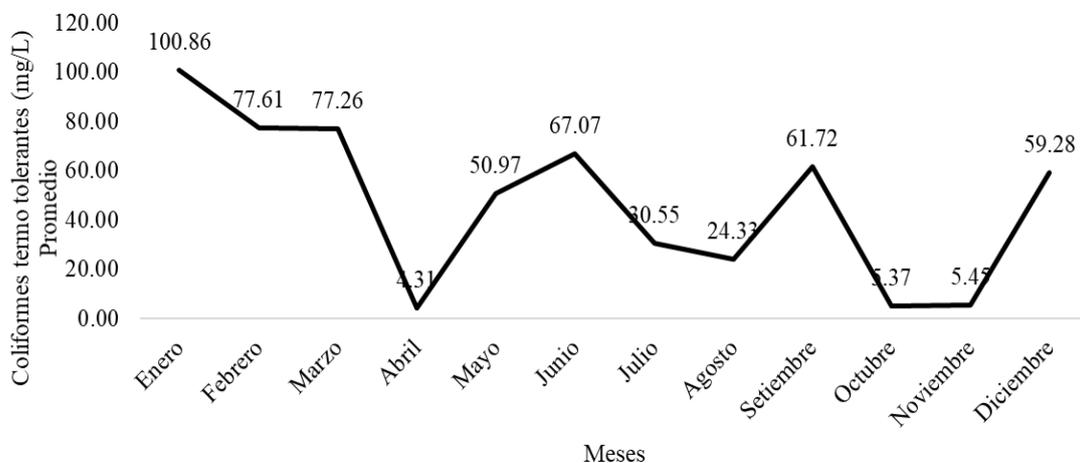


Figura 17. Valores promedio de coliformes termo tolerantes por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 6 y figura 17 muestra que la concentración de coliformes termo tolerantes promedio en el periodo de estudio, se puede notar que no tiene un patrón estacional, pueden observarse valores bajos en los meses de abril, octubre y noviembre. Un valor que se puede entender como anómalo se presenta en el mes de setiembre.

Análisis de la tendencia. La figura siguiente muestra la tendencia de la concentración de coliformes termo tolerantes de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

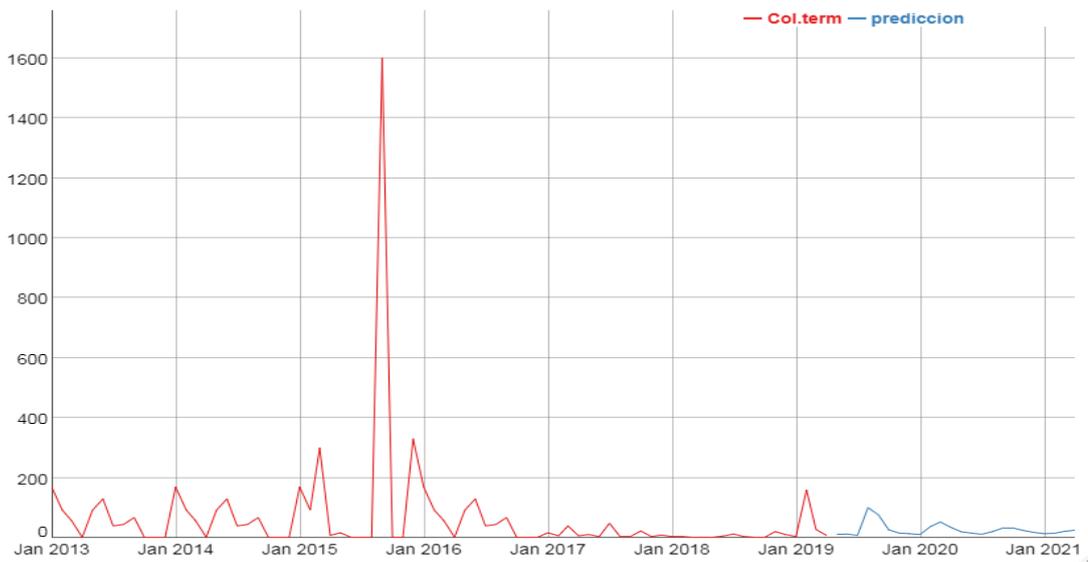


Figura 18. Tendencia y valores predichos para el parámetro coliformes termo tolerantes de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro coliformes termo tolerantes de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica un patrón repetido a través del tiempo, excepto el año 2015. la tendencia muestra que se seguirá manteniendo para los dos próximos años. Para obtener el modelo de red neuronal se usó 1,2,3,4,5, 6 y 7 retardos y 2 nodos, con un RMSE de 151.6256292.

Los termotolerantes diferentes de *Escherichia coli* pueden proceder a aguas orgánicamente enriquecidas como efluentes industriales, de materias vegetales y suelos en descomposición. Por contacto directo pueden infectar heridas, mucosas de ojos y oídos. Por ingestión ocasionan gastroenteritis aguda. A1: Basados en la Comunidad Económica Europea – CEE y Ecuador se propone 20 NMP/100 ml como estándar. A2: Basados en la Comunidad Económica Europea – CEE se propone 2000 NMP/100 ml como estándar (DIGESA). Solo tres meses reportan valores menores a 20 NMP y bajo

las normas MINAM las aguas no estarían aptas para el consumo Humano (0 NMP/100 ml).

4.7 Análisis descriptivo y tendencia del parámetro pH.

El pH es un importante parámetro operativo de la calidad del agua. Las aguas demasiado ácidas disuelven los metales empleados en las conducciones (plomo, cobre, zinc), los cuales, al ser ingeridos, afectan negativamente la salud. El pH aceptable para agua potable varía entre 6.5 a 8.5 como valor guía (Jiménez, 2001). Según Galvín (2003), para las aguas de consumo humano, los valores extremos pueden causar irritación en las mucosas, irritación en órganos internos y hasta procesos de ulceración.

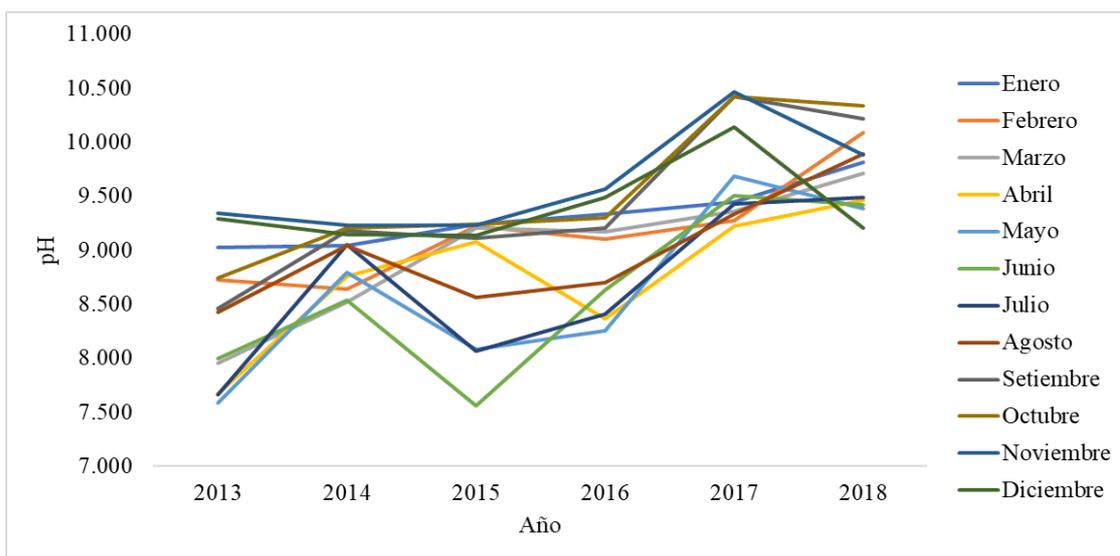


Figura 19. valores observados para pH de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de mayo, junio y julio y el valor máximo en el mes de noviembre y diciembre. La tendencia de los valores de pH ha ido en aumento hacia el año 2017 y 2018 que presentan valores mayores a 9.

La tabla 7 contiene los principales estadísticos descriptivos de la variable observada pH, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 7

Estadísticos descriptivos para pH por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Prom	9.449	9.252	9.070	8.856	8.633	8.610	8.686	8.994	9.435	9.543	9.622	9.403
De	0.438	0.509	0.619	0.649	0.804	0.766	0.752	0.552	0.744	0.680	0.484	0.386
Max	10.237	10.088	9.713	9.462	9.684	9.508	9.490	9.895	10.428	10.423	10.467	10.144
Min	9.026	8.643	7.958	7.669	7.587	7.563	7.666	8.431	8.464	8.745	9.231	9.139

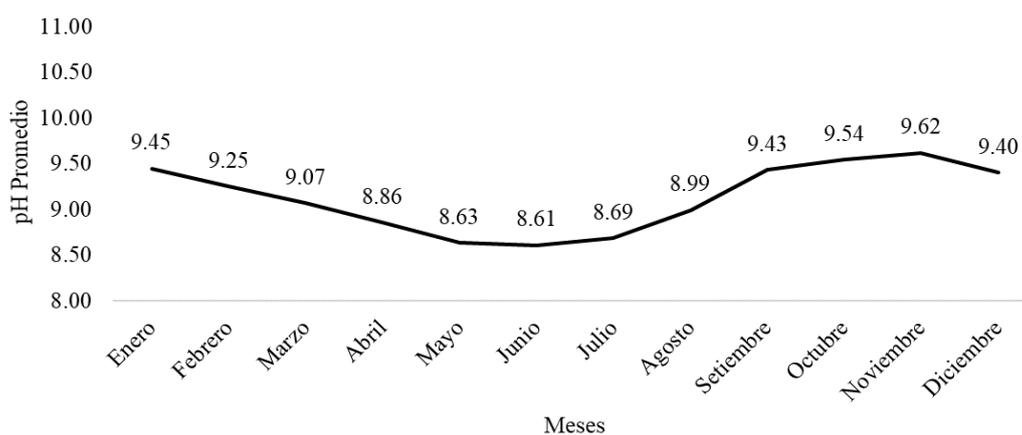


Figura 20. Valores promedio de pH por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 7 y figura 20 muestra que la concentración de pH promedio en el periodo de estudio, se puede notar que tiene un patrón estacional, pueden observarse valores bajos en meses de abril, mayo junio julio y agosto y valore altos en los meses de setiembre, octubre, noviembre, diciembre, enero y febrero.

Análisis de la tendencia. La figura siguiente muestra la tendencia de la concentración de pH de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.



Figura 21. Tendencia y valores predichos para el parámetro pH de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro pH de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica que en el tiempo los valores han aumentado a partir del año 2016, muy por encima del valor neutro. La tendencia tiende a mantenerse y disminuir paulatinamente. Para obtener el modelo de red neuronal se usó 1,5,8,10,11 y 12 retardos y 5 nodos, con un RMSE de 0.020126897.

Hallasi (2018), reporta un pH de 8.38 en la Isla los Uros, valores similares a la toma de puerto muelle de puno que en todo caso sobrepasan el nivel máximo permitido 8.5 (Jiménez, 2001) (DIGESA, 2011) que indica no apta para consumo humano.

4.8 Análisis descriptivo y tendencia del parámetro Nitrito

Las heces de los animales contienen proteína no asimilada (nitrógeno orgánico) y las proteínas que queda en los cuerpos de los animales y plantas que mueren se convierten en gran medida en amoniaco por acción de las bacterias heterótrofas, en condiciones aeróbicas y anaeróbicas. El amoniaco liberado por la acción de las bacterias sobre la urea y las proteínas es utilizado por las plantas. Si se libera en exceso es oxidado por las bacterias(nitrosomas) que incondiciones aeróbicas convierten el amoniaco a nitrito (DIGESA, 2011).

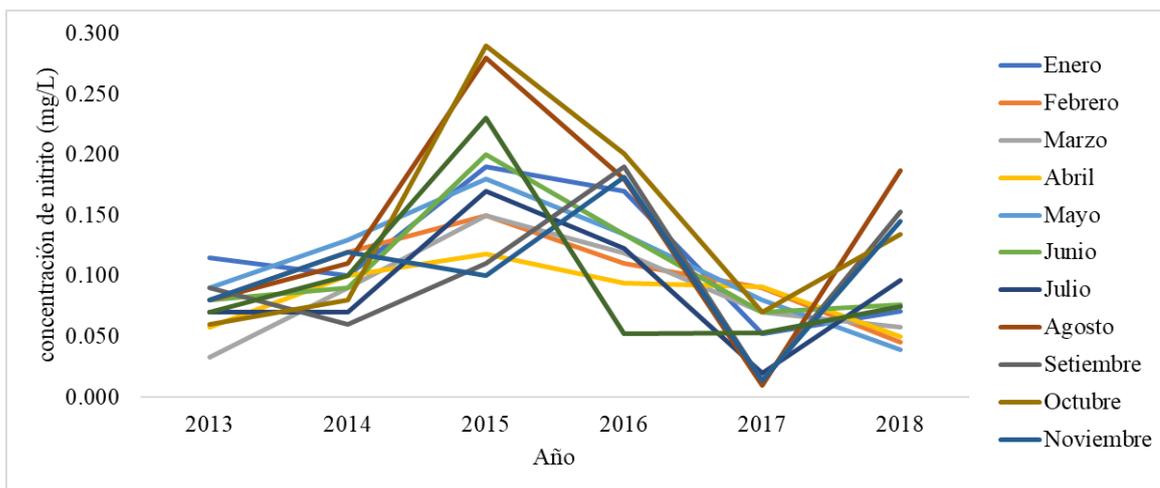


Figura 22. valores observados para nitrito de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de marzo, abril y octubre y el valor máximo en el mes de noviembre y enero. Los valores de nitrito aumentaron en el año 2015 para luego disminuir ala año 2017, la tendencia es nuevamente al aumento de nitrito en el lago.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada Nitrito, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 8

Estadísticos descriptivos para Nitrito por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	0.11	0.09	0.10	0.09	0.11	0.11	0.09	0.14	0.10	0.14	0.11	0.10
de	0.05	0.03	0.04	0.03	0.05	0.05	0.05	0.09	0.06	0.08	0.05	0.06
max	0.19	0.15	0.16	0.13	0.18	0.20	0.17	0.28	0.19	0.29	0.18	0.23
min	0.05	0.05	0.03	0.05	0.04	0.07	0.02	0.01	0.01	0.06	0.01	0.05

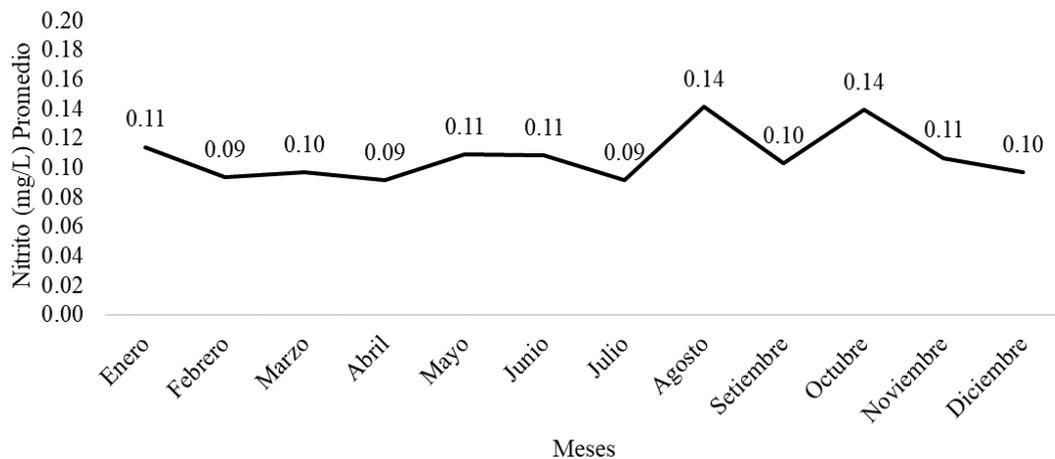


Figura 23. Valores promedio de Nitrito por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 8 y figura 23 muestra que la concentración de Nitrito promedio en el periodo de estudio, se puede notar que no tiene un patrón estacional, con valores relativamente estabilizados.

Análisis de la tendencia. La figura siguiente muestra la tendencia de la concentración de Nitrito de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente.

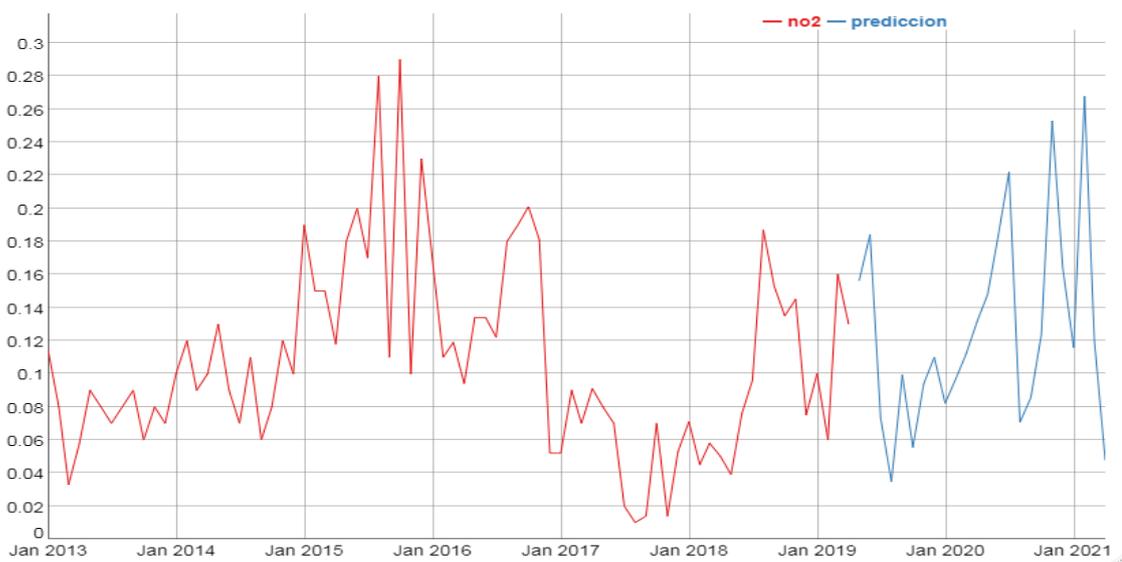


Figura 24. Tendencia y valores predichos para el parámetro Nitrito de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro Nitrito de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica que en el tiempo los valores han aumentado a partir del año 2016, muy por encima del valor neutro. La tendencia tiende a mantenerse con similitud al 2015. Para obtener el modelo de red neuronal se usó 1,2,4 y 5 retardos y 8 nodos, con un RMSE de 0.021025395.

DIGESA (2011) indica que los nitritos presentes en aguas superficiales y que luego pasen por un tratamiento, sus concentraciones se verán reducidas debido a la oxidación por la cloración. valores A1: valores aceptables 1 mg/L N-nitrito, B1: se recomienda 1 mg/L Nitrito. Bajo este parámetro, las aguas del puerto muelle de Puno contienen valores menores a 1 mg/l.

4.9 Análisis descriptivo y tendencia del parámetro conectividad eléctrica

La conductividad es una expresión numérica de la capacidad de una solución para transportar una corriente eléctrica. Esta capacidad depende de la presencia de iones y de su concentración total, de su movilidad, valencia y concentraciones relativas, así como la temperatura de la medición.

El agua pura tiene muy poca conductividad, por lo que la medida de la conductividad de un agua nos da una idea de los sólidos disueltos en la misma. (DIGESA, 2011)

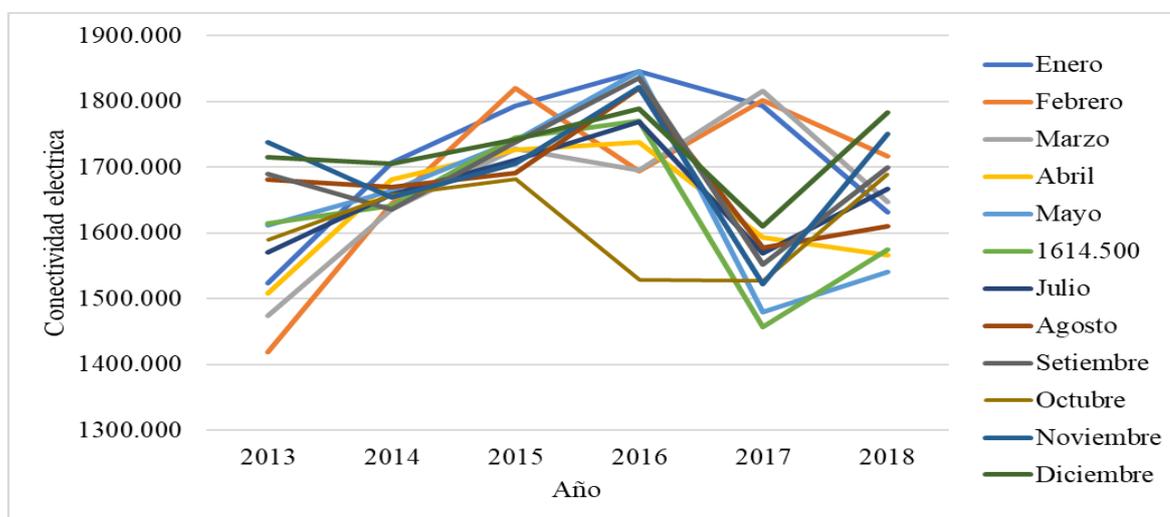


Figura 25. valores observados para la conectividad eléctrica de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de febrero y marzo y abril y el valor máximo en el mes de julio y diciembre. Los valores de conectividad eléctrica aumentaron en el año 2014 para luego disminuir al año 2017, nuevamente se incrementaron los valores en el año 2018.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada conectividad eléctrica, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 9

Estadísticos descriptivos para conectividad eléctrica por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	1711.06	1695.95	1678.75	1651.85	1647.11	1633.94	1657.96	1675.04	1692.10	1612.30	1698.99	1724.14
de	110.80	137.17	109.22	95.61	133.49	115.05	78.39	84.00	95.56	74.29	102.60	65.76
max	1846.00	1819.90	1816.39	1748.97	1846.10	1770.57	1769.16	1820.94	1836.23	1688.74	1821.60	1789.61
min	1524.52	1418.79	1474.58	1508.33	1480.19	1457.20	1569.68	1577.94	1551.83	1527.03	1522.27	1609.74

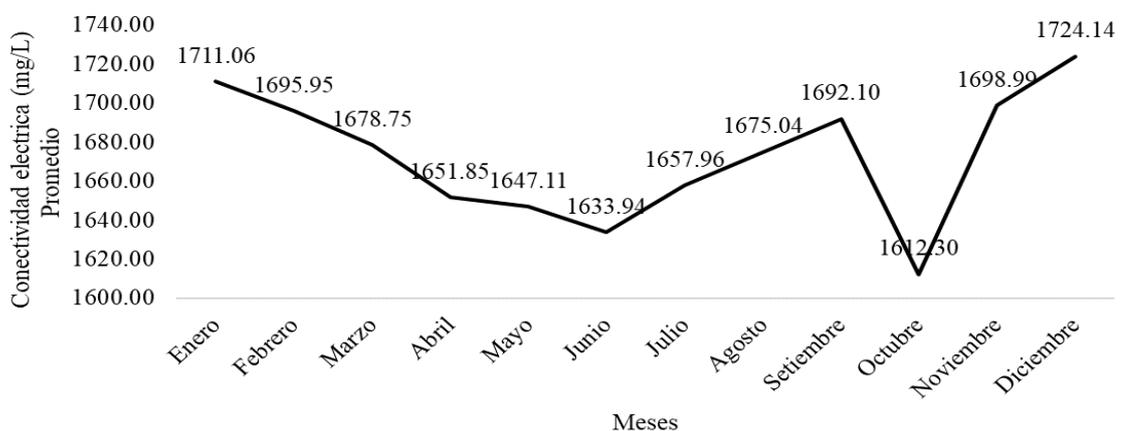


Figura 26. Valores promedio de conectividad eléctrica por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019

La tabla 9 y figura 26 muestra que la concentración de conectividad eléctrica promedio en el periodo de estudio, se puede notar que tiene un patrón estacional, excepto el mes de octubre que presenta un valor anómalo la tendencia es decreciente hacia los meses de

abril, mayo, junio llegando al punto más bajo y luego presenta una tendencia positiva desde el mes de julio.

Análisis de la tendencia. La figura siguiente muestra la tendencia de la concentración de conectividad eléctrica de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

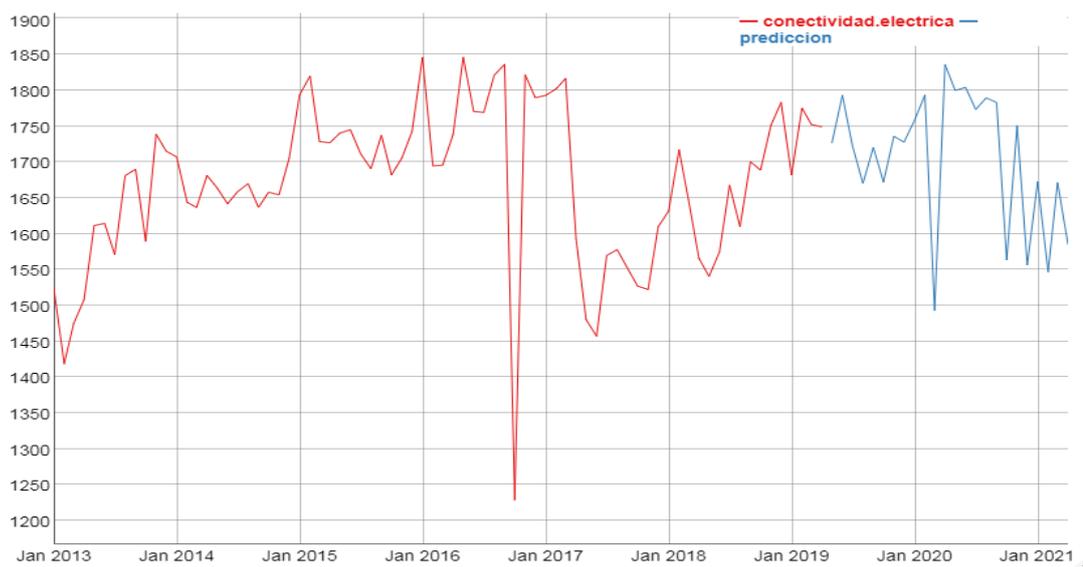


Figura 27. Tendencia y valores predichos para el parámetro conectividad eléctrica de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro conectividad eléctrica de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica una forma oscilatoria, la proyección es similar al año 2018. Para obtener el modelo de red neuronal se usó 1,2,4 y 5 retardos y 8 nodos, con un RMSE de 0.021025395.

La conductividad eléctrica de un agua se utiliza como una medida indirecta de su concentración de sólidos disueltos totales o de minerales en el agua. DIGESA propone los siguientes valores A1/A2/A3: se recomienda concentraciones de 1000 us/cm. las aguas del puerto muelle sobrepasan este valor por lo que existe concentración de solidos totales o minerales.

