

UCUENCA

**Facultad de Ciencias Químicas
Carrera de Ingeniería Ambiental**

**“ESTUDIO COMPARATIVO DE LA PREDICTIBILIDAD DEL CAUDAL DE LA
CUENCA DEL RÍO PAUTE BASADO EN LOS ÍNDICES DE TELECONEXIÓN Y LOS
MODOS DE VARIABILIDAD”**

Trabajo de titulación previo a la obtención del
título de Ingeniero Ambiental

Autora:

María Daniela González Salto

C.I: 0150268753

Correo electrónico: mardangosa@hotmail.com

Director:

Alex Manuel Avilés Añazco

C.I: 0102247186

Tutor:

Darío Xavier Zhiña Villa

C.I: 0105044481

Cuenca, Ecuador

15-julio-2022

Resumen

La cuenca del río Paute formada por varias subcuencas que suministran agua para diferentes sectores urbanos, agrícolas y energético presenta diferentes comportamientos hidrológicos que dependen del clima y la intervención antropogénica. Debido al cambio climático y a la variabilidad climática que se produce de forma natural a través del tiempo, afectan los distintos procesos hidrológicos, generando eventos extremos (inundaciones y sequías) que impactan los ecosistemas, la productividad de las centrales hidroeléctricas y la calidad de vida de las personas que dependen de los recursos hídricos. Esta problemática demuestra la importancia de conocer con anticipación el comportamiento hidrológico que permita tomar decisiones adecuadas con respecto a adaptación y la resiliencia en ambientes extremos cambiantes. El objetivo de este estudio es explorar la predictibilidad de los caudales en la cuenca del río Paute a partir de los índices de teleconexión a nivel global. El proceso se dividió en varias etapas que incluyen, la recopilación y relleno de datos, exploración gráfica, aplicación de análisis de componentes principales, usando desfases mensuales y análisis de multicolinealidad en modelos de regresión múltiple. Todo esto a través del software R Studio, utilizando métricas estadísticas para comparar el rendimiento de cada modelo propuesto. Se demostró que el escenario con análisis de componentes principales obtuvo los mejores ajustes, específicamente al incluir 41 índices y 20 componentes. Para el escenario usando retrasos mensuales el mejor desfase ocurre con un solo mes, para la mayoría de estaciones. Finalmente, con el escenario de análisis de multicolinealidad se obtuvo mejores resultados usando 41 índices, aunque esencialmente el desempeño corresponde a la cantidad y a los índices de cada modelo. Los índices de teleconexión no son suficientes al utilizarlos como única variable de entrada para modelamiento y predicción de caudales, dando en su mayoría resultados poco satisfactorios. Sin embargo, existe una clara tendencia que vincula el comportamiento de caudales e índices, pudiéndose mejorar los modelos a partir de más variables climáticas o con otros métodos predictivos.

Palabras clave: Predicción de caudales. Índices de teleconexión. Análisis de componentes principales. Modelos de regresión múltiple. Análisis de multicolinealidad.

Abstract

The Paute river basin, formed by several sub - basins that supply water for different urban, agricultural and energy sectors, presents different hydrological behaviors that depend on the climate and anthropogenic intervention. Due to climate change and climate variability that occurs naturally over time, they affect the different hydrological processes, generating extreme events (floods and droughts) that impact ecosystems, the productivity of hydroelectric power plants and the quality of life of people who depend on water resources. This problem demonstrates the importance of knowing in advance the hydrological behavior that allows making appropriate decisions regarding adaptation and resilience in changing extreme environments. The objective of this study is to explore the predictability of flows in the Paute river basin from global teleconnection rates. The process was divided into several stages, including data collection and filling, graphical exploration, application of principal components analysis, using monthly delays, and multicollinearity analysis in multiple regression models. All this through the R Studio software, using statistical metrics to compare the performance of each proposed model. It was shown that the principal component analysis scenario obtained the best predictational fits, specifically by including 41 indices and 20 components. For the scenario using monthly delays, the best delay occurs with a single month, for most stations. Finally, with the multicollinearity analysis scenario, better results were obtained using 41 indices, although essentially the performance corresponds to the quantity and the indices of each model. Teleconnection indices are not sufficient when used as the only input variable for flow modeling and prediction, giving mostly unsatisfactory results. However, there is a clear trend that links the behavior of flows and indices, and it is possible to improve the models based on more climatic variables or with other predictive methods.

Keywords: Streamflow prediction. Teleconnection indices. Principal component analysis. Multiple regression models. Multicollinearity analysis.

Índice

INTRODUCCIÓN	16
OBJETIVOS	19
Objetivo general:	19
Objetivos específicos:	19
1.1 Índices de Teleconexión.....	20
1.2 Variabilidad Climática	32
1.2.1 Modos de Variabilidad	32
1.3 Antecedentes	33
1.3.1 Fenómenos hidrometeorológicos en el Ecuador	33
1.3.2 Funcionamiento de los índices climáticos en Ecuador	34
1.4 Influencia global de los índices climáticos atmosféricos y oceánicos	35
1.5 Influencia de los índices de teleconexión en los desastres naturales	37
1.5.1 Australia.....	37
1.5.2 Brasil.....	38
1.5.2 China.....	38
1.5.3 Pacífico oriental.....	38
1.5.4 Europa.....	38
1.5.5 México	39
1.5.6 Atlántico norte	39
1.5.7. Noroeste del Pacífico.....	39
1.5.8 Sudáfrica.....	40
1.5.9 Sur de Asia.....	40
1.5.10 Sudeste Asiático	40
1.5.11 Estados Unidos	40

1.6 Modelación para predictibilidad de caudales	41
CAPITULO 2: MATERIALES Y MÉTODOS	44
2.1 Descripción del área de estudio.....	44
2.3 Recolección de datos.....	46
2.3.1 Recolección de datos hidrometereológicos	46
2.3.2 Recolección de índices climáticos atmosféricos y oceánicos.....	47
2.4 Relleno de datos	48
2.4.1 Relleno de datos de estaciones de caudal	48
2.4.1 Relleno de datos de índices climáticos atmosféricos y oceánicos	48
2.5 Análisis exploratorio gráfico.....	49
2.5.1 Análisis gráfico de series de tiempo.....	49
2.5.2 Análisis diagrama de dispersión.....	50
2.6 Prueba de Homogeneidad de datos	51
2.6.1 Curva de masa	51
2.7 Calibración y Validación de datos	52
2.8 Modelación para predicción de caudales con series temporales de los índices climáticos atmosféricos y oceánicos.....	52
2.8.1 Modelación usando el Análisis de Componentes Principales (ACP).....	52
2.8.2 Modelación usando Índices climáticos atmosféricos y oceánicos y retrasos mensuales	55
2.8.3 Modelación usando análisis de multicolinealidad	55
2.8.3.1 Modelos de Regresión Múltiple (MLR)	56
2.9 Obtención de resultados de predicción de series de tiempo de caudales	58
2.10 Análisis estadístico de resultados mediante métricas.....	58
2.10.1 Valor P o P-value.....	59

2.10.2 Eficiencia de Nash-Sutcliffe.....	60
2.10.3 Eficiencia de Kling-Gupta.....	61
2.10.4 Coeficiente de determinación.....	61
2.10.5 Raíz del error cuadrático medio.....	62
CAPITULO 3: RESULTADOS Y DISCUSION	63
3.1 Análisis exploratorio gráfico.....	63
3.1.1 Gráficas de series de tiempo.....	63
3.2 Prueba de homogeneidad de datos	68
3.2.1 Curva de masa	68
3.3 Calibración y Validación de datos	68
3.4 Modelación para predicción de caudales con series temporales de los índices climáticos atmosféricos y oceánicos.....	69
3.4.1 Modelación usando el Análisis de Componentes Principales (ACP).....	69
3.4.2 Modelación usando Índices climáticos atmosféricos y oceánicos y retrasos mensuales	79
3.4.3 Modelación usando análisis de multicolinealidad	89
.....	94
3.5 Análisis estadístico de resultados mediante métricas.....	98
Elaborado por: María Daniela González.....	99
3.5.1 Valor P o P-value.....	102
3.5.2 Eficiencia de Nash-Sutcliffe.....	102
3.5.3 Eficiencia de Kling-Gupta.....	103
.....	105
3.5.4 Coeficiente de determinación.....	105
3.5.5 Error cuadrático medio.....	106

CAPITULO 4: CONCLUSIONES Y RECOMENDACIONES.....	109
4.1 Conclusiones	109
4.2 Recomendaciones.....	110
REFERENCIAS BIBLIOGRÁFICAS.....	113
CAPITULO 5: ANEXOS	129
5.1 Anexo 1: Resultados de series de tiempo en la etapa de calibración para los escenarios ACP y usando retrasos mensuales.	129
5.2 Anexo 2: Resultados de diagramas de dispersión en la etapa de calibración para los escenarios ACP y usando retrasos mensuales.	135
5.3 Anexo 3: Resultados de series de tiempo en la etapa de validación para los escenarios ACP, usando retrasos mensuales y VIF.	141
5.4 Anexo 4: Resultados de diagramas de dispersión en la etapa de validación para los escenarios ACP, usando retrasos mensuales y VIF.	153
5.5 Anexo 5: Resultados de las métricas en la etapa de validación.	165

Índice de figuras

Figura 1. Ubicación del sitio de estudio, así como la ubicación de las diferentes subcuencas que conforman la cuenca del río Paute.	45
Figura 2. Ubicación de estaciones de caudal y precipitación en la cuenca del río Paute.	46
Figura 3 . Esquema general de una gráfica de serie de tiempo y sus componentes.....	50
Figura 4. Esquema general de un diagrama de dispersión y sus componentes.....	51
Figura 5. Esquema general de un gráfico de curva de masa.....	52
Figura 6. Series de tiempo de las estaciones de caudal seleccionadas para el estudio.	64
Figura 7. Series de tiempo de las estaciones de precipitación usadas en el relleno de datos.	67
Figura 8. Prueba de homogeneidad usando curvas de masa.....	68
Figura 9. Proporción de la varianza acumulada para los escenarios de predicciones de caudal usando ACP	70

Figura 10. Resultados del escenario utilizando 31 índices y 90% de varianza con el método ACP (etapa de calibración).....	71
Figura 11. Resultados del escenario utilizando 41 índices y 90% de varianza con el método ACP (etapa de calibración).....	72
Figura 12. Correlación entre los índices de teleconexión y los componentes ocho y 14 de la modelación ACP utilizando 31 índices.....	76
Figura 13. Correlación entre los índices de teleconexión y los componentes 10 y 20 de la modelación ACP utilizando 41 índices.....	77
Figura 14. Resultados del análisis de dispersión entre el escenario utilizando 31 índices con el método ACP y 90% de varianza (etapa de calibración).	78
Figura 15. Resultados del análisis de dispersión entre el escenario utilizando 41 índices con el método ACP y 90% de varianza (etapa de calibración).	79
Figura 16. Grafica de representación de la correlación existente entre las estaciones de caudal y los Índices de teleconexión.	80
Figura 17. Resultados del escenario utilizando 31 índices y un mes de retraso (etapa de calibración).	83
Figura 18. Resultados del escenario utilizando 41 índices y un mes de retraso (etapa de calibración).	84
Figura 19. Ubicación de los caudales seleccionados para el estudio y regímenes de precipitación en la cuenca del río Paute.....	86
Figura 20. Resultados del análisis de dispersión el escenario utilizando 31 índices con un mes de retraso (etapa de calibración).....	88
Figura 21. Resultados del análisis de dispersión entre el escenario utilizando 41 índices con un mes de retraso (etapa de calibración).....	89
Figura 22. Frecuencia de aparición de los índices de teleconexión en los MLR a nivel general en el escenario VIF.....	90
Figura 23. Resultados del escenario utilizando 31 índices y análisis de multicolinealidad (etapa de calibración).	94
Figura 24. Resultados del escenario utilizando 41 índices y análisis de multicolinealidad (etapa de calibración).	95

Figura 25. Resultados del análisis de dispersión entre la serie de tiempo predicha utilizando 31 índices con el análisis de multicolinealidad (etapa de calibración).	97
Figura 26. Resultados del análisis de dispersión entre la serie de tiempo predicha utilizando 41 índices con el análisis de multicolinealidad (etapa de calibración).	97
Figura 27. Frecuencia de aparición de los índices de teleconexión en los escenarios VIF y con retrasos.	101
Figura 28. Resultados de NSE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.	103
Figura 29. Resultados de KGE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.	105
Figura 30. Resultados de R^2 para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.	106
Figura 31. Resultados de RMSE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.	108
Figura 32. Resultados del escenario con 31 índices y 70% de varianza en el análisis ACP.	129
Figura 33. Resultados del escenario con 41 índices y 70% de varianza en el análisis ACP.	130
Figura 34. Resultados del escenario con 31 índices y dos meses de retraso.	131
Figura 35. Resultados del escenario con 31 índices y tres meses de retraso.	132
Figura 36. Resultados del escenario con 41 índices y dos meses de retraso.	133
Figura 37. Resultados del escenario con 41 índices y tres meses de retraso.	134
Figura 38. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 70% de varianza.	135
Figura 39. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 70% de varianza.	136
Figura 40. Resultados del análisis de dispersión el escenario con 31 índices con dos meses de retraso.	137
Figura 41. Resultados del análisis de dispersión el escenario con 31 índices con tres meses de retraso.	138
Figura 42. Resultados del análisis de dispersión el escenario con 41 índices con dos meses de retraso.	139

Figura 43. Resultados del análisis de dispersión el escenario con 41 índices con tres meses de retraso.....	140
Figura 44. Resultados del escenario con 31 índices y 70% de varianza en el análisis ACP.	141
Figura 45. Resultados del escenario con 31 índices y 90% de varianza en el análisis ACP.	142
Figura 46. Resultados del escenario con 41 índices y 70% de varianza en el análisis ACP.	143
Figura 47. Resultados del escenario con 41 índices y 90% de varianza en el análisis ACP.	144
Figura 48. Resultados del escenario con 31 índices y un mes de retraso.	145
Figura 49. Resultados del escenario con 31 índices y dos meses de retraso.	146
Figura 50. Resultados del escenario con 31 índices y tres meses de retraso.	147
Figura 51. Resultados del escenario con 41 índices y un mes de retraso.	148
Figura 52. Resultados del escenario con 41 índices y dos meses de retraso.	149
Figura 53. Resultados del escenario con 41 índices y tres meses de retraso.	150
Figura 54. Resultados del escenario con 31 índices y análisis de multicolinealidad.....	151
Figura 55. Resultados del escenario con 41 índices y análisis de multicolinealidad.....	152
Figura 56. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 70% de varianza.	153
Figura 57. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 90% de varianza.	154
Figura 58. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 70% de varianza	155
Figura 59. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 90% de varianza	156
Figura 60. Resultados del análisis de dispersión el escenario con 31 índices con un mes de retraso.	157
Figura 61. Resultados del análisis de dispersión el escenario con 31 índices con dos meses de retraso.....	158
Figura 62. Resultados del análisis de dispersión el escenario con 31 índices con tres meses de retraso.....	159
Figura 63. Resultados del análisis de dispersión el escenario con 41 índices con un mes de retraso.	160

Figura 64. Resultados del análisis de dispersión el escenario con 41 índices con dos meses de retraso.....	161
Figura 65. Resultados del análisis de dispersión el escenario con 41 índices con tres meses de retraso.....	162
Figura 66. Resultados del análisis de dispersión entre la serie de tiempo predicha con 31 índices con el análisis de multicolinealidad.	163
Figura 67. Resultados del análisis de dispersión entre la serie de tiempo predicha con 41 índices con el análisis de multicolinealidad.	164
Figura 68. Resultados de NSE para cada escenario de predicción de las estaciones de caudal seleccionadas.....	165
Figura 69. Resultados de KGE para cada escenario de predicción de las estaciones de caudal seleccionadas.....	165
Figura 70. Resultados de R^2 para cada escenario de predicción de las estaciones de caudal seleccionadas.....	166
Figura 71. Resultados de RSME para cada escenario de predicción de las estaciones de caudal seleccionadas.....	166

Índice de tablas

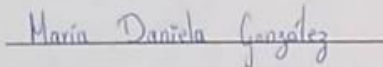
Tabla 1. Descripción de los índices de teleconexión.	21
Tabla 2. Estaciones de Caudal usadas para el estudio.	47
Tabla 3. Métricas usadas del paquete HydroGOF.	59
Tabla 4. Valores de correlación de los índices usados para los modelos de regresión múltiple de cada estación de caudal.	80
Tabla 5. Modelos de represi3n linear múltiple para cada estación de caudal usando el método VIF.	92
Tabla 6. Resultados estadísticos de las métricas aplicadas a los modelos para evaluar su funcionamiento.	99
Tabla 7. Promedio del valor P de cada estación de caudal con sus distintos escenarios de predicción.....	102

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

María Daniela González Salto en calidad de autora y titular de los derechos morales y patrimoniales del trabajo de titulación "ESTUDIO COMPARATIVO DE LA PREDICTIBILIDAD DEL CAUDAL DE LA CUENCA DEL RÍO PAUTE BASADO EN LOS ÍNDICES DE TELECONEXIÓN Y LOS MODOS DE VARIABILIDAD", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio Institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 15 de julio del 2022



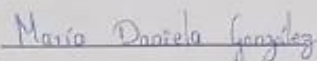
María Daniela González Salto

C.I: 0150268753

Cláusula de Propiedad Intelectual

María Daniela González Salto autora del trabajo de titulación "ESTUDIO COMPARATIVO DE LA PREDICTIBILIDAD DEL CAUDAL DE LA CUENCA DEL RÍO PAUTE BASADO EN LOS ÍNDICES DE TELECONEXIÓN Y LOS MODOS DE VARIABILIDAD", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autora.

Cuenca, 15 de julio del 2022



María Daniela González Salto

C.I.: 0150268753

Agradecimientos

Expreso mi sincera gratitud y agradecimiento a mi familia, mi mamá Mónica, y mi hermana Emilia, quienes me han apoyado con cariño y valentía en esta etapa académica. Le agradezco especialmente a mi papito Flavio, por su apoyo y sabiduría durante la realización de este trabajo.

Agradezco a mi director el Ing. Alex Avilés, y a mi tutor el Ing. Darío Zhiña quienes además de ser profesionales de excelencia, son personas brillantes y gentiles que siempre estuvieron dispuestos a ayudarme y enseñarme para cumplir con los objetivos de este trabajo, brindándome también una sensación de confianza y ánimos altos. Agradezco también a todos los docentes que han sido parte de mi crecimiento académico desde mi niñez, y a la Universidad de Cuenca por ser el espacio en el que pude culminar una importante etapa de mi preparación para el mundo.

Dedicatoria

Le dedico este proyecto de graduación a mi familia. Especialmente a mi papa Flavio, por siempre ser en mi vida una luz cálida en la incertidumbre oscura que es la vida, su apoyo como: amigo, maestro, padre, sustento y sobre todo hogar es lo que me ha permitido culminar mi carrera profesional con enseñanzas de gran valor que van más allá de lo científico y técnico.

Un cariño especial dedicado a mis tres angelitos: Sinjun, Tigris y Sucucus, mis fieles compañeros felinos. Y a mi callejero, mi perro Tiberiano. Ellos son verdaderamente la inspiración para construir y trabajar en la mejora del mundo, para poder cobijar a los seres más frágiles que forman parte de nuestro entorno.

INTRODUCCIÓN

El cambio climático alrededor del mundo ha provocado fuertes alteraciones en el ecosistema, en cada nivel de funcionamiento de la vida. Su influencia en los ciclos hidrológicos ha ocasionado la aparición de fenómenos en las regiones tropicales del Océano Atlántico y Pacífico. Estos fenómenos naturales pueden causar daños incommensurables en los lugares donde se producen, con grandes pérdidas materiales y vidas humanas. Ya sea por la acción directa del viento (huracanes), o de forma indirecta por las inundaciones provocadas por las lluvias y ondas de tormenta (Capurro, 2002). En la actualidad está comprobado que las inundaciones son las catástrofes naturales que mayor número de víctimas producen en el mundo. Se cuantificó que aproximadamente tres millones de personas murieron en el siglo XX a causa de estos eventos. Es decir que en una perspectiva global las inundaciones han ocasionado más de la mitad de muertes del valor total atribuido a los desastres naturales de todo tipo (Gómez, 2014). Toda la región de Latinoamérica y el Caribe está catalogada como la más vulnerable a los desastres naturales. Entre 1980 - 2010, los daños económicos asociados a estos eventos sobrepasaron el total de los daños registrados durante las ocho décadas anteriores (1900 - 1980). Esto sumado a que las poblaciones más pobres son las más afectadas y a los factores que influyen negativamente estos impactos, como la falta de planificación territorial, la degradación ambiental y la carencia de estabilidad política y gubernamental (Fernandez & Buss, 2016).

En el Ecuador los desastres naturales de mayor frecuencia son los deslizamientos de tierra, las inundaciones y sequías, siendo estos dos últimos muy influyentes en las condiciones de vida dependiendo la zona donde ocurran (PNUMA et al., 2008). El problema se traduce en términos económicos y humanos y se puede abordar este tema desde dos criterios. La sequía crónica que existe en la zona Andina y que restringe el flujo de alimentación de la principal central hidroeléctrica del país. Lo cual genera por defecto un aumento en la producción de energía termoeléctrica y desestabiliza la distribución en la red por determinados periodos de tiempo. Y, por otro lado, en la Costa ecuatoriana ocurre el tema de las inundaciones periódicas. Esto debido a la falta de drenaje de las aguas-lluvia y a la concentración de las precipitaciones en pocos meses, este último factor agravado por la presencia del evento El Niño Oscilación del Sur (ENOS) (G. Gómez et al., 1996). Aún en la actualidad los gobiernos y entidades encargadas de manejar estos

temas no tienen la suficiente información para crear planes de contingencia que reduzcan los impactos ya mencionados en el país, debido a las condiciones hidrológicas de la región.

Las familias damnificadas que debido a eventos extremos del fenómeno ENOS perdieron sus viviendas y propiedades fueron en total 14.000, 9% de esta cantidad únicamente en el Oriente. Las afectaciones en términos económicos en el Oriente se ubican en un 7% del daño total en el país (Corporación Andina de Fomento, 1998).

En la actualidad en Sudamérica el fenómeno ENOS tiene una tendencia de inundaciones y sequías a lo largo de la costa occidental tropical de Ecuador / Perú y Colombia, respectivamente. El impacto que esta teleconexión pueda tener depende de los modos de variabilidad climática (decadal, interanual, Temperatura Superficial del Mar (TSM) del Atlántico Norte y del Océano Índico) y especialmente de las interacciones entre cuencas entre el Pacífico y el Atlántico. La variabilidad a escala decadal, tanto en lo que respecta al ENOS como a su variabilidad, agrega complejidad al diagnóstico de las teleconexiones, especialmente al registro instrumental relativamente corto. En la actualidad aún hacen falta investigaciones meteorológicas y climáticas continuas, desarrollos de modelos e investigaciones específicas. Extender los registros instrumentales de tiempo usados en los estudios puede ayudar a entender mejor la complejidad de los impactos de ENOS en el pasado y poder proyectar los mismos en el futuro. Existen varios sesgos y grados de incertidumbre al querer simular ENOS en Sudamérica, los más importantes son los errores sistemáticos al trabajar con series de tiempo extensas (centro del Pacífico ecuatorial, este del Océano Índico ecuatorial, regiones que contienen los sistemas de corriente de límite, incluidos el Pacífico tropical y el Atlántico). Las proyecciones más realistas sugieren que al incluir el efecto invernadero en las simulaciones se puede esperar impactos del ENOS más extremos en el futuro (Cai et al., 2020).

En Ecuador los estudios demuestran que además de la influencia del fenómeno ENOS y la variabilidad de la TSM, los cuales se encuentran muy asociados con el comportamiento de los caudales a través de todo el Ecuador. ENOS y sus modos pueden detectar anomalías en el flujo de corrientes en todo el país (Córdoba Machado et al., 2015).

La cuestión central que desea resolver en este estudio es el uso de índices de teleconexión para predecir el comportamiento de las condiciones hidrológicas (caudales). Sin embargo, además de

utilizar los índices de teleconexión, hacen falta estudios con otro tipo de evaluaciones como los modos de variabilidad que pueden arrojar mejores resultados.

Cada año que transcurre, los fenómenos hidrológicos que afectan a las distintas zonas del Ecuador incrementan sus impactos debido al cambio climático y esto se traduce en pérdidas humanas y materiales. Por tal razón, es fundamental que, si no es posible eliminar por completo el riesgo o impactos que tienen las inundaciones o sequías en las zonas de mayor vulnerabilidad, se realice un plan de contingencia. Este estudio basado en los índices climáticos y los modos de variabilidad servirá como una línea base para entender de forma profunda los fenómenos hidrometeorológicos que suceden en Ecuador y poder desarrollar a futuro un plan de contingencia que reduzca en gran medida la incertidumbre de los posibles y potenciales problemas que generan estos eventos. Con esto se podría mejorar las condiciones de vida de los pobladores que se ven afectados directa o indirectamente por los eventos de inundación.

OBJETIVOS

Objetivo general:

- Determinar la predictibilidad de caudal en la cuenca del río Paute mediante el uso de índices de teleconexión y modos de variabilidad.

Objetivos específicos:

- Recopilar y procesar la información sobre los índices de teleconexión y modos de variabilidad que afectan a los caudales en la cuenca del río Paute.
- Generar predicciones de caudal en la cuenca del río Paute mediante el uso de índices de teleconexión y los modos de variabilidad.
- Comparar las predicciones generadas por los índices de teleconexión y los modos de variabilidad con los datos registrados en las estaciones hidrológicas en el periodo de estudio.

CAPITULO 1: REVISIÓN DE LA LITERATURA

1.1 Índices de Teleconexión

Los índices de teleconexión son correlaciones estadísticamente significativas de las anomalías circulatorias recurrentes en la atmosfera, a nivel planetario o hemisférico que ocurren en regiones adyacentes y lejanas, a veces de forma simultánea (Hatzaki et al., 2007). El descubrimiento de las teleconexiones es atribuido a Sir Gilbert Walker en 1924, quien, en un estudio en la India, identificó una serie de sistemas de circulación de baja frecuencia dentro del clima global. El término teleconexión fue utilizado por primera vez en 1935 por Angstrom al describir a la Oscilación del Atlántico Norte (NAO) (Quiring & Blair, 2000).

En 1970 la importancia de las teleconexiones cobró reconocimiento luego de una extensa investigación sobre el evento ENOS, actualmente este es el índice más conocido y estudiado, el cual implica interacciones océano-atmósfera a gran escala. Además, se sabe que los patrones de la TSM del Atlántico y del Pacífico tropical tienen mucha influencia en la circulación atmosférica de los trópicos, y que los efectos de este patrón se transportan al resto del sistema climático como una serie de ondas a escala global (Quiring & Blair, 2000).

El estudio de las teleconexiones climáticas tiene un gran interés por sus aplicaciones en el campo de predicciones estacionales, de caudal y precipitación. También ayuda a mejorar la gestión de los recursos hídricos y la planificación anual de riesgos (inundaciones y sequías). Así como la distribución para otros servicios que dependen de estos recursos, la generación de energía eléctrica, planificación de cultivos, entre otros, en una escala de aplicación global (Pastor, 1999).

Existen distintos índices de teleconexión identificados en varias zonas de acuerdo a su origen, transporte y área de impactos alrededor del mundo, los mismos se mencionan en la Tabla 1.

Tabla 1. Descripción de los índices de teleconexión.

ÍNDICE DE TELECONEXIÓN	ABREVIATURA	RANGO	DESCRIPCIÓN
OSCILACIÓN ANTÁRTICA (ANTARCTIC OSCILLATION)	AAO	Latitudes medias y altas del hemisferio Sur	Es una variación estacional de presión y altura geopotencial en las latitudes medias y altas del hemisferio sur. Se identifica como una oscilación a nivel del mar entre los cinturones esta misma zona (alternancia a gran escala de masa atmosférica) (Alves et al., 2017). También se conoce como El modo anular del sur o Southern Anular Mode (SAM).
MODO MERIDIONAL ATLÁNTICO (ATLANTIC MERIDIONAL MODE)	AMM	Atlántico tropical	El Modo Meridional del Atlántico se caracteriza por generar un gradiente de temperatura meridional de TSM cerca de la ubicación de la Zona de Convergencia Intertropical Climatológica (ZCIT). También produce vientos que fluyen hacia el agua anómalamente más cálida y giran hacia la derecha (izquierda) en el hemisferio norte (sur) (con efecto Coriolis) (Vimont & Kossin, 2007).
OSCILACIÓN MULTIDECADAL ATLÁNTICA NO SUAVIZADA VERSIÓN LARGA (ATLANTIC MULTIDECADAL OSCILLATION UNSMOOTHED LONG VERSION)	AMO (AMOU)	Oceano Atlántico del Norte	Este índice tiene efecto en el hemisferio Norte, se conoce como Variabilidad Multidecadal del Atlántico. Es un modo interno de variabilidad climática que se origina en los cambios en la circulación del Océano Atlántico, es una anomalía en la TSM promedio del Atlántico norte menos la TSM media global. Tiene varias fases positivas que duran de 40 a 20 años y dos fases negativas que duran aproximadamente 30 años (Cuevas et al., 2019). Contiene valores sin tendencia desde el año 1948.

<p>OSCILACIÓN MULTIDECADAL ATLÁNTICA SUAVIZADA VERSIÓN LARGA (ATLANTIC MULTIDECADAL OSCILLATION SMOOTHED LONG VERSION)</p>	<p>AMO (AMOS)</p>	<p>Oceano Atlántico del Norte</p>	<p>Similar al índice AMO, pero contiene valores desde el año de 1856 y mantiene una tendencia (Cuevas et al., 2019).</p>
<p>OSCILACIÓN ÁRTICA (ARCTIC OSCILLATION)</p>	<p>AO</p>	<p>Hemisferio Norte, medio oriente.</p>	<p>La Oscilación del Ártico (AO), representa un balancín a gran escala, domina las latitudes medias y la masa atmosférica del Ártico, influye en el clima de Eurasia, América del Norte, el este de Canadá, el norte de África y el Medio Oriente, especialmente durante el invierno boreal. Es una función ortogonal de la de presión al nivel del mar (He et al., 2017).</p>
<p>FUNCIÓN EMPÍRICA ORTOGONAL DE LA TSM DEL ATLÁNTICO TRIPOLAR (ATLANTIC TRIPOLE SST EOF)</p>	<p>AT SST EOF</p>	<p>Atlántico Norte</p>	<p>El índice de la función empírica ortogonal de la TSM del Atlántico Tripolar tiene una variabilidad que se puede representar por el índice tripolar de Czaja y Marshall, este se reconoce como una anomalía de la TSM promediada sobre una caja norte (40°N – 55°N, 60°O – 40°O) menos el de un cuadro del sur (25°–35°N, 80°–60°O) (Fan & Schneider, 2012).</p>
<p>SERIE TEMPORAL ENOS BIVARIADA VERSIÓN LARGA (BIVARIATE ENSO TIMESERIES LONGER VERSION)</p>	<p>BEST</p>	<p>Pacífico Tropical Oriental Central, Australia.</p>	<p>Se calcula a partir de la serie de tiempo SOI y la TSM del índice Niño 3.4 con sus valores estandarizados. Al incluir el índice SOI al Niño 3.4 los procesos atmosféricos explícitos son incluidos de forma directa. También se reducen los sesgos en datos de TSM que fueron reconstruidos (Smith & Sardeshmukh, 2000).</p>

<p>ÍNDICE DEL CARIBE (CARIBBEAN INDEX)</p>	<p>CAR</p>	<p>Zona del Caribe</p>	<p>El Índice del Caribe está basado en pronósticos de la TSM reconstruida y ampliada con datos de la NOAA, empleando modelos lineales inversos y buscando calcular las anomalías relativas a la climatología de 1981 - 2010. Cuando estas anomalías se promediaron sobre la zona del caribe (limitado por 26°N, 80°W y la costa este de América Central) se crea el valor del índice CAR, que representa la desviación de la TSM actual en toda la región del Caribe a partir de promedios históricos (Gingold et al., 2014).</p>
<p>PRECIPITACIÓN EN LA INDIA CENTRAL (REGIÓN MONZÓN DEL NÚCLEO) (CENTRAL INDIAN PRECIPITATION (CORE MONSOON REGION))</p>	<p>CIP</p>	<p>India</p>	<p>Este índice se obtiene a partir de 306 estaciones distribuidas uniformemente (30 subdivisiones meteorológicas), para las cuales existen datos desde el año 1871. Las series mensuales de precipitación ponderada por área de todo el año de cada una de las 30 subdivisiones meteorológicas se preparan asignando un peso según el tamaño de área del distrito como el peso para cada estación pluviométrica en esa subdivisión (Kumar et al., 2010).</p>
<p>ÍNDICE DEL ATLÁNTICO ESTE- OESTE DE RUSIA (THE EAST ATLANTIC-WEST RUSSIA)</p>	<p>EA/WR</p>	<p>Europa, norte de China, centro del Atlántico norte y norte del Mar Caspio.</p>	<p>Es un conjunto de anomalías del Atlántico. La fase positiva está asociada con anomalías de altura positivas ubicadas en Europa y el norte de China, y anomalías de altura negativas ubicadas en el centro del Atlántico norte y norte del Mar Caspio. Durante la fase positiva, se observan condiciones más húmedas en el este de China y más secas en Europa central y la región del Mediterráneo. Se sabe que este índice tiene una fuerte influencia sobre las precipitaciones en la región mediterránea (Barnston & Livezey, 1987).</p>

<p>ÍNDICE DE PRECIPITACIÓN DE EL NIÑO OSCILACIÓN DEL SUR (ENOS PRECIPITATION INDEX)</p>	<p>ENOS</p>	<p>Cuenca del Océano Pacífico tropical.</p>	<p>Es un fenómeno interanual conocido como El Niño Oscilación del Sur. En su fase cálida afecta el Océano Pacífico tropical provocando aguas anómalamente cálidas en el este, alisios débiles y presión baja acompañada de lluvias intensas y condiciones contrarias en el oeste. La fase fría se conoce como La Niña, que es un evento equivalente, pero de signo contrario (Curtis & Adler, 2000)</p>
<p>OSCILACIÓN DEL PACÍFICO ESTE/ PACÍFICO NORTE (EAST PACIFIC/NORTH PACIFIC OSCILLATION)</p>	<p>EP/NP</p>	<p>Alaska, oeste de Canadá, Pacífico norte central y el este de América del Norte.</p>	<p>Se basa en un patrón primavera-verano-otoño. Se presenta como anomalías positivas de altura ubicadas sobre Alaska/ oeste de Canadá, y anomalías negativas sobre el Pacífico norte central y el este de América del Norte. Las anomalías se asocian con la intensificación de corriente del Pacífico y un ascenso de temperatura y precipitación (Schulte & Lee, 2017)</p>
<p>ÍNDICE DEL PROMEDIO DE LA TEMPERATURA GLOBAL DEL OCÉANO/ LAN (GLOBAL MEAN LAN/OCEAN TEMPERATURE INDEX)</p>	<p>GB</p>	<p>Escala global áreas oceánicas.</p>	<p>Es índice es una medida del cambio de temperatura del aire en la superficie, sirve para cuantificar el cambio climático global. Se combinan redes de estaciones meteorológicas terrestres y marinas. Aunque solo puede ser medido en áreas oceánicas (Hansen et al., 1996).</p>
<p>MOMENTO GLOBALMENTE ANGULAR INTEGRADO (GLOBALLY INTEGRATED ANGULAR MOMENTUM)</p>	<p>GIAM</p>	<p>Escala global</p>	<p>El momento angular global integrado refleja tanto la rotación ligada a la de la Tierra como la rotación debida a los vientos. Sirve para estudiar la dinámica en diferentes escalas temporales y espaciales. Como índice de la circulación atmosférica refleja muchos aspectos de la firma del clima y el tiempo (Salstein, 2015).</p>

<p>ÍNDICE BIMESTRAL ENOS MULTIVARIADO (MULTIVARIATE ENOS INDEX)</p>	<p>MEI V2</p>	<p>Pacífico tropical (región de “El Niño”)</p>	<p>Es una combinación lineal de seis variables atmosféricas observadas sobre el Pacífico tropical: presión a nivel del mar, componentes zonal y meridional del viento en la superficie, temperatura de la superficie del mar, temperatura del aire en la superficie y nubosidad del cielo. Los valores mensuales de MEI están estandarizados con respecto a un periodo de referencia 1950 - 1993 (Mazzarella et al., 2013).</p>
<p>OSCILACIÓN DEL ATLÁNTICO NORTE (NORTH ATLANTIC OSCILLATION)</p>	<p>NAO</p>	<p>Costa este de los Estados Unidos hasta Siberia y desde el Ártico hasta el Atlántico subtropical.</p>	<p>Es uno de los patrones más influyentes y frecuentes de variabilidad. Se refiere a una redistribución de la masa atmosférica entre el Ártico y el Atlántico subtropical, produce grandes cambios en la velocidad y dirección media del viento sobre el Atlántico, y los otros continentes, así como también modifica el transporte de calor, humedad (intensidad), el número de tormentas, sus trayectorias y su clima (Hurrell et al., 2003).</p>
<p>OSCILACIÓN DEL ATLÁNTICO NORTE JONES (NORTH ATLANTIC OSCILLATION JONES)</p>	<p>NAO (Jones)</p>	<p>Costa este de los Estados Unidos hasta Siberia y desde el Ártico hasta el Atlántico subtropical.</p>	<p>Extensión del índice NAO con nuevas observaciones instrumentales para el periodo 1823 - 1996 (Jones et al., 1997).</p>

<p>ANOMALÍA DE PRECIPITACIONES EN EL NORESTE DE BRASIL (NORTHEAST BRAZIL RAINFALL ANOMALY)</p>	<p>NBRA</p>	<p>Noreste de Brasil</p>	<p>El Índice de anomalías de lluvia del noreste de Brasil se usa para el estudio de la atmosfera y el océano y tiene un periodo de 1849 - 2002. Este índice combina los datos de series temporales de precipitaciones normalizadas medidas en la zona costera de Fortaleza y 100 km tierra adentro en la región de Quixeramobim. Esta área únicamente esta influenciada por la variabilidad de la ZCIT del Atlántico por lo que es una buena zona de estudio para analizar la variabilidad de las precipitaciones o la migración estacional (Brito, 2011).</p>
<p>TSM DEL PACÍFICO TROPICAL ORIENTAL EXTREMO (EXTREME EASTERN TROPICAL PACIFIC SST)</p>	<p>Niño 1+2</p>	<p>Costa de Sudamérica.</p>	<p>La región Niño 1+2 es la más angosta y oriental de las regiones del Niño, se ubica en la región costera de América del Sur (0°S - 10°S, 90°W - 80°W) donde El Niño fue reconocido por primera vez por las poblaciones locales. Este índice tiende a tener la mayor variación de los índices de Niño TSM Se asocia con el incremento de temperaturas y precipitaciones (Van Oldenborgh et al., 2021)</p>
<p>TSM DEL PACÍFICO TROPICAL ORIENTAL (EASTERN TROPICAL PACIFIC SST)</p>	<p>Niño 3</p>	<p>Pacífico Tropical Oriental</p>	<p>Tiene el mismo funcionamiento atmosférico que el Niño 1+2, diferenciándose de este por su ubicación en: 5°N - 5°S 150°W. 90°W, cubre la región de la franja fría (Van Oldenborgh et al., 2021).</p>
<p>TSM DEL PACÍFICO TROPICAL ORIENTAL CENTRAL (EAST CENTRAL TROPICAL PACIFIC SST)</p>	<p>Niño 3.4</p>	<p>Pacífico Tropical Oriental Central</p>	<p>Se ubica en: 5°N–5°S 120°W–170°W, es el índice de teleconexión actualmente más usado para monitorear la intensidad de los fenómenos del Niño y la Niña, usa una media móvil de cinco meses (Tyaquiçã, 2017).</p>

TSM DEL PACÍFICO TROPICAL CENTRAL (CENTRAL TROPICAL PACIFIC SST)	Niño 4	Pacífico Tropical Central	<p>Tiene el mismo funcionamiento atmosférico que el Niño 1+2, diferenciándose de este por su ubicación en: 5°N - 5°S 160°E - 150°W, está relacionado con muchos índices de teleconexión globales (Tyaquiçã, 2017).</p>
ÍNDICE DE OSCILACIÓN DEL NORTE (NORTHERN OSCILLATION INDEX)	NOI	<p>Alto del Pacífico Norte, el Pacífico noreste y cerca de Darwin, Australia.</p>	<p>Este índice se basa en la diferencia de las anomalías de la presión del nivel del mar en el Alto del Pacífico Norte, el Pacífico noreste y cerca de Darwin, Australia. Representa una amplia gama de eventos climáticos tropicales y extratropicales que impactan el Pacífico norte en escalas intraestacionales, interanuales y decadales asociados con eventos de El Niño y La Niña (Schwing et al., 2002).</p>
PATRÓN DEL PACÍFICO NORTE (NORTH PACIFIC PATTERN)	NP	<p>Pacífico norte occidental y central, este de Siberia, Alaska y América del Norte.</p>	<p>El patrón del Pacífico Norte es prominente de marzo a julio. Tiene un centro de anomalías primario, que abarca las latitudes centrales del Pacífico norte occidental y central, y una región más débil de signo opuesto, que se extiende por el este de Siberia. Se caracteriza por una intensificación de la corriente en chorro del Pacífico y circulaciones ciclónicas y anticiclónicas</p>
NORTE DEL ATLÁNTICO TROPICAL (NORTH TROPICAL ATLANTIC)	NTA	<p>Norte del Atlántico Tropical</p>	<p>La NTA influye en la TSM alcanzando su punto máximo en la primavera boreal, con un calentamiento en su fase positiva. La NTA también influye profundamente en la precipitación en el noreste de Brasil. Este índice surge de anomalías en el flujo de calor latente asociadas con intercambios anómalos del noreste (Yang et al., 2021).</p>

ÍNDICE DE NIÑO OCEÁNICO (OCEANIC NIÑO INDEX)	ONI	Océano Pacífico ecuatorial centro-este, que se llama la región Niño 3.4 (5°S a 5°N; 170°W a 120°W)	Es una serie de tiempo de las anomalías de la temperatura superficial marina calculadas a partir de mediciones in situ efectuadas en la región Niño 3.4 (5°N - 5°S, 120°O - 170°O). El periodo base abarca de 1971 a 2000 y los datos se obtienen con el método de medias móviles aplicado a periodos de tres meses (Jiménez Quiroz, 2014).
OSCILACIÓN DECADAL DEL PACÍFICO (PACIFIC DECADAL OSCILLATION)	PDO	Océano Pacífico Norte y Sur, Australia y América del Sur	Este índice resulta de un análisis de EOF (Función empírica ortogonal) de las anomalías mensuales residuales de la TSM del Pacífico Norte, hacia los polos de 20°N para el periodo de registro de 1900 - 1993. Los “residuales” se definen como la diferencia entre las anomalías observadas y las anomalías promedio mensuales globales de la TSM. Tiene una persistencia multianual y multidecadal con cambios de signo (Mantua Nathan J. & Hare Steven R., 2002).
PATRÓN DEL PACÍFICO Y AMÉRICA DEL NORTE (PACIFIC NORTH AMERICAN INDEX)	PNA	Norteamérica y el Océano Pacífico	Es un conjunto de variaciones en la circulación atmosférica que se sitúa sobre Norteamérica y el océano pacífico, muy influyente en el invierno. Se identifica como anomalías en los campos de altura geopotencial, es decir, alteraciones de la distancia vertical existente entre el nivel medio del mar y ciertos niveles de presión (700 y 500 hPa). (Rincon, 2017)
REGIÓN DE LA PISCINA CÁLIDA DEL PACÍFICO (PACIFIC WARMPOOL REGION)	PWR	Ecuador, Sur de la India, Sumatra, Java, Borneo, Nueva Guinea y Océano Pacífico central.	La piscina cálida del Indo-Pacífico se define como la región del océano que contiene la isoterma con una TSM de: 27,5°C o 28°C, que son las temperaturas umbral requeridas para la convección atmosférica profunda. Esta región se mueve a través de las circulaciones de Walker y Hadley y sirve como una fuente importante de calor y vapor de agua (Kim et al., 2012).

<p>OSCILACIÓN CUASI-BIENAL (QUASI-BIENNIAL OSCILLATION)</p>	<p>QBO</p>	<p>Zona del Ecuador terrestre</p>	<p>La Oscilación Cuasi-Bienal es el patrón dominante de la variabilidad en la estratosfera ecuatorial. Se conoce que, entre tres y 100 hPa, el viento zonal en el ecuador tiene un patrón de regiones de cizalladura descendentes del este y del oeste, con cambios de dirección del viento aproximadamente cada 14 meses (Schenzinger et al., 2016).</p>
<p>FLUJO DE RADIO SOLAR DE 10,7 CM (SOLAR FLUX (10,7CM))</p>	<p>SF</p>	<p>Escala Global</p>	<p>Este índice se conoce como el flujo de radio solar de 10,7 cm (2800 MHz), sirve para indicar la actividad solar, posee un gran registro observacional y está directamente relacionado con la cantidad de manchas solares y algunos registros de radiación ultravioleta (UV) y solar visible en el rango 3,4 de onda. Es fácilmente medible bajo todas las condiciones climáticas (Zhang et al., 2022).</p>
<p>ÍNDICE DE OSCILACIÓN DEL SUR (SOUTHERN OSCILLATION INDEX)</p>	<p>SOI</p>	<p>Pacífico tropical entre Tahití y Darwin, Australia.</p>	<p>Representa las fluctuaciones a gran escala de la presión media del nivel del mar en el Pacífico tropical entre Tahití y Darwin, Australia. Este índice normalmente se expresa como un número que varía de -30 a +30. Los periodos prolongados de valores de SOI negativos se denominan eventos de El Niño, donde la presión del aire está por debajo de lo normal en Tahití y por encima de lo normal en Darwin. La condición opuesta se denomina La Niña.</p>
<p>LLUVIA DEL SAHEL (SAHEL RAINFALL)</p>	<p>SR</p>	<p>Norte del continente africano</p>	<p>La zona del Sahel es una región ecoclimática y biogeográfica de transición entre sus límites, al norte con el desierto del Sahara y al Sur con la sabana sudanesa. La temporada de lluvia de esta zona sucede de junio a octubre, y se toman las medias para esos meses. El índice se calcula a partir de datos con una resolución de 0,25 grados. Y el promedio de la región se determinó mediante un</p>

			análisis de componentes principales rotados de la precipitación africana (Mitchell, 2018).
LLUVIA EN LA REGIÓN MONZÓN DEL SUR OESTE (SW MONSOON REGION RAINFALL)	SW MRR	Arizona y Nuevo México	Este índice se obtiene al calcular usando el conjunto de datos de la división climática de National Climate Data Center (NCDC). Los valores de precipitación mensual para cada una de las divisiones climáticas en Arizona y Nuevo México se promedian para producir un único valor mensual (National Oceanic and Atmospheric Administration, 2022).
ÍNDICE DEL ATLÁNTICO NORTE TROPICAL (TROPICAL NORTHERN ATLANTIC INDEX)	TNA	Océano Norte tropical oriental	El índice de anomalías TNA es un indicador de la temperatura superficial del Océano Atlántico Norte tropical oriental. Es calculado con las TSM en el recuadro 55°W - 15°W, 5°N - 25°N (Tyaquiçã, 2017).
INDICE TRANS NIÑO (TRANS-NIÑO INDEX)	TNI	Pacífico hasta la costa estadounidense.	El índice resulta del promedio de anomalías normalizadas de TSM en las regiones Niño 1+2 y Niño 4 (Trenberth & Stepaniak, 2001).
FUNCIÓN EMPÍRICA ORTOGONAL DE LA TSM DEL PACIFICO TROPICAL (TROPICAL PACIFIC SST EOF)	TP SST EOF	Escala global	La función empírica ortogonal de la TSM del Pacífico tropical se reconoce como la única estructura espacial de la variabilidad interanual de la TSM. Este patrón tiene una evolución mensual en amplitud y fase desde el año de 1963 al 1989 (Hoerling et al., 2001).
ÍNDICE TRIPOLAR PARA LA OSCILACIÓN INTERDECADAL DEL PACÍFICO (TRIPOLE INDEX FOR THE INTERDECADAL PACIFIC OSCILLATION)	TPI (IPO)	Pacífico ecuatorial central, noroeste y suroeste del Pacífico.	El índice se basa en la diferencia entre las anomalías de la TSM A promediado sobre el Pacífico ecuatorial central y el promedio de las anomalías de la TSM en el noroeste y suroeste del Pacífico. Es una medida de la variabilidad interdecadal en el Pacífico (Henley et al., 2015).

<p>ÍNDICE DEL ATLÁNTICO SUR TROPICAL (TROPICAL SOUTHERN ATLANTIC INDEX)</p>	<p>TSA</p>	<p>Golfo de Guinea, el sur tropical oriental del Océano Atlántico</p>	<p>El índice de anomalías TSA es un indicador de las temperaturas de la superficie ubicadas en el Golfo de Guinea, el sur tropical oriental del Océano Atlántico. Se calcula con las TSM en el recuadro 30°W - 10°E, 20°S - 0 ° (Tyaquicã, 2017).</p>
<p>PISCINA CALIENTE DEL HEMISFERIO OCCIDENTAL (WESTERN HEMISPHERE WARM POOL)</p>	<p>WHWP</p>	<p>Hemisferio Occidental y Océano Pacífico.</p>	<p>La piscina caliente del hemisferio Occidental está compuesta por la temperatura de la superficie del mar (TSM) más cálida de los océanos del mundo. Se expande hacia el este y está asociada a grandes cambios en la convección atmosférica. Es altamente influyente en la evolución del fenómeno ENOS (Wang & Enfield, 2001).</p>
<p>ÍNDICE DEL PACÍFICO OCCIDENTAL (WESTERN PACIFIC INDEX)</p>	<p>WP</p>	<p>El Lejano Oriente y el Pacífico Norte occidental.</p>	<p>Es un modo primario de variabilidad de baja frecuencia en el Pacífico Norte en todos los meses. En invierno y primavera el patrón se presenta como anomalías de altura geopotencial en la península de Kamchatka y partes del sudeste de Asia y el Pacífico Norte subtropical occidental. Durante todas las estaciones se manifiesta al este del Pacífico Norte y el suroeste de los Estados Unidos (Tanaka et al., 2016).</p>

Elaborado por: María Daniela González



1.2 Variabilidad Climática

La secuencia de oscilaciones alrededor de los valores normales, se conoce como variabilidad climática y su valoración se logra mediante la determinación de las anomalías. La variabilidad climática es la respuesta a las fluctuaciones del clima, es decir que es el estado de transición del mismo, y que tiene características específicas de la escala de tiempo y espacio a la que pertenece (IDEAM - UNAL, 2018).

Todos los comportamientos que tienen una escala más amplia que la de los fenómenos meteorológicos, se consideran variaciones climáticas. La variabilidad puede producirse debido a procesos del sistema climático (interno) o a variaciones forzadas antropogénicas (externo). Normalmente las fluctuaciones de clima en las series de tiempo se producen por modificaciones en la interacción de distintos componentes climatológicos y por cambios en los factores radiactivos forzantes (Ramamasy & Baas, 2007).

Aunque la variabilidad está íntimamente ligada con el cambio climático, la diferencia es que la variabilidad climática se evidencia en periodos de tiempo relativamente cortos, y puede identificarse en variables climatológicas como la temperatura y la precipitación. En general, se ha determinado que el cambio climático intensifica los fenómenos de la variabilidad climática (Pabón et al., 2004).

Existen varias escalas temporales de variabilidad climática: escala intraestacional (veranillos e intensificación de las lluvias), escala estacional (estaciones lluviosas y secas), escala interanual (teleconexiones y ciclos de más de un año de duración) y escala interdecadal (fluctuaciones climáticas) (Alzate et al., 2015).

1.2.1 Modos de Variabilidad

Los modos de variabilidad se definen a partir de subcálculos o variaciones temporales en el comportamiento de índices de teleconexión. Por ejemplo, el fenómeno ENOS cuenta con varios modos de variabilidad que afectan a la región del Pacífico, estos corresponden a escalas interanuales e interdecenales. Los modos de variabilidad comparten características comunes con los índices de teleconexión asociados, es decir producen procesos físicos comunes, pero mantienen un origen común para estos climas específicos (Dima & Lohmann, 2004).



1.3 Antecedentes

1.3.1 Fenómenos hidrometeorológicos en el Ecuador

Ecuador está ubicado en la ZCIT, que es un cinturón de baja presión que se sitúa rodeando al globo terrestre a la altura de la región ecuatorial y está formado por la convergencia de aire cálido y húmedo.

Por este motivo el Ecuador en sus distintas regiones tiene influencia de masas de aire con distintas características: Masas tropicales marítimas o MATM (MATM, temperatura entre 24 - 28°C y humedad mayor al 80%), y las Masas tropicales continentales o MATC, que se originan en las planicies del Litoral y el Oriente (MATC, temperatura entre 12 - 24°C y humedad entre el 70 - 80%), las templadas, se sitúan en los valles interandinos (temperatura entre 0 - 12°C y la humedad entre 70 - 60°C), finalmente las de aire frío, se asientan en las mesetas andinas y sobre las montañas que superan los 3000m de altura (temperatura menor o igual a 0°C, humedad menor a 60°C) (PNUMA et al., 2008).

En la región costanera central el clima está fuertemente influenciado por la corriente fría de Humboldt. Entre junio y septiembre la TSM disminuye, ya que masas de aire marítimo frío ingresan a la faja costanera, produciéndose neblinas y lloviznas. La región de la Sierra, se encuentra bajo la influencia alterna de MATM y MATC. Desde inicios de septiembre, la ZCIT que viene desde el norte, rechaza los vientos alisios del sudeste y traslada las MATM al continente. Al unirse los vientos alisios del Noroeste a estos, inicia la estación lluviosa. A fines de diciembre y comienzos de enero, la ZCIT provoca en el norte, una reducción muy sutil de precipitación, esto corresponde al “veranillo de El Niño”. Mientras tanto, debido a las fuertes temperaturas, las MATC en la llanura amazónica siguen intensificándose, se integran al callejón interandino y producen un segundo pico lluvioso desde marzo. La región amazónica, fundamentalmente se encuentra bajo la influencia de las MATC, que funcionan como perturbaciones tropicales (Portilla, 2018)

En el país, los eventos hidrometeorológicos han ocasionado distintos impactos. Las inundaciones son los desastres más frecuentes y de mayor rango de afectación, también se producen sequías, heladas y marejadas. La presencia de ENOS en el Ecuador produce cambios en los fenómenos estivales ya que de manera irregular de dos a siete años altera el clima en toda cuenca del Pacífico.



En el periodo 1988 - 1998 el país sufrió episodios de intensa pluviosidad debido a ENOS, este régimen afecto a diversas provincias de la Costa que son las comúnmente más afectadas por las inundaciones (Santa Elena, Manabí, Los Ríos, y Esmeraldas) (PNUMA et al., 2008).

1.3.2 Funcionamiento de los índices climáticos en Ecuador

La región andina tropical, ha sido identificada como un área de alta vulnerabilidad tanto al cambio climático como a las amenazas climáticas relacionadas. Ecuador tiene un clima influenciado por diversas dinámicas topográficas y atmosféricas contrastantes entre sí que ocurren en la región, esto crea una gran variabilidad que dificulta determinar los efectos del cambio climático (Morán et al., 2016). Como ya se sabe, el país está altamente influenciado por la ZCIT, sujeta a vientos alisios, la humedad directa de la cuenca amazónica y la circulación de Hadley y Walker (Zhiña, 2021). La circulación de Hadley describe la circulación de escala meridional, y la circulación de Walker a escala longitudinal la cual esta espacialmente influenciada por ENOS en la región del océano Pacífico (Gastineau et al., 2009). Por lo que ENOS es el principal patrón de variabilidad climática en Ecuador y los Andes tropicales y controla de forma periódica las fluctuaciones de la TSM y anomalías en la presión del aire superficial en la troposfera media. A su vez este fenómeno produce variabilidad interanual y se divide en dos fases la cálida conocida como El Niño y la fase fría o La Niña.

