

Camera based Water Stage and Discharge Prediction with Machine Learning

Ernesto Adrián Álvarez Salazar A00227490

Carlos Javier Leal Beltrán A01741355

Carlos Moisés Chávez Jiménez A01637322

Luis Armando Salazar López A01114901

Septiembre 2022 - Diciembre 2022

1 Introducción

La medición y el modelado precisos del nivel y la descarga de la corriente son importantes para la gestión diaria del agua, el pronóstico y la gestión de inundaciones, la evaluación del cumplimiento de los acuerdos de uso del agua y el diseño de embalses, sistemas de suministro de agua, puentes y alcantarillas (Boiten, 2008). Los datos continuos de series de tiempo de medidores de flujo también son críticos para calibrar y/o validar modelos de aguas superficiales y subterráneas, mientras que las brechas en estos datos aumentan la incertidumbre en las predicciones de un modelo. El nivel de la corriente normalmente se mide con sensores flotantes, de presión, ópticos y acústicos (Turnipseed y Sauer, 2010). Estos sensores tradicionales pueden fallar y requerir un mantenimiento regular, los cuales son costosos. Como resultado, es posible que se produzcan lagunas en los registros de flujo y descarga debido a una instalación incorrecta (p. ej., especialmente durante estudios a corto plazo, cuando las características del sitio no son bien conocidas), fallas en los equipos y/o brechas en el financiamiento de los programas de monitoreo. Por supuesto, las cámaras también pueden fallar, pero pueden proporcionar una redundancia económica con información que no está disponible en los sensores que emiten mediciones escalares individuales. Por ejemplo, las imágenes proporcionan una verificación visual de las condiciones hidrológicas, incluida la presencia de hielo, obstrucciones o cambios importantes en la geometría del canal. En este estudio, nos enfocamos en un enfoque de monitoreo pasivo que utiliza imágenes de lapso de tiempo que se pueden combinar con mediciones de sensores tradicionales que son adecuadas para llenar los vacíos en los registros de caudales.

2 Tratamiento Inicial de los Datos

Para comenzar a trabajar con los datos, es necesario que pasen por un proceso de preparación que nos permita obtener la mejor parte de ellos. Este proceso se divide en tres etapas: Limpieza, Transformación y Visualización. A continuación desglosaremos las fases involucradas a este proceso:

2.1 Limpieza de los datos

La limpieza es la primera y una etapa fundamental del tratamiento de la información. Aquí se busca eliminar la mayor cantidad de imperfecciones que pudieramos llegar a encontrar. Cosas como valores faltantes, datos fuera de rango, dividir la información disponible en "entrenamiento" y "pruebas", eliminar columnas innecesarias para el análisis, etc.

SensorTime	CaptureTime	Filename
Agency	SiteNumber	TimeZone
Stage	Discharge	CalcTimestamp
width	height	exposure
fNumber	isoSpeed	shutterSpeed
grayMean	graySigma	entropyMean
entropySigma	hMean	hSigma
sMean	sSigma	vMean
vSigma	areaFeatCount	grayMean 0
graySigma 0	entropyMean 0	entropySigma 0
hMean 0	hSigma 0	sMean 0
sSigma 0	vMean 0	vSigma 0
grayMean 1	graySigma 1	entropyMean 1
entropySigma 1	hMean 1	hSigma 1
sMean 1	sSigma 1	vMean 1
vSigma 1	WeirAngle	WeirPt1X
WeirPt1Y	WeirPt2X	WeirPt2Y
WwRawLineMin	WwRawLineMax	WwRawLineMean
WwRawLineSigma	WwCurveLineMin	WwCurveLineMax
WwCurveLineMean	WwCurveLineSigma	

Clases disponibles en nuestro archivo de datos.