4.10 Análisis descriptivo y tendencia del parámetro total de sólidos en suspensión

Los sólidos en suspensión son productos de la erosión de los suelos, detritus orgánico y plancton. Los sólidos suspendidos, tales como limo, arena y virus, son generalmente responsables de impurezas visibles. La materia suspendida consiste en partículas muy pequeñas, que no se pueden quitar por medio de deposición. Pueden ser identificadas con la descripción de características visibles del agua, incluyendo turbidez y claridad, gusto, color y olor del agua (DIGESA, 2011).

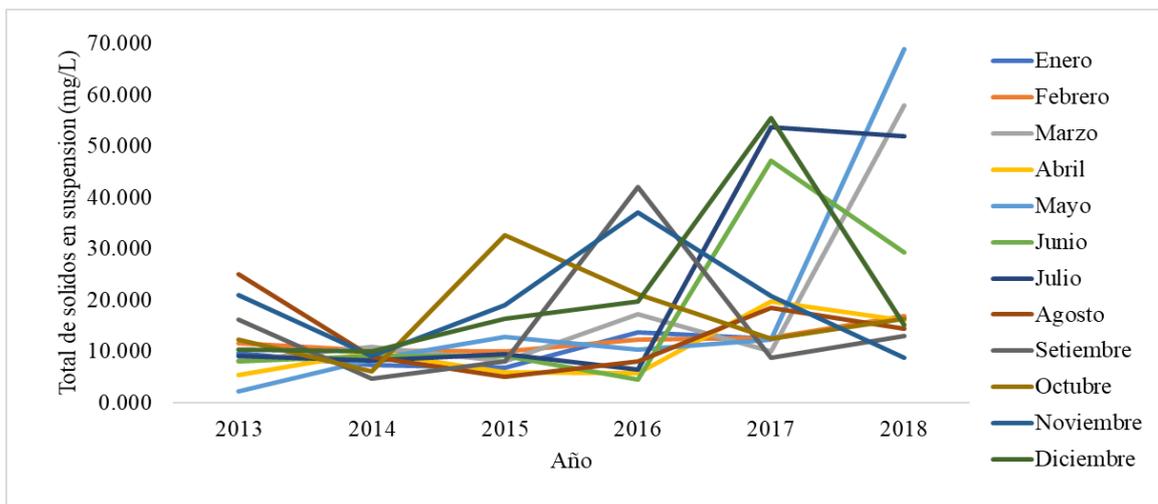


Figura 28. valores observados para el total de sólidos en suspensión de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de mayo y abril y el valor máximo en el mes de agosto. Los valores de total de sólidos en suspensión han ido variando según los meses, tal es el caso de que el valor más alto se presentó en el mes de mayo y marzo.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada total de sólidos en suspensión, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 10

Estadísticos descriptivos para la variable total de sólidos en suspensión por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	90.22	92.24	149.62	94.80	115.16	107.38	139.01	80.07	92.74	101.31	115.83	127.01
DE	5.86	3.24	19.07	9.85	24.69	16.87	23.04	7.47	13.61	9.18	10.35	17.19
max	23.50	18.32	58.00	32.00	69.00	47.25	53.75	25.00	42.00	32.67	37.11	55.43
min	6.80	10.02	8.10	5.50	2.33	4.50	6.45	5.00	4.67	6.20	8.86	10.05

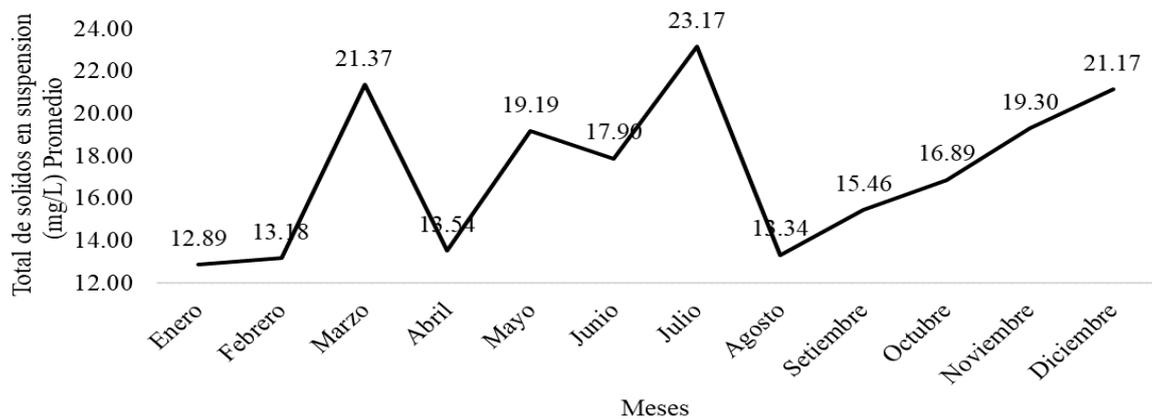


Figura 29. Valores promedio de total de sólidos en suspensión por meses de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019

La tabla 10 y figura 29 muestra que el total de sólidos en suspensión promedio en el periodo de estudio, se puede notar que tiene un patrón no estacional, pero se observa que, a partir del mes de agosto, los valores se incrementan paulatinamente hasta diciembre.

Análisis de la tendencia. La figura siguiente muestra la tendencia del total de sólidos en suspensión de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

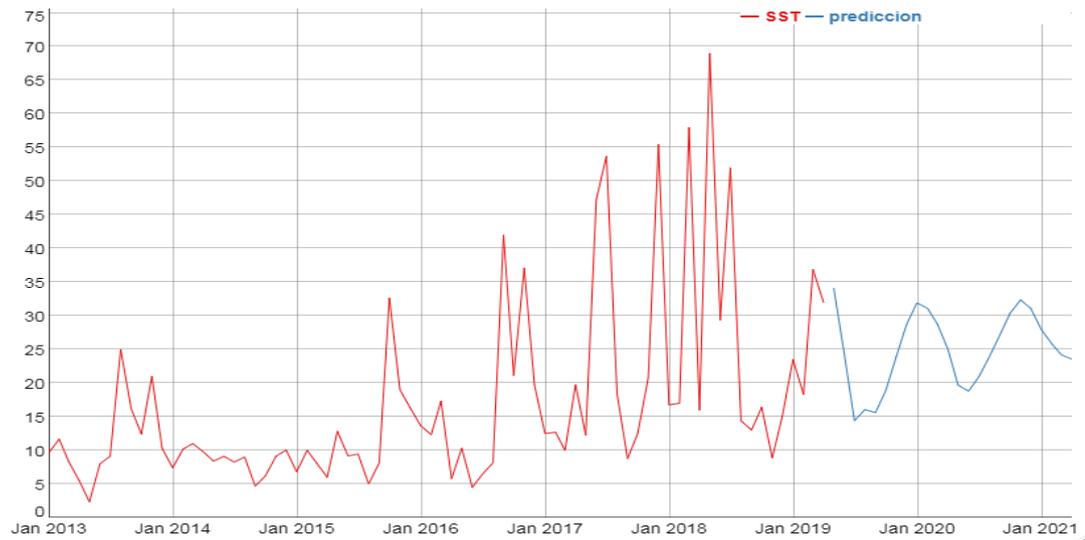


Figura 30. Tendencia y valores predichos para el parámetro total de sólidos en suspensión de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro total de sólidos en suspensión de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. La tendencia indica una forma oscilatoria irregular desde los años 2017 y 2018, la proyección muestra valores mayores a 15 en forma más estabilizada. Para obtener el modelo de red neuronal de 1,2,3,4,5,6,7,8,9,10 retardos y 2 nodos, con un RMSE de 8.640392117.

Hallasi (2108) indica que Solidos totales disueltos 11147,80 mg/l Al confrontar estos valores con los ECAs (MINAM, 2017) en la Categoría 1 A-1 se aprecia que nuestros valores superan ampliamente lo establecido en la norma, ya que aguas con fines de consumo humano no deberían pasar los SDT de 1000 mg/l, por lo tanto, estas aguas no son adecuadas en este parámetro para el consumo humano. teniendo en cuenta los valores de la SUNASS en el cual indica que para consumo humano se recomienda 25 mg/L, las aguas del puerto muelle no serían aptas para consumo, ya que sobrepasan el valor recomendado en A1 y A2 (50 mg/L)

4.11 Análisis descriptivo y tendencia del parámetro Clorofila-a

La clorofila-a es el principal pigmento fotosintético de las algas. Se considera como un proxy para la biomasa de fitoplancton presente en la columna de agua. Con las partículas minerales en suspensión, contribuye a la turbidez del agua. Cuantos más aportes de

nutrientes (nitrógeno y fósforo) son importantes, más el ecosistema se vuelve propicio para el desarrollo del fitoplancton (Rojas, 2016).

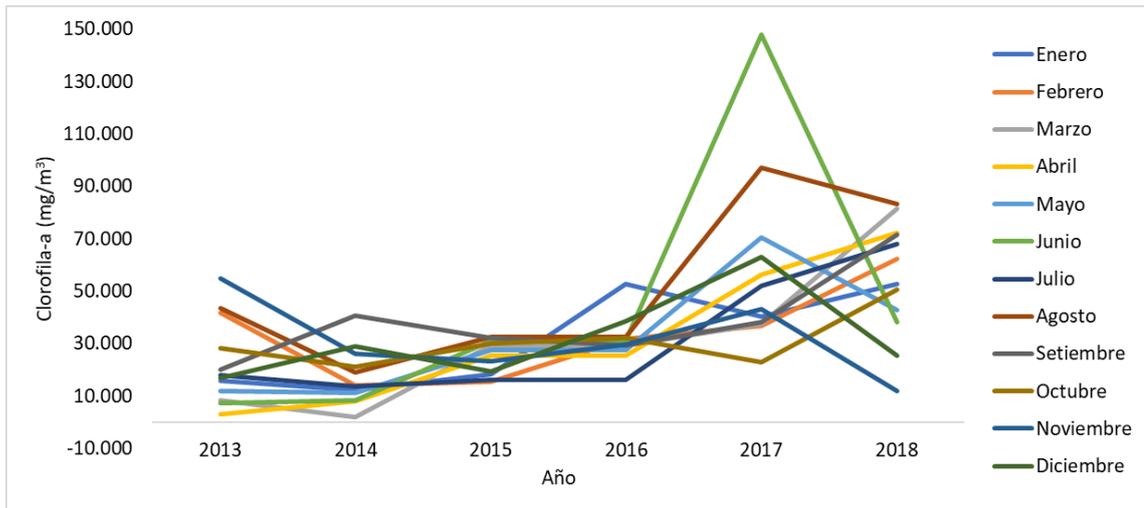


Figura 31. valores observados para Clorofila-a de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de marzo y abril y el valor máximo en el mes de noviembre. Los valores de Clorofila han ido variando según los meses, tal es el caso que estos valores se incrementan a partir del año 2015, 2106 y 2017, en el 2108 se observa una pequeña disminución.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada clorofila, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 11

Estadísticos descriptivos para clorofila-a por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	30.85	33.36	36.36	31.91	31.93	43.98	30.71	51.36	38.72	30.74	31.50	32.12
de	17.44	16.37	28.93	24.78	22.26	52.46	23.34	31.34	17.64	10.54	15.22	17.03
max	52.84	62.21	81.52	72.16	70.49	147.80	68.11	97.09	71.54	50.48	54.78	63.18
min	12.45	14.12	2.05	3.08	11.20	7.31	13.80	19.15	20.15	21.10	12.00	16.90

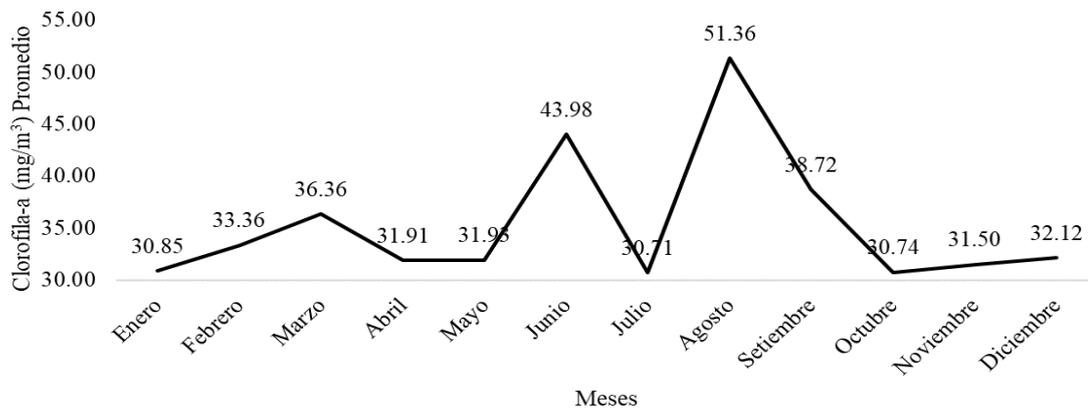


Figura 32. Valores promedio para clorofila-a por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2109.

La tabla 11 y figura 32 muestran los valores de clorofila-a promedio en el periodo de estudio, se puede notar que tiene un patrón no estacional, pero se observa que en los meses de junio y agosto presentan valores altos en comparación a los otros meses.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de clorofila-a de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

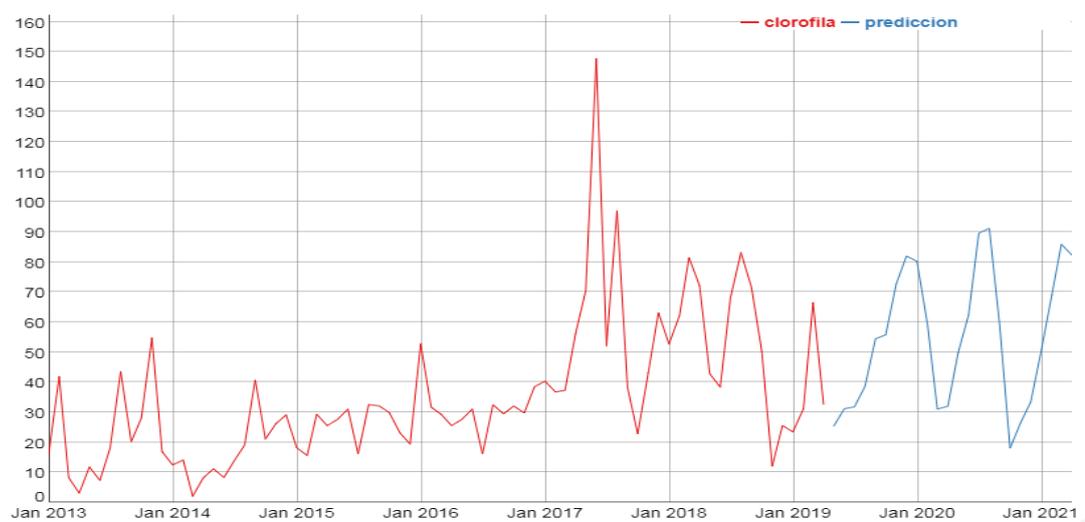


Figura 33. Tendencia y valores predichos para el parámetro clorofila-a de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro clorofila-a de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del

2019. los valores han aumentado a partir del año 2017, con pequeños bajos, la predicción muestra que la tendencia continuara con valores similares al año 2018.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10 retardos y 2 nodos, con un RMSE de 8.640392117.

La concentración de clorofila-a se utiliza para estimar en forma indirecta la biomasa de las comunidades fitoplanctónicas, debido a que es el principal pigmento fotosintético presente en las algas (Gregor y Marsálek, 2004). La clorofila-a también es un indicador del grado de contaminación de los ecosistemas acuáticos y un importante índice del estado fisiológico del fitoplancton (Pinto, Von Sperling y Moreira, 2001).

4.12 Análisis descriptivo y tendencia de la Demanda bioquímica de oxígeno

La DBO₅ expresan la materia orgánica en términos generales, pero no indican su composición, la cual es muy variada. Como su origen proviene de organismos, y sus productos de degradación o de metabolismo, se puede afirmar que la componen proteínas, carbohidratos y lípidos y/o sus productos de degradación: aminoácidos, monosacáridos, hidrocarburos, ácidos grasos, alcoholes, más otros componentes propios de los vegetales como pigmentos DBO: Es el parámetro de contaminación orgánica. Es el resultado de la degradación de tres tipos de materiales: materiales orgánicos carbónicos (microorganismos aerobios), nitrógeno oxidable (nitrosomas y nictrobacter), compuestos químicos reductores (se oxidan con el OD) (DIGESA, 2011).

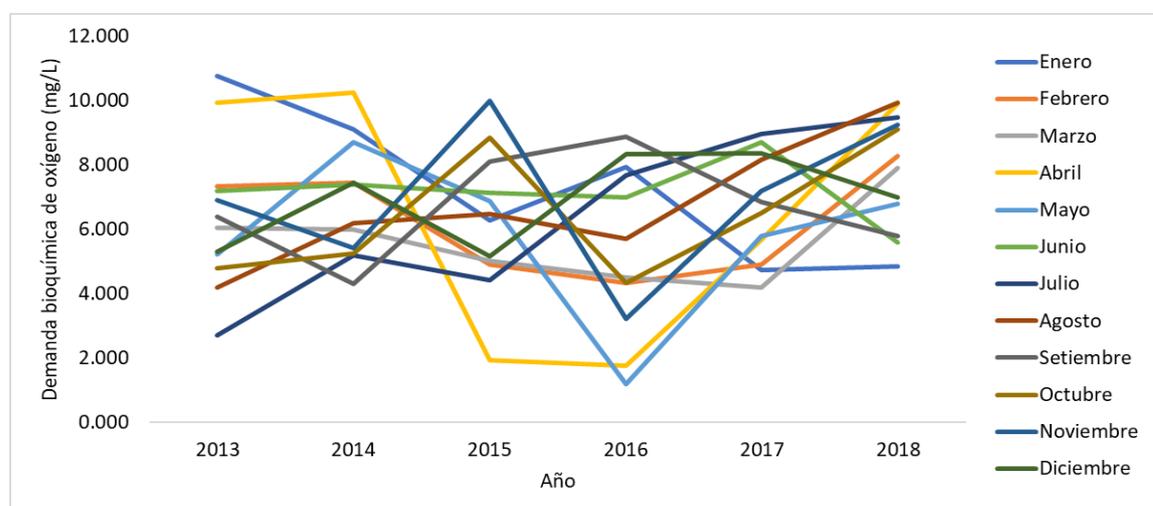


Figura 34. valores observados para la demanda bioquímica de oxígeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

La figura 34 muestra que en el año 2013 los valores mínimos se presentaron en los meses de julio y agosto y el valor máximo en el mes de enero. Los valores de Clorofila han ido variando según los meses, presentando una disminución en el año 2106 (mayo). el comportamiento es similar entre todos los años.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada demanda bioquímica de oxígeno, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 12

Estadísticos descriptivos para la demanda bioquímica de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	6.94	6.84	6.25	7.29	5.77	7.17	6.41	6.79	6.72	6.48	7.00	6.94
DE	2.39	2.27	2.11	4.13	2.54	0.99	2.71	2.01	1.64	2.07	2.48	1.43
max	10.76	10.64	10.14	11.53	8.72	8.70	9.48	9.94	8.88	9.12	10.00	8.38
min	4.74	4.34	4.18	1.76	1.20	5.60	2.70	4.20	4.31	4.34	3.23	5.16

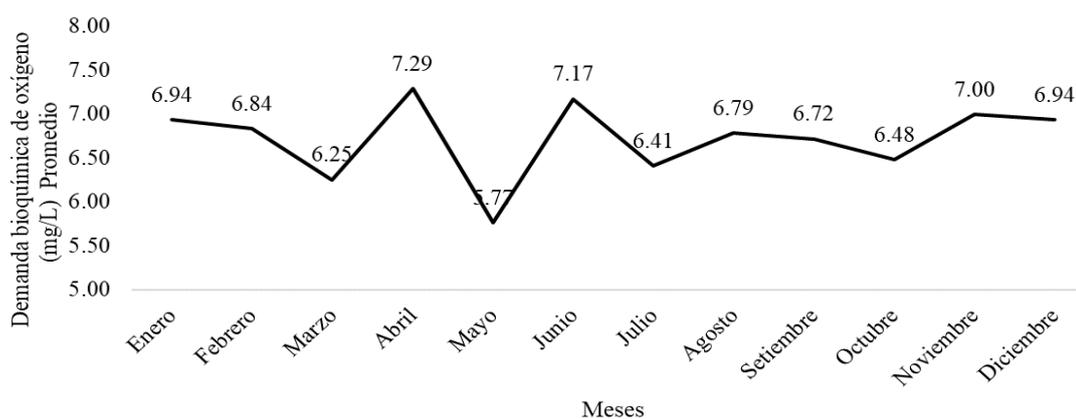


Figura 35. Valores promedio para demanda bioquímica de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019

La tabla 12 y figura 35 muestran los valores de demanda bioquímica de oxígeno promedio en el periodo de estudio, se puede notar que tiene un patrón no estacional, con valores que aumentan y bajan en el periodo.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de demanda bioquímica de oxígeno de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

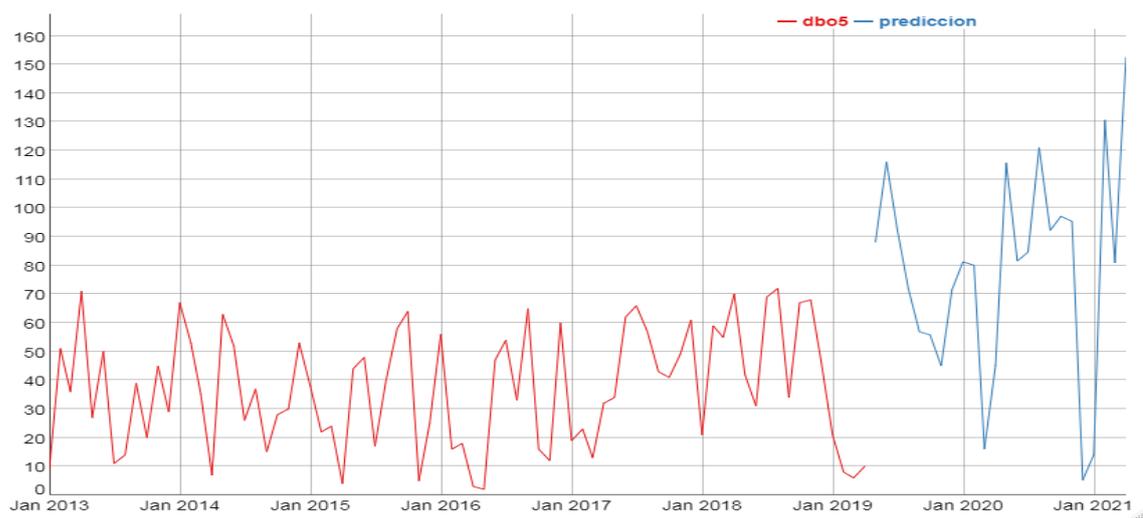


Figura 36. Tendencia y valores predichos para el parámetro demanda bioquímica de oxígeno de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro demanda bioquímica de oxígeno de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran un periodo estacional con valores máximos y mínimos oscilantes, la predicción muestra que la tendencia continuara con valores incrementados.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10 retardos y 4 nodos, con un RMSE de 7.324894685.

DIGESA (2011) recomienda los parámetros A1: La DBO es la cantidad de oxígeno consumido por microorganismo para la degradación biológica en un medio aeróbico de la materia orgánica un medio aeróbico, a mayor cantidad de materia orgánica mayor

contenido de microorganismos en una muestra de agua por lo que se recomienda bajas concentraciones. Algunas legislaciones recomiendan DBO5 de 3 mgO₂/L. A2: se recomienda concentraciones aceptables entre 5 mgO₂/L y 10 mgO₂/L para aguas dulces. los valores de la toma, superan los 5 mg/l.

4.13 Análisis descriptivo y tendencia del parámetro Demanda química de oxígeno

La demanda química de oxígeno (DQO) es un parámetro que mide la cantidad de sustancias susceptibles de ser oxidadas por medios químicos que hay disueltas o en suspensión en una muestra líquida. Se utiliza para medir el grado de contaminación y se expresa en miligramos de oxígeno diatómico por litro (mg O₂/L). Aunque este método pretende medir principalmente la concentración de materia orgánica, sufre interferencias por la presencia de sustancias inorgánicas susceptibles de ser oxidadas (sulfuros, sulfitos, yoduros...), que también se reflejan en la medida.

La demanda bioquímica de oxígeno, DBO, se define como la cantidad de oxígeno usado por los microorganismos no fotosintéticos a una temperatura de 20°C, para metabolizar los compuestos orgánicos degradables biológicamente (Cisterna, s.f).