Otro impulsor importante de la variabilidad climática en esta zona es la PDO, un patrón que se encuentra en las aguas superficiales del Océano pacífico al norte (20°N), varía entre fases cálidas y frías en series interdecenales. Se cree que el cambio climático ha incrementado la influencia de este índice en América del Sur (Morán et al., 2016). Otros índices asociados con ENOS como Niño 3.4 o SOI afectan al Ecuador en el aumento de las precipitaciones. TSM también tiene una alta correlación con esta variable y es causal al Niño 3.4, NP y WP, tienen poca influencia en las lluvias en general del continente suramericano, se encuentran más distribuidos en Brasil y están presentes en las sequías. AO y AAO no se relacionan con la variabilidad climática del Ecuador debido a su lejanía espacial, han interferido con la humedad de la región de forma muy dispersa y con influencias leves. TNA y TSA están directamente relacionados con la presencia de sequías en zonas extensas. NAO no se relaciona en el Ecuador y en general en la zona ni con sequías ni excesos de lluvia (Giddings & Soto, 2006).



1.4 Influencia global de los índices climáticos atmosféricos y oceánicos

Los estudios a través del tiempo han determinado que ENOS es el modo más comúnmente estudiado de variabilidad climática interanual, influye en la región del Pacífico, pero sus impactos también se pueden sentir en gran parte del mundo a través de las teleconexiones, posiblemente de naturaleza asincrónica. Dentro de los estudios relacionados a las teleconexiones globales entre ENOS y los caudales, se hallaron impactos de flujo fuertes y regionalmente consistentes en: Australia y Nueva Zelanda, América del Sur y Central, y señales más débiles en algunas partes de África y América del Norte (Morán et al., 2016).

Además, se ha propuesto la hipótesis de que la relación lluvia-ENOS es menos fuerte que caudal-ENOS (de modo que un efecto de amplificación toma lugar), debido a las propiedades integradoras del proceso de flujo fluvial (Kundzewicz et al., 2019). Está claro que el comportamiento de descarga en una serie de tiempo en un río puede diferir de una temporada de recuperación intensa. Esto ocurre porque los grandes ríos tienen cuencas que ocupan muchas hectáreas y existe una transformación de la lluvia en escorrentía fluvial, lo que se traduce como un retraso en el transporte, sin olvidar la influencia humana. Existe una gran cantidad de investigaciones destinadas a conectar las inundaciones con la variabilidad climática a escala continental. Se han realizado muchos estudios relevantes en: Asia, América del Norte, América del Sur, Australia y Europa, utilizando índices como: ENOS, NAO, AMO, TSA, TNA, AO, PDO entre otros (Giddings & Soto, 2006).

Un estudio del impacto del índice ENOS a nivel global asocia la producción agrícola del cereal con las condiciones climáticas del mismo. Lo que a su vez se traduce en implicaciones socioeconómicas a nivel mundial. Una de estas investigaciones se realizó en el periodo 1962 - 2009 considerando la producción anual de cultivos a nivel de los países en toda la región del Pacífico Occidental y en las regiones costeras occidentales de América Central y del Sur. Los resultados indicaron un efecto positivo de ENOS sobre la producción. Esto se debe principalmente a sus características de precipitación y temperatura idóneas para el suelo de cultivo. También se pudo observar una relación negativa directa con la producción y los eventos El Niño y La Niña. (Abdolrahimi, 2016).



Se realizó una investigación usando el índice AMO para medir su influencia en la productividad agrícola regional de los estados costeros de México y varios de los cultivos más importantes de cada estado. El análisis realizado a escala anual y decadal mostró la influencia AMO sobre las variables de temperatura y lluvia, en el periodo 1980 - 2015. La temperatura mínima es la variable que presenta mayor correlación significativa, considerando los distintos cultivos de cada estado. Además, se pudo conocer cómo afectará en los años posteriores la producción total, y en el mismo estudio se ha sugerido que con los índices PDO y NAO se realice análisis conjuntos en el área de silvicultura y agricultura ya que se ve una gran utilidad económica en el tema (Cuevas et al., 2019).

Se realizó también una investigación sobre la variabilidad interanual de las vías de contaminación desde los continentes del hemisferio norte hasta el Ártico. El estudio se llevó a cabo mediante un modelo de simulación de 15 años de la dispersión de trazadores pasivos representativos de las emisiones antropogénicas (campos de viento) y el mismo demostró que el índice NAO tiene un fuerte control en el transporte de la contaminación, particularmente en invierno y primavera. Se determinó que Europa es el continente más influyente en los niveles de la contaminación. Esto explicaría por qué el cambio climático en el Ártico fue probablemente más rápido que en cualquier otro lugar del mundo (aerosoles y las nubes delgadas calientan la superficie al aumentar la emisividad de onda larga) (Eckhardt et al., 2003).

En Australia se realizó un estudio que vinculaba las precipitaciones directamente controladas por los índices climáticos con la estructura y productividad del paisaje, la cual representa 25% del continente. La investigación se desarrolló en el periodo 1900 - 2010 tomando 16 lugares distintos a lo largo del Transecto tropical del norte de Australia. Se encontró que todos los índices dentro del proceso atmosférico de ENOS tienen una correlación positiva con la precipitación y la conservación de la sabana en todo el país. También se observó la baja o nula correlación de los Índices SOI y El Niño para todas las estaciones. Aunque en la investigación también se menciona la gran falta de estudios en el tema de los índices climáticos analizando los efectos en un largo y corto plazo y se centra en únicamente considerar otros factores como incendios y prácticas antropogénicas para predecir el estado de los ecosistemas (Rogers & Beringer, 2017).

Aunque los estudios se limitan en espacio solo a cuencas mínimamente alteradas, para centrarse en los cambios impulsados por el clima en lugar de los cambios debido a alteraciones de la cuenca.



Se ha demostrado que los cambios temporales en la ocurrencia de inundaciones importantes están dominados por la variabilidad del índice AMO, más que por las tendencias a largo plazo. También se plantea que las fases particulares de la variabilidad climática pueden causar riesgo de inundaciones con posibilidades mayores al promedio, esta correlación puede variar en el tiempo, incluso invirtiendo su naturaleza de influencia. A esto se le conoce como comportamiento de época, y está asociado con la teoría de que el cambio climático está modificando los sistemas de teleconexiones (Kundzewicz et al., 2019).

1.5 Influencia de los índices de teleconexión en los desastres naturales

Con el propósito de describir la correlación entre los índices de teleconexión y su influencia en los eventos extremos alrededor de todo el planeta, se han identificado seis regiones que representan en forma general la dinámica atmosférica de su entorno (Stephoe et al., 2018).

1.5.1 Australia

Las tormentas que se producen en esta región están influenciadas en gran medida por los ciclones tropicales que se forman en los océanos Índico Sur o Pacífico Sur. Estos a su vez están afectados por los valores de TSM más cálidos que resultan en un aumento de la intensidad de la superficie. El índice SOI tiene una alta capacidad predictiva en escala mensual. Se determinó que las ocurrencias son más frecuentes en los años de La Niña. Los impulsores de las tormentas de viento son las conexiones atmosféricas de ENOS/SOI y AAO, por lo que se recomienda estudiar la estacionariedad de estos índices (Dowdy et al., 2013).

Las lluvias extremas que suceden al norte de Australia están fuertemente influenciadas por ENOS, IOD (Dipolo del Océano Índico) y la Niña. AAO tiene una correlación ampliamente positiva con las precipitaciones mensuales de la mayoría de las estaciones del país (Stephoe et al., 2018). NP tiene una correlación significativa con las sequías de la zona, a su vez se relaciona con la TSM que impulsa la circulación anómala de Walker que suprime la convección y reduce la lluvia (Song et al., 2017). La actividad de los incendios australianos tiene una correlación significativa con AAO, cuando este se encuentra en su fase positiva (reducción de la precipitación). Son similares los índices: ENOS, La Niña, SOI y la TSM (Mariani & Fletcher, 2016). Para IOD este se relaciona de



forma positiva con la incidencia de incendios, contribuye a la reducción de las precipitaciones y al aumento de las temperaturas en los meses anteriores a la temporada de incendios del verano austral (White et al., 2014).

1.5.2 Brasil

La relación entre las inundaciones y el índice ENOS muestra la influencia en la variabilidad espacial en Brasil. ENOS está relacionado con sequías en la Amazonia y noreste de Brasil e inundaciones al Sureste (Cavalcanti, 2012). También se han identificado vínculos importantes de La Niña con inundaciones en los años 1988 - 1998 y en el periodo 1995 - 1996 (Satyamurty et al., 2013). AAO al vincularse con la Zona de Convergencia Sudamericana influye en las áreas anómalas de precipitación intensificándolas (Thompson & Wallace, 2000).

1.5.2 China

ENOS positivo disminuye la ocurrencia e intensidad de lluvias en la primavera en el sureste de la cuenca del río Yangtze. Para las lluvias de otoño, el ENOS negativo del mismo año tiende a aumentar la ocurrencia de lluvias en la parte sur del mismo sitio (Xiao et al., 2015). El Niño está asociado con un aumento de precipitaciones al Sur de China durante el verano. La AO mantiene interacciones indirectas a través de las influencias en la convección local (Qu et al., 2017). AAO tiene una correlación positiva con las precipitaciones de China durante el verano. La NAO y la NP muestran correlaciones positivas en direcciones opuestas, el vínculo no es estacionario y sea incrementado su intensidad a través de los años (Wu et al., 2009). IOD presenta correlación significativa positiva con las precipitaciones al sur de China aunque sus efectos aún se consideran lejanos (Qiu et al., 2015).

1.5.3 Pacífico oriental

El Niño aumenta la frecuencia e intensidad de los ciclones tropicales que ocurren en el Pacífico Oriental, se cree que están asociados con la reducción de cizalladura del viento y con un retraso de calor sustancial generado por ENOS que alcanza su punto máximo durante la temporada de estos eventos (Chu & Wang, 1997). PNA provee una gran capacidad predecible al reflejar cambios en la temperatura local y el calor de la TSM (Caron et al., 2015).

1.5.4 Europa



NAO tiene una correlación con la ocurrencia de tormentas de viento explosivas y profundas, existe una influencia entre el movimiento del chorro polar en la trayectoria de las tormentas. También se intensifican la ocurrencia de ciclones (Pinto et al., 2009). El índice AO mantiene un comportamiento similar a NAO. SCP (el patrón escandinavo) en su fase positiva explica un bloqueo al norte de Europa y una desaceleración de los ciclones (debilitación y desvío) (Zubiate et al., 2017).

Por otra parte en lo que respecta a precipitaciones la NAO mantiene un vínculo positivo con las precipitaciones extremas en invierno al centro y norte de Europa, y un vínculo negativo al sur (Willems, 2013). SCP tiene una correlación negativa que provoca una acumulación de aire frío y disminución de las precipitaciones al noreste de Europa (Lavers et al., 2012). IOD en su fase positiva se relaciona a precipitaciones reducidas en todo el país (Saji & Yamagata, 2003).

1.5.5 México

ENOS tiene una influencia muy significativa en el comportamiento de la precipitación en México, este índice es más intenso en regiones de convección climatológicamente fuerte, esto se comprueba al ver la ubicación de la convección estacional de Niño 3.4 (Bhattacharya & Chiang, 2014). La NP tiene una influencia positiva sobre las precipitaciones al oeste de México sin embargo no se identifica una correlación específicamente (Linkin & Nigam, 2008).

1.5.6 Atlántico norte

En los años de La Niña, la actividad de huracanes en el Atlántico aumenta, con el índice El Niño sucede de forma contraria (aplica también para ciclones tropicales) (Camargo et al., 2010). Juntos los índices AMM y La Niña en su fase positiva producen ciclones tropicales atlánticos extremadamente intensos, apoyados en gran parte por una humedad troposférica media por encima del promedio (Patricola et al., 2014). La NAO afecta significativamente la TSM del Atlántico y la presión del nivel de mar, creando una influencia significativa para génesis de los ciclones (Steptoe et al., 2018).

1.5.7. Noroeste del Pacífico

ENOS es el principal impulsor de los ciclones de esta zona, las variaciones climáticas son influenciadas por cambios atmosféricos (Niño 3.4) y no solo la TSM asociada con este índice



(Sobel et al., 2016). La trayectoria de la génesis varía la trayectoria de destino los ciclones generador por el Niño se mueven hacia el oeste y norte en comparación con La Niña, estas curvaturas a su vez están fuertemente relacionadas con SOI (Nakano et al., 2016). La NP tiene una importancia en el chorro impulsado por remolinos lo que lo hace más extenso e intenso (Tanaka et al., 2016).

1.5.8 Sudáfrica

Los eventos ocasionados por ENOS positivo están relacionados con la alta presión al Sur de África provocando un descenso anómalo en la tropósfera media, divergencia en la humedad y precipitación más baja del promedio (Crétat et al., 2012). IOD complementa o interrumpe los efectos de ENOS en la precipitación, cuando ocurren años La Niña las lluvias aumentan, con el Niño ocurre lo contrario (Hoell et al., 2017). Durante enero-marzo el índice AAO muestra una correlación positiva débil en gran parte de Sudáfrica (Gillett et al., 2006).

1.5.9 Sur de Asia

ENOS reduce la precipitación en la mayoría de las áreas asociada con el monzón de verano del sur de Asia (SASM) (Cash et al., 2003). Las nevadas euroasiáticas extremas en invierno están asociadas con la fase negativa de AO, las cuales inducen una serie de ondas meridionales sobre el Atlántico norte tropical, lo que provoca un enfriamiento en esta región, aunque esto no influye en la precipitación (Prabhu et al., 2017). AAO por otra parte durante febrero-marzo da lugar a una circulación meridional anómala que sucede en el Pacífico central y provoca un enfriamiento en la misa región, debilitando las lluvias monzónicas sobre el subcontinente indio (Gillett et al., 2006).

1.5.10 Sudeste Asiático

ENOS tiene una relación negativa significativa con las lluvias monzónicas de verano sobre el sur este asiático (Räsänen et al., 2016). El Niño con origen en el Pacífico oriental ha influenciado en la intensificación de la circulación de Walker sobre Tailandia e Indonesia, lo que reduce significativamente la convección y por tanto la precipitación. Existe una tendencia de sequías y una inversión de ENOS en Filipinas con eventos extremos en sus fases positivas y negativas (Cruz et al., 2013).

1.5.11 Estados Unidos



PNA en fase positiva se asocia con una frecuencia mayor de tormentas en las latitudes medias de los EE.UU., además esta teleconexión es la impulsora más importante de las inundaciones en el centro de este país. Esto sucede porque en la costa oeste el flujo de energía de las olas superiores (una medida de los niveles de agua) tiene una correlación positiva con PNA, la cual está muy relacionada con NAO (Mallakpour & Villarini, 2016). Durante El Niño el agua del Océano Pacífico oriental se convierte en una fuente de humedad precipitable sobre el suroeste de los Estados Unidos y causa que los caudales aumenten debido al fortalecimiento del chorro subtropical durante el invierno (Mo & Higgins, 1998).

La Niña produce condiciones extremadamente secas al suroeste de EE.UU. y se extiende hasta las grandes llanuras. En cuanto a los incendios ENOS tiene una fuerte correlación con los incendios forestales de esta zona referente al calentamiento de la TSM del Pacífico Norte y el enfriamiento de la misma en el Pacífico tropical Oriental. La relación ocurre por la forma en la que el índice ENOS modula la humedad del suelo (Stephoe et al., 2018).

1.6 Modelación para predictibilidad de caudales

Los modelos hidrológicos son esenciales para comprender el comportamiento de los ciclos hidrológicos en regiones de importancia por su vulnerabilidad y además para aportar información valiosa que ayude a distribuir los recursos hídricos de una manera óptima. Se puede pronosticar eventos extremos como sequías, inundaciones y efectos del cambio climático. En la actualidad el pronóstico hidrológico se apoya especialmente en sistemas de información geográfica (Madsen et al., 2020).

Dentro de los tipos de modelación para cuencas estos se clasifican espacialmente en: distribuidos y agrupados, se entienden como agrupados si se asume que todas sus características son constantes y la cuenca es homogénea (propiedades del suelo y flujo). Por otro lado, la modelación distribuida divide en grillas a la región de estudio y realiza un balance hídrico con las propiedades de cada grilla representando mejor los efectos del entorno (Cabrera, 2012).



Si la clasificación depende de la conceptualización de los procesos básicos, entonces el modelo puede ser empírico o de regresión, y su funcionamiento se describe como de parámetro, si la relación está calibrada entre entradas y salidas. El modelo se conoce como basado en procesos si se conocen todas las dinámicas involucradas en la transformación, y para explicar su funcionamiento se presentan ecuaciones físicas asociadas a todos los procesos internos que se pretenden simular y se anula el proceso de calibración. Los modelos son conceptuales si se conocen algunos de los procesos internos del modelo y para entender los restantes se usan leyes físicas en forma simplificada (Cabrera, 2012)

De acuerdo a la naturaleza de los algoritmos básicos el modelo puede ser determinístico si todas las variables incluidas se consideran libres de la variación aleatoria o incertidumbre, es decir que ninguna tiene una distribución probabilística. Los modelos también pueden ser estocásticos (Cabrera, 2012), como en nuestro caso de estudio. La definición de este tipo de modelamiento se basa en un grupo de procedimientos para calcular cantidades de interés hidrológico a través del análisis de un registro de otra variable hidrológica (precipitación, caudales, niveles de lagos, entre otros), y estas observaciones se consideran aleatorias debido a la misma naturaleza de la variable, es decir una variable sujeta a leyes probabilísticas. Esta definición se extiende a las relaciones entre una o más variables aleatorias y a incluir una dimensión espacial en la que las series de tiempo registradas pertenecen a dos puntos separados. Por lo que los cálculos a partir de estos modelos están sujetos a incertidumbre (Clarke, 2002).

Dentro de la modelación hidrológica se describen etapas para integrar todo el proceso:

Primero se propone definir los objetivos de la modelación, en esta etapa se determinan los objetivos, alcance y nivel de profundidad del estudio. Se deben considerar los requerimientos, información necesaria, y recursos tecnológicos y virtuales disponibles. Como segunda etapa se deben recopilar los datos e información para la modelación, se identifican el tipo y cantidad de datos disponibles en la zona del estudio. Además, se determina que tan representativos son en términos de escala y resolución. También se consideran diferentes métodos para el relleno de datos. Es importante que si se trata con modelos estadísticos las series de tiempo sean largas, cualquiera que sea la variable de interés. La tercera etapa se realiza conjuntamente, definiendo las escalas de



longitud espacial y tiempo de observación, se considera el principio de correspondencia espacio-temporal (eventos de pequeña extensión son intensos, y viceversa) (Vargas & Giraldo, 2018).

La cuarta etapa es un análisis de la calidad de datos e información, se identifica el nivel de error e incertidumbre y se aplican medidas para disminuir los sesgos. Es un gran condicionante en la creación de los modelos. La quinta etapa es describir de forma más específica dentro de las series de tiempo donde ocurren, se conoce el comportamiento general de los datos. La sexta etapa es construir un modelo conceptual, definiendo el área de estudio y los fenómenos que ocurren en el entorno de importancia hidrológica. La séptima etapa se enfoca en la selección del modelo se realiza considerando el tiempo de observación de la variable y la disponibilidad de información, además de la forma de procesamiento del modelo (rango de aplicabilidad, parámetros y datos requeridos), a este paso se integra la octava etapa la cual consiste en realizar un análisis de sensibilidad para lograr una mejor calibración (Vargas & Giraldo, 2018).

La novena etapa es seleccionar criterios de desempeño, o métricas estadísticas que ayuden a evaluar el funcionamiento del modelo aplicado a los distintos escenarios hidrológicos. Antes de crear el modelo se divide el conjunto de datos en dos etapas: calibración y validación, esto sirve para probar la legitimidad del modelo aplicado a distintos valores y condiciones. Por último, también se recomienda que después de creado el modelo y presentados los resultados de relevancia, este se mantenga actualizándose constantemente (optimización del modelo) ya que las variables del entorno varían a través del tiempo (Vargas & Giraldo, 2018).



CAPITULO 2: MATERIALES Y MÉTODOS

2.1 Descripción del área de estudio

La cuenca del río Paute tiene una superficie aproximada de 6.437 km² hasta su desembocadura en el río Upano, y se localiza en la cadena montañosa de los Andes, con pendientes que se encuentran en el rango del 25 al 50 % (CELEC EP, 2013). Esta cuenca se encuentra ubicada en cuatro provincias: Azuay, Cañar, Chimborazo y Morona Santiago. En un plano más amplio es parte de la cuenca del Río Santiago.

El clima de la cuenca es muy heterogéneo, en las zonas donde alcanza grandes alturas tiene un clima ecuatorial frío de alta montaña, también se caracteriza por tener un clima frío de los páramos, en general su climatología es templada y subtropical en las hoyas y plenamente tropical en las llanuras de nivel más bajo. Todo esto se debe a que la cuenca en si atraviesa ubicaciones con orografía muy diversa, como ya se mencionó anteriormente las altitudes que considera en su trayecto son desde 500 msnm (zona amazónica) a 4.600 msnm (zona interandina).

En cuanto al uso de suelos y cobertura vegetal actual de la cuenca del río Paute, esta se divide en: pasto y vegetación leñosa (hierba/bosque) en un área de 0,43 ha, pastos y cultivos esencialmente de maíz (paisaje mixto de pastos y cultivos con parcelas compartidas) en un área de 366,26 ha, pastizal natural (pastizal de monte, casas y cultivos) ocupan 1.538,95 ha, formación herbácea perenne (paramo) se ubica extendida en 373,18 ha, formación arbustiva arbórea (matorral) en un espacio de 985,69 ha y finalmente formación arbórea (monte y selva) ocupa la mayor cantidad de espacio con 11256,17 ha (CELEC EP, 2013).

La precipitación media anual máxima se encuentra entre 2.500 – 3.000 mm en el extremo oriental. En la cordillera Occidental en cambio los valores oscilan entre 1.200 - 1.500 mm. La precipitación tiene tres regímenes conocidos de precipitación: uno trimodal que es el sistema presente en toda la cuenca, un régimen bimodal que se halla habitualmente en los valles, y un régimen unimodal que afecta las laderas de la cordillera oriental. El régimen trimodal de alta montaña presenta mayor cantidades de precipitación anual, su precipitación media mensual es: 160 mm/m², el régimen bimodal alto interandino tiene 140 mm/m² y el unimodal 350 mm/m² (Instituto de Estudios de Régimen et al., 2017).

Se considera que existen dos periodos de precipitación frecuentes en la zona explicados por el ZCIT, de junio a agosto es la época húmeda para el régimen unimodal, de marzo a mayo, para el régimen bimodal, las estaciones secas corresponden de forma viceversa a ambos regímenes, y el régimen trimodal es una mezcla entre los anteriores (Campozano et al., 2016). Las subcuencas del río Paute y su ubicación en el Ecuador se pueden observar en la Figura 1.

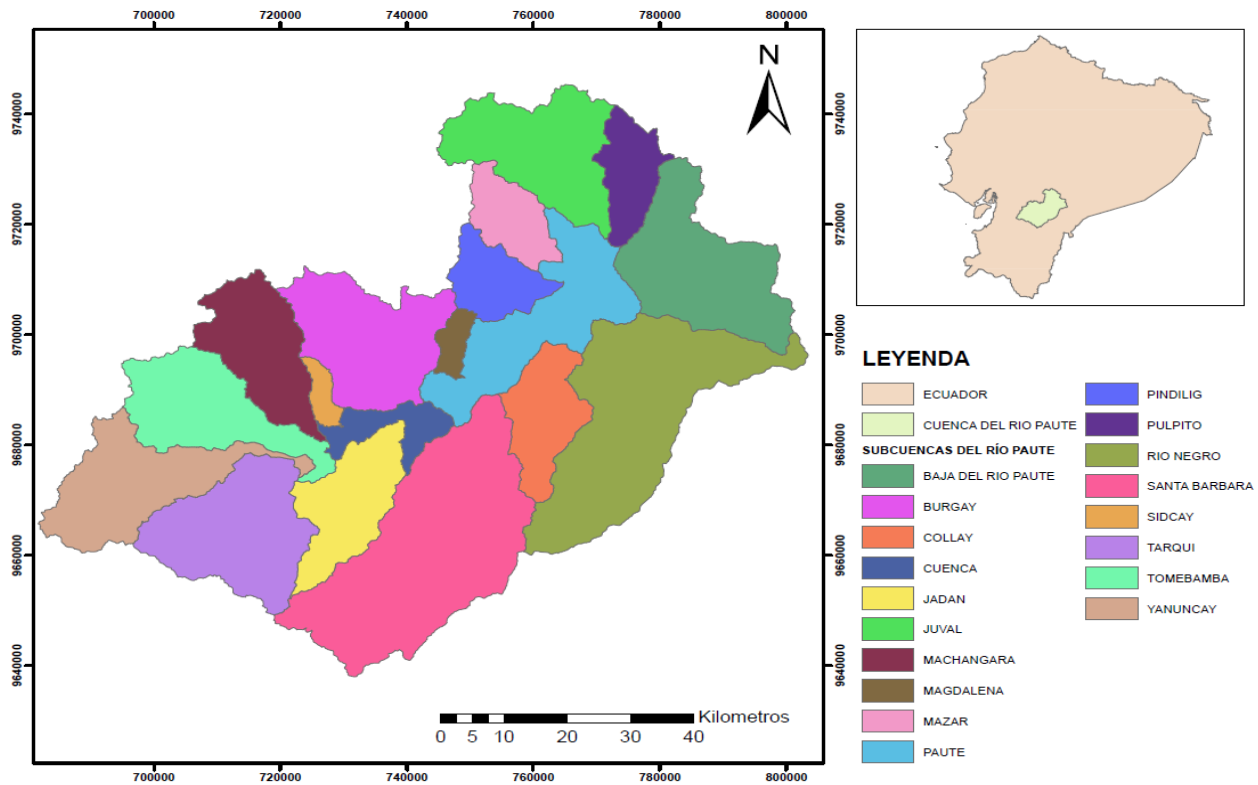


Figura 1. Ubicación del sitio de estudio, así como la ubicación de las diferentes subcuencas que conforman la cuenca del río Paute.

La red de drenaje superficial de la cuenca está formada por los ríos: Tarqui, Yanuncay, Machángara, Matadero, Burgay, Jadán y Gualaceo, entre otros. De esta manera, la cuenca del Paute consta de 18 subcuencas hidrográficas (que cubren una superficie aproximada de 6.437 km²). Las cuencas del río Negro y la cuenca Baja del río Paute son las que alimentan a la estación Molinos; el resto de cuencas aportan al embalse de Amaluzá (CELEC EP, 2013). En la Figura 2 se pueden observar las respectivas ubicaciones de estaciones de caudal y precipitación pertenecientes al *Instituto Nacional de Meteorología e Hidrología* (INAMHI) a lo largo de la cuenca del río Paute.

El caudal de la cuenca del río Paute mediante un proceso de caída alimenta a las centrales hidroeléctricas con un valor de caudal de: Mazar ($141,10 \text{ m}^3/\text{s}$), Molino ($200 \text{ m}^3/\text{s}$), Sopladora ($150 \text{ m}^3/\text{s}$) y Cardenillo ($180 \text{ m}^3/\text{s}$) (Orbes & Peralta, 2017). La cantidad de energía producida por cada una de las centrales mencionadas se detalla a continuación: Amaluza (1075 MW), El Labrado y Chanlud (38,4MW), Mazar (162,6 MW) y Sopladora (500 MW). Se calcula que en total produce un 40% de la producción hidroeléctrica a nivel del país (Instituto de Estudios de Régimen et al., 2017).

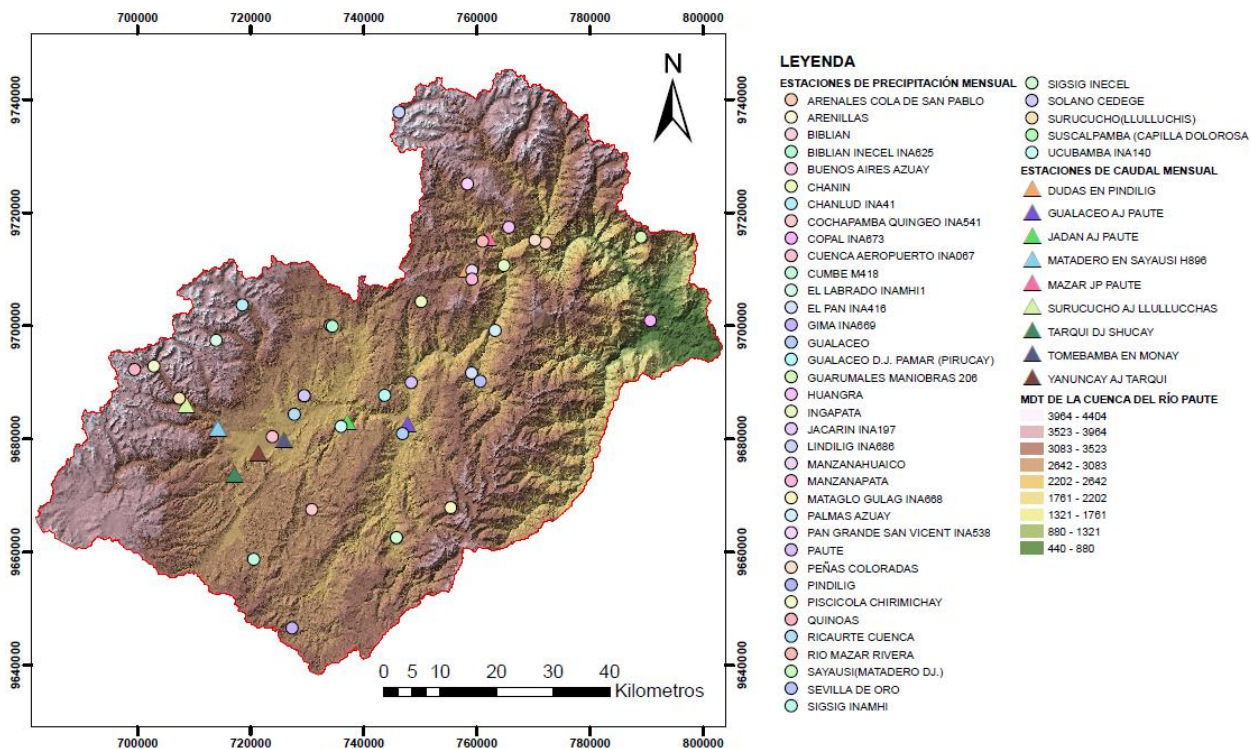


Figura 2. Ubicación de estaciones de caudal y precipitación en la cuenca del río Paute.

2.3 Recolección de datos

2.3.1 Recolección de datos hidrometeorológicos

Los datos recolectados corresponden a estaciones distribuidas en toda la cuenca del río Paute (Figura 2). Los mismos pertenecen al INAMHI y se consideraron los datos de estaciones de caudal de escala diaria y mensual en el periodo de 1995 – 2015, es decir 20 años de registros. Para este



estudio se consideraron únicamente las estaciones que contuvieran al menos un 40% de datos, para evitar que la ausencia de los mismos interfiriera con los resultados reales de las predicciones de caudal. De acuerdo a los estudios realizados este porcentaje es aceptable estadísticamente para un relleno de datos (McNeish, 2017). De las 27 estaciones de caudal disponibles, solo nueve cumplían con los requisitos iniciales, las cuales se detallan en la Tabla 2. Con un total de 252 datos para cada estación.

Tabla 2. Estaciones de Caudal usadas para el estudio.

NOMBRE	ID ESTACIÓN	PERIODO		ESCALA	% DE VALORES FALTANTES
		FECHA INICIO	FECHA FINAL		
PAUTE AJ DUDAS	H900	1/8/1963	16/3/2009	Diario	53
GUALACEO AJ PAUTE	H0931	1/1/1982	31/12/2015	Diario	60
TOMBAMBA EN MONAY	INA895	16/1/1982	16/12/2008	Mensual	38
PAUTE EN PAUTE (DJ GUALACEO)	H894	1/1/1973	16/12/2015	Diario	35
SURUCUCHO AJ LLULLUCCHAS	INA897	16/1/1982	16/12/2015	Mensual	18
MATADERO EN SAYAUSI	INA896	16/1/1971	16/12/2015	Mensual	14
COLLAY AJ PAUTE	H929	21/1/1985	16/12/2015	Diario	50
DUDAS EN PINDILIG	H902_R	1/1/1965	16/12/2015	Diario	40
MAZAR JP PAUTE	H0892	1/1/1972	31/12/2015	Diario	59

Elaborado por: María Daniela González

2.3.2 Recolección de índices climáticos atmosféricos y oceánicos

Los datos para las series de tiempo de los índices de teleconexión y modos de variabilidad están disponibles en la base de datos: *National Oceanic and Atmospheric Administration* (NOAA). Se puede acceder mediante: <https://psl.noaa.gov/data/climateindices/list/>. Las señales utilizadas están enlistadas en el sitio principal de acceso abierto, estos índices están descritos en la Tabla 1, su funcionamiento, zona de influencia y siglas.



2.4 Relleno de datos

2.4.1 Relleno de datos de estaciones de caudal

Se realizó un relleno de datos para el cual, mediante el software Arcmap 10.5 se identificaron las 10 estaciones de caudal y precipitación más cercanas a cada una de las estaciones seleccionadas para el estudio. Se procedió a obtener el coeficiente de correlación de Pearson entre las estaciones cercanas, en el software R Studio 1.4.1106. La ecuación viene dada por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (1)$$

Donde x_i y y_i son los valores de las variables x y y para cada uno de los valores individuales que toman las mismas (i -ésimo) (Mukaka, 2012). Los coeficientes de correlación sirven para medir la asociación lineal bidireccional que existe entre dos variables aleatorias continuas. Su valor puede estar entre uno negativo hasta uno positivo (uno positivo es una correlación positiva perfecta) (Mukaka, 2012). El coeficiente de Pearson se utiliza para variables con una distribución normal, puede verse afectado por valores extremos que pueden influenciar positiva o negativamente en la fuerza de la relación (Restrepo & González, 2007).

Posteriormente se evaluaron las combinaciones de variables con los resultados más altos y se realizaron modelos de regresión lineal múltiple (MLR) para rellenar los datos faltantes en las series de tiempo. Ya que diversos estudios sugieren que la regresión múltiple arroja óptimos resultados (Tencaliec et al., 2015). Como las estaciones de precipitación seleccionadas por su cercanía a las zonas de estudio no se encontraban completas, en caso de obtener aun datos faltantes después del primer relleno, se obtuvieron promedios de cada mes durante toda la serie de tiempo y se procedió a completar los valores.

2.4.1 Relleno de datos de índices climáticos atmosféricos y oceánicos

Además, se realizó un relleno de los valores faltantes para los índices, usando el promedio de los meses a través de los 20 años de las series de tiempo. A partir de esto se utilizaron los índices con sus valores originales sin alteración para realizar las predicciones. Estos se dividieron en dos grupos, los índices sin rellenar (31), y los índices con relleno (41), buscando comprobar cuál es la



influencia del relleno de datos en el estudio (Uvo et al., 2021). Los índices que se retiraron sin utilizar relleno para el escenario de 31 índices fueron: EP/NP, GIAM, NAO (Jones), NBRA, NOI, TP SST EOF, SW MRR, TPI(IPO), AT SST EOF y CIP. Cada una de las modelaciones se usó en ambos grupos.

2.5 Análisis exploratorio gráfico

Considerando que una serie de tiempo de datos hidrológicos está formada por una larga secuencia de observaciones registradas en un determinado lugar, se sabe que mientras más largo es el periodo de observación más probabilidad de error existe. Es difícil identificar esto observando una tabla con valores, por lo que para verificar inconsistencias (errores sistemáticos, saltos y tendencias) se realizó un análisis exploratorio de los datos. El análisis exploratorio gráfico de los datos iniciales sirve para observar el comportamiento de los mismos, comprobando tendencias positivas o negativas o cambios en las series temporales de forma visual. Es el primer análisis que se aplica antes de otros procedimientos cuantitativos y antes de utilizar la información climatológica oceánica y atmosférica para modelar o simular (Castro & Carvajal, 2010).

2.5.1 Análisis gráfico de series de tiempo

Estas gráficas representan los datos ordenados cronológicamente a través del tiempo (datos en ordenadas y secuencia temporal en las abscisas) como se muestra en la Figura 3, es fácil identificar tendencias, cambios bruscos, intermitencias entre otros. Permite realizar un análisis rápido del comportamiento y consistencia de la información a utilizarse e identificar valores muy altos o muy bajos en relación con el promedio de la serie, lo que puede indicar errores sistemáticos (Castro & Carvajal, 2010).

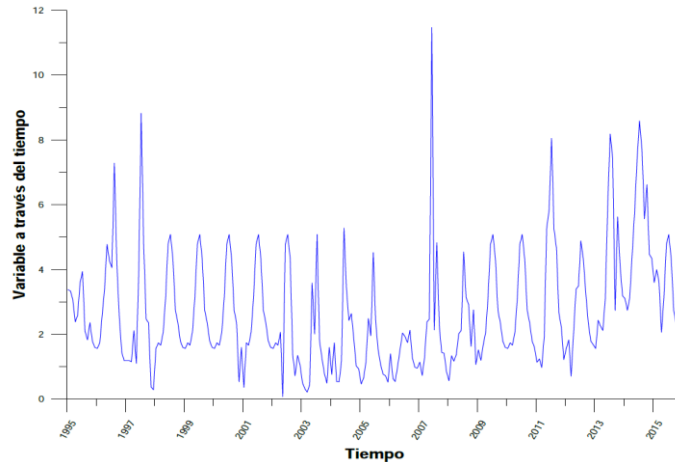


Figura 3. Esquema general de una gráfica de serie de tiempo y sus componentes.

2.5.2 Análisis diagrama de dispersión

Este gráfico (Figura 4) permite identificar de forma visual que tan dependientes o independientes estadísticamente son dos variables una de la otra. El gráfico consiste en un sistema de ejes en el que las ordenadas y las abscisas son dos variables relacionadas entre sí, sus valores se representan como puntos y se puede observar cuanta información acerca de la distribución de cada variable aportan las mismas. Se mide la relación si es positiva o negativa, también se puede integrar al gráfico una línea de 45 grados o línea de referencia para observar el comportamiento de la relación entre las variables (Herrera, 2018).

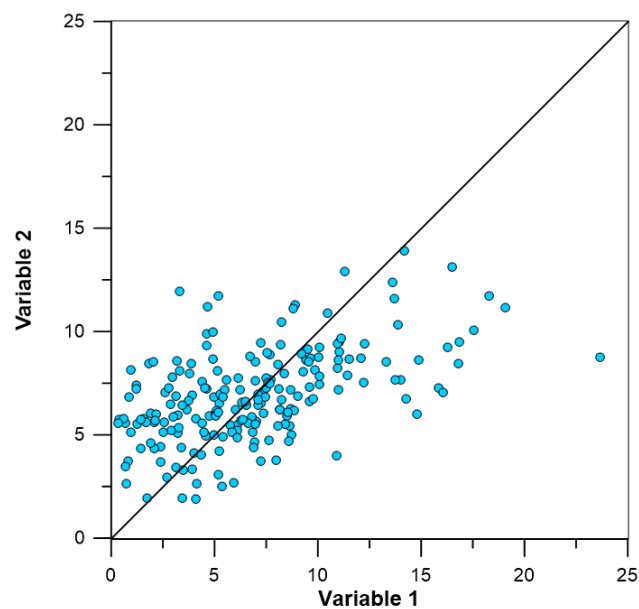




Figura 4. Esquema general de un diagrama de dispersión y sus componentes.

2.6 Prueba de Homogeneidad de datos

Cualquier serie de tiempo que involucre elementos ambientales puede estar sujeta a homogeneidad o heterogeneidad, es decir cambios aleatorios no naturales. Estos cambios pueden suceder por la variabilidad entre periodos de tiempo. Así también pueden estar relacionadas con modificaciones o errores en la instrumentación, preparación de equipos o procesamiento de datos, pudiendo aportar un error a las investigaciones (Gudmundsson et al., 2017).

Por esto se realiza una prueba de homogeneidad que demuestra la confianza que se puede obtener de los datos para identificar las influencias no climáticas dentro de la serie de tiempo. El concepto determina que los patrones climáticos y las observaciones de una determinada región serán similares. Estas también reflejarán un mismo patrón ya que por el espacio donde se ubican deben estar altamente correlacionadas. En recursos hídricos especialmente, se aplica este principio como la base del análisis de la frecuencia regional (Guttman, 1998). La prueba de homogeneidad que se usó en el estudio es la curva de masa.

2.6.1 Curva de masa

Es una gráfica del caudal acumulado (Figura 5), por tanto, siempre es creciente. La pendiente representa los valores de cual a través del tiempo (altura de los caudales acumulados por unidad de tiempo). Este método considera que la proporcionalidad de los datos en las estaciones se puede representar gráficamente. La representación consiste en determinar las estaciones a monitorear y obtener los valores anuales o mensuales acumulados. En un sistema de ejes ortogonales se grafican los valores de las estaciones en las ordenadas y en las abscisas ocurre la serie de tiempo. Dependiendo de si todas las estaciones se alinean con una pendiente única y uniforme se mide la heterogeneidad de los datos (Quispe et al., 2021).

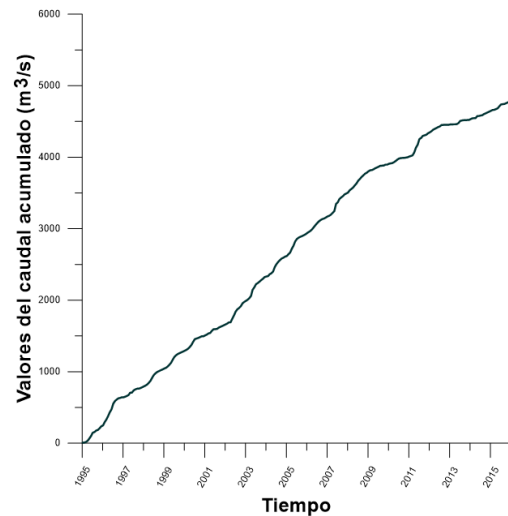


Figura 5. Esquema general de un gráfico de curva de masa.

2.7 Calibración y Validación de datos

La calibración de un modelo hidrológico es el proceso por el cual se ajusta al modelo para que las predicciones representen las condiciones reales del sitio donde se aplican. Los parámetros del modelo se optimizan para aumentar la precisión del modelo. La validación sirve para demostrar que el modelo calibrado funciona de forma precisa bajo otras condiciones de modelado (Daggupati et al., 2015). Para este estudio se consideró un 80% de los datos para la etapa de calibración y el 20% restante se usó en la etapa de validación. Los datos de enero del año 1995 hasta septiembre del 2011 se usaron directamente para la calibración y desde septiembre del 2011 a diciembre del 2015 se usaron para la validación.

2.8 Modelación para predicción de caudales con series temporales de los índices climáticos atmosféricos y oceánicos

2.8.1 Modelación usando el Análisis de Componentes Principales (ACP)

Luego de realizar los análisis preliminares para explorar el comportamiento general de los datos seleccionados para el estudio de predicciones se realizó la técnica de componentes principales con la finalidad de reducir la dimensionalidad de los múltiples índices climáticos usados.



El método de los componentes principales es una importante técnica en el análisis de datos que sirve para reducir dimensionalidad de un gran conjunto de datos con variables interrelacionadas entre sí, sin embargo, esta reducción debe conservar la variación de los datos para no perder su representatividad. Algebraicamente transforma al conjunto de datos en variables no correlacionadas generando nuevos sistemas de coordenadas para maximizar la varianza o S^2 (la carga de cada componente a la varianza acumulada que explica el comportamiento de los datos se reduce). El primer componente principal indica la dirección espacial de la mayoría de la varianza de los datos. Cada componente principal tiene su propio vector que explica la varianza (Shabri & Shuhaida, 2014).

Suponiendo que los valores de p -variables en n -elementos dentro de una población se disponen en una matriz X de dimensiones $n \times p$, las (columnas p -variables, filas n -elementos). Dentro de esta matriz a cada variable se le resta el valor de la media, por lo que la media es 0 y su matriz de covarianza se presenta como $1/n X'X$.

Espacialmente esto puede ejemplificarse como un conjunto de puntos de los que se obtiene una recta que pretende pasar cerca de los mismos y que la distancia entre los puntos y la recta se mantenga. Al considerar un punto x_i y una dirección $a_1 = (a_{11}, \dots, a_{1p})'$, que está basada en un vector a_1 de norma unidad, la proyección del punto x_i encima de esta dirección se traduce en un escalar (Peña, 2002).

El análisis de componentes principales también funciona como un método para determinar un patrón en los datos. La ecuación 2 representa el escalar y muestra la forma inicial del ACP.

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (2)$$

Donde Z_i representa a los componentes principales, a_i es un valor relacionado con los eigen vectores de la transformación espacial y X_i son las variables de entrada. La ecuación 2 se obtiene de la ecuación 3 al resolver:

$$|R - I\lambda| = 0 \quad (3)$$



Donde I es la matriz unitaria, R la matriz de varianza-covarianza y λ es el eigen valor. Mediante los eigen valores podemos obtener todos los eigen vectores (Dehghan et al., 2014).

La representación de puntos- p dimensionales con una pérdida de información ínfima es equivalente a sustituir las variables- p por nuevas variables z_1 que compendian la información de forma más optima. Para esto se toma la variable que mantenga la mayor variabilidad (máxima correlación con las variables originales). La variable escalar determinada al proyectar los puntos sobre la recta funciona para prever el conjunto de datos originales. Esta recta minimiza las distancias ortogonales entre los puntos y la recta y se encuentra entre las rectas de regresión de los puntos inicial y final.

Al referirnos a los datos de dimensionalidad dos, se calcula el plano que mejor se aproxima a los puntos. La nueva dirección del vector unitario a_2 puede tomarse de forma ortogonal en referencia de a_1 , y únicamente debe verificar que la proyección de un punto sobre este nuevo eje siga maximizando las distancias entre los puntos iniciales igualmente proyectados, esto representa la segunda variable z_2 , no correlacionada con z_1 y con la varianza máxima entre todas las combinaciones posibles lineales de las p -variables originales e la matriz X . Por tanto, el primer componente principal esta dado como: $z_1 = Xa_1$ (Peña, 2002).

Para este procedimiento se usó el Software R, el paquete stats 3.6.2, la función “prcomp”, cuya función es realizar el método de análisis de componentes principales en una matriz de entrada. En este procedimiento se centran y escalan los valores para evitar que los índices tengan distintos pesos en el análisis. La función devuelve varios elementos como resultado de la operación:

- “sdev”, las desviaciones estándar de los componentes principales, es decir las raíces cuadradas de los valores singulares de la matriz de datos.
- “rotación”, es la matriz con los eigen valores o varianza explicada de cada componente principal.
- “X”, devuelve el valor de los datos rotados, es decir que al aplicar un cálculo de covarianza a estos valores (matriz) y obtiene la diagonal con los componentes principales o eigenvectores (Rodrigo, 2017).



También se obtuvo la varianza del método para cada componente, son 31 y 41 columnas respectivamente ya que la cantidad de componentes es similar a las columnas de entrada. Se obtuvo la varianza acumulada que en total suma el 100%. Y se trabajó con los componentes principales que explicaran el 70% y 90% de varianza (Rea & Rea, 2016).

2.8.2 Modelación usando Índices climáticos atmosféricos y oceánicos y retrasos mensuales

Para descartar posibles procesos de retraso entre el inicio de la circulación atmosférica de cada índice y sus efectos en la precipitación y caudal, se establecieron tres posibles escenarios de predicción. Incluyendo uno, dos y tres meses de retraso de los índices (predictores) en comparación con los valores de caudal de cada una de las estaciones. Es decir, los valores de los índices ascendieron un mes, de febrero a enero, marzo a febrero y abril a marzo.

También se realizó un test de correlación y se usaron únicamente los cinco índices climáticos atmosféricos y oceánicos con los coeficientes más altos de correlación de Pearson en referencia a cada una de las estaciones de caudal, y se procedió a aplicar el cálculo de predicciones.

2.8.3 Modelación usando análisis de multicolinealidad

Para usar los índices con sus valores originales en la predicción se consideró el factor de inflación de la varianza (VIF) con objetivo de identificar multicolinealidad. La cual describe una dependencia lineal (correlación o similitud) entre las variables que se usan para la predicción. Y causa interferencia en el efecto que tiene cada predictor sobre la variable dependiente, ya que no se puede determinar el peso individual de cada índice sobre los caudales (Vega Vilca & Guzman, 2011). Se considero cinco como un valor adecuado de VIF para cada índice, las variables que obtuvieron valores superiores al límite establecido, fueron descartadas. El umbral se eligió debido a que valores de VIF mayores a cinco o 10 indican alta colinealidad y coeficientes de regresión asociados mal estimados (Lujano, 2019). Se calcula a partir de la ecuación 4:

$$VIF = \frac{1}{1 - R^2} = \frac{1}{Tolerancia} \quad (4)$$



Donde R^2 es el coeficiente de determinación, la tolerancia es la definición inversa de VIF, mientras menor es este valor existe más probabilidad de multicolinealidad entre las variables (Shrestha, 2020).

Considerando el siguiente modelo de regresión múltiple en la ecuación 5:

$$y = X\beta + \varepsilon \quad (5)$$

Donde y es un vector de resultados $n \times 1$, X es una matriz $n \times p$ de variables de regresión, β es un vector $p \times 1$ de constantes desconocidas y ε es un vector de errores aleatorios $n \times 1$. Consecuentemente En consecuencia, $X'X$ se entiende como una matriz de correlaciones $p \times p$ entre las variables de regresión y $X'y$ es un vector de correlación $p \times 1$ entre las variables de regresión y el resultado.

Al considerar la j -ésima columna de la matriz X como X_j de modo que si $X = [X_1, X_2, \dots, X_p]$. X_j contiene todos los n niveles de las variables regresores. La multicolinealidad se define como la dependencia lineal de las columnas de X . Los vectores son linealmente dependientes si existe un conjunto de constantes t_1, t_2, \dots, t_p , donde no todas toman el valor de 0, por lo que se deduce en la ecuación 6:

$$\sum_{j=1}^p t_j X_j = 0 \quad (6)$$

Si la ecuación 6 se cumple de forma exacta para un subconjunto de columnas en la matriz X , el rango en la matriz $X'X$ es menor que p y $(X'X)^{-1}$ es inexistente. Si esta misma ecuación no se cumple en totalidad, existe una dependencia casi lineal en $X'X$ y se observa multicolinealidad (Ranjit, 2006).

2.8.3.1 Modelos de Regresión Múltiple (MLR)

Estos modelos son una extensión del modelo de regresión lineal simple, sirven para tratar datos con múltiples variables predictoras, y obtener un mismo resultado $(x_{i1}, x_{i2}, \dots, x_{i,p-1}, y_i)$. Donde $i = 1, 2, \dots$, hasta n unidades de observación. Captura una relación estadística que se produce simultáneamente entre el único resultado a obtener (Y) y las variables predictoras X_k ($k = 1, 2, \dots$).



., $p - 1$), la relación se muestra en la ecuación 7. La equivalencia de esta misma relación se muestra en la ecuación 8, donde se incluye σ^2 la desviación estándar:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad (7)$$

$$\varepsilon_i \stackrel{indep}{\sim} N(0, \sigma^2) \quad (8)$$

donde β_0 es la intersección es decir la media de Y cuando las variables predictoras $X_k = 0$. Por su parte cada β_k (coeficiente de regresión parcial), representa una pendiente con respecto a X_k , es decir cuál es la magnitud de valor de cambio en la media de Y si X_k es mayor a la unidad (con las demás variables constantes). Por lo tanto, la equivalencia del resultado Y se define como la ecuación 9.

$$y_i \stackrel{indep}{\sim} N(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1}, \sigma^2) \quad (9)$$

Los supuestos de la MLR son:

y_i son independientes entre sí, mantienen una distribución normal, la media de esta distribución es una función lineal de cada X_{ik} ; y la varianza de esa distribución es la misma para todo y_i (varianza constante u homocedasticidad) (Le & Eberly, 2003).

Para la obtención de modelos se usó el criterio de stepwise con la técnica Forward, la cual agrega variables al modelo y mantiene únicamente las se consideran estadísticamente significativas, realizando iteraciones hasta que alcanza un resultado óptimo. Esto se realizó mediante el uso del software RStudio, y el paquete MASS. La función utilizada fue leapforward. Las métricas que se utilizaron para evaluar cada uno de los modelos generados para las estaciones de caudal (únicamente para la creación de modelos de este escenario) son parte de la función “glance” del paquete broom se extraen los valores del modelo de: r cuadrado ajustado, AIC (Criterio de información de Akaike), BIC (Criterio de Información Bayesiano), deviance (desviación), Grados de libertad residuales, logLik (log-verosimilitud del modelo), nobs (número de observaciones usadas), p.value (Valor de p correspondiente al estadístico de prueba) r cuadrado estadístico, sigma



(error estándar estimado de los residuos), estadística (test estadístico), d.f (grados de libertad del numerador) (Robinson, 2022). Los modelos elegidos fueron los que arrojaron gran significancia estadística, menor error, y mayor r cuadrado ajustado, la función “train” del paquete caret, realiza esta comparación automática con todas las variables y presenta en sus resultados el modelo más apropiado (Zevross, 2017).

2.9 Obtención de resultados de predicción de series de tiempo de caudales

Para realizar la predicción de las distintas series de tiempo de acuerdo a la variación en la selección de índices se utilizó una misma metodología. Con el paquete tidy, se usó la función “predict”, la cual usa el resultado de un modelo ajustado (función “lm” en RStudio), y devuelve predicciones para cada uno de los puntos de la serie de tiempo del modelo (Dalpiaz, 2020). El 80% de datos ya mencionados se usaron directamente para la predicción con el software R Studio. Mientras que para el 20% de los datos de validación se tomaron los coeficientes e interceptos de los modelos usados en la predicción y se realizaron regresiones lineales en Excel. Con excepción para los escenarios con retrasos en donde los meses se adecuaron a los porcentajes considerados para calibración y validación.

Para comprender mejor el comportamiento de cada escenario de modelación en los resultados predichos se realizaron gráficos de dispersión y series de tiempo ya descritos anteriormente.

2.10 Análisis estadístico de resultados mediante métricas

Para determinar la precisión de las predicciones realizadas, se usaron una serie de métricas procesadas en el software R Studio, mediante el paquete hydroGOF, con la función “gof” (goodness of fit en inglés), este análisis sirve para obtener medidas numéricas de bondad de ajuste entre valores observados y simulados, una serie de índices para comparar dos matrices de datos (Mauricio Zambrano Bigiarini, 2020). Se procedió a analizar los valores originales con cada uno de los distintos escenarios con los índices aplicados. Las métricas que se obtienen como resultado se describen en la Tabla 3.

**Tabla 3.** Métricas usadas del paquete HydroGOF.

MÉTRICA	NOMBRE	DESCRIPCIÓN	RANGO DE BUENOS RESULTADOS
KGE	Eficiencia de Kling-Gupta (Kling Gupta Efficiency)	Este índice es una descomposición del diagnóstico de Nash-Sutcliffe, y analiza varios componentes como la correlación, variabilidad y sesgo. El coeficiente de correlación usado es el de Pearson, beta o la relación entre la media de los valores simulados y observados, v_r es la relación de variabilidad (desviación estándar).	-Inf - 1 (1 valor optimo).
NSE	Eficiencia de Nash-Sutcliffe (Nash-Sutcliffe Efficiency)	Es una estadística normalizada que determina la magnitud relativa de la varianza residual (ruido) en comparación con la varianza de los datos medidos.	-Inf - 1 (1 valor optimo)
R²	Coefficiente de determinación (Coefficient of Determination)	Representa la proporción de la varianza total de la variable explicada por la regresión. Expresa la bondad del ajuste de un modelo de la variable que pretende explicar.	0 - 1
RSME	Raíz del error cuadrático medio (Root Mean Squared Error)	Sirve para medir la cantidad de error que hay entre dos conjuntos de datos. Es la raíz cuadrada del MSE, se conoce también como Desviación Cuadrática Media.	0 - +Inf (valores más bajos indican mejor resultado)

Elaborado por: María Daniela González

Las métricas usadas en el estudio fueron: P-Value, Eficiencia de Nash-Sutcliffe, Eficiencia de Kling-Gupta, Coeficiente de determinación y Error cuadrático medio de la raíz, los cuales se describen a continuación con su fórmula explicada.