2.2 Transformación de los datos

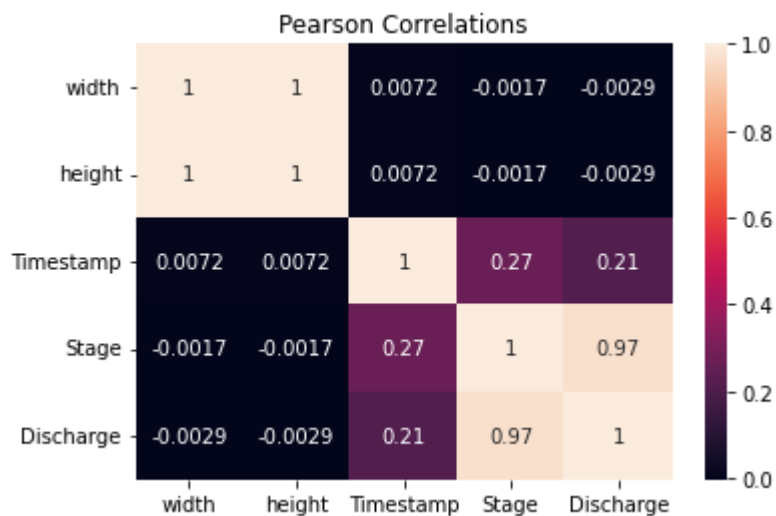
Una vez revisados los datos, entendemos que debemos transformarlos para generar gráficos y diagramas que faciliten la búsqueda de patrones y la determinación de si será necesario un modelo predictor de regresión o de clasificación.

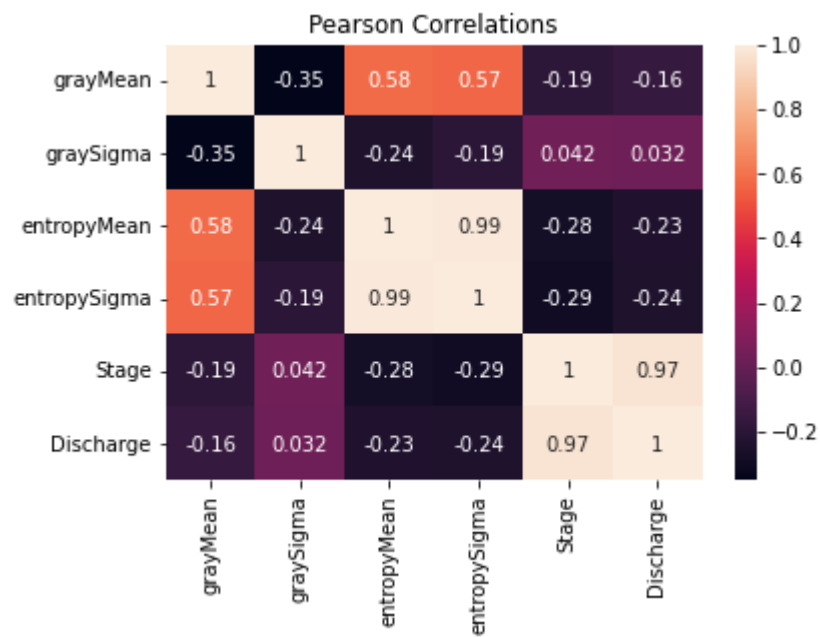
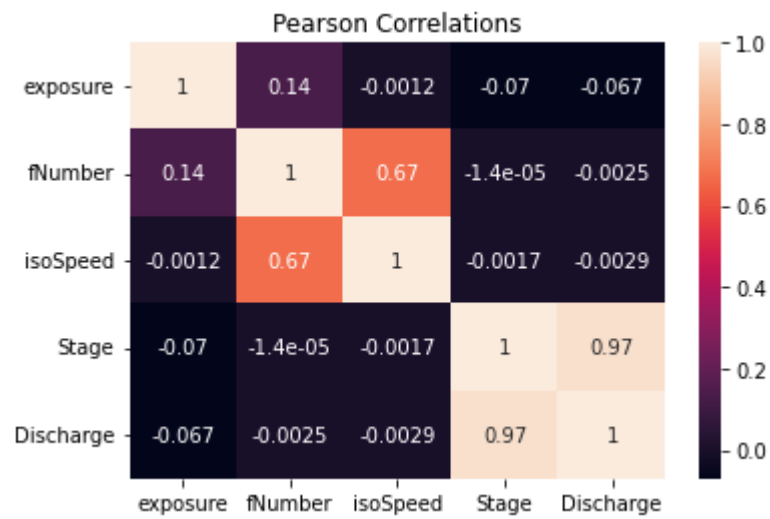
Para esto, revisamos el dataframe con la información proporcionada. Encontramos que primero debemos ordenar los datos con base en el timestamp. Esta variables que indican el tiempo con el que fueron realizadas las capturas