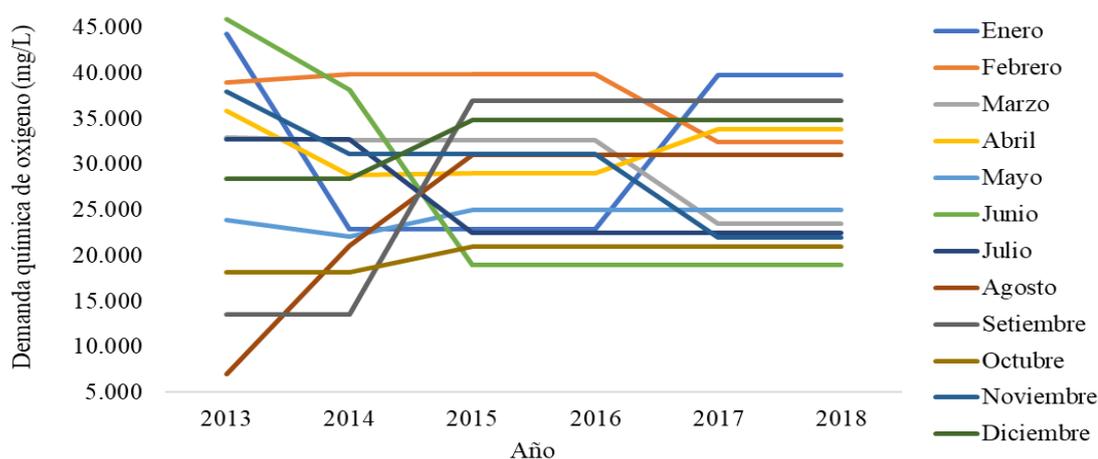


Figura 37. valores observados para la demanda química de oxígeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 los valores mínimos se presentaron en los meses de agosto y setiembre y el valor máximo en el mes de junio. Los valores de demanda química de oxígeno han incrementado a partir del año 2105.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada demanda química de oxígeno, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 13

Estadísticos descriptivos para la variable demanda química de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
Prom	33.20	36.58	28.80	32.06	24.33	26.68	25.93	25.35	29.17	20.07	29.23	32.73
de	9.77	3.87	4.96	3.01	1.18	12.15	5.32	9.82	12.14	1.45	6.20	3.36
max	44.30	39.90	33.00	35.90	25.00	45.90	32.80	31.00	37.00	21.00	38.00	34.90
min	22.90	32.46	23.50	28.80	22.10	19.00	22.50	7.00	13.50	18.20	22.00	28.40

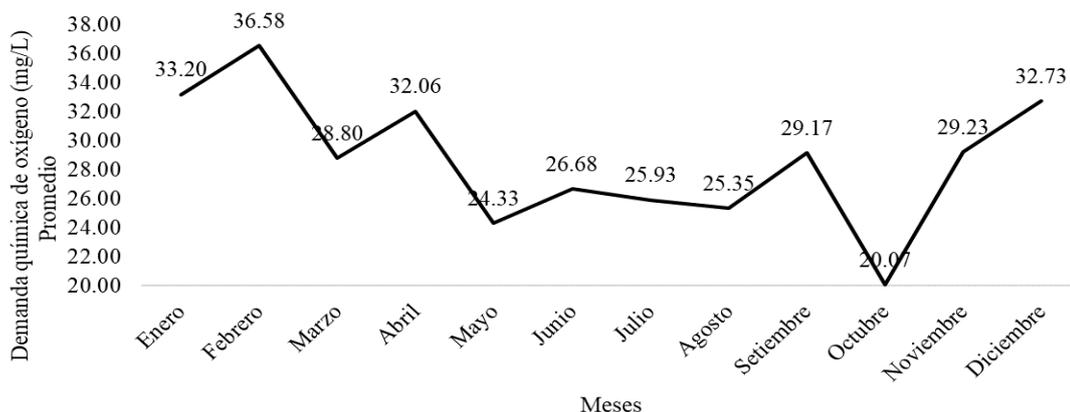


Figura 38. Valores promedio para demanda química de oxígeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 13 y figura 38 muestran los valores de demanda química de oxígeno promedio en el periodo de estudio, se puede notar que tiene un patrón descendente hasta los meses de mayo, junio, julio agosto, setiembre y octubre, luego sube en noviembre, diciembre, enero y febrero.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de demanda química de oxígeno de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

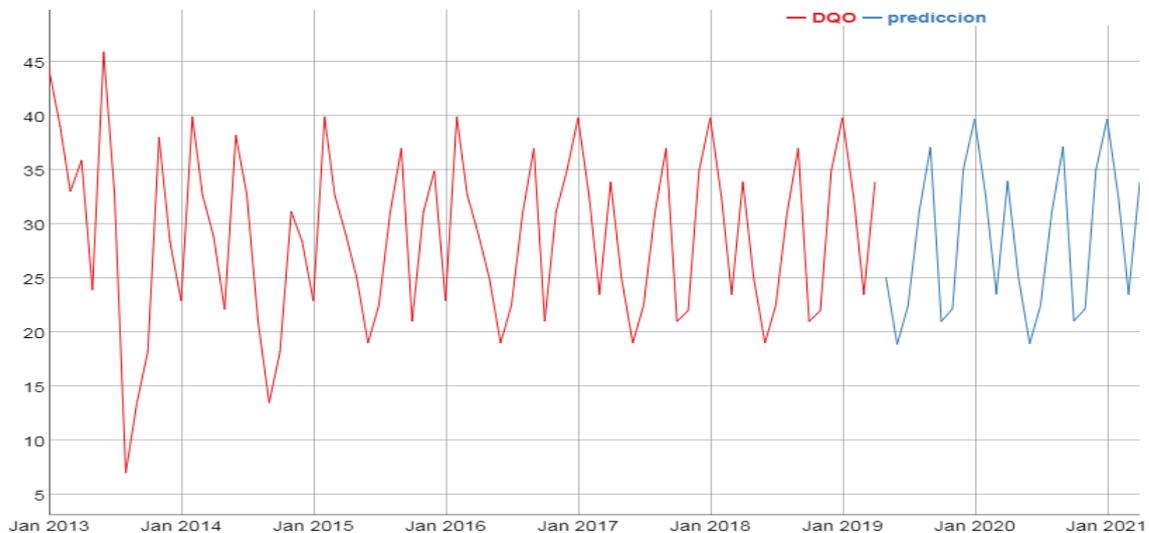


Figura 39. Tendencia y valores predichos para el parámetro demanda química de oxígeno de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro demanda química de oxígeno de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran un periodo estacional con valores máximos y mínimos oscilantes, la predicción muestra que la tendencia continuara bajo el mismo patrón.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10 retardos y 4 nodos, con un RMSE de 7.324894685.

DIGESA propone A1: algunas normas consideran valores de 10 mg/L. A2: se recomiendan concentraciones de 20 mg/L. A3: concentraciones aceptables de 30 mg/L. los valore2 obtenidos superan los parámetros A1 y A2.

4.14 Análisis descriptivo y tendencia del parámetro Nitrate

¿Qué es el nitrato (NO₃)? El nitrato es un contaminante común que se encuentra en el agua subterránea y que puede provocar efectos nocivos si se consume en altos niveles. El

nitrate is odorless and colorless. Low concentrations of nitrate are normal, but high amounts can contaminate our drinking water source. In areas with pure water, such as shallow groundwater not affected by human activities, concentrations are usually less than 2 milligrams per liter (mg/L) of nitrate.

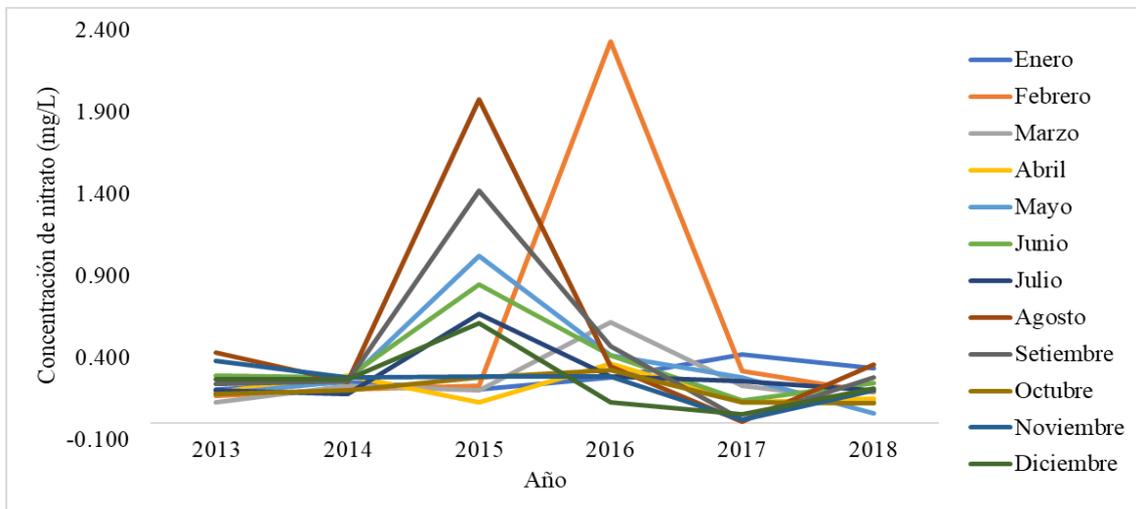


Figura 40. valores observados para la Concentración de nitrato de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

In 2013 there is not much difference in the values of nitrate concentration, we observe high values in the year 2015 months of August, July and May, decreasing towards the year 2017 and 2018 to values similar to 2014 and 2013.

The following table contains the main descriptive statistics of the variable observed nitrate concentration, which is part of the structural model and the relationship of the physical-chemical parameters of the inner bay of Lake Titicaca, Puno port. For this effect, the entire study period has been consolidated.

Tabla 14

Estadísticos descriptivos para la concentración de nitrato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	0.30	0.53	0.25	0.20	0.37	0.37	0.30	0.56	0.45	0.21	0.24	0.26
de	0.08	0.79	0.17	0.09	0.34	0.25	0.19	0.71	0.50	0.08	0.12	0.19
max	0.42	2.33	0.62	0.37	1.02	0.85	0.67	1.98	1.42	0.32	0.38	0.61
min	0.21	0.17	0.13	0.13	0.06	0.14	0.18	0.01	0.02	0.12	0.02	0.06

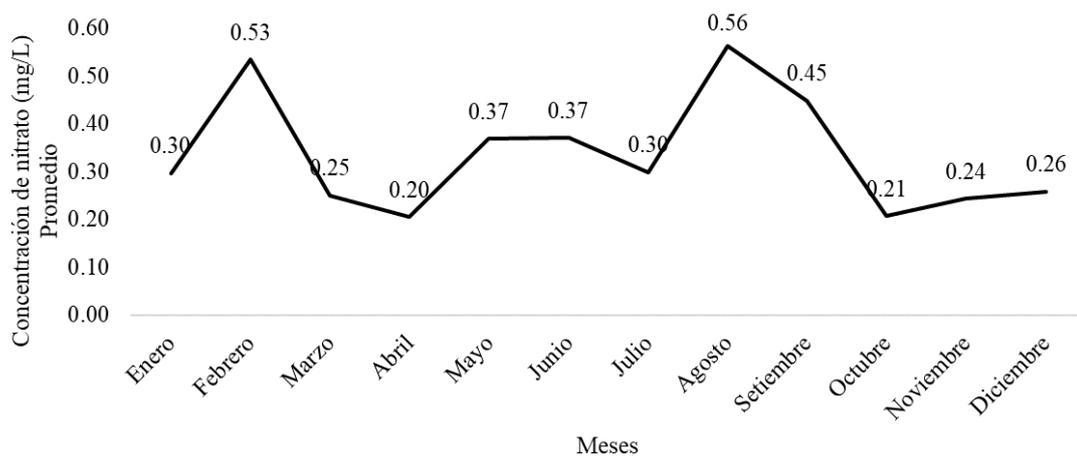


Figura 41. Valores promedio para la concentración de nitrato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 14 y figura 41 muestran los valores de concentración de nitrato promedio en el periodo de estudio, se puede notar que tiene un patrón oscilatorio entre meses, no sigue un patrón anual.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de concentración de nitrato de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

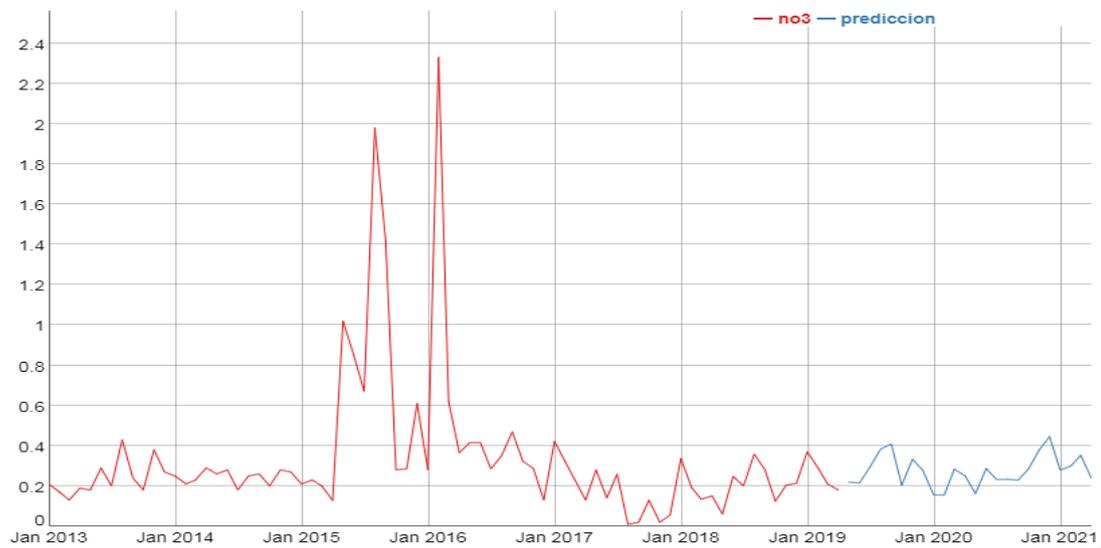


Figura 42. Tendencia y valores predichos para el parámetro concentración de nitrato de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de nitrato de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran que a partir del año 2016 estos valores muestran similitud con descensos en el año 2017 segundo semestre y valores altos en el año 2105 y 2106. La predicción muestra que la tendencia continuara bajo el mismo patrón.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10 retardos y 4 nodos, con un RMSE de 7.324894685.

DIGESA (2011) propone que debido a que las practicas convencionales de tratamiento no modifican en forma apreciable los niveles de nitrato en agua y como sus concentraciones no varían en forma notable en los sistemas de distribución, las concentraciones detectadas en aguas superficiales son iguales a las aguas de consumo. A1/A2: la EPA recomienda que la concentración de nitratos no debe excederse de 45 mg/L, dado los efectos sobre la salud de los infantes otras legislaciones establecen de 10 y 50 mg/L de N nitrato. A3: concentraciones aceptables de 50mg/L nitrato. B1: se recomienda 10 mg/L. los valores no exceden A1 y menos a A2.

4.15 Análisis descriptivo y tendencia del parámetro fosfato

Los fosfatos son nutrientes para las plantas. Tienen aplicaciones industriales diversas y como fertilizantes. Los vertidos de fosfatos a las aguas naturales pueden causar eutrofización.

De la utilización de los fosfatos y nitratos, presentes en la materia orgánica de la basura, de los detergentes hechos a base de fosfatos ocurre el proceso eutrofización, que son arrastrados o arrojados a los ríos y lagos son un problema muy grave para las aguas estancadas cerca de los centros urbanos o agrícolas. Durante las épocas cálidas la sobrecarga de estos productos químicos, que sirven de nutrientes, genera el crecimiento acelerado de vegetales como algas, cianobacterias, lirios acuáticos y lenteja de agua, las cuales al morir y ser descompuestas por las bacterias aeróbicas provocan el agotamiento del oxígeno disuelto en la capa superficial de agua y causan la muerte de los diferentes tipos de organismos acuáticos que consumen oxígeno, en las aguas de los lagos y ríos.

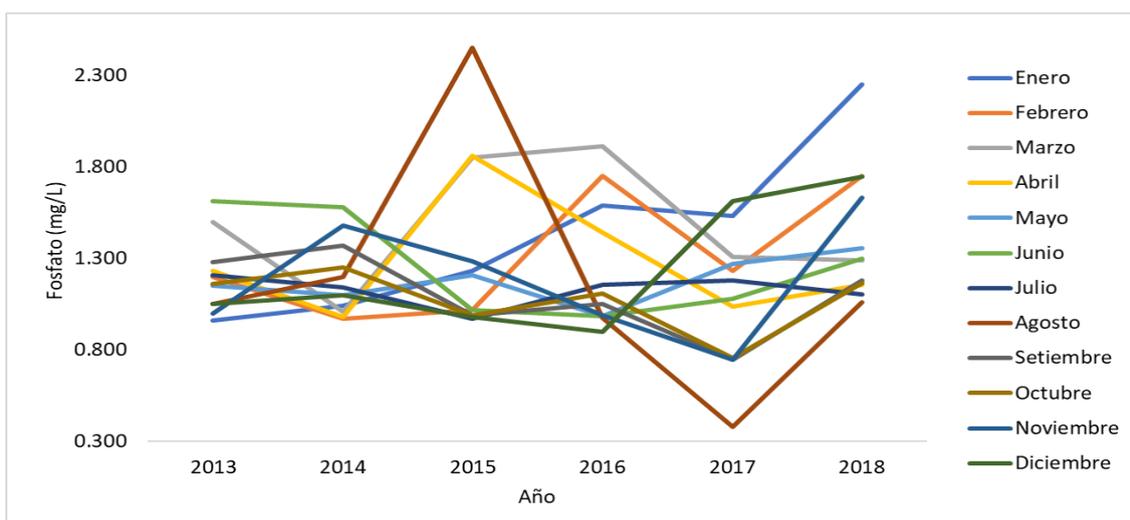


Figura 43. valores observados para la variable fosfato de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 en los meses de noviembre y enero se presentaron los niveles más bajos de fosfato y en el mes de junio el valor más alto, en el mes de agosto a lo largo del periodo se muestra una tendencia bastante irregular mostrando puntos altos y bajos, al año 2018 los niveles son similares a los años anteriores excepto al año 2017.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada fosfato, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 15

Estadísticos descriptivos para fosfato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	1.37	1.36	1.57	1.42	1.18	1.26	1.13	1.19	1.10	1.07	1.19	1.23
de	0.46	0.33	0.39	0.46	0.13	0.28	0.08	0.68	0.22	0.18	0.34	0.35
max	2.25	1.75	2.09	2.22	1.36	1.61	1.21	2.45	1.37	1.25	1.63	1.74
min	0.96	0.97	1.01	0.98	0.98	0.98	0.97	0.38	0.75	0.75	0.75	0.90

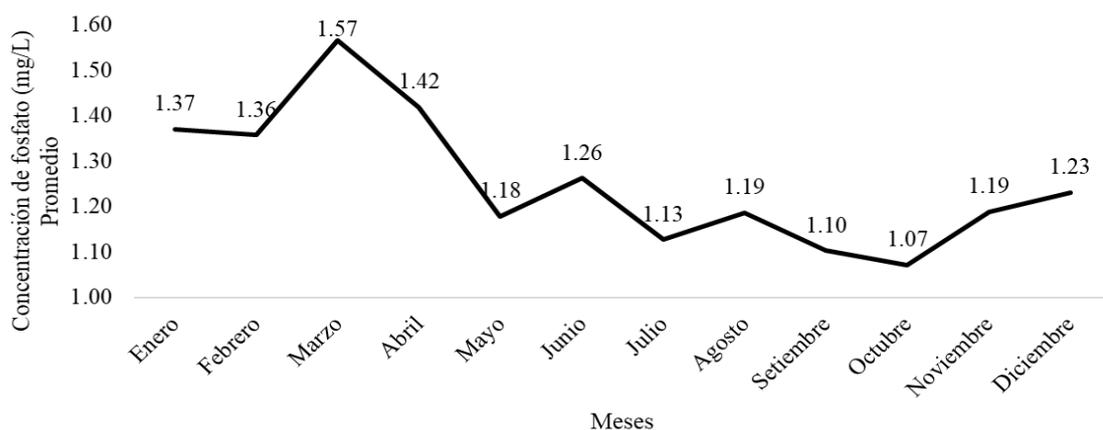


Figura 44. Valores promedio para la concentración de fosfato por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 15 y figura 44 muestran los valores de concentración de fosfato promedio en el periodo de estudio, se puede notar que en los meses de mayo a diciembre estos valores disminuyen en comparación a los meses de enero a mayo.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de concentración de fosfato de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

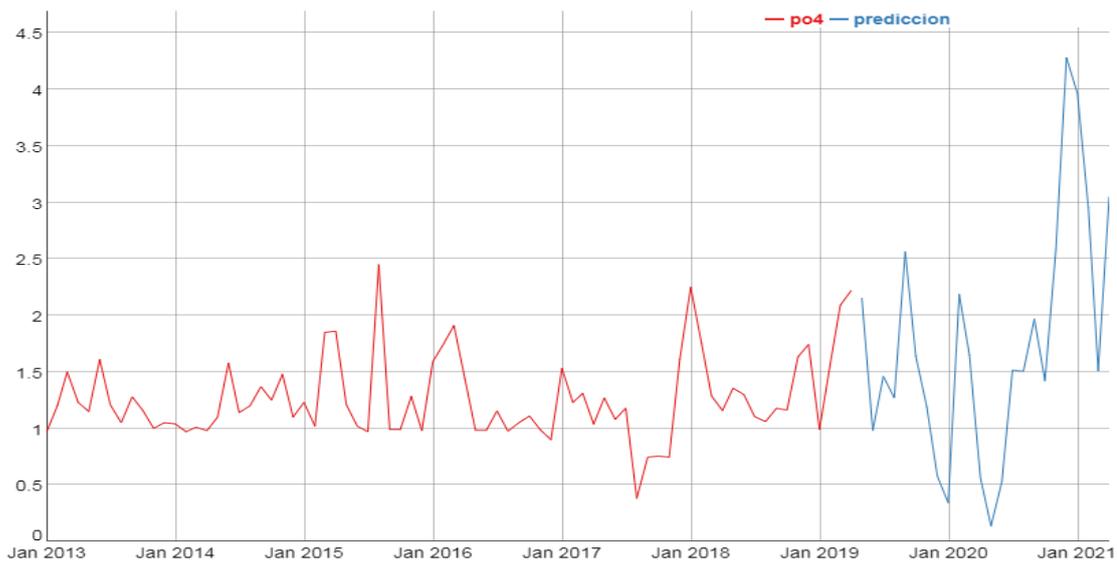


Figura 45. Tendencia y valores predichos para el parámetro concentración de fosfato de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de nitrato de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran que la tendencia es similar a lo largo de los años, excepto que la predicción es a aumentar en el año 2019 y 2020.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10 retardos y 5 nodos, con un RMSE de 0.107015831.

DIGESA propone A1: se recomienda valores de 0.4 mg/L P₂O₅. A2/A3: concentraciones aceptables 0.7 mg/L P₂O₅. B1/B2: (No Aplicable). los valores superan lo mínimo permitido.

Hallasi (2018) obtiene valores de Fosfato 0,04 mg/l, las aguas cumplen con la Normativa de los ECAs, por lo tanto, son aptas para consumo humano ya que el límite evidenciado en el (MINAM, 2017) es de 0.1 mg/l para A-1. que difieren de los niveles de fosfato en aguas de la toma muelle y los UROS.

4.16 Análisis descriptivo y tendencia del parámetro Nitrógeno

El nitrógeno (N) es un contaminante presente en las aguas residuales que debe ser eliminado por múltiples razones: reduce el oxígeno disuelto de las aguas superficiales, es tóxico para el ecosistema acuático, entraña un riesgo para la salud pública y junto al

fósforo (P), son responsables del crecimiento desmesurado de organismos fotosintéticos (eutrofización).

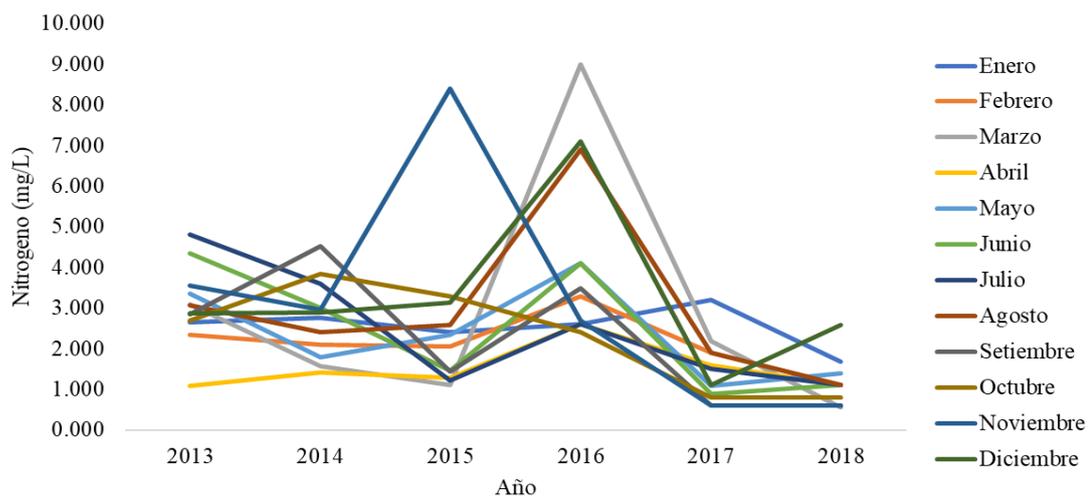


Figura 46. valores observados para la variable nitrógeno de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 en el mes de abril se presentó el nivel más bajo de nitrógeno y en el mes de julio el valor más alto, en los meses de enero y marzo del año 2015 y 2016 respectivamente, se presentaron valores altos de nitrógeno, la tendencia general de los parámetros al 2018 es a disminuir.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada nitrógeno, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 16

Estadísticos descriptivos para nitrógeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	2.55	2.12	2.82	1.74	2.35	2.49	2.48	3.00	2.25	2.31	3.13	3.29
de	0.46	0.65	2.85	0.78	1.17	1.53	1.48	2.02	1.62	1.27	2.86	2.00
max	3.20	3.30	9.00	3.08	4.10	4.35	4.80	6.90	4.52	3.85	8.40	7.10
min	1.68	1.12	0.56	1.09	1.10	0.90	1.12	1.12	0.60	0.80	0.60	1.12

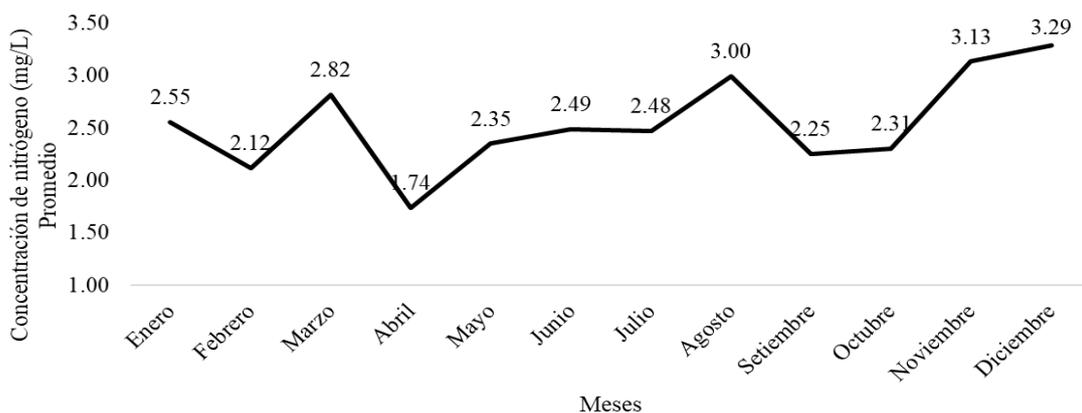


Figura 47. Valores promedio para la concentración de nitrógeno por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 16 y figura 47 muestran los valores de concentración de nitrógeno promedio en el periodo de estudio, se puede notar que no existe una tendencia anual pero el valor mínimo se dio en el mes de abril.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de concentración de nitrógeno de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

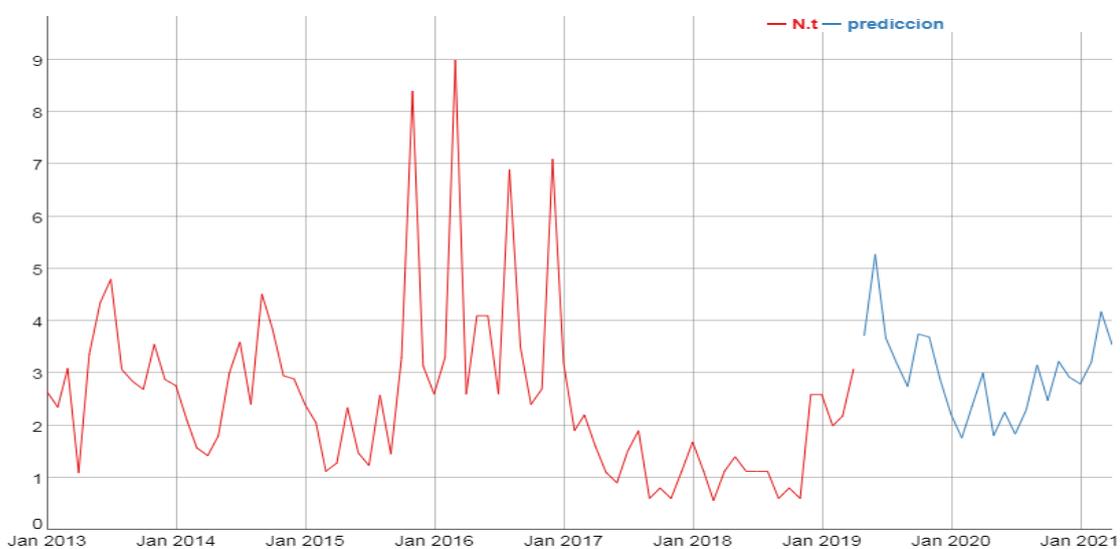


Figura 48. Tendencia y valores predichos para el parámetro concentración de nitrógeno de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de nitrógeno de la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran un incremento de los valores en el año 2016 y luego a partir del año 2017 el valor disminuye, con tendencia a volver a incrementarse.