2.10.1 Valor P o P-value

Es un criterio estadístico creado por Ronald Fisher en los años 20, se denomina Prueba de significación y sirve para analizar hipótesis estadísticas. La prueba de significación se basa en determinar un valor p, que es a su vez una medida de la veracidad de la evidencia de los datos en



contra de la hipótesis nula (hipótesis contraria a la que se plantea en el estudio). Cuando se obtiene en una investigación un valor P pequeño o ínfimo se interpreta como evidencia en contra de la hipótesis nula. Considerando que el p-value se calcula suponiendo que la hipótesis nula es cierta, el umbral de medida del valor de p es 0,05. Es decir que, si la hipótesis nula es cierta, en el 5% de los estudios que se realicen se obtendrá un valor $p < 0,05$ sólo como consecuencia del azar y que estadísticamente el modelo que se está poniendo a prueba no es significativo (Gutiérrez, 2012).

La fórmula para calcular valor P ese puede apreciar en la ecuación 10:

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 \frac{(1 - p_0)}{n}}} \quad (10)$$

Donde \hat{p} es la proporción de la muestra, p_0 es la proporción hipotética y n es el tamaño de la muestra (Massey & Miller, 2016).

2.10.2 Eficiencia de Nash-Sutcliffe

La eficiencia de Nash-Sutcliffe, fue propuesta en 1970 por Nash y Sutcliffe. Se conceptúa como uno menos la suma de las diferencias absolutas al cuadrado entre los valores pronosticados y observados normalizados de un modelo, multiplicado por la varianza de los valores observados durante el periodo de estudio. Su función es cuantificar el rendimiento de un modelo hidrológico, el rango de resultados se halla entre infinito negativo y uno. Donde uno indica un ajuste perfecto, y un número negativo indica que el valor promedio de la serie original es un mejor predictor que el modelo evaluado. La principal desventaja de este criterio es que la diferencia entre valores observados y predichos se calculan al cuadrado, por tanto, existe una tendencia a sobrestimar y despreciar valores extremadamente altos o bajos (rendimiento). La ecuación 11 describe la fórmula para obtener este coeficiente:

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_o^t - Q_m^t)^2}{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2} \quad (11)$$

Donde \bar{Q}_o es el promedio de los valores observados, Q_m es el valor pronosticado por el modelo, T es el número total de observaciones y Q_o^t es el valor observado en el tiempo t (Krause et al., 2005).



2.10.3 Eficiencia de Kling-Gupta

El índice Kling Gupta tiene la misma función que el índice NSE, evaluar el rendimiento de un modelo hidrológico, sin embargo, aborda varias deficiencias. Fue creado por Gupta en 2009. De igual manera su rango va desde el infinito negativo a uno, siendo uno un ajuste perfecto y mientras más negativo el resultado indica un modelo no satisfactorio. Usa tres parámetros estadísticos, correlación lineal entre valores observados y simulados, la desviación estándar y un término de sesgo para calcular la proporción entre los valores observados y simulados (Knoben et al., 2019). La eficiencia es calculada mediante la ecuación 12:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (12)$$

Donde r es la correlación lineal entre observaciones y simulaciones, α es una medida del error de la variabilidad de flujo de la serie y β el término que mide el sesgo.

A su vez α y β se calculan a partir de las ecuaciones 13 y 14.

$$\beta = \frac{\mu_s}{\mu_o} \quad (13)$$

$$\alpha = \frac{\sigma_s}{\sigma_o} \quad (14)$$

Donde μ_s es el promedio de los valores simulados, μ_o representa el promedio de los valores observados; Y σ_s conjuntamente con σ_o , son los valores de la desviación estándar de la serie simulada y observada respectivamente (Gupta et al., 2009).

2.10.4 Coeficiente de determinación

El coeficiente de determinación denotado por R^2 , es un indicador sin unidades, es un coeficiente que indica la correlación lineal de un modelo entre x y y . El rango de resultados puede ir de uno negativo a uno positivo, considerando a uno positivo como una dependencia lineal exacta, uno negativo una dependencia lineal exacta pero inversa y cero como una falta de relación entre las variables. Funciona para cuantificar la bondad del ajuste de un modelo y se conceptúa como la



proporción de variabilidad de la variable Y explicada por X (Montoro, 2015). La ecuación 15 describe el cálculo simplificado.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SSE}{SS_{yy}} \quad (15)$$

Donde cada observación es i , y_i es la distancia entre cada observación y el promedio \bar{y} . \hat{y}_i es el valor ajustado. Cabe aclarar que la distancia entre cada observación y el promedio, es equivalente a la distancia de cada dato entre su valor ajustado más el residuo y la media. SSE es la suma de cuadrados residuales y SS_{yy} es la suma total de los cuadrados (Novales, 2010).

2.10.5 Raíz del error cuadrático medio

Se denomina RMSE por sus siglas en inglés, se considera como una métrica estándar que mide el rendimiento de un modelo. Su cálculo cuantifica la dispersión promedio de un conjunto de observaciones a partir de un valor conocido. Sus resultados varían en un rango de cero al infinito positivo, mientras menores sean los valores mejor se considera el ajuste del modelo.

Se calcula como la ecuación 16:

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (16)$$

Donde μ es el valor correcto conocido, n es el número de observaciones de μ , y x_i es solo una observación del conjunto de n observaciones (serie) (Meyer, 2010).



CAPITULO 3: RESULTADOS Y DISCUSION

3.1 Análisis exploratorio gráfico

3.1.1 Gráficas de series de tiempo

Al observar la Figura 6 del análisis exploratorio se pudo observar dos grupos de estaciones, cuatro estaciones contienen caudales altos en un rango de 0 a 200 m³/s, las cuales son: Paute AJ Dudas, Paute DJ Gualaceo, Gualaceo y Tomebamba. Las dos últimas estaciones alcanzan su máximo valor en aproximadamente 100 m³/s; El resto de estaciones se mantienen en un rango de 0 a 25 m³/s máximo, estas son: Collay, Matadero, Mazar, Dudas, Surucucho.

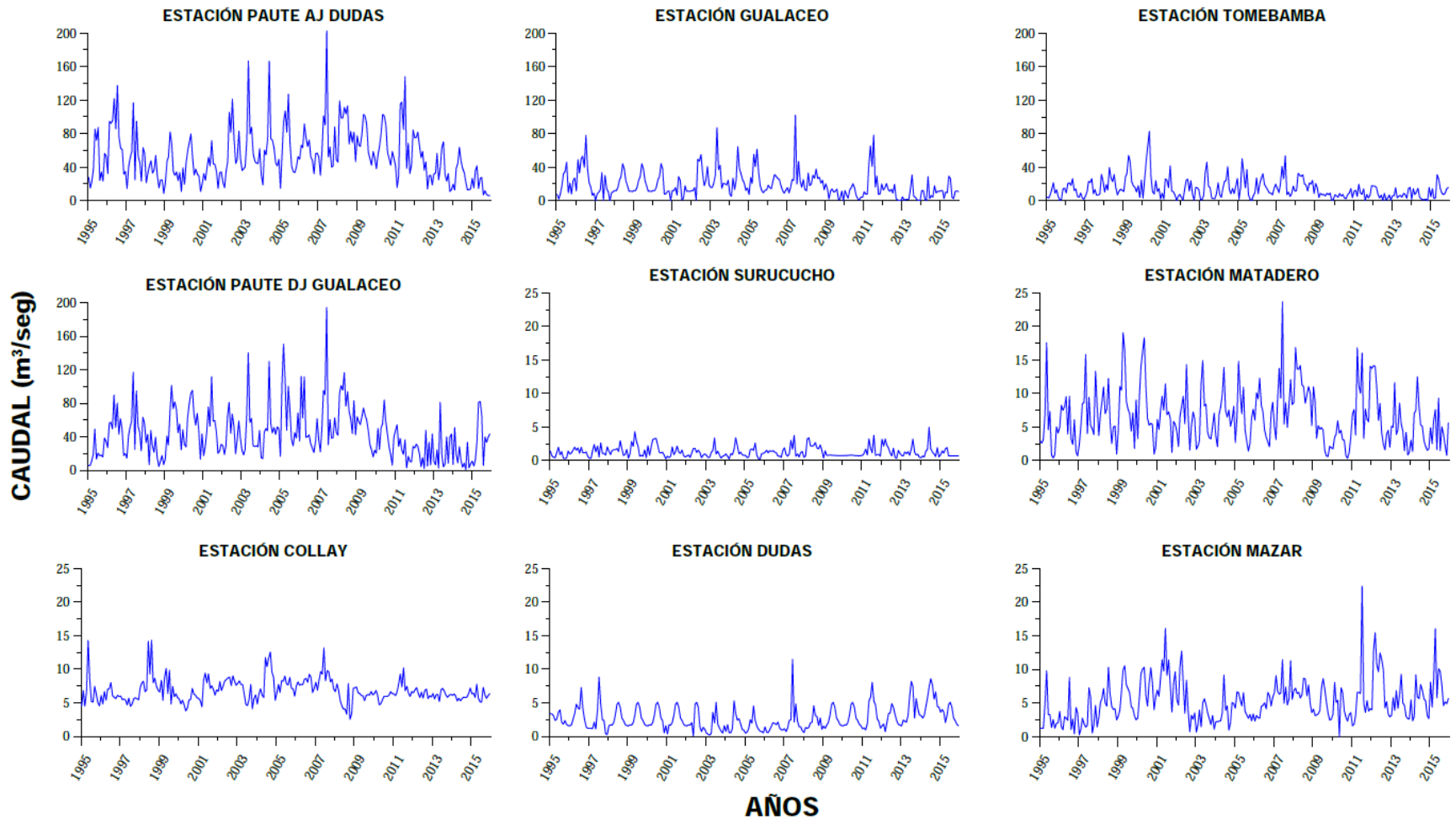


Figura 6. Series de tiempo de las estaciones de caudal seleccionadas para el estudio.



Para la estación Paute AJ Dudas se puede observar un máximo de 202,56 m³/s, el mínimo de 6,06 m³/s y un caudal promedio de 53,96 m³/s. Esta estación tiene el promedio más alto entre todas las series de tiempo analizadas, por tanto, se asume que también contiene valores de caudal más altos. La unión de los ríos Collay y Paute son medidos antes de confluir con el río Dudas en esta estación (CELEC EP & CFE-ACOTENIC, 2016).

Para la estación Gualaceo se puede observar un máximo de 102,10 m³/s, el mínimo de 0,01 m³/s y un caudal promedio de 18,96 m³/s. El río Gualaceo están conformado por cuatro ríos que vienen de la cordillera central: Sigsig, Palmar, Shío y San Francisco, la estación del mismo nombre se encarga de medir el caudal conjunto que se mantiene alto por sus diversos afluentes (CELEC EP & CFE-ACOTENIC, 2016).

En la estación de Tomebamba se observa un valor de caudal máximo de 82,71 m³/s, el mínimo de 0,10 m³/s y un caudal promedio de 14,02 m³/s. Las estaciones de Tomebamba y Matadero miden los escurrimientos de la subcuenca del río Tomebamba (CELEC EP & CFE-ACOTENIC, 2016)

El comportamiento de caudal en la estación Paute DJ Gualaceo se puede describir con un máximo de 194,24 m³/s, el mínimo de 0,97 m³/s y un caudal promedio de 43,86 m³/s. Esta estación mide el caudal del río Paute en la parte media de la cuenca, cuando el río Tomebamba y Gualaceo se han unificado (CELEC EP & CFE-ACOTENIC, 2016).

Dentro de la estación Surucucho se mantienen valores muy bajos en la serie, esto se confirma al observar un máximo de 4,96 m³/s, el mínimo de 0,23 m³/s y un caudal promedio de 1,29 m³/s. Esta estación tiene el promedio más bajo y esto es visible en la gráfica. Mide el caudal en la zona de transición del río Matadero al Tomebamba (CELEC EP & CFE-ACOTENIC, 2016).

En la estación Matadero, la cual mide los escurrimientos del río Namangoza antes de unirse con el río Tomebamba (unión del río Paute y Negro) (CELEC EP & CFE-ACOTENIC, 2016). Se percibe un caudal máximo de 23,68m³/s, el mínimo de 0,34 m³/s y un caudal promedio de 6,59 m³/s. La subcuenca del río Matadero tiene un sistema lagunar que regula los escurrimientos que forman el río a través de una quebrada, a través de caudales pequeños (Zebrowski & Roura, 1997).

La estación Collay es notable entre el resto de estaciones debido a que se puede observar que el caudal jamás llega a 0 m³/s y es constante a través del tiempo. Su caudal máximo es de 14,32 m³/s,



el mínimo de $2,53 \text{ m}^3/\text{s}$ y un caudal promedio de $6,83 \text{ m}^3/\text{s}$. Esta estación mide el caudal del río con su mismo nombre (CELEC EP & CFE-ACOTENIC, 2016).

Al examinar a la estación Dudas, primero se puede determinar que tiene una serie de tiempo con valores bajos. Mantiene un caudal máximo de $11,47 \text{ m}^3/\text{s}$, el mínimo de $0,08 \text{ m}^3/\text{s}$ y un caudal promedio de $2,68 \text{ m}^3/\text{s}$. Mide el río Dudas antes de unificarse a otras corrientes y convertirse en el río Pindilig (CELEC EP & CFE-ACOTENIC, 2016).

Finalmente, en la estación Mazar se identifica un valor de caudal máximo de $22,37 \text{ m}^3/\text{s}$, el mínimo de $0,05 \text{ m}^3/\text{s}$ y un caudal promedio de $5,27 \text{ m}^3/\text{s}$. Esta estación mide el caudal del río Mazar antes de unirse al río Paute. (CELEC EP & CFE-ACOTENIC, 2016). Las subcuencas de los ríos Dudas y Mazar, tienen zonas que reciben una precipitación baja pero constante todo el año, por eso los caudales se mantienen dentro del rango 0 a $25 \text{ m}^3/\text{s}$, sin llegar a la sequía (zonas lluviosas) (Zebrowski & Roura, 1997).

Las estaciones de precipitación usadas para el relleno de datos se pueden observar en las Figura 7. La época húmeda para las estaciones de caudal seleccionadas en el estudio se encuentra en el rango de los meses abril a julio, con mayo y junio siendo los más altos y frecuentes. La época seca se considera para los meses de septiembre a enero, siendo este último el mes con los valores más bajos registrados. En las estaciones de precipitación utilizadas se observa en mismo patrón mensual en cuanto a la época húmeda, con los meses abril y marzo con más frecuencia registrando los valores más altos, sin embargo, la época seca se modifica para los meses de septiembre, agosto (el más frecuente) y enero. La temporalidad de las estaciones de caudal está fuertemente relacionada con el régimen unimodal que tiene la temporada de junio, julio y agosto como la más lluviosa, sin embargo, la mayoría de estaciones de precipitación se asemejan más al régimen bimodal que funciona de forma viceversa al régimen unimodal, es decir que su temporada seca funciona en los meses de junio, julio y agosto, y aunque la temporada seca es un poco más aleatoria, tiende a tener una temporada humedad en los meses de septiembre a enero (Campozano et al., 2016).

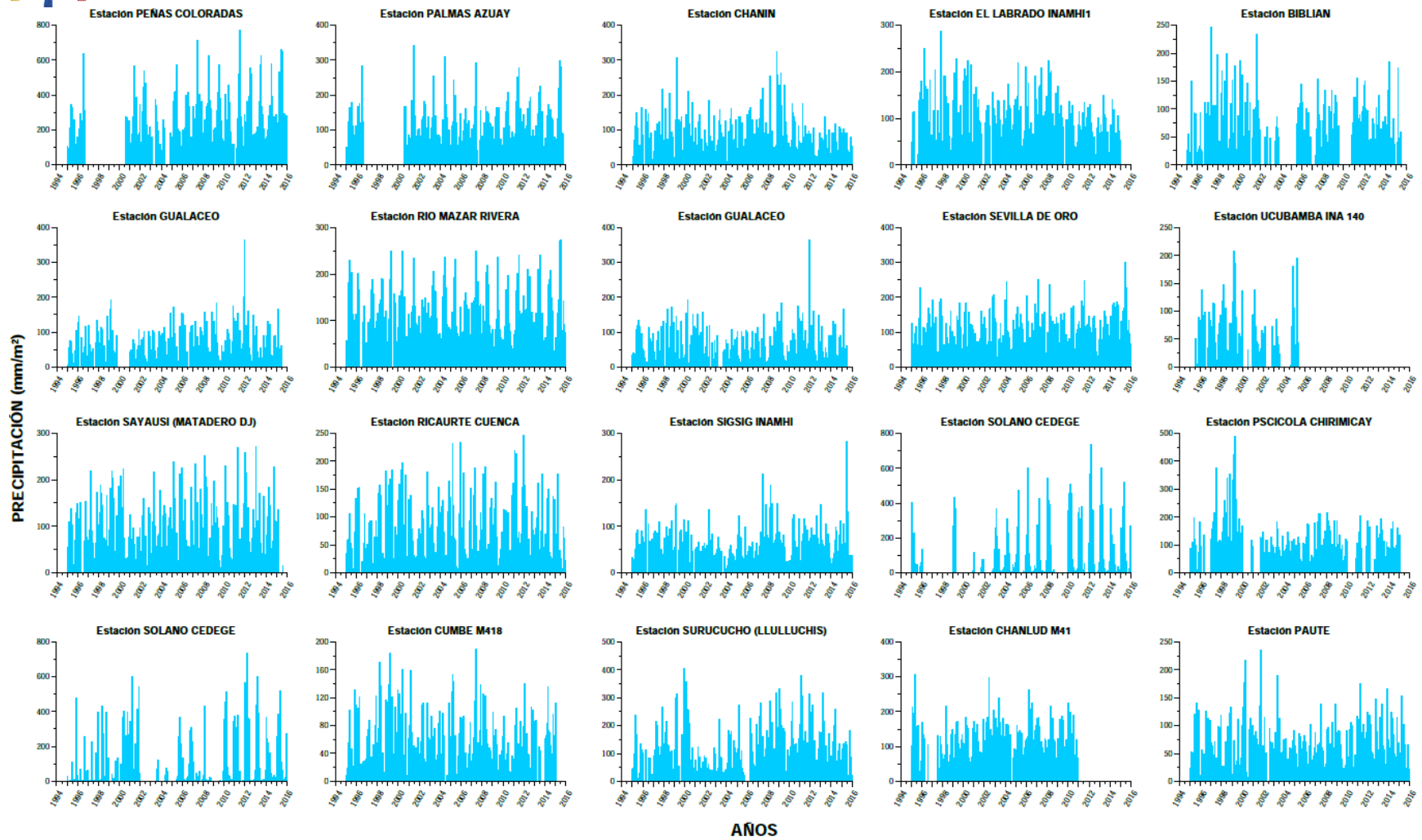


Figura 7. Series de tiempo de las estaciones de precipitación usadas en el relleno de datos.

3.2 Prueba de homogeneidad de datos

3.2.1 Curva de masa

Al observar las distintas curvas de caudal acumulado de las estaciones seleccionadas en la Figura 8 se puede observar que todas mantienen una tendencia similar, a pesar de que tienen una gran dispersión unas de otras. Esto sucede por los rangos de valores que cada una de las estaciones mide. De forma general se puede observar que las estaciones cumplen el criterio de homogeneidad en el periodo 1995 – 2015. Es decir que no se puede observar una gran influencia de actividades antrópicas o de la variabilidad climática (Pirnia et al., 2019). Los resultados obtenidos concuerdan con anteriores estudios realizados que describen el caudal del río Paute como homogéneo, considerando su temporalidad de los meses junio, julio y agosto en donde tiende a llegar a su caudal máximo (Quishpe et al., 2016).

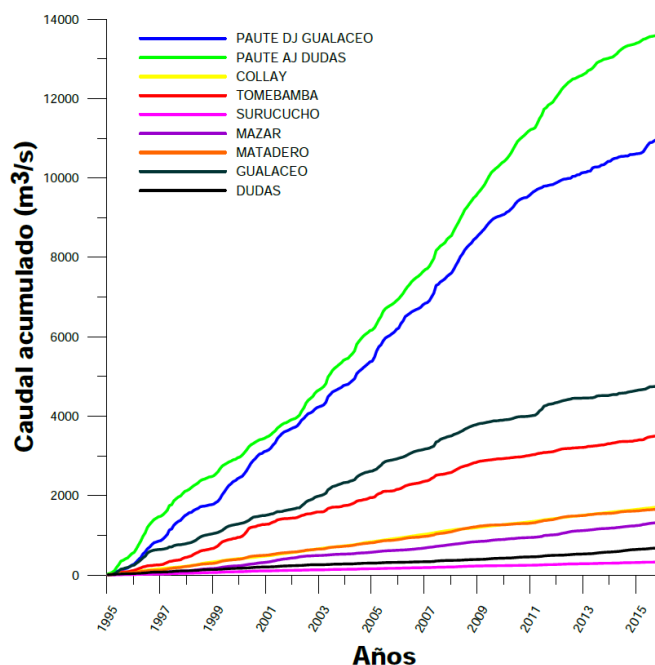


Figura 8. Prueba de homogeneidad usando curvas de masa.

3.3 Calibración y Validación de datos

De acuerdo a cada escenario propuesto para las predicciones, considerando una cantidad original de 252 datos se realizó la distribución de los porcentajes de datos usados en calibración (80%) y



validación (20%). Con excepción del escenario con retrasos en los cuales, para uno, dos y tres meses, los datos se redujeron a 251, 250 y 249 respectivamente en cada serie (50 datos para validación). Todos los modelos en etapa de calibración presentaron malos rendimientos por lo que las gráficas con los resultados se incluyeron únicamente en los Anexos. El Anexo 3 contiene series de tiempo, Figuras 44 a 47 (ACP), 48 a 53 (usando retrasos mensuales), 54 y 55 (VIF). El Anexo 4 integra los diagramas de dispersión, Figuras 59 a 59 (ACP), 60 a 65 (desfases mensuales), 66 y 67 (VIF).

En el caso de la etapa de calibración, para cada escenario se presentan el mejor modelo para 31 y 41 índices, el resto de los resultados de series de tiempo se encuentra en el Anexo 1, Figuras 32 y 33 (ACP) y 34 a 37 (retardos mensuales). En el caso de los diagramas de dispersión se agrupan en el Anexo 2, Figuras 38 y 38 (ACP), y 40 a 43 (retardos mensuales).

3.4 Modelación para predicción de caudales con series temporales de los índices climáticos atmosféricos y oceánicos

3.4.1 Modelación usando el Análisis de Componentes Principales (ACP)

Al realizar el análisis de componentes principales en los 41 índices (con relleno), y los 31 índices (sin relleno), se obtuvieron resultados distintos sobre la varianza acumulada explicada para cada escenario, en la Figura 9 se puede observar el comportamiento porcentual de la varianza (S^2 o σ^2).

Se consideraron dos casos para cada escenario, en el caso de los 41 índices se seleccionaron los componentes 10 y 2. La varianza acumulada determina que en estos componentes ya se explica el 70% y 90% del comportamiento de los datos en el espacio tal como lo indica Rea & Rea, (2016), es decir que estos porcentajes son significativos para el estudio. Para el escenario de 31 índices estos porcentajes se observan en los componentes ocho y 14. Puede notarse que se reducen a la mitad de columnas de entrada iniciales. Se considera 70% un alto porcentaje de la varianza total, por este motivo se usaron los datos de los componentes que explican este porcentaje y para evaluar si mejoraba el escenario de predicciones con mayor porcentaje de varianza también se usó el 90% recomendado por Rojas et al., (2020).

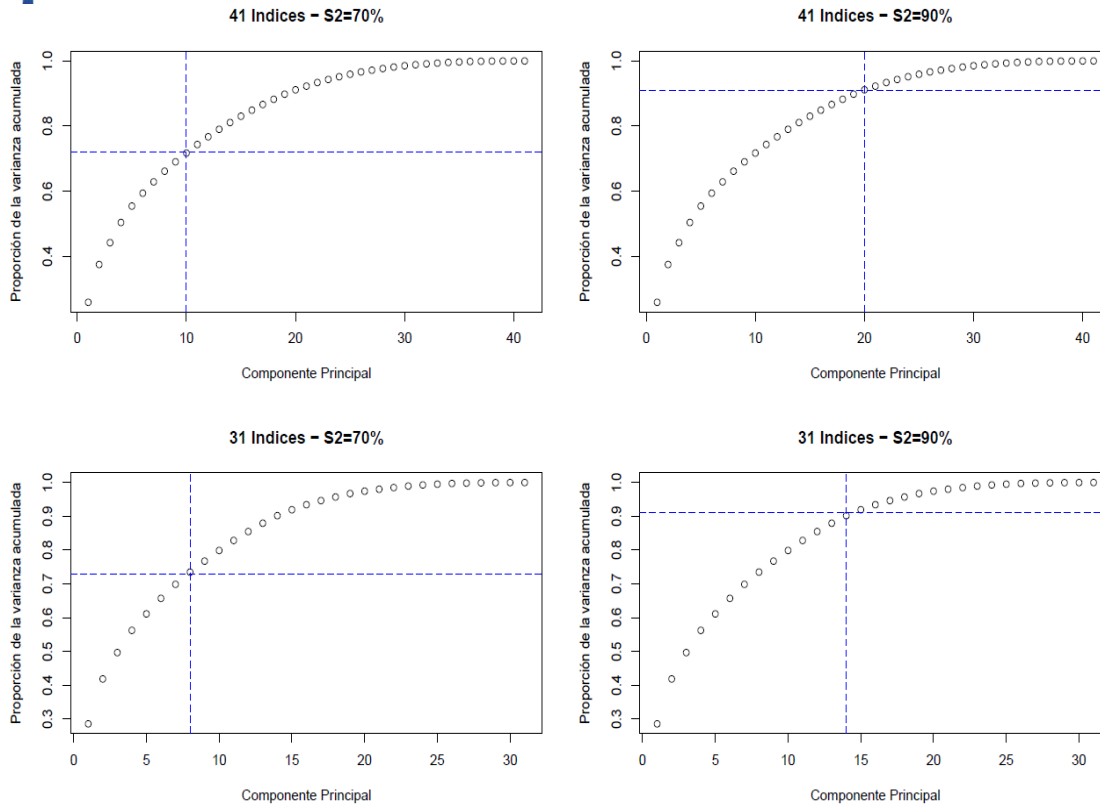


Figura 9. Proporción de la varianza acumulada para los escenarios de predicciones de caudal usando ACP

Los resultados obtenidos con el análisis ACP se graficaron en figuras de series de tiempo y diagramas de dispersión. Los mejores resultados se obtuvieron con los escenarios con la mayor varianza explicada, es decir con 14 componentes para los 31 índices (sin relleno) y con 20 componentes para los 41 índices, como se puede observar en la Figura 10 y 11.

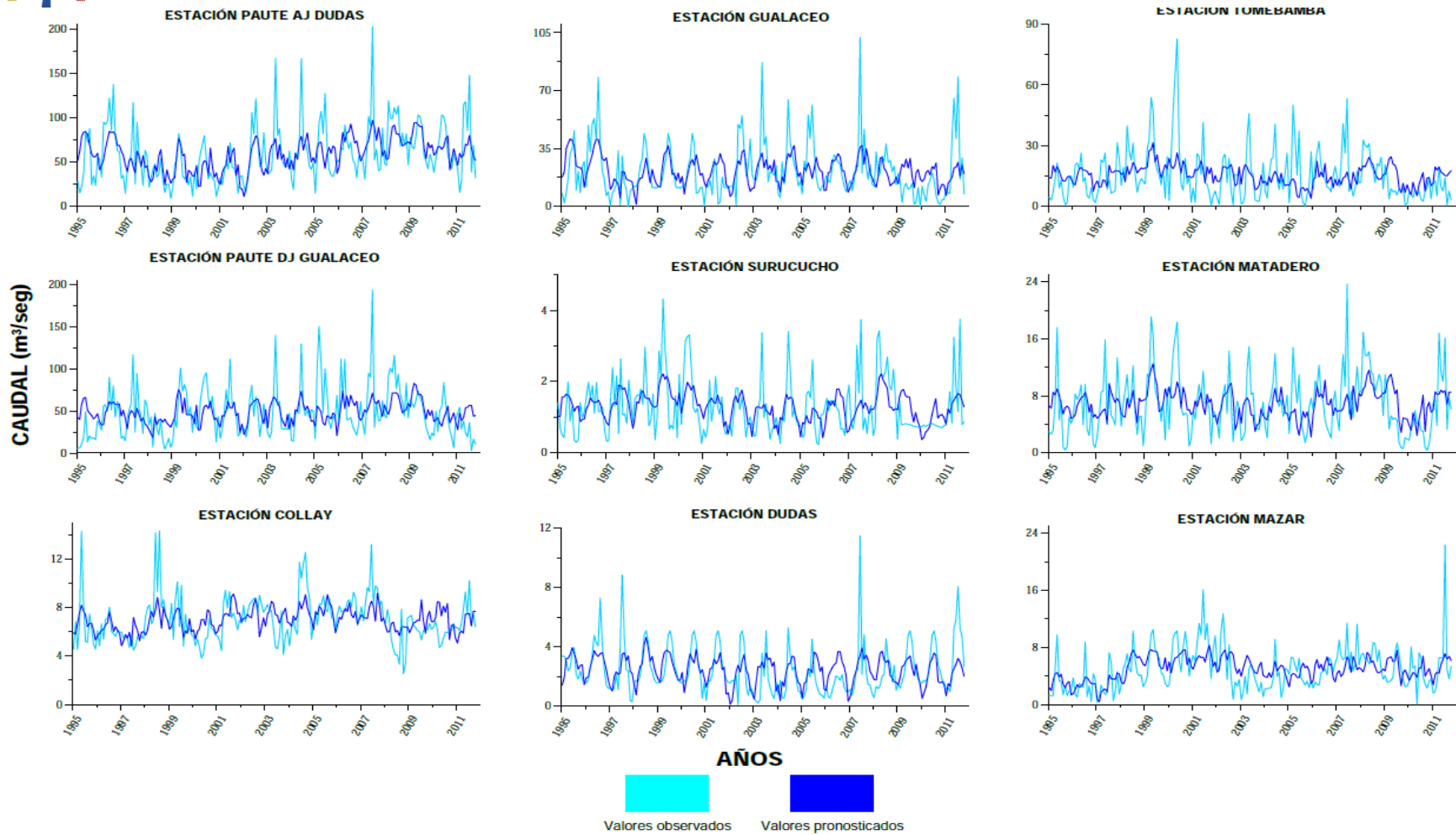


Figura 10. Resultados del escenario utilizando 31 índices y 90% de varianza con el método ACP (etapa de calibración).

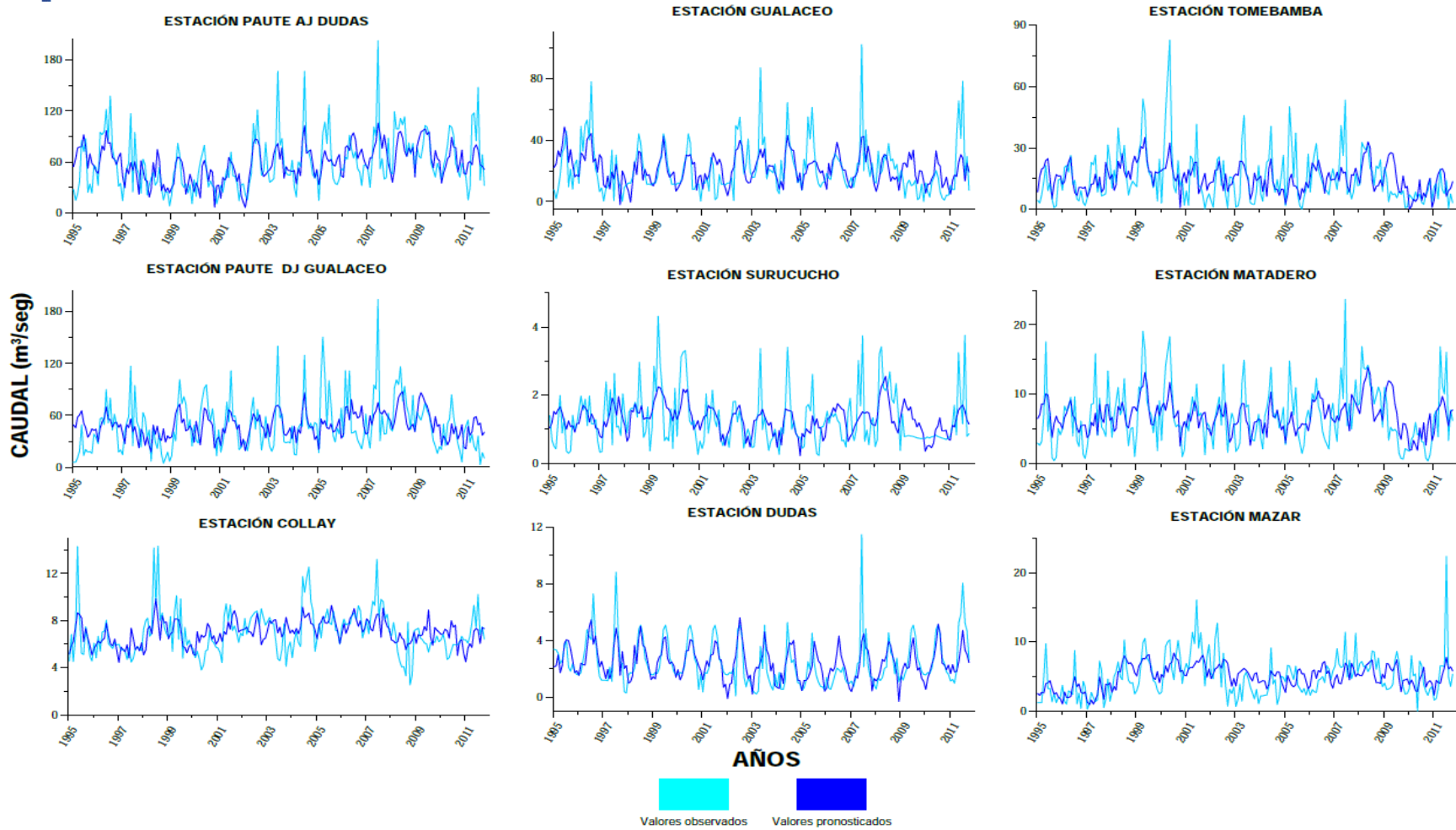


Figura 11. Resultados del escenario utilizando 41 índices y 90% de varianza con el método ACP (etapa de calibración).



Se puede observar en la Figura 10 que la serie de tiempo con valores predichos logra mantener la misma secuencia de incrementos y reducciones de caudal. Sin embargo, no se acerca en ningún momento a los valores altos observados en las series de tiempo originales, consigue determinar mejor el descenso de los valores de caudal. Funciona de forma poco precisa para todas las estaciones sin considerar el rango de valores individual. Es decir que la modelación propiamente no se ve afectada por el conjunto de estaciones, sino que procesa individualmente las predicciones.

En la Figura 11 usando 41 índices y 20 componentes se observa que los resultados mejoran en las estaciones Tomebamba y Surucucho, la predicción se ajusta un poco más a la serie original sin llegar a un resultado aceptable. El comportamiento de la serie es similar al escenario con 31 índices y 14 componentes en cuanto a tendencia.

En la Figura 12 se puede observar que los índices con mayor correlación para el escenario con 31 índices son: AAO (positiva y negativa), AMOS, EA/WR, PNA, QBO (correlación positiva), PDO y SR (correlación negativa). Esta relación no supera el 0,50 ni -0,50, por lo que en nuestro estudio se puede notar que la influencia es débil pero notoria debido al número de variables. Las correlaciones más altas se producen en QBO y SR al alcanzar 0,50 y -0,50 respectivamente.

En la Figura 13 se presentan los resultados de la carga factorial del modelo ACP con 41 índices (correlación). Los índices más correlacionados a los componentes principales son: AMOS, QBO, SF, SR y SWMRR con valores positivos, PDO, NAO (Jones), TP SST EOF de forma negativa. Las correlaciones más altas ocurren con QBO (0,57) y SWMRR (0,39). La puntuación de estas cargas indica una correlación fuertemente positiva si alcanza uno, una correlación fuertemente negativa si alcanza uno negativo, y una neutralidad al mantenerse en cero o cercano (Winter & Dodou, 2016).

En un estudio realizado en Irán, se usaron 25 patrones de teleconexión para explicar la variabilidad de la precipitación a través del método ACP, de las cuales se obtuvo ocho componentes principales que explicaban un 80% de varianza, los índices y modos de variabilidad seleccionados de acuerdo a la correlación con cada componente fueron: AMO, AMM, BEST, Niño 3.4, El Niño 4, NTA, SOI y TNA (Choubin et al., 2016). En nuestro estudio únicamente el índice AMO se mantiene entre los más correlacionados.



La cantidad de índices que muestran mayor influencia (a partir de 0,25) es similar con 31 y 41 índices varía entre siete y ocho respectivamente. Comparten la tendencia con cuatro índices (AMOS, QBO, PDO y SR) aunque en los últimos varía la dirección de la correlación. La tendencia entre las relaciones mayormente positivas o negativas no tiene ninguna relación con el número de variables usadas en el escenario. Las correlaciones en términos de rango son similares al escenario con 31 índices.

En comparación con los escenarios con retrasos y con análisis VIF ningún índice se repite en todos los escenarios, sin embargo, lo que están presentes en la mayoría de modelos son: AMOS, SR, PDO y AMM, el resto de índices actúa esporádicamente con el modelo de cada estación.

En la Figura 12 también se puede notar que el componente ocho captura la variabilidad de algunos índices que no tienen una coincidencia espacial con la zona de estudio. O que están más asociados al índice ENOS, como: EA/WR, PDO, así lo confirman Sun et al., (2016) y Mezzina et al., (2020) respectivamente. Conjuntamente con SF y GB que funcionan a nivel global, AAO que se ubica en la antártica y finalmente: QBO y AMOS que se encuentran circulando a través del Ecuador mantienen también una fuerte correlación con este componente. Sin embargo, en el componente 14 los índices con coincidencia espacial como: SF y AMOS siguen presentes, además de PNA, AAO que se ubican más dispersos del área de estudio y se manifiesta una fuerte correlación con SR y TSA que funcionan sobre África (véase Tabla 1).

En la Figura 13 con el modelo que incluye 41 índices, se produce una situación similar, el componente 10, captura la variabilidad de los índices asociados directamente con ENOS: NOI, NP, PDO, EP/NP, y otros que funcionan a nivel global o en otras zonas del planeta: SF, SR, SWMRR y TSA, únicamente dos índices que afectan directamente a la región de Ecuador presentan una fuerte correlación: QBO y NBRA. En el componente 20 más bien sucede que la mayoría de índices que proporcionan la variabilidad funcionan en otras regiones como: AT SST EOF, NAO (Jones), SWMRR y TP SST EOF, con únicamente PNA ubicado dentro de la zona de estudio. Aunque este comportamiento no es estricto para todos los índices, muestra un patrón claro en las correlaciones más altas.

Es visible que aunque el índice QBO funciona sobre la estratosfera del Ecuador, su comportamiento no está involucrado con la precipitación ni caudal de la zona, por este motivo



aunque presenta una gran correlación en el escenario ACP, en los componentes para 31 y 41 índices que explican un 70% de varianza, no se obtienen resultados satisfactorios de estos modelos, y en cambio los modelos que no incluyen este índice como uno de los principales influyentes con 90% de varianza alcanzan el mejor ajuste de todo el estudio.

En un estudio realizado en Colombia que estudia la variabilidad climática del río Cauca, al utilizar ACP en los índices: CCC (corriente en chorro del Choco, índice propio de la zona de estudio), ONI, PDO, EMI (El Niño Midoki), SST, SOI y MEI; Obtuvo que dos componentes explicaban el 80% de su varianza, el primer componente con solo 70% de varianza explicada, al analizar sus eigen vectores (los eigen vectores mantienen su proporción en los factores de carga). Se identificó que la variabilidad de este componente explicaba más a los índices y modos de variabilidad asociados a ENOS y a pesar de que el segundo componente explicaba más a EMI que aún es un índice dependiente del fenómeno ENOS, también capturaba mucho mejor la variabilidad de CCC, es decir del lugar de estudio (Sedano, 2017). En nuestro estudio esta dinámica es similar.

Se puede notar también que en los estudios mencionados de Choubin et al., (2016) y Sedano, (2017), el porcentaje de datos de entrada que se necesitan para explicar un 70% de varianza es de 25 a 30%, y para un 90% se necesita de 35 hasta 50% de los mismos. Al analizar la cantidad de índices que muestran una influencia visible en cada componente, se puede notar que este número decae al incrementar el porcentaje de varianza explicada, es decir que focaliza gran parte de este peso en unos cuantos índices a pesar de generar más componentes.

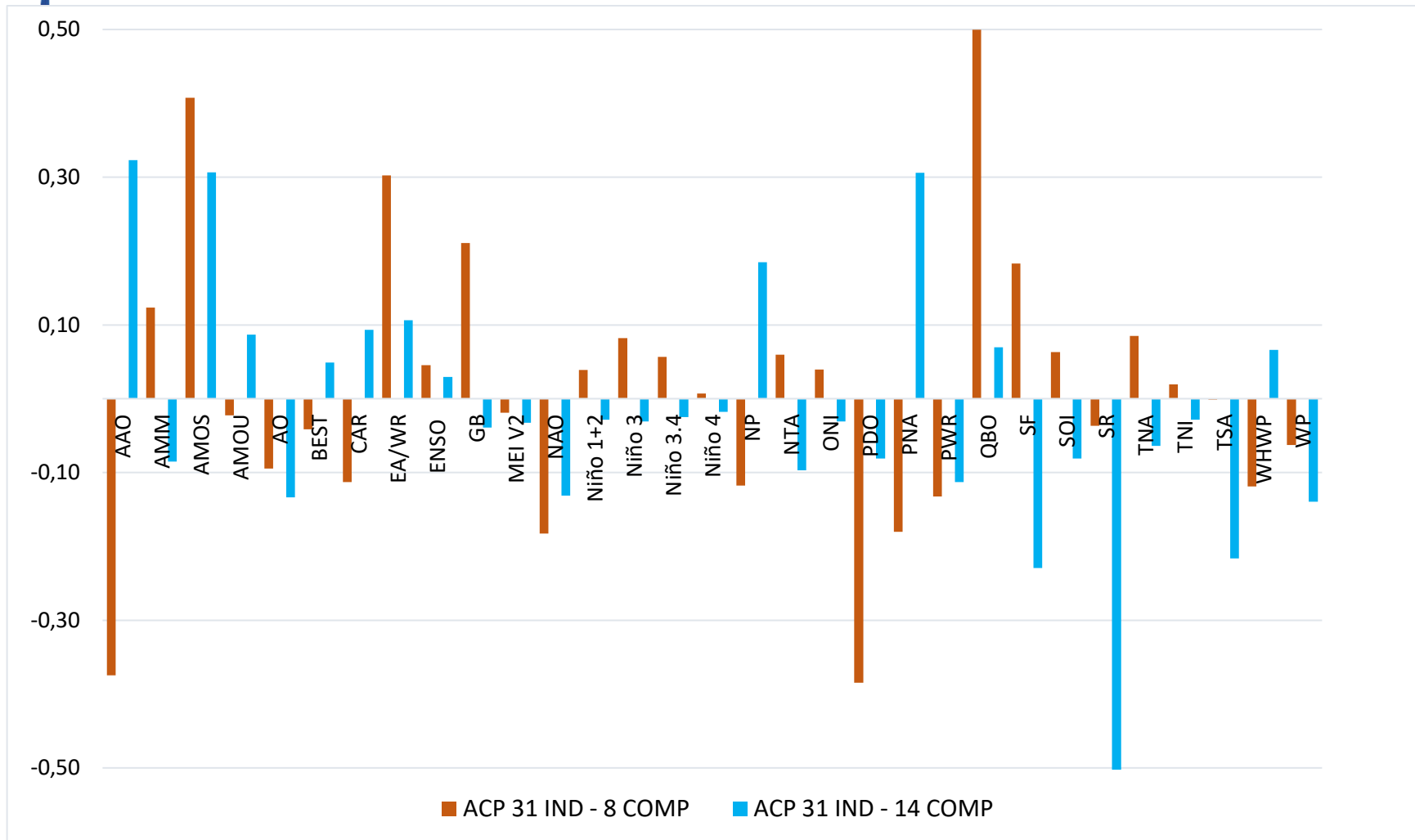


Figura 12. Correlación entre los índices de teleconexión y los componentes ocho y 14 de la modelación ACP utilizando 31 índices.

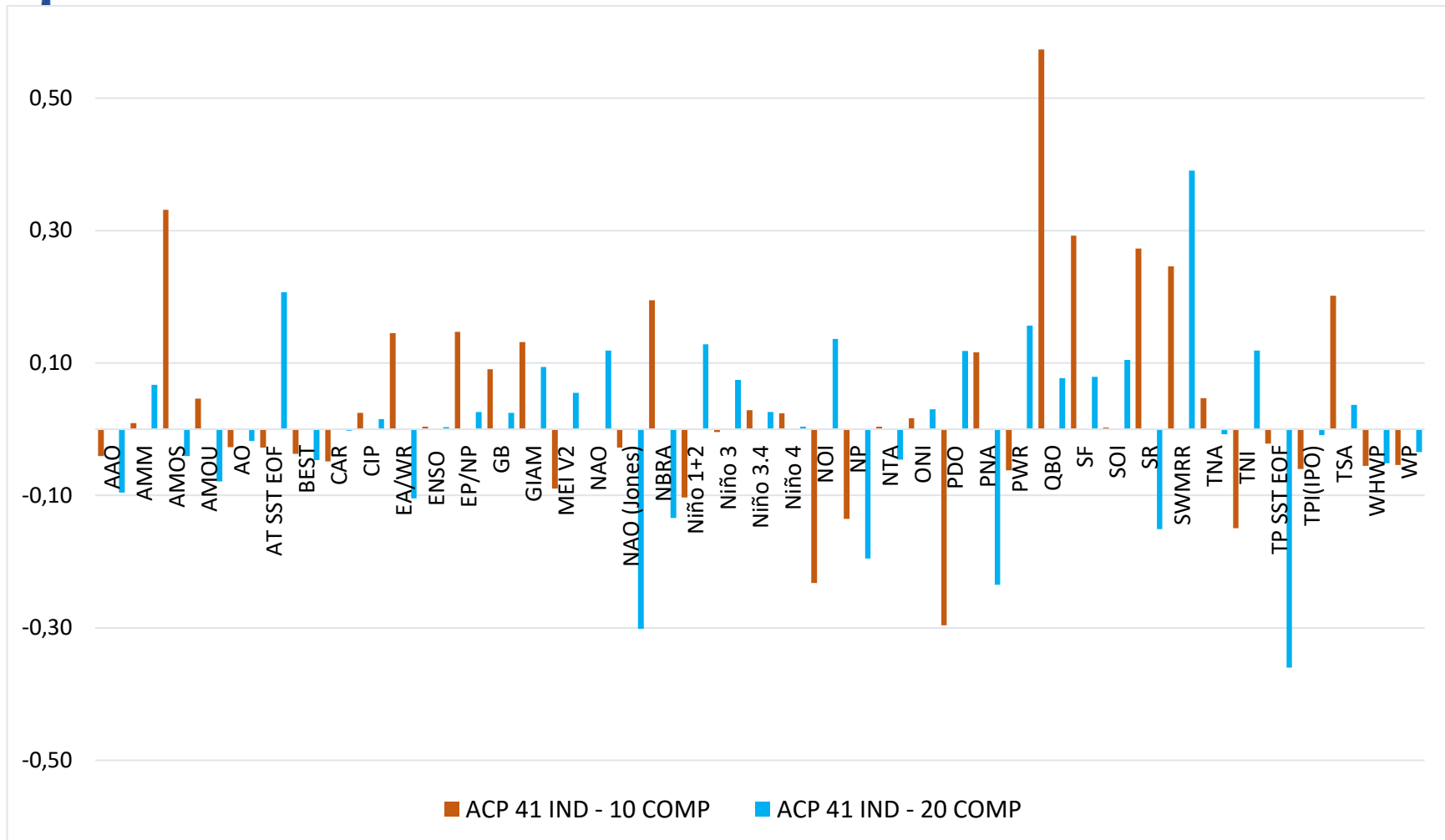


Figura 13. Correlación entre los índices de teleconexión y los componentes 10 y 20 de la modelación ACP utilizando 41 índices

En la Figura 14 se pueden apreciar los diagramas de dispersión de la modelación ACP con 31 índices y 14 componentes. La estación con R^2 más alto es Paute AJ Dudas con 0,32, seguida de Dudas con 0,30. El peor resultado se produce en Tomebamba con 0,14. En el caso del escenario con 41 índices y 20 componentes (Figura 15), todas las estaciones presentan una mejoría en su correlación, la estación Dudas alcanza 0,48, y Paute AJ Dudas siendo la segunda mejor estación tiene 0,41. El valor más bajo es 0,23 en la estación Paute DJ Gualaceo.

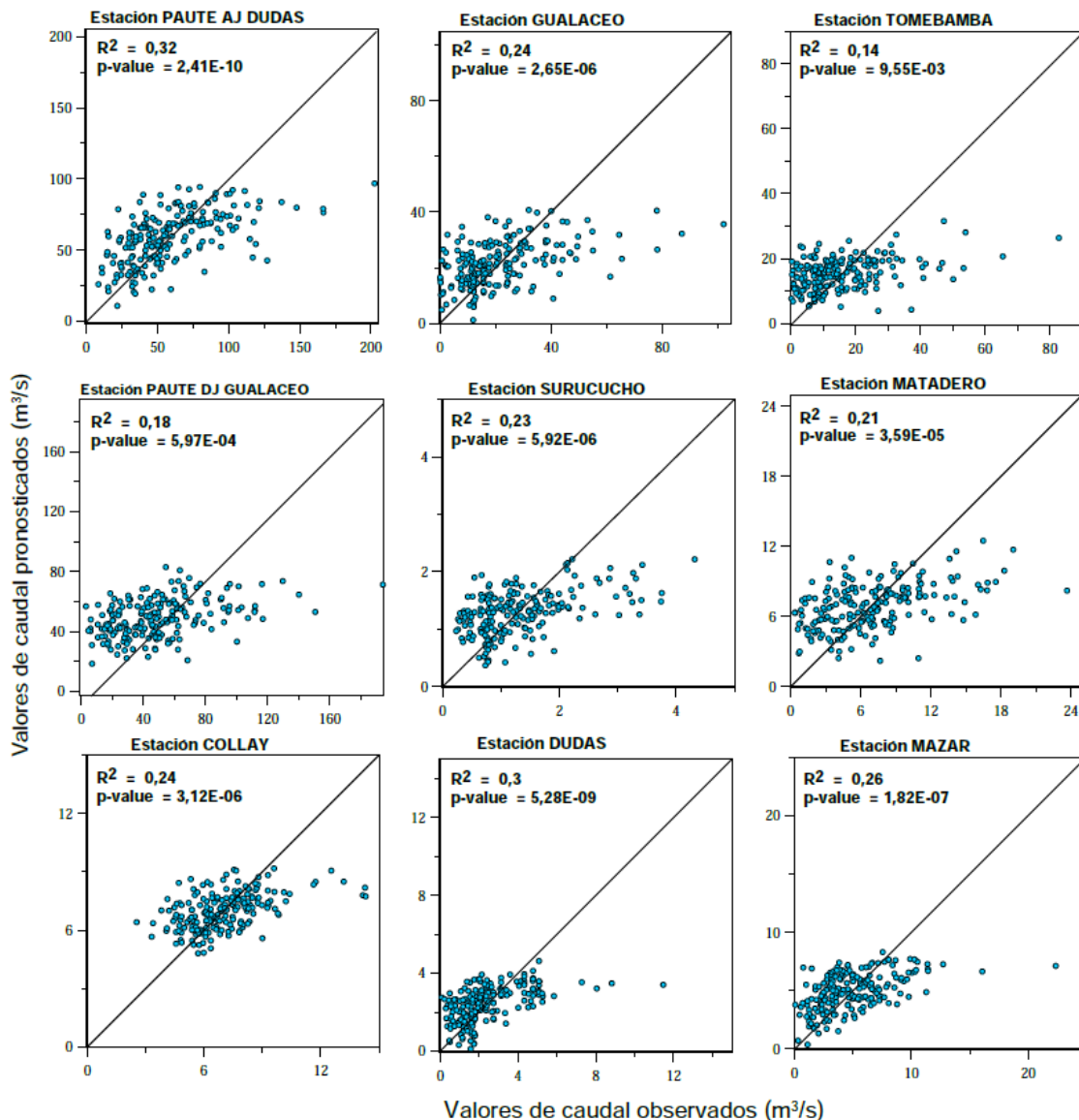


Figura 14. Resultados del análisis de dispersión entre el escenario utilizando 31 índices con el método ACP y 90% de varianza (etapa de calibración).

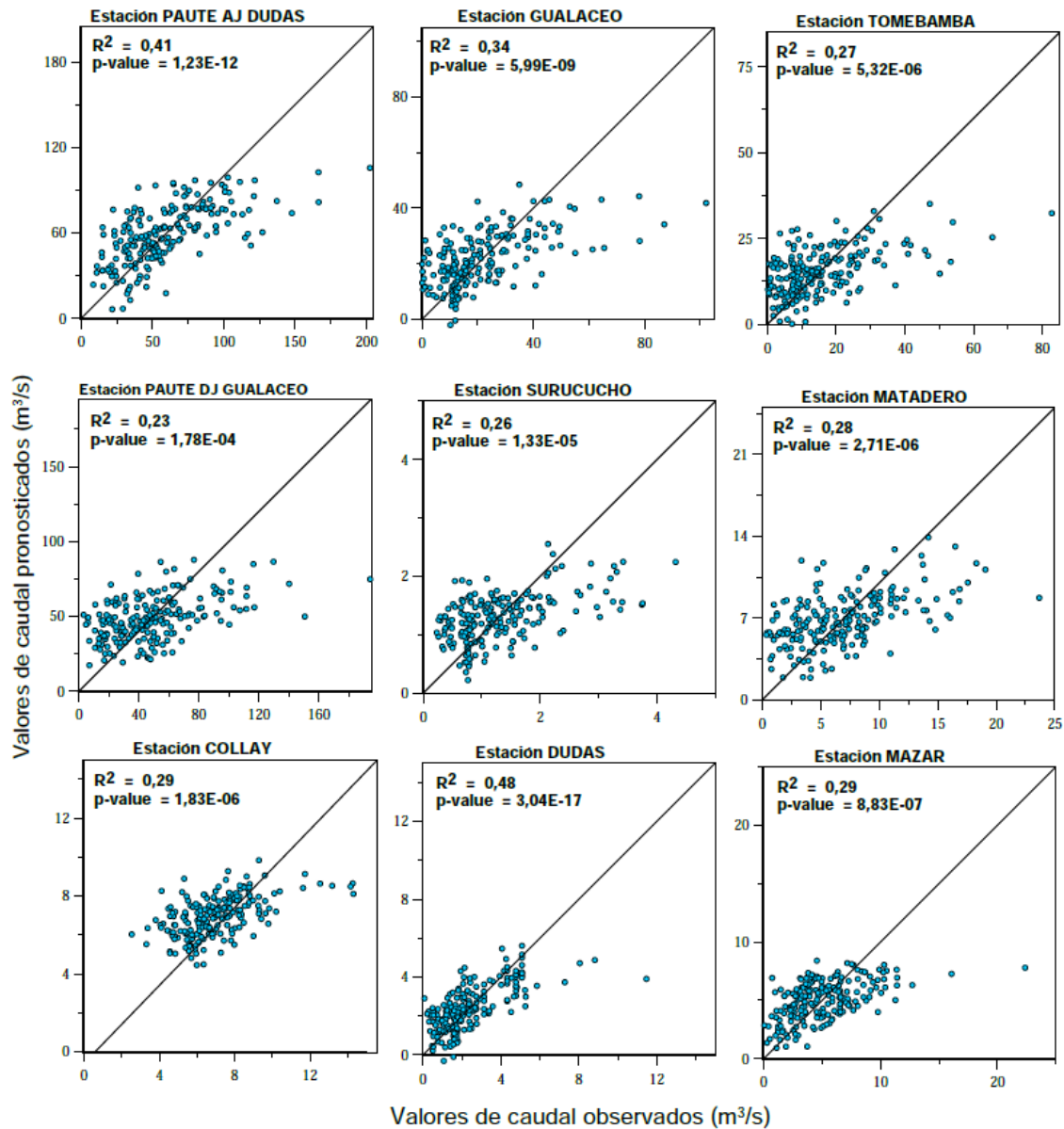


Figura 15. Resultados del análisis de dispersión entre el escenario utilizando 41 índices con el método ACP y 90% de varianza (etapa de calibración).