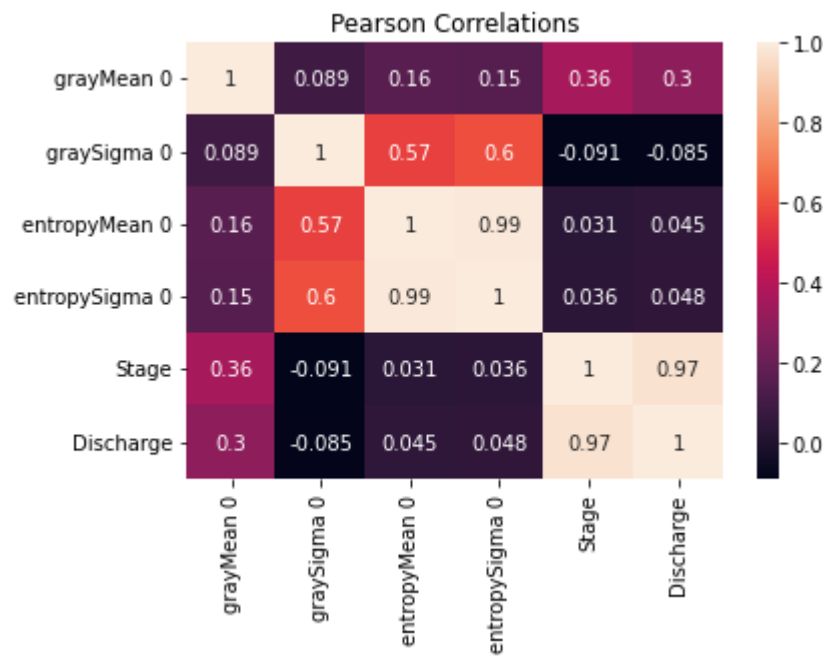
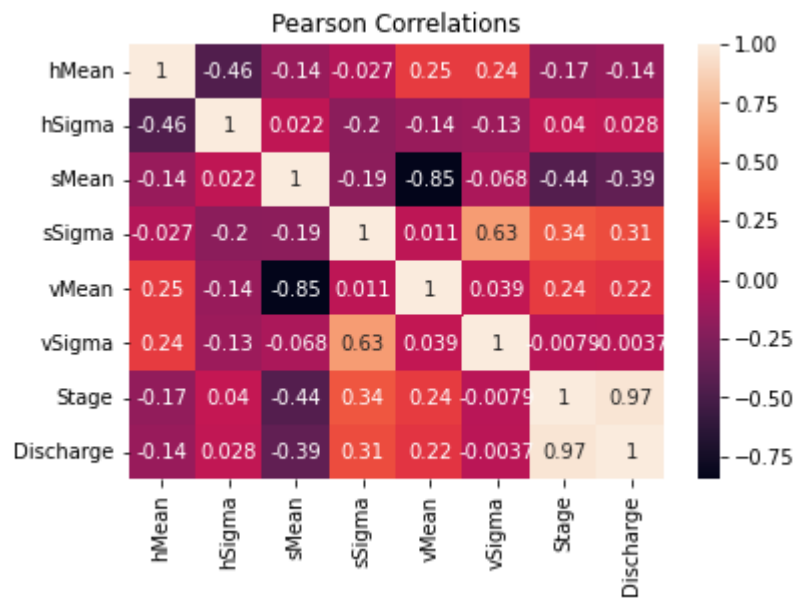
de datos. Dentro del dataframe se proporcionan tres variables para indicar el tiempo de captura: "CaptureTime", "SensorTime" y "CalcTimestamp", y cada una de estas contiene un formato distinto y a primera instancia nada fácil de trabajar, lo que indica que tenemos que separarlas para que queden enumeradas de una en una. Desarrollamos un método para tratar las fechas consiguiendo así, un orden en el formato de la captura de datos. En cuanto a lo demás, los datos venían en buenas condiciones, por lo que no tuvimos que arreglar valores nulos.