Para obtener el modelo de red neuronal se usó 1,3,4,8,9 retardos y 8 nodos, con un RMSE de 0.504065548.

DIGESA propone A1: se recomienda concentraciones de 1 mg/L N. las aguas del muelle de Puno superan este valor propuesto.

4.17 Análisis descriptivo y tendencia del parámetro Fosforo

El fósforo ocurre en aguas naturales y residuales principalmente en forma particulada inorgánica u orgánica y en una menor fracción en forma disuelta, orgánica (nucleótidos, polinucleótidos, sustancias húmicas, poli, meta y ultra fosfatos) e inorgánica (PO₄-3, HPO₄-2, H₂PO₄-) (Arocena & Conde 1999; Wetzel 2001). La concentración de fósforo total mide la cantidad de fósforo disponible en forma orgánica e inorgánica, disuelta y particulada en los sistemas acuáticos.

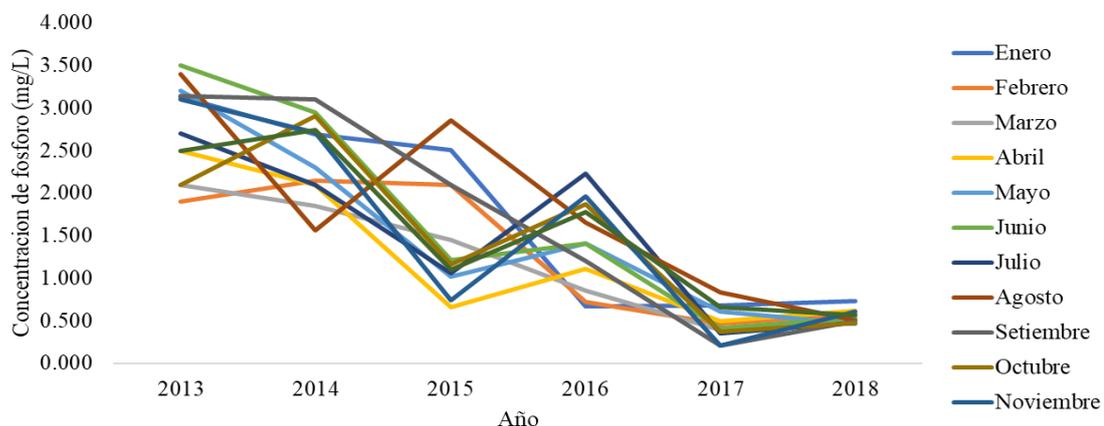


Figura 49. valores observados para la variable fosforo de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 en el mes de febrero se presentó el nivel más bajo de fosforo y en el mes de junio el valor más alto, la tendencia de todos los valores a partir del año 2103 tienden a disminuir hasta el año 2017, se nota un pequeño incremento en el año 2018.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada fosforo, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 17

Estadísticos descriptivos para fosforo por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	1.61	1.23	1.18	1.25	1.50	1.67	1.49	1.80	1.71	1.48	1.56	1.56
de	1.11	0.78	0.65	0.78	1.06	1.28	0.99	1.13	1.27	0.99	1.20	0.93
max	3.13	2.15	2.10	2.50	3.20	3.50	2.70	3.40	3.14	2.91	3.10	2.74
min	0.67	0.47	0.39	0.50	0.46	0.41	0.35	0.51	0.21	0.37	0.21	0.57

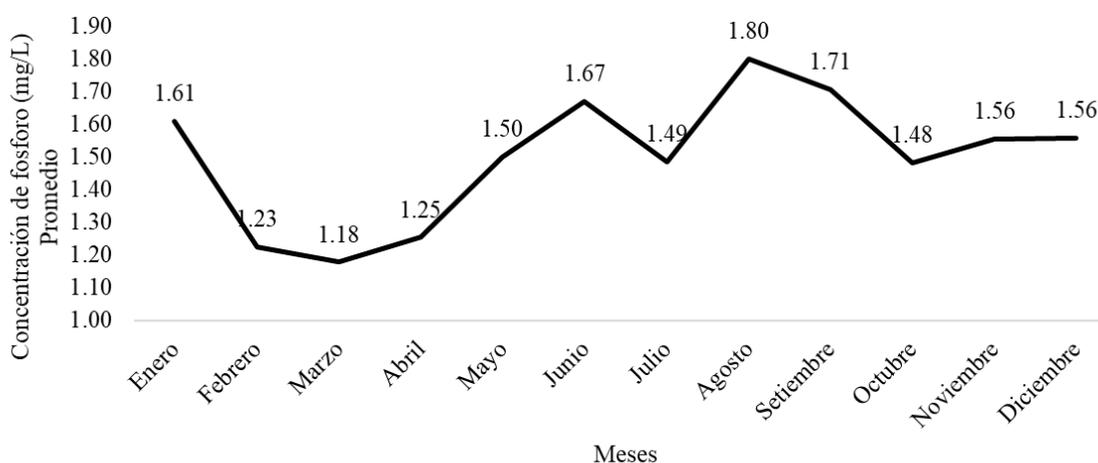


Figura 50. Valores promedio para la concentración de fosforo por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 17 y figura 50 muestran los valores de concentración de fosforo promedio en el periodo de estudio, se puede notar que existe una tendencia a disminuir en los meses de febrero, marzo y abril.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de concentración de fosforo de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la

inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

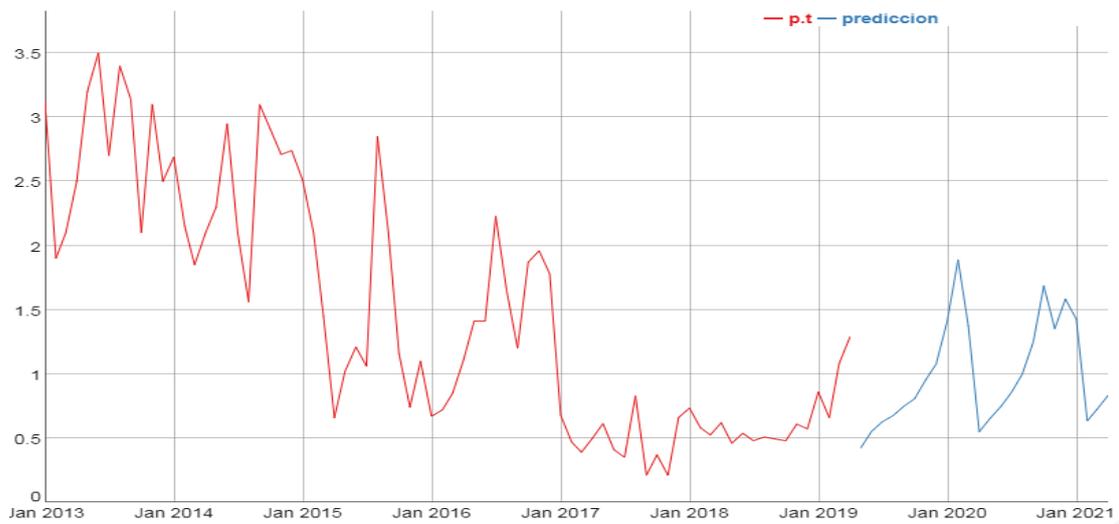


Figura 51. Tendencia y valores predichos para el parámetro concentración de fosforo de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de fosforo en la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran una disminución de los valores en el año 2017, 2018 y 2019, la tendencia es a aumentar para el año 2020.

Para obtener el modelo de red neuronal se usó 1,10 retardos y 5 nodos, con un RMSE de 0.406302059.

Estándar de calidad del agua establecido en la clase 3 del Decreto 253/79, el cual determina una concentración igual o menor a 0,025 mg/L. Los valores superan el nivel especificado.

4.18 Análisis descriptivo y tendencia del parámetro Nitrógeno amoniacal

En una contaminación reciente la mayor parte de nitrógeno está presente en forma de nitrógeno orgánico (proteína) y amoniacal, a medida que pasa el tiempo, el nitrógeno orgánico se convierte gradualmente a nitrógeno amoniacal y más tarde si existen las condiciones aeróbicas, ocurre la oxidación del amoniacal a nitritos y nitratos (DIGESA).

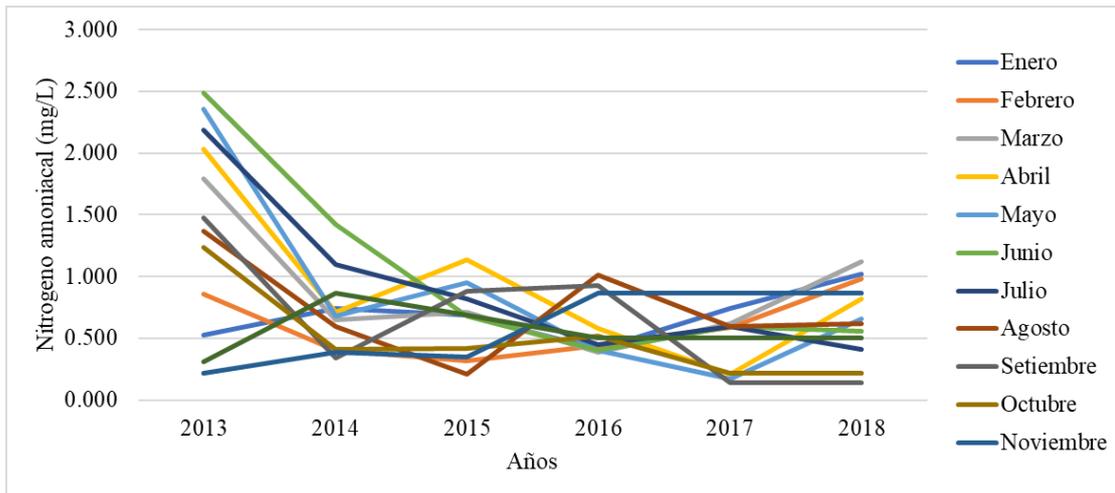


Figura 52. valores observados para nitrógeno amoniacal de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 en el mes de junio y julio se presentaron los niveles más bajos de nitrógeno amoniacal y en el mes de junio el valor más alto, la tendencia de todos los valores a partir del año 2014 tienden a disminuir hasta el año 2017 con un pequeño incremento para el año 2018.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada nitrógeno amoniacal, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 18

Estadísticos descriptivos para nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	0.69	0.61	0.91	0.90	0.87	1.03	0.93	0.74	0.65	0.51	0.60	0.56
de	0.19	0.25	0.47	0.57	0.78	0.80	0.67	0.40	0.54	0.38	0.31	0.19
max	1.02	0.98	1.79	2.03	2.36	2.49	2.19	1.37	1.48	1.24	0.87	0.87
min	0.41	0.32	0.39	0.21	0.17	0.40	0.41	0.21	0.14	0.22	0.22	0.31

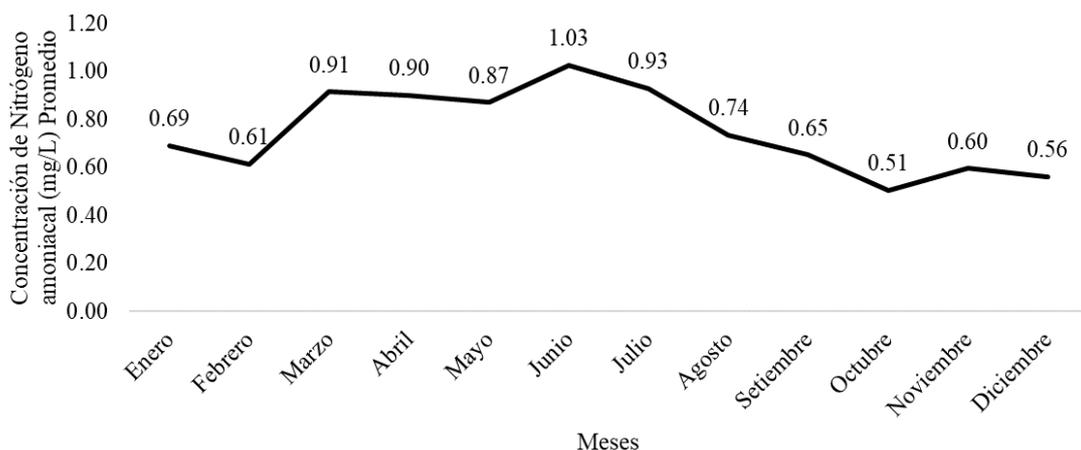


Figura 53. Valores promedio para la concentración de nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 18 y figura 53 muestran los valores de concentración de nitrógeno amoniacal promedio en el periodo de estudio, se puede notar que existe una tendencia a disminuir en los meses de agosto hasta el mes de febrero, subiendo sus valores en los meses de marzo, abril, mayo y junio.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de concentración de nitrógeno amoniacal de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

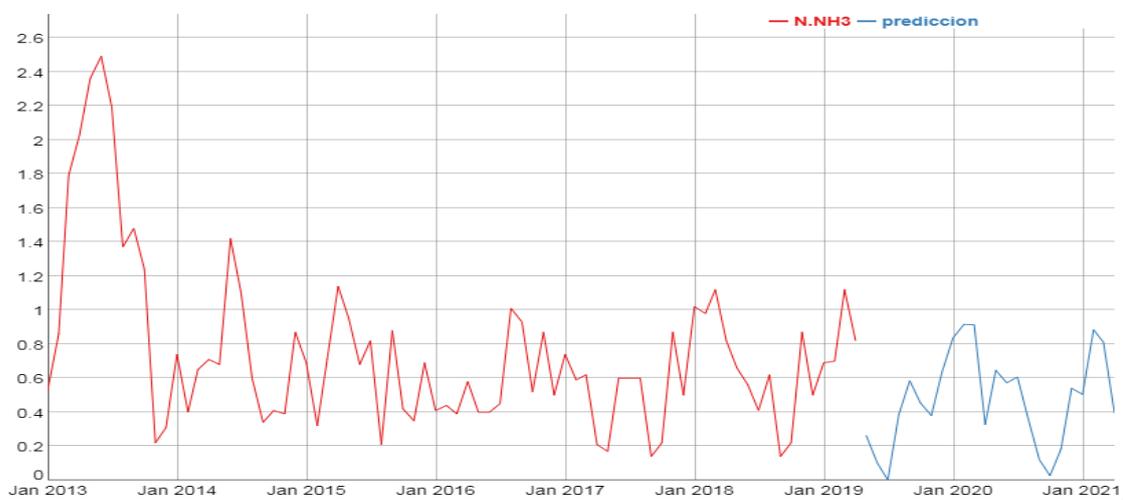


Figura 54. Tendencia y valores predichos para el parámetro nitrógeno amoniacal de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de nitrógeno amoniacal en la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran una tendencia anual, la predicción es similar a la tendencia actual.

Para obtener el modelo de red neuronal se usó 1,2,7,8,12 retardos y 4 nodos, con un RMSE de 0.170244081.

DIGESA indica que el nitrógeno amoniacal se oxida conforme avanza el tiempo, luego se forma nitritos por la acción de las bacterias nitrosomonas, los nitritos son compuesto toxico, si el nitrógeno amoniacal con concentraciones a 10 mg/L con un pH > a 8 hace que el amoniaco libre sea toxico a partir de concentraciones mayores 0.02 mg/L. A partir de concentraciones menores de 1 mg/L con un pH < 8 el amoniaco libre no presenta efectos toxicos. A1: La concentración para otros países de nitrógeno amoniacal de 0.5 mg/L. los valores superan lo establecido.

4.19 Análisis descriptivo y tendencia del parámetro Transparencia

La luz y la transparencia La luz (radiación solar incidente) es un factor abiótico esencial para la fotosíntesis (producción de oxígeno disuelto) por parte de los productores primarios, como las macrófitas acuáticas y el fitoplancton. La transparencia del agua es un parámetro físico clave para el desarrollo biológico acuático. La transparencia del agua disminuye al aumentar la cantidad de materia suspendida en ella (ejemplo: fitoplancton). En un lago poco profundo la luz entra en la zona bentónica. La presencia de luz estimula la actividad y el desarrollo del fitoplancton y, por lo tanto, la producción de oxígeno en el agua (esencial para la actividad biológica).

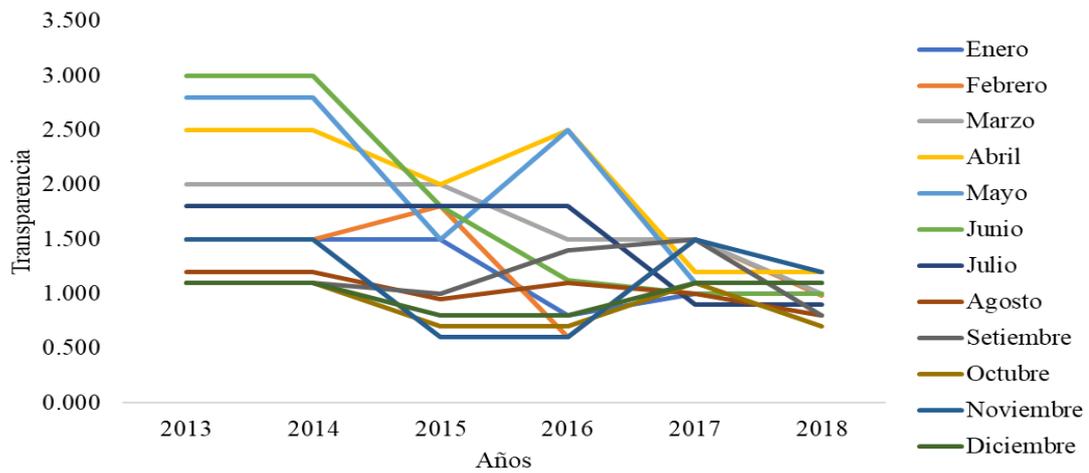


Figura 55. valores observados para la variable transparencia de la bahía interior del lago Titicaca toma puerto muelle Puno por mes según año, periodo enero 2013 – diciembre 2018.

En el año 2013 en el mes de octubre y diciembre se presentaron los niveles más bajos de transparencia y en el mes de junio el valor más alto, la tendencia de todos los valores a partir del año 2015 tienden a disminuir hasta el año 2018 con un pequeño incremento para el año 2016.

La Tabla siguiente contiene los principales estadísticos descriptivos de la variable observada transparencia, que forma parte del modelo de ecuaciones estructurales y la relación de los parámetros físico químicos de la bahía interior del lago Titicaca toma puerto muelle Puno. Para este efecto, se ha consolidado todo el periodo de estudio.

Tabla 19

Estadísticos descriptivos para la variable nitrógeno amoniacal por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

Est.	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
prom	1.16	1.30	1.49	1.81	1.97	1.82	1.50	1.04	1.15	0.90	1.15	1.00
de	0.33	0.40	0.59	0.75	0.82	0.96	0.46	0.16	0.26	0.22	0.44	0.15
max	1.50	1.80	2.00	2.50	2.80	3.00	1.80	1.20	1.50	1.10	1.50	1.10
min	0.80	0.60	0.45	0.75	1.10	1.00	0.90	0.80	0.80	0.70	0.60	0.80

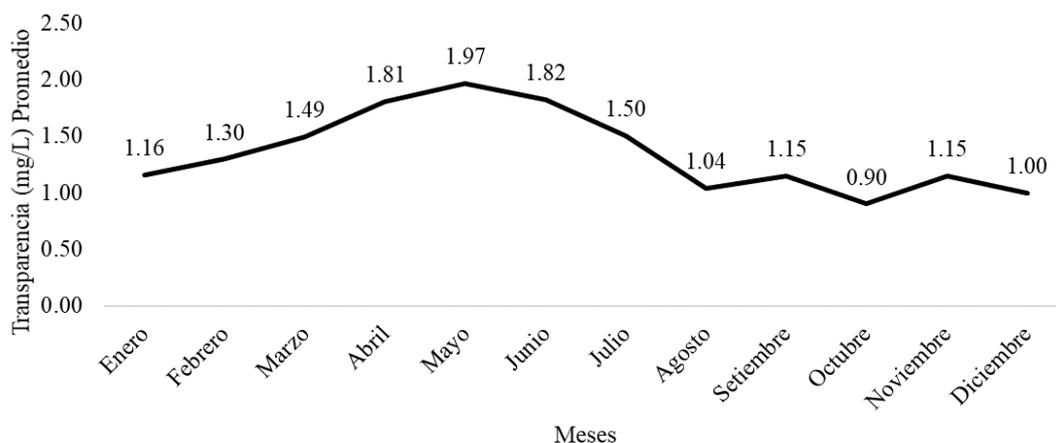


Figura 56. Valores promedio para la transparencia por mes de la bahía interior del lago Titicaca toma puerto muelle Puno, periodo enero 2013 – abril 2019.

La tabla 19 y figura 56 muestran los niveles de transparencia promedio en el periodo de estudio, se puede notar que existe una tendencia a disminuir desde el mes de agosto hasta marzo e incrementar sus valores en los meses de abril, mayo, junio y julio, con un punto más alto en el mes de mayo.

Análisis de la tendencia. La figura siguiente muestra la tendencia de los valores de transparencia de la muestra tomada en el puerto lacustre de la ciudad de Puno, para realizar el ajuste de los datos y las proyecciones se hizo uso de la una técnica de la inteligencia artificial cual es las redes neuronales artificiales ya que es un modelo que se ajusta adecuadamente los datos.

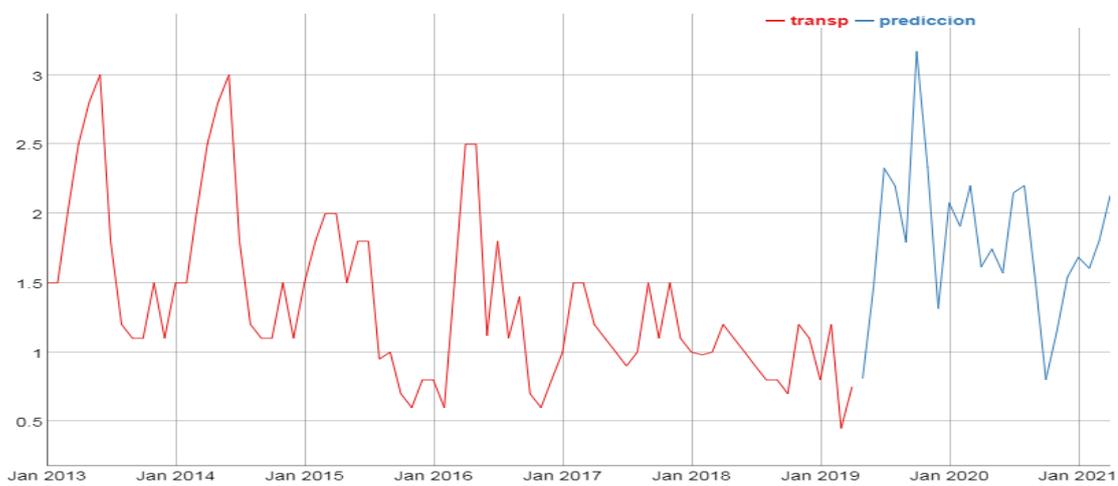


Figura 57. Tendencia y valores predichos para el parámetro nitrógeno amoniaco de la bahía interior del lago Titicaca, Toma Puerto muelle.

La figura muestra la tendencia del parámetro concentración de nitrógeno amoniacal en la bahía interior del lago Titicaca toma puerto muelle, en forma mensual desde enero del 2013 hasta marzo del 2019. Los valores muestran una tendencia anual, la predicción es similar a la tendencia actual.

Para obtener el modelo de red neuronal se usó 1,2,3,4,5,6,7,8,9,10,11 retardos y 5 nodos, con un RMSE de 0.060381584.

4.20 Análisis del principal supuesto básico (normalidad)

Los diagramas presentados en el anexo muestran los diagramas de barras, los diagramas de cajas y el grafico de cuantiles para explorar la normalidad de los datos en cada una de las variables.

Pruebas de Hipótesis de normalidad.

De los resultados que se presentan a continuación, se hará uso de la prueba de Kolmogorov-Smirnov ya que es una prueba de Hipótesis adecuada para muestras grandes, El siguiente esquema de prueba de Hipótesis se planteó para todas las variables en estudio.

1) Hipótesis estadística para comprobar la normalidad:

Ho: los datos tienden a una distribución normal

Ha: los datos no tienden a una distribución normal

2) Nivel de significancia: $\alpha = 0.01$

3) Prueba estadística: Prueba de Kolmogorov- Smirnov

4) Decisión: si $p < \alpha$, se rechaza la Hipótesis nula al 1% de significancia

Los gráficos y las pruebas de Hipótesis planteadas, muestran que 4 variables: Oxígeno disuelto, conectividad eléctrica, DBO5 y DQO tienden a la distribución normal. Al presentar falta de normalidad en los datos, es posible optar por la linealización utilizando diferentes fórmulas para el caso, para el presente trabajo se ha probado con 7 formas de linealización, las cuales mejoran el comportamiento lineal de los datos, es decir superan la falta de normalidad.

Las tablas siguientes muestran los resultados de las pruebas de Hipótesis de normalidad transformando la data y corrigiendo los outliers para poder utilizar la estadística paramétrica y los modelos adecuados.

En la tabla 20, se muestra los resultados de la prueba de normalidad, en todas las variables la significancia es mayor al 1%, es decir, $p > \alpha$, esto indica que existe normalidad en todas las variables.

4.21 Prueba de normalidad multivariante:

La prueba de normalidad multivariante, prueba la normalidad para todos los datos simultáneamente, para este caso se ha tomado dos pruebas estadísticas de Mardia y la prueba estadística de Royston.

Tabla 20

Prueba de normalidad multivariantes Mardia-tests y Royston.

	Estadístico	Sig.	Conclusión
Mardia Skewness	25,918798	0,171	SI
Mardia Kurtosis	1,2973399	0,1953	SI
MVN	NA	NA	SI
Royston	7.151419	0,0602	SI

El tests de Mardia y Royston indican que existe normalidad multivariante en los datos ya que los valores probabilísticos son mayores al nivel de significancia del 1%.

Tabla 21

Pruebas de normalidad univariantes

Parámetros físico químicos	Estadístico	Gl	Sig.
Temperatura	0,069	76	0,010
Oxígeno disuelto	0,107	76	0,031
PH	0,106	76	0,010
Conectividad eléctrica	0,192	76	0,179
SST	0,106	76	0,012
Clorofila-a	0,103	76	0,014
DBO5	0,163	76	0,200
DQO	0,119	76	0,010
Col. Totales	0,114	76	0,010
Col. Term.	0,166	76	0,025
NO2	0,129	76	0,051
NO3	0,182	76	0,032
PO4	0,161	76	0,010
N – T	0,136	76	0,010
P – T	0,152	76	0,031
N - NH3	0,159	76	0,042
Transp	0,186	76	0,032

Según Bollen (2001), existe normalidad multivariada cuando el coeficiente de Mardia es inferior a $p(p + 2)$, donde p es el número de variables observadas en el modelo, y el coeficiente de Mardia obtuvo un valor, para el modelo corregido, de 74.325, y en consecuencia podemos afirmar que existe normalidad multivariada de los datos. $p(p + 2) = 17 * 19 = 323$, que es menor al test de Mardia. También es necesario estudiar las relaciones de correlación bivalente que se puedan establecer entre las variables observadas. La Tabla 22 muestra la matriz de correlaciones y sus pruebas de significancia estadística. La prueba estadística utilizada para este caso es la correlación de Pearson por tratarse de variables cuantitativas medidas en escalas numéricas.