3.4.2 Modelación usando Índices climáticos atmosféricos y oceánicos y retrasos mensuales

Para determinar las variables de cada modelo de predicción para las estaciones de caudal, se realizó una correlación para los escenarios con y sin relleno de datos de los índices de teleconexión, y considerando el valor absoluto de las correlaciones. Se utilizaron los primeros cinco índices con valores más altos. Se pueden observar los resultados de la correlación para cada estación en la Figura 16.

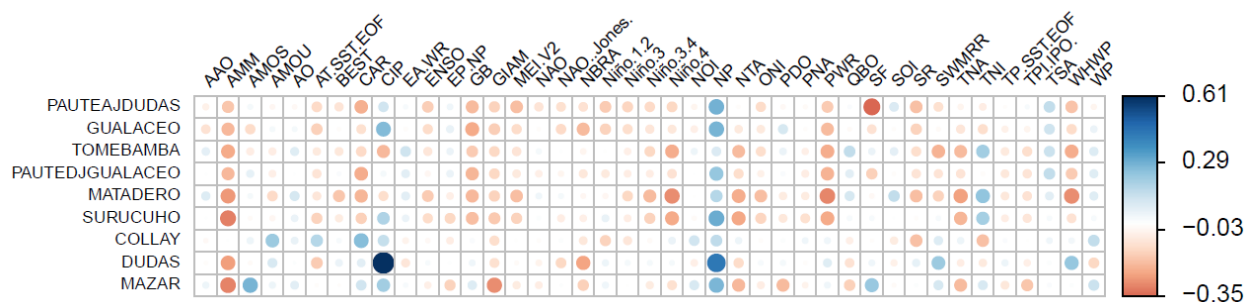


Figura 16. Grafica de representación de la correlación existente entre las estaciones de caudal y los Índices de teleconexión.

En la Tabla 4 se puede observar con que índices se emparejó cada estación, para la mayoría de índices los escenarios con relleno de datos no tuvieron ninguna influencia. Resultando en los mismos índices como variables para el modelo, en otras estaciones de caudal los cambios fueron mínimos, manteniendo la mayoría de índices similares entre sí. Por lo que se considera el peso de la variabilidad del índice en ajuste con el caudal, pero la frecuencia es similar. Se puede identificar también que la correlación se mantiene mayormente en un rango de 0,20 a 0,35. La correlación más alta la tiene Dudas con CIP, con 0,61, y la más baja pertenece a la misma estación con el índice NTA con 0,11. Las tres estaciones que tienen distintas variables dependiendo del escenario son: Gualaceo, Dudas y Collay.

Tabla 4. Valores de correlación de los índices usados para los modelos de regresión múltiple de cada estación de caudal.

ESTACIONES	Nº DE INDICES	VALORES DE CORRELACIÓN				
		SF	NP	CAR	GB	WHWP
PAUTE AJ DUDAS	31/41	0,35	0,29	0,21	0,19	0,18
		NP	GB	AMM	PWR	Niño 1+2
GUALACEO	31	0,28	0,23	0,20	0,19	0,14



	41	NP	CIP	GB	AMM	PWR
		0,28	0,26	0,23	0,20	0,19
TOMBAMBA	31/41	AMM	WHWP	Niño 4	PWR	TNI
		0,23	0,22	0,22	0,22	0,21
PAUTE DJ GUALACEO	31/41	NP	CAR	PWR	GB	AMM
		0,23	0,22	0,21	0,20	0,19
SURUCUCHO	31/41	AMM	NP	Niño 4	PWR	NTA
		0,31	0,30	0,23	0,23	0,23
MATADERO	31/41	PWR	WHWP	Niño 4	AMM	TNA
		0,29	0,28	0,28	0,26	0,24
COLLAY	31	CAR	AMOU	TNI	SR	WP
		0,26	0,22	0,18	0,18	0,15
	41	CAR	AMOU	TNI	SR	AT SST EOF
		0,26	0,22	0,18	0,18	0,17
DUDAS	31	NP	AMM	WHWP	WP	NTA
		0,44	0,25	0,23	0,13	0,11
	41	CIP	NP	AMM	NBRA	WHWP
		0,61	0,44	0,25	0,24	0,23
MAZAR	31/41	AMM	NP	AMOS	SF	NTA
		0,30	0,28	0,28	0,28	0,23

Elaborado por: María Daniela González

Un estudio que analizó la correlación espacial de los índices de teleconexión y modos de variabilidad de ENOS en el Ecuador, encontró que teleconexión positiva más alta pertenece al Niño 1+2, seguida de Niño 3 y Niño 4 con correlación negativa y positiva. De acuerdo a la distancia media, las teleconexiones: SOI, El Niño 1+2, Niño 3.4, Niño 4, TNI y NP mantuvieron un valor de 0,5 de correlación con una distancia de 100km. MEI, el Niño 3 y ONI alcanzaron este valor en una distancia más reducida. Las teleconexiones del Atlántico mostraron un comportamiento similar a las del Pacífico, AMO, AMM, TP SST EOF y TSA alcanzaron 0,5 de correlación dentro de un radio de 100 km. CAR y PNA a una distancia menor (Ballari et al., 2020).

Estos resultados encajan con nuestro análisis de correlación para los modelos predictivos, sin embargo, al usar todos los índices disponibles de NOAA, también se observa su influencia, AMM es el más frecuente (nueve repeticiones), seguido de NP (ocho), PWR (seis) y WHWP (cinco).

Los mejores resultados determinados por las métricas, fueron los obtenidos con la serie con solo un mes de retraso. En un estudio realizado en 2007 en todo el sur del Ecuador, se determinó que las variables multi-escala sirven para evaluar las variables climáticas en todos los rangos de



tiempo, ya que se pueden identificar rápidamente las condiciones climáticas y su intensidad a corto o largo plazo. También dependencias o correlaciones retrasadas se manifiestan espontáneamente al comparar las variables de teleconexión atmosférica (Emck, 2007).

En la Figura 17 podemos observar que la serie pronosticada con el retraso de un mes se ajusta bien a la tendencia de cada estación, sobre todo en valores medios de la serie y valores bajos, sin embargo, presenta el mismo problema que la serie pronosticada con el método ACP, todos los valores son indiferentes a los resultados del modelo y no se acercan a estos de forma individual.

Es decir, considerando que cada estación mantenía rangos distintos de valores en sus series de tiempo originales, la capacidad de predicción del modelo en general es mala para integrar los valores altos. Además, estos no dependen de una cifra específica para medir su alcance (si el valor más alto es cinco o $200 \text{ m}^3/\text{s}$ no influye en el comportamiento del modelo).

En la Figura 18, se observan los resultados del análisis usando 41 índices, los cuales son similares en su mayoría a la Figura 16 con 31 índices, la predictibilidad no mejora usando más diversidad de variables (no se reconocen valores altos). Collay se mantiene similar, aunque mantuvo distintos índices en cada escenario. Dudas y Gualaceo adoptan una forma mucho más homogénea y tienden a mantener la misma oscilación con un valor cercano entre sí para cada año, perdiendo la capacidad de predecir valores bajos de las series de tiempo.

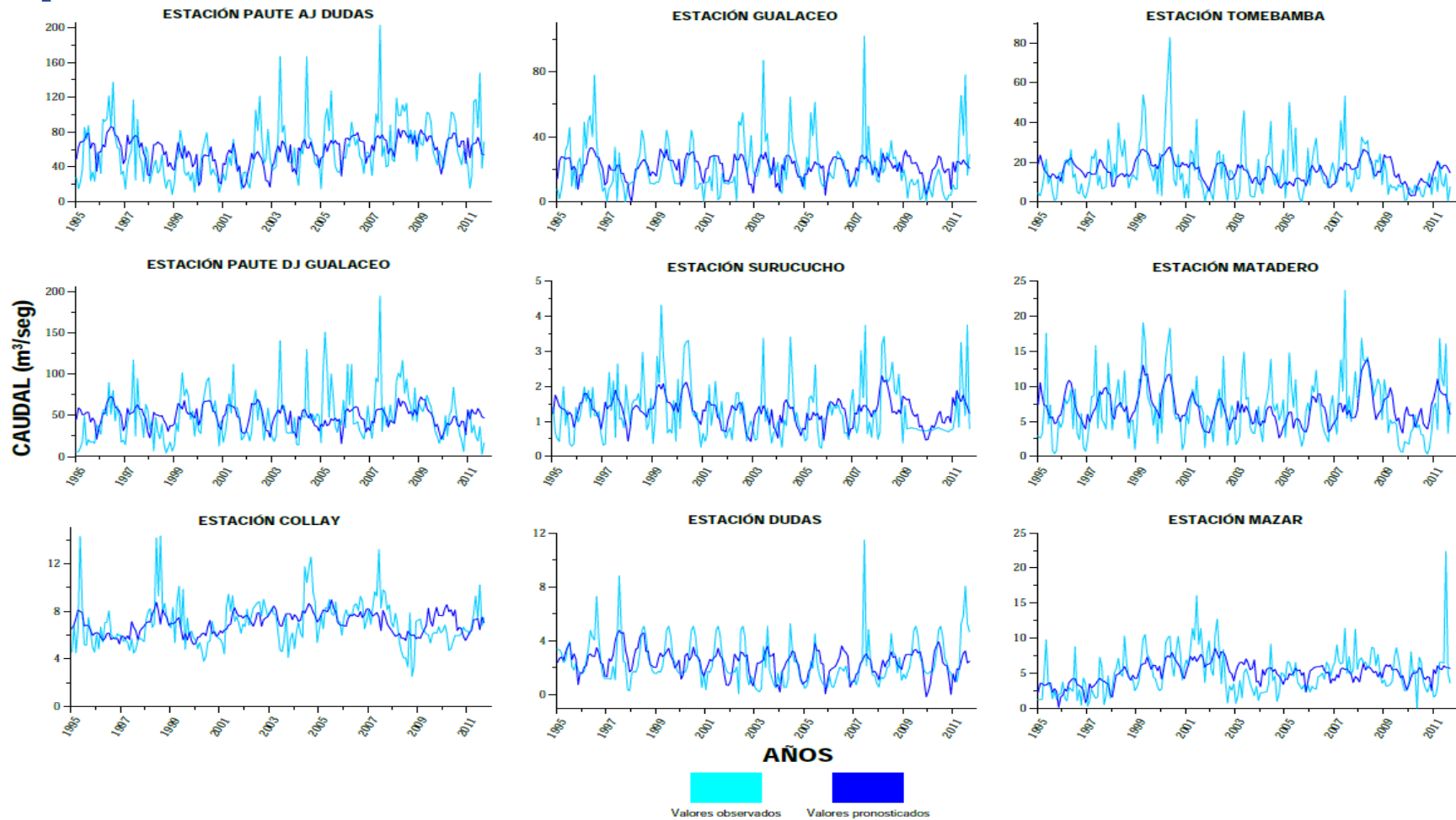


Figura 17. Resultados del escenario utilizando 31 índices y un mes de retraso (etapa de calibración).

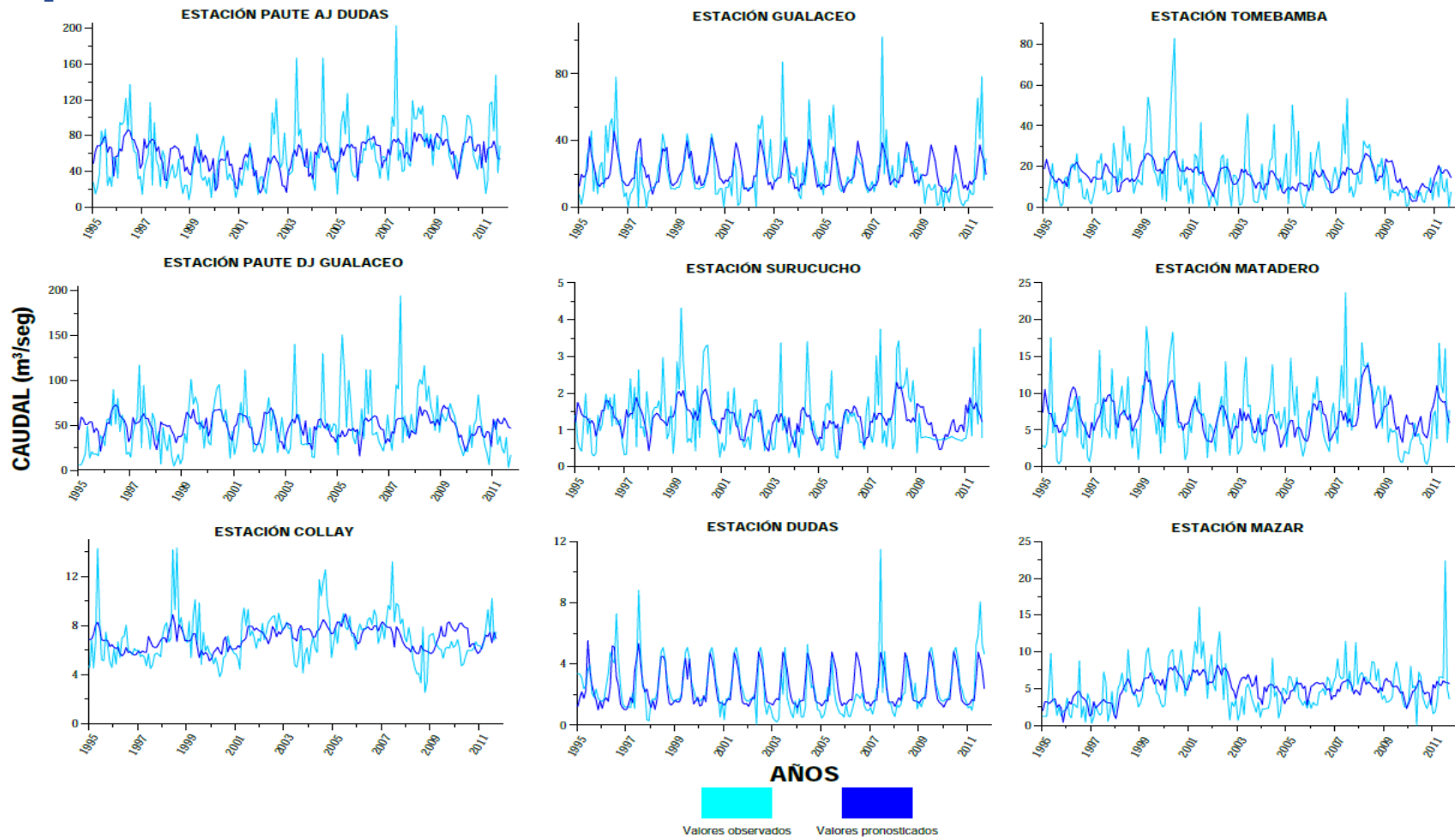


Figura 18. Resultados del escenario utilizando 41 índices y un mes de retraso (etapa de calibración).



En 2018 una investigación sobre la evolución de los sedimentos de varios lagos en el parque Nacional El Cajas determinó que con un mes de retraso para el índice ENOS los resultados de la correlación con las precipitaciones cercanas a las estaciones del lugar fueron sólidos. Probando que existe un vínculo positivo fuerte de las lluvias intensas y por consiguiente un aumento de caudal entre: La Niña, Niño 3.4 (explica la variabilidad de los andes ecuatorianos), El Niño 4 (relación más intensa), El Niño 1+2 (explica la variabilidad de las llanuras costeras) y un vínculo más neutral con El Niño (Schneider et al., 2018).

El desfase de un mes en nuestro estudio prueba que existe un retraso en los efectos de los índices de teleconexión en las condiciones climáticas de la cuenca del río Paute. Y que la variabilidad de los índices encaja con el nivel de correlación hallado considerando la configuración espacial de los estudios. En un estudio realizado en Brasil sobre la relación entre las llanuras aluviales y la conexión con las señales (modos de variabilidad) de teleconexión de ENOS (El Niño, Niño 3.4, SOI), se determinó que para la cuenca amazónica (cruza parte del Ecuador), existe un retardo de dos meses. Este sucede de enero a marzo y provoca precipitaciones más bajas, causando una disminución de caudal (Schöngart et al., 2004). Sin embargo una investigación realizada en Perú sobre la relación entre la lluvia mensual y la TSM identifico que en las áreas del Pacífico ecuatorial ENOS tiene un retraso de un mes en la estación húmeda y cero en la estación seca (Bazo et al., 2013).

Dado que en nuestro estudio la modelación con 41 índices y un mes de retraso es la segunda mejor en capacidad predictiva (mejor escenario en usando desfases). Esto en comparación con los estudios mencionados ratifica los resultados debido a que los índices con mayor correlación para este escenario son: AMM, NP, PWR y WHWP, los cuales están fuertemente influenciados por los modos de variabilidad de ENOS.

PWR está condicionado por El Niño (Carreric, 2020); WHWP se ve igualmente afecta de forma directa por El Niño en el verano, a través del aumento de temperatura de TNA que sucede por el invierno de El Niño en el Pacífico (Wang & Enfield, 2001). NP está dominado por las variaciones interanuales de los eventos El Niño y La Niña (Espino Sánchez, 2014); AMM por su parte produce anomalías en la TSM al norte de Ecuador, pero se produce durante la primavera boreal (marzo)

con mayor intensidad. En la Figura 19 se pueden apreciar las estaciones seleccionadas y el régimen de precipitación en el que se encuentran.

Los valores extremos de AMM se encuentran en sincronía con la época húmeda del régimen bimodal (cinco estaciones), y con la época seca del régimen unimodal (cuatro estaciones) por lo que, al generar los retrasos de uno, dos y tres meses teóricamente la correlación debería verse afectada negativamente. Por lo que a la razón que se apela es a la coincidencia espacial en la zona del estudio, ya que incluso su relación con el fenómeno ENOS es ínfima (Vimont, 2012).

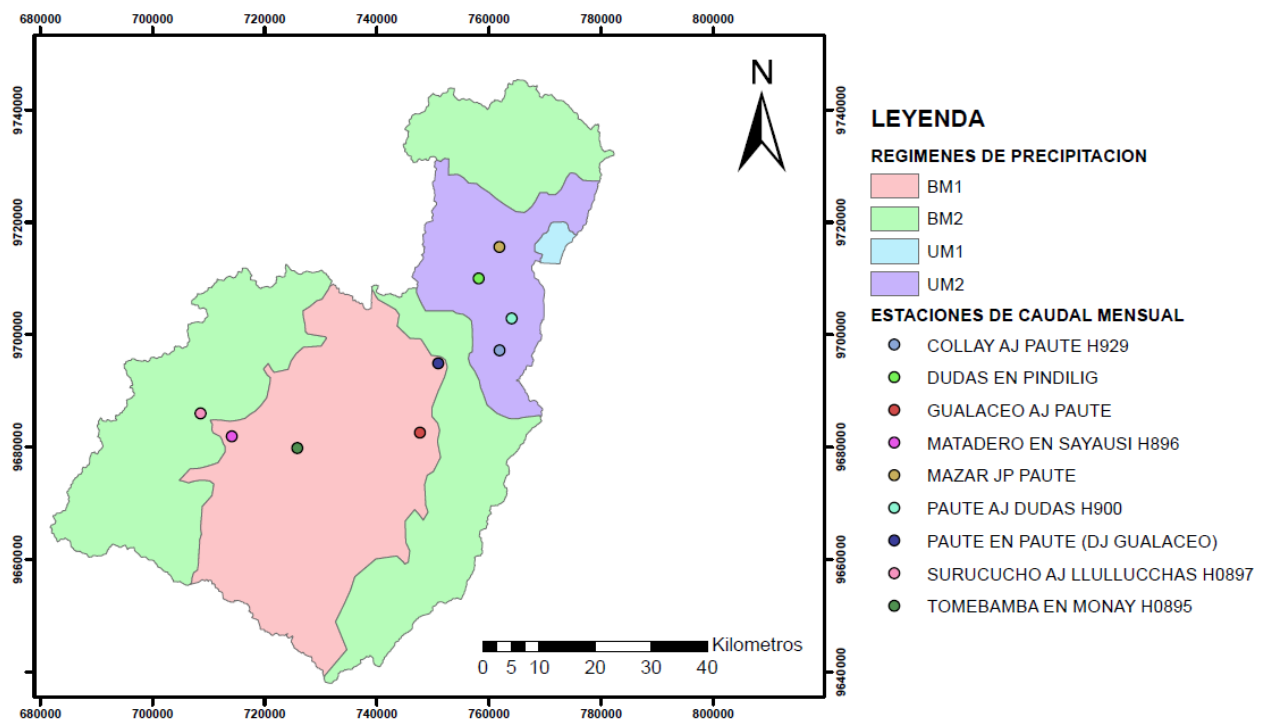


Figura 19. Ubicación de los caudales seleccionados para el estudio y regímenes de precipitación en la cuenca del río Paute.

En las Figuras 20 y 21 se puede observar que para todas las estaciones el criterio es similar, mientras más altos tienden a ser los datos de caudal de la serie de tiempo original de cada estación, los valores predichos se mantienen y empiezan a decaer. La correlación más alta (R^2) para el escenario de 31 índices sucede en Dudas, con 0,32, continua Matadero con 0,28 y Paute AJ Dudas con 0,24, esto prueba que los modelos funcionan de forma similar a pesar de que el rango de valores varía.

Ya que Dudas tiene un caudal muy bajo, Matadero un caudal un poco más alto (mediano), y Paute AJ Dudas el caudal más alto de las nueve estaciones. Para la modelación con 41 índices, se obtiene una muy buena correlación en Dudas con 0,53 lo que confirma la capacidad de los modelos para capturar los promedios. En esta serie de tiempo los valores de caudal extremos ocurren solo en tres ocasiones durante los 15 años de observación usados en el estudio. Dudas obtuvo el mejor comportamiento predictivo en ambos escenarios. El segundo mejor desempeño ocurre en la estación Gualaceo con un coeficiente de 0,31 y Matadero con 0,28, en este caso dos de las estaciones variaron en los índices utilizados lo que represento una gran mejoría para ambos casos.

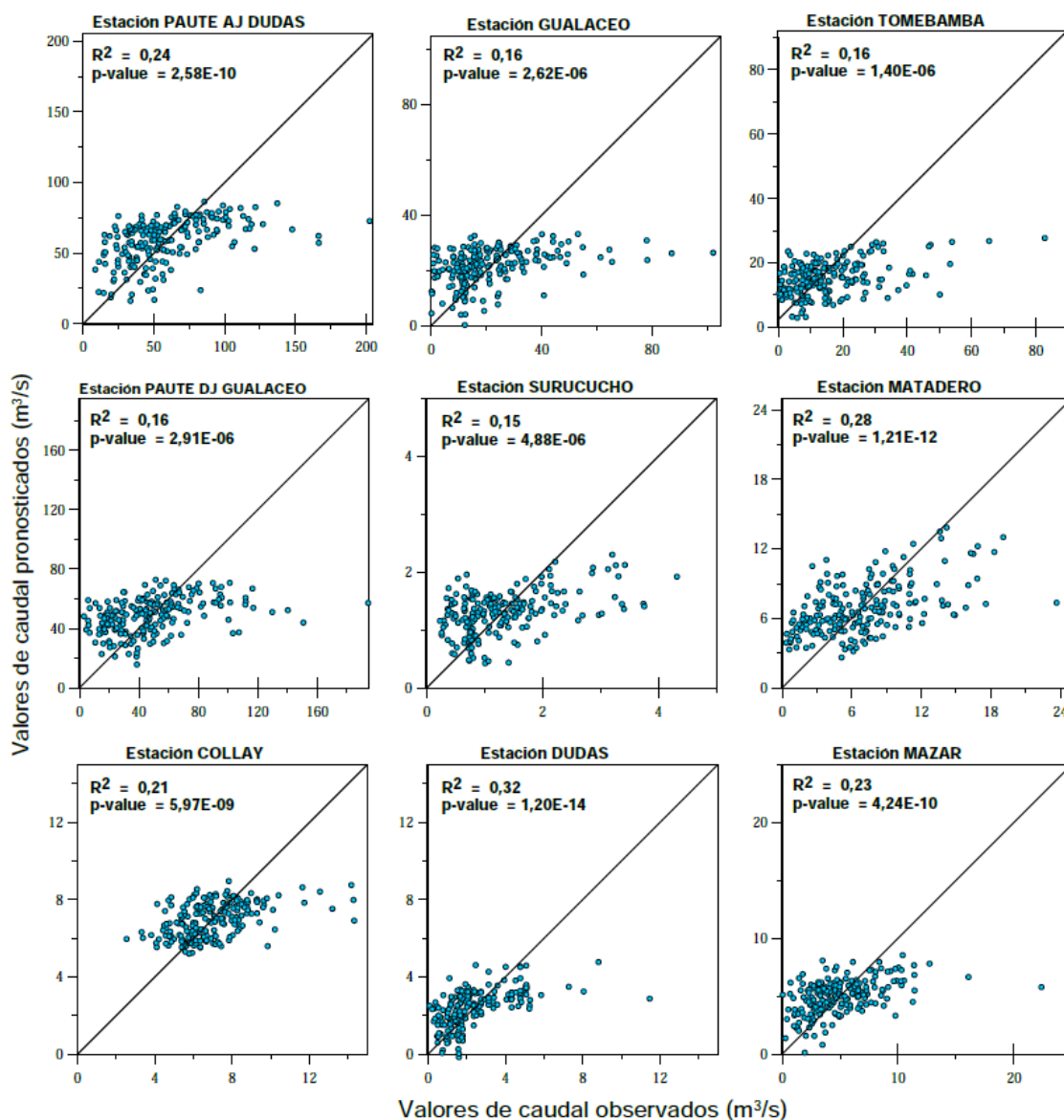


Figura 20. Resultados del análisis de dispersión el escenario utilizando 31 índices con un mes de retraso (etapa de calibración).

El valor más bajo es el de la estación Surucucho en el escenario con 31 índices con 0,15. En el escenario de 41 índices los índices con menor correlación son Paute DJ Gualaceo y Tomebamba ambos con un valor de 0,16. Es notorio que los mejores resultados se obtuvieron utilizando 41 índices para la correlación inicial de los MLR. Tampoco se ha identificado ningún patrón especial en los índices utilizados y la correlación lograda en cada uno de los modelos para las estaciones de caudal.

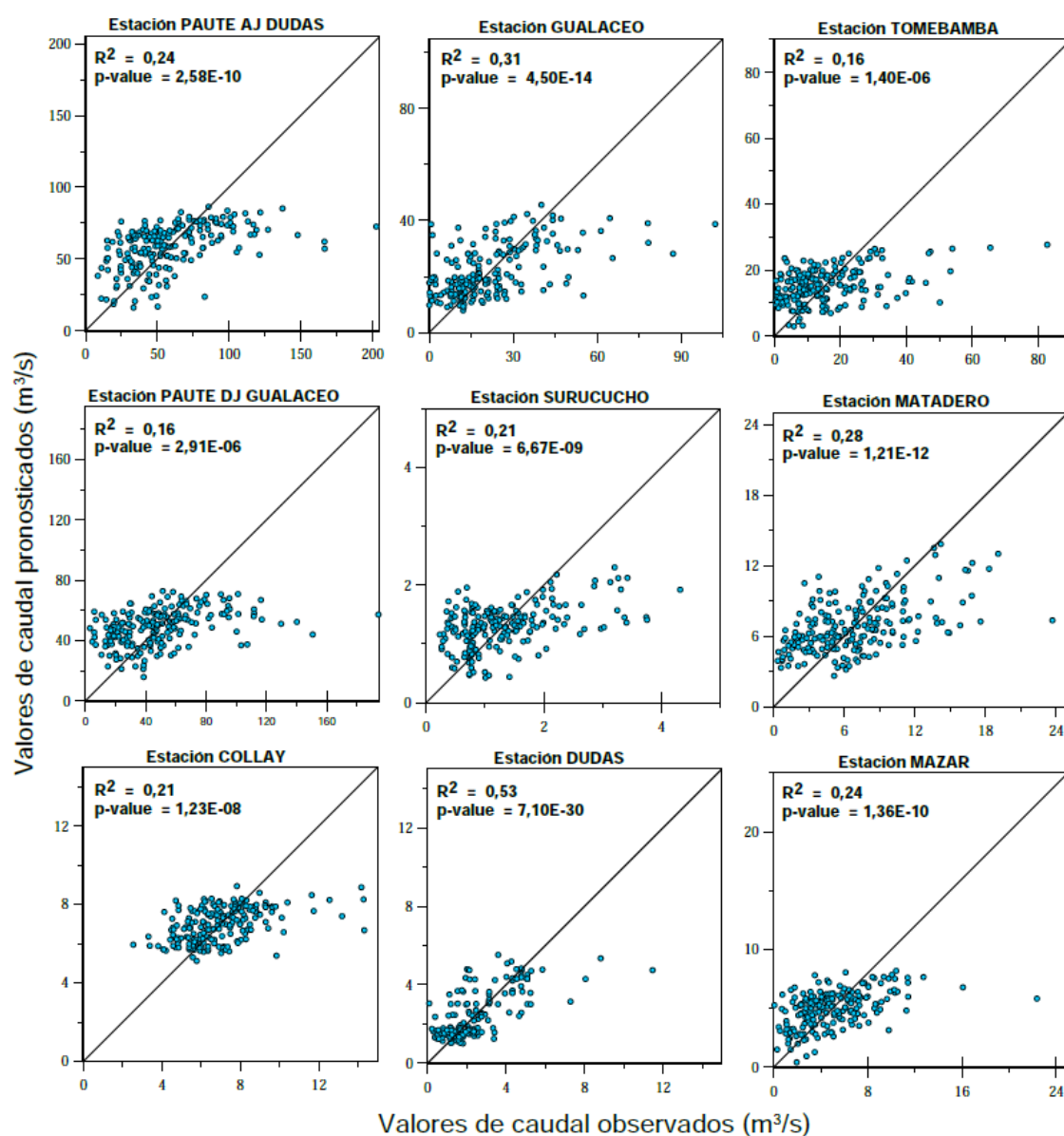


Figura 21. Resultados del análisis de dispersión entre el escenario utilizando 41 índices con un mes de retraso (etapa de calibración)

Al analizar el coeficiente de determinación se puede notar que únicamente el modelo con 41 índices en la estación Dudas es aceptable según la bibliografía. Ya que los valores inferiores a 0,5 indican en un modelo hidrológico que existen una gran varianza del error integrada al modelo y que no se puede explicar por el mismo. Es decir que la utilizar solo los valores de los índices en un modelo de regresión múltiple no se obtienen buenos resultados para los caudales de la cuenca del río Paute. Existiendo variables locales y regionales que podrían integrarse al modelo mejorando la capacidad predictiva (Moriassi et al., 2007). No muestra ninguna tendencia particular al analizar los valores más altos de caudal que suelen influir en el cálculo de R^2 produciendo errores, ya que los mejores valores no solo se producen en estaciones de bajo caudal en ambos escenarios.

El p-value, en todos los casos es estadísticamente significativo manteniendo cifras ínfimas, lo que quiere decir que los modelos pueden replicarse obteniendo respuestas consistentes y que condiciones externas al azar no están afectando su comportamiento. Lo que es importante mencionar debido a la escala espacial que involucra correlacionar los patrones de teleconexión con estaciones de una zona específica del Ecuador. Como ya se revisó en la literatura un p-value menor a 0,05 expresa la importancia del modelo y que este tiene un nivel de confianza de 95%, por lo que los índices de teleconexión son buenos predictores, y el enfoque de mejora debería centrarse en el modelo (Chithra et al., 2016).

3.4.3 Modelación usando análisis de multicolinealidad

El análisis VIF en R Studio se realizó para dos escenarios referentes al relleno de datos, en el caso de los 31 índices sin relleno, los índices representativos con un resultado de menos cinco (colinealidad baja) fueron: PNA, WP, EA/WR, NAO, TSA, PDO, NP, AO, AAO, PWR, CAR, AMOS, QBO, SR, SF y GB. Al observar los resultados del escenario con 41 índices, se determina que son similares al escenario sin relleno, sin embargo, al modelo se suman los índices: EP/NP, NAO (Jones), NOI, CIP, NBRA, SWMRR y AT SST EOF (22 índices de 41).

Ninguno de los modos de variabilidad ni señales relacionadas directamente a ENOS (MEI V2, Niño 1+2, Niño 3, Niño 3.4, Niño 4, SOI, ONI) se ven reflejados en los resultados. De forma indirecta hallamos a NP, lo que indica su alta influencia en el resto de índices y su comportamiento

íntimamente conectado y similar. Esto también ocurre con los índices: AMO UNSMOOTHED, AMM y AMO SMOOTHED, ya que sólo uno se conserva después del análisis VIF. Después de realizar los modelos para cada estación y para los escenarios de 31 y 41 índices, en la Figura 22 se observa la frecuencia de aparición de los índices en los modelos de regresión lineal múltiple.

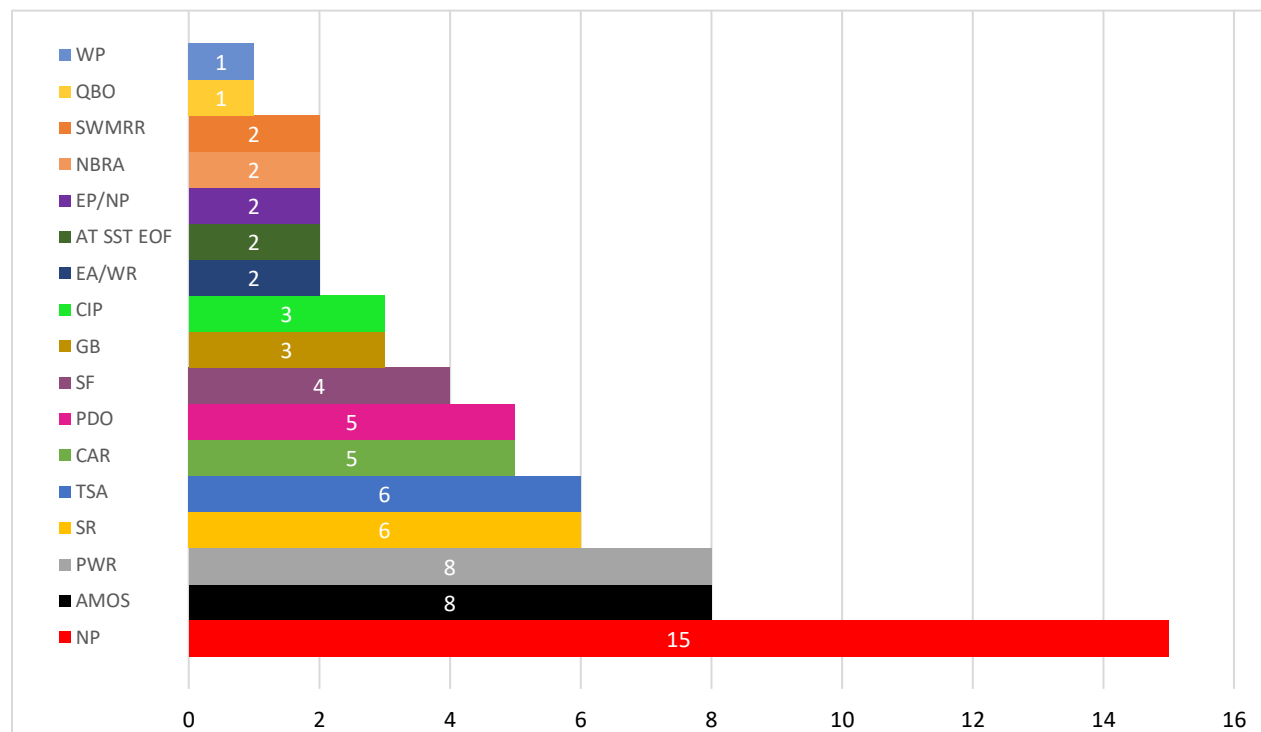


Figura 22. Frecuencia de aparición de los índices de teleconexión en los MLR a nivel general en el escenario VIF.

En la Figura 22 se puede observar que el índice más frecuente en todos los modelos es NP (15), seguido de AMOS (ocho) y PWR (ocho). QBO y WP en cambio son los índices que solo son utilizados una vez dentro de todos los modelos creados. Estos resultados concuerdan con el escenario utilizando retrasos en las series de tiempo, NP, PWR, QBO y WP mantienen un comportamiento similar en frecuencia.

En un estudio realizado en Colombia, sobre la variabilidad climática, al determinar la correlación de múltiples índices de teleconexión con las series de precipitación de la Amazonia; Se halló que



NP presentaba un buen grado de correlación en los resultados con casi todas las estaciones de la zona (Montealegre Bocanegra, 2009). Esto concuerda con la frecuencia de aparición de este índice en nuestro estudio. Una posible explicación a esto podría ser que este patrón de teleconexión está dominado por las variaciones interanuales de los eventos El Niño y La Niña, además de que influye directamente en la corriente del chorro del Pacífico (Espino Sánchez, 2014). En otoño e invierno, provoca un debilitamiento en el sistema de alta presión que resulta en una disminución en el flujo hacia el Ecuador, además también modifica los vientos y TSM. Estos cambios se perciben como variaciones en las contribuciones de masas de agua en los primeros 300m, es decir, el flujo agua subártica (Valencia Gasti et al., 2015).

En el caso de PWR, se ve influenciado por las condiciones de El Niño que reducen la intensidad de los vientos alisios de este patrón y debilita el afloramiento en el Ecuador (en los periodos de diciembre a febrero y marzo a mayo). Se produce un calentamiento anómalo de la capa superficial oceánica frente a las costas de Ecuador y Perú. A su vez se reduce el contraste de la TSM a lo largo del Ecuador, lo que reduce aún más los vientos alisios, esto acelera el desarrollo del Niño (Carreric, 2020). La alta correlación con las estaciones con el índice PWR, podría tener su explicación en la ausencia del índice el Niño entre las posibles variables del modelo. Al no estar presente, PWR toma su lugar ya que está vinculado al mismo índice y se ve influenciado de forma indirecta en su funcionamiento.

El índice QBO tiene por su naturaleza un comportamiento poco predecible que requiere de múltiples variables para integrar en un solo concepto. Pocos estudios se han realizado en el país sobre su influencia, se sabe que tiene un retraso de cuatro a 12 meses de manifestación. Sin embargo, múltiples estudios realizados en Colombia (parte de la región ecuatorial), demuestran que la señal de este índice de los vientos estratosféricos, no influye de forma directa o lineal en la precipitación. Por tanto en los caudales, de la zona, y en caso de hacerlo a nivel de cuencas no supera el 10% de explicación al comportamiento hidrológico de la misma (Aguirre et al., 2018; Yustres Quintero et al., 2020; Beltrán & Díaz, 2020; Zuluaga et al., 2000).

Esto se puede confirmar en la presencia casi nula de este índice dentro de las correlaciones realizadas para las estaciones de la cuenca del río Paute. El patrón WP no tiene ninguna influencia en la zona por su localización eso explica su aparición esporádica.



El índice AMM en el caso de las correlaciones ocurridas en el escenario con retrasos, estas ocurren porque la zona de influencia directa de esta señal es el Ecuador. Un estudio que analizó las proyecciones de eventos climáticos en el Ecuador continental probó que existe una relación muy alta con todos los modelos predictivos y el índice AMM sobre todo en la zona de los Andes y la región Amazónica (Chimborazo, 2018). La influencia del índice AMM sobre el Ecuador se ve reflejado por un gradiente de temperaturas que sucede en la TSM y la variación en la fuerza de los vientos superficiales que cruzan todo el Ecuador. Estos vientos refuerzan a su vez el gradiente ya mencionado de la TSM. Esto afecta subsecuentemente el movimiento meridional de la ZCIT y su banda de fuertes lluvias y nubosidad lo que lo vincula con el comportamiento de los caudales (Loua et al., 2019).

De forma similar la señal AMO funciona en parte sobre el Ecuador, aunque varios estudios han determinado que la relación entre este índice y las precipitaciones es complejo dependiendo de la zona de funcionamiento y de su fase (Knudsen et al., 2011; Knight et al., 2006; Ferrero Martín, 2019; Secretaría de Medio Ambiente y Ordenamiento Territorial, 2021). Por lo que la justificación más apropiada sobre su alta correlación es la ausencia del índice AMM y la tendencia acoplada a la serie de tiempo de valores originales.

En la Tabla 5, se puede observar que los modelos en su mayoría se mantienen dentro de un rango de uno a cinco índices (10 modelos de 15 en total). Siendo el mínimo uno y el máximo número de variables en el modelo 13. Cambia en la mayoría de estaciones dependiendo del número de índices inicial al aplicar el análisis VIF (41 o 31). Estos modelos fueron usados para la validación de este escenario de predicciones con el 20% de datos, sumado el valor del intercepto.

Tabla 5. Modelos de regresión lineal múltiple para cada estación de caudal usando el método VIF.

ESTACIONES	ESCENARIO	N° DE VARIABLES	INDICES
PAUTE AJ DUDAS	31	7	AMOS + CAR + NP + PDO + SF + SR + TSA



			AMOS + AT SST EOF + CIP + EP/NP + GB + NBRA + NP + PDO + PWR + SF + SR + SWMRR + TSA
	41	13	
GUALACEO	31/41	1	NP
TOMEBAMBA	31	1	PWR
	41	5	TSA + NP + PWR + CIP + SWMRR
PAUTE DJ GUALACEO	31	9	AMOS + CAR + EA/WR + GB + NP + PDO + PWR + QBP + TSA
	41	5	AMOS + CAR + EP/NP + NP + PDO + PWR + SR + TSA + WP
SURUCUCHO	31/41	2	NP + PWR
MATADERO	31	9	AMOS + EA/WR + GB + NP + PDO + PWR + SR + TSA + WP
	41	1	PWR
COLLAY	31	3	CAR + NP + SR
	41	2	CAR + SR
DUDAS	31	1	NP
	41	7	AMOS + AT SST EOF + CIP + NBRA + NP + SR + TSA
MAZAR	31/41	3	AMOS + NP + SF

Elaborado por: María Daniela González

En las Figuras 23 y 24 se presentan los resultados de las predicciones realizadas bajo el análisis VIF para la multicolinealidad. El escenario usando 31 índices da resultados insatisfactorios, aunque aún se mantiene la tendencia, la precisión se ve reducida, tanto para datos pequeños o muy altos. La estación Tomebamba visualmente presenta el peor resultado con predicciones que no están correlacionadas en absoluto con la serie de tiempo real, usando el índice PWR. Para el escenario con 41 índices los resultados mejoran para casi todas las estaciones, ya que Matadero presenta malos resultados al igual que Tomebamba con el índice PWR, teniendo solo a este como predictor del modelo. La serie de tiempo predicha de Dudas encaja bastante bien con la serie original, visualmente es el mejor resultado, como en los anteriores escenarios de modelación.

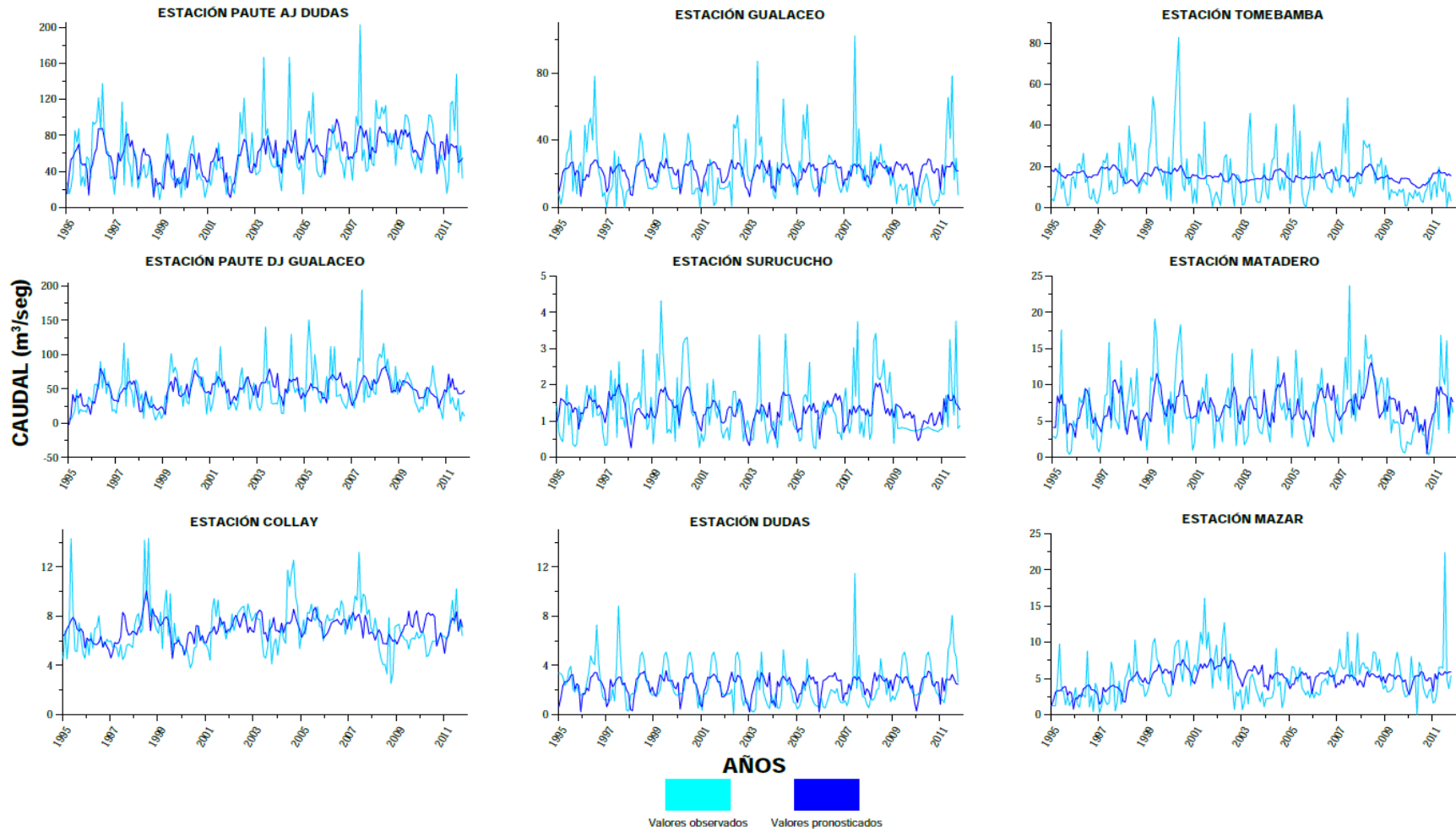


Figura 23. Resultados del escenario utilizando 31 índices y análisis de multicolinealidad (etapa de calibración).

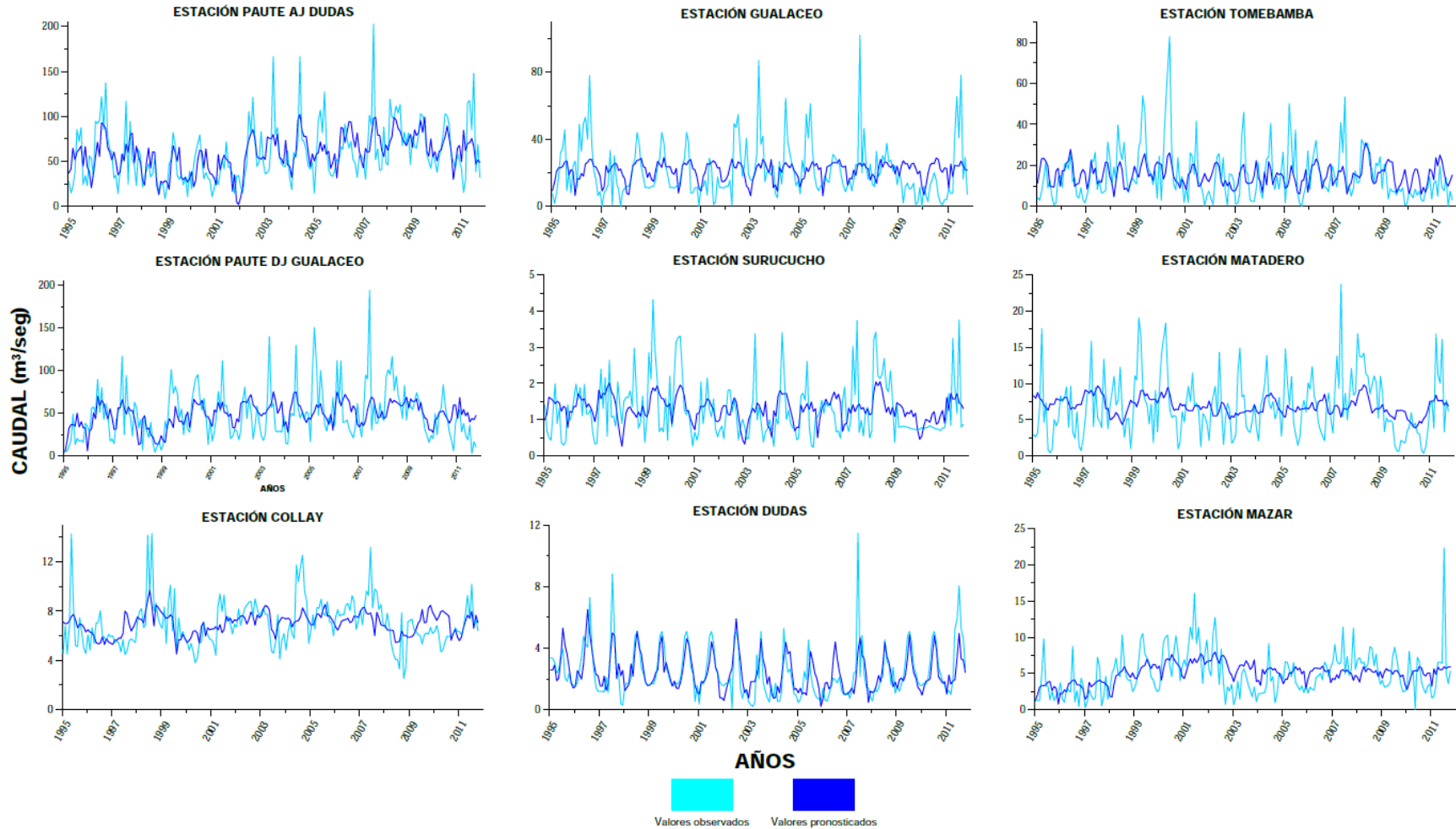


Figura 24. Resultados del escenario utilizando 41 índices y análisis de multicolinealidad (etapa de calibración).

De acuerdo a lo que se puede observar en la Figura 25 con el escenario de 31 índices y el análisis VIF, la correlación más alta la alcanza Paute AJ Dudas con 0,34, y la más baja la tiene Tomebamba con 0,03. El p-value en todos los casos es ínfimo, es decir que todos los modelos son estadísticamente representativos.

El comportamiento es similar a los resultados anteriores, a medida que aumenta el valor de la serie de tiempo original, la serie predicha tiende a mantenerse en el mismo nivel promedio o a descender. En la Figura 26 la estación Paute AJ Dudas es la segunda mejor correlación con 0,41, solo superada por Dudas con 0,52. La estación con menor correlación es Matadero con 0,08 y Gualaceo con 0,11.

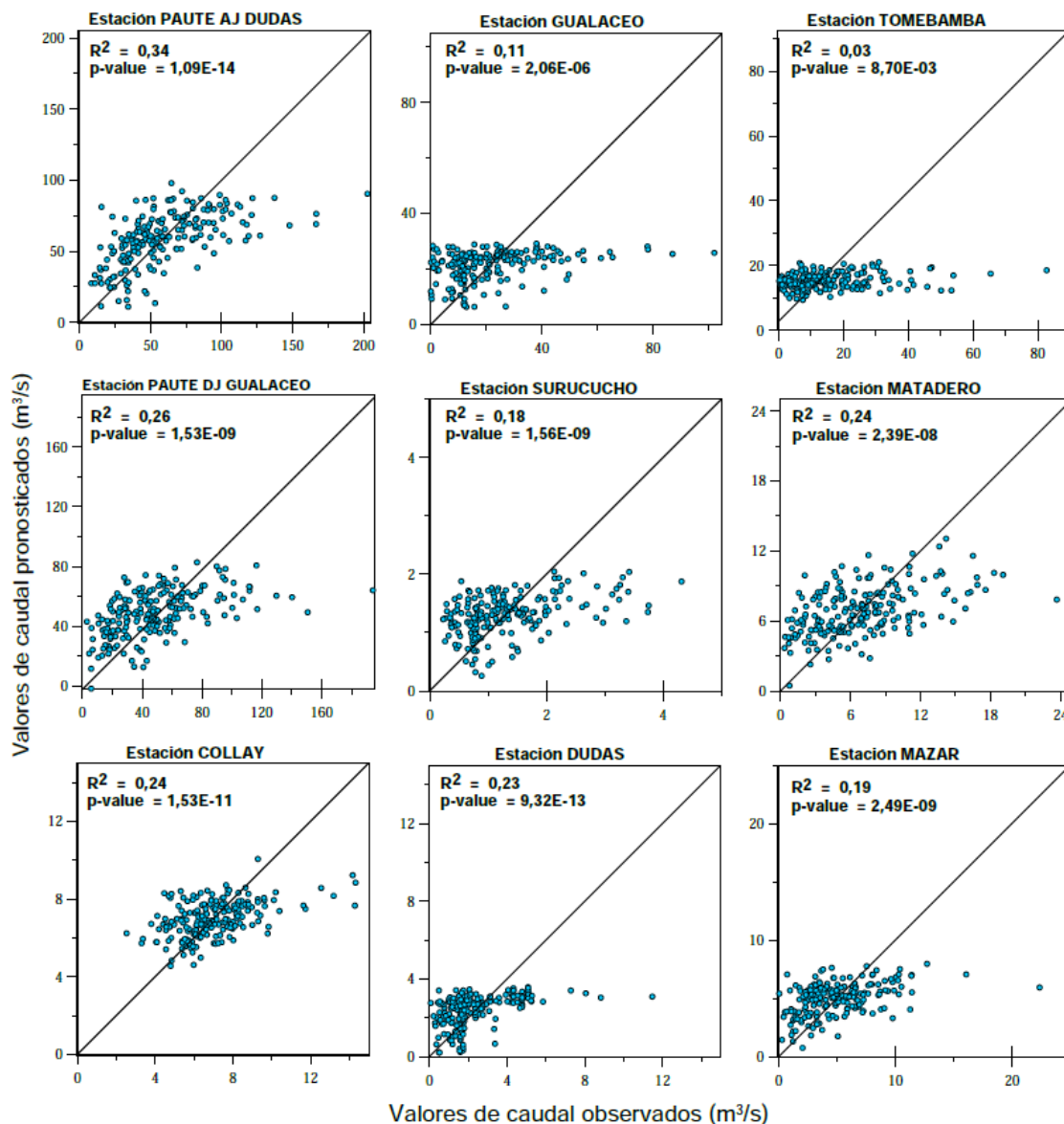


Figura 25. Resultados del análisis de dispersión entre la serie de tiempo predicha utilizando 31 índices con el análisis de multicolinealidad (etapa de calibración).