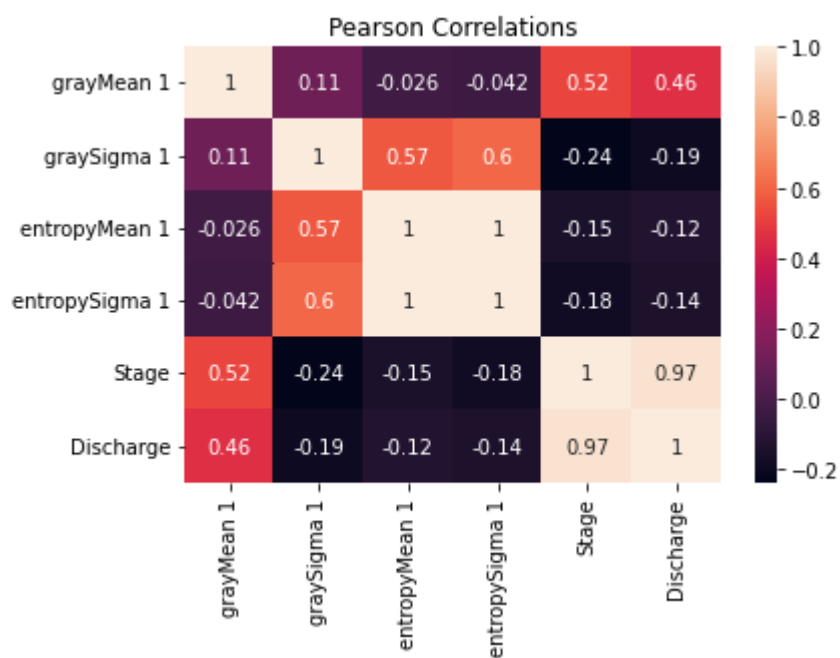
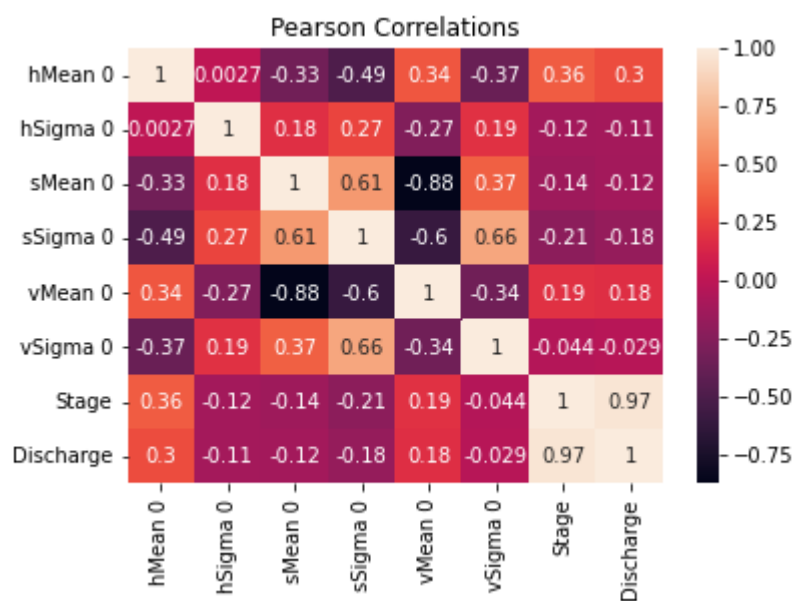
2.3 Visualización de los datos