Tabla 22

correlaciones de las variables observadas.

	Temp	Oxígeno disuelto	PH	Conect	SST	Cloro-fila	DBO5	DQO	Col. Totales	Col. Term.	NO2	NO3	PO4	N - T	P - T	N - NH3	Tramp	
Temperatura	1	,236*	,408**	,147	-,041	-,103	,040	,296**	-,550**	-,010	-,053	-,087	,174	,041	-,156	-,248*	-,243*	
Oxígeno disuelto		1	,653**	,206	,728	,374	,733	,009	,000	,931	,647	,455	,134	,724	,180	,031	,035	
PH			1	,160	,228*	,478**	,094	-,092	-,435**	,037	,213	,173	,078	-,123	-,457**	-,647**	-,598**	
conectividad				,166	,047	,000	,421	,427	,000	,754	,065	,136	,505	,289	,000	,000	,000	
SST				1	-,035	,377**	,230*	,065	-,650**	-,070	-,171	-,233*	,003	-,298**	-,584**	-,546**	-,606**	
Clorofila-a					,766	,001	,000	,576	,000	,550	,140	,043	,980	,009	,000	,000	,000	
DBO5					1	-,138	-,203	,046	,198	,131	,332**	,207	,156	,261*	-,023	-,146	-,027	
DQO						,234	,078	,693	,087	,258	,003	,073	,177	,023	,842	,210	,817	
Col. Totales						1	,521**	-,122	-,338**	-,153	-,126	-,185	,022	-,180	-,406**	-,097	-,423**	
Col. Term.							,000	,296	,003	,187	,278	,110	,851	,120	,000	,405	,000	
NO2							1	-,125	-,396**	-,116	-,175	-,111	-,026	-,286*	-,504**	-,273*	-,489**	
NO3								,284	,000	,316	,131	,340	,821	,012	,000	,017	,000	
PO4								,075	-,135	,064	-,026	-,148	-,017	-,172	-,107	-,046	-,189	
N - T								,518	-,244	,582	,821	,203	,883	,138	,359	,690	,101	
P - T								1	-,050	,161	-,076	,131	,160	,111	-,008	,068	,150	
N - NH3									,667	,166	,515	,259	,168	,339	,943	,559	,196	
Tramp									1	,310**	,219	,233*	-,125	,233*	,336**	,370**	,519**	
										,006	,058	,043	,283	,043	,003	,001	,000	
										1	,085	,343**	-,064	-,030	,153	,067	,015	
											,467	,002	,586	,797	,186	,563	,899	
											1	,416**	,126	,102	,098	-,182	-,133	
												,000	,279	,380	,398	,115	,252	
												1	,019	,119	,073	-,116	-,130	
													1	,306	,533	,319	,265	
														,073	-,046	,136	-,050	
														,531	,693	,240	,670	
														1	,331**	,085	-,006	
															,004	,467	,961	
															1	,403**	,417**	
																,000	,000	
																1	,468**	
																	,000	,000
																	1	1

*. La correlación es significativa en el nivel 0,05 (bilateral).

En la tabla podemos observar muchas variables no correlacionadas y otras medianamente correlacionadas que en el argot estadístico se resumiría como variables incorrelacionadas, por esta razón es necesario buscar la formación de variables latentes para explicar la correlación entre estas variables observadas.

posterior al estudio de la distribución univariada, es pertinente estudiar las relaciones de correlación que se puedan establecer entre las variables observadas no latentes. analizando los coeficientes de correlación de Pearson comprobaremos la existencia de una importante estructura de correlación entre las variables observadas, aunque no muy altas, que en el argot estadístico se resumiría como variables incorrelacionadas. Por tanto, en los datos se comprueba la existencia de una estructura de correlación compleja ente las variables observadas, lo cual a su vez hace que sea pertinente pensar en el diseño de un modelo de ecuaciones estructurales y otro análisis estadístico.

Luego de observar las correlaciones entre las variables, para observar la discriminación de las variables se ha hecho uso del análisis de clúster jerárquico, El análisis de conglomerados jerárquicos (en inglés clúster) es una técnica multivariante que permite agrupar los casos o las variables en función del parecido o la similaridad entre ellos. Es un análisis por tanto que partiendo de los elementos individualmente va creando grupos hasta llegar a la formación de un único grupo o conglomerado constituido por todos los elementos de la muestra.

Esta del análisis de conglomerados es una técnica es bastante similar al análisis factorial o los componentes principales, pero es más flexible en sus supuestos (no exige normalidad ni simetría, admite variables cuantitativas y categóricas al mismo tiempo) y en la forma de calcular usa la matriz de distancias a partir de la cual se realizan los conglomerados. Para nuestro caso, se ha recurrido al método de enlace promedio (dentro de grupos) y correlación de Pearson, obteniendo los siguientes conglomerados entre las variables.

4.22 Modelamiento con análisis clúster.

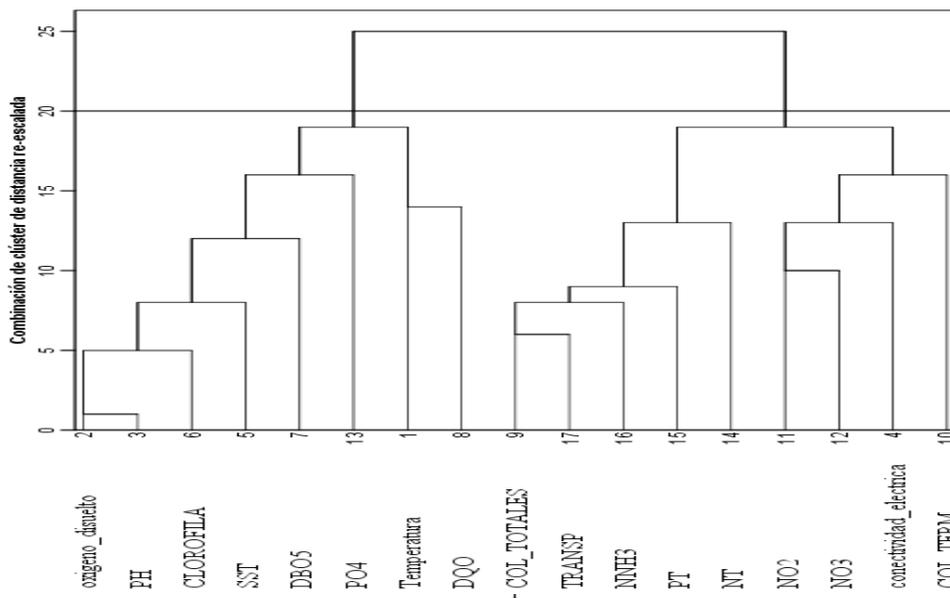


Figura 58. Dendograma para los parámetros físico químicos de la bahía interior del lago Titicaca: Toma puerto muelle.

Haciendo un corte a una distancia re-escalada de 20, la combinación de clúster indica la siguiente asociación de variables:

Cluster 1: variables en el conglomerado 1.

Oxígeno disuelto, PH, Clorofila-a, sólidos totales en suspensión, demanda bioquímica de oxígeno, fosfatos, Temperatura y demanda química de oxígeno.

Cluster 2: variables en el conglomerado 2.

Coliformes Totales, Transparencia, nitrógeno amoniacal, fosforo, Nitrógeno, Nitrito, Nitrato, Conectividad Eléctrica y Coliformes Termorreguladores.

Para el obtener las relaciones entre las variables, los modelos se formularán para las siguientes variables consideradas dependientes:

Variables Dependientes:

- Temperatura
- Oxígeno disuelto (mh/ml):
- PH

- Conectividad eléctrica
- SST (mg/L): Total de sólidos en suspensión en miligramos por litro
- Clorofila-a (mg/m³): pigmento fotosintético de las algas
- DBO5 (mg/L): Demanda bioquímica de oxígeno en miligramos por litro
- DQO (mg/L): Demanda química de oxígeno en miligramos por litro
- Transp (m): Transparencia

Variables Independientes:

Todas las demás variables incluidas las dependientes según sea el caso.

4.23 Modelamiento mediante árboles de predicción:

Los árboles de predicción son técnicas Machine learnig, que optimizan la relación y asociación de variables. Considerando una inicial falta de normalidad en los datos, se ha elegido la técnica estadística de árboles de predicción de tal manera que se pueda establecer relaciones entre las variables. Para todos los modelos, se ha utilizado 80% de los datos para el entrenamiento del modelo y 20% para la validación.

Los métodos predictivos como la regresión lineal o polinómica generan modelos globales en los que una única ecuación se aplica a todo el espacio muestral. Cuando el estudio implica múltiples predictores, que interaccionan entre ellos de forma compleja y no lineal, es muy difícil encontrar un modelo global que sea capaz de reflejar la relación entre las variables. Existen métodos de ajuste no lineal (step functions, splines,...) que combinan múltiples funciones y que realizan ajustes locales, sin embargo, suelen ser difíciles de interpretar. Los métodos estadísticos basados en árboles engloban a un conjunto de técnicas supervisadas no paramétricas que consiguen segmentar el espacio de los predictores en regiones simples, dentro de las cuales es más sencillo manejar las interacciones.

El primer modelo está basado en la variable temperatura como variable dependiente:

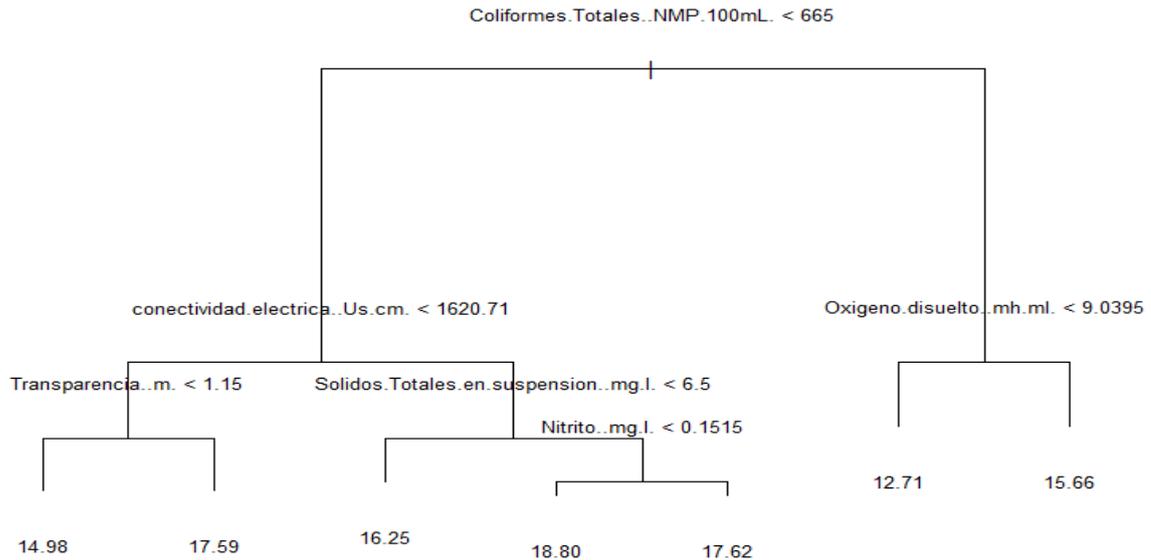


Figura 59. árbol de predicción para establecer la relación entre la variable temperatura y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en la temperatura del lago son: Coliformes totales, Conectividad eléctrica, oxígeno disuelto, transparencia, sólidos totales en suspensión, concentración de nitrito en miligramos por litro.

Predicciones:

- 1) Si los coliformes termotolerantes tienen valores $>$ a 665 NMP/100mL, los valores de oxígeno disuelto $>$ a 9.0395 mg/m³, la temperatura alcanza 15.66 °C.
- 2) Si los coliformes termotolerantes tienen valores $>$ a 665 NMP/100mL, los valores de oxígeno disuelto $<$ a 9.0395 mg/m³, la temperatura del lago alcanza 12.71 °C.
- 3) Si los coliformes termotolerantes tienen valores $<$ a 665 NMP/100mL, los valores de conectividad eléctrica $>$ a 1620.71 $\mu\text{s}/\text{cm}$, los sólidos totales $>$ 6.5 mg/l y la concentración de nitrito $>$ 0.1515 mg/l, la temperatura alcanza 17.62 °C.
- 4) Si los coliformes termotolerantes tienen valores $<$ a 665 NMP/100ml, los valores de conectividad eléctrica $>$ a 1620.71 $\mu\text{s}/\text{cm}$, los sólidos totales en suspensión $>$ 6.5 mg/l y la concentración de nitrito $<$ 0.1515 mg/l, la temperatura alcanza 18.80°C.

- 5) Si los coliformes termotolerantes tienen valores $< a$ 665 NMP/100ml, los valores de conectividad eléctrica $> a$ 1620.71 $\mu s/cm$, los sólidos totales en suspensión < 6.5 mg/l la temperatura del lago alcanza 18.80 °C.
- 6) Si los coliformes termotolerantes tienen valores $< a$ 665 NMP/100ml, los valores de conectividad eléctrica son $< a$ 1620.71 $\mu s/cm$, la transparencia del lago > 1.15 m, la temperatura del lago alcanza 17.59 °C.
- 7) Si los coliformes termotolerantes tienen valores $< a$ 665 NMP/100ml, los valores de conectividad eléctrica son $< a$ 1620.71 $\mu s/cm$, la transparencia del lago < 1.15 m, la temperatura del lago alcanza 14.98 °C.

Error de test (mse) del árbol de regresión tras podado: 5.78

El siguiente modelo, está basado en la variable oxígeno disuelto:

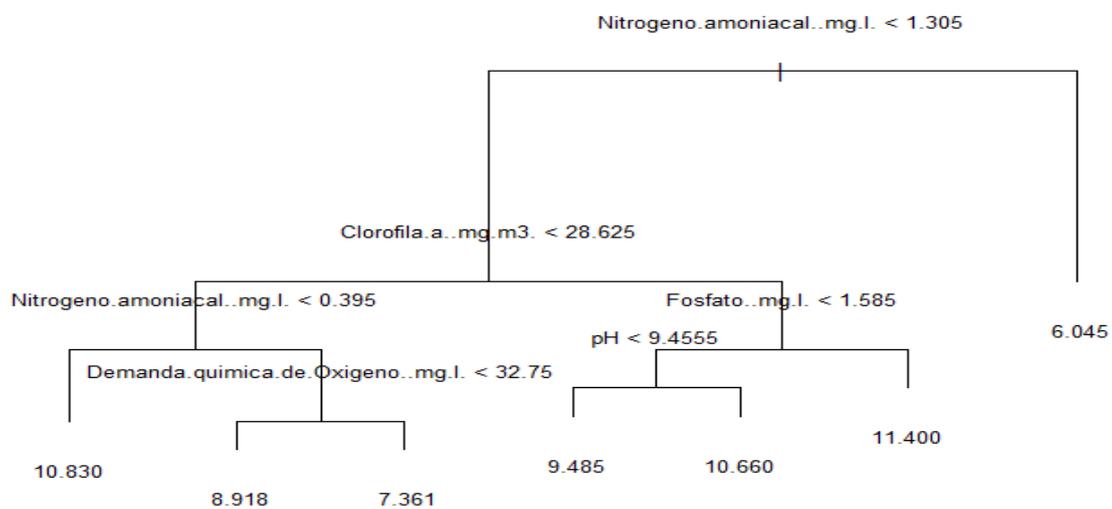


Figura 60. árbol de predicción para establecer la relación entre la variable oxígeno disuelto y los demás componentes del modelo.

Las variables más influyentes en el parámetro oxígeno disuelto son: Nitrógeno Amoniacal, Clorofila, Fosfato, pH, demanda química de oxígeno. Predicciones:

- 1) si los valores de nitrógeno amoniacal son $> a$ 1.305 mg/l, los valores de oxígeno disuelto serán de 6.045 mh/ml.
- 2) si los valores de nitrógeno amoniacal son $< a$ 1.305 mg/l, los valores de clorofila $> a$ 28.625 mg/m^3 , los valores de fosfato $> a$ 1.585 mg/l, el oxígeno disuelto será 11.400 mh/ml.

- 3) si los valores de nitrógeno amoniacal son $< 1.305 \text{ mg/l}$, los valores de clorofila $> 28.625 \text{ mg/m}^3$, los valores de fosfato $< 1.585 \text{ mg/l}$ y el pH > 9.4555 , el oxígeno disuelto será 10.660 mh/ml .
- 4) si los valores de nitrógeno amoniacal son $< 1.305 \text{ mg/l}$, los valores de clorofila $> 28.625 \text{ mg/m}^3$, los valores de fosfato $< 1.585 \text{ mg/l}$ y el pH < 9.4555 , el oxígeno disuelto será 9.485 mh/ml .
- 5) si los valores de nitrógeno amoniacal son $< 1.305 \text{ mg/l}$, los valores de clorofila $< 28.625 \text{ mg/m}^3$, los valores de nitrógeno amoniacal $> 0.395 \text{ mg/l}$ y la demanda química de oxígeno $> 32.75 \text{ mg/l}$, el valor de oxígeno disuelto será 7.361 mh/ml .
- 6) si los valores de nitrógeno amoniacal son $< 1.305 \text{ mg/l}$, los valores de clorofila $< 28.625 \text{ mg/m}^3$, los valores de nitrógeno amoniacal $> 0.395 \text{ mg/l}$ y la demanda química de oxígeno $< 32.75 \text{ mg/l}$, el valor de oxígeno disuelto será 8.918 mh/ml .
- 7) si los valores de nitrógeno amoniacal son $< 1.305 \text{ mg/l}$, los valores de clorofila $< 28.625 \text{ mg/m}^3$, los valores de nitrógeno amoniacal $< 0.395 \text{ mg/l}$, el valor de oxígeno disuelto será 10.830 mh/ml . "Error de test (mse) del árbol de regresión tras podado: 2.29".

El modelo siguiente está basado en la variable pH como variable dependiente:

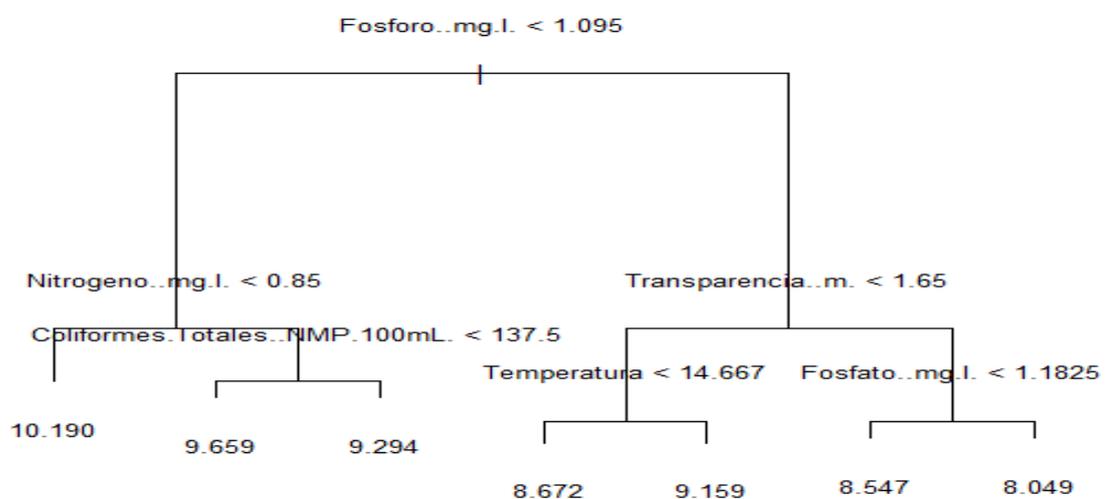


Figura 61. árbol de predicción para establecer la relación entre la variable pH y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro pH en las aguas del lago son: Fosforo, Nitrógeno, Transparencia, Coliformes totales, temperatura, fosfato.

Predicciones:

- 1) Si los niveles de fosforo son $>$ a 1.095 mg/l, la transparencia $>$ 1.165 m y los niveles de fosfato $>$ 1.1825 mg/l, el pH alcanza un nivel de 8.049.
- 2) Si los niveles de fosforo son $>$ a 1.095 mg/l, la transparencia $>$ 1.165 m y los niveles de fosfato $<$ 1.1825 mg/l, el pH alcanza un nivel de 8.547.
- 3) Si los niveles de fosforo son $>$ a 1.095 mg/l, la transparencia $<$ 1.165 m y los niveles de temperatura $>$ 14.667 °C, el pH alcanza un nivel de 9.159
- 4) Si los niveles de fosforo son $>$ a 1.095 mg/l, la transparencia $<$ 1.165 m y los niveles de temperatura $<$ 14.667 °C, el pH alcanza un nivel de 8.672.
- 5) Si los niveles de fosforo son $<$ a 1.095 mg/l, los niveles de nitrógeno $>$ 0.85 mg/l y los niveles de coliformes totales $>$ 137.5 NMP/100ml, el pH alcanza un nivel de 9.294.
- 6) Si los niveles de fosforo son $<$ a 1.095 mg/l, los niveles de nitrógeno $>$ 0.85 mg/l y los niveles de coliformes totales $<$ 137.5 NMP/100ml, el pH alcanza un nivel de 9.659.
- 7) Si los niveles de fosforo son $<$ a 1.095 mg/l, los niveles de nitrógeno $<$ 0.85 mg/l, el pH alcanza un nivel de 10.190.

"Error de test (mse) del árbol de regresión tras podado: 0.39".

El modelo está basado en la variable conectividad eléctrica como variable dependiente:

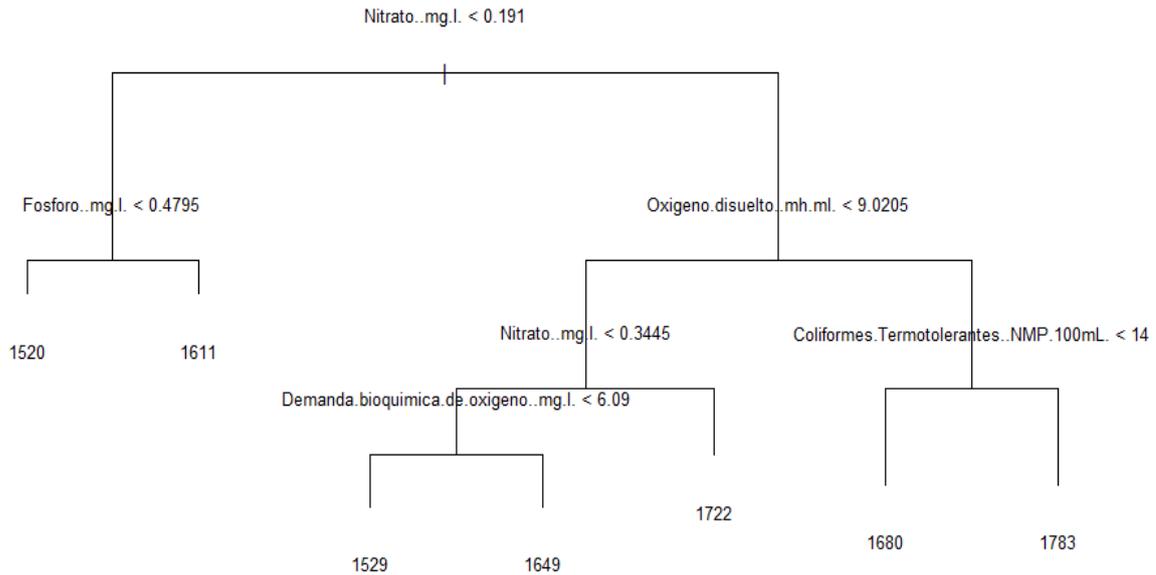


Figura 62. árbol de predicción para establecer la relación entre la variable conectividad eléctrica y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro conectividad eléctrica en las aguas del del lago son: Nitrato, Fosforo, Oxígeno disuelto, coliformes termo tolerantes, demanda bioquímica de oxígeno.

Predicciones:

- 1) si los valores de nitrato $>$ a 0.191 mg/l, los valores de oxígeno disuelto $>$ a 9.0205 mh/ml y los valores de coliformes termotolerantes $>$ a 14 NMP/100ml, la conectividad eléctrica será 1783 Us/cm.
- 2) si los valores de nitrato $>$ a 0.191 mg/l, los valores de oxígeno disuelto $>$ a 9.0205 mh/ml y los valores de coliformes termo tolerantes $<$ a 14 NMP/100ml, la conectividad eléctrica será: 1680 Us/cm.
- 3) si los valores de nitrato son $>$ a 0.191 mg/l, los valores de oxígeno disuelto $<$ a 9.0205 mh/ml y los valores de nitrato $>$ a 03445 mg/l, la conectividad eléctrica será 1722 US/cm.
- 4) si los valores de nitrato $>$ a 0.191 mg/l, los valores de oxígeno disuelto $<$ a 9.0205 mh/ml y los valores de nitrato $<$ a 03445 mg/l y la demanda bioquímica de oxígeno $>$ a 6.09 mg/l, el nivel de la conectividad eléctrica será 1649 Us/cm.

- 5) si los valores de nitrato $>$ a 0.191 mg/l, los valores de oxígeno disuelto $<$ a 9.0205 mh/ml y los valores de nitrato $<$ a 03445 mg/l y la demanda bioquímica de oxígeno $<$ a 6.09 mg/l, el nivel de la conectividad eléctrica será 1529 Us/cm.
- 6) si los valores de nitrato $<$ a 0.191 mg/l, los niveles de fosforo $>$ a 0.4795 mg/l, el nivel de la conectividad eléctrica será 1611 Us/cm.
- 7) si los valores de nitrato $<$ a 0.191 mg/l, los niveles de fosforo $<$ a 0.4795 mg/l, el nivel de la conectividad eléctrica será 1520 US/cm.

"Error de test (mse) del árbol de regresión tras podado: 13294.06"

El modelo está basado en la variable solidos totales en suspensión como variable dependiente.