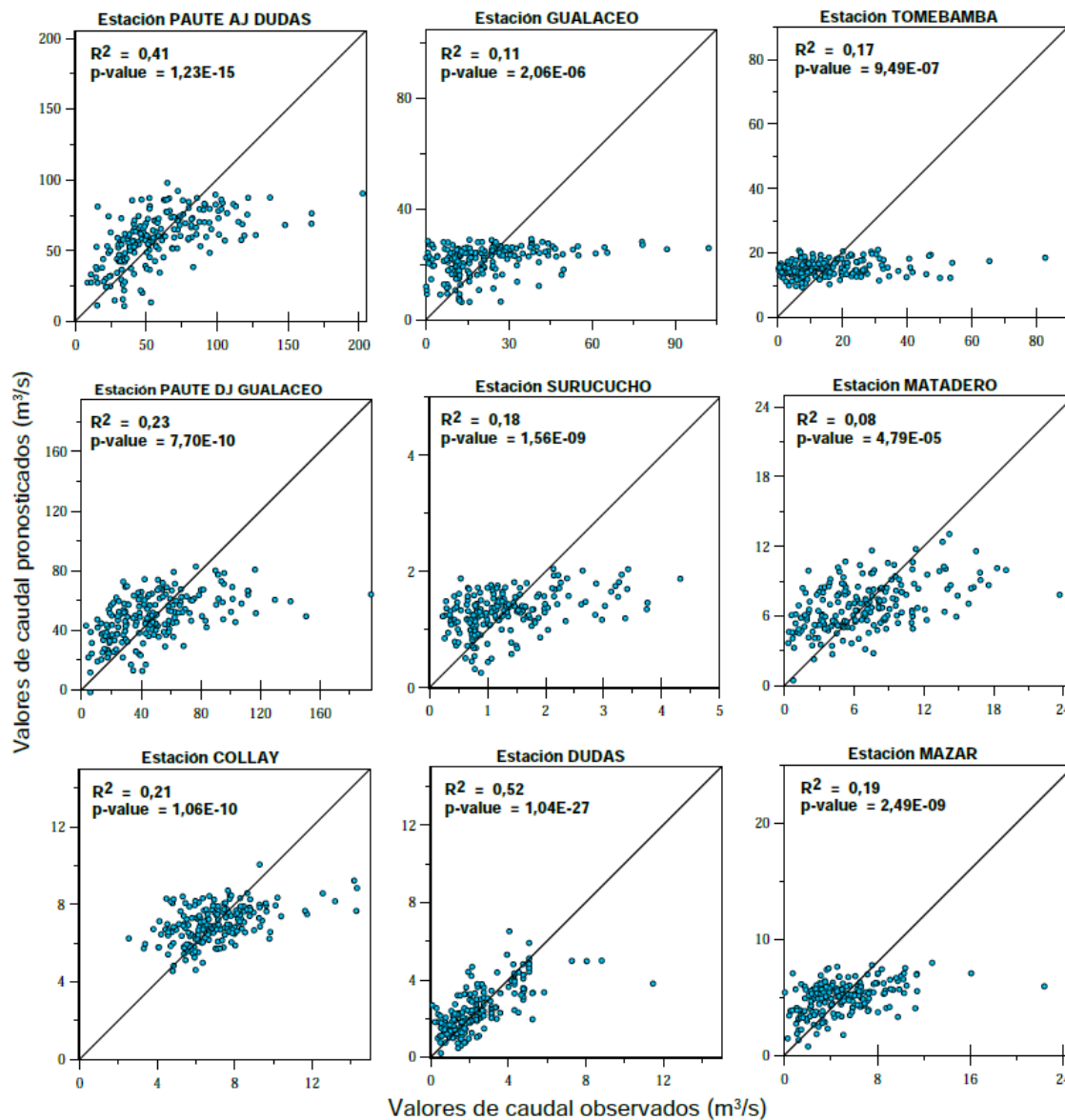


Figura 26. Resultados del análisis de dispersión entre la serie de tiempo predicha utilizando 41 índices con el análisis de multicolinealidad (etapa de calibración).

Al comparar estos resultados con el número de índices usados en cada modelo se puede determinar que, en una misma estación, los mejores resultados se obtienen al aumentar el número de variables utilizadas para la creación del modelo, sin embargo, esto no interfiere entre estaciones, ya que estaciones que utilizan un solo índice pueden obtener una mejor correlación que otras que utilicen más de cinco para el mismo propósito. El comportamiento aquí está directamente ligado entre



cómo se acoplan los valores de la serie de tiempo del índice con la serie de datos del caudal de forma individual. De igual manera no existe un patrón claro sobre los valores altos o bajos, esto no tiene interferencia sobre el resultado de cada modelo, ya que no se puede observar un patrón.

3.5 Análisis estadístico de resultados mediante métricas

En la Tabla 6 se observa a modo de resumen la gran diferencia entre los resultados obtenidos de la calibración usando un 80% del total de los datos de cada serie de tiempo y la validación con 20%. Las Figuras usadas para la evaluación de desempeño a través de métricas pertenecen a la etapa de calibración, los resultados de validación se encuentran en el Anexo 5, Figuras 68 a 71 (NSE, KGE, R^2 y RSME respectivamente).

La máxima correlación alcanzada en la calibración es la de Dudas y Paute AJ Dudas (por la forma en que estas estaciones mantienen en la mayor parte de la serie de tiempo la tendencia sin llegar a eventos extremos) con 0,32 en el caso de la validación ocurre en las mismas estaciones, pero con un valor menor de 0,23. NSE alcanzan 0,32 para Dudas y Paute AJ Dudas, sin embargo, para la validación el valor más cercano a uno y el único positivo es 0,03 de la estación Matadero, seguido de -0,18 en Mazar. Para KGE los resultados son similares, en Matadero se obtiene 0,08 y en Mazar -0,07 siendo estos los resultados más cercanos al óptimo (uno). El RMSE más bajo lo obtienen Dudas y Paute AJ Dudas, se puede confirmar que los mejores modelos sucedieron en estas estaciones.

Los peores resultados ocurren en Tomebamba y Paute DJ Gualaceo, con R^2 de 0,16, NSE del mismo valor para el 80% de datos. En el caso de la validación se obtienen 0,01 y 0,02 en Collay y Paute DJ Gualaceo, aunque los valores en general se mantienen bajos para la mayoría de las estaciones. NSE al igual que KGE presenta resultados negativos en casi todas las estaciones, lo que quiere decir que para la cantidad de 50 datos el modelo es menos efectivo prediciendo que al usar el promedio mensual de la serie para el mismo propósito. En el caso de NSE Surucucho y Collay tienen -16,02 y -12,19 siendo los peores modelos.

KGE muestra mejores resultados abordando los defectos de NSE al no sobreestimar valores por el cálculo al cuadrado del mismo, los peores rendimientos son Surucucho y Paute DJ Gualaceo con



-0,75 y -0,57, que, aunque indican un mal ajuste del modelo son cantidades más cercanas al rendimiento óptimo que las indicadas por NSE. EL RMSE más alto ocurre en las estaciones Paute DJ Gualaceo y Gualaceo con 34,37 y 13,61 respectivamente indica una gran dispersión de valores entre la serie predicha y la observada. Se puede determinar que Paute DJ Gualaceo, Collay y Surucucho fueron las estaciones con modelos de menor capacidad predictiva.

Tabla 6. Resultados estadísticos de las métricas aplicadas a los modelos para evaluar su funcionamiento.

MÉTRICAS/ ESTACIONES	CALIBRACIÓN				VALIDACION			
	R ²	NSE	KGE	RMSE	R ²	NSE	KGE	RMSE
PAUTE AJ DUDAS	0,32	0,32	0,36	1,37	0,23	-3,81	-0,18	3,17
GUALACEO	0,19	0,19	0,19	14,70	0,03	-1,75	-0,39	13,61
TOMBAMBA	0,16	0,16	0,15	11,52	0,07	-1,30	-0,11	8,37
PAUTE DJ GUALACEO	0,16	0,16	0,13	27,31	0,02	-1,63	-0,57	34,37
SURUCUCHO	0,19	0,19	0,20	0,72	0,09	-16,02	-0,75	1,81
MATADERO	0,21	0,21	0,23	3,73	0,11	0,03	0,08	3,67
COLLAY	0,18	0,19	0,19	1,72	0,01	-12,19	-0,15	4,50
DUDAS	0,32	0,32	0,36	1,37	0,23	-3,81	-0,18	3,17
MAZAR	0,23	0,23	0,26	2,68	0,16	-0,18	-0,07	3,51

Elaborado por: María Daniela González

Un estudio realizado en Estados Unidos sobre la importancia de la calibración de datos en la predicción de caudales obtuvo resultados similares a los de nuestro estudio en lo que respecta a calibración y validación y al desempeño NSE. Ya que para todos los escenarios de predicción la calibración mostró mejoras en la precisión del pronóstico, aunque también menciona que disminuir el sesgo de error de los datos influye en esos resultados (Shi et al., 2008).

Generalmente sucede que los resultados en las métricas de desempeño del modelo disminuyen en el caso de la validación por esto se recomienda aplicar métricas de evaluación en ambos periodos (calibración y validación), para medir la robustez del modelo (Muhammad et al., 2017). Esto a su vez sucede en modelos hidrológicos porque en un periodo de uno a cinco años el efecto de la diversidad de los caudales y los índices no es considerable en el modelo (no captura la variabilidad hidrológica temporal).

También es importante señalar que para periodos de calibración de más de 15 años, como en caso de nuestro estudio, el promedio de la eficiencia de la validación debería ser similar al promedio de



la eficiencia de la calibración y esto es señalado en múltiples estudios (Merz et al., 2009; Patel et al., 2016). Lo que por consiguiente indica que los modelos que solo dependen de los índices como variable no tienen robustez.

Un estudio comparativo realizado en Irán, demostró que los modelos multicapas o algoritmos de inteligencia artificial se desempeñan mejor que el modelamiento convencional de regresión lineal múltiple para predicción de precipitación usando múltiples índices (AMO, AMM, BEST, Niño 3.4, Niño 4, NTA, SOI, TNA). Sin embargo, esto se extiende también a la predicción de caudales, esto puede suceder porque las entradas de un MLR no consideran contextos ambientales, sino simplemente la serie de valores. Al igual que en nuestro estudio también se observó que el modelo de regresión lineal no predice la desviación estándar o fluctuaciones, es decir no pronostica años húmedos o temporadas de sequía (Choubin et al., 2016; Niu et al., 2019; Tsakiri et al., 2018).

Se realizaron un total de 108 modelos, cuatro para el escenario ACP, dos componentes para 31 índices y dos para 41 índices (70 y 90% de varianza explicada), dos modelos con análisis VIF (31 y 41 índices), y seis modelos con retrasos de uno, dos y tres meses para 31 y 41 índices. Un total de 12 modelos para cada estación, en la Figura 27 se puede apreciar la cantidad de apariciones de los índices en los escenarios VIF y con retrasos mensuales. El índice con mayor número de apariciones es: NP (51), seguido de AMM (42), PWR (38) y WHWP (24). En comparación con los índices que tienen una gran correlación con los componentes en ACP, coinciden: PWR y AMM.

Los índices que no se ven reflejados en la correlación con los caudales de los 41 totales en ningún escenario del estudio son: AO, BEST, ENSO, GIAM, MEI V2, NAO, Niño 3, Niño 3.4, ONI, SOI, TPI (IPO), en los modelos propuestos la correlación de estos índices no supera 0,20. La mayoría están asociados al fenómeno ENOS, por lo que se puede determinar que las correlaciones enfocan los valores de las relaciones directas o indirectas de ENOS con pocos índices que mantienen esta tendencia (PWR, NP).

El resto de índices tiene diferencias de escala en el funcionamiento, que no reflejan efectos en las series mensuales (TPI (IPO), GIAM), AO y NAO mantienen una ubicación diferente a la de la zona de estudio y su región. En el caso de NAO es importante determinar que a diferencia de NAO (Jones) que, si se ve reflejado en el estudio, este último mantiene observaciones instrumentales

entre 1823 - 1996 que integran mejor la variabilidad anual y decadal, por lo que esto pudo causar la diferencia en las correlaciones halladas (Jones et al., 1997).

Esto también pudo ocurrir debido a la actuación focalizada de cada índice y la pérdida de visibilidad en los efectos al aumentar la distancia de estudio. De acuerdo a Ballari et al., (2020) la distancia media de: SOI, Niño 3.4, mantiene 0,5 en 100 km. MEI, el Niño 3 y ONI alcanzan este valor en una distancia más reducida. AMO, AMM, TP SST EOF y TSA alcanzan 0,5 en un radio de 100 km (alta correlación).

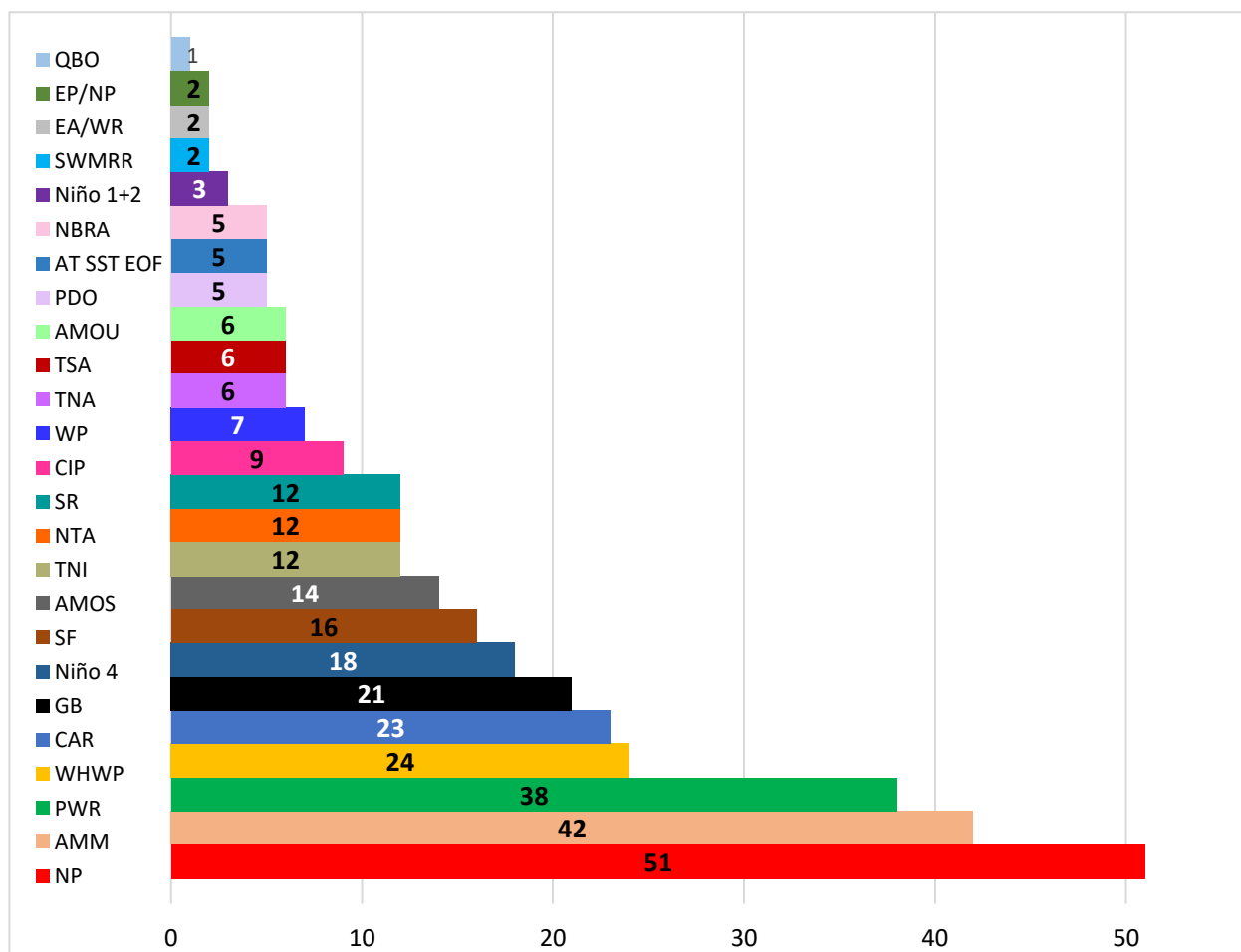


Figura 27. Frecuencia de aparición de los índices de teleconexión en los escenarios VIF y con retrasos.



3.5.1 Valor P o P-value

Como se observa en la Tabla 7 el promedio de todas las estaciones es un valor muy bajo. Es decir que incluso con los resultados insatisfactorios de los modelos generados, estos son estadísticamente muy significativos, aunque estos resultados no se asocian con el buen desempeño de los modelos, ya que las métricas no tienen ninguna relación directa.

Tabla 7. Promedio del valor P de cada estación de caudal con sus distintos escenarios de predicción.

INDICES/ ESTACIONES	P VALUE	
	41	31
DUDAS	5,75E-05	1,83E-05
GUALACEO	1,33E-06	3,32E-03
MATADERO	6,64E-04	8,10E-05
MAZAR	9,11E-07	2,92E-07
SURUCUCHO	2,35E-05	2,15E-05
TOMBAMBA	8,52E-04	6,01E-03
COLLAY	2,59E-04	4,07E-04
PAUTE AJ DUDAS	3,52E-08	4,75E-08
PAUTE DJ GUALACEO	6,11E-03	1,29E-03

Elaborado por: María Daniela González

3.5.2 Eficiencia de Nash-Sutcliffe

En la Figura 28 se pueden observar los resultados de la métrica NSE, Dudas tiene los resultados más altos alcanzando cerca de 0,6, un resultado satisfactorio. De acuerdo a la clasificación de Moriasi et al. (2007), la cual describe el desempeño de un modelo como: muy bueno, bueno, satisfactorio e insatisfactorio si los valores de NSE son mayores a: 0,75, entre 0,65 y 0,75, entre 0,5 y 0,65 y menores a 0,5 respectivamente. El siguiente mejor desempeño NSE es Paute AJ Dudas llegando a 0,40, el ajuste entre las series pronosticadas y las series originales no es bueno, el rango de varianza es muy amplio.

Un estudio realizado en Nueva Gales, Australia que uso los índices de ENOS: MEI, IOD, Niño 3.4, PDO y TPI para predecir caudales de la zona, obtuvo valores de NSE que en su mayoría se mantienen entre 0,15 a 0,55, lo que indica que solamente los índices no parecen ser suficientes como predictores de caudal (Esha et al., 2019). Los mejores modelos ocurren en el escenario con

ACP usando 41 índices y un total de 20 componentes (90% de varianza explicada) también usando los 41 índices con uno y dos meses de retraso. Y en el caso del análisis de multicolinealidad el mejor escenario se presenta al usar 41 índices.

Es decir que, al retirar los índices del relleno, se reduce la variabilidad que puede explicar el comportamiento de los caudales. La tendencia de crecimiento se repite para el escenario con 31 índices y 14 componentes, pero con resultados inferiores.

Los peores casos ocurren en al utilizar 31 índices en el análisis de multicolinealidad, también al usar 31 y 41 índices con tres meses de retraso, es decir que el desfase no ocurre con este tiempo entre los índices y su efecto climático en los caudales. Las estaciones con los resultados menos ajustados son Tomebamba, Paute DJ Gualaceo y Gualaceo.

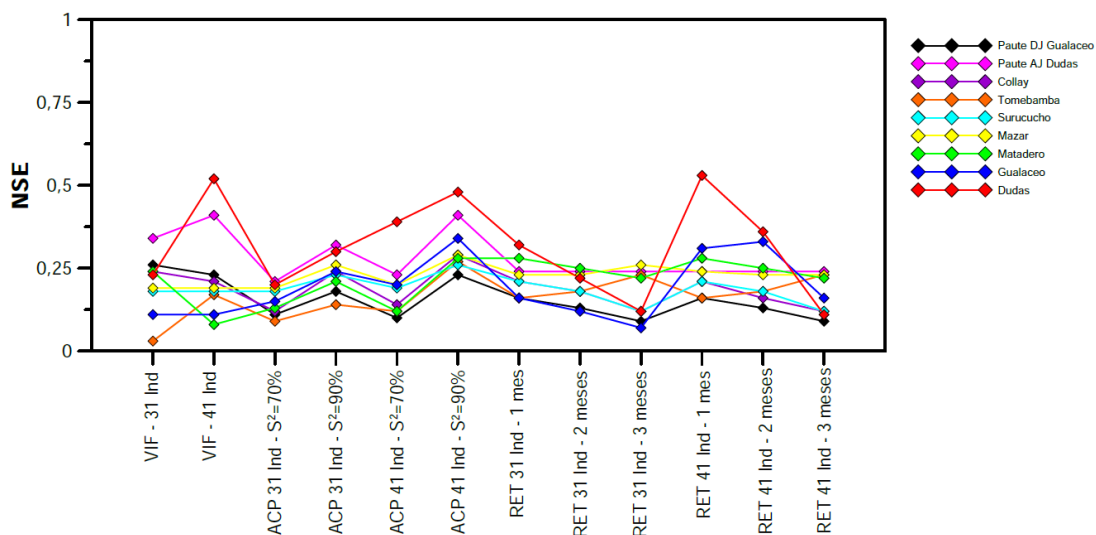


Figura 28. Resultados de NSE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.

3.5.3 Eficiencia de Kling-Gupta

En la Figura 29 se muestran los resultados de KGE, Dudas es la estación que lidera el mejor desempeño entre todos los modelos, conjuntamente con Paute AJ Dudas y Gualaceo, los valores máximos de estas estaciones son 0,61, 0,57 y 0,39 respectivamente, dos llegan al resultado satisfactorio. Los valores llegan a cifras más altas, pero también más bajas, el caso de Tomebamba, Gualaceo y Matadero, entre otros, que llegan a cifras negativas. El valor más bajo obtenido es -



0,15, determinando que tienen modelos inaceptables y que en esos casos el promedio mensual es mejor predictor que los índices usados (Krause et al., 2005).

La tendencia se mantiene en cuanto a rendimiento de cada estación, y en cuanto a escenarios también. El mejor desempeño se logra en la modelación con retrasos de uno y dos meses, al usar ACP con 41 índices y 20 componentes principales y al usar el análisis VIF de multicolinealidad con 41 índices. Es decir que los índices: EP/NP, GIAM, Nao (Jones), NBRA, NOI, TP SST EOF, SW MRR, TPI(IPO), AT SST EOF y CIP, que fueron los 10 excluidos del relleno de datos tienen peso en la variabilidad de los caudales. Esto se demuestra en los análisis de correlación usados para la creación de modelos con análisis VIF y usando desfases mensuales, en los que, en el caso del escenario con 41 índices, aparecen múltiples veces.

Un estudio realizado en Indonesia, que usó los índices: SOI, Niño 3.4 y IOD para analizar los regímenes de caudal de Java, identificó que los resultados de los modelos de regresión múltiple daban mejores resultados de KGE mientras más variables incluía el modelo (Nugroho et al., 2022). Esto se puede verificar en nuestro estudio, al observar que los mejores resultados se obtuvieron con el análisis ACP con 20 componentes principales en el escenario con 41 índices.

Mientras con el análisis de multicolinealidad para el escenario de 41 índices los mejores resultados KGE son: Dudas, Paute AJ Dudas y Paute AJ Gualaceo los cuales integran siete, cinco y 13 variables en sus modelos respectivamente. Para el escenario con 31 índices las mejores estaciones son Paute AJ Dudas, Matadero y Paute AJ Gualaceo, estas contienen siete, nueve y nueve respectivamente.

Y en el caso de los resultados menos ajustados, el número de índices usados en los modelos de cada estación fueron uno, (Matadero en el escenario de 41 índices), uno y cinco para Tomebamba (31 – 41 índices usados), dos índices en ambos casos para Matadero y un solo índice en ambos casos para Gualaceo. Aunque este comportamiento no es una regla, a modo general muestra una clara relación con el índice KGE.

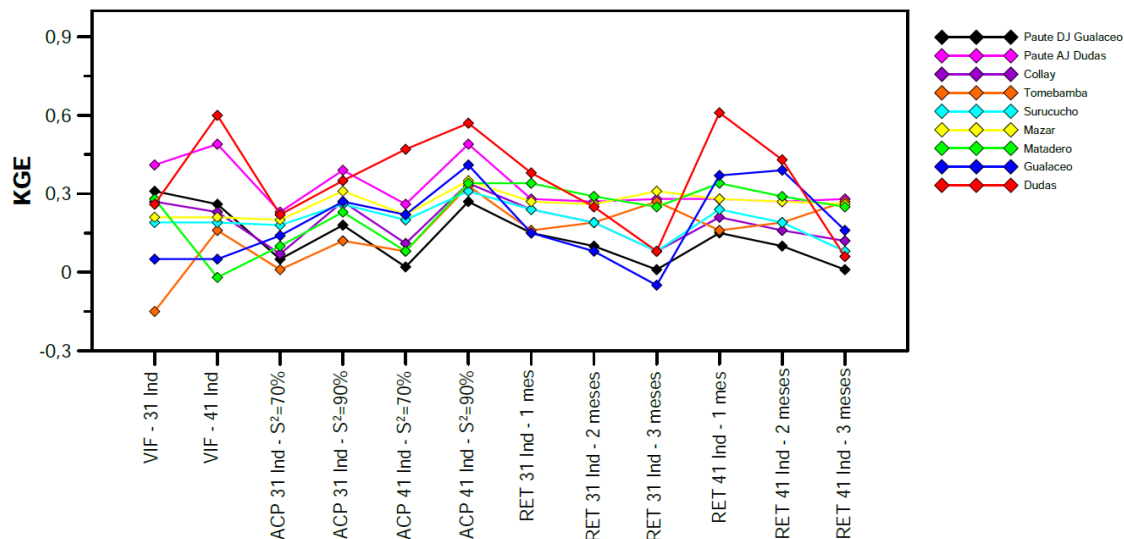


Figura 29. Resultados de KGE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.

3.5.4 Coeficiente de determinación

En la Figura 30 se puede observar el resultado del análisis con el coeficiente de correlación entre las series de tiempo originales y las series predichas para cada caudal. El máximo valor de R^2 llega a 0,53 en la estación Dudas, y el menor valor a 0,03 en Tomebamba. Similar a las otras métricas Dudas demuestra tener el mayor número de modelos que mejor se ajustan para cada escenario. Paute AJ Dudas es la siguiente estación con 0,41 como el valor más alto de R^2 , en el caso de Gualaceo se produce una gran variación ya que mantiene ajustes muy diferentes en los escenarios, su valor más alto es de 0,33 y el más bajo de 0,07.

Dentro del rango del criterio de R^2 , los resultados mayores a 0,85 reflejan un ajuste muy bueno, entre 0,85 y 0,75 es bueno, entre 0,60 y 0,75 es satisfactorio y por debajo de 0,60 es malo (Pérez et al., 2019). Es decir que en este estudio ningún modelo llega a ser satisfactorio, los modelos no logran explicar la variabilidad de los valores en las series de tiempo observadas.

En cuanto al desempeño de cada escenario, los modelos realizados con un mes de retraso y usando el análisis VIF con 41 índices en ambos casos son los escenarios que mejor capacidad tienen de predecir los valores de caudal. En el caso del ACP con 41 índices y 20 componentes todas las estaciones con excepción de Dudas, muestran tener el mejor resultado en este escenario con un rango entre 0,23 y 0,48. Con la modelación usando ACP 31 índices y 14 componentes también

ocurre una mejora notoria (tendencia) en todas las estaciones, aunque los ajustes más altos no se reflejen en este escenario. Esta observación se repite en NSE y KGE.

Este comportamiento en comparación con el uso reducido de variables para los MLR en los otros escenarios con mejor o igual coeficiente de determinación podría indicar la presencia de ruido en los componentes al necesitar de 14 y 20 componentes mínimo para integrar toda la variabilidad de los índices, siendo que en los otros casos basta de apenas cinco índices o menos (VIF) para conseguir los resultados más óptimos. Los peores modelos se desarrollan en el análisis VIF con 31 índices, y en los dos escenarios al usar tres meses de retraso.

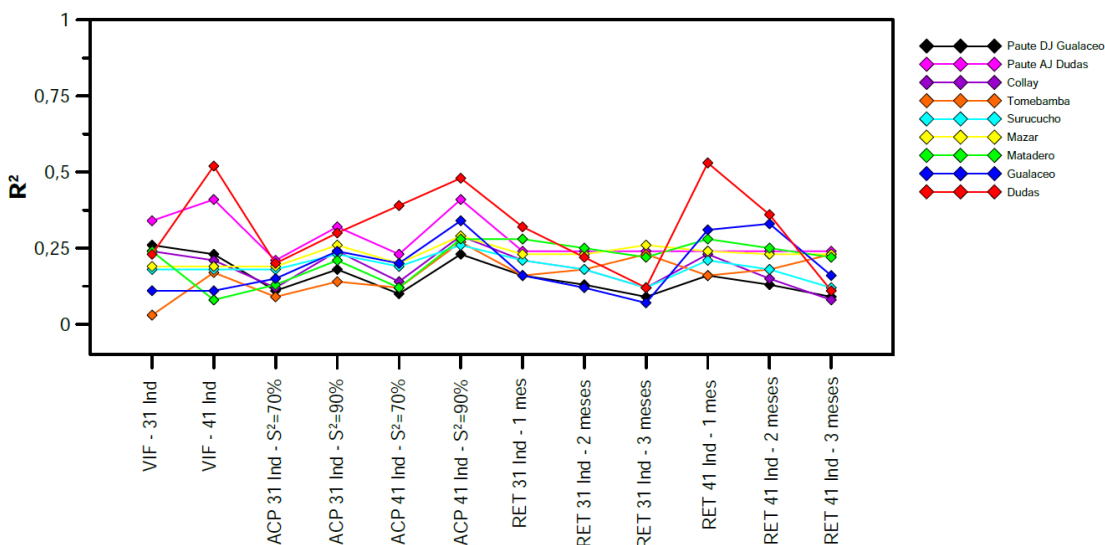


Figura 30. Resultados de R^2 para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.

3.5.5 Error cuadrático medio

Los resultados de RSME para cada escenario de modelación se presentan en la Figura 31. Se puede notar una diferencia pronunciada en el valor de esta métrica que está directamente relacionado con el valor de las series de tiempo. Ya que divide en tres grupos a las estaciones, las estaciones con el caudal más alto son las que a su vez presentan un gran RSME, así mismo las estaciones con caudal mediano y bajo.



No se presenta ninguna tendencia relacionada con el desempeño de las estaciones ya que Dudas y Paute AJ Dudas fueron las estaciones con mejor ajuste en la mayoría de casos y ambas mantienen un RSME muy distinto. En el caso de Dudas su valor más bajo es 1,15 para el caso del escenario VIF con 31 índices y usando retraso de un mes con 41 índices y el valor más alto se presenta usando un retraso de tres meses con 41 índices con 1,56. Para la estación Paute AJ Dudas el resultado más óptimo tiene un error de 24,35 en el escenario de ACP con 41 índices y 20 componentes, y el peor ajuste con 28,22 sucede al usar ACP con 31 índices y ocho componentes. El número de variables no parece marcar una diferencia en el error de cada estación.

Los resultados más óptimos ocurren en la estación Surucucho, que tiene un rango de 0,68 a 0,73 de RSME, en ACP con 41 índices y 20 componentes y retraso de tres meses con 41 índices respectivamente. Esta estación contiene los valores de caudal más bajos entre todas las series seleccionadas. Los modelos con más error se producen a su vez en la estación Paute DJ Gualaceo, que contiene el segundo promedio más alto de valores de caudal después de Paute AJ Dudas, con un rango entre 25,56 hasta 28,32 en los escenarios de VIF con 31 índices, y ACP con 41 índices y 10 componentes.

Aunque los valores de RSME entre las estaciones Paute DJ Gualaceo y Paute AJ Dudas son muy similares, sus desempeños son completamente distintos, desligando al funcionamiento del modelo de la magnitud de los valores de las series originales de tiempo de los caudales.

Al comparar los valores se observa que el escenario ACP con 41 índices y 20 componentes (90% de varianza) contiene en la mayoría de las estaciones el menor valor de RSME. Es decir que las predicciones más cercanas ocurren en esta modelación, esto se confirma con los resultados de KGE, NSE y R^2 . En el caso de ACP en su modelo influyen varios índices que no coinciden en su aparición con los modelos del resto de escenarios en el estudio.

Considerando que los resultados de esta metodología son los mejores producidos en el estudio, se puede inferir que, aunque existe la posibilidad de ruido o multicolinealidad. Los índices que se integran en estos modelos contienen más variabilidad explicada de los caudales del río Paute. Y estos no necesariamente se ven vinculados a ENOS o sus distintos modos de variabilidad, o incluso a índices que no tienen una ubicación de funcionamiento sobre el Ecuador, sino más bien que

generan efectos indirectos. Esto sucede porque cada señal de teleconexión tiene distintas formas de influir en el clima no siendo necesariamente en la precipitación o hidrología de la zona.

Se puede notar una pequeña mejora entre usar el 70% de varianza y el 90%, al usar este último porcentaje, RSME disminuye. También es importante mencionar que, aunque la mejoría no es cuantificable porque varía en cada estación, existe un aumento significativo del desempeño en todos los modelos ACP verificable en las métricas al variar entre 70 y 90% de varianza explicada para la selección de componentes.

Fuera de esta observación no se puede determinar ningún otro patrón especial bien definido como en el resto de métricas. Los valores no se relacionan de caudal a caudal, sin embargo, dentro del mismo modelo el rango de RSME no tiende a tener diferencias notorias, sino que se mantiene a no más de cinco cifras de distancia ya sea el valor mayor o menor.

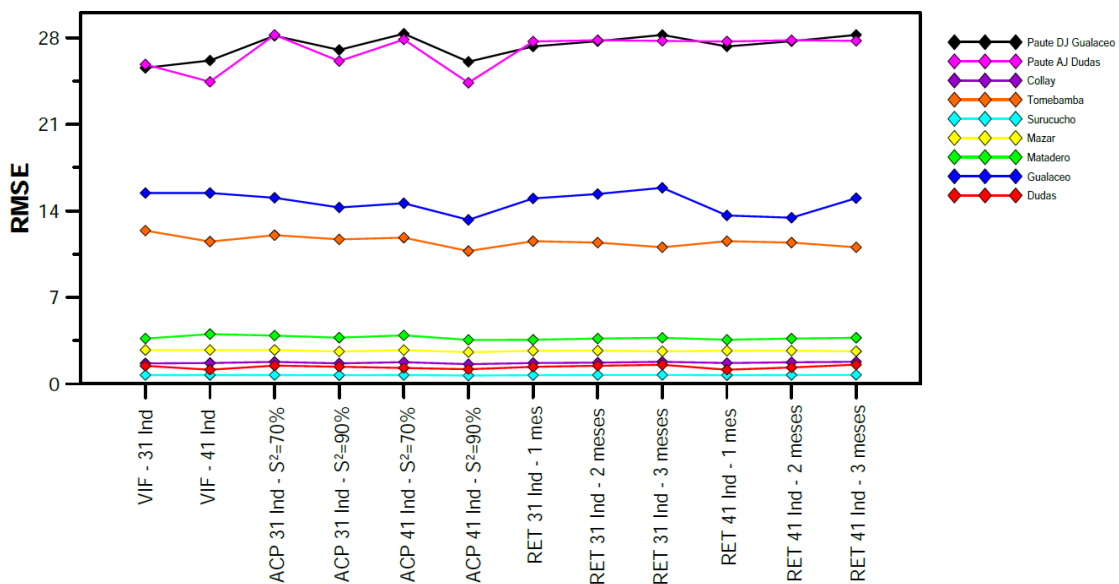


Figura 31. Resultados de RMSE para cada escenario de predicción (etapa de calibración) de las estaciones de caudal seleccionadas.



CAPITULO 4: CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

Para cubrir errores causados por el ruido y desfases de tiempo se realizaron los escenarios VIF y con retrasos, un total de 108 modelos, 12 para cada estación. Todo el análisis fue realizado sobre los valores del periodo de calibración para tener resultados representativos. El estudio realizado sobre la cuenca del río Paute usando índices de teleconexión bajo distintos escenarios para comprobar la capacidad predictiva de los mismos, sobre los caudales de la zona; Determinó que los mejores modelos predictivos tienen la capacidad de seguir la tendencia de las series de tiempo originales pero que no integran valores muy bajos o muy altos.

Se identificó que los 10 índices faltantes en el caso de los escenarios realizados con 31 índices presentan una variabilidad perdida dentro del análisis para cada caudal, esto se mide en el peso de los datos, más no en la frecuencia. Los mejores escenarios suceden al utilizar ACP con 41 índices y 20 componentes, y 41 índices con un mes de retraso, las estaciones con mejor desempeño a su vez son Dudas y Paute AJ Dudas debido a su tendencia que no frecuenta a mantener eventos extremos. Mientras que los peores ajusten ocurren al usar 31 índices con el análisis VIF o 31 y 41 índices con tres meses de retraso en las estaciones Paute DJ Gualaceo y Tomebamba.

Los valores de desempeño de todas las métricas sugieren un resultado insatisfactorio que en pocos casos alcanza más de 0,50 (NSE, KGE, R^2). NSE y R^2 se mantienen entre 0 a 0,55, KGE obtiene valores negativos de -0,15 a 0,61 lo que indica que el valor promedio de la serie original es un mejor predictor que el modelo evaluado en el caso de los modelos con el análisis VIF con 41 y 31 índices. Por lo que se puede determinar que los índices no parecen ser suficientes como predictores de caudal.

El RSME de cada modelo se asoció más a la magnitud de los valores de entrada que a los modelos propiamente los valores se acercaron desde cero a 30. Entre todos los escenarios propuestos solo ACP con 41 índices y 20 componentes mostró una tendencia a disminuir el RSME en todas las estaciones. El p-value fue muy bajo en todos los modelos, es decir que son estadísticamente muy significativos.



Al analizar los periodos de calibración y validación se puede observar un gran decaimiento del desempeño para este último lo cual es usual en la modelación hidrológica. Este comportamiento indica poca robustez en el modelo.

También es importante mencionar que por la naturaleza de nuestro estudio, al utilizar MLR los resultados no fueron satisfactorios. En cuanto a los índices y la variabilidad que representan en los distintos modelos, se identificó que el índice con mayor número de apariciones es: NP (51), seguido de AMM (42), PWR (38) y WHWP (24). Para la modelación con ACP, PWR y AMM coinciden en esta frecuencia. Los índices que no representan ninguna influencia importante en la zona son: AO, BEST, ENSO, GIAM, MEI V2, NAO, Niño 3, Niño 3.4, ONI, SOI, TPI (IPO).

Finalmente es importante destacar que siendo una primera aproximación al uso de todos los índices de teleconexión para medir su capacidad predictiva en la cuenca del río Paute, los resultados son muy interpretables. Es fundamental mencionar que los índices de teleconexión (señales y modos de variabilidad) además de aportar información global, mantienen una circulación atmosférica y oceánica bien estudiada. Esto permite no solo observar la variabilidad climática a gran escala espacial y temporal, sino también analizar variables dependientes de estos fenómenos, sin tener que recurrir a su fraccionamiento para un solo estudio.

4.2 Recomendaciones

Se recomienda mejorar la calidad de la información de entrada, tanto en rango de tiempo y disposición espacial, así como en la cantidad de datos que llega al registro de cada estación, para poder utilizar más estaciones sin incertidumbre o cambios en la serie por la metodología para el relleno de datos.

Es recomendable variar los periodos de validación para obtener modelos que reflejen la variabilidad del desempeño que se está aplicando a la calibración y conocer si el modelo tiene robustez.

En cuanto a la metodología se recomienda mantener la misma cantidad de índices, ya que los diversos escenarios han probado que la correlación existente entre los caudales y los índices no depende de un vínculo espacial. O una relación al fenómeno ENOS (modos de variabilidad) que



predomina el comportamiento de Ecuador. El análisis ACP puede mejorarse si se realiza con cada caudal para captar mejor la variabilidad de los índices asociada a cada estación. El análisis de multicolinealidad puede mejorarse al agrupar los índices de acuerdo a su ubicación.

Los modelos MLR dan resultados regulares (no predice eventos extremos) por lo que se sugiere optar por metodologías con algoritmos de inteligencia artificial que puedan integrar los procesos del contexto ambiental en los modelos de caudal, basándose en la zona de estudio. Si el estudio es estocástico es importante incluir más variables ambientales que estén directamente relacionadas con el caudal como la precipitación, para que distribuyan mejor el peso en los ajustes.

De ser posible se recomienda utilizar estas técnicas en combinación con otro tipo de datos, como los son los datos de las variables meteorológicas in situ. El uso de estas variables en conjunto con los índices climáticos podría elevar el grado de predictibilidad de los diferentes modelos.





REFERENCIAS BIBLIOGRÁFICAS

- Abdolrahimi, M. (2016). *The Effect of El Niño Southern Oscillation (ENSO) on World Cereal Production*. May, 133.
- Aguirre, A. D., Almanza, S., & Ivanova, Y. (2018). Influencia del fenómeno de Oscilación Cuasi-Bienal en las afluencias del sistema sur de abastecimiento de Bogotá. *Ingeniería, Investigación y Desarrollo*, 18(1), 16–24.
- Alzate, D., Mosquera, E., & Jemay, R. (2015). Cambio Climático Y Variabilidad Climática Para El Periodo 1981-2010 En Las Cuencas De Los Ríos Zulia Y Pamplonita, Norte De Santander – Colombia. *Luna Azul*, 1(40), 127–153.
<https://doi.org/10.17151/luaz.2015.40.10>
- Ballari, D., Campozano, L., Samaniego, E., & Orellana, D. (2020). Spatial association to characterize the climate teleconnection patterns in Ecuador based on satellite precipitation estimates. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W12, 101–106. <https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-101-2020>
- Bazo, J., Lorenzo, M. D. L. N., & Porfirio Da Rocha, R. (2013). Relationship between monthly rainfall in NW peru and tropical sea surface temperature. *Advances in Meteorology*, 2013. <https://doi.org/10.1155/2013/152875>
- Beltrán, L., & Díaz, D. C. (2020). Influence of the macroclimatic oscillations in the gachaneca river basin weather patterns; Boyaca-Colombia. *Revista Brasileira de Meteorologia*, 35(2), 171–185. <https://doi.org/10.1590/0102-7786351012>
- Bhattacharya, T., & Chiang, J. C. H. (2014). Spatial variability and mechanisms underlying El Niño-induced droughts in Mexico. *Climate Dynamics*, 43(12), 3309–3326.
<https://doi.org/10.1007/s00382-014-2106-8>
- Cabrera, J. (2012). Modelos Hidrológicos. In *Universidad Nacional de Ingeniería* (p. 8). Universidad Nacional de Ingeniería.
http://www.imefen.uni.edu.pe/Temas_interes/modhidro_1.pdf



- Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T., Jo, H.-S., Marengo, J. A., Alves, L. M., Osman, M., Li, S., ... Vera, C. (2020). Climate impacts of the El Niño–Southern Oscillation on South America. *Nature Reviews Earth & Environment*, 1(4), 215–231. <https://doi.org/10.1038/s43017-020-0040-3>
- Camargo, S. J., Sobel, A. H., Barnston, A. G., & Klotzbach, P. J. (2010). The Influence of Natural Climate Variability on Tropical Cyclones, and Seasonal Forecasts of Tropical Cyclone Activity. In *Global Perspectives on Tropical Cyclones, from Science to Mitigation* (Vol. 56, pp. 325–360). https://doi.org/10.1142/9789814293488_0011
- Campozano, L., Célleri, R., Trachte, K., Bendix, J., & Samaniego, E. (2016). Rainfall and Cloud Dynamics in the Andes: A Southern Ecuador Case Study. *Advances in Meteorology*, 2016(1), 16. <https://doi.org/10.1155/2016/3192765>
- Capurro, L. (2002). Huracanes, tifones, baguíos, willy-willies y ciclones. *Revista Del Centro de Investigación*, 19(5), 13–22. <https://www.redalyc.org/articulo.oa?id=34251902>
- Caron, L.-P., Boudreault, M., & Camargo, S. J. (2015). On the Variability and Predictability of Eastern Pacific Tropical Cyclone Activity*. *Journal of Climate*, 28(24), 9678–9696. <https://doi.org/10.1175/JCLI-D-15-0377.1>
- Carreric, A. (2020). *Enso diversity and global warming* (p. 329). Universidad Paul Sabatier.
- Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jäger, J., & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, 100(14), 8086–8091. <https://doi.org/10.1073/pnas.1231332100>
- Castro, L., & Carvajal, Y. (2010). Análisis De Tendencia Y Homogeneidad De Series Climatológicas. *Ingeniería de Recursos Naturales y Del Ambiente*, 9, 15–25.
- Cavalcanti, I. F. A. (2012). Large scale and synoptic features associated with extreme precipitation over South America: A review and case studies for the first decade of the 21st century. *Atmospheric Research*, 118, 27–40. <https://doi.org/10.1016/j.atmosres.2012.06.012>



- CELEC EP. (2013). *Actualización Del Estudio De Impacto Ambiental Y Plan De Manejo Ex Post De La Central Paute Molino*.
- CELEC EP, & CFE-ACOTENIC. (2016). *Caracterización y diagnóstico del área referencial del PHS (línea base)* (p. 281). CELEC EP. <http://acotecnic.com/wp-content/uploads/2016/01/5.1-LINEA-BASE-AREA-3-V5.pdf>
- Chimborazo, O. (2018). *Projected changes in climate, elevation-dependent warming, and extreme events over continental Ecuador for the period 2041–2070*. Universidad de Albania.
- Chithra, S., Kumar, S. R. R. S., Chinnaraju, K., & Alfin Ashmita, F. (2016). A comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks. *Construction and Building Materials*, 114, 528–535. <https://doi.org/10.1016/j.conbuildmat.2016.03.214>
- Choubin, B., Khalighi-sigaroodi, S., Malekian, A., & Kişi, Ö. (2016). Multiple linear regression , multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, 61(6), 1001–1009. <https://doi.org/10.1080/02626667.2014.966721>
- Chu, P.-S., & Wang, J. (1997). Tropical Cyclone Occurrences in the Vicinity of Hawaii: Are the Differences between El Niño and Non–El Niño Years Significant?*. *Journal of Climate*, 10(10), 2683–2689. [https://doi.org/10.1175/1520-0442\(1997\)010<2683:TCOITV>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<2683:TCOITV>2.0.CO;2)
- Clarke, R. (2002). Stochastic Hydrology Revisited. *Revista Brasileira de Recursos Hídricos*, 7(4), 94–104. <https://doi.org/10.21168/rbrh.v7n4.p94-104>
- Córdoba Machado, S., Palomino Lemus, R., Gámiz Fortis, S. R., Castro Díez, Y., & EstebanParra, M. J. (2015). Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia. *Global and Planetary Change*, 124, 41–61. <https://doi.org/10.1016/j.gloplacha.2014.11.003>
- Corporación Andina de Fomento. (1998). El Fenomeno El Niño 1997-98. *Bulletin de l'Institut*



Français d'études Andines, 27(3), 304.

- Crétat, J., Richard, Y., Pohl, B., Rouault, M., Reason, C., & Fauchereau, N. (2012). Recurrent daily rainfall patterns over South Africa and associated dynamics during the core of the austral summer. *International Journal of Climatology*, 32(2), 261–273.
<https://doi.org/10.1002/joc.2266>
- Cruz, F. T., Narisma, G. T., Villafuerte, M. Q., Cheng Chua, K. U., & Olaguera, L. M. (2013). A climatological analysis of the southwest monsoon rainfall in the Philippines. *Atmospheric Research*, 122, 609–616. <https://doi.org/10.1016/j.atmosres.2012.06.010>
- Cuevas, A., Azuz, I., & González, C. (2019). Predicting the Temporal Structure of the Atlantic Multidecadal Oscillation (AMO) for Agriculture Management in Mexico's Coastal Zone. *Coastal Education and Research Foundation*, 35(1), 210–226.
<https://doi.org/10.2112/JCOASTRES-D-18-00030.1>
- Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K. R., Zeckoski, R. W., Jeong, J., Parajuli, P. B., Saraswat, D., & Youssef, M. A. (2015). A recommended calibration and validation strategy for hydrologic and water quality models. *Transactions of the ASABE*, 58(6), 1705–1719. <https://doi.org/10.13031/trans.58.10712>
- Dalpiaz, D. (2020). *Chapter 21 The caret Package | R for Statistical Learning*.
<https://davidalpiaz.github.io/r4sl/the-caret-package.html>
- Dehghan, M., Saghafian, B., Rivaz, F., & Khodadadi, A. (2014). Uncertainty Analysis of Monthly Streamflow Forecasting. *Current World Environment*, 9(3), 894–902.
<https://doi.org/10.12944/CWE.9.3.40>
- Dima, M., & Lohmann, G. (2004). Fundamental and derived modes of climate variability: concept and application to interannual time-scales. *Tellus A: Dynamic Meteorology and Oceanography*, 56(3), 229–249. <https://doi.org/10.3402/tellusa.v56i3.14415>
- Dowdy, A. J., Mills, G. A., Timbal, B., & Wang, Y. (2013). Changes in the Risk of Extratropical Cyclones in Eastern Australia. *Journal of Climate*, 26(4), 1403–1417.
<https://doi.org/10.1175/JCLI-D-12-00192.1>



- Eckhardt, S., Stohl, A., Beirle, S., Spichtinger, N., James, P., Forster, C., Junker, C., Wagner, T., Platt, U., & Jennings, S. G. (2003). The North Atlantic Oscillation controls air pollution transport to the Arctic. *Atmospheric Chemistry and Physics*, 3(5), 1769–1778.
<https://doi.org/10.5194/acp-3-1769-2003>
- Esha, R. I., Imteaz, M. A., & Nazari, A. (2019). Assessing Gene Expression Programming as a technique for seasonal streamflow prediction: A case study of NSW. *IOP Conference Series: Earth and Environmental Science*, 351(1). <https://doi.org/10.1088/1755-1315/351/1/012004>
- Espino Sánchez, M. A. (2014). *Patrones de variabilidad ambiental y las pesquerías en el Pacífico Sud Este*. Universidad Nacional Mayor de San Marcos.
- Fernandez, C., & Buss, S. (2016). Ocurrencia y gestión de inundaciones en América Latina y el Caribe: Factores claves y experiencia adquirida | Publications. *Bid*, 1–27.
<https://publications.iadb.org/publications/spanish/document/Ocurrencia-y-gestión-de-inundaciones-en-América-Latina-y-el-Caribe-Factores-claves-y-experiencia-adquirida.pdf>
- Ferrero Martín, C. (2019). *Variabilidad multianual de los ciclones tropicales en el Atlántico Norte* [Universidad de Cádiz]. <http://hdl.handle.net/10498/21602>
- Gastineau, G., Li, L., & Le Treut, H. (2009). The Hadley and Walker Circulation Changes in Global Warming Conditions Described by Idealized Atmospheric Simulations. *Journal of Climate*, 22(14), 3993–4013. <https://doi.org/10.1175/2009JCLI2794.1>
- Giddings, L., & Soto, M. (2006). 2. Metodología 2.1. Fuentes de datos. *Revista de Climatología*, 6(1), 13–20. <http://www.cdc.noaa.gov/ClimateIndices/List/>
- Gillett, N. P., Kell, T. D., & Jones, P. D. (2006). Regional climate impacts of the Southern Annular Mode. *Geophysical Research Letters*, 33(23), L23704.
<https://doi.org/10.1029/2006GL027721>
- Gómez, G., Cadier, E., & Rossel, F. (1996). Las inundaciones en la zona costera ecuatoriana: causas, obras de protección existentes y previstas. *Bulletin de l'Institut Francais d'études Andins*, 25(3), 399–420.



- Gómez, R. (2014). Clasificación de tormentas tropicales según lluvias asociadas: (1) estado del arte. *Ingeniería Hidráulica y Ambiental*, 35(2), 35–51.
- Gudmundsson, L., X. H., Leonard, M., Westra, S., & Sonia, S. (2017). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality Control, Time-series Indices and Homogeneity Assessment. In *Earth System Science Data* (Issue September).
<https://doi.org/10.5194/essd-2017-104>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91.
<https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gutiérrez, P. M. (2012). ¿Criterio mágico para resolver cualquier problema o leyenda urbana? *Universitas Scientiarum*, 17(2), 203–215.
- Guttman, N. (1998). Homogeneity, data adjustments and climatic normals. In *National Climatic Data Center. Asheville, United ...* (Issue March).
<http://www.stat.washington.edu/peter/7IMSC/Normals.pdf>
- Hatzaki, M., Flocas, H. A., Asimakopoulos, D. N., & Maheras, P. (2007). The eastern Mediterranean teleconnection pattern: identification and definition. *International Journal of Climatology*, 27(6), 727–737. <https://doi.org/10.1002/joc.1429>
- Herrera, V. (2018). Revista Internacional de Investigación e Innovación Tecnológica. *Revista Internacional de Investigación e Innovación Tecnológica*, 6(32), 1–20.
- Hoell, A., Funk, C., Zinke, J., & Harrison, L. (2017). Modulation of the Southern Africa precipitation response to the El Niño Southern Oscillation by the subtropical Indian Ocean Dipole. *Climate Dynamics*, 48(7–8), 2529–2540. <https://doi.org/10.1007/s00382-016-3220-6>
- IDEAM - UNAL. (2018). Variabilidad Climática y el cambio climático en Colombia. In *Bogota, D.C.* <http://documentacion.ideam.gov.co/openbiblio/bvirtual/023778/variabilidad.pdf>
- Instituto de Estudios de Régimen, Al, S. del E.-A., Investigaciones, V. de, & Azuay, de la U.



- del. (2017). CONFibSIG - XVI Conferencia Iberoamericana de Sistemas de Información Geográfica. *XVI Conferencia Iberoamericana de Sistemas de Información Geográfica*, 16, 475.
- Jones, P. D., Jonsson, T., & Wheeler, D. (1997). Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *International Journal of Climatology*, 17(13), 1433–1450.
[https://doi.org/10.1002/\(SICI\)1097-0088\(19971115\)17:13<1433::AID-JOC203>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P)
- Knight, J. R., Folland, C. K., & Scaife, A. A. (2006). Climate impacts of the Atlantic multidecadal oscillation. *Geophysical Research Letters*, 33(17), 2–5.
<https://doi.org/10.1029/2006GL026242>
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Knudsen, M. F., Seidenkrantz, M. S., Jacobsen, B. H., & Kuijpers, A. (2011). Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature Communications*, 2(1). <https://doi.org/10.1038/ncomms1186>
- Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97.
<https://doi.org/10.5194/adgeo-5-89-2005>
- Kumar, V., Jain, S. K., & Singh, Y. (2010). Analyse des tendances pluviométriques de long terme en Inde. *Hydrological Sciences Journal*, 55(4), 484–496.
<https://doi.org/10.1080/02626667.2010.481373>
- Kundzewicz, Z. W., Szwed, M., & Pińskwar, I. (2019). Climate variability and floods-A global review. *Water (Switzerland)*, 11(7). <https://doi.org/10.3390/w11071399>
- Lavers, D. A., Villarini, G., Allan, R. P., Wood, E. F., & Wade, A. J. (2012). The detection of atmospheric rivers in atmospheric reanalyses and their links to British winter floods and the large-scale climatic circulation. *Journal of Geophysical Research: Atmospheres*, 117(D20).



<https://doi.org/10.1029/2012JD018027>

Le, C., & Eberly, L. (2003). *Introductory Biostatistics* (J. W. & S. Inc. (ed.); Second edi).

Linkin, M. E., & Nigam, S. (2008). The North Pacific Oscillation–West Pacific Teleconnection Pattern: Mature-Phase Structure and Winter Impacts. *Journal of Climate*, *21*(9), 1979–1997. <https://doi.org/10.1175/2007JCLI2048.1>

Loua, R. T., Bencherif, H., Mbatha, N., Bègue, N., Hauchecorne, A., Bamba, Z., & Sivakumar, V. (2019). Study on temporal variations of surface temperature and rainfall at Conakry airport, Guinea: 1960–2016. *Climate*, *7*(7), 1–26. <https://doi.org/10.3390/CLI7070093>

Lujano, E. (2019). *Análisis regional de frecuencia para la estimación de caudales máximos en la cuenca amazónica del Perú*. Universidad Nacional del Altiplano.