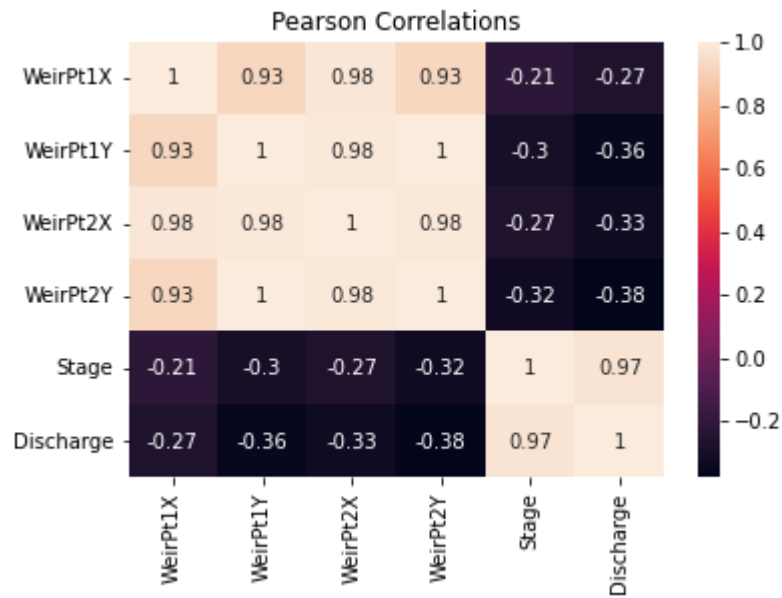
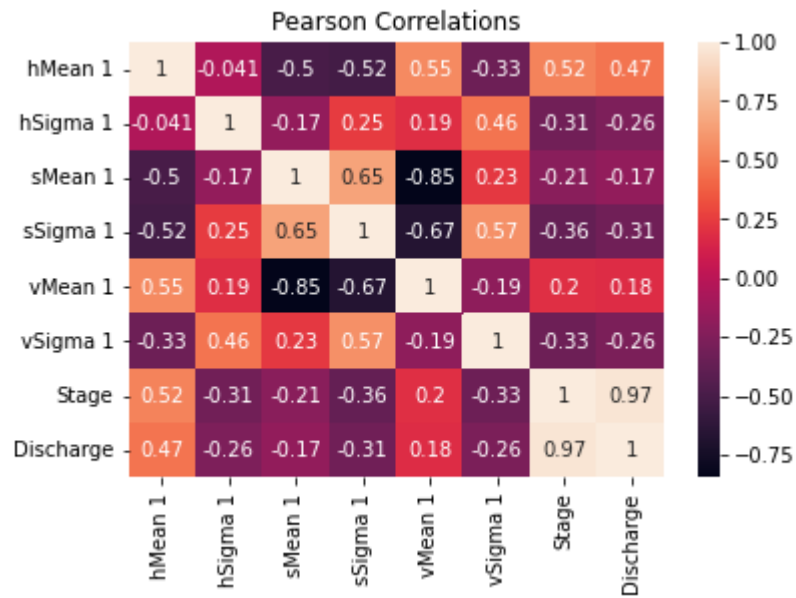
Siendo las variables "Stage" y "Discharge" las que nos interesan predecir con un modelo, las comparamos con el resto de las variables para ir descartando las de menor correlación.