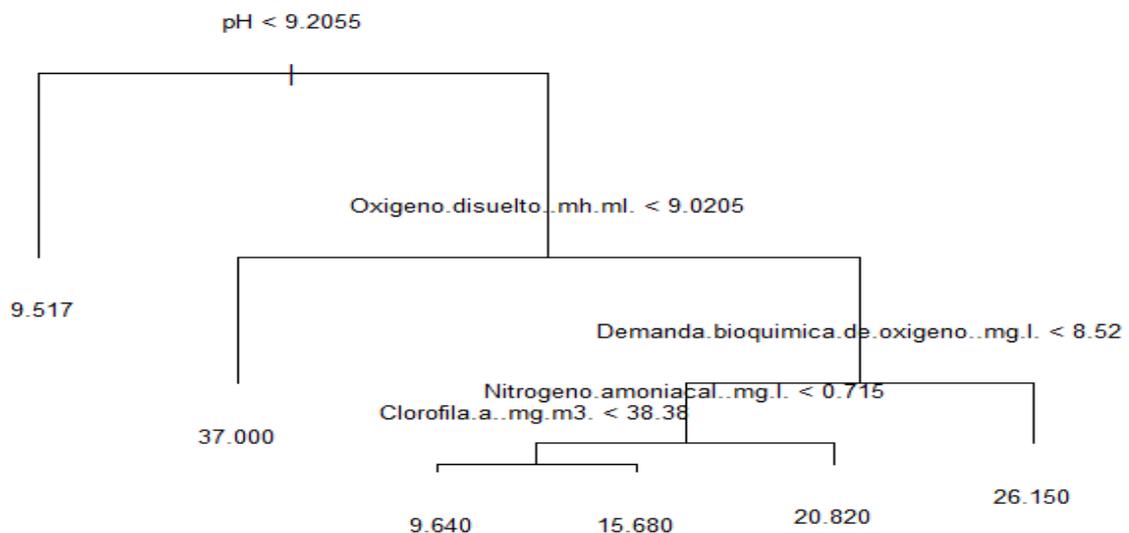


Figura 63. árbol de predicción para establecer la relación entre la variable solidos totales en suspensión y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro solidos totales en suspensión en las aguas del del lago son: pH, Oxígeno disuelto, Demanda bioquímica de oxígeno, nitrógeno amoniacal, Clorofila

Predicciones:

- 1) Si el pH $>$ 9.2055, los niveles de oxígeno disuelto $>$ 9.0205 mh/ml y la demanda bioquímica de oxígeno $>$ 8.52 mg/l, los niveles de solidos totales en suspensión alcanzan un valor de 26.150 mg/l.

- 2) Si el pH > 9.2055, los niveles de oxígeno disuelto > 9.0205 mh/ml, la demanda bioquímica de oxígeno < 8.52 mg/l y, los niveles de nitrógeno amoniacal > a 0.715 mg/l, los sólidos totales en suspensión alcanzan un valor de 20.820 mg/l.
- 3) Si el pH > 9.2055, los niveles de oxígeno disuelto > 9.0205 mh/ml, la demanda bioquímica de oxígeno < 8.52 mg/l, los niveles de nitrógeno amoniacal < a 0.715 mg/l y los niveles de clorofila > a 38.38 mg/m³, los sólidos totales en suspensión alcanzan un valor de 15.680 mg/l.
- 4) Si el pH > 9.2055, los niveles de oxígeno disuelto > 9.0205 mg/l, la demanda bioquímica de oxígeno < 8.52 mg/l y, los niveles de nitrógeno amoniacal < a 0.715 mg/l y los niveles de clorofila-a < a 38.38 mg/m³, los sólidos totales en suspensión alcanzan un valor de 9.640 mg/l.
- 5) Si el pH > 9.2055, los niveles de oxígeno disuelto < a 9.0205 mg/l, los sólidos totales en suspensión alcanzan un valor de 37.000 mg/l.
- 6) Si el pH < 9.2055, los sólidos totales en suspensión alcanzan un valor de 37.000 mg/l.

Error de test (mse) del árbol de regresión tras podado: 295.74"

El modelo está basado en la variable clorofila-a como variable dependiente.

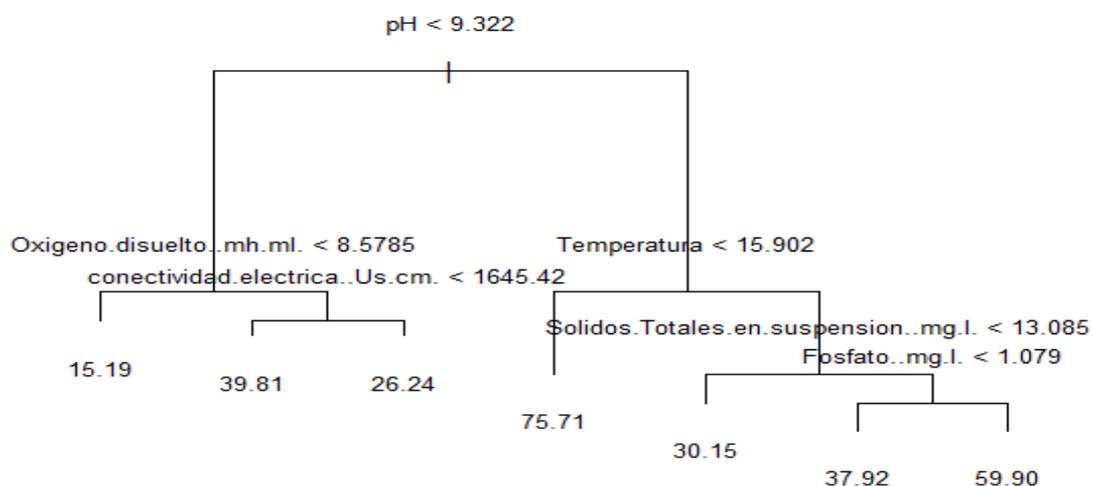


Figura 64. árbol de predicción para establecer la relación entre la variable clorofila-a y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro clorofila-a en las aguas del del lago son: pH, Oxígeno disuelto, Temperatura, Conectividad eléctrica, Solidos totales en suspensión, Fosfato

Predicciones:

- 1) Si los niveles de pH son > 9.322 , la temperatura > 15.902 °C, los sólidos totales en suspensión > 13.085 mg/l y el fosfato > 1.079 mg/l, los valores de la clorofila-a será 59.90 mg/m³.
- 2) Si los niveles de pH son > 9.322 , la temperatura > 15.902 °C, los sólidos totales en suspensión > 13.085 mg/l y el fosfato < 1.079 mg/l, los valores de la clorofila-a será 37.92 mg/m³.
- 3) Si los niveles de pH son > 9.322 , la temperatura > 15.902 °C, los sólidos totales en suspensión < 13.085 mg/l, los valores de la clorofila-a será 30.15 mg/m³.
- 4) Si los niveles de pH son > 9.322 , la temperatura < 15.902 °C, los valores de la clorofila-a será 75.71 mg/m³
- 5) Si los niveles de pH son < 9.322 , los valores de oxígeno disuelto > 8.5785 mh/ml y la conductividad eléctrica > 1645.42 Us/cm, los valores de la clorofila-a será 26.24 mg/m³.
- 6) Si los niveles de pH son < 9.322 , los valores de oxígeno disuelto > 8.5785 mh/ml y la conductividad eléctrica < 1645.42 Us/cm, los valores de la clorofila-a será 39.81 mg/m³.
- 7) Si los niveles de pH son < 9.322 , los valores de oxígeno disuelto < 8.5785 mh/ml, los valores de la clorofila-a será 15.19 mg/m³

"Error de test (mse) del árbol de regresión tras podado: 188.58"

El modelo está basado en la variable demanda bioquímica de oxígeno como variable dependiente.

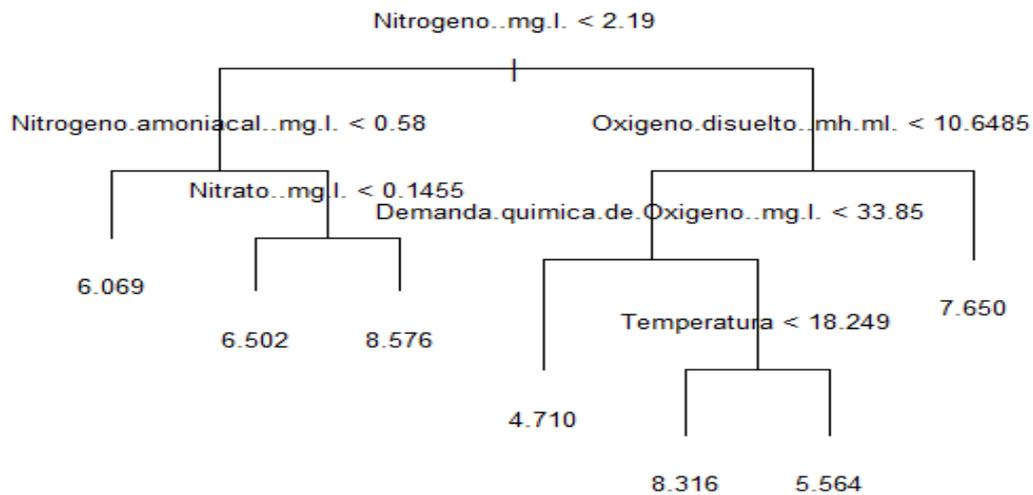


Figura 65. árbol de predicción para establecer la relación entre la variable demanda bioquímica de oxígeno y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro demanda bioquímica de oxígeno en las aguas del del lago son: Nitrógeno, Nitrógeno amoniacal, Oxígeno disuelto, Nitrato, Demanda Química de oxígeno, Temperatura.

Predicciones:

- 1) Si los valores de nitrógeno superan 2.19 mg/l y los valores de oxígeno disuelto también > a 10.6485 mg/l, la demanda bioquímica de oxígeno alcanza un nivel de 7.650 mg/l.
- 2) Si los valores de nitrógeno superan (>) 2.19 mg/l, los valores de oxígeno disuelto < a 10.6485 mh/ml, la demanda química de oxígeno > 33.85 mg/l y la temperatura > a 18.249 °C, la demanda bioquímica de oxígeno alcanza un nivel de 5.564 mg/l.
- 3) Si los valores de nitrógeno superan (>) 2.19 mg/l, los valores de oxígeno disuelto < a 10.6485 mh/ml, la demanda química de oxígeno > 33.85 mh/ml y la temperatura < a 18.249 °C, la demanda bioquímica de oxígeno alcanza un nivel de 8.316 mg/l.
- 4) Si los valores de nitrógeno superan (>) 2.19 mg/l, los valores de oxígeno disuelto < a 10.6485 mh/ml, la demanda química de oxígeno < 33.85 mg/l, la demanda bioquímica de oxígeno alcanza un nivel de 4.710 mg/l.

- 5) Si los valores de nitrógeno son < 2.19 mg/l, los valores de nitrógeno amoniacal > 0.58 mg/ml y los valores de nitrato > 0.1455 mg/l, la demanda bioquímica de oxígeno alcanza un nivel de 8.576 mg/l.
- 6) Si los valores de nitrógeno son < 2.19 mg/l, los valores de nitrógeno amoniacal > 0.58 mg/l y los valores de nitrato < 0.1455 mg/l, la demanda bioquímica de oxígeno alcanza un nivel de 6.502 mg/l.
- 7) Si los valores de nitrógeno son < 2.19 mg/l, los valores de nitrógeno amoniacal < 0.58 mg/l, la demanda bioquímica de oxígeno alcanza un nivel de 6.069 mg/l.

"Error de test (mse) del árbol de regresión tras podado: 7.17"

La variable demanda química de oxígeno como variable dependiente:

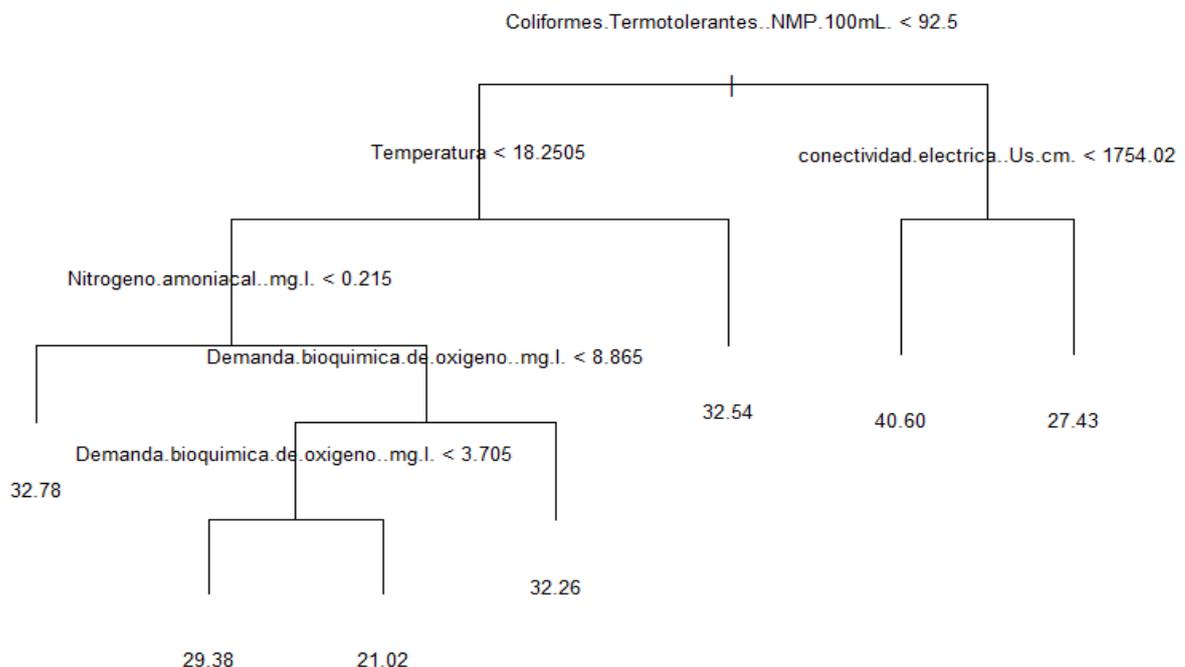


Figura 66. Árbol de predicción para establecer la relación entre la variable demanda química de oxígeno y los demás componentes del modelo.

En la figura podemos observar que: las variables más influyentes en el parámetro demanda química de oxígeno en las aguas del del lago son: Coliformes termotolerantes, Temperatura, Conectividad eléctrica, Nitrógeno amoniacal, Demanda bioquímica de oxígeno.

Predicciones:

- 1) Si los coliformes termotolerantes alcanzan valores > 92.5 NMP/100ml, la conectividad eléctrica tiene valores $>$ a 1754.02 Us/cm, la demanda química de oxígeno alcanza 27.43 mg/l
- 2) Si los coliformes termotolerantes alcanzan valores > 92.5 NMP/100ml, la conectividad eléctrica tiene valores $<$ a 1754.02 Us/cm, la demanda química de oxígeno alcanza 40.60 mg/l.
- 3) Si los coliformes termotolerantes alcanzan valores < 92.5 NMP/100ml, la temperatura tiene valores $>$ a 18.2505 °C, la demanda química de oxígeno alcanza 32.54 mg/l.
- 4) Si los coliformes termotolerantes alcanzan valores < 92.5 NMP/100ml, la temperatura tiene valores $<$ a 18.2505 °C, los valores de nitrógeno son > 0.215 mg/l y la demanda bioquímica de oxígeno es $>$ a 8.865 mg/l, la demanda química de oxígeno alcanza 32.26 mg/l.
- 5) Si los coliformes termotolerantes alcanzan valores < 92.5 , la temperatura tiene valores $<$ a 18.2505 °C, los valores de nitrógeno amoniacal son > 0.215 mg/l, la demanda bioquímica de oxígeno es $>$ a 8.865 mg/l y la demanda bioquímica de oxígeno > 3.705 mg/l, la demanda química de oxígeno alcanza 21.02 mg/l.
- 6) Si los coliformes termotolerantes alcanzan valores < 92.5 , la temperatura tiene valores $<$ a 18.2505 °C, los valores de nitrógeno amoniacal son > 0.215 mg/l, la demanda bioquímica de oxígeno es $>$ a 8.865 mg/l y la demanda bioquímica de oxígeno < 3.705 mg/l, la demanda química de oxígeno alcanza 29.38 mg/l.
- 7) Si los coliformes termotolerantes alcanzan valores < 92.5 NMP/100ml, la temperatura tiene valores $<$ a 18.2505 °C, los valores de nitrógeno amoniacal son < 0.215 mg/l, la demanda química de oxígeno alcanza 32,78 mg/l.

"Error de test (mse) del árbol de regresión tras podado: 58.02"

El modelo está basado en la variable transparencia como variable dependiente:

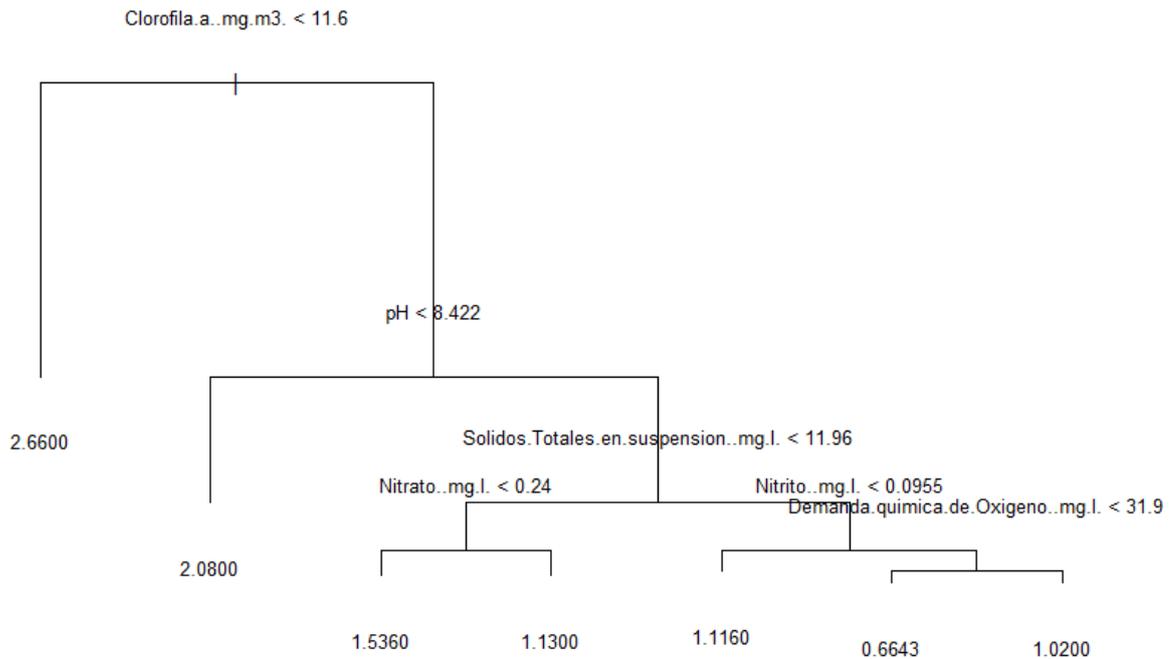


Figura 67. Árbol de predicción para establecer la relación entre la variable transparencia y los demás componentes del modelo.

En la figura 67 podemos observar que: las variables más influyentes en el parámetro transparencia en las aguas del del lago son: Clorofila-a, pH, Sólidos totales en suspensión, Nitrito, Nitrito, Demanda química de oxígeno.

Predicciones:

- 1) Si los valores de clorofila-a > 11.6 mg/m³, los valores de pH > a 8.422, los sólidos totales en suspensión > 11.9 mg/l, los valores de nitrito > a 0.0955 mg/l y la demanda química de oxígeno > 31.9 mg/l, la transparencia tendrá un valor de 1.0200 m.
- 2) Si los valores de clorofila-a > 11.6 mg/m³, los valores de pH > a 8.422, los sólidos totales en suspensión > 11.9 mg/l, los valores de nitrito > a 0.0955 mg/l y la demanda química de oxígeno < 31.9 mg/l, la transparencia tendrá un valor de 0.6643 m.
- 3) Si los valores de clorofila-a > 11.6 mg/m³, los valores de pH > a 8.422, los sólidos totales en suspensión > 11.9 mg/l, los valores de nitrito < a 0.0955 mg/l, la transparencia tendrá un valor de 1.1160 m.

- 4) Si los valores de clorofila-a $> 11.6 \text{ mg/m}^3$, los valores de pH > 8.422 , los sólidos totales en suspensión $< 11.9 \text{ mg/l}$, los valores de nitrato $> 0.24 \text{ mg/l}$, la transparencia tendrá un valor de 1.1300 m.
- 5) Si los valores de clorofila-a $> 11.6 \text{ mg/m}^3$, los valores de pH > 8.422 , los sólidos totales en suspensión $< 11.9 \text{ mg/l}$, los valores de nitrato $< 0.24 \text{ mg/l}$, la transparencia tendrá un valor de 1.5360 m.
- 6) Si los valores de clorofila-a $> 11.6 \text{ mg/m}^3$, los valores de pH < 8.422 , la transparencia tendrá un valor de 2.0800 m.
- 7) Si la clorofila-a $< 11.6 \text{ mg/m}^3$, la transparencia tendrá un valor de 2.6600 m.

"Error de test (mse) del árbol de regresión tras podado: 0.13"

4.24 Análisis Exploratorio de datos, formación de variables latentes

Para determinar las variables latentes se hizo uso de la técnica estadística multivariada análisis factorial exploratorio, obteniendo los siguientes resultados:

Análisis factorial para las variables dependientes:

Los resultados muestran que la determinante de variables observadas es $det = 0.065$.

La matriz de correlaciones indica que la determinante no es cero, comparando con la teoría estadística que indica que la matriz debe ser próxima a cero, pero no cero, indica que estamos por buen camino al utilizar el análisis factorial para determinar las variables latentes, para confirmar esta afirmación se recurre a la prueba estadística de Kaiser-Meyer-Olkin y Bartlett

Tabla 23

Resultados de la Prueba de KMO y Bartlett

Estadísticos	
Medida Kaiser-Meyer-Olkin de adecuación de muestreo	0,727
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado 194,990
	Grados de libertad 36
	Significancia 0,000

La prueba estadística de KMO (Medida Kaiser-Meyer-Olkin) muestra un valor superior a 0.5 con un chi-cuadrado significativo $p(0.00) < \alpha(0.05)$, que indica la idoneidad del uso del análisis factorial y que las variables latentes estarán bien discriminadas en grupos factoriales.

Para observar el aporte de las variables observadas se recurre a la matriz anti-imagen donde se espera valore mayores a 0.5 en la diagonal principal de las correlaciones anti-imagen.

La matriz anti-imagen muestra una estructura de las variables adecuada, todas las correlaciones anti imagen, excepto conectividad son mayores a 0.5, esto confirma la buena relación entre las variables observadas y que en los componentes de la matriz habrá buena discriminación.

Tabla 24

Varianza total explicada de las variables observadas

Componente	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,125	34,722	34,722	3,125	34,722	34,722
2	1,595	17,725	52,446	1,595	17,725	52,446
3	1,176	13,063	65,510	1,176	13,063	65,510
4	,861	9,567	75,076			
5	,687	7,630	82,706			
6	,602	6,692	89,398			
7	,375	4,165	93,563			
8	,326	3,628	97,191			
9	,253	2,809	100,000			

Método de extracción: análisis de componentes principales.

La tabla muestra que, utilizando los autovalores donde se especifica como valor mínimo la unidad (1), muestra que las variables observadas pueden asociarse en tres componentes y éstas explican el 65.5% de la varianza total explicada. El primer componente explica el

34.722%, el segundo componente explica el 17.725% y el tercer componente explica el 13.063% respectivamente. Si observamos el cuarto y de más componentes, las varianzas totales explicadas, no superan el 10%, razón por la cual se toma la decisión de formar tres componentes para obtener las variables latentes para el análisis de ecuaciones estructurales. Las sumas de cargas al cuadrado de la extracción son similares al uso de autovalores iniciales. La tabla siguiente, muestra las componentes discriminadas.

Tabla 25

Formación de componentes con las variables observadas

	Componente 1	Componente 2	Componente 3
Temperatura	,282	,771	,172
Oxígeno disuelto	,769	,223	-,332
PH	,839	,257	,063
Conectividad eléctrica	-,054	,519	-,493
SST	,627	-,390	,178
Clorofila-a	,737	-,381	,018
DBO5	,359	-,126	,579
DQO	-,094	,547	,623
Transp	-,820	-,055	,179

Método de extracción: análisis de componentes principales.

a. 3 componentes extraídos.

Los resultados claramente muestran el poder discriminatorio del análisis:

- Componentes 1 o constructo 1: Oxígeno disuelto, pH, Sólidos totales en suspensión, Clorofila-a y transparencia.
- Componentes 2 o constructo 2: Temperatura y conectividad eléctrica
- Componentes 3 o constructo 3: Demanda Bioquímica de Oxígeno y Demanda Química de Oxígeno.

Esto explica que existen tres componentes internamente muy correlacionados entre ellos y externamente muy incorrelacionados, se presume que aparezcan 3 variables latentes.

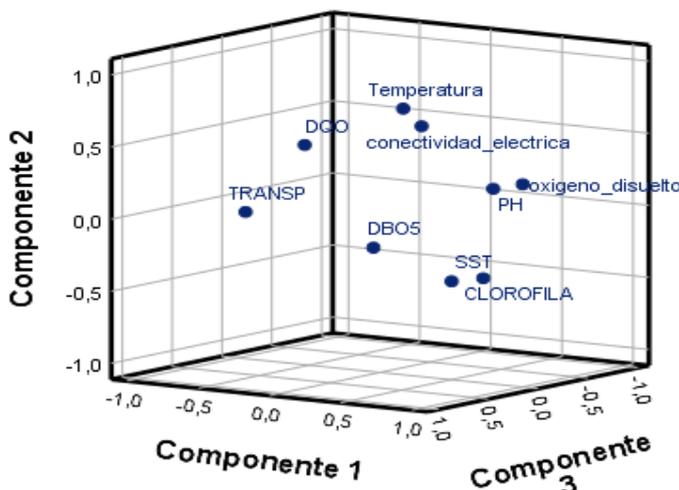


Figura 68. Matriz de componentes para formar las variables latentes.

La figura muestra gráficamente los tres componentes correlacionados entre las variables observadas que forman las variables latentes en el estudio.

Análisis factorial para las variables Independientes:

Los resultados muestran que la determinante de la Matriz $det = 0.246$.

La matriz de correlaciones indica que la determinante no es cero, comparando con la teoría estadística que indica que la matriz debe ser próxima a cero, pero no cero, indica que estamos por buen camino al utilizar el análisis factorial para determinar las variables latentes, para confirmar esta afirmación se recurre a la prueba estadística de Kaiser-Meyer-Olkin y Bartlett

Tabla 26

Resultados de la Prueba de KMO y Bartlett

Estadísticos	
Medida Kaiser-Meyer-Olkin de adecuación de muestreo	0,498
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado 100,228
	Grados de libertad 28
	Significancia 0,000

La prueba estadística de KMO muestra un valor menor a 0.5, pero próximo a ello, con un chi-cuadrado significativo $p(0.00) < \alpha(0.05)$, esto indica la idoneidad del uso del análisis factorial y que las variables latentes estarán bien discriminadas.

Para observar el aporte de las variables observadas se recurre a la matriz anti-imagen donde se espera valore mayores a 0.5 en la diagonal principal de las correlaciones anti-imagen.

La matriz anti-imagen muestra una estructura de las variables es adecuada, todas las correlaciones anti imagen, excepto PO4, son mayores a 0.5.

Observemos la varianza total para verificar la formación de grupos:

Tabla 27

Varianza total explicada de las variables observadas independientes

Componente	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
	1	2,128	26,595	26,595	2,128	26,595
2	1,584	19,805	46,399	1,584	19,805	46,399
3	1,176	14,699	61,099	1,176	14,699	61,099
4	1,017	12,712	73,811			
5	,740	9,245	83,055			
6	,609	7,614	90,669			
7	,434	5,431	96,100			
8	,312	3,900	100,000			

Método de extracción: análisis de componentes principales.

La tabla muestra que, utilizando los autovalores donde se especifica como valor mínimo la unidad (1), muestra que las variables observadas pueden asociarse en cuatro componentes, el porcentaje de explicación aumenta a 73.811%, pero existen variables observadas que no se asociación con otras, debido a esto se decide agrupar en 3

componentes. La varianza total indica que las variables estas explicadas en el 61.099% con tres componentes. El primer componente explica el 26.595%, el segundo componente explica el 19.805 % y el tercer componente explica el 14.699 respectivamente.