Madsen, T., Franz, K., & Hogue, T. (2020). Evaluation of a distributed streamflow forecast model at multiple watershed scales. *Water (Switzerland)*, *12*(5), 1–18. <https://doi.org/10.3390/W12051279>

Mallakpour, I., & Villarini, G. (2016). Investigating the relationship between the frequency of flooding over the central United States and large-scale climate. *Advances in Water Resources*, *92*, 159–171. <https://doi.org/10.1016/j.advwatres.2016.04.008>

Mariani, M., & Fletcher, M. (2016). The Southern Annular Mode determines interannual and centennial-scale fire activity in temperate southwest Tasmania, Australia. *Geophysical Research Letters*, *43*(4), 1702–1709. <https://doi.org/10.1002/2016GL068082>

Massey, A., & Miller, S. J. (2016). Tests of Hypotheses Using Statistics. In *Mathematics Department, Brown University, Providence, RI 2912* (pp. 1–32). Universidad de Brown.

Mauricio Zambrano Bigiarini. (2020). Package ‘hydroGOF.’ In *Hydrological Processes* (Vol. 1, Issue 4, pp. 2723–2725).

McNeish, D. (2017). Missing data methods for arbitrary missingness with small samples. *Journal of Applied Statistics*, *44*(1), 24–39. <https://doi.org/10.1080/02664763.2016.1158246>

Merz, R., Parajka, J., & Blo, G. (2009). Scale effects in conceptual hydrological modeling. *Water*



Resources Research, 45(1), 1–15. <https://doi.org/10.1029/2009WR007872>

- Meyer, T. (2010). Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. *Surveying and Land Information Science*, 72(5), 672.
- Mezzina, B., García-Serrano, J., Bladé, I., & Kucharski, F. (2020). Dynamics of the ENSO teleconnection and NAO variability in the North Atlantic-European late winter. *Journal of Climate*, 33(3), 907–923. <https://doi.org/10.1175/JCLI-D-19-0192.1>
- Mo, K., & Higgins, W. (1998). Tropical influences on California precipitation. *Journal of Climate*, 11(3), 412–430. [https://doi.org/10.1175/1520-0442\(1998\)011<0412:TIOCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<0412:TIOCP>2.0.CO;2)
- Montealegre Bocanegra, J. E. (2009). Estudio de la variabilidad climática de la precipitación en Colombia asociada a procesos oceánicos y atmosféricos de meso y gran escala. In *IDEAM*.
- Montoro, D. (2015). Capítulo 9. Regresión lineal simple. In *Métodos Estadísticos en La Ingeniería* (1st ed., p. 11). Universidad de Jaén.
<http://www4.ujaen.es/~dmontoro/Metodos/Tema 9.pdf>
- Morán, E., Bazo, J., López, J., Aguilar, E., Azorín, C., Sanchez, A., Martínez, R., Nieto, J., Mejía, R., Martín, N., & Vicente, S. (2016). Climate trends and variability in Ecuador (1966–2011). *International Journal of Climatology*, 36(11), 3839–3855.
<https://doi.org/10.1002/joc.4597>
- Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Muhammad, W., Mani, N., Andiego, G., & Muhammad, U. (2017). A review of criteria of fit for hydrological models. *International Research Journal of Engineering and Technology*, 4(11), 11.
- Mukaka, M. (2012). Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient. *Malawi Medical Journal*, 24(3), 69–71. <https://pubmed.ncbi.nlm.nih.gov/23638278/>
- Nakano, S., Ito, K., Suzuki, K., & Ueno, G. (2016). Decadal-scale meridional shift of the



- typhoon recurvature latitude over five decades. *International Journal of Climatology*, 36(11), 3819–3827. <https://doi.org/10.1002/joc.4595>
- Niu, W., Feng, Z., Feng, B., Min, Y., & Cheng, C. (2019). *Neural Network , Extreme Learning Machine , and Support Vector Machine in Deriving Operation Rule of Hydropower Reservoir*. <https://doi.org/10.3390/w11010088>
- Novales, A. (2010). *Análisis de Regresión*. Universidad Complutense.
- Nugroho, A. R., Tamagawa, I., & Harada, M. (2022). Spatiotemporal Analysis on the Teleconnection of ENSO and IOD to the Stream Flow Regimes in Java, Indonesia. *Water (Switzerland)*, 14(2). <https://doi.org/10.3390/w14020168>
- Orbes, J., & Peralta, T. (2017). *Estado del arte en Manejo de Sedimentos en cuencas Andinas en el Ecuador, caso de estudio: cuenca del Río Paute*. Universidad de Cuenca.
- Pabón, J., Zea, J., León, G., Hurtado, G., Gonzáles, O. C., & Montealegre, J. (2004). La atmósfera , el tiempo y el clima. In *El Medio Ambiente en Colombia* (pp. 92–113). <http://documentacion.ideam.gov.co/openbiblio/bvirtual/000001/cap3-ii.pdf>
- Pastor, J. A. G. (1999). Teleconexiones climáticas y precipitaciones en la España mediterránea. In C. M. en I. Balears (Ed.), *La climatología española en los albores del siglo XXI* (pp. 243–252). Congreso de la Asociación de Climatología.
- Patel, S., Hardaha, M. K., Seetpal, M. K., & Madankar, K. K. (2016). Multiple Linear Regression Model for Stream Flow Estimation of Wainganga River. *American Journal of Water Science and Engineering*, 2(1), 1–5. <https://doi.org/10.11648/j.ajwse.20160201.11>
- Patricola, C. M., Saravanan, R., & Chang, P. (2014). The Impact of the El Niño–Southern Oscillation and Atlantic Meridional Mode on Seasonal Atlantic Tropical Cyclone Activity. *Journal of Climate*, 27(14), 5311–5328. <https://doi.org/10.1175/JCLI-D-13-00687.1>
- Peña, D. (2002). Componentes principales. In Interamericana de España (Ed.), *Análisis de datos multivariantes* (Mcgraw-Hil).
- Pérez, J., Senent, J., Segura, F., Pulido, D., & Srinivasan, R. (2019). Evaluating hydrological models for deriving water resources in peninsular Spain. *Sustainability (Switzerland)*,



11(10), 1–36. <https://doi.org/10.3390/su11102872>

- Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., & Ulbrich, U. (2009). Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate Dynamics*, 32(5), 711–737. <https://doi.org/10.1007/s00382-008-0396-4>
- Pirnia, A., Golshan, M., Darabi, H., Adamowski, J., & Rozbeh, S. (2019). Using the Mann–Kendall test and double mass curve method to explore stream flow changes in response to climate and human activities. *Journal of Water and Climate Change*, 10(4), 725–742. <https://doi.org/10.2166/wcc.2018.162>
- PNUMA, FLACSO, & MAE. (2008). Geo Ecuador 2008. In *Geo Ecuador 2008* (Vol. 1). <http://www.flacsoandes.edu.ec/libros/digital/41444.pdf>
- Portilla, F. (2018). *Agroclimatología del Ecuador* (p. 647). Editorial Universitaria Abya-Yala.
- Prabhu, A., Oh, J., Kim, I., Kripalani, R. H., Mitra, A. K., & Pandithurai, G. (2017). Summer monsoon rainfall variability over North East regions of India and its association with Eurasian snow, Atlantic Sea Surface temperature and Arctic Oscillation. *Climate Dynamics*, 49(7–8), 2545–2556. <https://doi.org/10.1007/s00382-016-3445-4>
- Qiu, Y., Cai, W., Guo, X., & Ng, B. (2015). The asymmetric influence of the positive and negative IOD events on China’s rainfall. *Scientific Reports*, 4(1), 4943. <https://doi.org/10.1038/srep04943>
- Qu, J., Gong, D., Mao, R., Yang, J., & Li, S. (2017). Possible influence of Arctic oscillation on precipitation along the East Asian rain belt during boreal spring. *Theoretical and Applied Climatology*, 130(1–2), 487–495. <https://doi.org/10.1007/s00704-016-1900-0>
- Quiring, S. M., & Blair, D. (2000). The utility of global teleconnection indices for long-range crop forecasting on the Canadian prairies. In J. I. Romanowski (Ed.), *Prairie Perspectives: Geographical Essays* (Vol. 3, pp. 23–61). University of Manitoba.
- Quishpe, C., García, M., Gámiz, S. R., Castro, Y., & Esteban, M. J. (2016). Variabilidad estacional de los caudales en Ecuador en el periodo 1964-2014. In *Clima, sociedad, riesgos*



- y ordenación del territorio (pp. 97–107). Servicio de Publicaciones de la UA.
<https://doi.org/10.14198/XCongresoAECALicante2016-09>
- Quispe, S., Tatiana, S., & Mar, M. (2021). Capítulo 7 Metodología para el tratamiento de datos meteorológicos e hidrométricos a escala de cuenca. In ECORFAN (Ed.), *CIERMMI Women in Science TXVI Engineering and Technology* (pp. 1–39). Handbooks.
<https://doi.org/10.35429/H.2021.16.107.145>
- Ramamasy, S., & Baas, S. (2007). Climate variability and change: adaptation to drought in Bangladesh. In *Case Study - Institutions for Rural Development, FAO* (Issue No.9).
cabi:20083071262
- Ranjit, P. (2006). Multicollinearity: causes, effects and remedies. *IASRI, 1*(1), 58–65.
- Räsänen, T. A., Lindgren, V., Guillaume, J. H. A., Buckley, B. M., & Kumm, M. (2016). On the spatial and temporal variability of ENSO precipitation and drought teleconnection in mainland Southeast Asia. *Climate of the Past, 12*(9), 1889–1905.
<https://doi.org/10.5194/cp-12-1889-2016>
- Rea, A., & Rea, W. (2016). *How Many Components should be Retained from a Multivariate Time Series PCA ?*
- Restrepo, L., & González, J. (2007). De Pearson a Spearman. *Rev Col Cienc Pec, 20*(2), 183–192.
- Robinson, D. (2022). Package “broom.” In *CRAN*.
- Rodrigo, J. A. (2017). *Análisis de Componentes Principales (Principal Component Analysis , PCA) y t-SNE. Detección de Anomalías: Autoencoders y PCA.*
- Rogers, C. D. W., & Beringer, J. (2017). Describing rainfall in northern Australia using multiple climate indices. *Biogeosciences, 14*(3), 597–615. <https://doi.org/10.5194/bg-14-597-2017>
- Saji, N., & Yamagata, T. (2003). Possible impacts of Indian Ocean Dipole mode events on global climate. *Climate Research, 25*, 151–169. <https://doi.org/10.3354/cr025151>
- Satyamurthy, P., da Costa, C. P. W., Manzi, A. O., & Candido, L. A. (2013). A quick look at the



- 2012 record flood in the Amazon Basin. *Geophysical Research Letters*, 40(7), 1396–1401. <https://doi.org/10.1002/grl.50245>
- Schneider, T., Hampel, H., Mosquera, P. V., Tylmann, W., & Grosjean, M. (2018). Paleo-ENSO revisited: Ecuadorian Lake Pallcacocha does not reveal a conclusive El Niño signal. *Global and Planetary Change*, 168, 54–66. <https://doi.org/10.1016/j.gloplacha.2018.06.004>
- Schöngart, J., Junk, W. J., Piedade, M. T. F., Ayres, J. M., Hüttermann, A., & Worbes, M. (2004). Teleconnection between tree growth in the Amazonian floodplains and the El Niño-Southern Oscillation effect. *Global Change Biology*, 10(5), 683–692. <https://doi.org/10.1111/j.1529-8817.2003.00754.x>
- Secretaría de Medio Ambiente y Ordenamiento Territorial. (2021). *Actualización del Diagnóstico Climatológico del Estado de Guanajuato*. <http://cuentame.inegi.org.mx/monografias/informacion/gto/poblacion/>
- Sedano, R. (2017). *Influencia de la variabilidad climática y factores antrópicos en los extremos hidrológicos en el Valle Alto del río Cauca, Colombia*. Universitat Politècnica de Valencia.
- Shabri, A., & Shuhaida, I. (2014). Stream Flow Forecasting using Principal Component Analysis and Least Square Support Vector Machine. *Journal of Applied Science and Agriculture*, 9, 170–180.
- Shi, X., Wood, A. W., & Lettenmaier, D. P. (2008). How essential is hydrologic model calibration to seasonal stream flow forecasting? *Journal of Hydrometeorology*, 9(6), 1350–1363. <https://doi.org/10.1175/2008JHM1001.1>
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Sobel, A. H., Camargo, S. J., Barnston, A. G., & Tippett, M. K. (2016). Northern hemisphere tropical cyclones during the quasi-El Niño of late 2014. *Natural Hazards*, 83(3), 1717–1729. <https://doi.org/10.1007/s11069-016-2389-7>
- Song, L., Chen, S., Chen, W., & Chen, X. (2017). Distinct impacts of two types of La Niña events on Australian summer rainfall. *International Journal of Climatology*, 37(5), 2532–



2544. <https://doi.org/10.1002/joc.4863>

- Steptoe, H., Jones, S. E. O., & Fox, H. (2018). Correlations Between Extreme Atmospheric Hazards and Global Teleconnections: Implications for Multihazard Resilience. *Reviews of Geophysics*, *56*(1), 50–78. <https://doi.org/10.1002/2017RG000567>
- Sun, C., Li, J., & Ding, R. (2016). Strengthening relationship between ENSO and western Russian summer surface temperature. *Geophysical Research Letters*, *43*(2), 843–851. <https://doi.org/10.1002/2015GL067503>
- Tanaka, S., Nishii, K., & Nakamura, H. (2016). Vertical Structure and Energetics of the Western Pacific Teleconnection Pattern. *Journal of Climate*, *29*(18), 6597–6616. <https://doi.org/10.1175/JCLI-D-15-0549.1>
- Tencaliec, P., Favre, A.-C., Prieur, C., & Mathevet, T. (2015). Reconstruction of missing daily streamflow data using dynamic regression models. *Water Resources Research*, *51*(12), 9447–9463. <https://doi.org/10.1002/2015WR017399>
- Thompson, D. W. J., & Wallace, J. M. (2000). Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability*. *Journal of Climate*, *13*(5), 1000–1016. [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2)
- Tsakiri, K., Marsellos, A., & Kapetanakis, S. (2018). Artificial Neural Network and Multiple Linear Regression for Flood Prediction in Mohawk River, New York. *Water*, *10*(9), 20. <https://doi.org/10.3390/w10091158>
- Uvo, C. B., Foster, K., & Olsson, J. (2021). The spatio-temporal influence of atmospheric teleconnection patterns on hydrology in Sweden. *Journal of Hydrology: Regional Studies*, *34*, 100782. <https://doi.org/10.1016/j.ejrh.2021.100782>
- Valencia Gasti, J. A., Baumgartner, T., & Durazo, R. (2015). Efectos del clima oceánico sobre el ciclo de vida y distribución de peces pelágicos menores en el Sistema de la Corriente de California, frente a Baja California. *Ciencias Marinas*, *41*(4), 315–348. <https://doi.org/10.7773/cm.v41i4.2571>
- Vargas, N. O., & Giraldo, M. V. (2018). Protocolo de Modelacion Hidrológica e Hidráulica. In



Minambiente Colombia.

http://documentacion.ideam.gov.co/openbiblio/bvirtual/023833/Protocolo_Modelacion_HH.pdf

Vega Vilca, J. C., & Guzman, J. (2011). Regresion PLS y PCA Como Solución al Problema de Multicolinealidad en Regresion Multiple. *Revista de Matemática: Teoría y Aplicaciones*, 18(1), 9. <https://doi.org/10.15517/rmta.v18i1.2111>

Vimont, D. J. (2012). Analysis of the Atlantic meridional mode using linear inverse modeling: Seasonality and regional influences. *Journal of Climate*, 25(4), 1194–1212. <https://doi.org/10.1175/JCLI-D-11-00012.1>

Wang, C., & Enfield, D. B. (2001). The tropical western hemisphere warm pool. *Geophysical Research Letters*, 28(8), 1635–1638. <https://doi.org/10.1029/2000GL011763>

White, C. J., Hudson, D., & Alves, O. (2014). ENSO, the IOD and the intraseasonal prediction of heat extremes across Australia using POAMA-2. *Climate Dynamics*, 43(7–8), 1791–1810. <https://doi.org/10.1007/s00382-013-2007-2>

Willems, P. (2013). Multidecadal oscillatory behaviour of rainfall extremes in Europe. *Climatic Change*, 120(4), 931–944. <https://doi.org/10.1007/s10584-013-0837-x>

Winter, J. C. F., & Dodou, D. (2016). Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations. *Communications in Statistics - Simulation and Computation*, 45(1), 299–321. <https://doi.org/10.1080/03610918.2013.862274>

Wu, Z., Wang, B., Li, J., & Jin, F.-F. (2009). An empirical seasonal prediction model of the east Asian summer monsoon using ENSO and NAO. *Journal of Geophysical Research*, 114(D18), D18120. <https://doi.org/10.1029/2009JD011733>

Xiao, M., Zhang, Q., & Singh, V. P. (2015). Influences of ENSO, NAO, IOD and PDO on seasonal precipitation regimes in the Yangtze River basin, China. *International Journal of Climatology*, 35(12), 3556–3567. <https://doi.org/10.1002/joc.4228>

Yustres Quintero, L. F., Riveros Zarate, Z. C., & Ivanova, J. (2020). Efecto del fenómeno de la



Oscilación Cuasi-Bienal en los afluentes del embalse de Betania. *Revista EIA*, 17(34), 1–9.
<https://doi.org/10.24050/reia.v17i34.1246>

Zebrowski, E., & Roura, X. (1997). *Inundaciones y sequias Cooperacion* (p. 220). INAMHI.
http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers12-12/010014173.pdf

Zevross. (2017). *Zevross Know your data*. Predictive Modeling and Machine Learning in R with the Caret Package. <http://zevross.com/blog/2017/09/19/predictive-modeling-and-machine-learning-in-r-with-the-caret-package/>

Zhiña, D. (2021). *Factors controlling the stable isotopic composition of precipitation in the tropical alpine highlands of south Ecuador* [Universidad de Cuenca].
<http://dspace.ucuenca.edu.ec/handle/123456789/36734>

Zubiate, L., McDermott, F., Sweeney, C., & O'Malley, M. (2017). Spatial variability in winter NAO-wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 552–562. <https://doi.org/10.1002/qj.2943>

Zuluaga, M., Posada, J., & Poveda, G. (2000). *Relación entre las fases de la Oscilación Cuasi-Bienal (QBO), la precipitación mensual y la radiación de onda larga emergente en Colombia*. Universidad Nacional de Colombia.

CAPITULO 5: ANEXOS

5.1 Anexo 1: Resultados de series de tiempo en la etapa de calibración para los escenarios ACP y usando retrasos mensuales.

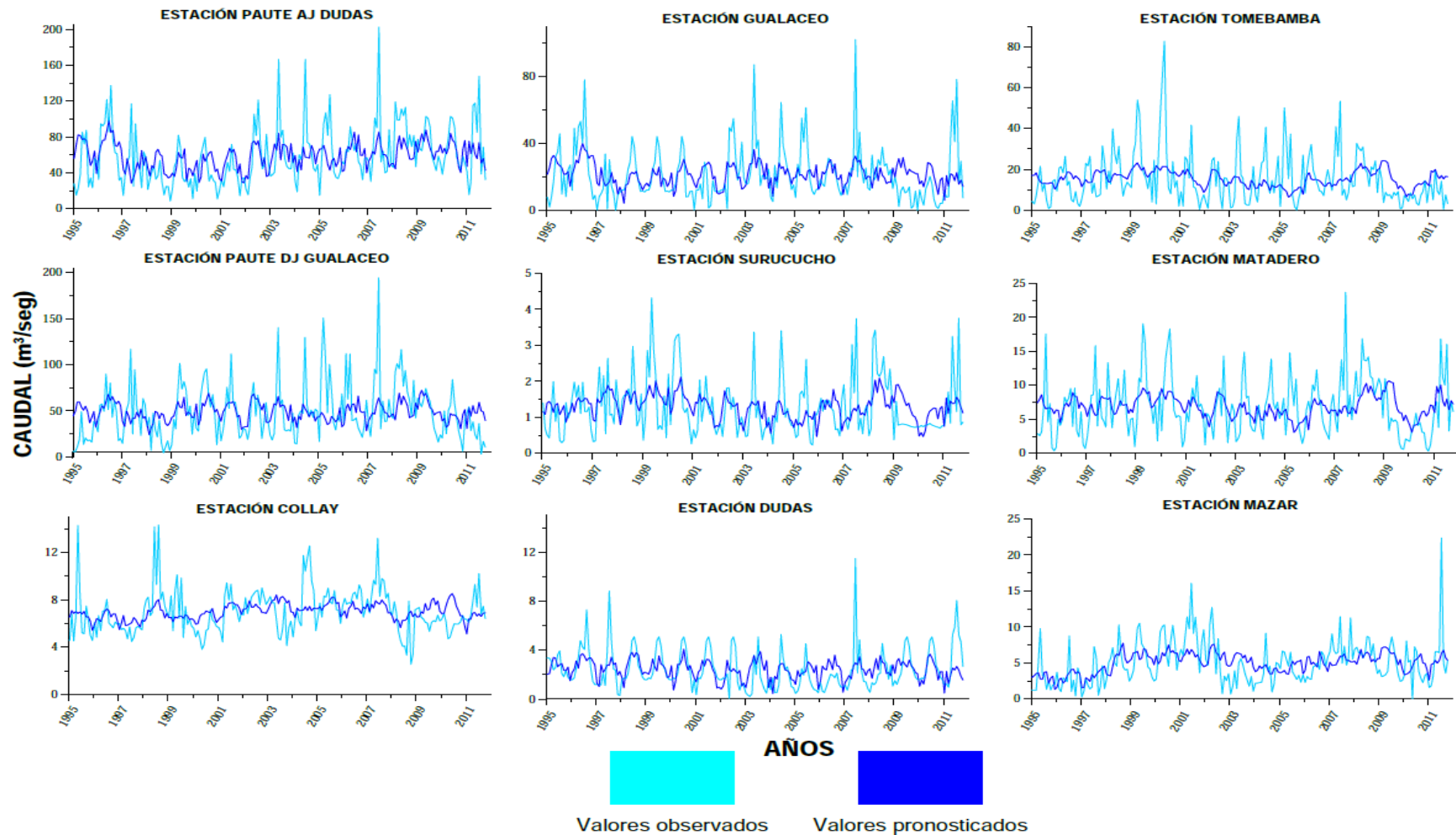


Figura 32. Resultados del escenario con 31 índices y 70% de varianza en el análisis ACP.

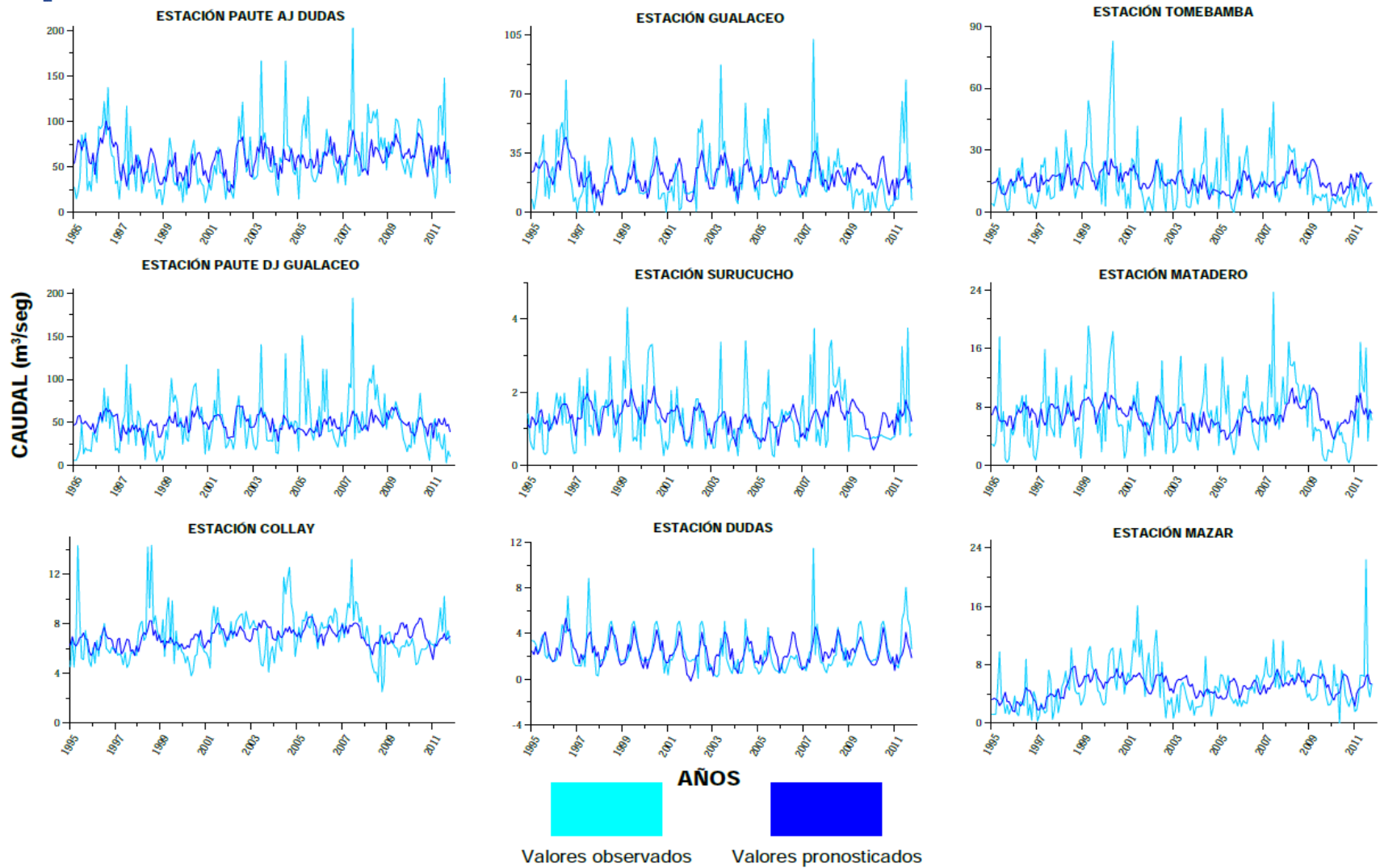


Figura 33. Resultados del escenario con 41 índices y 70% de varianza en el análisis ACP.

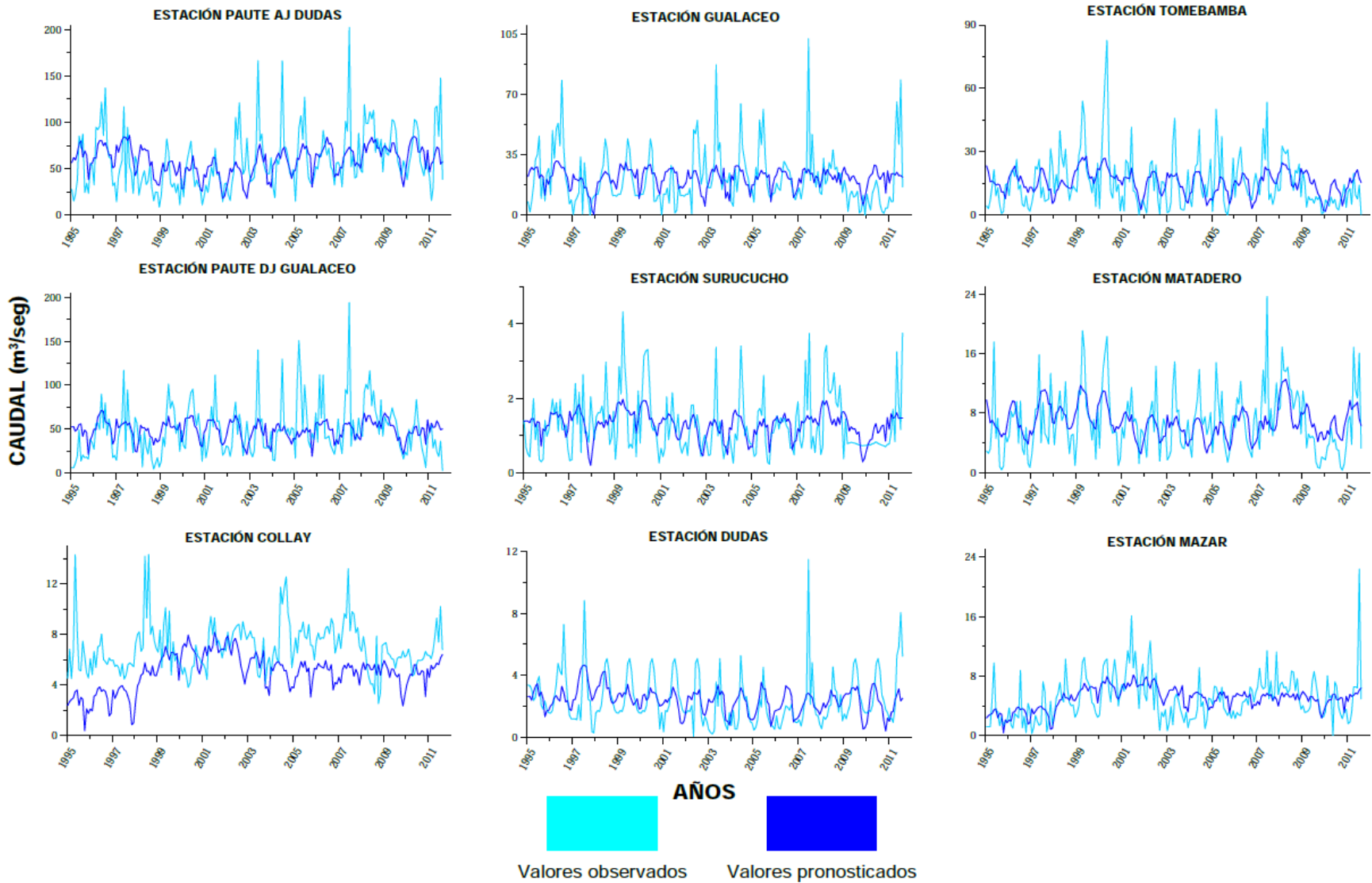


Figura 34. Resultados del escenario con 31 índices y dos meses de retraso.

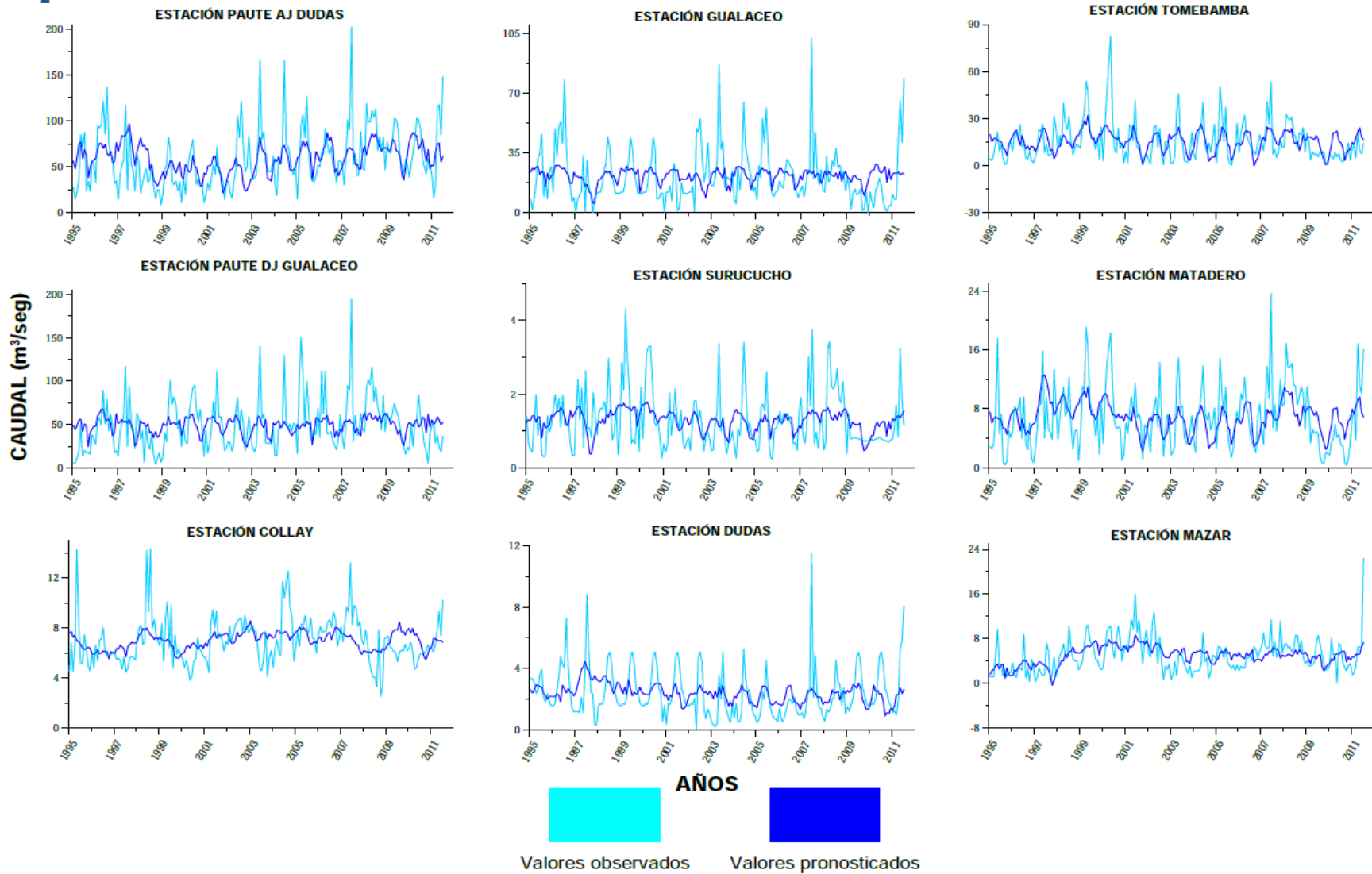


Figura 35. Resultados del escenario con 31 índices y tres meses de retraso.

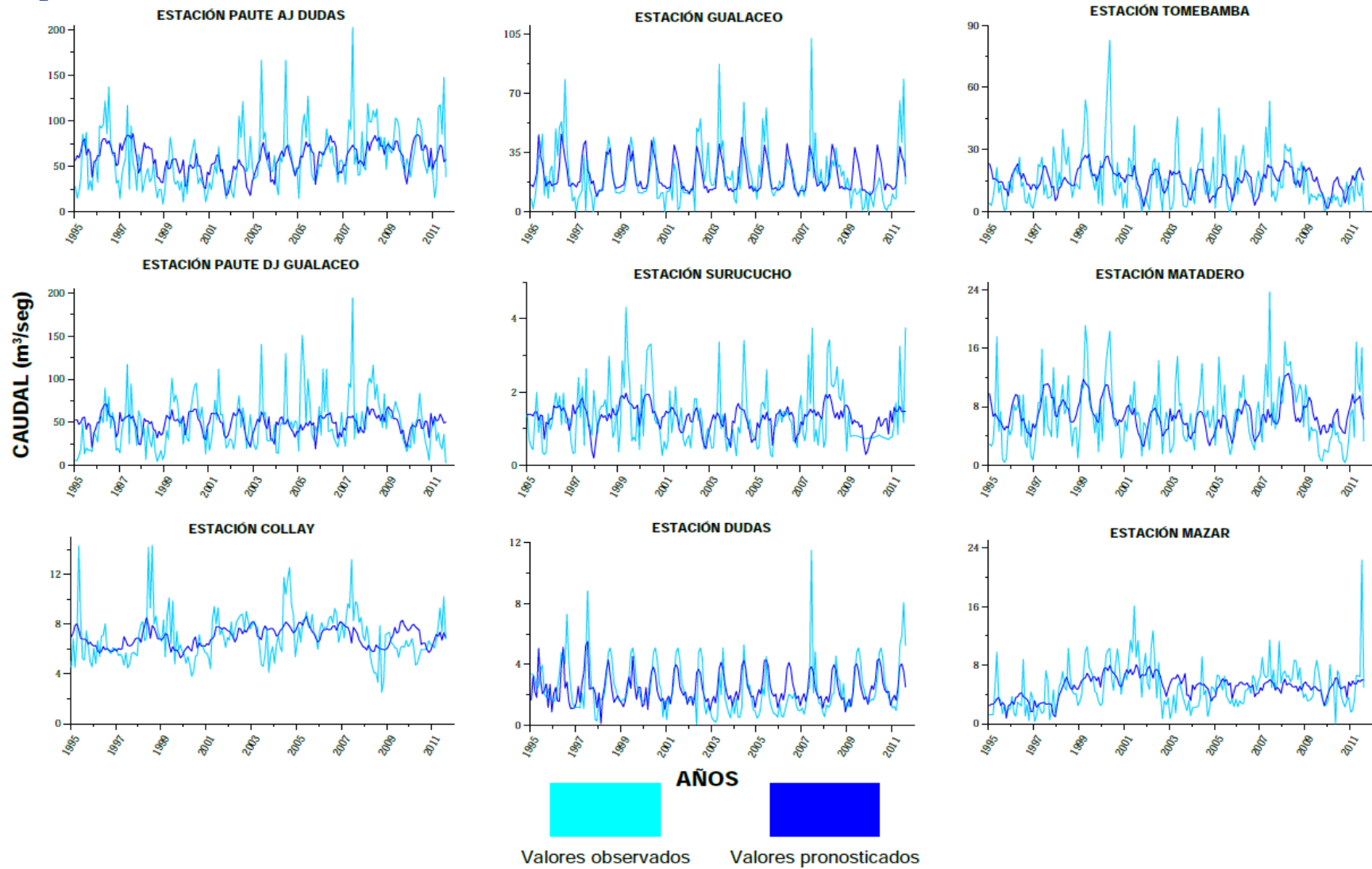


Figura 36. Resultados del escenario con 41 índices y dos meses de retraso.

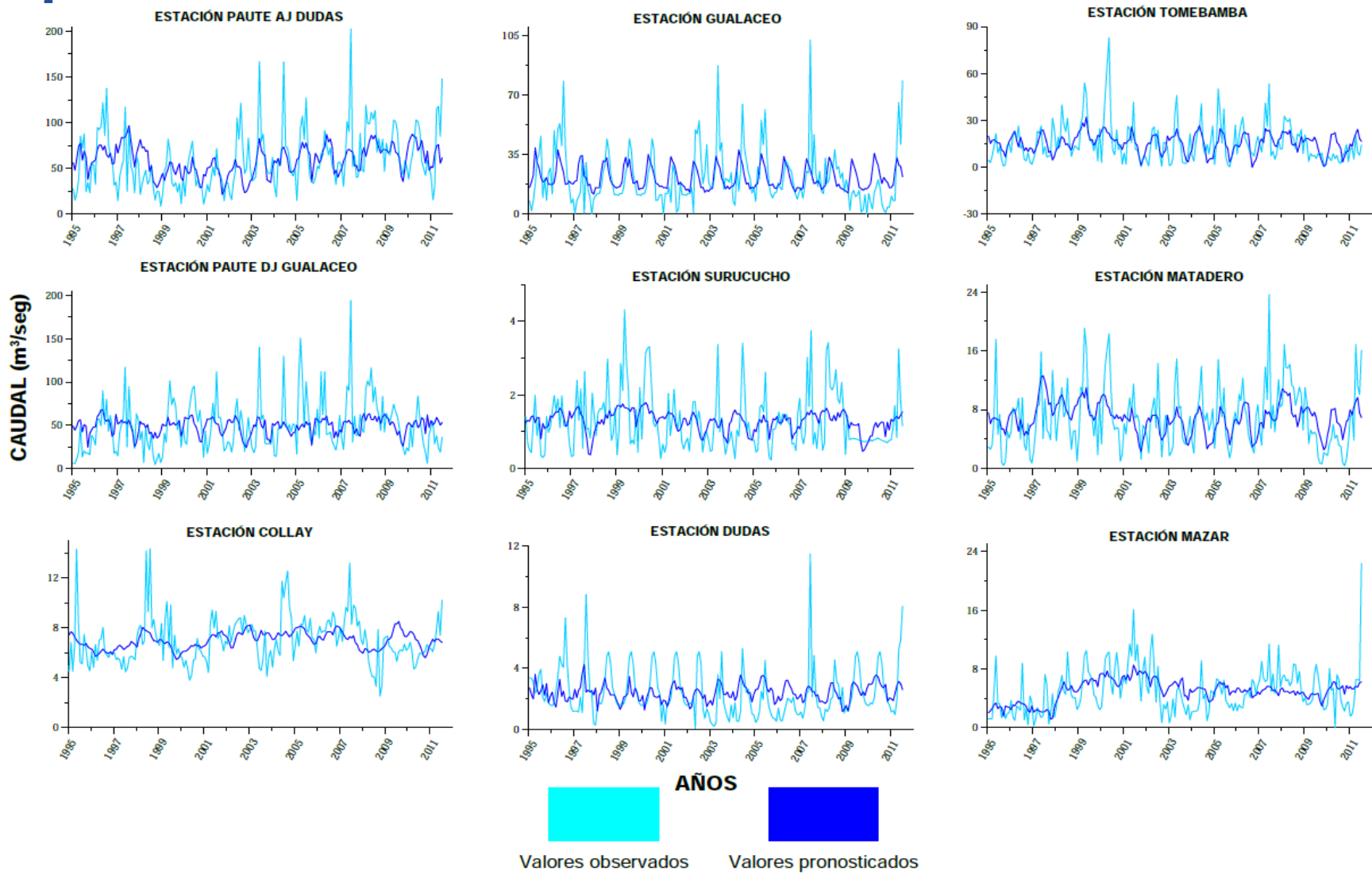


Figura 37. Resultados del escenario con 41 índices y tres meses de retraso.

5.2 Anexo 2: Resultados de diagramas de dispersión en la etapa de calibración para los escenarios ACP y usando retrasos mensuales.

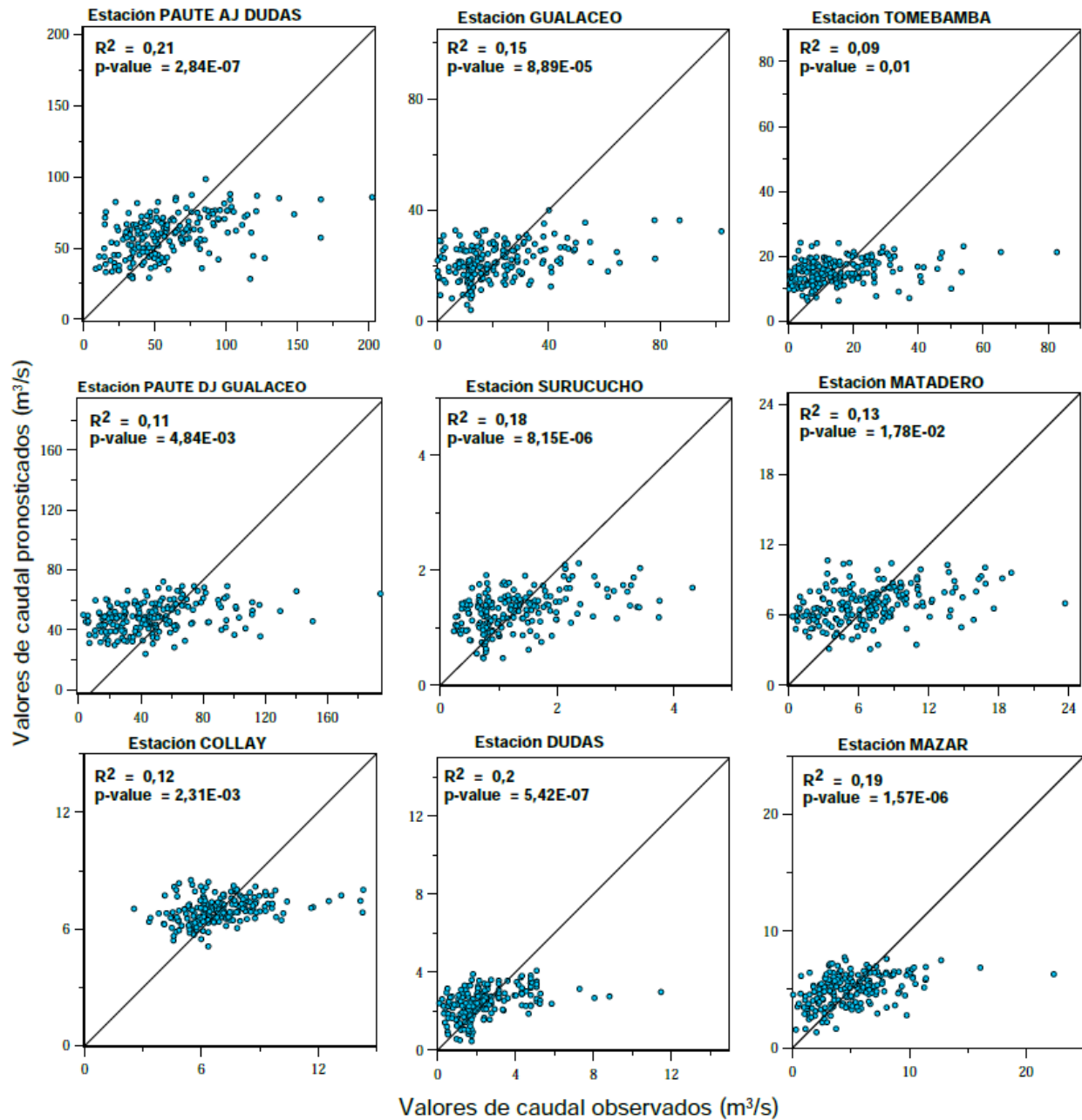


Figura 38. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 70% de varianza.

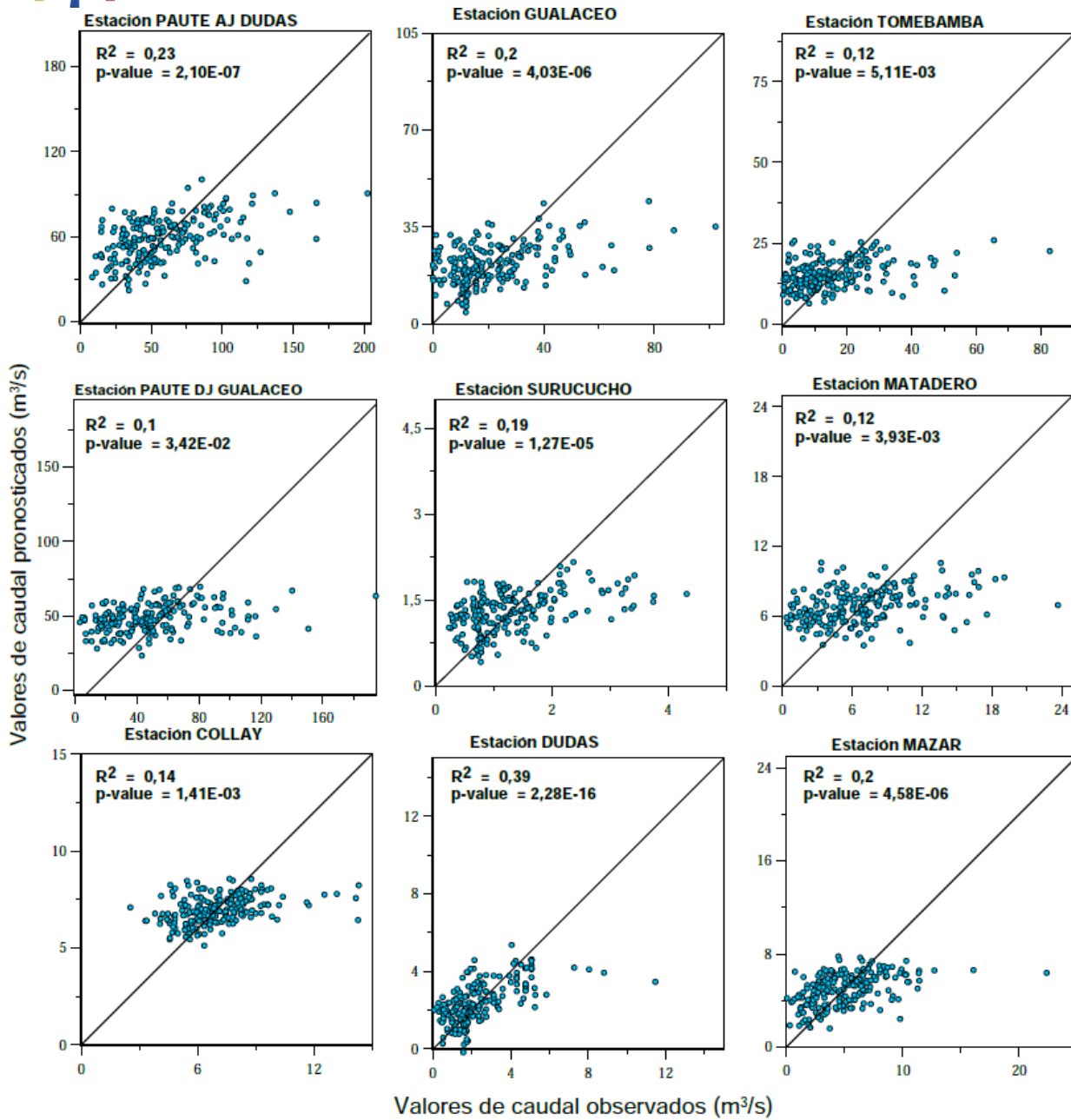


Figura 39. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 70% de varianza.

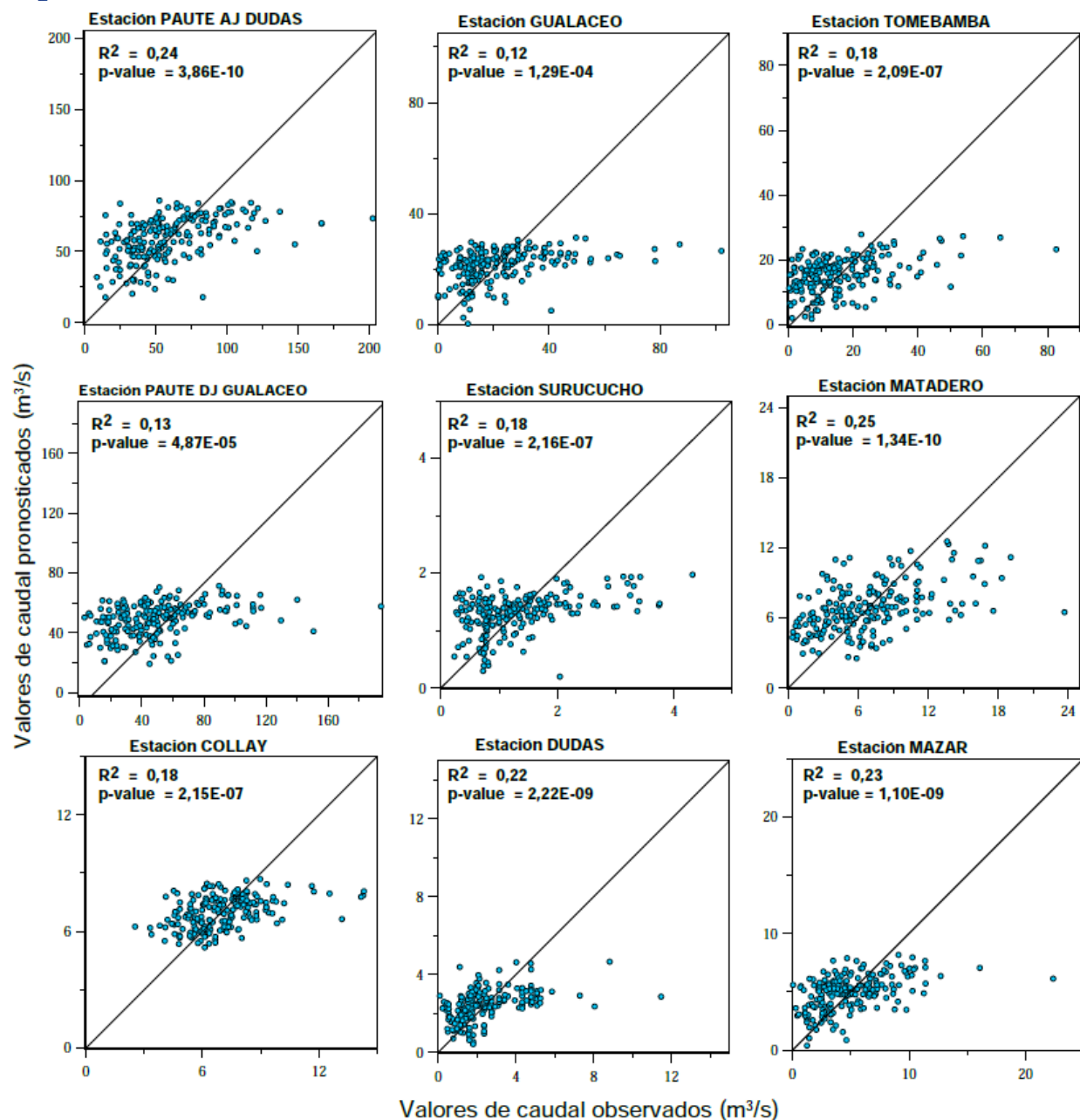


Figura 40. Resultados del análisis de dispersión el escenario con 31 índices con dos meses de retraso.

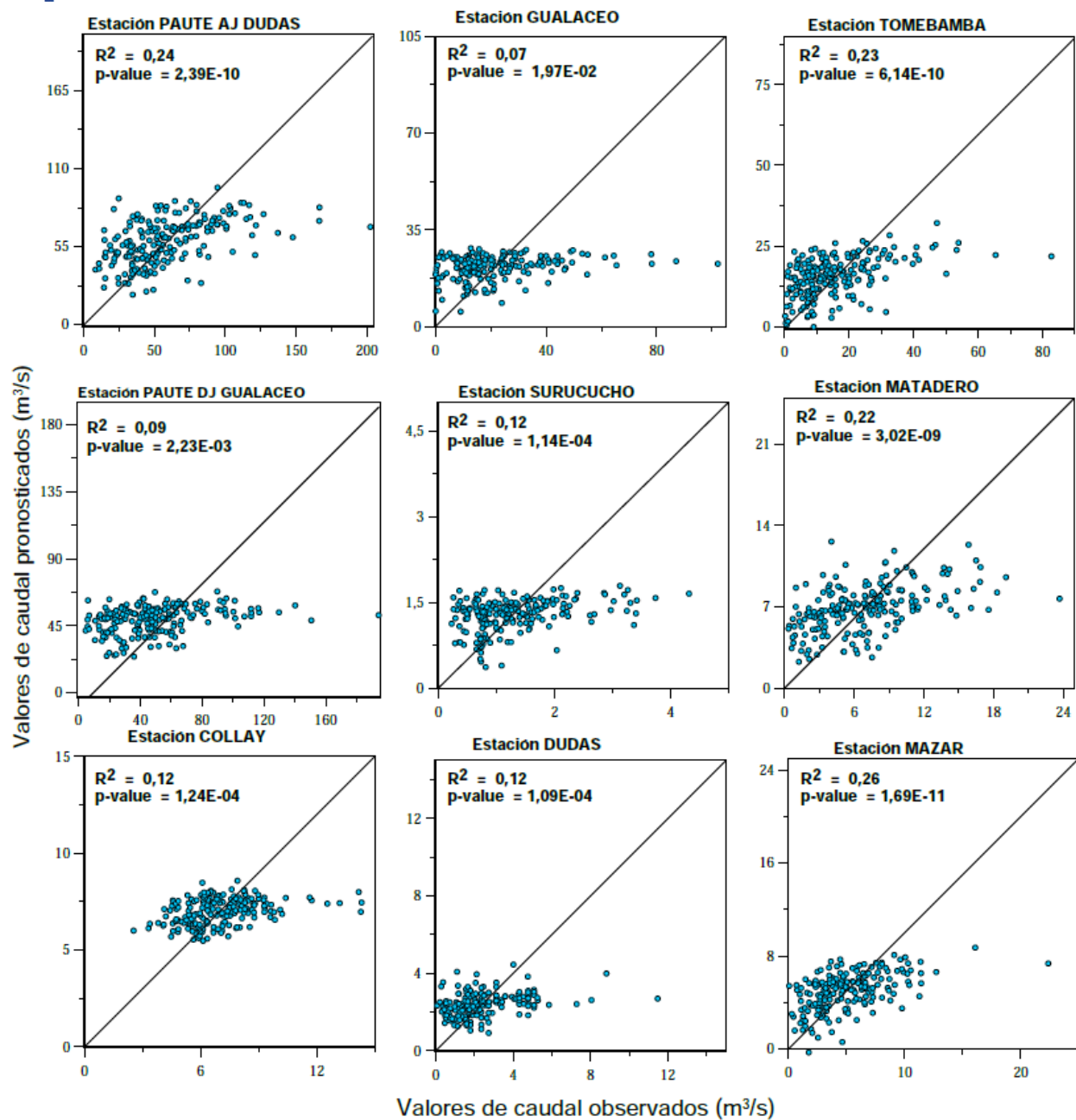


Figura 41. Resultados del análisis de dispersión el escenario con 31 índices con tres meses de retraso.