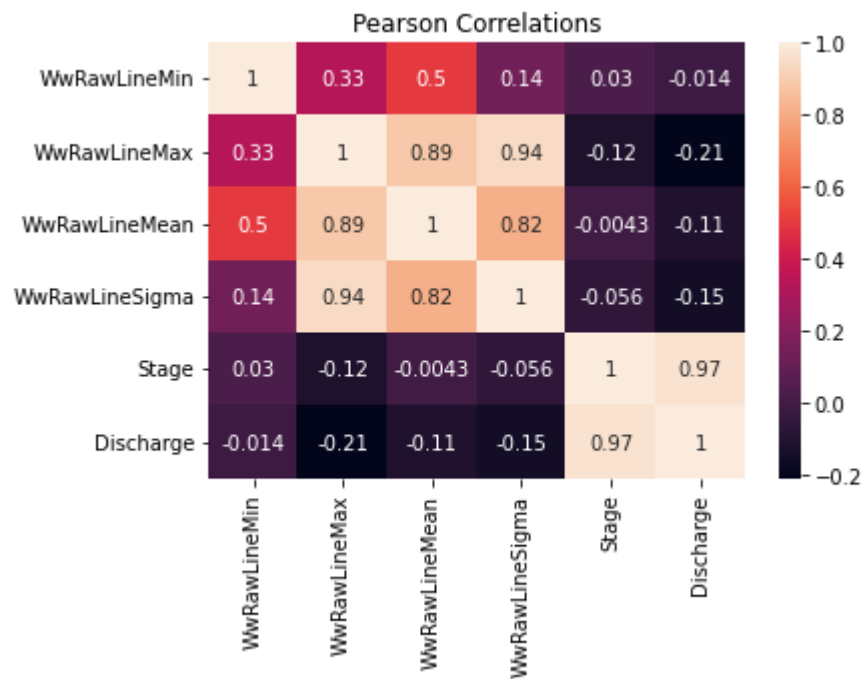








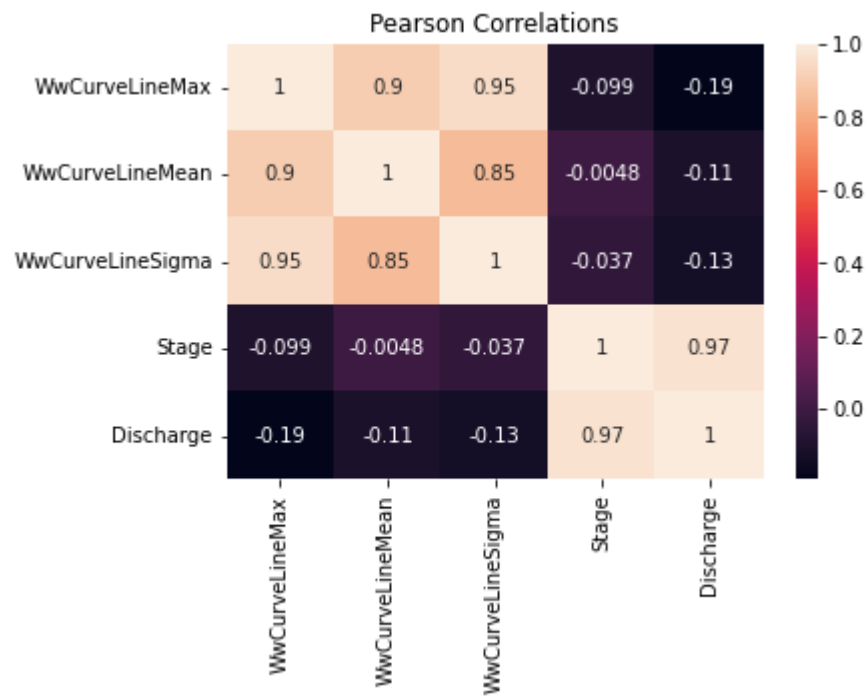




Tras analizar los resultados optamos por las siguientes variables:

- sSigma
- grayMean 0
- hMean 0
- grayMean 1
- hMean 1
- sMean
- sSigma 1
- WeirPt1Y
- WeirPt2X
- WeirPt2Y

Podemos observar sus correlaciones en este último gráfico



3 Desarrollo del modelo con los datos