La tabla siguiente, muestra las componentes discriminadas.

Tabla 28

Formaci3n de componentes con las variables observadas independientes

	Componente 1	Componente 2	Componente 3
COL. TOTALES	,751	-,179	-,256
COL. TERM.	,497	,172	-,569
NO2	,424	,611	,054
NO3	,538	,665	,011
PO4	,121	,324	,747
N – T	,469	-,125	,413
P – T	,651	-,413	,109
N - NH3	,435	-,644	,209

M3todo de extracci3n: analisis de componentes principales.

Los resultados claramente muestran el poder discriminatorio del analisis:

- Componentes 1: Coliformes Totales, Nitr3geno y Fosforo
- Componentes 2: Nitrito, Nitrato y Nitr3geno amoniacal
- Componentes 3: Coliformes Termotolerantes y fosfato

Esto explica que existen tres componentes internamente muy correlacionados entre ellos y externamente muy incorrelacionados.

Para observar de una manera mas objetiva, se hace uso de graficos de componentes en una escala tridimensional.

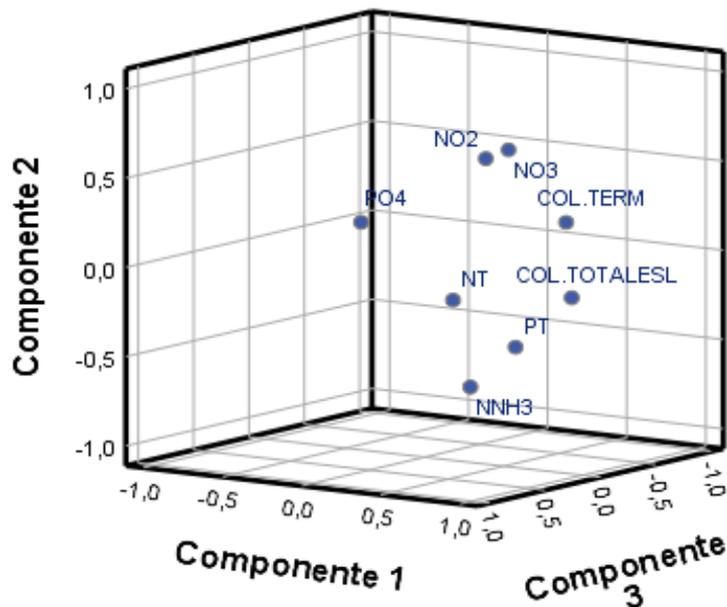


Figura 69. distribución tridimensional de los componentes del modelo factorial

La figura muestra los componentes en un espacio tridimensional en el cual se puede observar los factores asociados a las componentes.

El Gráfico siguiente nos ayuda a comprender de una forma rápida y sencilla cómo se construyen estas variables latentes a partir de las variables observadas que entran en el modelo de ecuaciones estructurales inicial. Tanto las relaciones de covariación como de regresión entre las variables latentes se analizarán detenidamente en los siguientes puntos; así como también los parámetros a estimar del modelo.

Modelo de Ecuaciones Estructurales Inicial.

El modelo de ecuaciones estructurales inicial se construye a partir de las 17 variables observadas. A partir de ellas se crean 6 variables latentes diferentes a través de análisis factorial confirmatorio. Una vez creadas las variables latentes, se determinan las relaciones que éstas establecen entre sí, pudiendo ser relaciones con correlación y/o, relaciones de regresión. En el Gráfico inicial se pueden apreciar, además de las relaciones que se establecen entre las variables, algunos parámetros a estimar.

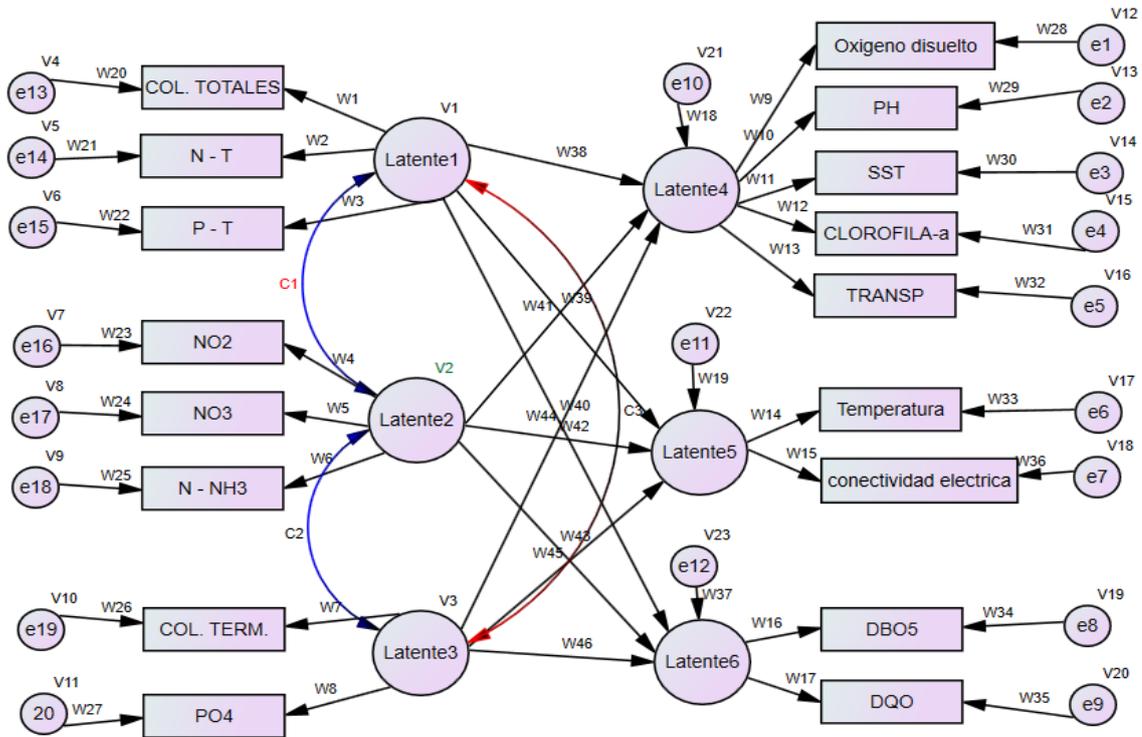


Figura 70. Modelo de ecuaciones estructurales inicial

Siguiendo con un análisis más pormenorizado del Gráfico inicial, los componentes del modelo de ecuaciones estructurales inicial son los siguientes:

a. Variables latentes:

- Exógenas:

- Latente 1: ξ_1 ; Latente 2: ξ_2 ; Latente 3: ξ_3

- Endógenas

- Latente 4: η_1 ; Latente 5: η_2 ; Latente 6: η_3

b. Variables observadas:

- Endógenas

Temperatura, Oxígeno disuelto (mh/ml), pH, Conectividad eléctrica, SST (mg/L), Clorofila-a (mg/m³), DBO5 (mg/L), DQO (mg/L), Transparencia.

- Exógenas

Col Totales (NMP/100mL), Col Term, NO₂ (mg/L), NO₃ (mg/L), PO₄ (mg/L), N- T (mg/L), P-T: (mg/L), N-NH₃ (mg/L).

c. Errores de medida:

- De las variables observadas exógenas

e13, e14, e15, e16, e17, e18, e19, e20 (en el modelo como δ)

- De las variables observadas exógenas

e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11, e12 (en el modelo como δ)

d. Términos de perturbación:

ζ_1, ζ_2 que incluyen los efectos de las variables omitidas, los errores de medida y la aleatoriedad del proceso especificado, la variación en el término de perturbación se denota por ψ y la covariación entre los términos de perturbación i-esimo y j-esimo se denota por ψ_{ij} .

e. Coeficientes de regresión:

Son las que relacionan las variables latentes con las observadas λ_x, λ_y (W_i).

f. coeficientes de regresión:

γ_{ij} representan la relación entre variables latentes y una endógena, β_{ij} (relacionan las variables latentes endógenas entre sí) y, ϕ_{ij} representan la covariación entre las variables latentes exógenas).

Recordemos que los modelos de ecuaciones estructurales están compuestos por el modelo estructural, propiamente dicho y un modelo de medida. Tal y como ya se mencionó en el punto anterior, la representación del modelo estructural sigue la expresión.

$$\eta = \beta\eta + \Gamma\xi + \zeta$$

Identificación del modelo inicial. Para el modelo representado en el gráfico anterior, el número total de variables endógenas es $p = 9$, variables exógenas $q = 8$, por lo que $s = 17$; en virtud de lo cual, el número de elementos no redundantes en Σ es: $\Sigma = \frac{s(s+1)}{2} = \frac{17(18)}{2} = 153$

En este caso, el número total de parámetros del modelo a ser estimados, denotados por t , asciende a $t = 49$. como $t(49) < \sum = 153$, podemos afirmar que el modelo se encuentra sobre-identificado, por lo que el modelo cumple la condición necesaria para estar identificado. Además, es importante señalar que estamos ante un modelo recursivo, puesto que no existe ninguna correlación entre residuos.

$$\text{además, } g = \frac{s(s+1)}{2} - t = 153 - 49 = 104 > 0,$$

donde g son los grados de libertad,

Mientras más grados de libertad, más parsimonioso es el modelo.

Recuentos variables (Grupo número 1)

Número de variables en su modelo: 43

Número de variables observadas: 17

Número de variables no observadas: 26

Número de variables exógenas: 23

Número de variables endógenas: 20

Estimación y ajuste. Presentación de resultados en este acápite, hay examinar la existencia de posibles estimaciones infractoras, o lo que es lo mismo, coeficientes estimados tanto en los modelos de medida como en los estructurales que excedan los límites aceptables. Esto es:

- Varianzas de error negativas o varianzas de error no significativas para cualquiera de los constructos.
- Coeficientes estandarizados cuyos valores sobrepasen 1.000.
- Errores estándar muy elevados asociados con cualquier coeficiente estimado.

El método de estimación elegido es el de máxima verosimilitud.

En primer lugar, corresponde analizar el ajuste del modelo de medida, así como las relaciones entre las distintas variables que conforman el Modelo 1, partiendo de los distintos parámetros, estimados por máxima verosimilitud, que aparecen reflejados en el Gráfico.

Los valores que aparecen junto a las flechas que van desde los constructos a las variables de medida indican la “carga factorial” de las variables indicadoras o directamente observables en el factor común. Por su parte, en la parte superior de los rectángulos que representan a las variables indicadoras encontramos el valor de la varianza explicada en la variable por el constructo (R^2), que podemos entender como la fiabilidad de la medida.

Tabla 29

Fiabilidad de la medida.

Variables y factores latentes	Estimate
L6D	.018
L5D	.401
L4D	1.087
Demanda Química de Oxígeno	.002
Demanda Bioquímica de oxígeno	3.415
Conectividad eléctrica	.006
Temperatura	7.037
Transparencia	.487
Clorofila-a	.389
Sólidos Totales en Suspensión	.232
pH	.786
Oxígeno disuelto	.518
Coliformes termotolerantes	-.050
Fosfato	-.115
Nitrito	-.005
Nitrato	-.046
Nitrógeno amoniacal	-.206
Coliformes totales	.555
Nitrógeno	.036
Fosforo	.305

Fuente: Salida AMOS en base a los datos

- 1) Se estima que los predictores de L6D explican 1.8 por ciento de su varianza. En otras palabras, la varianza de error de L6D es aproximadamente el 98.2 por ciento de la varianza de L6D.
- 2) Se estima que los predictores de L5D explican el 40.1 por ciento de su varianza. En otras palabras, la varianza de error de L5D es aproximadamente el 59.9 por ciento de la varianza de L5D.

- 3) Se estima que los predictores de L4D explican 108.7 por ciento de su varianza. En otras palabras, la varianza de error de L4D es aproximadamente -8.7 por ciento de la varianza de L4D.
- 4) Se estima que los predictores de la Demanda Química de Oxígeno explican el 0.2 por ciento de su varianza. En otras palabras, la varianza de error de Demanda Química de Oxígeno es aproximadamente el 99.8 por ciento de la varianza de Demanda Química de Oxígeno.
- 5) Se estima que los predictores de Demanda Bioquímica de Oxígeno explican el 341.5 por ciento de su varianza. En otras palabras, la varianza de error de Demanda Bioquímica de Oxígeno es aproximadamente -241.5 por ciento de la varianza de Demanda Bioquímica de Oxígeno en sí.
- 6) Se estima que los predictores de Conectividad eléctrica explican 0.6 por ciento de su varianza. En otras palabras, la varianza de error de Conectividad eléctrica es aproximadamente el 99.4 por ciento de la varianza de Conectividad eléctrica.
- 7) Se estima que los predictores de Temperatura explican el 703.7 por ciento de su varianza. En otras palabras, la varianza de error de Temperatura es aproximadamente -603.7 por ciento de la varianza de Temperatura misma.
- 8) Se estima que los predictores de Transparencia explican el 48.7 por ciento de su varianza. En otras palabras, la varianza de error de Transparencia es aproximadamente el 51.3 por ciento de la varianza de Transparencia.
- 9) Se estima que los predictores de Clorofila-a explican el 38.9 por ciento de su varianza. En otras palabras, la varianza de error de Clorofila-a es aproximadamente el 61.1 por ciento de la varianza de Clorofila-a.
- 10) Se estima que los predictores de Sólidos Totales en Suspensión explican el 23.2 por ciento de su varianza. En otras palabras, la varianza del error de Sólidos Totales en Suspensión es aproximadamente el 76.8 por ciento de la varianza de la Sólidos Totales en Suspensión.
- 11) Se estima que los predictores de pH explican el 78.6 por ciento de su varianza. En otras palabras, la varianza de error de pH es aproximadamente el 21.4 por ciento de la varianza de pH en sí.



- 12) Se estima que los predictores de Oxígeno disuelto explican el 51.8 por ciento de su varianza. En otras palabras, la varianza de error de Oxígeno disuelto es aproximadamente el 48.2 por ciento de la varianza de Oxígeno disuelto en sí.
- 13) Se estima que los predictores de Coliformes termotolerantes explican -5 por ciento de su varianza. En otras palabras, la varianza de error de Coliformes termotolerantes es aproximadamente el 105 por ciento de la varianza de Coliformes termotolerantes.
- 14) Se estima que los predictores de Fosfato explican -11.5 por ciento de su varianza. En otras palabras, la varianza de error de Fosfato es aproximadamente el 111.5 por ciento de la varianza de Fosfato.
- 15) Se estima que los predictores de Nitrito explican -0.5 por ciento de su varianza. En otras palabras, la varianza de error de Nitrito es aproximadamente el 100.5 por ciento de la varianza de Nitrito en sí.
- 16) Se estima que los predictores de Nitrato explican -4,6 por ciento de su varianza. En otras palabras, la varianza de error de Nitrato es aproximadamente el 104.6 por ciento de la varianza de Nitrato en sí.
- 17) Se estima que los predictores de Nitrógeno amoniacal explican -20.6 por ciento de su varianza. En otras palabras, la varianza de error de Nitrógeno amoniacal es aproximadamente el 120.6 por ciento de la varianza de Nitrógeno amoniacal.
- 18) Se estima que los predictores de Coliformes Totales explican el 55.5 por ciento de su varianza. En otras palabras, la varianza de error de Coliformes Totales es aproximadamente el 44.5 por ciento de la varianza de Coliformes Totales.
- 19) Se estima que los predictores de Nitrógeno explican el 3.6 por ciento de su varianza. En otras palabras, la varianza de error de Nitrógeno es aproximadamente el 96.4 por ciento de la varianza de Nitrógeno en sí.
- 20) Se estima que los predictores de Fosforo explican el 30.5 por ciento de su varianza.

Tabla 30

Pesos de regresión estandarizados

Variables y factores latentes		Estimate
L4D	<--- L1I	-.558
L5D	<--- L1I	-1.505
L6D	<--- L1I	-.046
Fosforo	<--- L1I	.553
Nitrógeno	<--- L1I	.190
Coliformes Totales	<--- L1I	.745
Oxígeno disuelto	<--- L4D	.720
pH	<--- L4D	.887
Solidos Totales en Suspensión	<--- L4D	.482
Clorofila-a	<--- L4D	.624
Transparencia	<--- L4D	-.698
Temperatura	<--- L5D	2.653
Conectividad eléctrica	<--- L5D	.077
Demanda Bioquímica de oxígeno	<--- L6D	1.848
Demanda Química de oxígeno	<--- L6D	.049

Fuente; Salida AMOS en base a los datos

- 1) Cuando L1I sube 1 desviación estándar, L4D baja 0,558 desviaciones estándar.
- 2) Cuando L1I sube 1 desviación estándar, L5D baja 1,505 desviaciones estándar.
- 3) Cuando L1I sube 1 desviación estándar, L6D baja 0,046 desviaciones estándar.
- 4) Cuando L1I aumenta en 1 desviación estándar, PT aumenta en 0.553 DS.
- 5) Cuando L1I aumenta en 1 desviación estándar, NT aumenta en 0.19 DS.
- 6) Cuando L1I aumenta en 1 desviación estándar, COL_TOTALES aumenta en 0.745 DS.
- 7) Cuando L4D aumenta 1 desviación estándar, oxígeno disuelto aumenta 0.72 DS.

- 8) Cuando L4D aumenta en 1 desviación estándar, PH aumenta en 0.887 DS.
- 9) Cuando L4D aumenta en 1 desviación estándar, SST aumenta en 0.482 DS.
- 10) Cuando L4D aumenta 1 desviación estándar, CLOROFILA aumenta 0,624 DS.
- 11) Cuando L4D sube 1 desviación estándar, TRANSP baja 0,698 DS.
- 12) Cuando L5D aumenta en 1 desviación estándar, Temperatura aumenta en 2.653 DS.
- 13) Cuando L5D aumenta en 1 desviación estándar, conectividad eléctrica aumenta en 0.077 DS.
- 14) Cuando L6D aumenta 1 desviación estándar, DBO5 aumenta 1,848 DS.
- 15) Cuando L6D aumenta 1 desviación estándar, DQO aumenta 0,049 DS.

El siguiente paso en el análisis de ajuste del modelo comienza comprobando que el Modelo 1 se confirma, es decir, llevando a cabo la evaluación global del mismo, de forma que se pueda determinar en qué medida este modelo inicial reproduce adecuadamente las relaciones que existen en la matriz de covarianzas de los datos empíricos. En la Tabla siguiente se muestran los índices para analizar plausibilidad de dicho modelo (NPAR).

Tabla 31

Ajuste del modelo

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	63	252.187	107	.000	2.357
Saturated model	170	.000	0		
Independence model	34	544.597	136	.000	4.004

Fuente: Salida AMOS en base a los datos

NPAR es el número de parámetros distintos (q) que se estiman. Dos parámetros (dos pesos de regresión, por ejemplo) que deben ser iguales entre sí cuentan como un solo parámetro, no dos.

CMIN es el valor mínimo, de la discrepancia. DF es el número de grados de libertad

P es la probabilidad de obtener una discrepancia tan grande como ocurrió con la muestra actual (bajo supuestos distributivos apropiados y suponiendo un modelo especificado correctamente). Es decir, P es un "valor p" para probar la hipótesis de que el modelo encaja perfectamente en la población.

Para el modelo predeterminado, es la discrepancia dividida por grados de libertad es $252.187/107 = 2.357$.

Default model: El modelo predeterminado tiene 63 parámetros, una discrepancia de 252.187 con 107 grados de libertad. P, Suponiendo que el modelo de modelo predeterminado es correcto, la probabilidad de obtener una discrepancia tan grande como 252.187 es de 0.000. $p(0.000) < \alpha(0.01)$, es decir, la prueba significativa.

Saturated model: El modelo saturado tiene 170 parámetros, una discrepancia de .000 y 0 grados de libertad.

Independence model: El modelo de independiente tiene 34 parámetros, una discrepancia de 544.597, 136 grados de libertad y suponiendo que el modelo de Independencia es correcto, la probabilidad de obtener una discrepancia tan grande como 544.597 es de 0.000.

La siguiente tabla, contiene las estimaciones de los parámetros del modelo (Modelo 1-ML), el error estándar aproximado (S.E.), la proporción crítica (C.R.) y las estimaciones estandarizadas de los parámetros.

La proporción crítica (C.R.) es el cociente entre la estimación del parámetro y la estimación del error estándar. Si se reúnen las suposiciones de distribución apropiadas, este estadístico sigue una distribución normal estándar bajo la hipótesis nula de que el parámetro tiene un valor de cero. Es decir, si a una estimación le corresponde una proporción crítica mayor que 2 en valor absoluto, más concretamente 1.96, el parámetro es significativamente diferente de cero al nivel 0.05.

A la vista de los resultados contenidos en dicha tabla, se comprueba que no todos los valores de proporción crítica son mayores que dos, en consecuencia, no todos los parámetros estimados son significativos.

Tabla 32

Valores significativos

			Estimate	S.E.	C.R.	P	Label
L4D	<---	L1I	-1.409	.845	-1.667	.095	W12
L5D	<---	L1I	-18.816	64.282	-.293	.770	W13
L6D	<---	L1I	-.360	.791	-.454	.649	W14
L4D	<---	L2I	-2.319	1.687	-1.375	.169	W15
L5D	<---	L2I	31.207	116.882	.267	.789	W16
L6D	<---	L2I	-1.307	1.474	-.886	.375	W17
L4D	<---	L3I	-6.143	3.765	-1.632	.103	W18
L5D	<---	L3I	84.161	320.465	.263	.793	W19
L6D	<---	L3I	-3.200	3.530	-.907	.365	W20
PT	<---	L1I	1.000				
NT	<---	L1I	.586	.355	1.652	.099	W1
Coliformes totales	<---	L1I	807.521	159.491	5.063	***	W2
NNH3	<---	L2I	1.000				
NO3	<---	L2I	.354	.156	2.267	***	W3
NO2	<---	L2I	.018	.014	1.292	.196	W4
PO4	<---	L3I	1.000				
Coliformes termotol	<---	L3I	338.423	177.908	1.902	.057	W5
Oxígeno disuelto	<---	L4D	1.000				
pH	<---	L4D	.446	.061	7.302	***	W6
Solidos Totales en sus	<---	L4D	5.025	1.258	3.994	***	W7
Clorofila-a	<---	L4D	11.173	2.159	5.175	***	W8
Transparencia	<---	L4D	-.309	.053	-5.798	***	W9
Temperatura	<---	L5D	1.000				
Conectividad eléctrica	<---	L5D	1.292	8.947	.144	.885	W10
DBO5	<---	L6D	1.000				
DQO	<---	L6D	.093	1.649	.056	.955	W11

Fuente: Salida AMOS en base a los datos

Parámetros estimados significativos: Transparencia, Clorofila-a, Solidos Totales en suspensión, pH, NO3 y Coliformes Totales.

Interpretando el primer valor:

-1.409, es el valor con el cual la variable latente exógena (L1I) influye sobre la variable latente endógena (L4D). dicho de otra manera, la influencia de la variable latente L1I es negativa sobre la variable latente endógena (L4D).

Cuando L1I sube en 1, L4D baja en 1.409, La estimación del peso de regresión, -1,409, tiene un error estándar de aproximadamente 0,845. Al dividir la estimación del peso de

regresión por la estimación de su error estándar se obtiene $z = -1.409 / .845 = -1.667$. En otras palabras, la estimación del peso de regresión es 1.667 errores estándar por debajo de cero. La probabilidad de obtener una relación crítica tan grande como 1,667 en valor absoluto es 0,095. En otras palabras, el peso de regresión para LII en la predicción de L4D no es significativamente diferente de cero en el nivel de 0.05 (dos colas).

Interpretando las variables significativas

Coliformes totales. Cuando LII aumenta en 1, Coliformes Totales aumenta en 807.521. La estimación del peso de regresión, 807.521, tiene un error estándar de aproximadamente 159.491. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z=807.521/159.491 = 5.063$. En otras palabras, la estimación del peso de regresión es 5.063 errores estándar por encima de cero.

La probabilidad de obtener una relación crítica tan grande como 5.063 en valor absoluto es menor que 0.001. En otras palabras, el peso de regresión para LII en la predicción de Coliformes Totales es significativamente diferente de cero en el nivel de 0.001.

pH. Cuando L4D aumenta en 1, pH aumenta en 0.446. La estimación del peso de regresión, .446, tiene un error estándar de aproximadamente .061. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z = 0.446 / 0.061 = 7.302$. En otras palabras, la estimación del peso de regresión es 7.302 errores estándar por encima de cero. La probabilidad de obtener una relación crítica tan grande como 7.302 en valor absoluto es menor que 0.001. En otras palabras, el peso de regresión para L4D en la predicción de pH es significativamente diferente de cero en el nivel de 0.001 (dos colas).

Sólidos Totales en Suspensión. Cuando L4D aumenta en 1, SST aumenta en 5.025. La estimación del peso de regresión, 5.025, tiene un error estándar de aproximadamente 1.258. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z = 5.025 / 1.258 = 3.994$. En otras palabras, la estimación del peso de regresión es 3.994 errores estándar por encima de cero. La probabilidad de obtener una relación crítica tan grande como 3.994 en valor absoluto es menor que 0.001. En otras palabras, el peso de regresión para L4D en la predicción de SST es significativamente diferente de cero en el nivel de 0.001 (dos colas).

Clorofila-a. Cuando L4D aumenta en 1, Clorofila-a aumenta en 11.173. La estimación del peso de regresión, 11.173, tiene un error estándar de aproximadamente 2.159. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z = 11.173 / 2.159 = 5.175$. En otras palabras, la estimación del peso de regresión es 5.175 errores estándar por encima de cero. La probabilidad de obtener una relación crítica tan grande como 5.175 en valor absoluto es menor que 0.001. En otras palabras, el peso de regresión para L4D en la predicción de Clorofila-a es significativamente diferente de cero en el nivel de 0.001 (dos colas).

Transparencia. Cuando L4D aumenta en 1, Transparencia disminuye en 0.309. La estimación del peso de regresión, -.309, tiene un error estándar de aproximadamente 0.053. Relación crítica para el peso de regresión. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z = -.309 / .053 = -5.798$. En otras palabras, la estimación del peso de regresión es 5.798 errores estándar por debajo de cero. La probabilidad de obtener una relación crítica tan grande como 5.798 en valor absoluto es menor que 0.001. En otras palabras, el peso de regresión para L4D en la predicción de Transparencia es significativamente diferente de cero en el nivel de 0.001.

NO3. Cuando L2I aumenta en 1, NO3 aumenta en 0.354. The regression weight estimate, .354, has a standard error of about .156. Al dividir la estimación del peso de regresión por la estimación de su error estándar se obtiene $z = .354 / .156 = 2.267$. En otras palabras, la estimación del peso de regresión es 2.267 errores estándar por encima de cero. La probabilidad de obtener una relación crítica tan grande como 2.267 en valor absoluto es .023. En otras palabras, el peso de regresión para L2I en la predicción de NO3 es significativamente diferente de cero en el nivel de 0.05 (dos colas).