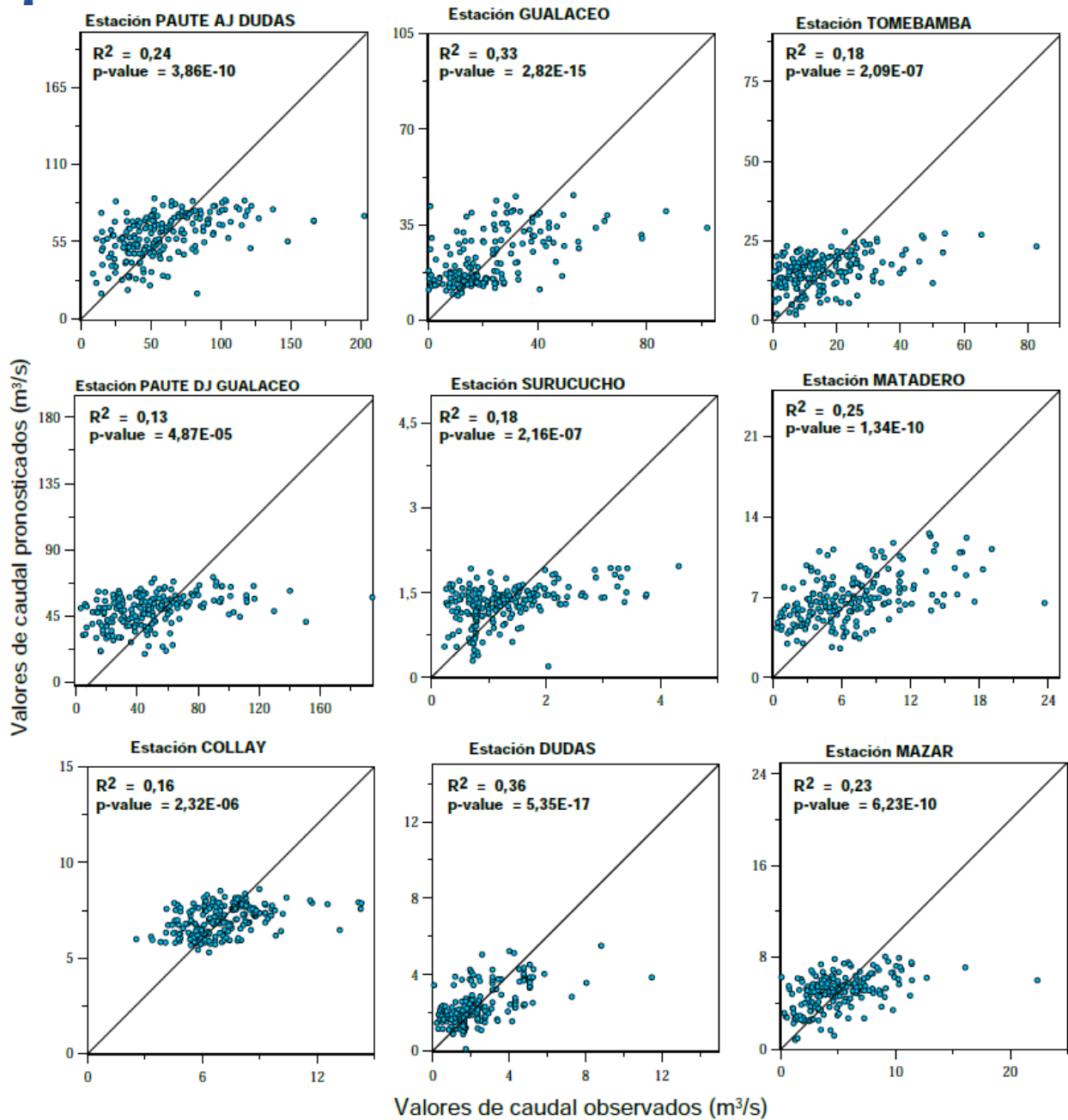


Figura 42. Resultados del análisis de dispersión el escenario con 41 índices con dos meses de retraso.

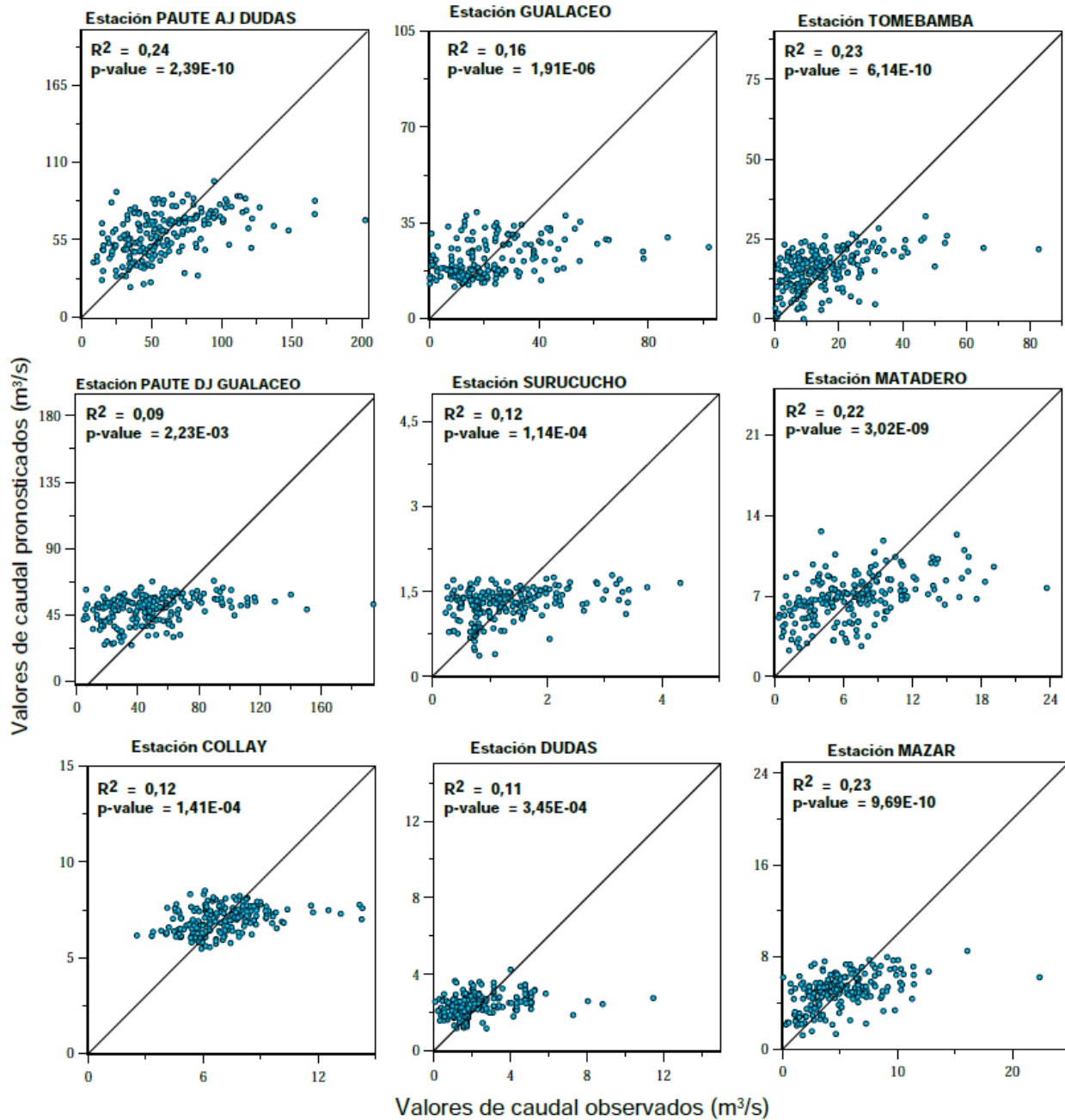


Figura 43. Resultados del análisis de dispersión el escenario con 41 índices con tres meses de retraso.

5.3 Anexo 3: Resultados de series de tiempo en la etapa de validación para los escenarios ACP, usando retrasos mensuales y VIIF.

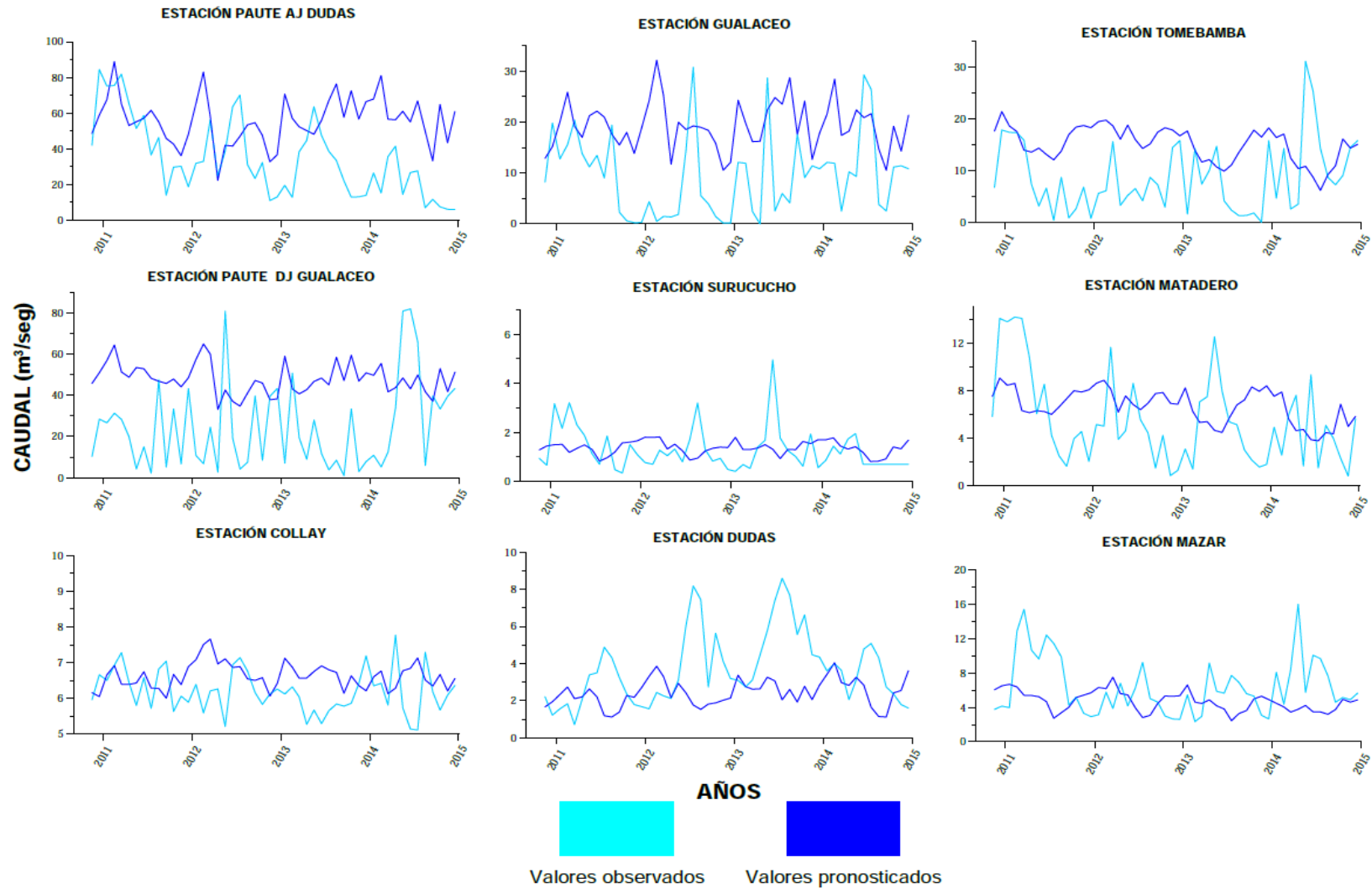


Figura 44. Resultados del escenario con 31 índices y 70% de varianza en el análisis ACP.

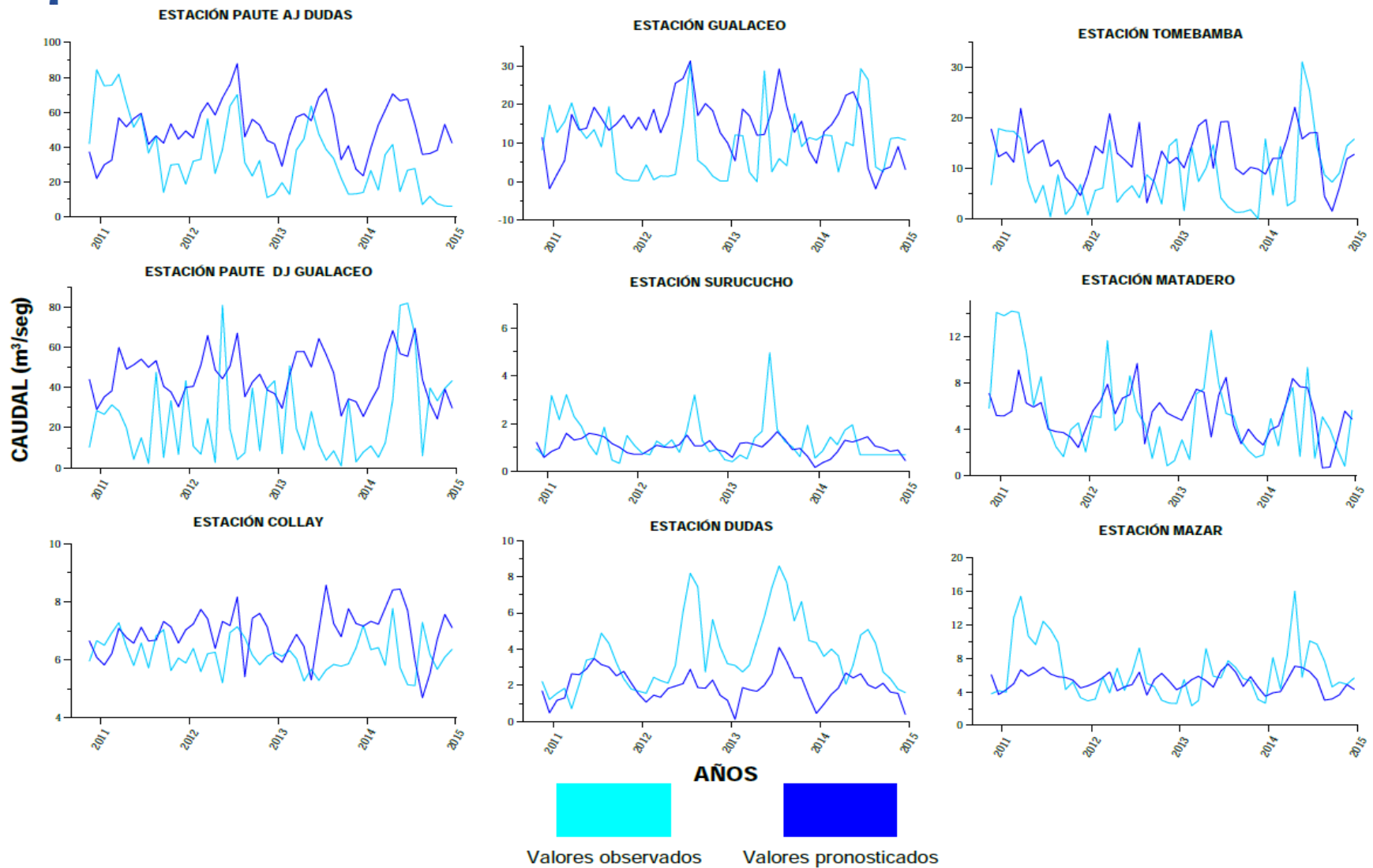


Figura 45. Resultados del escenario con 31 índices y 90% de varianza en el análisis ACP.

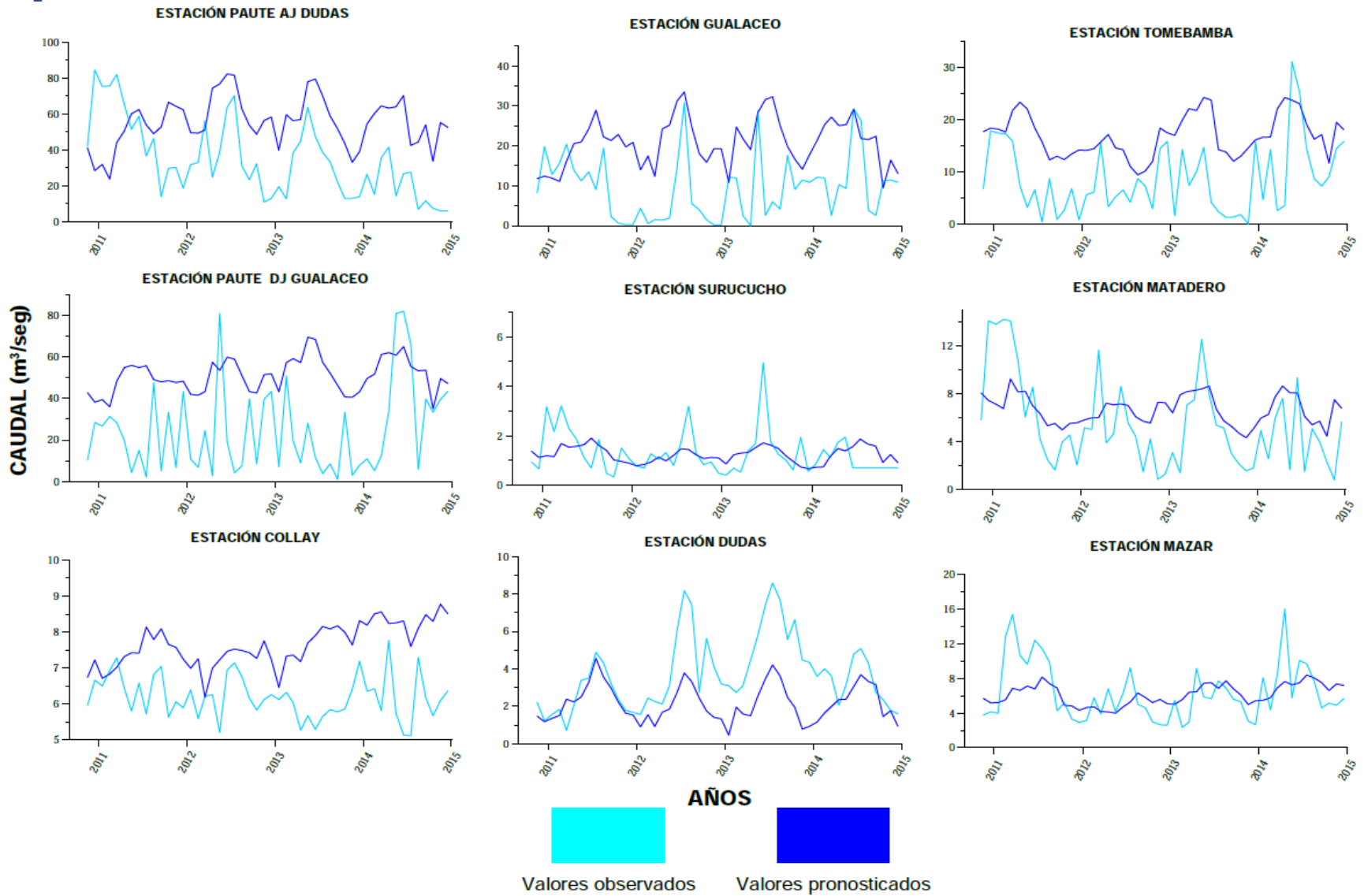


Figura 46. Resultados del escenario con 41 índices y 70% de varianza en el análisis ACP.

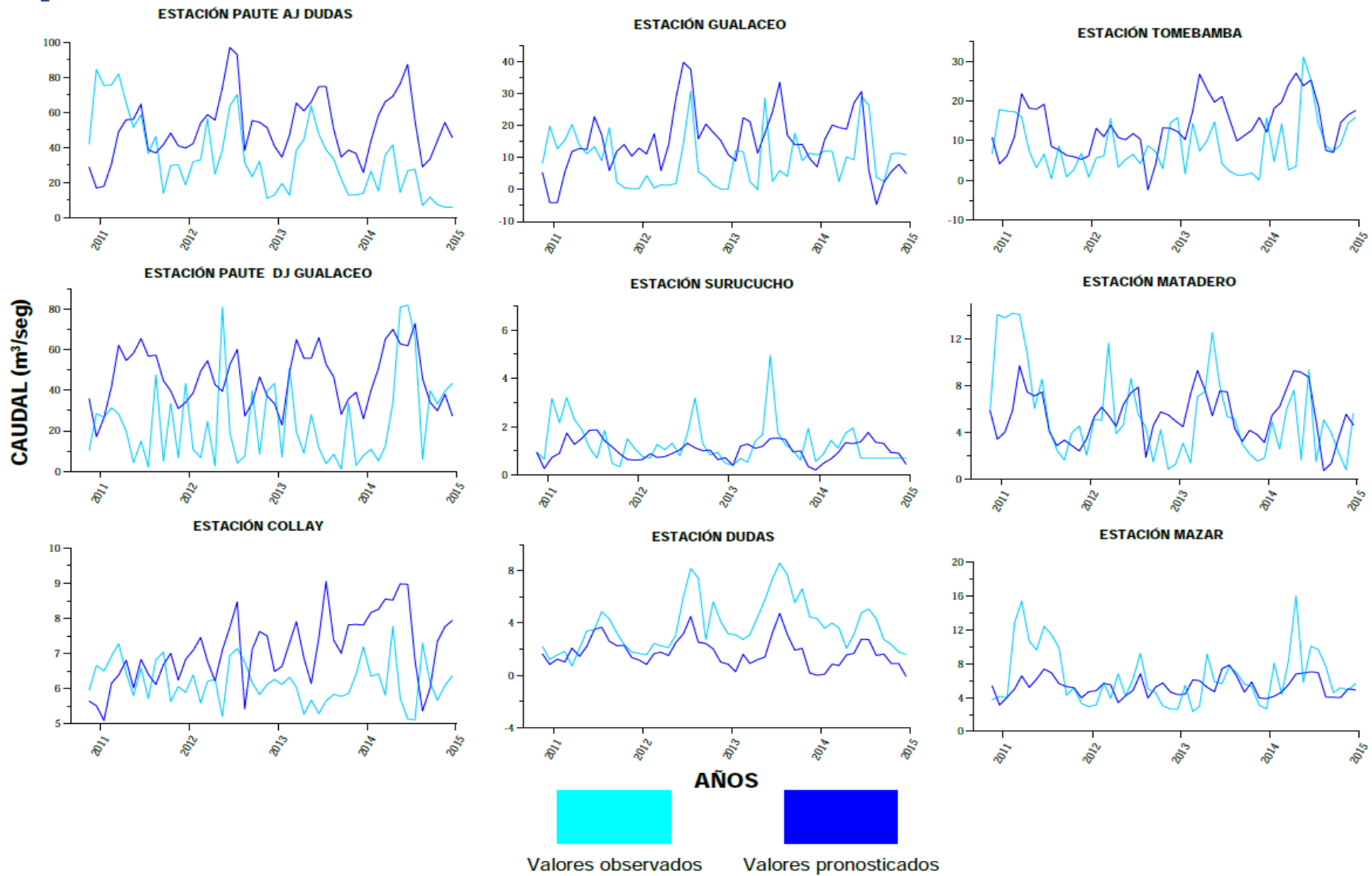


Figura 47. Resultados del escenario con 41 índices y 90% de varianza en el análisis ACP.

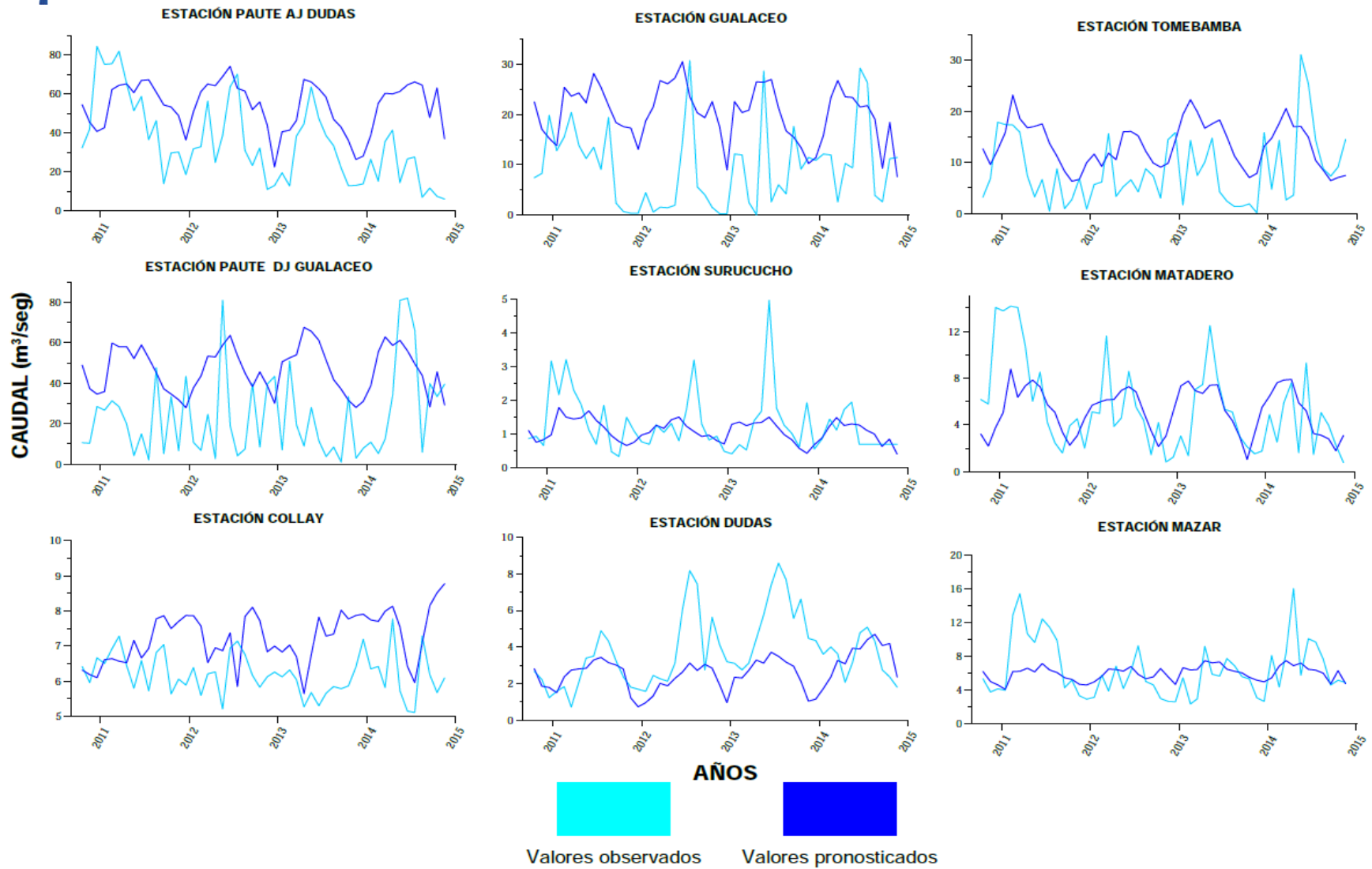


Figura 48. Resultados del escenario con 31 índices y un mes de retraso.

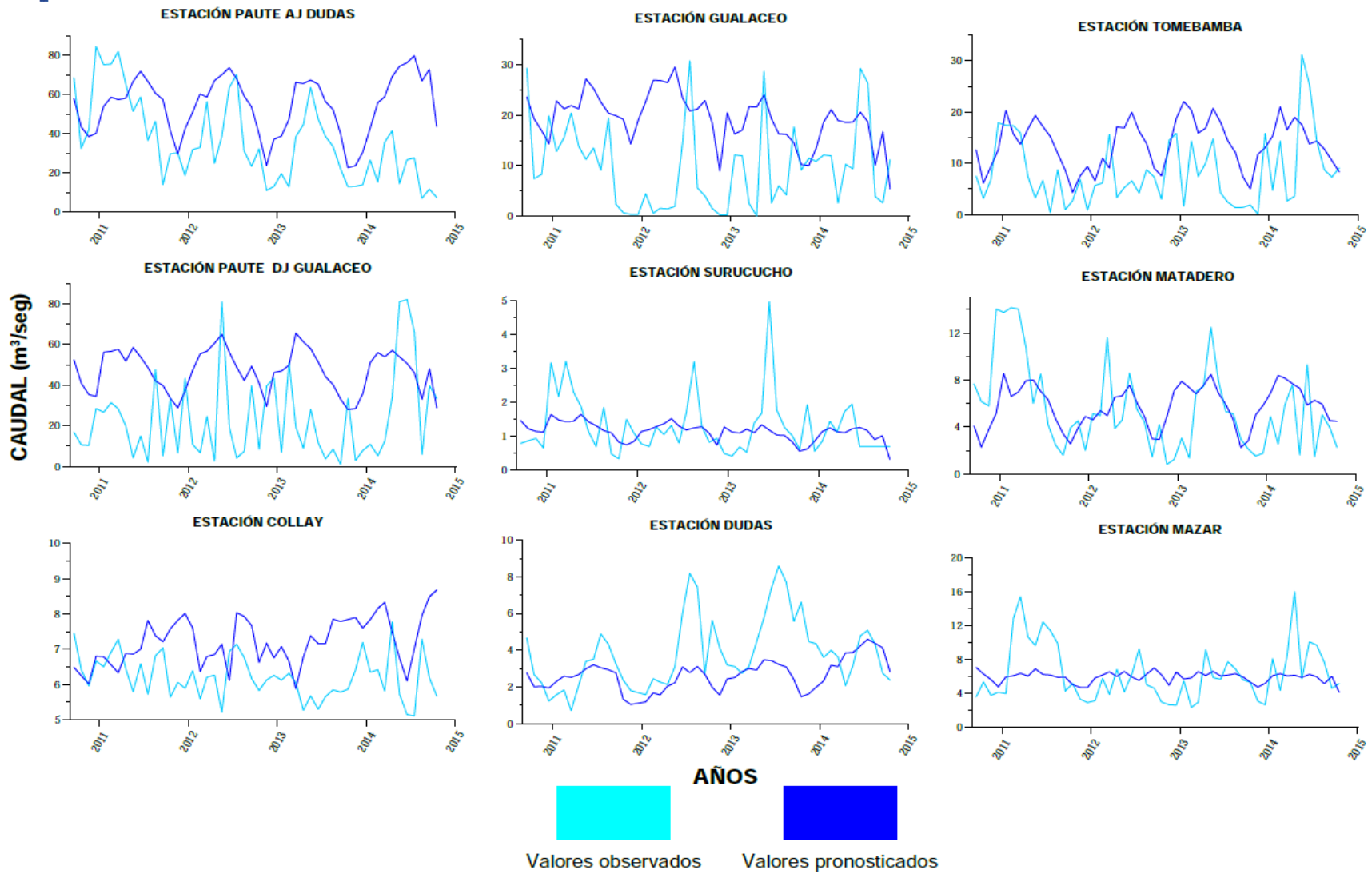


Figura 49. Resultados del escenario con 31 índices y dos meses de retraso.

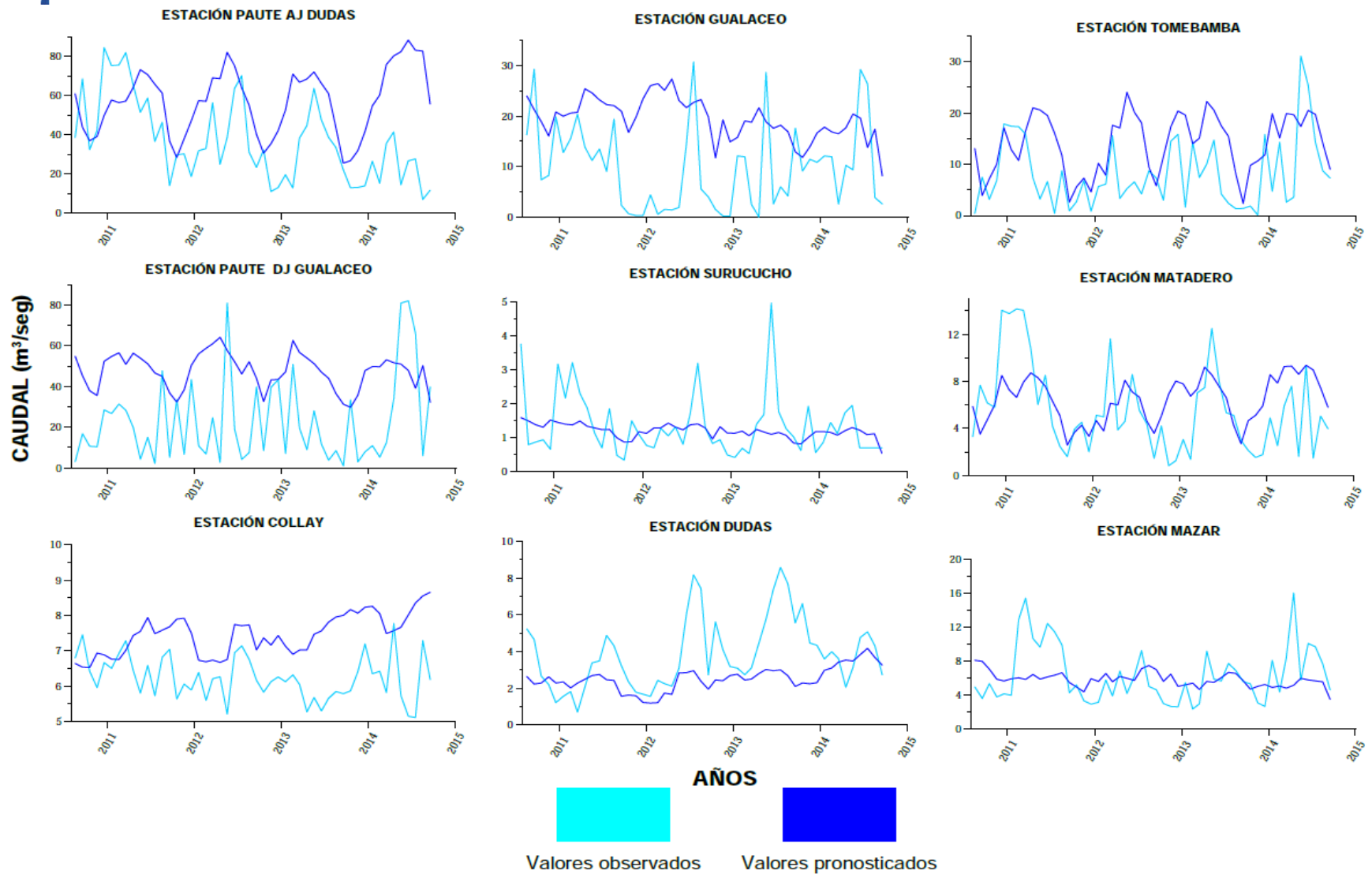


Figura 50. Resultados del escenario con 31 índices y tres meses de retraso.

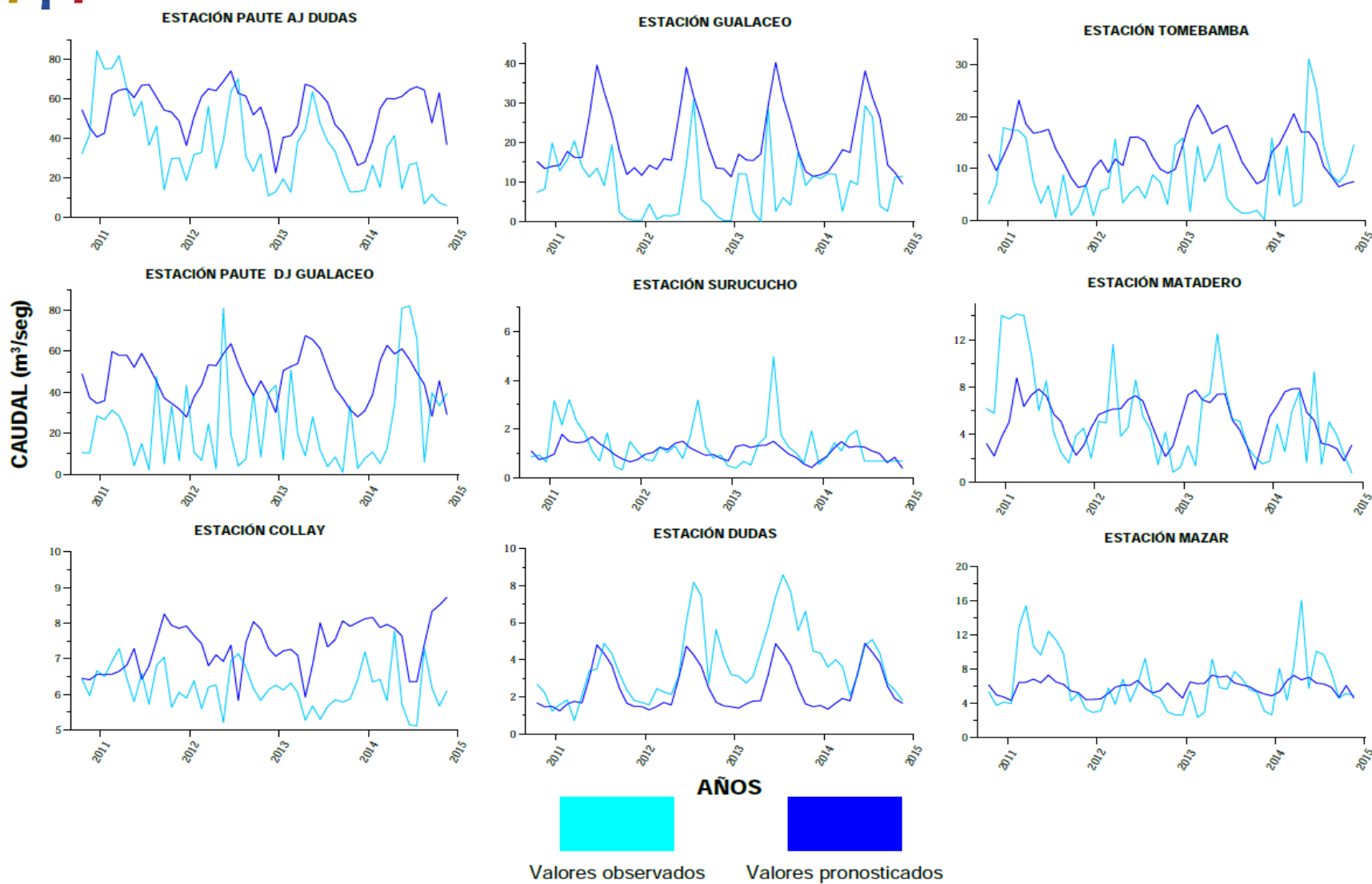


Figura 51. Resultados del escenario con 41 índices y un mes de retraso.

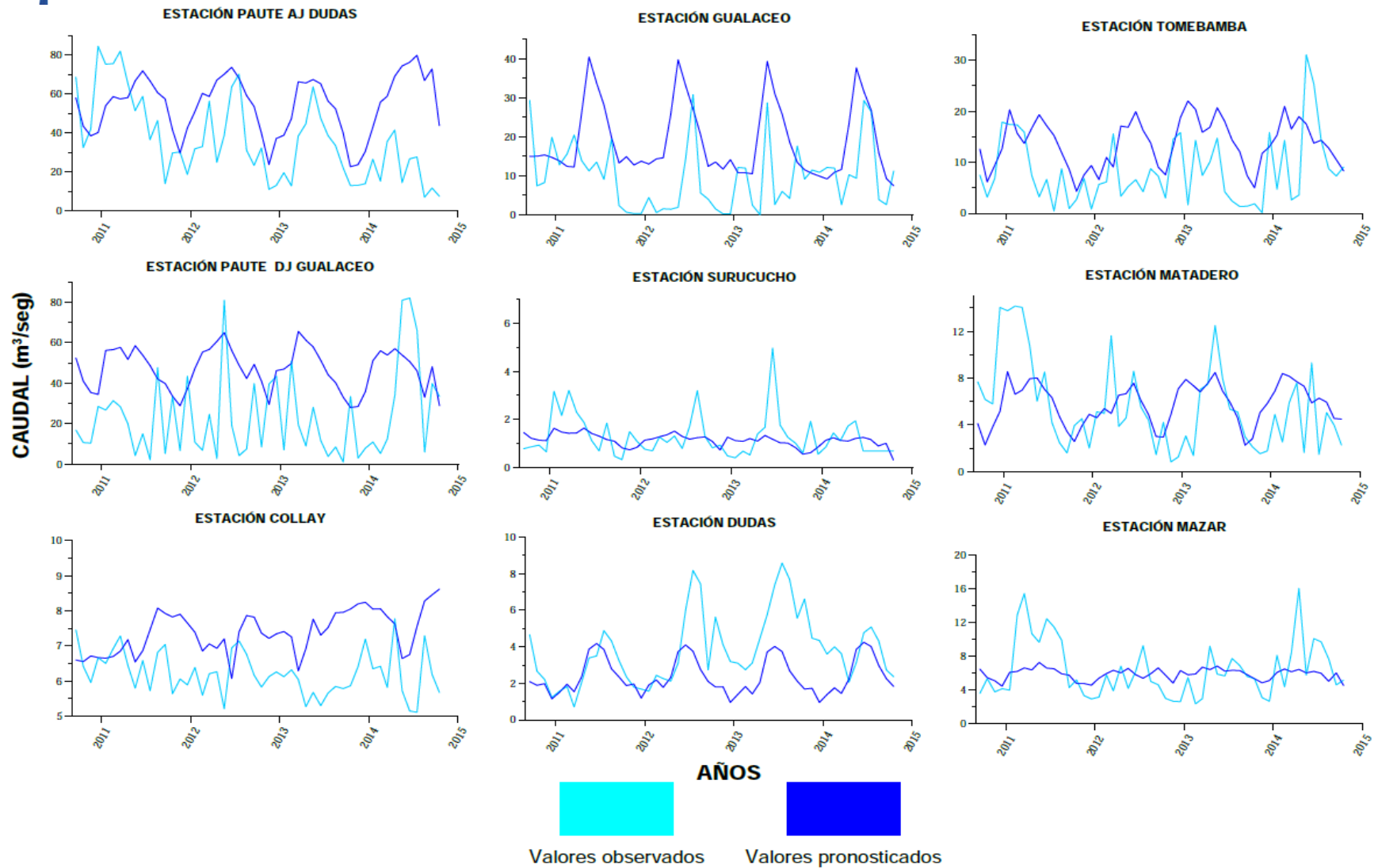


Figura 52. Resultados del escenario con 41 índices y dos meses de retraso.

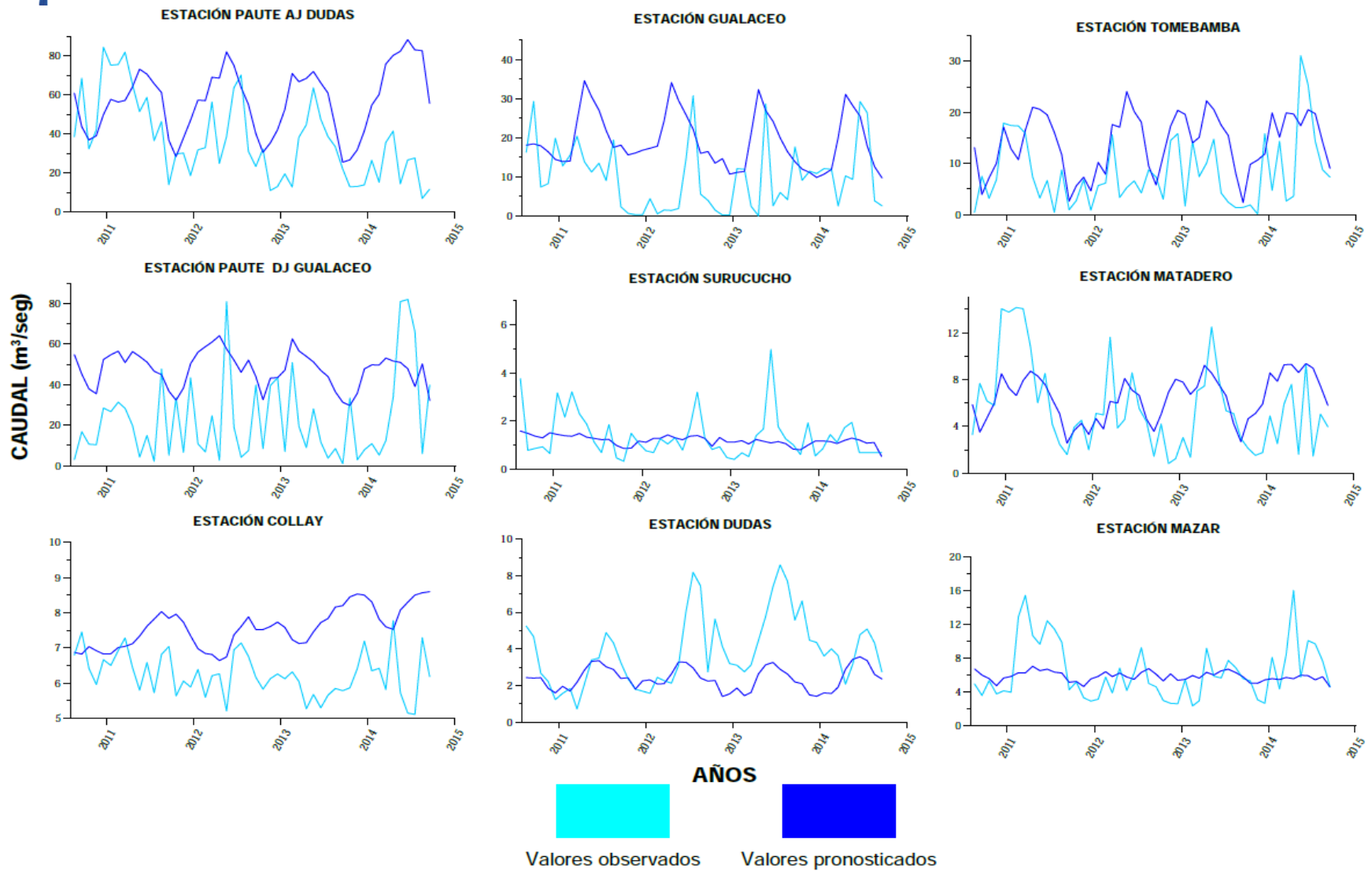


Figura 53. Resultados del escenario con 41 índices y tres meses de retraso.

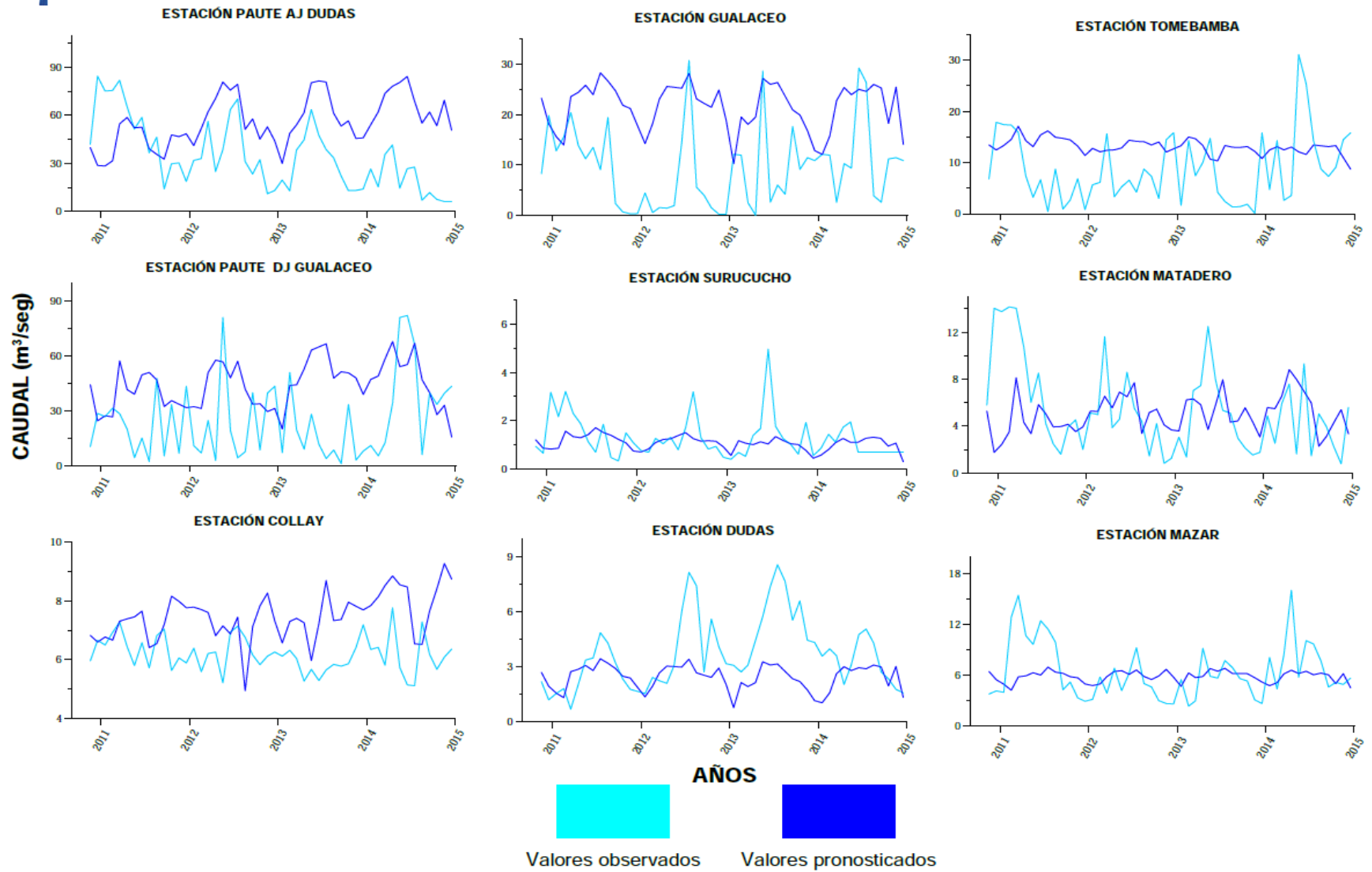


Figura 54. Resultados del escenario con 31 índices y análisis de multicolinealidad.

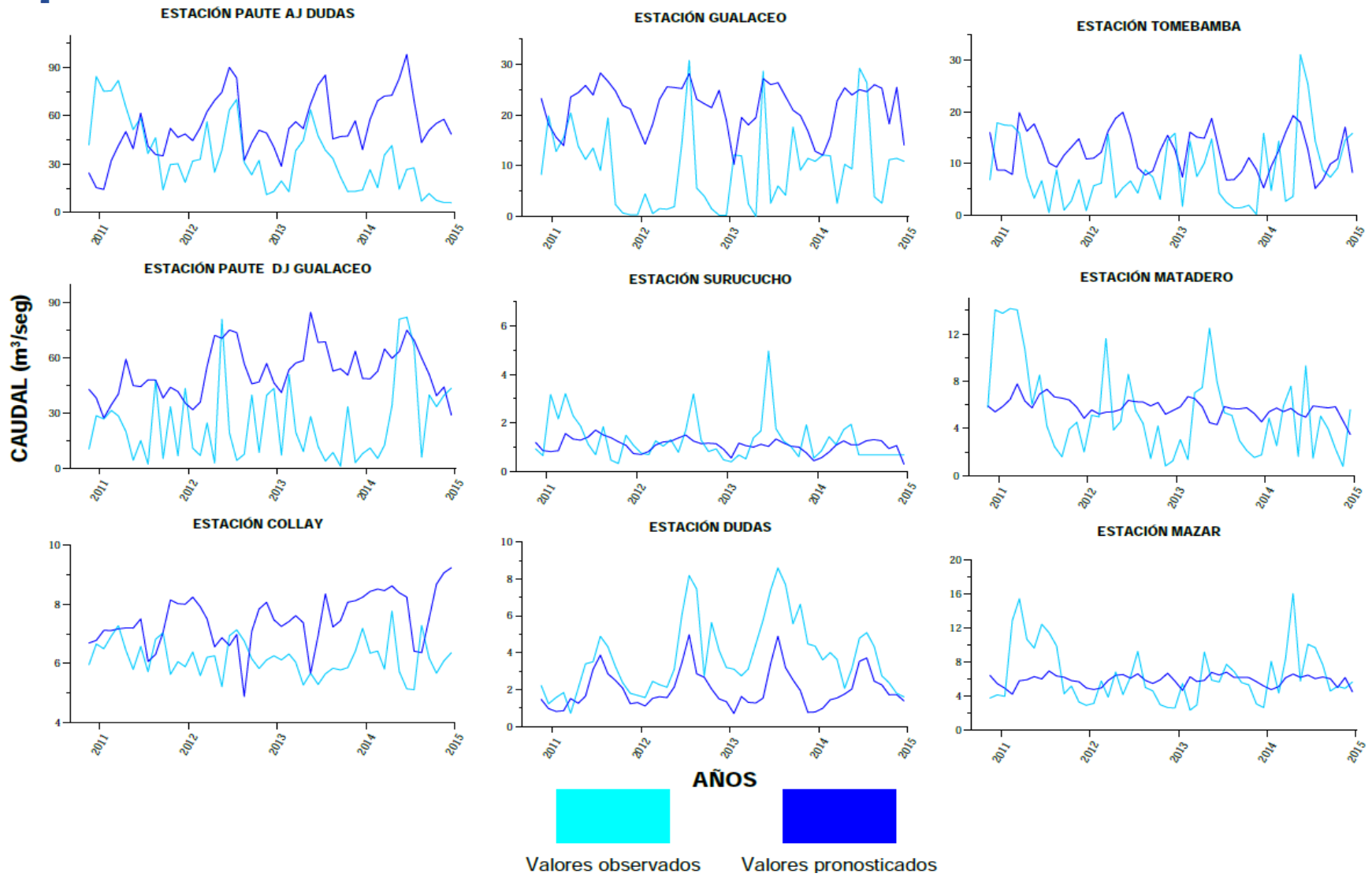


Figura 55. Resultados del escenario con 41 índices y análisis de multicolinealidad.

5.4 Anexo 4: Resultados de diagramas de dispersión en la etapa de validación para los escenarios ACP, usando retrasos mensuales y VIF.