Tabla 33

Las ecuaciones estructurales quedan definidas como:

Coliformes Totales	<---	L1I	807.521	159.491	5.063	***	W2
NO3	<---	L2I	.354	.156	2.267	***	W3
pH	<---	L4D	.446	.061	7.302	***	W6
Solidos Totales en Suspensión	<---	L4D	5.025	1.258	3.994	***	W7
Clorofila-a	<---	L4D	11.173	2.159	5.175	***	W8
Transparencia	<---	L4D	-.309	.053	5.798	***	W9

Fuente: Salida AMOS en base a los datos



Modelos de ecuaciones estructurales:

$$COL_{TOTALES} = 807.521 \times L1I$$

$$NO3 = 0.354 \times L2I$$

$$PH = 0.446 \times L4D$$

$$SST = 5.025 \times L4D$$

$$Clorofila = 11.173 \times L4D$$

$$Transp = -0.309 \times L4D$$

CONCLUSIONES

- 1) Las variables temperatura, coliformes totales, oxígeno disuelto, pH, conectividad eléctrica y transparencia tiene una tendencia estacional disminuyendo en los meses de abril, mayo, junio y julio y las variables coliformes termotolerantes, nitrito, total de sólidos en suspensión, clorofila-a, demanda bioquímica de oxígeno, demanda química de oxígeno, nitrato, fosfato, nitrógeno, fosforo y nitrógeno amoniacal, no tienen tendencia estacional anual.
- 2) Los valores de temperatura en los meses de mayo, junio, Julio y agosto están por debajo de 15°C. Las variables: Coliformes totales, la concentración de oxígeno disuelto, la concentración de coliformes termotolerantes, el pH, los sólidos totales disueltos, la demanda bioquímica de oxígeno, la demanda química de oxígeno y el nitrógeno superan los límites permitidos en todos los meses del año. Las variables fosforo, nitrógeno amoniacal, nitrato, fosfato no exceden los límites permitidos en todos los meses del año y los valores de clorofila-a se encuentran entre 2,05 y 51.36 al igual que la transparencia entre 0.45 y 1.97 Es decir, Las aguas no son aptas para el consumo humano.
- 3) La temperatura está relacionada con coliformes totales, la conectividad eléctrica, oxígeno disuelto, transparencia, sólidos totales en suspensión y el nitrito. La variable oxígeno disuelto está relacionado con nitrógeno amoniacal, clorofila-a, fosfato, demanda química de oxígeno y pH. El pH está relacionado con fosforo, nitrógeno, transparencia, coliformes totales, temperatura, fosfato. La conectividad eléctrica está relacionada con Nitrato, Fosforo, oxígeno disuelto, coliformes termotolerantes y demanda bioquímica de oxígeno. Los sólidos totales en suspensión están relacionados con pH, oxígeno disuelto, demanda bioquímica de oxígeno, nitrógeno amoniacal y clorofila-a. La clorofila-a esta relacionada con

pH, oxígeno disuelto, temperatura, conectividad eléctrica sólidos totales en suspensión y fósforo. La demanda bioquímica de oxígeno está relacionada con nitrógeno, nitrógeno amoniacal, oxígeno disuelto, nitrato, demanda química de oxígeno y temperatura. La demanda química de oxígeno está relacionada con: coliformes termotolerantes, temperatura, conectividad eléctrica, nitrógeno amoniacal y demanda bioquímica de oxígeno. La transparencia está relacionada con clorofila-a pH, sólidos totales en suspensión, nitrato, nitrito y demanda química de oxígeno.

- 4) las variables dependientes se concentran en Componentes 1: Oxígeno disuelto, pH, Sólidos totales en suspensión, Clorofila-a y transparencia. Componentes 2: Temperatura y conectividad eléctrica. Componentes 3: Demanda Bioquímica de Oxígeno y Demanda Química de Oxígeno. Las variables dependientes se concentran en: Componentes 1: Coliformes Totales, Nitrógeno y Fósforo. Componentes 2: Nitrito, Nitrato y Nitrógeno amoniacal y Componentes 3: Coliformes Termotolerantes y fósforo.
- 5) Los Modelos de ecuaciones estructurales son:

$$\begin{aligned} COL_{TOTALES} &= 807.521 \times L1I; & NO3 &= 0.354 \times L2I; & pH &= 0.446 \times L4D; \\ SST &= 5.025 \times L4D; & Clorofila - a &= 11.173 \times L4D; & Transp &= \\ & & & & -0.309 \times L4D \end{aligned}$$



RECOMENDACIONES

1. Los árboles de predicción son una buena opción cuando los datos no tienden a una distribución normal o no se cumplen los supuestos de una regresión lineal múltiple.
2. Existe diferencia en los valores de los parámetros entre meses del año y no es recomendable promediarlos por trimestre, semestre o anual porque estos varían de acuerdo a diferentes factores en los parámetros físico-químicos.
3. Los datos que se utilizaron en la presente investigación pueden ser tratados por otros modelos y ser comparados en cuanto a la predicción.
4. Obtener modelos de Box Jenkins para comparar con las predicciones realizadas en la presente investigación.
5. El IMARPE es una institución que permanentemente realiza investigación y toma de datos, es necesario analizar estos datos con técnicas estadísticas modernas.

BIBLIOGRAFÍA

- Aguilar, I. L. (2019). *Análisis comparativo en la implementación de la red neuronal backpropagation usando el método de componentes principales y el método clásico*. Piura – Perú. Recuperada de:
<http://repositorio.unp.edu.pe/handle/UNP/1583>
- Arohuanca, C.C.A. (2015). *Evaluación de la carga de nitrógeno y fósforo en las principales fuentes puntuales que vierten al lago Titicaca como fuente de Eutrofización Puno 2015*. (Tesis de grado). Universidad Nacional de San Agustín de Arequipa.
- Autoridad binacional del lago Titicaca (ALT) (14 de diciembre 2019). Características físicas del lago Titicaca. Recuperada de:
http://www.alt-perubolivia.org/web_lago/WEB_LT/Finales/2_carac_fisicas/carac.htm
- Beltrán, F.D. F., Palomino, C.R.P., Moreno, T.E.G., Peralta, C.G., Montesinos, D.B. (2011). *Calidad de agua de la bahía interior de Puno, lago Titicaca durante el verano del 2011*. Revista Peruana de Biología, vol.22 no.3 Lima. Scielo.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: J. Wiley & Sons.
- Buolange, y Aquize, J. (1981) *Morphologie, hydrographie et climatologie du lac Titicaca et de son bassin versant*. Revista Hydrobiol.
- Calvo, P.C., Martinez, F.A. y Juanatey, B.O. (2013). *Análisis de dos modelos de ecuaciones estructurales alternativos para medir la intención de compra*. Revista investigación de operaciones. vol 34, Nun 3. Cuba.

- Callata, T.F.E. (2015). *Monitoreo y evaluación del cuerpo de agua de la bahía interior de Puno - Lago Titicaca*. (Tesis de grado). Universidad Nacional del Altiplano.
- Carpio, V.B.O. (2016). *Contaminación por metales pesados en macrofitas de los principales ríos tributarios del lago Titicaca*. (Tesis doctorado). Universidad Nacional del Altiplano. Recuperado de:
<http://www.bvsde.paho.org/bvsaidis/chile13/trab-12.pdf>
- Cisterna, P. (s.f.) *Determinación de la relación DQO/DBO5 en aguas residuales de comunas con población menor a 25.000 habitantes en la VIII región*. Analista Universidad Tec. Fed. Sta María.
- Crombet, G.S., Abalos, R.A., Zamora, M.L. (2018). *Determinación de los parámetros ambientales de mayor incidencia en las aguas residuales de la comunidad universitaria Antonio Maceo*. Rev. Scielo Cuba.
- Cuadras, C.M. (2019). *Nuevos métodos de análisis multivariantes*. CMC Editions. Barcelona España.
- Cuevas, A.A.B. (2018). *Macroinvertebrados bentónicos como referentes de la calidad de aguas del lago Titicaca en el centro de investigación y tratamiento tecnológico Chucuito- Puno*. (tesis de grado). Universidad Nacional del Altiplano.
- Cupani, M. (2012). *Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación*. Revista tesis. Nro 1. p.p. 186-199.
- Diaz, C. E., Alvarado, G.A.R, Camacho, C.K.E. (2012). *El tratamiento de agua residual doméstica para el desarrollo local sostenible: el caso de la técnica del sistema unitario de tratamiento de aguas, nutrientes y energía (SUTRANE) en San Miguel Almaya*, Universidad Autónoma de México.
- Dirección General de Salud Ambiental e Inocuidad Alimentaria (DIGESA) (12 de Julio 2019). *Informes técnicos*. Recuperado de:
http://www.digesa.minsa.gob.pe/DEPA/informes_tecnicos/GRUPO%20DE%20USO%201.pdf
- Escobar, M.F. (2019). *Determinación de parámetros físico-químicos y niveles de metales pesados en agua y sedimentos en la zona de crianza de truchas (Oncorhynchus*

- mykiss*), bahía de Puno del lago Titicaca. (Tesis doctorado). Universidad Nacional del Altiplano.
- Galvín, R. (2003). *Fisicoquímica y microbiología de los medios acuáticos. Tratamiento y control de calidad de aguas*. Madrid: Editorial Díaz de Santos.
- García, I., Rodríguez J. G., López, F. y Tenorio, Y. M. (2010). *Transporte de Contaminantes en Aguas Subterráneas mediante Redes Neuronales Artificiales*. revista Información tecnológica Vol. 21(5).
- Gasco, M.J.M., Quesada, P. V. (2007). *Estudio estadístico de la calidad de las aguas en la cuenca hidrográfica del río EBRO*. (Tesis doctoral). Universidad Politécnica de Madrid.
- Gregor, J., Marsálek, B. *Freshwater Phytoplankton Quantification by Chlorophyll a: A Comparative Study of in vitro, in vivo and in situ Methods*. Revista Water Research, Vol. 38, Issue 3, February 2004, Pages 517-522.
- Hallasi, P.G. E. (2018). *Determinación de los parámetros microbiológicos y físico-químicos de las aguas de consumo humano en las islas flotantes uros del lago Titicaca*. (Tesis de grado). Universidad Nacional del Altiplano, Puno-Perú.
- Hernández, A.J.L. (2016). *Modelos de ecuaciones estructurales aplicados al análisis de fatiga*. Revista Ciencias de la Salud. vol. 14. Universidad de Rosario.
- Hu, L.T., y Bentler, P.M. (1995). *Evaluating model fit*. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and application* (pp. 77-99). Thousand Oaks. [https://www.scirp.org/\(S\(i43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=1015696](https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1015696)
- Instituto del mar del Perú (IMARPE) (2016). *Informe*. Vol 43 Nro. 3 Julio-Setiembre 2016. <http://biblioimarpe.imarpe.gob.pe/handle/123456789/3111>
- Jimenez, M.L.A. (2106). *Tratamiento de aguas eutrofizadas con el uso de dos macrofitos de la bahía interior de Puno*. (Tesis doctorado). Universidad Nacional del Altiplano.
- Jiménez, B. (2001). *La contaminación ambiental en México: causas, efectos y tecnología apropiada*. México: UNAM y FEMISCA

- Kaplan, D. (2000) *Structural Equation Modeling: A Multidisciplinary Journal*. Book Review: Structural Equation Modeling: Foundations and Extensions (Advanced Quantitative Techniques in the Social Sciences Series, Vol. 10. Publication Date: 01 April 2003. Recuperado de:
https://www.researchgate.net/publication/233337570_Book_review_of_D_Kaplan_2000_Structural_equation_modeling_Foundations_and_extensions
- Larios, M.J.F., Gonzales, T.C. y Morales, O.Y. (2015). *Las aguas residuales y sus consecuencias en el Perú*. (versión electrónica). Revista de la Facultad de Ingeniería de la USIL. Vol. 2, N° 2. Segundo semestre 2015. pp. 09-25.
- Larrea, M.J.A, Rojas, B.M.M, Romeu, Á.B, Rojas, H.N, Heydrich, P.M.M. (2013) *Bacterias indicadoras de contaminación fecal en la evaluación de calidad de las aguas: revisión de la literatura*. La Habana, cuba.
- Levi, J. y Varela, J. (2004). *Análisis Multivariable para las ciencias sociales*. Edit Prentice Hall. Madrid España.
- Madrigal, M.F, Madrigal, M.S., Avila, F. (2019). *Modelos de ecuaciones estructurales en investigaciones sociales enfocadas al comportamiento de compra de millennials*, Espacios Vol 40.
- Ministerio de vivienda ordenamiento territorial y Medio Ambiente (MVOTMA) (25 de noviembre 2019) *Indicadores ambientales, Concentración de Fosforo Total*. Recuperado de:
https://www.dinama.gub.uy/indicadores_ambientales/ficha/oan-concentracion-de-fosforo-total/
- MINAM (2017). *Normas Legales, Estándares Nacionales de Calidad Ambiental (ECA) para Agua N° 004 - 2017 - MINAM, según la categoría A-1 "Aguas superficiales destinadas a la producción de agua potable"*. La Ley de Recursos Hídricos, Ley N° 29338. Lima.
- Dirección General de Salud (DIGESA). Ministerio de Salud (2011). *Reglamento de Calidad del Agua para Consumo Humano*. Lima - Perú. 46 p.
- Moreno, T.E., Argota, P.G., Alfaro, T.R., Aparicio, S.M., Atencio, L.S., Goyzueta, C.G. (2018). *Cuantificación de metales en sedimentos superficiales de la bahía interior, lago Titicaca-Perú*. Revista de Investigaciones Alto Andinas. Vol 20 Nro 1.

- Programa regulatorio agrícola recursos para agricultores (2013) *Nitrato en el agua potable*.
https://www.waterboards.ca.gov/centralcoast/water_issues/programs/ag_waivers/docs/spanish_forms/nitratos_en_el_agua.pdf
- Pereira, S.P., De Araujo, K.R y Moya, R.L. (2019), *Evaluation of the City Logistics Performance Through Structural Equations Model*. IEEE.
- Pinto, A. Von, S.E, Moreira, R. (2004) *Chlorophyll-a Determination Via Continuous Measurement of Plankton Fluorescence: Methodology Development*. Revista Water Res. 2001;35(16). Pag 517-522.
- Pulido, C.I., Roldan, C.J., Lopez, L.J., Gutierrez, E. J. (2002). *Técnicas de predicción a corto plazo de la demanda de agua. Aplicación al uso agrícola*. Revista ingeniería del agua. Vol 9, Nro 3. Universidad de Valencia.
- Raffo, E., Ruiz, E. (2014). *Caracterización de las aguas residuales y la demanda bioquímica de oxígeno*. Revista Diseño y tecnología. UNMSM
- Ramón, y Cajal, S. (1911/2007), *Histología del sistema nervioso del hombre y de los vertebrados*, Madrid, Ministerio de Sanidad y Consumo, CSIC.
- Paulino, R.C. (2016). *Estimación de concentración de clorofila-a en el lago Titicaca usando imágenes satelitales MERIS*. Recuperado de:
<http://biblioimarpe.imarpe.gob.pe/handle/123456789/3111>
- Romero, A.M. Colin, C.A, Sánchez, y Ortiz, H.M.L. (2009). *Tratamiento de aguas residuales por un sistema piloto de humedales artificiales*. Revista internacional de contaminación ambiental, vol.25, n.3, pp.157-167. Scielo.
- Romero, S.C.K. (2018). *Determinación de la relación de la comunidad Fito planctónica con los factores físicos y químicos del Sector puno de la reserva nacional del Titicaca, Región puno, Abril – julio 2014*. (Tesis de Grado). Universidad nacional de San Agustín de Arequipa.
- Rosenblatt, F. (1958). *The perceptron: a probabilistic model for information storage and organization in the brain*. Psychological Review. Vo. 65 Nro 6. 1958.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.335.3398&rep=rep1&type=pdf>

- Rumelhart, D., McClelland, J.L. (1986). *Procesamiento distribuido paralelo: exploraciones en la microestructura de la cognición*. MIT Press, Cambridge.
- Samboni, R.N.E., Carvajal Escobar Y., Escobar J.C. (2007). *Revisión de parámetros fisicoquímicos como indicadores de calidad y contaminación del agua*. Revista ingeniería e investigación vol.27 Nro 3. Universidad Nacional de Colombia.
- Shogo, O., Haruyo, K., Atsushi, Y., Kyoichi, K., Atsuko, T. (2018). *Layered modeling of affective and sensory experiences using structural equation modeling: Touch experiences of plastic surfaces as an example*. Member, IEEE.
- Rich, E., y Knight, K. (1994). *Inteligencia artificial* (2 ed.). (P. A. González Calero, & F. Trescastro Bodega, Trads.) McGraw-Hill.
- Ruiz, Y.C. (s.f.) *Aplicación del análisis de componentes principales como técnica para obtener índices sintéticos de calidad ambiental*. revistas ucv.
- Universidad de Salamanca (25 de junio 2018). *Redes Neuronales*. recuperado de:
<http://avellano.fis.usal.es/~lalonso/RNA/index.htm>
- Vias, M. J., Perles Rosello M.J., Andreo Navarro Bartolome. (2003). *Aplicación de un análisis clúster para la evaluación de la vulnerabilidad a la contaminación de los acuíferos*. Revista internacional de ciencia y tecnología de la información geográfica Nro. 3 p. 199-215.
- Wirmann, D. (1991) *Morfología y barimetría. el lago Titicaca*. ORSTOM-HISBOL, La PAZ p.31-37.
- Wright, S. (1921). *Correlation and causation*. Journal of Agricultural Research 20, 557–585.
- Wright, S. (1934). *The method of path coefficients*. Annals of Mathematical Statistics 5,161–215.

ANEXOS

Anexo 1. Valores pronosticados, usando redes neuronales

Mes	temperatura	oxígeno disuelto mb/ml	PH	conectividad eléctrica Us/cm	SST (mg/L)	CLOROFILA-a (mg/m3)	DBO5 (mg/L)	DQO (mg/L)	COL. TOTALES NMP/100mL	COL. TERM. NMP/100mL	NO2 (mg/L)	NO3 (mg/L)	PO4 (mg/L)	N - T (mg/L)	P - T (mg/L)	N - NH3 (mg/L)	TRANSP. (m)
15/06/2019	11.147	11.873	8.887	1792.977	24.3543391	31.170266	116.13194	18.8775428	12.4708992	382.115048	0.18421134	0.21507118	0.98016794	5.27461681	0.55250555	0.1037459	1.47054343
16/07/2019	12.087	11.031	8.858	1722.623	14.3925478	31.8044346	92.7114725	22.4248527	8.03290239	148.904054	0.07347403	0.29334958	1.461199201	3.67820489	0.62684146	0.00167161	2.32694613
16/08/2019	12.638	10.036	8.640	1670.305	16.062587	38.61757	72.0183551	31.1444186	100.715725	266.399185	0.03464485	0.38293317	1.26937294	3.18918505	0.67585188	0.3848536	2.19831834
17/09/2019	16.003	7.4857	10.40	1720.407	15.6022021	54.4967892	56.9091941	37.0940133	76.5982146	381.781494	0.09928043	0.40891296	2.56294842	2.74409059	0.74956629	0.58542743	1.7907626
16/10/2019	17.026	7.254	10.378	1672.071	18.8581953	55.7905625	55.8025304	20.9922542	26.9451929	156.250881	0.05551215	0.20362853	1.63869523	3.74992839	0.80708557	0.4550527	3.1698371
19/11/2019	18.672	7.2618	9.546	1735.879	23.9027992	72.6949916	45.0391463	22.1923204	15.8212294	275.176454	0.09398729	0.33316374	1.19350081	3.69148217	0.95257696	0.38056687	2.33339505
17/12/2019	17.702	7.2140	10.088	1727.675	28.6438491	82.0049854	71.4520175	35.0846709	14.0802869	379.799365	0.10994147	0.27802117	0.5782698	2.8960228	1.07818663	0.63495926	1.3137516
17/01/2020	17.254	11.247	9.931	1757.669	31.9268173	80.2964147	81.2902402	39.7075585	10.8980537	169.976757	0.08213783	0.15631539	0.33971023	2.22014231	1.40890286	0.83571367	2.07811163
18/02/2020	17.459	11.405	9.186	1793.277	31.1049943	59.2646792	80.087952	32.4703758	37.8536659	215.085985	0.09728408	0.1550108	2.18825705	1.75806919	1.88996029	0.91556346	1.90760703
19/03/2020	16.417	8.808	9.140	1493.393	28.7929546	31.0462631	16.0710178	23.5169438	53.1455189	341.453648	0.11205811	0.28374018	1.64712226	2.36855871	1.37576821	0.91288602	2.2001161
22/04/2020	15.880	9.877	8.961	1835.893	25.0459958	32.0179974	45.2110456	33.9450638	35.1601663	337.622603	0.13158174	0.2510485	0.56704559	3.00095668	0.54944818	0.32829261	1.61616854
15/05/2020	16.198	9.4805	9.225	1799.73	19.6980448	49.7223399	115.744636	25.0847457	20.5688872	297.632702	0.14794104	0.16236494	0.1354444	1.80244572	0.6518054	0.64635797	1.74367582
15/06/2020	12.544	9.773	9.386	1804.0260	18.7863547	62.4159204	81.5288103	18.9508793	15.6256708	359.971055	0.18440626	0.28669096	0.52927676	2.25056362	0.74532082	0.57187248	1.57165437
16/07/2020	11.981	7.4801	8.033	1773.226	20.917595	89.6166097	84.6812204	22.4791414	12.102657	228.347277	0.22215441	0.2320441	1.51487428	1.83338998	0.85495381	0.60725339	2.14892415
16/08/2020	13.659	7.456	8.177	1789.386	23.8842704	91.2534064	121.082113	31.068796	20.7939082	347.764325	0.0706218	0.23365704	1.5057144	2.29287492	0.99923279	0.35726422	2.20216446
17/09/2020	14.571	6.195	9.672	1783.032	27.1349039	58.3885987	92.1920476	37.1397658	32.9476263	366.689835	0.08553861	0.2287376	1.96969049	3.15802323	1.24872393	0.12086553	1.51960355
16/10/2020	18.478	8.804	8.963	1563.31201	30.3527808	18.0307004	97.1329324	21.0537486	32.9168333	208.904472	0.12367622	0.2836998	1.42200547	2.47713019	1.68632252	0.02752347	0.80174208
19/11/2020	17.886	8.806	9.320	1750.89681	32.3708731	26.4865372	95.3775423	22.1445795	24.5502769	332.571553	0.25280063	0.37804135	2.58365729	3.22231254	1.35255428	0.18287766	1.1485476
17/12/2020	17.008	11.842	8.634	1556.304	31.0963918	33.4670926	5.19523861	35.0147885	18.3918078	355.080812	0.16392636	0.44502752	4.28035217	2.92344604	1.58349003	0.54104601	1.54046736
17/01/2021	18.741	12.053	8.604	1672.668	27.9664743	50.6077138	14.0930358	39.6938779	14.0188531	243.506246	0.111576255	0.27878083	3.96476507	2.79393786	1.42741449	0.50427688	1.68464708
18/02/2021	14.981	12.167	9.255	1546.949	25.833717	69.0004471	130.603791	32.4523498	15.494201	357.357777	0.26774364	0.2999503	2.95220744	3.20533967	0.63520318	0.88464049	1.60450581
19/03/2021	18.255	9.9179	8.956	1671.494	24.1916196	85.9411432	81.0277391	23.5005553	21.7756284	329.96657	0.12010234	0.35287013	1.5106724	4.17766026	0.7253169	0.81031317	1.80286245
22/04/2021	13.415	7.500	8.742	1585.434	23.5522576	82.3487922	152.612496	33.8550986	25.686771	329.878174	0.04772194	0.23892786	3.04759077	3.54483863	0.83280063	0.39479769	2.12870028

Anexo 2. Covarianzas y correlaciones anti-imagen de las variables observadas dependientes

		Tempe- ratura	Oxígeno disuelto	PH	Conecti- vidad	SST	Clorofila- a	DBO5	DQO	Transp
Covaria nza anti- imagen	Temperatura	,642	-,004	-,178	-,054	,050	,164	,022	-,201	,091
	Oxígeno disuelto	-,004	,428	-,172	-,169	,096	-,126	,040	,065	,121
	PH	-,178	-,172	,378	,082	-,079	-,068	-,060	-,066	,086
	conectividad	-,054	-,169	,082	,850	-,001	,145	,029	-,039	,028
	SST	,050	,096	-,079	-,001	,631	-,180	-,126	,052	,112
	Clorofila-a	,164	-,126	-,068	,145	-,180	,489	-,021	-,025	,072
	DBO5	,022	,040	-,060	,029	-,126	-,021	,883	-,089	,033
	DQO	-,201	,065	-,066	-,039	,052	-,025	-,089	,831	-,121
	Transp	,091	,121	,086	,028	,112	,072	,033	-,121	,481
Correlac ión anti- imagen	Temperatura	,549^a	-,007	-,360	-,073	,079	,293	,029	-,276	,164
	Oxígeno disuelto	-,007	,717^a	-,428	-,281	,184	-,275	,066	,109	,266
	PH	-,360	-,428	,755^a	,145	-,161	-,158	-,104	-,117	,203
	conectividad	-,073	-,281	,145	,426^a	-,001	,225	,034	-,046	,044
	SST	,079	,184	-,161	-,001	,755^a	-,324	-,169	,072	,203
	Clorofila-a	,293	-,275	-,158	,225	-,324	,747^a	-,033	-,039	,149
	DBO5	,029	,066	-,104	,034	-,169	-,033	,797^a	-,104	,051
	DQO	-,276	,109	-,117	-,046	,072	-,039	-,104	,505^a	-,191
	Transp	,164	,266	,203	,044	,203	,149	,051	-,191	,838^a

a. Medidas de adecuación de muestreo (MSA)

Anexo 3. Covarianzas y correlaciones anti-imagen de las variables observadas independientes

		Matrices anti-imagen							
		Col. Totales	Col. Term.	NO2	NO3	PO4	N - T	P - T	N - NH3
Covarianza anti-imagen	Col. Totales	,614	-,146	-,162	-,109	,197	-,138	-,025	-,256
	Col. Term.	-,146	,783	,078	-,232	,079	,120	-,073	,009
	NO2	-,162	,078	,738	-,199	-,091	,024	-,098	,186
	NO3	-,109	-,232	-,199	,632	-,234	-,039	-,016	,125
	PO4	,197	,079	-,091	-,234	,772	-,088	,104	-,229
	N - T	-,138	,120	,024	-,039	-,088	,831	-,235	,084
	P - T	-,025	-,073	-,098	-,016	,104	-,235	,695	-,244
	N - NH3	-,256	,009	,186	,125	-,229	,084	-,244	,604
Correlación anti-imagen	Col. Totales	,546^a	-,211	-,241	-,176	,286	-,193	-,038	-,421
	Col. Term.	-,211	,552^a	,102	-,329	,101	,149	-,100	,012
	NO2	-,241	,102	,529^a	-,291	-,121	,031	-,137	,278
	NO3	-,176	-,329	-,291	,542^a	-,334	-,054	-,024	,203
	PO4	,286	,101	-,121	-,334	,270^a	-,109	,141	-,335
	N - T	-,193	,149	,031	-,054	-,109	,521^a	-,310	,119
	P - T	-,038	-,100	-,137	-,024	,141	-,310	,596^a	-,376
	N - NH3	-,421	,012	,278	,203	-,335	,119	-,376	,501^a

a. Medidas de adecuación de muestreo (MSA)