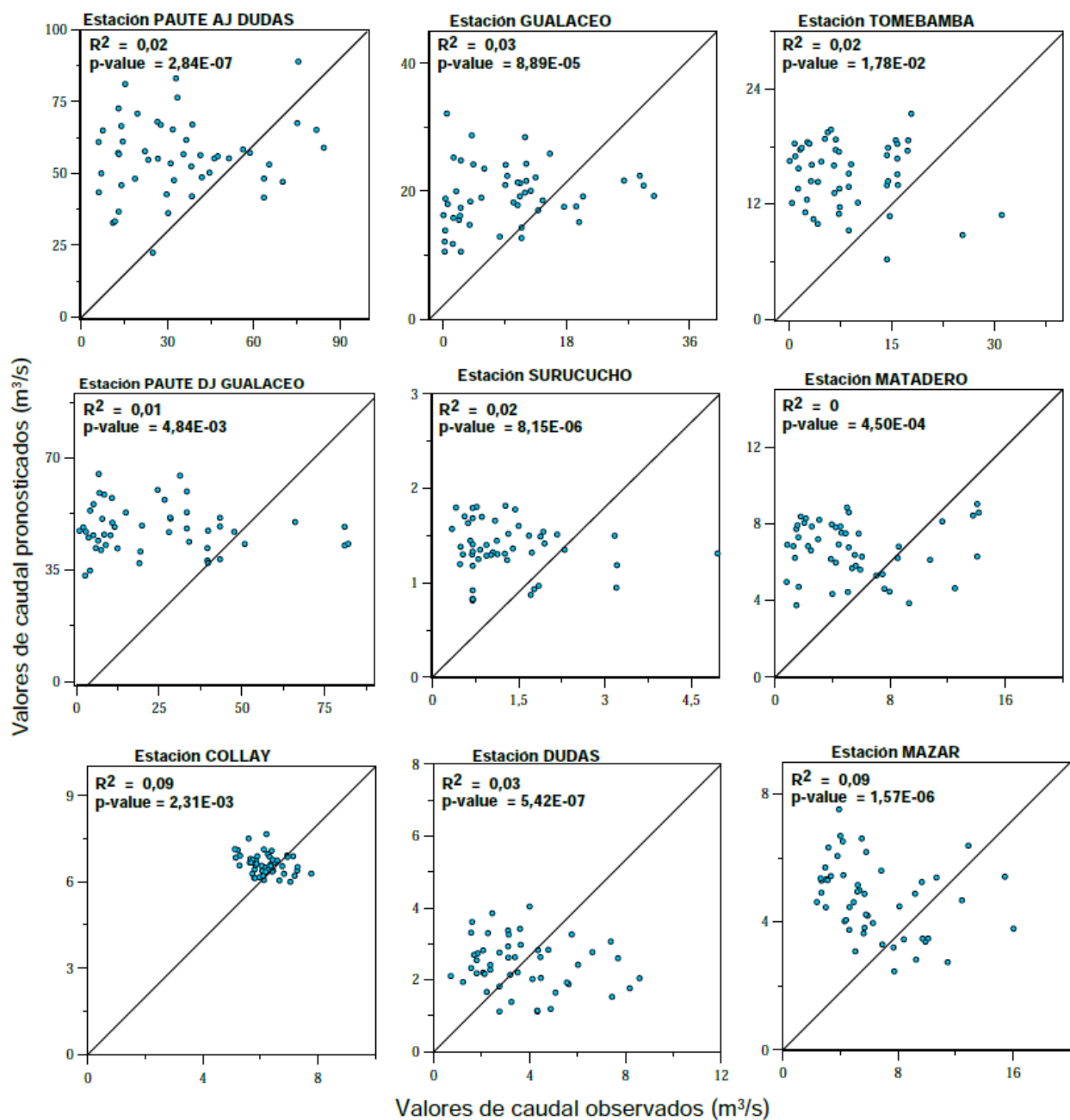


Figura 56. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 70% de varianza.

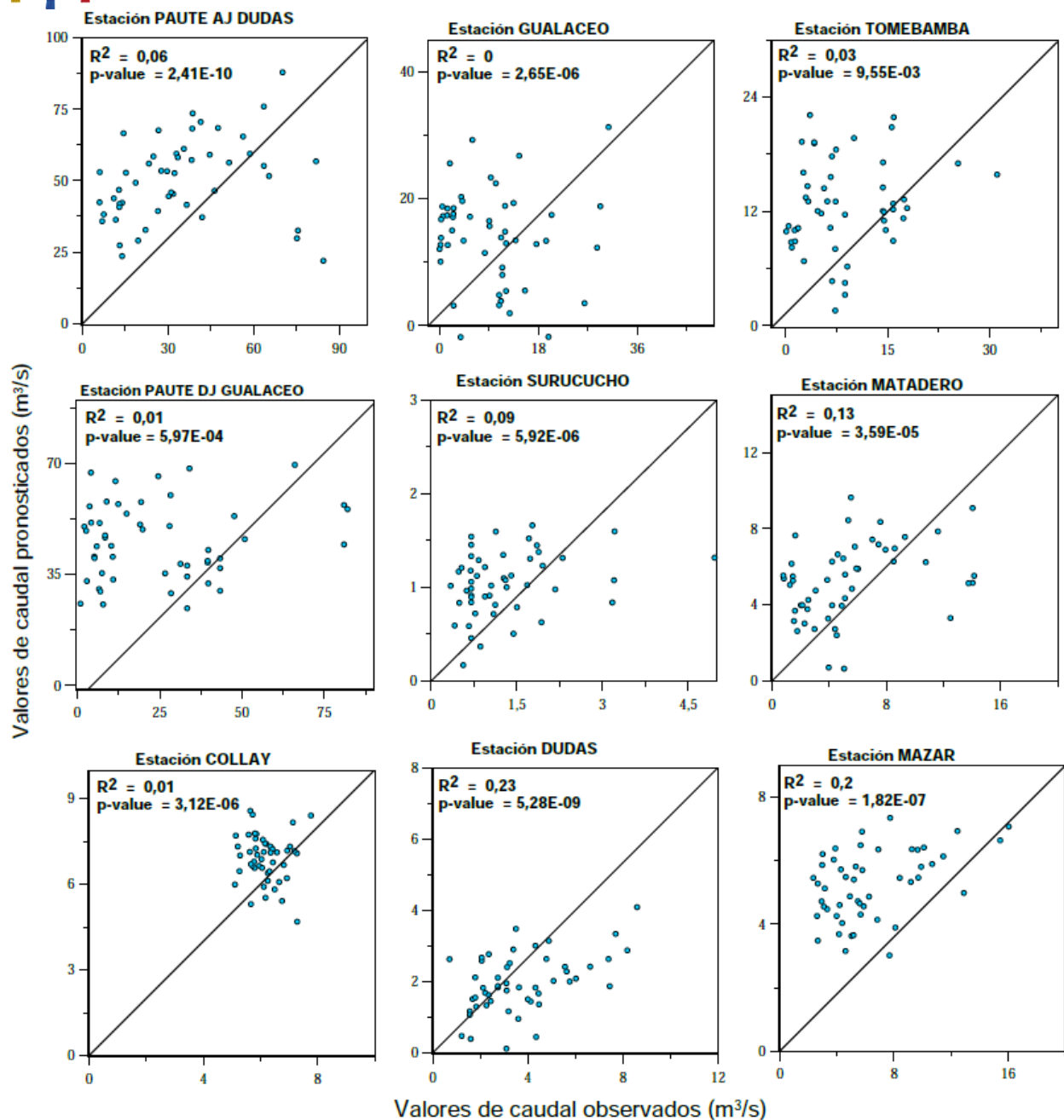


Figura 57. Resultados del análisis de dispersión entre el escenario con 31 índices con el análisis de ACP y 90% de varianza.

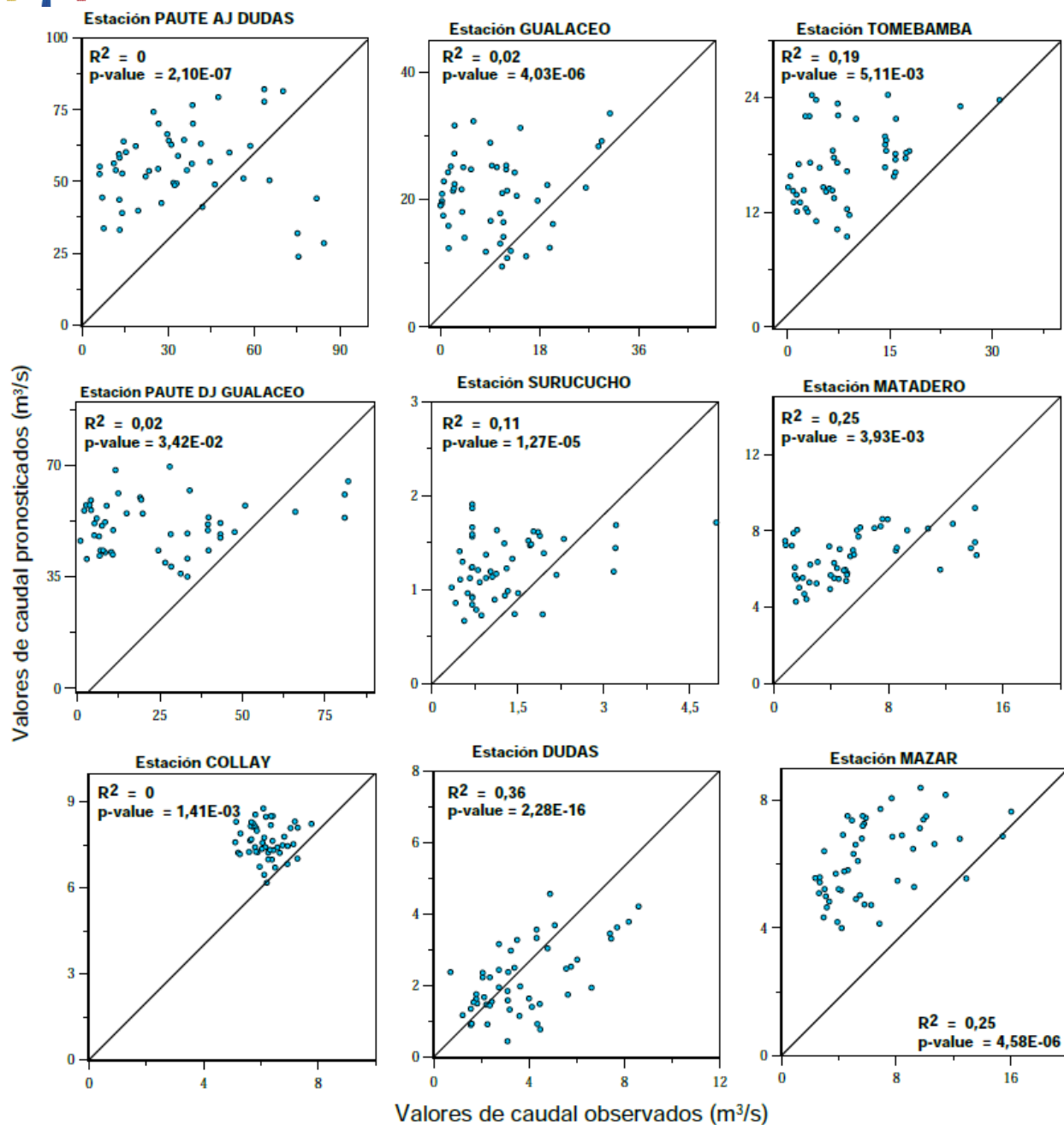


Figura 58. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 70% de varianza

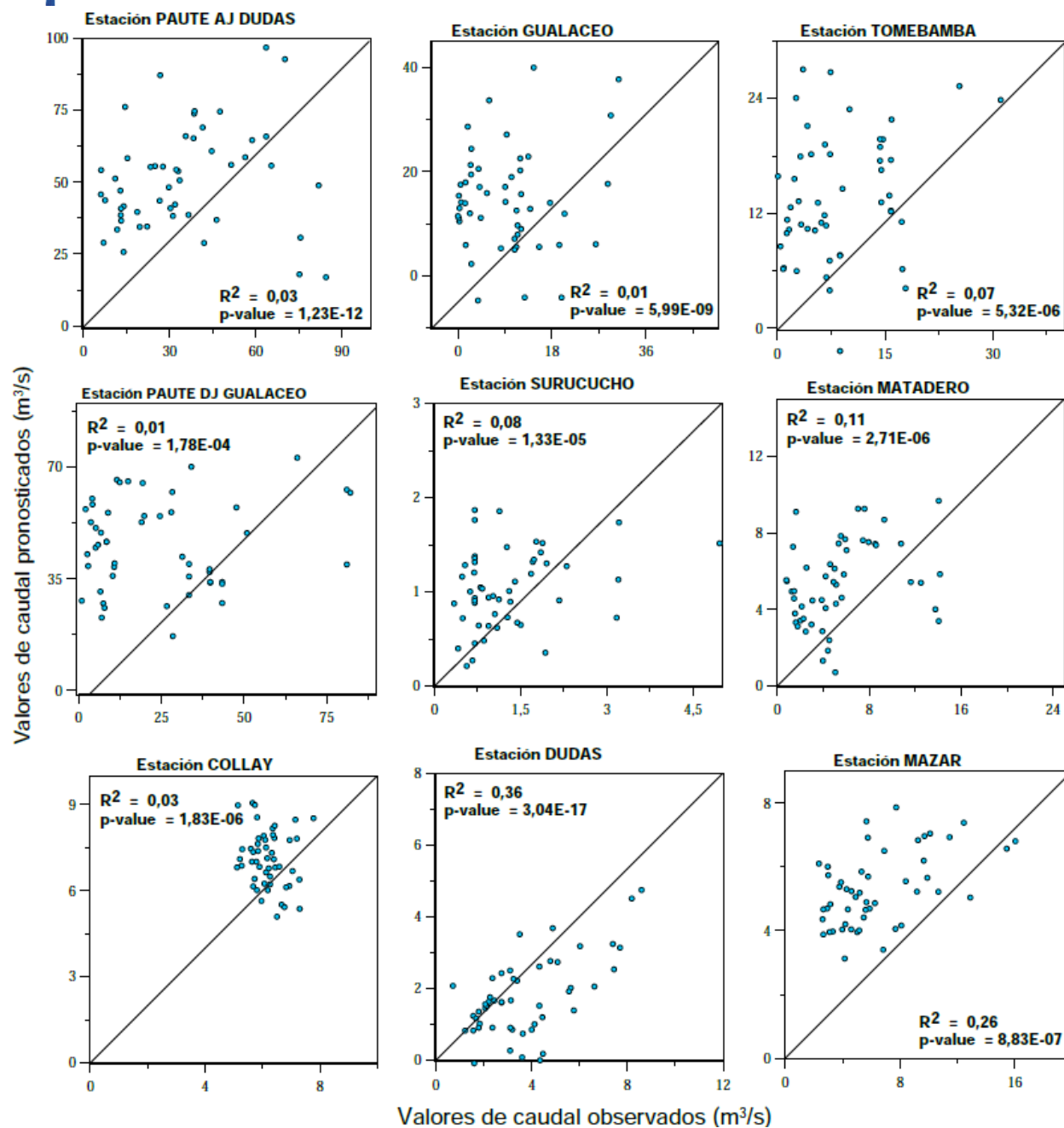


Figura 59. Resultados del análisis de dispersión entre el escenario con 41 índices con el análisis de ACP y 90% de varianza

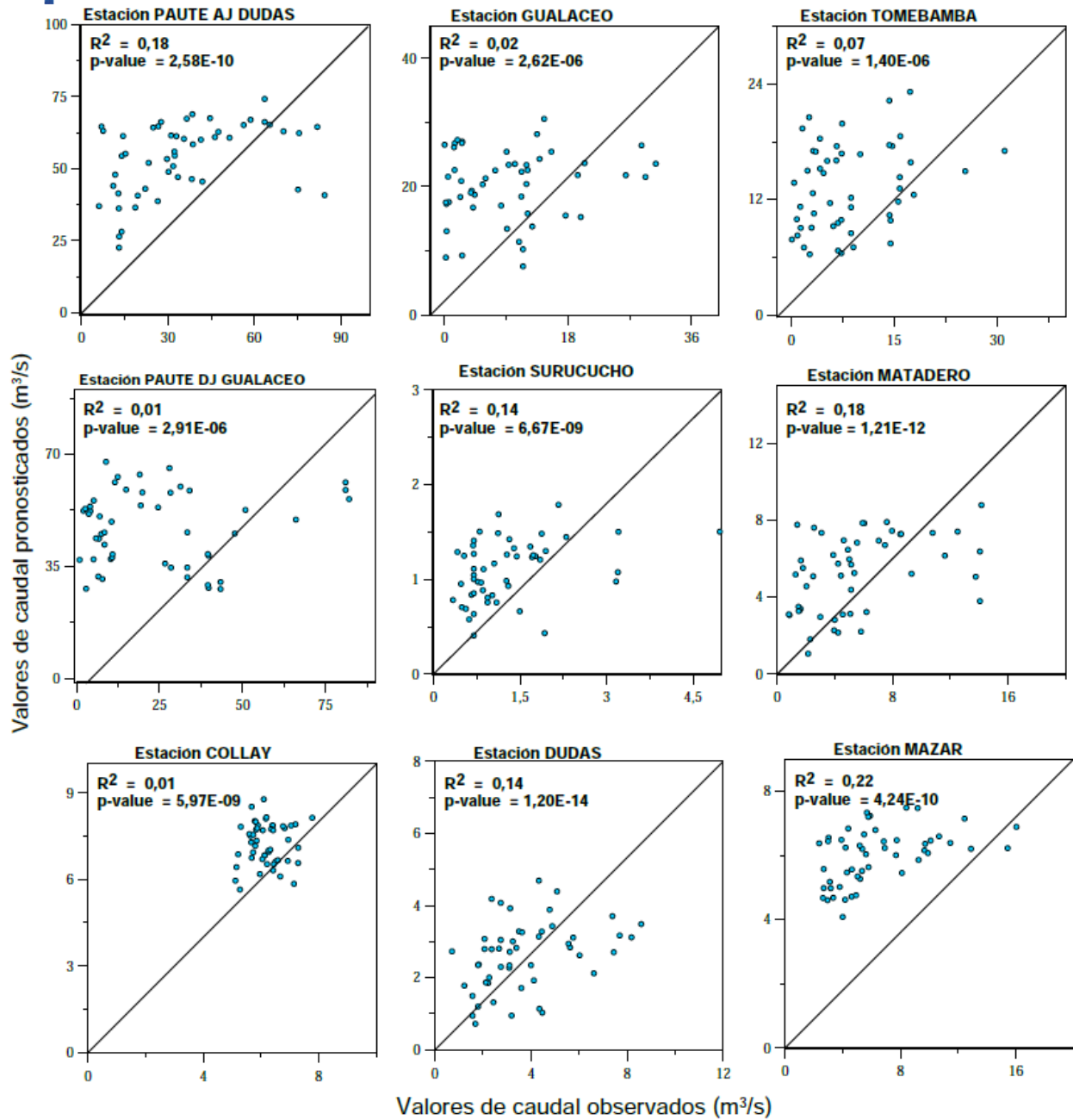


Figura 60. Resultados del análisis de dispersión el escenario con 31 índices con un mes de retraso.

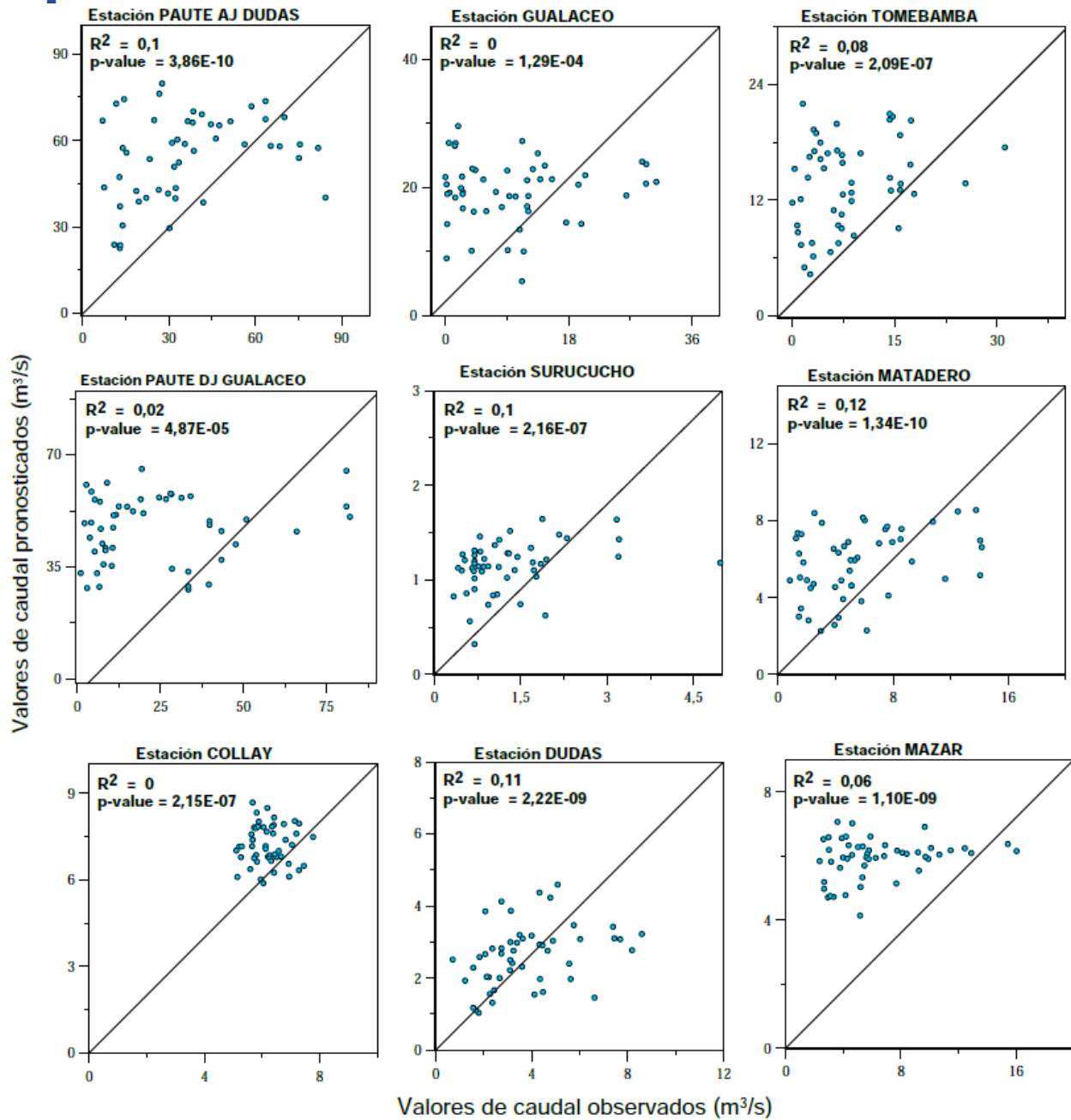


Figura 61. Resultados del análisis de dispersión el escenario con 31 índices con dos meses de retraso.

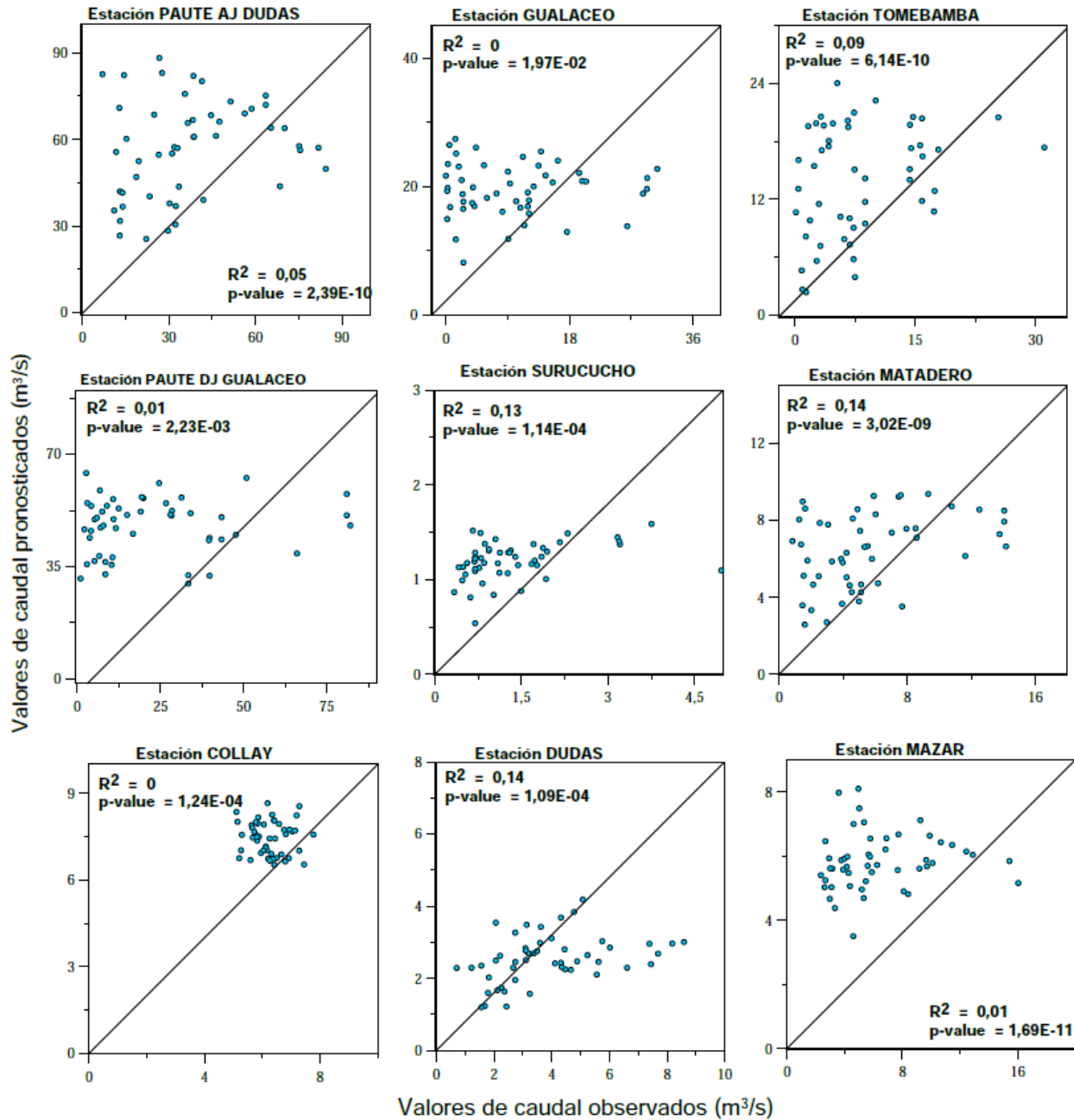


Figura 62. Resultados del análisis de dispersión el escenario con 31 índices con tres meses de retraso.

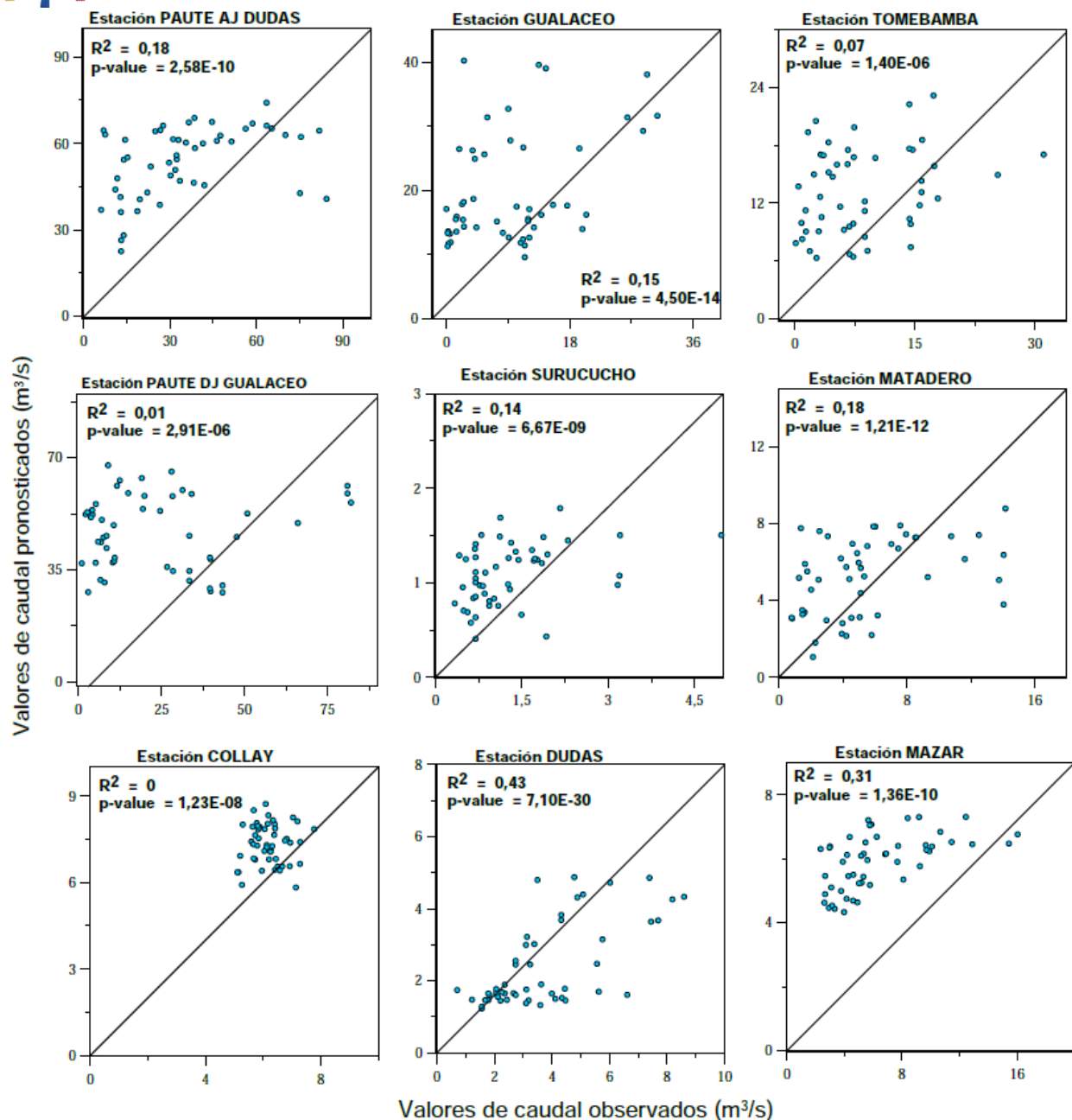


Figura 63. Resultados del análisis de dispersión el escenario con 41 índices con un mes de retraso.

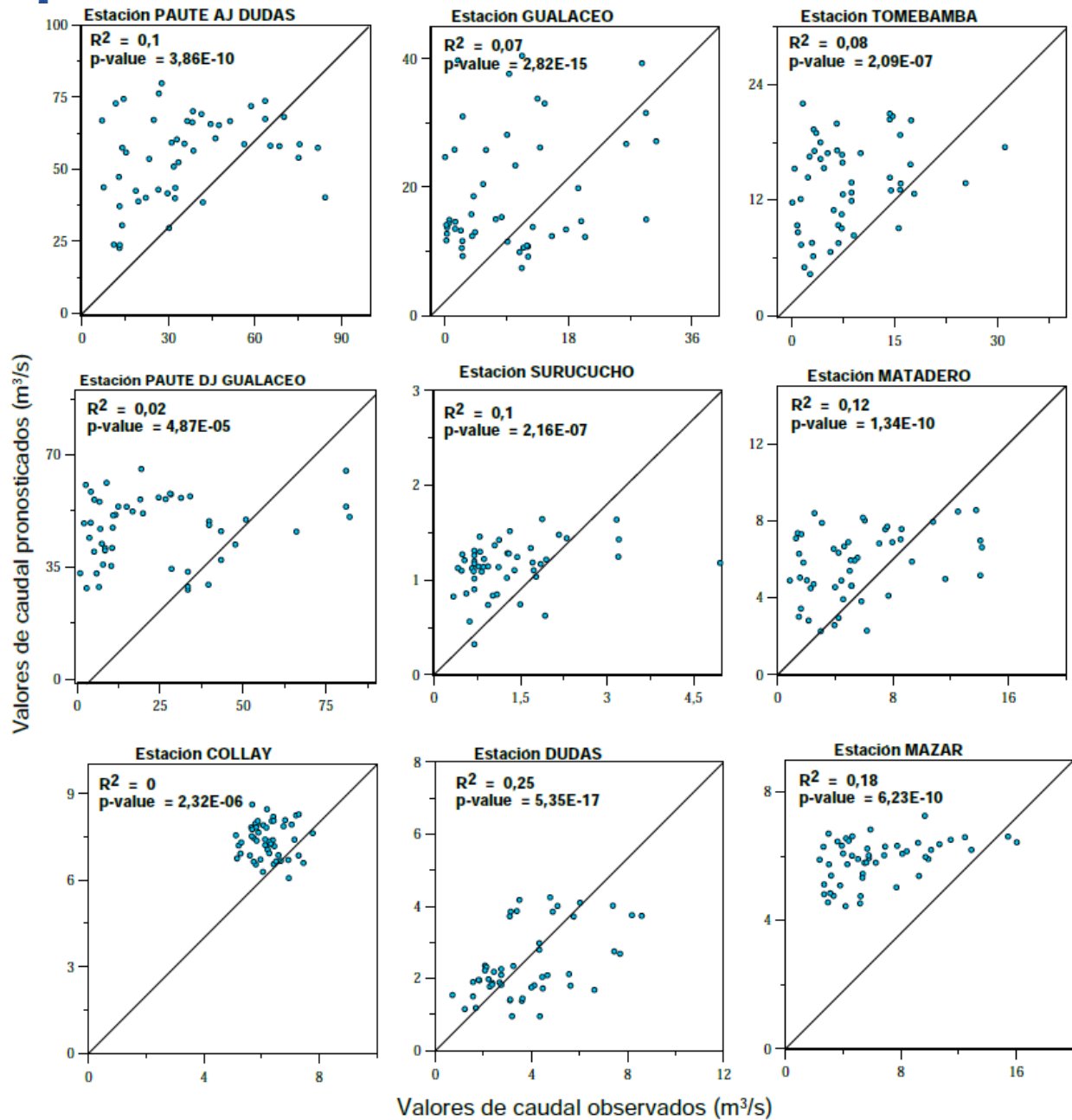


Figura 64. Resultados del análisis de dispersión el escenario con 41 índices con dos meses de retraso.

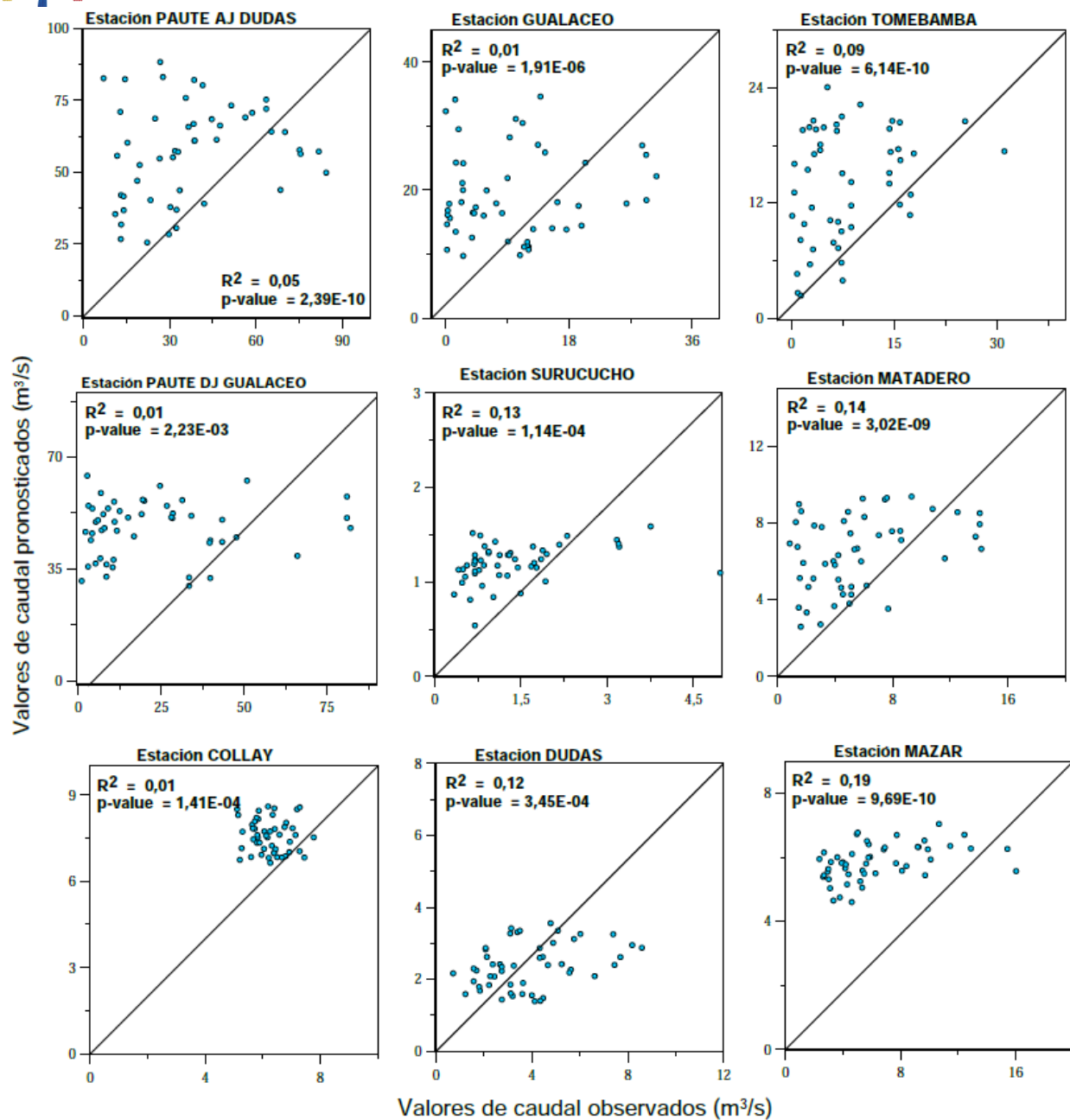


Figura 65. Resultados del análisis de dispersión el escenario con 41 índices con tres meses de retraso.

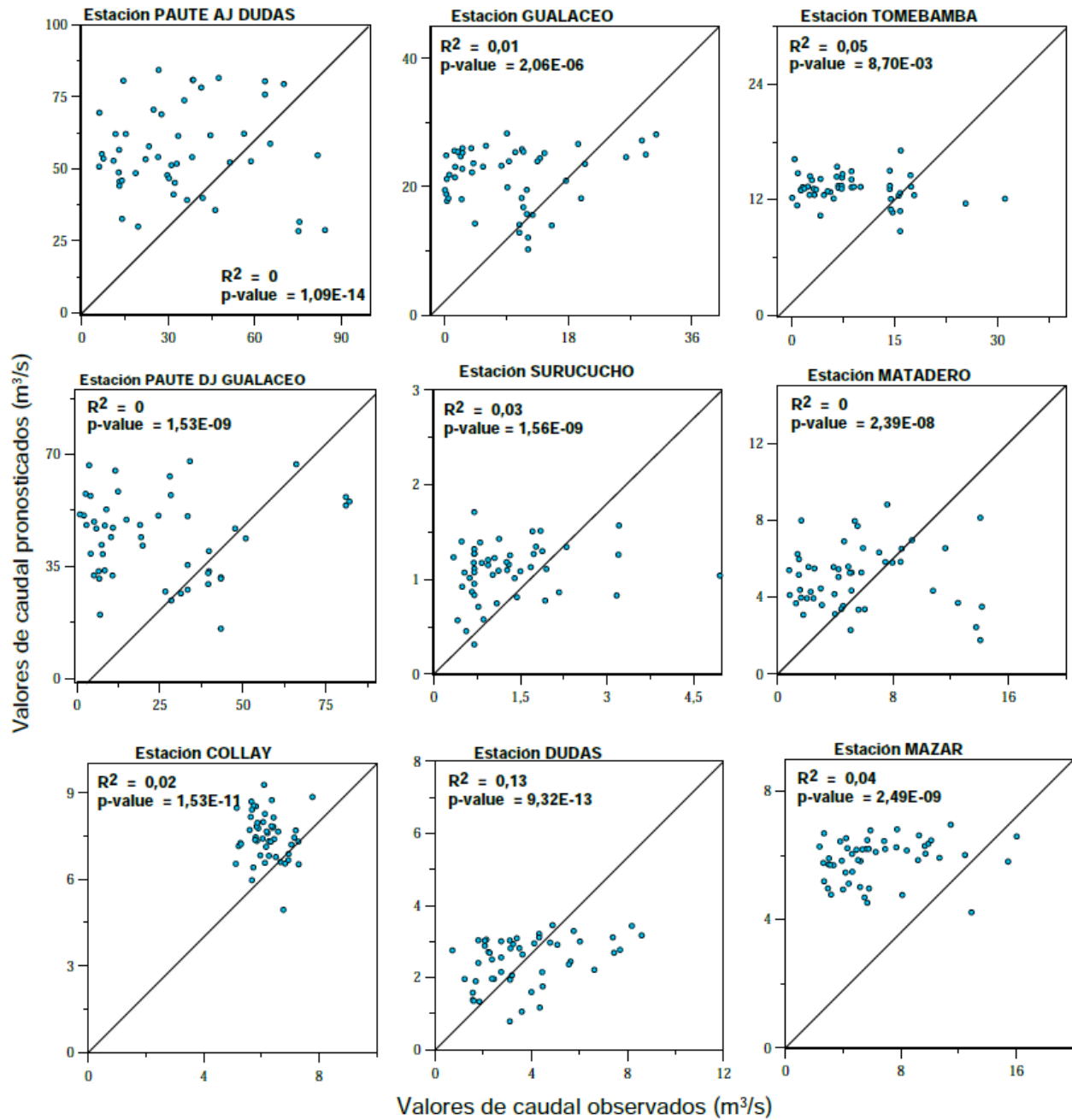


Figura 66. Resultados del análisis de dispersión entre la serie de tiempo predicha con 31 índices con el análisis de multicolinealidad.

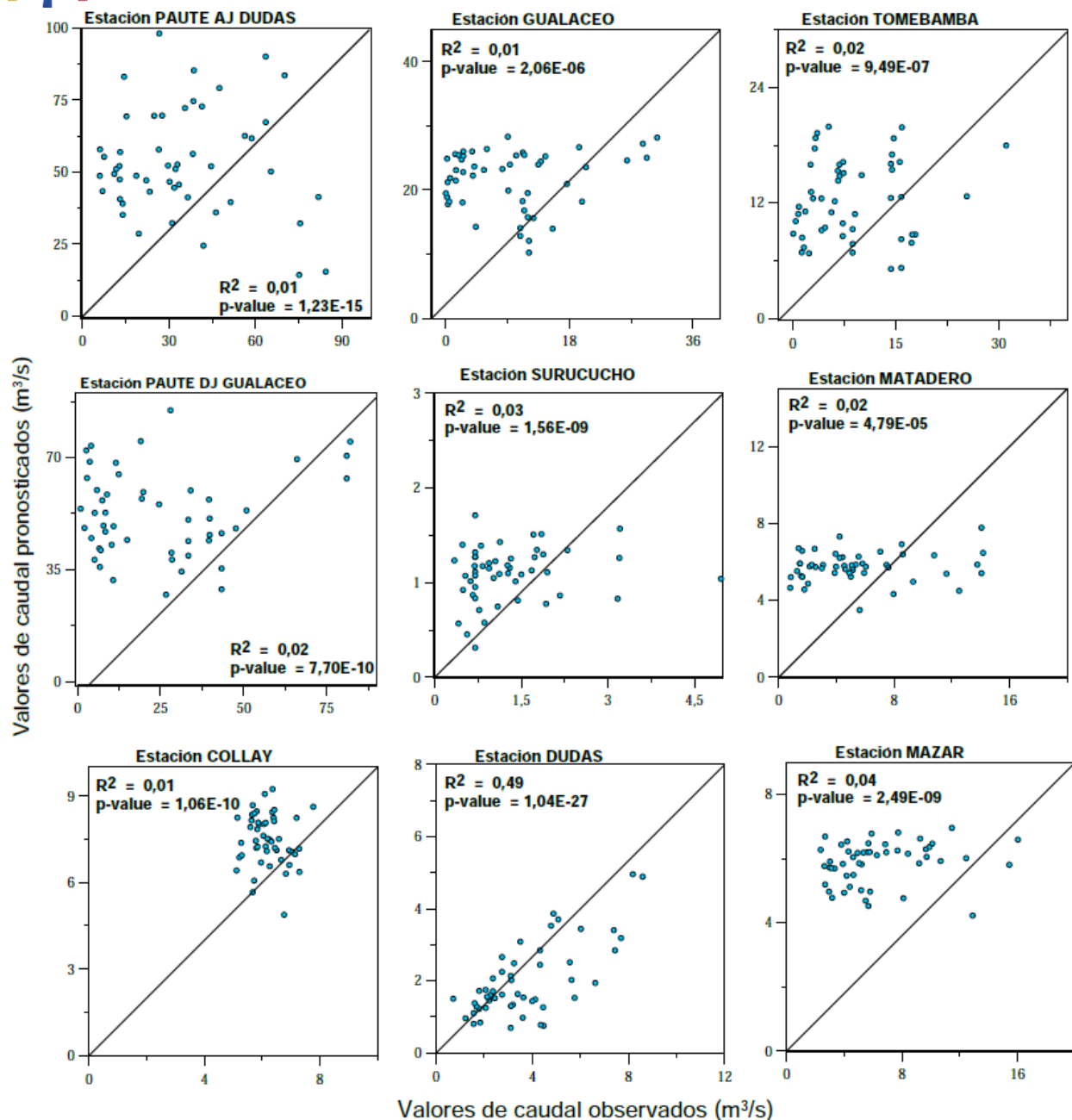


Figura 67. Resultados del análisis de dispersión entre la serie de tiempo predicha con 41 índices con el análisis de multicolinealidad.

5.5 Anexo 5: Resultados de las métricas en la etapa de validación.

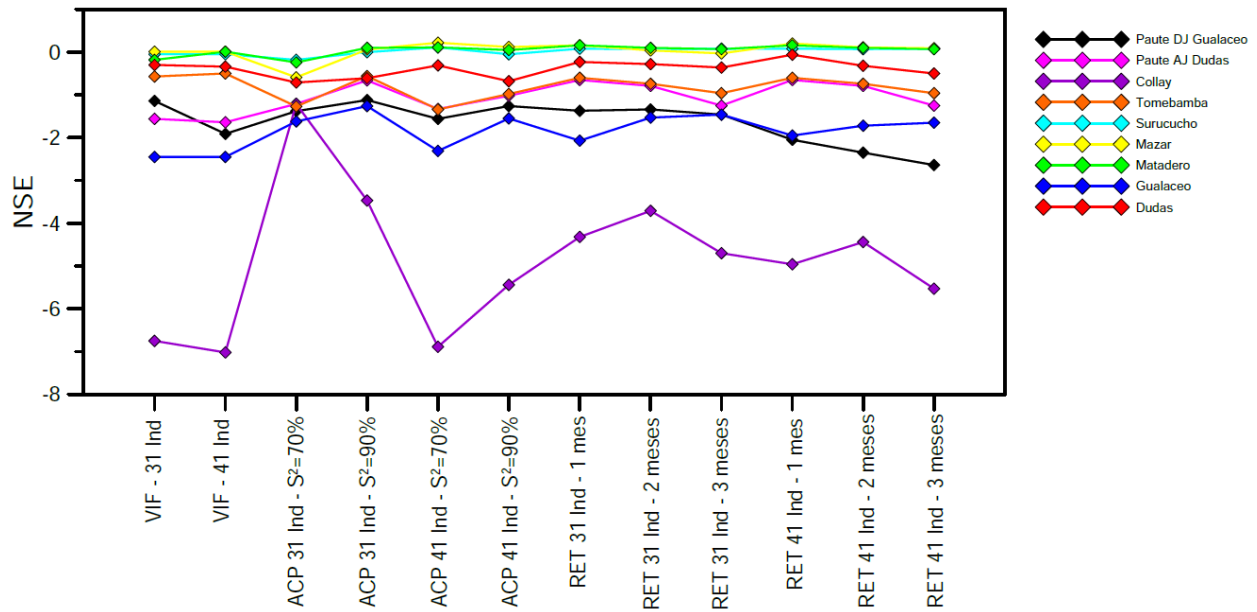


Figura 68. Resultados de NSE para cada escenario de predicción de las estaciones de caudal seleccionadas.

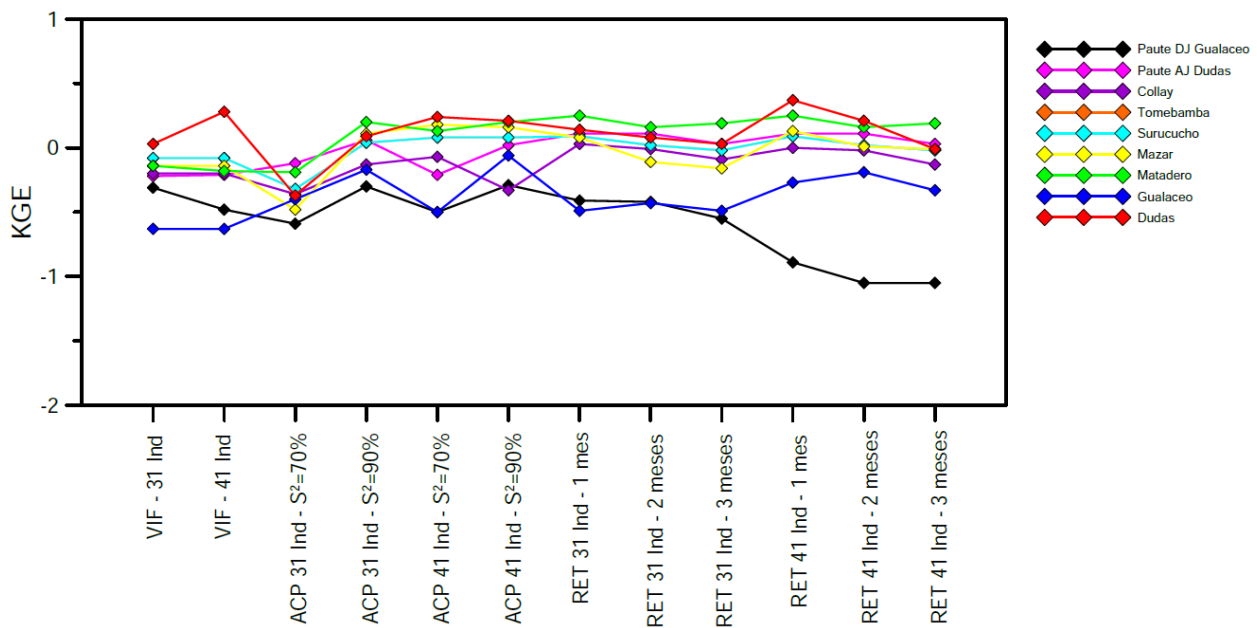


Figura 69. Resultados de KGE para cada escenario de predicción de las estaciones de caudal seleccionadas.

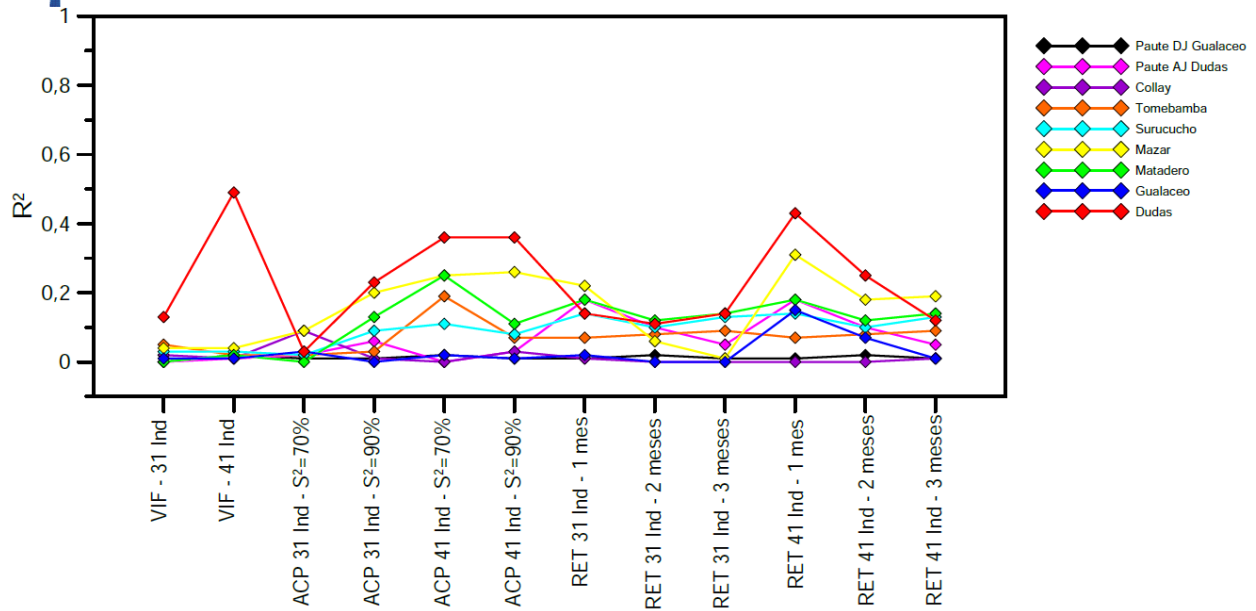


Figura 70. Resultados de R^2 para cada escenario de predicción de las estaciones de caudal seleccionadas.

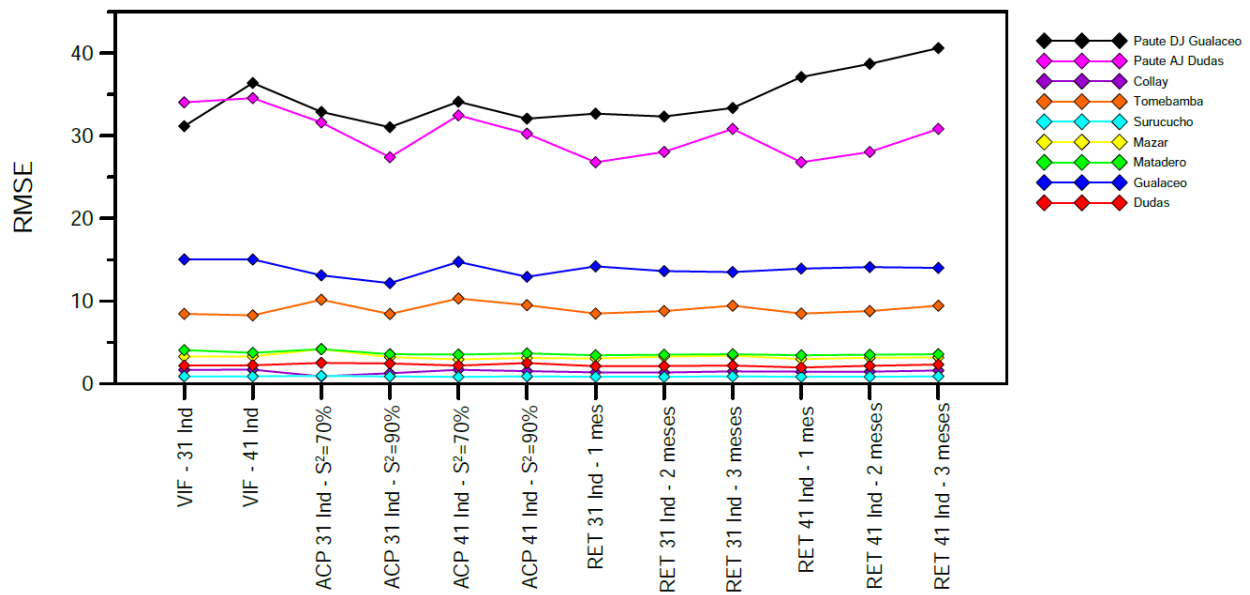


Figura 71. Resultados de RSME para cada escenario de predicción de las estaciones de caudal seleccionadas.